

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique
Département Hydraulique



Mémoire de projet de fin d'études
pour l'obtention du diplôme d'ingénieur d'état en hydraulique

**Modélisation du Transport Solide par Machine
Learning (Application au Bassin Versant de la
Medjerda)**

DJELLALI.Menad_BOUNJAD.Abderraouf

Sous la direction de M. Abdelmalek BERMAD Professeur

Présenté(e) et soutenue publiquement le (08/07/2020)

Composition du jury :

Président	Mr. El Hadi BENYOUSSEF	Professeur ENP
Promoteur	Mr. Abdelmalek BERMAD	Professeur ENP
Examineur	Mr. Rafik OULEBSIR	Doctorant USTHB
Examineur	Mr. Abdelouahab LEFKIR	MCA ENSTP

ENP 2020

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique
Département Hydraulique



Mémoire de projet de fin d'études
pour l'obtention du diplôme d'ingénieur d'état en hydraulique

**Modélisation du Transport Solide par Machine
Learning (Application au Bassin Versant de la
Medjerda)**

DJELLALI.Menad_BOUNJAD.Abderraouf

Sous la direction de M. Abdelmalek BERMAD Professeur

Présenté(e) et soutenue publiquement le (08/07/2020)

Composition du jury :

Président	Mr. El Hadi BENYOUSSEF	Professeur ENP
Promoteur	Mr. Abdelmalek BERMAD	Professeur ENP
Examineur	Mr. Rafik OULEBSIR	Doctorant USTHB
Examineur	Mr. Abdelouahab LEFKIR	MCA ENSTP

ENP 2020

Dédicace

Je dédie ce travail à :

*Ma très chère mère qui m'a toujours apporté
Son amour et son affection*

*Mon cher père, qui m'a toujours encouragé, conseillé
et
Soutenu dans mon travail*

Mes frères et mes sœurs

*Mes camarades de la promo d'hydraulique
De 2019/2020
Mes très chers amis*

Abderraouf

Dédicace

Je dédie ce modeste travail à mes chers parents.

Ma mère pour m'avoir mis au Monde et pour m'avoir accompagné tout le long de ma vie.

Mon père qui sans lui je ne serais pas arrivé jusqu'ici. J'espère toujours Rester fidèle aux valeurs morales qu'il m'a apprises.

Ma grande sœur et mon petit frère et toute ma famille.

Mes très chers amis et Mes camarades de la promo d'hydraulique 2019/2020.

Menad

Remerciement

En tout premier lieu, nous remercions dieu le tout puissant et miséricordieux, qui nous aide et nous donne la force, la patience et le courage durant ces longues années d'étude, d'accomplir ce modeste travail.

*Au terme de cette modeste étude, nous tenons à exprimer notre profonde reconnaissance et nos chaleureux remerciements à notre promoteur Monsieur **A.BERMAD** qui nous a guidé et orienté tout au long de la réalisation de ce travail en prodiguant ses précieux conseils et ses vifs encouragements. Nous le remercions vivement de sa disponibilité de tous les instants qui ont rendu ce travail possible.*

*Nous tenons à remercier Monsieur **R. OULEBSIR** pour le soutien, l'aide et les conseils qu'il nous a dispensé pour l'élaboration de ce présent mémoire.*

Nous remercions également le président et les membres du jury qui nous feront l'honneur de juger notre travail.

A tous les enseignants du département d'hydraulique qui ont contribué à notre formation.

Enfin, nous voudrions ensuite remercier tous nos proches et amis qui ont contribué de près ou de loin à ce mémoire (Bouchra BOUFARH...etc.), et tous les amis de notre classe.

ملخص

تتميز الجزائر بمناخ شبه قاري وهذا ما جعلها تكون عرضة لمشكل التآكل والرواسب والذي يحدث على مستوى الأحواض المائية ويهدد سلامة السدود ووظيفتها. في هذه الدراسة قمنا في المقام الأول بشرح مفصل لمشكل التآكل ونقل الرواسب وتأثيرها السلبي على انجراف التربة وزيادة نسبة ترسب المواد الصلبة في الطبقة السفلى في السدود والحواجز المائية (الطمي). أدى التطور الكبير الحاصل في السنوات الأخيرة في مجال النمذجة الهيدرولوجية إلى خلق مناهج جديدة لتقدير حجم التآكل وكمية المواد الصلبة المتحركة وهذا من خلال تقدير تركيز المواد الصلبة بدلالة التدفق المائي. النمذجة التي قمنا بها في هذا العمل تمت باستعمال نموذجين قائمين على استخدام الذكاء الاصطناعي، حيث قمنا بتطبيقهما على منطقتين ثم قمنا بمقارنة النتائج لاختيار النموذج الأفضل. وبعد تقييم النتائج تبين أن النموذج الأكثر كفاءة هو نموذج الشبكات العصبية الاصطناعية في حالة وجود عدد كبير من البيانات الموثوقة، أما النموذج الثاني يوفر مرونة أفضل في حالة نقص عدد البيانات.

الكلمات المفتاحية: انجراف التربة، التآكل، نقل الرواسب، الذكاء الاصطناعي، الشبكات العصبية الاصطناعية.

Abstract

In Algeria, soil erosion in watersheds is a major problem, which increased soils the function of dams. In our study, we did first a literature review of erosion and sediment transport and their impact on the silting of dams especially in Algeria. The progress of hydrological modeling field in recent years has provided new approaches for the quantification of erosion and sediment transport by estimating suspended matter concentration-liquid flow at the different basins slopes. The flow-MES model is made by two artificial intelligence methods, neural networks and gradient boosting. The application on the two watersheds (Medjerda and Colorado) and the evaluation of the performance of these two models showed that the most efficient model is that of artificial neural networks in the event of a large number of reliable data, in the case on the contrary, the Gradient Boosting offers better flexibility.

Key words: erosion, sediment transport, artificial intelligence, artificial neural networks, gradient boosting, Medjerda, Colorado.

Résumé

Le contrôle de l'envasement des retenues des barrages permet d'évaluer de façon globale l'importance de processus de l'érosion et de transport solide.

Dans notre mémoire, on a fait dans un premier lieu une recherche bibliographique concernant le problème majeur de l'érosion et de transport solide et leur impact sur l'envasement des barrages notamment en Algérie. L'évolution reconnue de domaine de la modélisation hydrologique dans les dernières années a donné des nouvelles approches pour la quantification de l'érosion et de transport solide par l'estimation des concentrations en matière en suspension en fonction de débit liquide au niveau des différents bassins versants.

La modélisation débit-MES a été faite par deux méthodes d'intelligence artificiels, les réseaux de neurones artificiels et le Gradient Boosting. L'application sur les deux bassins versant (Medjerda et Colorado) et l'évaluation des performances de ces deux modèles a montré que le modèle le plus performant est celui de réseaux de neurones artificiels en cas de nombre important de données fiable, dans le cas contraire le Gradient Boosting présente une meilleure souplesse.

Mots clés : érosion, transport solide, intelligence artificielle, réseaux de neurones artificiels, Gradient Boosting, Medjerda, Colorado.

SOMMAIRE

Liste des tableaux

Liste des figures

Introduction générale	13
Chapitre I :	15
Généralités sur l'érosion Hydrique et le Transport Solide	15
I.1 Introduction	16
I.2 Définition de l'érosion	16
I.3 L'érosion hydrique	16
I.3.1 Définition	16
I.3.2 Origine et mécanismes	16
I.3.3 Les formes de l'érosion hydrique	22
I.4 Les facteurs induisant l'érosion hydrique	24
I.4.1 La lithologie (nature de sols)	24
I.4.2 Le climat	25
I.4.3 Le couvert végétal	25
I.4.4 Facteurs topographiques	25
I.4.5 Occupation des terres	26
I.5 Dégâts causée par l'érosion	26
I.5.1 Sur site	26
I.5.2 Hors site	28
I.6 Le transport solide en Algérie	30
I.7 Différents modes de transport solide	31
I.7.1 Transport par charriage	31
I.7.2 Transport par saltation	32
I.7.3 Transport par suspension	32
I.8 Mesure du transport solide	33
I.8.1 Mesure de charriage	33
I.8.2 Mesure des suspensions	35
I.8.3 Méthodes de mesure en Algérie	38
I.9 Conclusion	39
Chapitre II :	40

Présentation de la zone d'étude	40
II.1 Situation géographique du bassin versant.....	41
II.1.1 Présentation du bassin versant de la Medjerda	41
II.1.2 Présentation du sous bassin d'Oued Mellègue.....	43
II.2 Caractéristiques morphométriques	44
III.3 Relief de la zone d'étude	44
II.4 Le réseau hydrographique	45
II.5 La géologie du secteur d'étude.....	48
II.6 Le climat	48
II.6.1 L'évaporation	49
II.6.2 La pluviométrie	50
II.6.3 Autres informations sur le climat.....	51
II.7 Présentation des données hydrométriques.....	52
II.7.1 Paramètres statistiques des données	53
II.7.2 Les données utilisées.....	54
II.8 Conclusion	54
Chapitre III :	55
Méthodes de Machines Learning.....	55
III.1 Introduction	56
III.2 Les Réseaux de Neurones Artificiels	56
III.2.1 Définition.....	56
III.2.2 Types des réseaux de neurones artificiels	56
III.2.3 Structure d'un réseau de neurones artificiel	58
III.2.4 Fonctions d'activation des RNA	61
III.2.5 Modes d'apprentissage des réseaux de neurones artificiels.....	62
III.2.6 Sélection de la structure du réseau	64
III.2.7 Sélection de l'algorithme d'apprentissage	64
III.3 Arbres de Décision	65
III.3.1 Fonctionnement des Arbres de Décision	65
III.3.2 Random Forest.....	68
III.3.3 Gradient Boosting	69
Chapitre IV :	71
Résultat et discussions.....	71

IV.1 Introduction	72
IV.2 Critères d'évaluation d'un modèle	72
IV.2.1 Critères graphiques	72
IV.2.2 Critères numériques	72
IV.3 Critique des données	75
IV.3.1 Traitement statistique des données	78
IV.4 Application des modèles Machine Learning au bassin versant de la Medjerda	78
IV.4.1 Réseaux de Neurones Artificiels	79
IV.4.2 Gradient Boosting	83
IV.4.3 Interprétation des résultats	86
IV.5 Bassin de Colorado :	86
On a recueilli les données du site américain USGS :	
https://waterdata.usgs.gov/nwis/inventory/?site_no=09251000&agency_cd=USGS	86
Voici le nom de la rivière et le code de la station de mesures :	87
USGS 09251000 YAMPA RIVER NEAR MAYBELL, CO	87
IV.5.1 Application des modèles Machine Learning au bassin versant du Colorado	88
IV.6 Comparaison et conclusion	95
Conclusion générale	98
REFERENCES BIBLIOGRAPHIQUES	100

LISTE DES TABLEAUX

Tableau II.1: Les caractéristiques morphométriques du bassin versant d'Oued Mellègue amont (Stamboul,2017)(Gouaidia, Laouar, Défaflia & Zenati, 2018).....	45
Tableau II.2 : Répartition mensuelle de l'évaporation (ANRH).....	50
Tableau II.3 : Répartition mensuelle de la pluie annuelle (ANRH).....	51
Tableau II.4 : Paramètres climatiques de la station de Tébessa (Khedidji, 2012).	53
Tableau IV.1 : Les critères d'évaluation pour le RNA de Medjerda.	79
Tableau IV.2 : Les critères d'évaluation pour le Gradient Boosting Medjerda.	82
Tableau IV.3 : Les critères d'évaluation pour le RNA Colorado.....	87
Tableau IV.4 : Les critères d'évaluation pour le GB Colorado.....	91

LISTE DES FIGURES

Figure I.1 : Les trois étapes par lesquelles passe l'érosion	18
Figure I.2 : Schématisation de la dégradation par compression	19
Figure I.3: Schématisation de la désagrégation Splash	20
Figure I.4 : Schématisation des forces de la friction et de la résistance au cisaillement	22
Figure I.5 : Différents modes de transport des particules	23
Figure I.6: Érosion en nappe	24
Figure I.7 : Ravins profonds	25
Figure I.8 : Terrain avec rigoles Tunis Tunisie	28
Figure I.9 : Érosion en ravin Agadir Maroc	29
Figure I.10 : Sédiments pénétrant dans une voie d'eau Kairouan Tunisie.	30
Figure I.11 : Inondation boueuse Tébessa Algérie	30
Figure I.12 : Envasement de la retenue d'eau Maroc	31
Figure I.13 : Modes du transport solide.....	33
Figure I.14 : Marquage des sédiments	34
Figure I.15 : Plage de dépôt	35
Figure I.16 : Schéma de principe de la mesure par acoustique passive	36
Figure I.17 : Bouteille de mesure de la suspension	38
Figure I.18 : Installation du turbidimètre dans le cours d'eau	39
Figure II.1: Les principaux cours d'eau du bassin versant de Medjerda	42
Figure II.2 : Situation géographique du bassin versant Medjerda	43
Figure II.3 : Situation géographique du bassin d'Oued Mellègue	44
Figure II.4 : Topographie de la région d'étude	46
Figure II.5 : Photo satellite montrant les différents Oueds de la région d'étude.....	47
Figure II.6 : Stations pluviométriques et hydrométriques	48
Figure II.7 : Répartition moyenne mensuelle de l'évaporation.....	50
Figure II.8 : Répartition mensuelle de la précipitation.....	51
Figure II.9 : Précipitations moyenne annuelles de la station de El Aouinet pour la période	52
Figure II.10 : Caractéristiques statistiques du débit liquide.....	54
Figure II.11 : Caractéristiques statistiques de la concentration.....	54
Figure III.1 : Structure d'un neurone simple.....	59
Figure III.2 : Réseau de neurones monocouche	60
Figure III.3 : Réseau de neurones multicouches	61
Figure III.4 : Fonction d'activation binaire à seuil	62
Figure III.5 : Fonction d'activation à rampe avec saturation	62
Figure III.6 : Fonction sigmoïde	62
Figure III.7 : Principe de classification des arbres de décision.....	66
Figure III.8 : Principe de fonctionnement des forêts aléatoires	69
Figure IV.1 : Evolution du débit liquide observé du bassin de la Medjerda.....	76
Figure IV.2 : Evolution de la concentration observée du bassin de la Medjerda.....	76
Figure IV.3 : Evolution de la hauteur observée du bassin de la Medjerda.....	77

Figure IV.4 : Evolution de débit solide du bassin de la Medjerda.	77
Figure IV.5 : Evolution du débit solide (mesurés simulés) en période de calage (Medjerda, RNA).....	80
Figure IV.6 : Evolution du débit solide (mesurés et simulés) en période de validation (Medjerda, RNA).	80
Figure IV.7 : Représentation de corrélation simulés-mesurés en période de calage (Medjerda, RNA).....	81
Figure IV.8 : Représentation de corrélation simulés-mesurés en période de validation (Medjerda, RNA).....	81
Figure IV.9 : Evolution du débit solide (mesurés et simulés) en période de calage (Medjerda, Gradient Boosting).....	83
Figure IV.10 : Evolution du débit solide (mesurés et simulés) en période de validation (Medjerda, GB).....	83
Figure IV.11 : Représentation de corrélation simulés-mesurés en période de calage (Medjerda, GB).....	84
Figure IV.12 : Représentation de corrélation simulés-mesurés en période de validation (Medjerda, GB).....	85
Figure IV.13 : Evolution du débit solide observé (en Kg/s) du bassin de Colorado.	86
Figure IV.14 : Evolution du débit liquide observé (en m ³ /s) du bassin de Colorado.	87
Figure IV.15 : Evolution du débit solide (mesurés et simulés) en période de calage (Colorado, RNA).....	88
Figure IV.16 : Evolution du débit solide (mesurés et simulés) en période de validation (Colorado, RNA).....	89
Figure IV.17 : Evolution du débit solide (mesurés et simulés) en période de calage (Colorado, RNA).....	90
Figure IV.18 : Représentation de corrélation simulés-mesurés en période de validation (Colorado, RNA).....	90
Figure IV.19 : Evolution du débit solide (mesurés et simulés) en période de calage (Colorado, GB).....	92
Figure IV.20 : Evolution du débit solide (mesurés et simulés) en période de validation (Colorado, GB).	92
Figure IV.21 : Représentation de corrélation simulés-mesurés en période de calage (Colorado, GB).....	93
Figure IV.22 : Représentation de corrélation simulés-mesurés en période de validation (Colorado, GB).	94

Introduction générale

Dans le contexte actuel des changements climatiques, le suivi du transport sédimentaire dans les rivières est important pour mettre l'accent sur les problèmes environnementaux souvent reliés à des changements observés dans le cycle hydrologique. Le suivi du transport sédimentaire permet également d'évaluer et de quantifier les transferts de matières des continents vers les mers et plus particulièrement les barrages. Ces derniers qui représentent des ouvrages importants pour le stockage des eaux superficielles arrivent à peine à satisfaire les besoins de la population surtout en zone aride et semi-aride, tel que l'Algérie où l'eau est au cœur des préoccupations de la société. A défaut d'emmagasiner cette source de vie, ces réservoirs se trouvent vite envahis de vase. L'estimation des volumes de la vase décantée au fond des barrages et la durée de vie de ces ouvrages nécessitent une bonne connaissance des apports solides. Ces derniers dépendent aussi bien des conditions d'écoulement au niveau des cours d'eau que de la nature et du volume des apports liquides générés par le bassin versant. Le phénomène s'est accru et s'est amplifié aussi bien dans l'espace que dans le temps, aggravé en maints endroits par le changement climatique.

L'Algérie est classée parmi les pays aux sols les plus érodables du monde, ces volumes immenses de terre arrachée chaque année sont la cause principale de l'envasement des barrages en Algérie, d'où la nécessité de bien quantifier ces volumes afin de prévenir ce phénomène, et c'est l'objectif principal de notre travail.

Le problème du nombre réduit de données relatives aux transports solides a toujours freiné les investigations. Les mesures de transport solide effectuées en Algérie, et les études ponctuelles détaillées restent insuffisantes pour traiter ce sujet. Le défi consiste donc à utiliser des méthodes innovantes pour la modélisation de ce phénomène.

Ce travail s'intéresse au bassin de la Medjerda en Algérie, afin d'étudier l'évolution du transport solide à son exutoire. Après une introduction générale, consacrée à la thématique étudiée à savoir le transport solide et à la problématique traitée, ce mémoire de fin d'études a été scindé en quatre chapitres :

Le premier, porte sur les phénomènes de l'érosion et du transport solide, causes, origines et conséquences.

INTRODUCTION GENERALE

Le chapitre deux, s'intéresse à la zone d'étude, on y trouve les caractéristiques hydrologiques, climatologiques et géologiques du bassin versant de la Medjerda, ainsi que la présentation et la critique des données recueillies.

Les définitions et présentation théorique des réseaux de neurones artificiels et du Gradient Boosting, utilisés comme techniques de modélisation du transport solide, ont fait l'objet du chapitre III.

Le dernier chapitre, a été réservé à la discussion des résultats obtenus par les deux modèles élaborés et à la comparaison entre ces deux modèles.

Enfin, une conclusion générale résume, tout le travail réalisé lors de ce mémoire, et plus particulièrement, les résultats obtenus à partir des deux modèles élaborés avec une comparaison et des perspectives pour la suite à donner à ce mémoire de fin d'études.

Chapitre I :
**Généralités sur l'érosion Hydrique et
le Transport Solide**

I.1 Introduction

Dans le domaine de la pédologie et de l'écologie, la régression et la dégradation sont des processus d'évolution associés à une perte d'équilibre d'un sol antérieurement stable. Ce type d'érosion commence généralement avec la destruction du couvert végétal.

La matière première du débit solide est l'érosion continentale, c'est-à-dire la désagrégation, sous toutes ses formes, du sol par les agents atmosphériques : le gel qui fait éclater les roches par congélation de l'eau absorbée soit en petit dans la structure de la roche, soit en grand dans les fissures, l'impact des gouttes de pluie, érosion Fluviale surtout active sur les sols meubles, le vent, les variations brutales de température, les actions chimiques de l'eau et le ruissellement (Roche, 1990).

I.2 Définition de l'érosion

C'est un phénomène spatio-temporel, qui est défini par le détachement et le transport des particules sous l'effet de la pluie, lorsque le sol n'est plus capable d'infiltrer l'eau. Cette situation se produit généralement sur des sols préalablement fragilisés, dans le cas d'une intensité de pluie supérieure aux capacités d'infiltration du sol (lors d'orages violents notamment), ou sur des sols gorgés d'eau (en périodes automnale et hivernale) (Zaabar, 2016).

I.3 L'érosion hydrique

I.3.1 Définition

L'érosion hydrique est le résultat d'un ensemble de processus complexes et interdépendants qui provoquent le détachement et le transport des particules du sol. Elle se définit comme la perte de sol due à l'eau qui arrache et transporte la terre vers un lieu de dépôt (Zaabar, 2016).

I.3.2 Origine et mécanismes

L'érosion hydrique est un phénomène complexe, qui menace particulièrement les potentialités en eau et en sol. Elle se définit comme le détachement et le transport de particules de sol de son emplacement d'origine par différents agents vers un lieu de dépôt. Donc, les trois étapes par lesquelles passe l'érosion sont le détachement, le transport et la sédimentation. Cependant, il est à signaler que la pluie et le ruissellement superficiel sont à l'origine du détachement, du transport

et du dépôt des particules du sol arraché comme schématisé dans la figure suivante : (HAMDI, 2016).

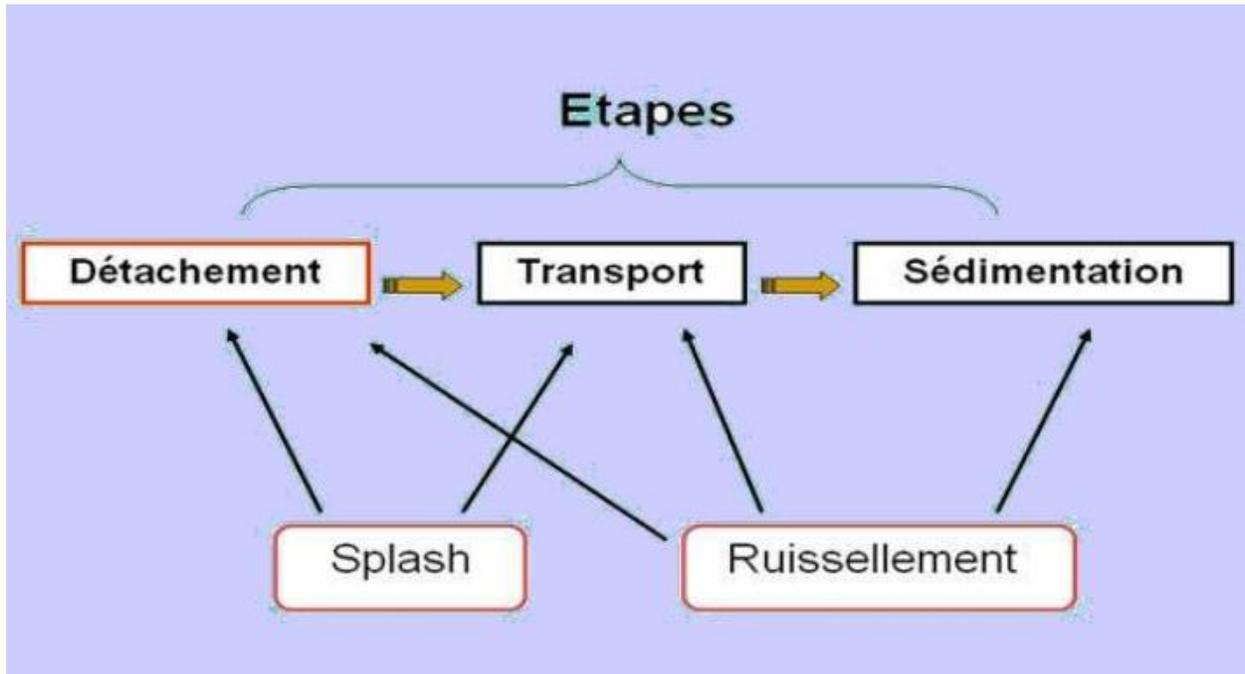


Figure I.1 : Les trois étapes par lesquelles passe l'érosion (HAMDI, 2016).

I.3.2.1 Le détachement Origine et mécanisme

Les principaux mécanismes conduisant au détachement sont :

L'humectation par l'effet de l'impact de gouttes de pluies

Les quatre processus qui peuvent être identifiés comme responsables de la désagrégation sont :

a) L'éclatement

Correspondant à la désagrégation par compression de l'air piégé lors de l'humectation. L'intensité de l'éclatement dépend entre autres, du volume d'air piégé, donc de la teneur en eau initiale des agrégats et de leur porosité. (HAMDI, 2016).

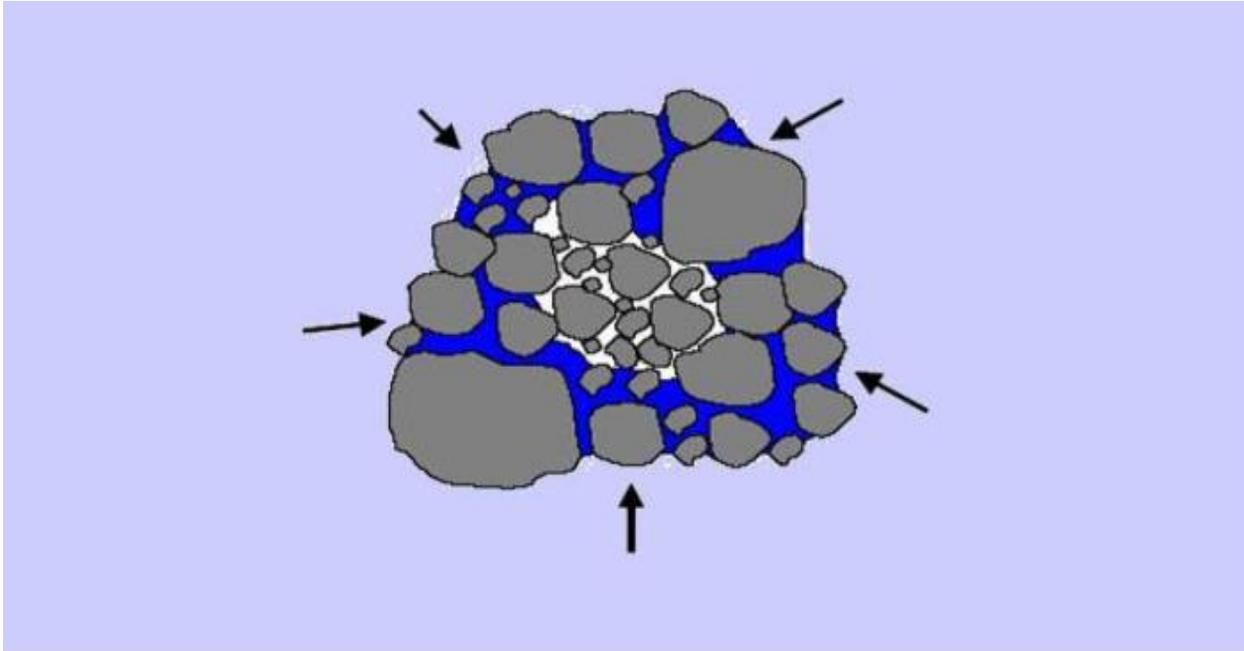


Figure I.2 : Schématisation de la dégradation par compression (HAMDI, 2016).

b) Le gonflement différentiel

Ce phénomène intervient suite à l'humectation et la dessiccation des argiles, entraînant des fissurations dans les agrégats. L'importance de ce mécanisme dépend en grande partie de la teneur et de la nature de l'argile des sols (HAMDI, 2016).

c) La dispersion physico-chimique

Elle correspond à la réduction des forces d'attraction entre particules colloïdales lors de l'humectation. Elle dépend de la taille et la valence des cations (particulièrement du sodium) pouvant lier les charges négatives dans le sol (HAMDI, 2016).

d) La désagrégation mécanique sous l'impact des gouttes de pluie

L'impact des gouttes de pluie peut fragmenter les agrégats et surtout détacher les particules de leur surface. Ce mécanisme intervient en général conjointement aux autres mécanismes cités précédemment et nécessite une pluie d'une certaine énergie qui est variable selon les sols. L'énergie cinétique des gouttes n'est plus absorbée mais est transformée en force de cisaillement qui provoque détachement et Splash (HAMDI, 2016).

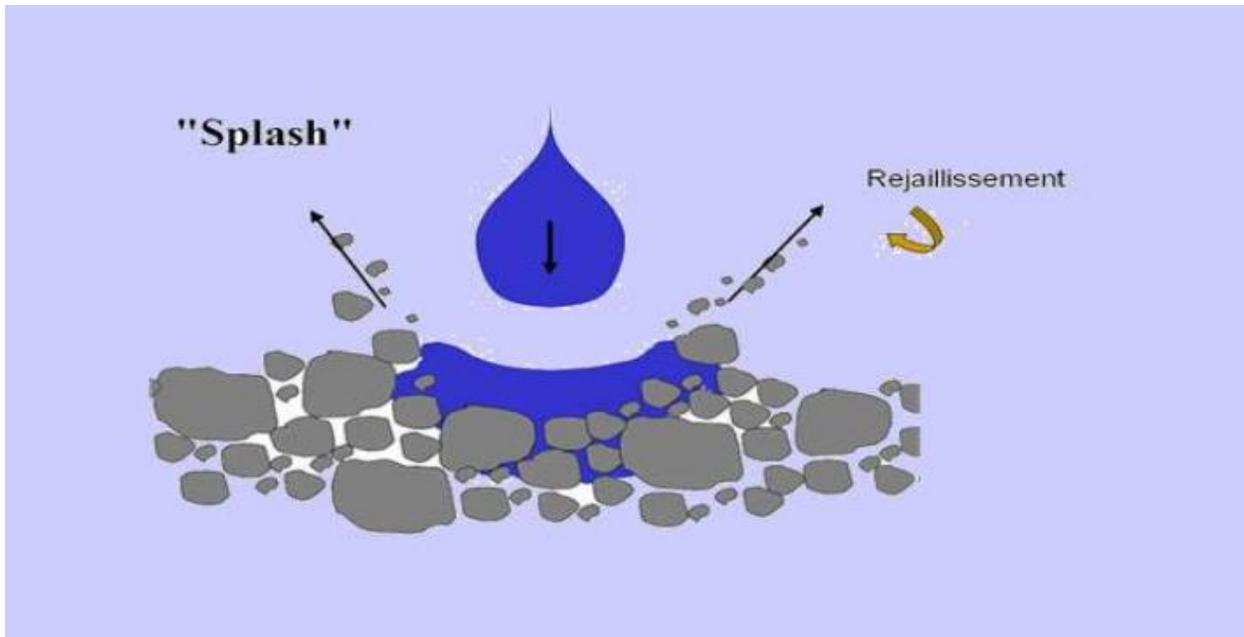


Figure I.3: Schématisation de la désagrégation Splash (HAMDI, 2016).

Cette action combinée de détachement, et déplacement par Splash des gouttes d'eau, est probablement la raison pour laquelle on l'a souvent considéré comme le seul processus à l'origine de la battance, et de l'érosion. Cependant, il ne faut pas forcément assimiler, Splash et désagrégation structurale, car le Splash peut dans certains cas déplacer des agrégats sans qu'aucune désagrégation n'intervienne. Les particules détachées par les gouttes de pluie sont généralement des micro-agrégats ou des particules élémentaires inférieures à 100 millimètres (HAMDI, 2016).

La taille et l'impact des gouttes sont des facteurs importants dans ce processus de destruction et d'arrachement (éclaboussement par effet Splash) (HAMDI, 2016).

L'énergie d'une seule goutte de pluie cause une érosion par éclaboussement ou rejaillissement qui peut déplacer les particules sur quelques dizaines de centimètre, la distance dépendant de la masse des particules et de l'angle d'incidence des gouttes de pluies par rapport à la surface (HAMDI, 2016).

L'énergie cinétique des gouttes qui tombent est généralement utilisée comme paramètre pour déterminer le pouvoir érosif des pluies. Cette énergie cinétique peut être très élevée dans les régions humides ou semi-arides. Au grand Maghreb, par exemple, elle peut être deux à six fois plus importante que dans les zones tempérées (HAMDI, 2016).

Les particules de sol très fines qui sont détachées de la surface par l'impact des gouttes sont piégées entre les éléments plus grossiers et peuvent obstruer les pores de la couche supérieure du sol et réduire considérablement le taux d'infiltration (battance) (HAMDI, 2016).

Cette obstruction augmente les risques d'érosion et de ruissellement en surface. Les sols limoneux sont particulièrement touchés par ce phénomène. La croûte de battance ainsi formée s'épaissit dans les petites dépressions où l'eau stagne, permettant la sédimentation des éléments fins. La perméabilité de la surface peut descendre en dessous de 2 mm/h en période humide (HAMDI, 2016).

Le micro relief s'estompe et le sol perd toute capacité de rétention d'eau superficielle. Lorsque la croûte de battance est formée, les pluies ultérieures, même si elles sont de faible intensité, engendreront du ruissellement (HAMDI, 2016).

Le ruissellement

L'érosion des sols se développe lorsque les eaux de pluie, ne pouvant plus s'infiltrer dans le sol, ruissellent sur la parcelle en emportant les particules de terre. Ce refus du sol d'absorber les eaux en excédent apparaît soit lorsque l'intensité des pluies est supérieure à l'infiltrabilité de la surface du sol ; soit lorsque la pluie arrive sur une surface partiellement ou totalement saturée par une nappe (ruissellement par saturation) (HAMDI, 2016).

Ces deux types de ruissellement apparaissent généralement dans des milieux très différents, bien que l'on observe parfois une combinaison des deux. Une fois le ruissellement déclenché sur la parcelle, l'érosion peut prendre différentes formes qui se combinent dans le temps et dans l'espace pouvant donner naissance soit une érosion diffuse et /ou soit une érosion concentrée (HAMDI, 2016).

Il est donc à noter qu'il y a détachement par ruissellement lorsque la force de friction de l'eau sur les particules du sol est supérieure à la résistance du sol au cisaillement comme schématisé sur le graphique suivant (HAMDI, 2016) :

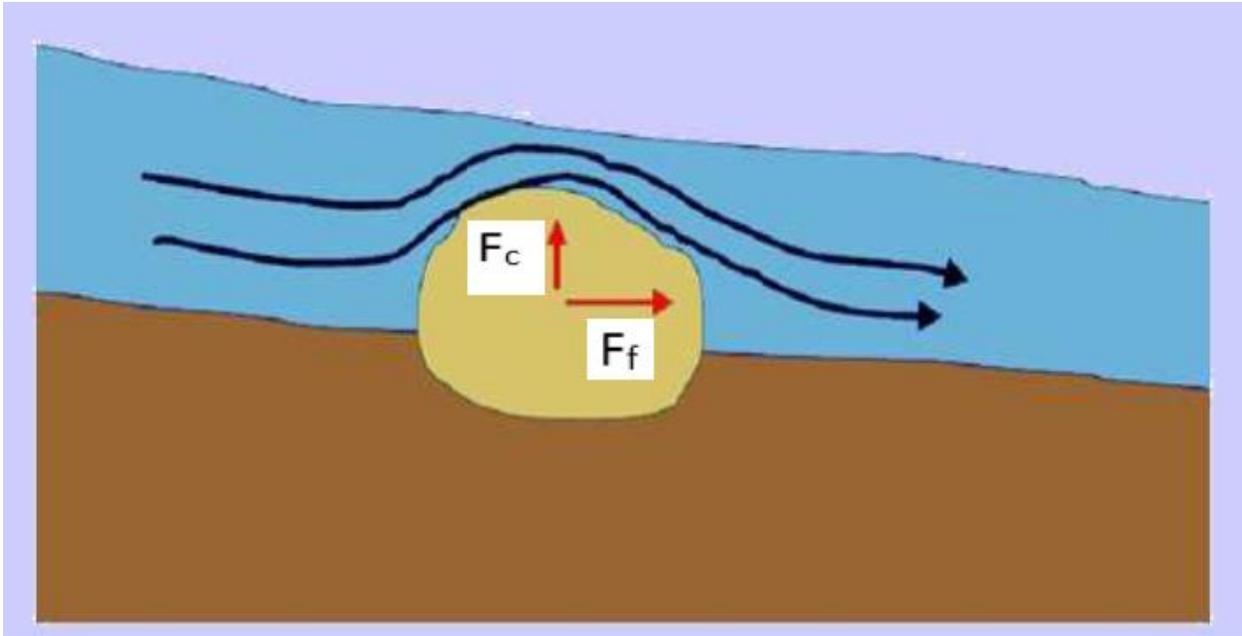


Figure I.4 : Schématisation des forces de la friction et de la résistance au cisaillement (HAMDI, 2016).

I.3.2.2 Le transport

Il est dû à la fois aux gouttes d'eau de pluie et aux eaux de ruissellement. Ainsi, le transport est assuré par ces eaux. Cependant, il est à signaler que le mode de transport par effet Splash est généralement négligeable sauf sur pente forte (HAMDI, 2016).

Alors que les eaux de ruissellement sont les plus responsables du transport des particules du sol détaché. Les modes de transport par ruissellement sont illustrés sur le graphique ci-dessous : (HAMDI, 2016).

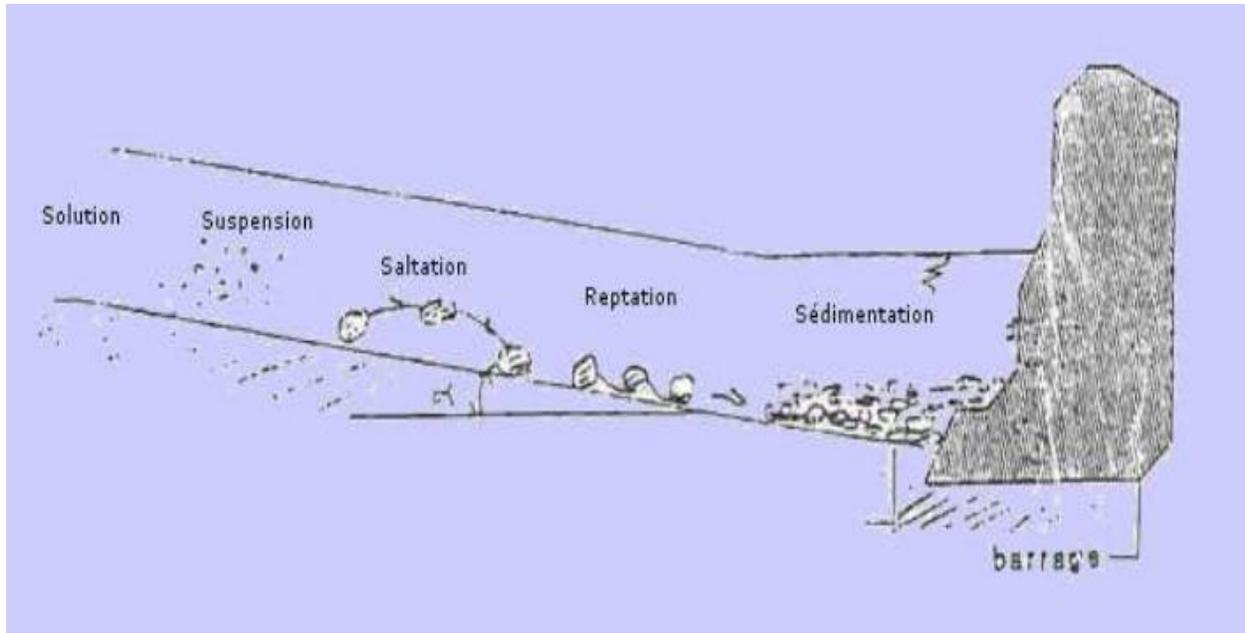


Figure I.5 : Différents modes de transport des particules (HAMDI, 2016).

I.3.2.3 La sédimentation

L'agent responsable de la sédimentation est l'eau de ruissellement. Les particules arrachées du sol se déposent entre le lieu d'origine et l'aval en fonction :

- a.** de leur dimension ;
- b.** de leur densité ;
- c.** de la capacité de transport du ruissellement ou du cours d'eau.

Les particules se déposent dans l'ordre suivant :

- a.** sable ;
- b.** sable fin ;
- c.** limon.

Les argiles et l'humus colloïdal sont généralement transportés jusqu'à l'embouchure du cours d'eau où il se dépose soit après évaporation de l'eau, soit après floculation (HAMDI, 2016).

I.3.3 Les formes de l'érosion hydrique

I.3.3.1 L'érosion en nappe

Ce type d'érosion s'effectue sur toute la surface du sol. À la surface des terres agricoles, forestières ou à pâturages, les particules de sol sont détachées par la pluie et emportées par le ruissellement. Ceci prend la forme d'érosion dite en nappe avec une mince lame d'eau s'écoulant sur une pente douce des hautes terres (Zaabar, 2016).

L'importance de l'érosion en nappe dépend à la fois :

- de l'intensité maximale des pluies qui déclenche le ruissellement,
- de l'énergie des pluies qui détache les particules susceptibles de migrer,
- de la durée des pluies et/ou de l'humidité avant les pluies.

Le transport de sédiments concerne particulièrement les particules fines : comme l'argile et le limon, les matières organiques de faible masse volumique, tels que les résidus de culture et les déjections animales ou encore les fertilisants épandus sur les parcelles cultivées (Zaabar, 2016).



Figure I.6: Érosion en nappe (Wikipédia)

Cette forme d'érosion est caractéristique des sommets de bassins versants. Le martèlement des pluies (Splash) détache les particules et les maintient en suspension par turbulence. L'érosion en nappe a un effet érosif maximal au sommet des versants ou à l'aval d'un obstacle (Zaabar, 2016).

I.3.3.2 L'érosion linéaire (micro-Channel ou Rill Erosion)

Lorsque l'intensité des pluies dépasse la capacité d'infiltration de la surface du sol, il se forme d'abord des flaques qui communiquent par des filets d'eau et lorsque ceux-ci atteignent une certaine vitesse, 25 cm par seconde d'après Hjulström (1935) (Zaabar, 2016).

Cette forme d'érosion se manifeste dès que les filets d'eau diffus se concentrent sur des lignes de plus forte pente et exercent ainsi sur le sol une force de cisaillement permettant de creuser des formes de plus en plus profondes dans le sol en arrachant des particules de plus en plus grosses: graviers, cailloux voir des blocs, de petits canaux naissent que l'on peut diviser en trois types :

griffes (quelques centimètres de profondeur), rigoles (la profondeur dépasse 10 cm), ravines (quelques mètres) (Zaabar, 2016).

I.3.3.3 L'érosion par ravinement

La ravine est une rigole approfondie (figure I.2) où se concentrent les filets d'eau. La rigole se transforme en ravine lorsque sa profondeur interdit son nivellement par des simples instruments aratoires (Zaabar, 2016).

Le ravinement constitue un stade avancé de l'érosion. Les ravines peuvent atteindre des dimensions considérables. L'approfondissement des ravines remonte du bas vers le haut de la pente (érosion régressive) (Zaabar, 2016).



Figure I.7 : Ravins profonds (comité 2011 Elbaraka)

I.4 Les facteurs induisant l'érosion hydrique

L'érosion hydrique des sols est fonction de multiples facteurs qui semblent être les mêmes partout dans le monde, selon Morgan (1986) :

I.4.1 La lithologie (nature de sols)

Les principales caractéristiques des sols qui permettent de déterminer le degré de sensibilité à l'érosion hydrique sont les suivantes : la profondeur, la piérrosité, la granulométrie, la teneur en matière organique (Roose et *al*, 1993), la nature minéralogique des argiles (Le Bissonnais et *al*, 1995), l'infiltrabilité (Papy et *al*, 1995) et la cohésion (Mériaux ,1961). Les sols méditerranéens à texture limoneuse sont parmi les plus sensibles à l'érosion hydrique (Osborn et *al*, 1976) ;

(Poesen, 1983), et cela malgré leur forte capacité de rétention en eau liée à la distribution favorable de leur porosité.

I.4.2 Le climat

Le climat est l'élément moteur de l'érosion. Les changements de température, l'intensité et la hauteur de précipitation ainsi que le vent ont des effets prépondérants sur la désintégration de la roche. Par ailleurs les variations saisonnières des orages, leur localisation, leur fréquence et le régime des vents exercent une influence sur l'érosion des sols. En Algérie sur des parcelles peu couvertes, pendant les orages d'automne le ruissellement journalier maximal a dépassé 19 à 32 % et jusqu'à 70 à 85 % des averses importantes en hiver sur des sols détrempés (Arabi et Roose, 1989).

La relation établie par Langbein et Schumm (1958) entre la hauteur de pluie efficace et l'érosion spécifique montre que l'érosion spécifique maximum (environ 290 t/km²/an) se produit en climat semi-aride avec un peu moins de 300 mm de pluie annuelle.

I.4.3 Le couvert végétal

Il est clair que le couvert végétal est efficace pour réduire le ruissellement. La couverture du sol, plus particulièrement celle assurée par la strate herbacée et les résidus végétaux, constitue la protection la plus efficace du sol. Elle intervient vis à vis du sol par :

- L'interception nette : une partie de la pluie qui tombe sur la végétation est directement reprise par l'évaporation. Elle ne participe alors, ni à l'infiltration, ni au ruissellement.
- La dissipation de l'énergie cinétique : une partie de la pluie interceptée par le feuillage continue sa course jusqu'au sol. L'énergie cinétique des gouttes s'en trouve alors modifiée.

L'expérience montre (Ruiz Figueroa, 1983) que l'efficacité d'un couvert végétal à l'égard de l'infiltration est d'autant plus marquée que le dernier élément récepteur est bas. Donc, l'amélioration du couvert végétal (densité, fertilisation, rotation avec les légumineuses, cultures fourragères) a réduit plus ou moins fortement les risques d'érosion et de ruissellement (Zaabar, 2016).

I.4.4 Facteurs topographiques

Les modifications des formes du terrain donnent un aperçu des facteurs topographiques agissant sur l'érosion. Les facteurs topographiques essentiels sont la pente du bassin versant, le relief, la densité de drainage, l'importance des vallées et des plaines d'inondation, l'orientation et

la taille du bassin versant. Sur les pentes fortes l'eau de pluie ruisselle vite et cause une érosion grave. L'augmentation de la pente développe une érosion en rigoles dix fois plus agressive que l'érosion en nappe (Roose, 1993).

Les effets de longueur de pente sur les risques d'érosion sont considérables, si bien que les eaux de ruissellement ont tendance à s'accumuler sur une pente longue, le temps d'absorption par le sol est plus grand, l'écoulement prend de la vitesse et de l'énergie et l'érosion s'intensifie (Martinez-Mena et al. 1999) ; (Megnounif, 2007).

I.4.5 Occupation des terres

L'occupation, sauf en des lieux éloignés et inhabités, est en rapport direct avec les activités humaines. Les activités qui ont la plus grande influence sur l'érosion des sols sont l'agriculture, l'élevage, les exploitations minières, l'abattage du bois, la construction de routes et l'urbanisation. Dans une zone de sols érodibles, la suppression de la végétation ou toute autre activité qui entraîne une modification de la densité du couvert végétal agira sur les phénomènes d'érosion. Les activités de construction peuvent avoir l'effet le plus important à cause à la fois de la suppression de la végétation, des modifications topographiques et de la destruction de l'équilibre du sol. Certaines modifications de l'occupation des terres peuvent ne pas modifier l'érosion ou même la réduire comme l'aplanissement d'une zone fortement ravinée (MAKKEB, 2016).

I.5 Dégâts causée par l'érosion

Parmi les signes et effets physiques visibles de la dégradation du sol causés par l'érosion peuvent être cités les éléments ci-dessous et ce, indépendamment des lieux à savoir les dégâts engendrés en amont et en aval (HAMDI, 2016) :

I.5.1 Sur site

C'est à dire en amont, on peut citer quelques exemples avec illustration par certaines photos:

- Baisse de rendement ... abandon des terres.
- Pertes en terre et en éléments nutritifs : les griffes, fines rigoles formées par l'eau, particulièrement en haut des pentes, sur le bord des pistes ou dans les champs sillonnés par les labours, elles deviennent des ravines par élargissement dû à la concentration de ruissellement excessif (HAMDI, 2016).
- Pertes d'engrais et de matière organique.
- Destruction de la structure du sol.

- Réduction de la profondeur du sol.
- L'érosion des rives entraîne non seulement le recul des rives des cours d'eau, menaçant ainsi la disparition d'habitats fauniques, mais également une augmentation de la charge particulaire des eaux du fleuve (HAMDI, 2016).

Une fois arrachées, les matières particulaires sont transportées dans les cours d'eau, parfois sur de longues distances et se déposent sur le lit pour former des sédiments jusque dans l'estuaire. De plus, si elles sont contaminées, elles contribuent à la contamination du milieu récepteur en aval (HAMDI, 2016).



Figure I.8 : Terrain avec rigoles Tunis Tunisie (Google).

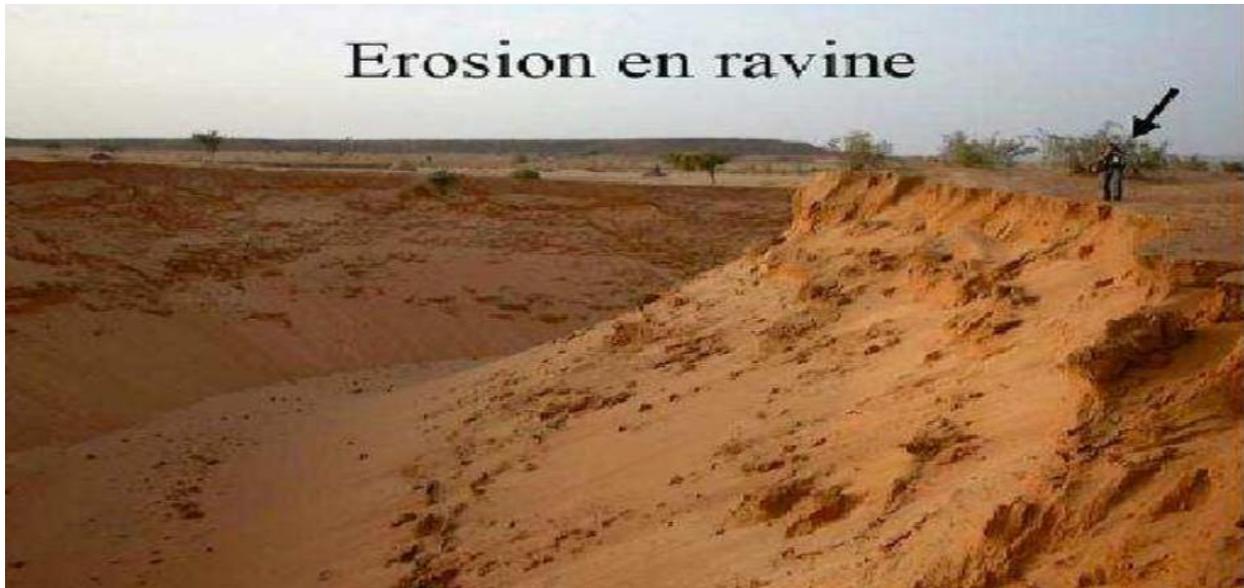


Figure I.9 : Érosion en ravin Agadir Maroc (HAMDI, 2016)

I.5.2 Hors site

À côté des dégâts bien visibles concernant les terres cultivées, il existe des dégâts en aval beaucoup plus insidieux provoqués par l'augmentation du ruissellement et l'entraînement des particules du sol (HAMDI, 2016).

Par ailleurs, pour les zones typiquement montagnardes notamment en milieu forestier, d'autres phénomènes tels que les glissements ou les éboulements de terrains de grande ampleur ou les laves torrentielles ainsi l'introduction massive de sédiments dans les cours d'eau peut être induit par l'érosion hydrique. Nous pouvons donc, illustrer certains de ces dégâts en aval à savoir : (HAMDI, 2016).

- Les inondations dues aux eaux de ruissellement par suite de la réduction de la capacité d'infiltration des sols dégradés, et la baisse de niveau ou assèchement des nappes et points d'eau à cause des pertes d'eau par ruissellement.

- Charger les rivières en M.E.S (matières en suspension). L'augmentation de la turbidité des eaux modifie l'équilibre trophique. L'entraînement des particules de sols dans les eaux superficielles s'accompagne également de celui des intrants agricoles (engrais, pesticides) et des polluants d'origine industrielle, urbaine et routière (HAMDI, 2016).

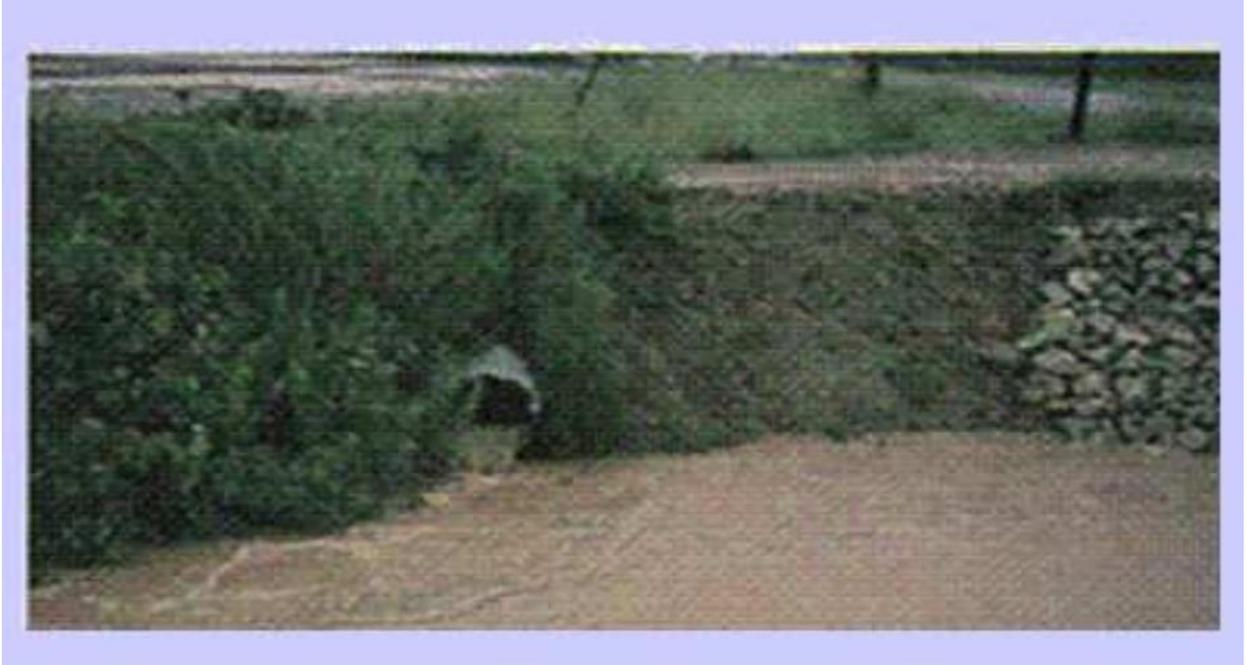


Figure I.10 : Sédiments pénétrant dans une voie d'eau Kairouan Tunisie.
(Google).

-Inondations boueuses.



Figure I.11 : Inondation boueuse Tébessa Algérie (Google).

- Eutrophisation des eaux de surface

L'apport important de sédiments dans les eaux de ruissellement a des effets biologiques et physiques néfastes sur la qualité de l'eau. Ces apports peuvent inclure des éléments azotés et

phosphatés et même des métaux lourds peuvent être également transportés. La qualité de l'eau est détériorée par eutrophisation à cause du réchauffement de la température de l'eau et l'intensification du développement d'algues et de bactéries causant le vieillissement prématuré des eaux des exutoires et, par le fait même, une perte de la biodiversité (HAMDI, 2016).

- Ensablement des lits de rivière.
- Envasement des retenues d'eau.



Figure I.12 : Envasement de la retenue d'eau Maroc (HAMDI, 2016).

- Dégâts aux infrastructures routières.

I.6 Le transport solide en Algérie

L'Algérie est située dans une région où règne un climat semi-aride à aride, ses bassins versants sont caractérisés par une pluviométrie saisonnière et interannuelle irrégulière. Cette irrégularité a un impact direct sur l'action érosive et par conséquent sur le transport solide en suspension. Ces pluies sont nettement fréquentes en automne lorsque le couvert végétal est absent et le sol ameubli par les labours, le transport solide atteint donc des valeurs maximales (MAKKEB, 2016).

L'intensité de l'érosion en Algérie varie d'un bassin à un autre et d'une région à l'autre. La partie Ouest, où l'érosion touche 47 % de l'ensemble des terres, est la région la plus érodée du pays, viennent ensuite les régions du Centre (27%) et de l'Est (26%) (Ministère de

l'environnement et l'aménagement du territoire, 2000). Le taux d'érosion spécifique atteint la valeur de 5000 t/km².an sur le bassin versant de l'Oued Agrioum (*Demmak*, 1982).

L'Algérie est classée parmi les pays aux sols les plus érodibles et cela est dû à la composition lithologique de la région, qui est principalement formée de plus de 75% des argiles et des limons. Ces deux matériaux sont les plus sensibles à l'arrachement par le ruissellement et leur transport s'effectue facilement, même pour de faibles vitesses de l'écoulement (*MAKKEB*, 2016).

L'érosion des bassins versants est très répandue dans la région du Maghreb, puisque toutes les conditions sont réunies pour déclencher et développer un tel processus : les irrégularités climatiques, la faible densité du couvert végétale, la nature des sols qui est peu résistante à l'écoulement et la violence des crues. En effet, la région enregistre les valeurs les plus élevées de la planète (*MAKKEB*, 2016).

Au début des années quatre-vingt, environ 120 millions de tonnes de sédiments arrachés annuellement au niveau des bassins versant du Nord Algérien (*Demmak*, 1982). Une partie de cette masse se dépose dans les 57 grands barrages que détenait l'Algérie. Le dépôt sédimentaire annuel dans ces barrages est évalué à 45 millions de m³ (*Remini et Hallouze*, 2007).

Le volume de la vase déposée dans les 57 grands barrages en exploitation avoisine la valeur de 1,1 milliards de m³, soit une perte de capacité de 16% de la réserve totale dans les années 2000 (*LARFI*, 2006).

I.7 Différents modes de transport solide

Le transport solide représente l'entraînement de granulats sous l'action d'un écoulement fluide. Le transport solide dans les cours d'eau constitue la seconde phase de l'érosion. Il relève essentiellement du ruissellement et met en jeu diverses formes d'écoulement de l'eau dans la nature (*Hug*, 1975).

Le mode de transport dépend essentiellement d'un paramètre fondamental : la taille du sédiment transporté, et de ce fait on peut distinguer trois types de transport solide pouvant être observés dans les bassins versants.

I.7.1 Transport par charriage

Le transport solide par charriage, représente la progression de sédiments grossiers (sables, gravier, galets et blocs), qui se déplacent par roulement ou glissement, au voisinage immédiat du

fond du cours d'eau, et parcourent des distances relativement petites et discontinues dans le temps et dans l'espace (Neboit, 1991).

Les interactions avec le lit sont alors prépondérantes et ce type de transport est directement lié à la morphologie du lit. D'autre part, il est nécessaire que les contraintes hydrauliques soient suffisamment élevées pour pouvoir assurer le transport des cailloux sur le fond, ce mécanisme est principalement régi par les forces de gravités (Bourouba, 1997).

I.7.2 Transport par saltation

Il correspond à des matériaux assez fins pour que les interactions avec le fond ne soient pas prépondérantes mais trop grossiers pour que les matériaux restent durablement dans l'écoulement. Dans les rivières, il s'agit essentiellement de sables. Dans le cas d'un bassin versant granitique, c'est ce type de transport qui est prépondérant (Neboit, 1991).

I.7.3 Transport par suspension

Le transport solide en suspension représente le transport de sédiments fins (argiles, limons, sables fins), maintenues en suspension sur toute la hauteur de la masse liquide du cours d'eau, sous l'action des fluctuations turbulentes de l'écoulement.

Les matériaux fins n'ont donc pas d'influence prépondérante sur le comportement de l'écoulement et la morphologie du lit mineur. Par contre, ils peuvent jouer un rôle prépondérant pour le développement des terrasses élevées ou l'envasement des retenues (Bourouba, 1997).

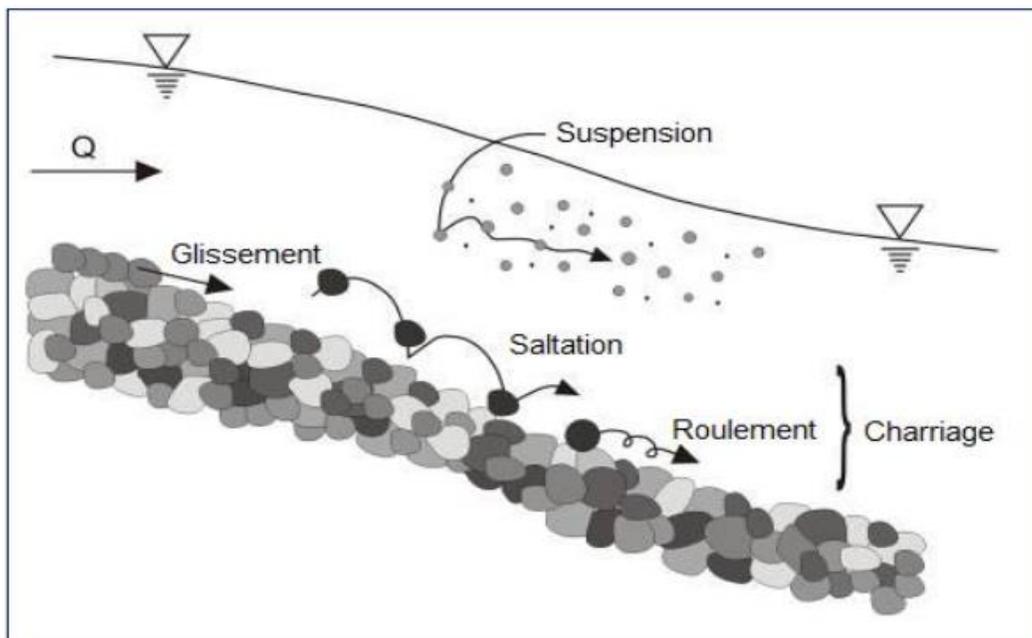


Figure I.13 : Modes du transport solide (Google)

I.8 Mesure du transport solide

I.8.1 Mesure de charriage

a. La mesure par échantillon

Cette mesure a été mise au point pour l'hydraulique fluviale, elle consiste à mettre des nasses dans le courant. Cette méthode semble peu adaptable aux torrents autant à cause des contraintes subies que des quantités et de la taille des éléments transportés (MAKKEB, 2016).

Une autre technique qui s'apparente dans son esprit à la précédente est celle du marquage des sédiments par un traceur radioactif. On injecte à un endroit précis du lit des sédiments marqués et on relève leur dispersion par les crues (MAKKEB, 2016).



Figure I.14 : Marquage des sédiments (Fanny, 2016).

b. La mesure globale

C'est la méthode la plus ancienne et la plus simple à mettre en œuvre dans les torrents. On utilise soit une fosse dans laquelle les sédiments se déposent, soit une plage de dépôts avec un barrage filtrant en travers du lit (MAKKEB, 2016).

La fosse à sédiments est utilisée lorsque les quantités transportées sont faibles et que la nature géologique du lit le permet. Sa forme géométrique facilite la mesure des dépôts (MAKKEB, 2016).

Le mode d'emploi est simple. Après chaque crue on procède à une mesure de la qualité de sédiments déposés. Lorsque la capacité de remplissage du piège est atteinte on procède à un curage, opération plus ou moins complexe suivant les quantités à évacuer (MAKKEB, 2016).



Figure I.15 : Plage de dépôt (Google).

c. La mesure en continu

Il existe deux systèmes qui permettent de mesurer en continu le charriage :

- La pesée quasi continue

Ce système est implanté dans une partie canalisée à l'amont d'un seuil. Sur le fond bétonné du canal une saignée oblique recueille les matériaux. L'ouverture d'une vanne latérale crée un vortex qui entraîne les matériaux en contrebas sur une trémie qui les sépare de l'eau. Un tamis assure une dernière réparation où n'est conservée que la fraction de dimension supérieure à 4mm. Un tapis roulant achemine les graviers vers une balance enregistreuse puis, après pesée, ils sont rejetés dans la rivière en aval du seuil (MAKKEB, 2016).

- L'acoustique passive

La mesure par acoustique passive consiste à écouter les bruits de la rivière. Pour cela, on utilise un capteur nommé hydrophone c'est à dire un microphone pour l'eau (GEAY, 2006).

Les sédiments, lorsqu'ils se déplacent au fond du lit, vont s'entrechoquer et générer du bruit dans des fréquences variant de quelques centaines de Hz à quelques centaines de kHz. Des études de laboratoire ont montré que les caractéristiques de ce bruit étaient liées à certaines caractéristiques du transport solide comme les flux ou la granulométrie. À ce jour, le potentiel de la méthode ainsi que ses limites ne sont pas connus (GEAY, 2006).

Cette méthode est relativement peu chère et facile à mettre en œuvre. Le grand avantage de l'acoustique réside dans sa résolution temporelle car la mesure est continue. Elle permettrait d'étudier le charriage pour des échelles de temps associées à la turbulence ou encore à la morphologie fluviale (GEAY, 2006).

L'interprétation du signal est cependant complexe et nécessite des connaissances à la croisée de différents champs scientifiques que sont le traitement du signal, l'acoustique et la physique de la rivière. Un schéma de principe de la mesure est proposé dans la figure I.16.

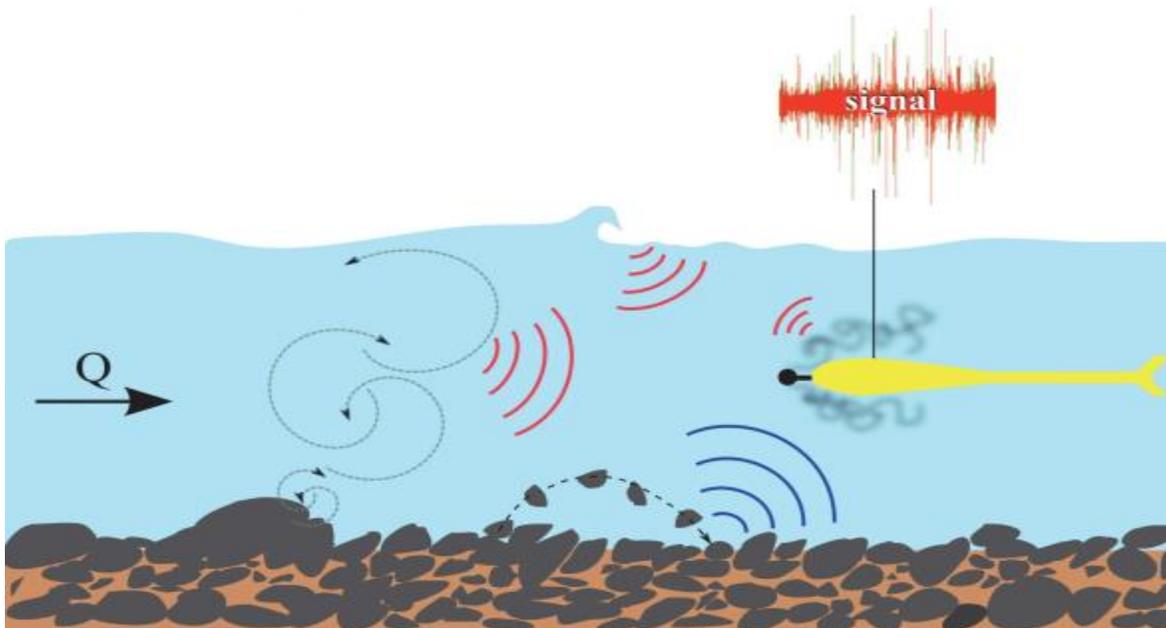


Figure I.16 : Schéma de principe de la mesure par acoustique passive (GEAY, 2006).

I.8.2 Mesure des suspensions

On retrouve pour les suspensions les mêmes principes de mesure que pour le charriage.

Il existe trois grands principes de mesures :

a. La mesure globale

Elle consiste à recueillir soit la totalité soit une fraction connue de l'écoulement et à faire une mesure globale des matières solides apportées par chaque évènement climatique (MAKKEB, 2016).

b. La mesure discontinue et ponctuelle

Elle permet la prise d'échantillons dans une section de cours d'eau. Le principe est simple, il consiste à prélever à un instant t une certaine quantité de liquide (en général 0,5 l) dans la section d'écoulement. On détermine ensuite la concentration de matières solides de l'échantillon et on

attribue cette valeur au débit qui est passé à cet instant, ce qui donne le flux de matières en suspension. En répétant cette mesure à intervalles adéquats, on arrive à suivre les variations de flux de matières en suspension avec le régime du cours d'eau (MAKKEB, 2016).

Pour que cette méthode donne des résultats représentatifs il faut que :

- La répartition des matières en suspension soit homogène dans la section de prélèvement ;
- L'instant de prélèvement soit déterminé précisément pour pouvoir être mis en relation avec le débit liquide ;
- Le nombre d'échantillons prélevés soit suffisamment important pour suivre les variations de régime du cours d'eau ;
- Le flux de suspension ne soit pas modifié localement par la prise d'échantillons.

Il existe différents procédés pour cette mesure :

➤ Les prélèvements manuels

C'est la méthode la plus ancienne. Le matériel consiste en bouteilles type « bouteille de lait », fermées par un clapet. La bouteille est positionnée dans la section mouillée à l'aide d'une perche, l'opérateur ouvre le clapet à l'aide d'un fil et retire la bouteille quand elle est pleine (MAKKEB, 2016).

Avantage de la méthode :

- Il est possible de faire des prélèvements en différents points de la section au même instant ou à des instants très proches.
- Il est possible de multiplier les points de prélèvements et de changer de section facilement.

Inconvénients :

- Le principal est la nécessité d'avoir des opérateurs sur place aux moments les plus intéressants, c'est-à-dire les crues.

➤ Les prélèvements automatiques

Le plus simple consiste à implanter un ou plusieurs pylônes dans le lit et d'accrocher des bouteilles (type « bouteille de lait ») à entrer calibrée à différentes hauteurs. Les bouteilles se remplissent au fur et à mesure de la montée de l'eau. On change les bouteilles pleines après chaque crue (MAKKEB, 2016).

Avantage de la méthode :

- Théoriquement, peut permettre de suivre plusieurs points d'une section en multipliant les implantations.

Inconvénients :

- Inutilisable dans les zones à fort charriage à cause des chocs de pierres ;
- Perturbe fortement l'écoulement à son voisinage ;
- Ne permet de suivre que la montée en crue. Aucune information sur la décrue et les régimes quasi stationnaires ;
- Les bouteilles peuvent être colmatées par des transports végétaux ;
- Le point de prélèvement varie avec la hauteur d'eau et ne concerne que la partie superficielle de l'écoulement.



Figure I.17 : Bouteille de mesure de la suspension (OSR4, 2015-2017).

c. La mesure en continu

Elle se fait à travers une section ou partie de section soit du cours d'eau, soit d'une dérivation du débit à l'aide d'un partiteur. Deux systèmes peuvent être utilisés pour suivre la teneur en MES dans les cours d'eau (MAKKEB, 2016):

- Les systèmes optiques (néphélomètres, turbidimètres) d'une part et l'absorption de rayonnement γ d'autre part.
- Les systèmes optiques se composent d'une source de lumière et d'une cellule photo-électrique qui mesure soit la lumière diffusée (néphélométrie) soit la lumière absorbée

(turbidimétrie) par le liquide. Un enregistreur-amplificateur recueille le courant produit par la cellule photo-électrique.

Ces appareils sont souvent utilisés et mis au point pour la chimie et la surveillance de MES dans les stations d'épuration.

L'avantage de ces appareils c'est que nous avons la possibilité d'avoir un suivi en continu des transports en suspension.

- Le principe de mesure avec une source radioactive artificielle est de mesurer l'atténuation du rayonnement γ émis par une source radioactive à travers le liquide. Le matériel se compose d'une source radioactive (en général un radio-isotope artificiel) et d'un détecteur à scintillation avec amplificateur (MAKKEB, 2016).

L'avantage de cette mesure est que l'analyse peut se faire sur pratiquement la totalité de l'écoulement en combinant plusieurs sondes si nécessaires.

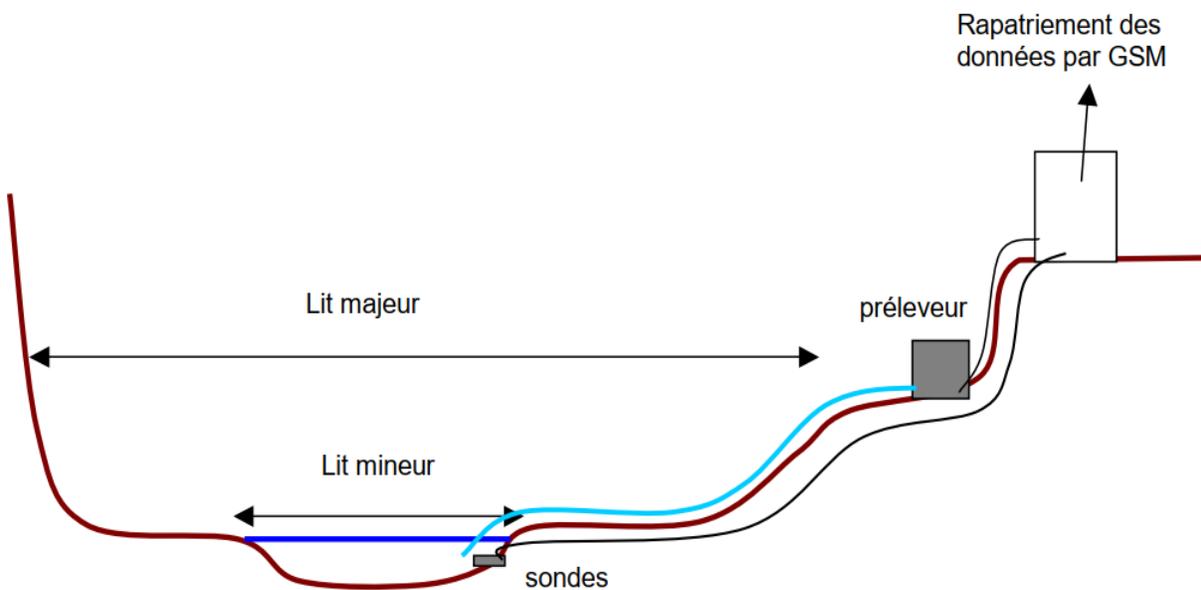


Figure I.18 : Installation du turbidimètre dans le cours d'eau (Wikipédia).

I.8.3 Méthodes de mesure en Algérie

En Algérie, la méthode d'échantillonnage systématique du transport solide sur la plupart des cours d'eau est l'échantillonnage uni ponctuel. Le prélèvement d'un échantillon se fait par l'observateur à l'aide de bouteilles de 0,5 à 1 litre de contenance. Les échantillons prélevés sont analysés par les procédés classiques (filtrage sur filtre millipore et la charge en suspension est mesurée par pesée des filtres) (MAKKEB, 2016).

Pour les débits liquides, les lectures de hauteur d'eau effectuées sont transformées en débits liquides grâce aux courbes d'étalonnage établies par les services de l'Agence Nationale des Ressources Hydrauliques. Ces courbes sont valables pour une période donnée dite période de validité (MAKKEB, 2016).

Quant au transport solide, l'évaluation durant une crue est effectuée de la même façon que pour les débits liquides, sur la base d'un dépouillement des lectures. Avant le tracé du turbidigramme, l'opération de report sur le limnigramme. Le nombre de prélèvements et leur échelonnement correct dans le temps conditionnent la fiabilité et la précision des graphes temporels de la concentration des sédiments en suspension (MAKKEB, 2016).

I.9 Conclusion

L'érosion des versants est un processus complexe et continu qui a causé d'énormes dégâts dans les pays arides et semi arides tels que l'Algérie, et un envasement important des retenues et des barrages. Dans ce chapitre nous avons détaillé les principaux processus et causes de l'érosion des sols. Notre travail sera consacré à l'étude du transport solide d'Oued Mellègue dans la Medjerda, et ceci pour une meilleure prévention et lutte contre ce phénomène.

Chapitre II :

Présentation de la zone d'étude

II.1 Situation géographique du bassin versant

II.1.1 Présentation du bassin versant de la Medjerda

Le bassin versant de la Medjerda se situe au Sud-Est du Nord Algérien, d'une superficie de 7785 Km², le bassin Medjerda-Mellègue est limité du côté Nord par le bassin des Côtiers Constantinois Est (03), au Sud par le bassin du Chott Melrhir (06), à l'Ouest par les bassins de la Seybouse (14) et des hauts plateaux Constantinois (07) et à l'Est par la frontière Tunisienne. Le bassin de la Medjerda couvre cinq wilayas « Souk Ahras- Tebessa-Khenchela-Oum El BouaghI.Guelma » (Hakima, 2018).

Le bassin étudié englobe cinq sous bassins versants, présentant un chevelu hydrographique moins développé et moins dense. Il comprend quatre grands Oueds « Medjerda – Chabro – Meskiana - Mellègue ». C'est un bassin septentrional, à écoulement relativement fourni, de type exoréique. Oued Mellègue, prend sa source des massifs de Bentaleb et des hautes plaines de Sedrata (Figure II.1, Photo 1). Oued Medjerda, est un cours d'eau dont la source est située en Algérie (monts de Souk-Ahras), et le lit se jette en Tunisie (Figure II.1, Photo 2). Il se caractérise par un écoulement permanent sur l'ensemble de son cours, ce qui lui donne le profil d'un fleuve (Hakima, 2018).



Photo 1: Oued Mellègue.



Photo 2: Oued Medjerda.

Figure II.1: Les principaux cours d'eau du bassin versant de Medjerda (Google).

CHAPITRE II. PRESENTATION DE LA ZONE

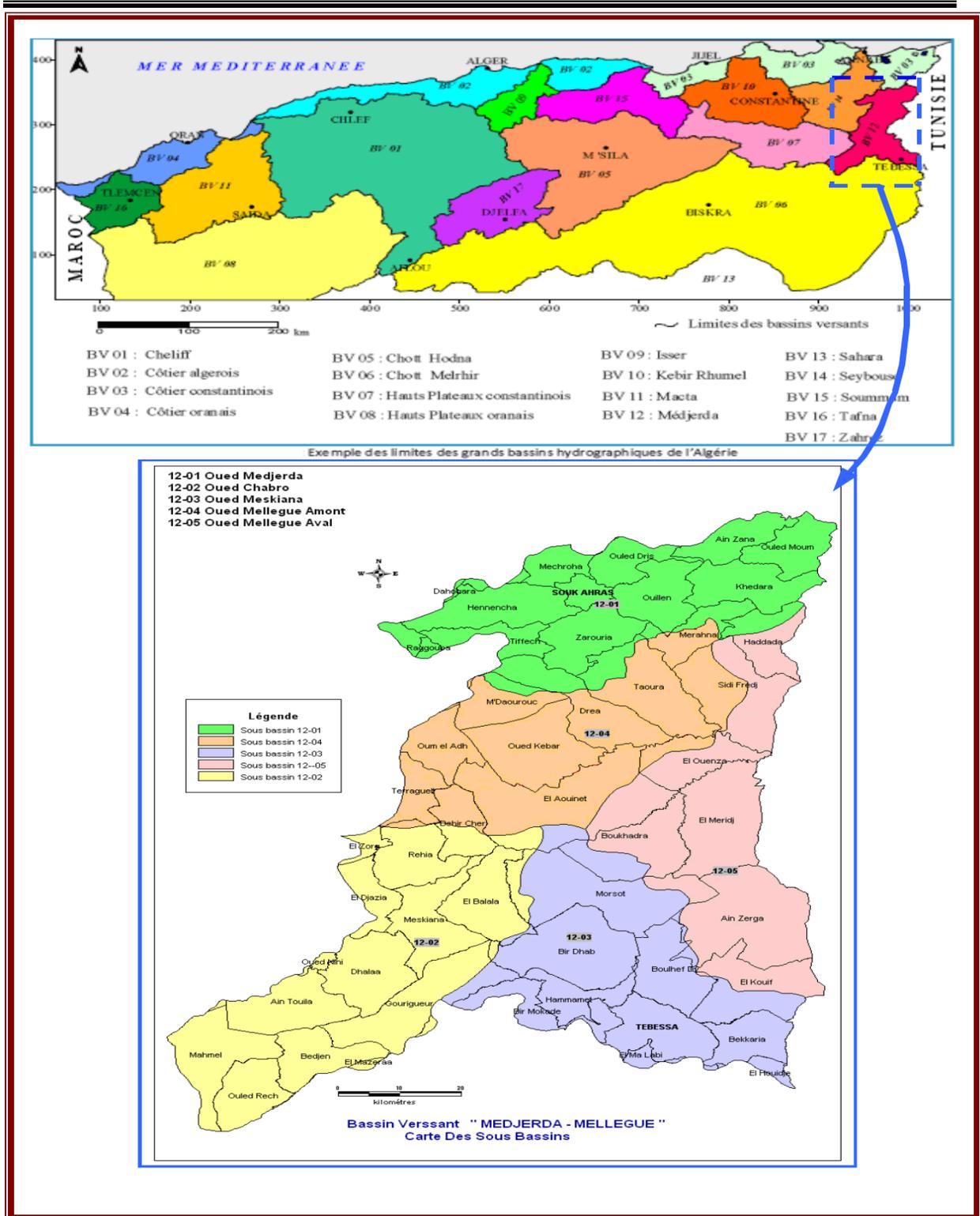


Figure II.2 : Situation géographique du bassin versant Medjerda (ABH Constantinois-Seybouse-Mellègue).

II.1.2 Présentation du sous bassin d'Oued Mellègue

Le sous bassin versant de Mellègue s'étend sur une surface de 3370 km² entre 35,13' et 35,51' Nord et 7,14' à 7, 50' Est. Comme le bassin est d'une forme allongée, il constitue une vaste région de transition entre l'étage bioclimatique tempéré au Nord et l'étage semi-aride au Sud (Gouaidia, Laouar, Défaflia & Zenati, 2018).

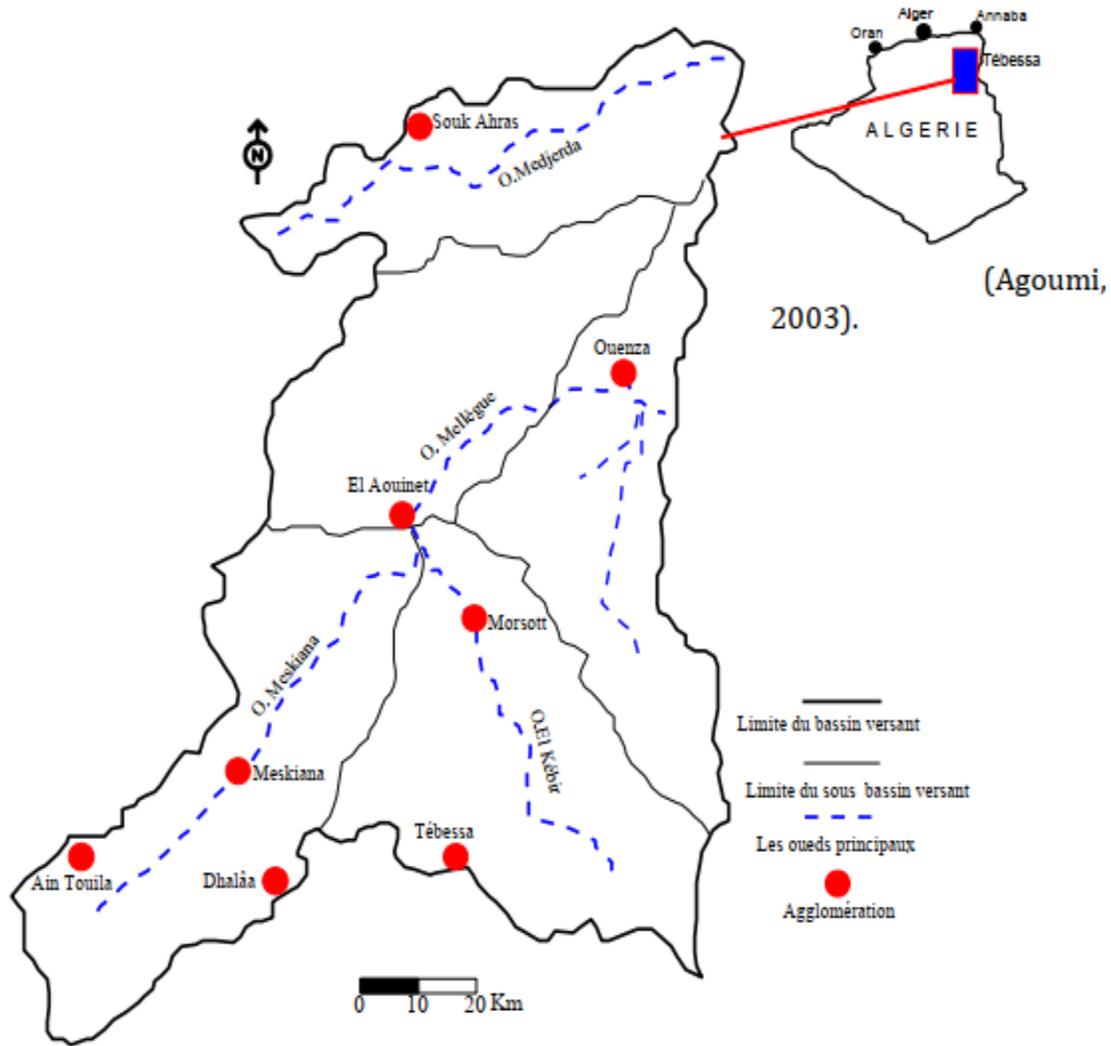


Figure II.3 : Situation géographique du bassin d'Oued Mellègue (Gouaidia, Laouar, Défaflia & Zenati, 2018).

II.2 Caractéristiques morphométriques

Tableau II.1: Les caractéristiques morphométriques du bassin versant d'Oued Mellègue amont (Stamboul, 2017) (Gouaidia, Laouar, Défaflia & Zenati, 2018).

Bassin versant	Oued Mellègue amont
Station	El Aouinet
Code de la station	12 04 04
Oued	Mellègue
Surface	3370 Km ²
Périmètre	290 Km
Altitude max	1626 m
Altitude moyenne	845 m
Altitude min	540 m
Longueur du Talweg principale	58.19 Km

III.3 Relief de la zone d'étude

La région d'étude présente une allure d'un plateau d'altitude moyenne d'environ 550 m. Ce plateau est limité au Sud-Ouest par Djebel Mesloul (1040m), Djebel Guelb (1132m), vers le Sud par Djebel Dzeroua (1040m) et Djebel Boukhadra (1463 m). Au Nord-Est l'altitude maximale est de 1288 m au niveau du Djebel Ouenza (juste au Sud-Ouest de la ville d'El Ouenza), suivi de Koudiat El Megsouma située au Nord de la ville d'El Ouenza avec 794 m et Kef Zerga (780m) et Djebel Def (712m) situés au Sud-Est (Stamboul, 2017).

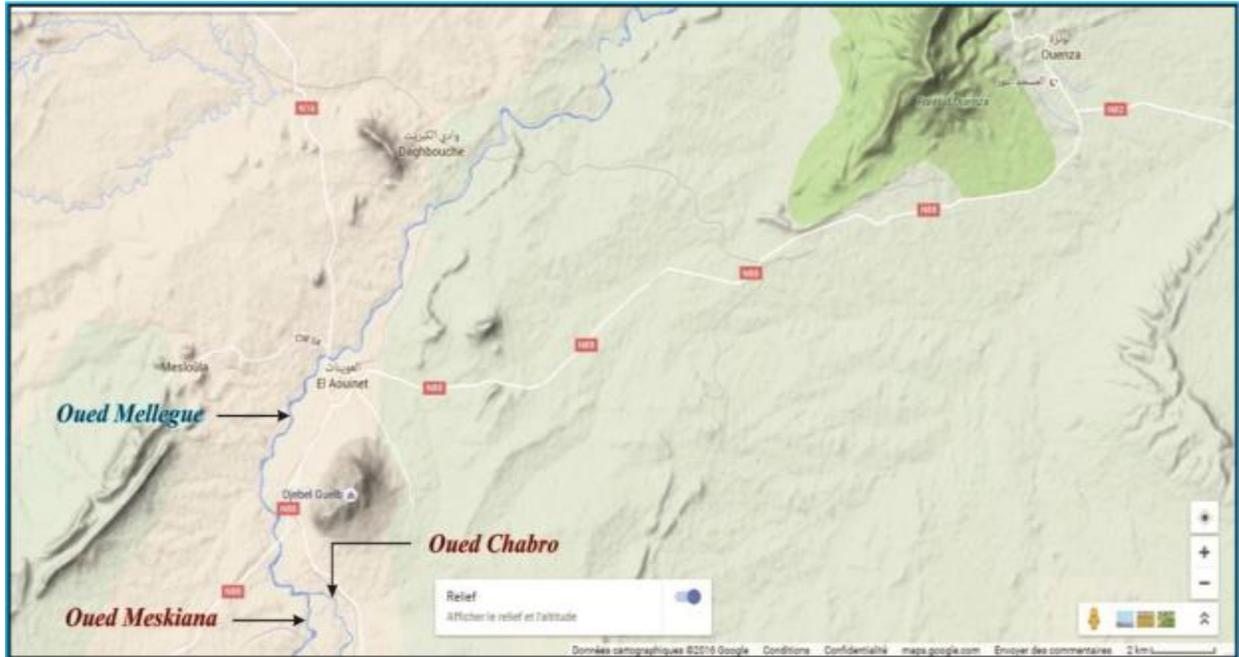


Figure II.4 : Topographie de la région d'étude (Google Maps).

II.4 Le réseau hydrographique

Le réseau hydrographique est bien développé et représenté avec des petits cours d'eau temporaires qui se déversent dans des grands Oueds présentant un écoulement pérenne : Oued Mellègue et Oued Kebarit à l'Ouest et Oued El Kseub à l'Est (Stamboul, 2017).

L'écoulement du système hydrographique se fait du Sud-Ouest vers le Nord-Est, caractère typique des Oueds de toute la région (Stamboul, 2017).

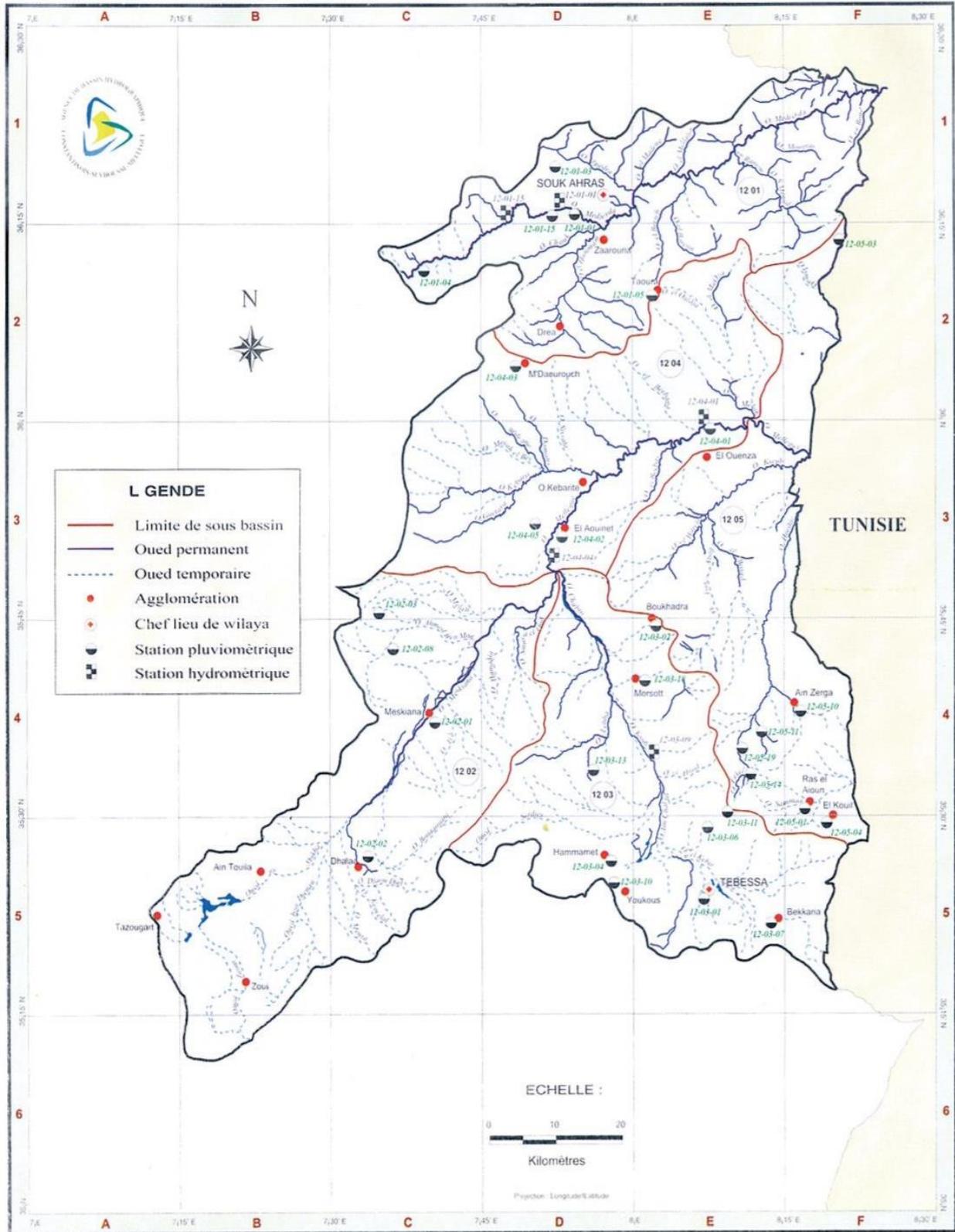


Figure II.6 : Stations pluviométriques et hydrométriques (Agence de bassin hydrographique Constantinois-Seybouse-Mellègue).

II.5 La géologie du secteur d'étude

Notre terrain d'étude fait partie de l'ensemble autochtone d'Algérie orientale et des confins Algéro-Tunisiens : C'est un ensemble de structures plissées à série sédimentaires mésozoïques épaisses s'étendant de la zone des diapirs à l'Est jusqu'aux monts du Hodna à l'Ouest en passant respectivement par les hautes plaines d'Ain Baida et Oum El Bouaghi, des Aurès et les Monts de Bélezma. La région d'étude s'intègre dans le secteur oriental Algéro-Tunisien (Stamboul, 2017).

Cette région englobe les monts de Morsott et de Mellègue. Il s'agit d'une vaste zone d'anticlinaux et de synclinaux orienté selon la direction atlasique SW-NE injectés ou percés par des diapirs d'évaporites. (Stamboul, 2017).

La série stratigraphique est constituée en majeure partie par des formations crétacées constituées de :

- Terrains anciens représentés par le Trias à faciès lagunaire, apparaissant en diapirs.
- Très importantes assises d'âge crétacé (Aptien et Maestrichtien) formées de marnes et de calcaires.
- Des grès et des calcaires gréseux marins formant la base du Miocène.
- Divers recouvrements continentaux représentant le Mio-Plio-Quaternaire.

II.6 Le climat

Dans un pays comme l'Algérie caractérisé par le contraste entre le climat méditerranéen de la bordure littorale et le climat désertique du Sud en passant par le climat des montagnes, des hauts plateaux, il est nécessaire de pouvoir préciser le type climatique du secteur que l'on étudie (Stamboul, 2017).

La région d'étude, sur le plan météorologique, constitue une zone de transition entre plusieurs phénomènes climatiques, notamment, le passage du régime climatique méditerranéen au climat saharien caractérisé par des influences desséchantes qui règnent, pendant une partie de l'année, sur le Sahara septentrional (Stamboul, 2017).

Le relief caractérisé par les derniers chaînons de l'Atlas Saharien et formant une barrière aux influences méditerranéennes accentue les contrastes climatiques entre l'Atlas Saharien et la Plate-Forme Saharienne (Stamboul, 2017).

II.6.1 L'évaporation

On dispose une série de mesures pour la station de Tébessa pendant la période 1971-2006.

Le tableau ci-dessous présente la répartition mensuelle de l'évaporation. L'effet de l'altitude et de la continentalité se traduit dans les valeurs de l'évaporation :

- En saison froide, l'altitude conduit à des évaporations relativement faibles.
- En saison chaude, la continentalité conduit à de fortes températures et à de fortes évaporations.

Tableau II.2 : Répartition mensuelle de l'évaporation (ANRH).

Mois	S	O	N	D	J	F	M	A	M	J	J	A	Annuel
Evaporation (mm)	221.2	161.4	109.2	93.32	88.89	106	133.5	175.1	224.7	280.4	368	333	2294.41
Evaporation (%)	9.64	7.03	4.76	4.08	3.87	4.62	5.82	7.63	9.79	12.22	16.03	14.51	100

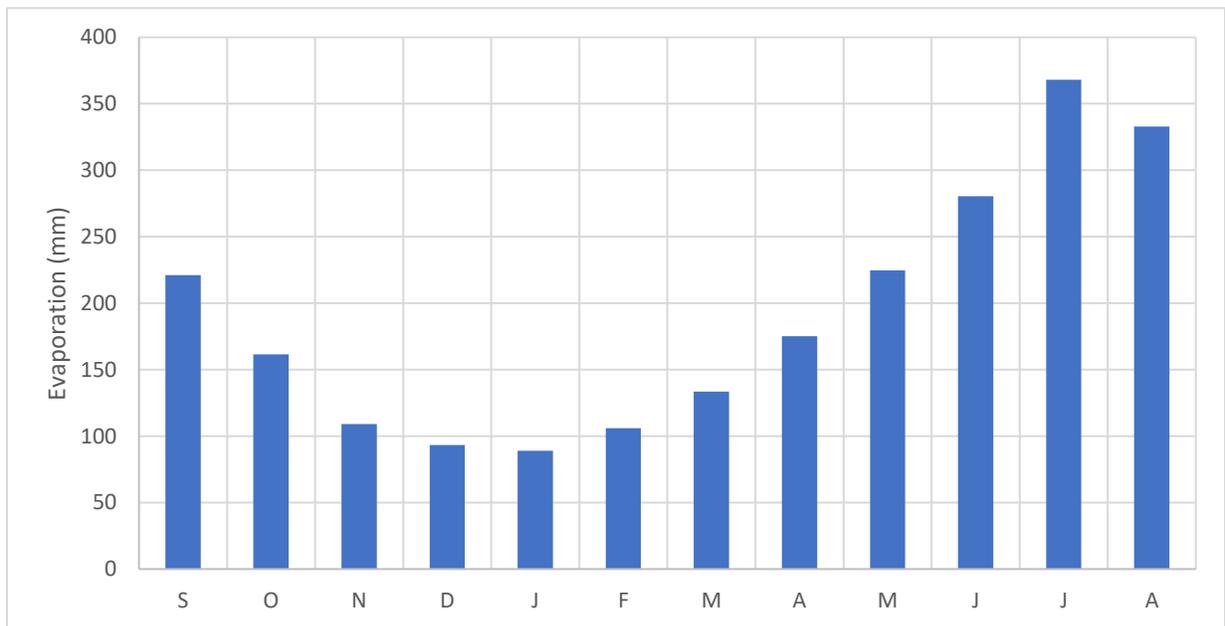


Figure II.7 : Répartition moyenne mensuelle de l'évaporation.

La figure montre, que l'évaporation la plus élevée, est observée pendant la période estivale, avec un maximum au mois de juillet de 367,95 mm, par contre, les valeurs les plus basses, s'observent pendant la période humide (hiver), avec un minimum de 88,89 mm, le mois de Janvier.

II.6.2 La pluviométrie

La pluviométrie demeure le paramètre le plus important, elle conditionne l'écoulement saisonnier, le régime des cours d'eau et l'alimentation des nappes d'eau souterraines.

Les hauteurs des précipitations moyennes annuelles sont fournies par l'ANRH. Les moyennes calculées pour la station d'El Aouinet sont reportées dans le tableau suivant pour la période 1987-2012.

Tableau II.3 : Répartition mensuelle de la pluie annuelle (ANRH).

Mois	S	O	N	D	J	F	M	A	M	J	J	A	Annuel
P (mm)	23.7	22.6	26	20	29.5	24.7	30.5	41.6	40.3	28.3	6.6	11.5	305.6

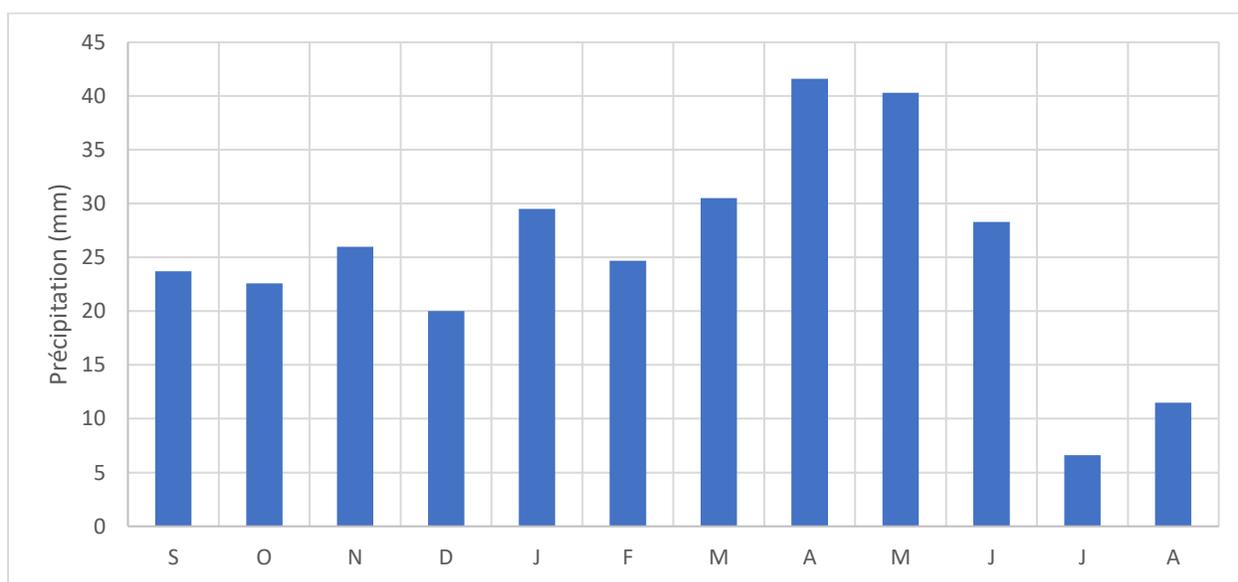


Figure II.8 : Répartition mensuelle de la précipitation.

D'après le tableau II.3, Avril et mai sont les mois les plus pluvieux, avec une moyenne de 41,64 mm pour avril et 40,34 mm pour le mois de mai, tandis que les précipitations moyennes les plus faibles sont enregistrées en Juillet avec 6,64 mm seulement.

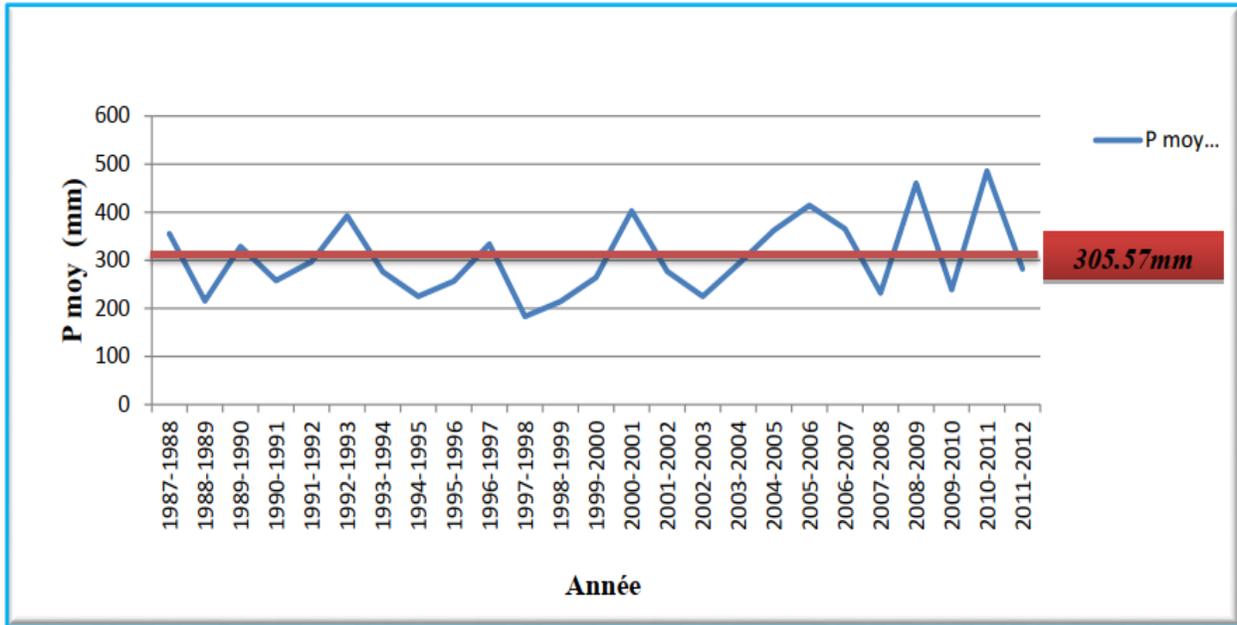


Figure II.9 : Précipitations moyennes annuelles de la station d’El Aouinet pour la période (1987/2012) (Stamboul, 2017).

II.6.3 Autres informations sur le climat

Le climat de la zone d’étude est un climat des Hautes Plaines Telliennes à tendances continentales et semi arides. Le sirocco est assez fréquent en été. La végétation naturelle est éparse.

Le tableau suivant résume les paramètres climatiques de la station de Tébessa.

On y donne :

- La température moyenne, T_{moy} .
- L’humidité relative, Hr.
- Le nombre de jours de pluie, N_j plu.
- La vitesse du vent, V_{vent} .
- Le nombre de jours de sirocco, N_j sir.
- Le nombre de jours de gelée blanche, N_j gel.

Tableau II.4 : Paramètres climatiques de la station de Tébessa (Khedidji, 2012).

Mois	S	O	N	D	J	F	M	A	M	J	J	A	An
$T_{moy}(^{\circ}c)$	23	16	11	7	6	7	10	14	18	23	27	26	16
Hr(%)	50	57	66	67	67	63	55	46	48	41	46	39	53
N_j plu	5	5	6	7	8	7	7	6	6	4	3	3	67
V_{ven} (m/s)	1.8	1.9	1.9	2.8	2	2.3	2.1	2.1	2.1	2	2.1	1.7	2
N_j sir	1	1	0	0	0	0	1	1	2	3	5	2	16

Nj gel	0	0	1	6	6	4	2	0	0	0	0	0	19
--------	---	---	---	---	---	---	---	---	---	---	---	---	----

II.7 Présentation des données hydrométriques

Pour les mesures du débit liquides, l'ANRH fait un prélèvement des hauteurs d'eau dans l'Oued à l'aide d'une échelle, ensuite cette hauteur est convertie en débit en utilisant la courbe de tarage (courbe : hauteur-débit).

La procédure du prélèvement du transport solide est différente, car ce dernier se compose de la somme du transport par suspension et du transport par charriage. Le premier est continu dans le temps, tant que l'écoulement existe. Le deuxième est discontinu, et n'apparait qu'à la faveur des crues.

La charge de la matière est obtenue à partir d'échantillons d'eau prélevés sur les rives de l'Oued au moyen d'un conteneur de 50 cl. La boue recueillie sur un papier filtre est pesée après séchage à l'étuve à 105° C pendant 30 minutes. On détermine ensuite la charge correspondante à un litre d'eau prélevé, ce qui établit la concentration, donnée en g/l. La fréquence des prélèvements effectués dépend de la variation de la hauteur d'eau.

Le bassin versant de la Medjerda contient un réseau hydrométrique implanté par l'ANRH équipé de 5 stations hydrométriques (Agence de Bassin Hydrographique Constantinois-Seybouse-Mellègue):

Souk Ahras (120101) ;

Ain Dalia (120115) ;

Morsott (120309) ;

El Ouenza (120401) ;

El Aouinet (120404) ;

Notre travail consiste à modéliser le transport solide au vu de son importance dans cette région, et le peu d'études effectuées, comparées aux Tunisiens.

L'ANRH nous a gracieusement fournie une série de données de la station d'El Aouinet (12 04 04) qui s'étale de 1973 jusqu'à 1996, contenant 1428 valeurs instantanés de débit liquide (Q_L en m^3/s), concentration (C en g/l) et les hauteurs d'eau (H en mm).

II.7.1 Paramètres statistiques des données

La station El Aouinet située sur Oued Mellègue amont, a permis de recueillir cette série de données qui est composée de débits liquides instantanés ainsi que de la concentration en MES.

Les variations statistiques de ces données sont représentées dans les figures suivantes :

Nous avons 1428 valeurs instantanées (débits liquide-solide).

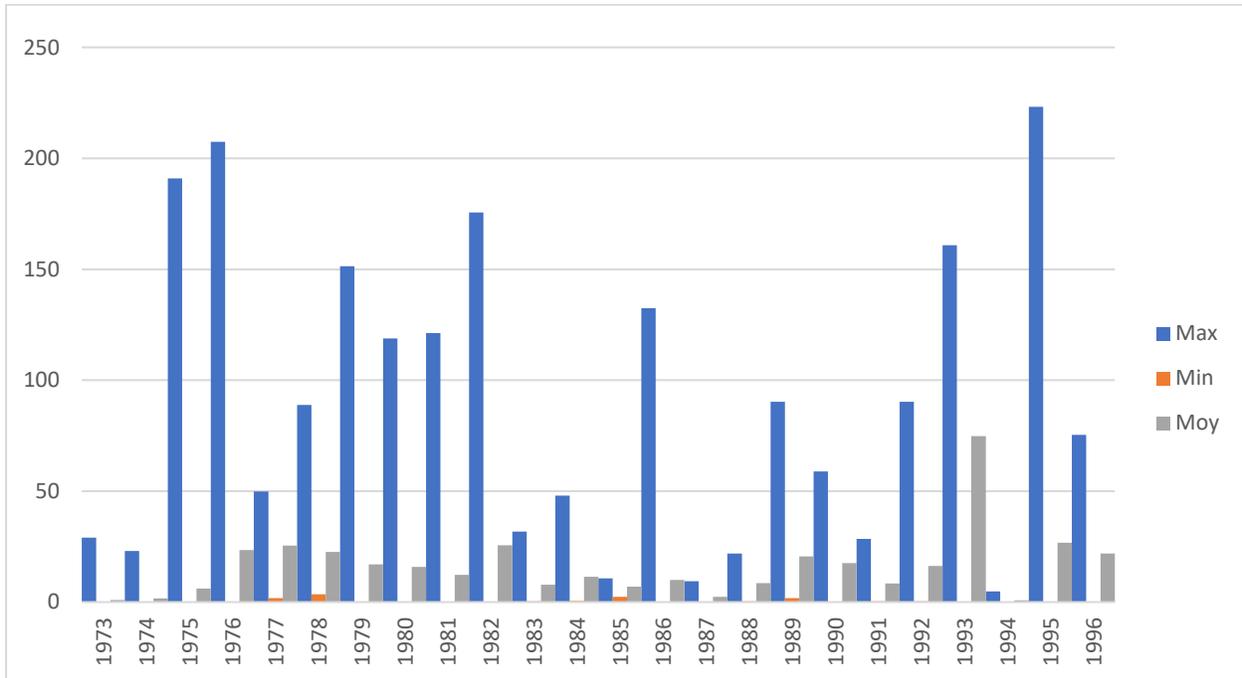


Figure II.10 : Caractéristiques statistiques du débit liquide.

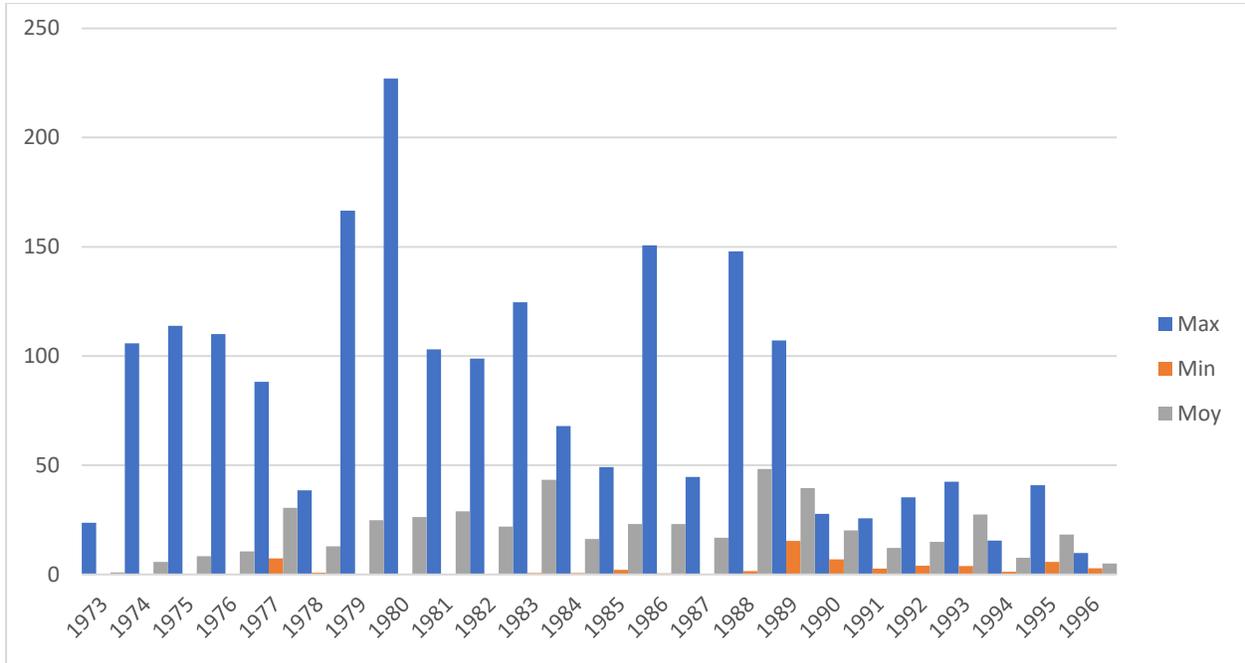


Figure II.11 : Caractéristiques statistiques de la concentration.

II.7.2 Les données utilisées

Nous avons travaillé avec les débits journaliers. Lorsqu'il y a une seule mesure on prend la valeur mesurée directement, mais lorsqu'il y a plusieurs mesures pendant la journée, on prend la moyenne.

II.8 Conclusion

Dans ce chapitre, nous avons détaillé les principales caractéristiques hydrologiques et climatologiques du bassin versant de l'Oued Mellègue, soumis à un climat aride et un régime hydrologique très irrégulier.

De par sa superficie et sa lithologie, le bassin versant d'Oued Mellègue est vulnérable à l'érosion et aux transports solides.

Ensuite, on a présenté les données recueillies auprès de l'ANRH, qui seront utilisées dans la modélisation à partir des techniques de Machine Learning.

Chapitre III :
Méthodes de Machines Learning

III.1 Introduction

Les techniques de traitement de l'information se sont largement développées ces dernières décennies avec le développement de l'informatique, tant sur le plan matériel que logiciel. La construction de logiciels s'appuie sur plusieurs approches. On citera deux des plus utilisées : l'approche algorithmique et l'approche par Machine Learning. Une approche algorithmique nécessite l'écriture du processus à suivre pour résoudre le problème. Malheureusement, les machines suivent à la lettre les instructions du programme dans une logique binaire qui ne permet pas la résolution des problèmes complexes.

La seconde approche possible est celle de l'intelligence artificielle (plus précisément Machine Learning), qui est présentée comme la vague du futur en informatique. C'est en effet des mécanismes d'auto-apprentissage qui ne nécessitent pas les compétences traditionnelles d'un programmeur (Lefkir, 2007).

Dans ce chapitre, on va présenter et comparer deux méthodes de Machine Learning : les Réseaux de Neurones Artificiels et la méthode du Gradient Boosting.

III.2 Les Réseaux de Neurones Artificiels

III.2.1 Définition

Les réseaux de neurones artificiels sont une formulation mathématique simplifiée des neurones biologiques. Ils ont la capacité de mémorisation, de généralisation et surtout d'apprentissage qui est le phénomène le plus important. (Hocini 2016).

III.2.2 Types des réseaux de neurones artificiels

Pour concevoir un réseau de neurones, nous devons établir des connexions entre les neurones. Nous avons quatre types principaux de connexion : directe, récurrente, latérale et à délais. Tous les réseaux de neurones utilisent la connexion directe pour acheminer l'information de l'entrée vers la sortie. La connexion récurrente, permet d'acheminer l'information, de la sortie des neurones des couches supérieures, vers les entrées des neurones précédents. Les réseaux de neurones qui doivent choisir un neurone gagnant, utilisent la connexion latérale, pour établir une relation entre les neurones de sortie, et la maintenir. Finalement, les problèmes temporels sont résolus par les modèles de réseaux dynamiques avec des connexions à délais.

Une couche est définie comme un ensemble de neurones, situé au niveau d'un réseau de neurones. Nous avons, par exemple, une couche de neurones de sortie, avec des couches situées

entre les entrées-sorties appelées couches cachées. Les réseaux de neurones possèdent une ou plusieurs couches de neurones, et leur dimension dépend du nombre de couches, et du nombre de neurones par couche. (Hocini 2016).

III.2.2.1 Perceptron

Le perceptron est la forme la plus simple d'un réseau de neurones, il modélise la perception visuelle. Il comprend trois principaux éléments : la rétine, les cellules d'association et les cellules de décision. La fonction d'activation utilisée dans ce réseau est de type tout ou rien (0 ou 1). L'apprentissage du perceptron peut se faire avec plusieurs méthodes déjà utilisées, il n'y a qu'une seule couche de poids modifiables entre les cellules d'association et les cellules de décision. Le perceptron est limité dans ses applications. Premièrement, il ne peut être applicable que dans la classification dont les variables sont linéairement séparables et deuxièmement la sortie ne peut être que 0 ou 1.

III.2.2.2 Perceptron multicouche

Cette classe est la plus importante des réseaux de neurones car elle représente la généralisation du perceptron monocouche avec une fonction d'activation de type sigmoïde et une ou plusieurs couches cachées. Le vecteur d'entrée se propage dans le réseau de couche en couche jusqu'à la sortie, l'entraînement de celui-ci se fait avec l'algorithme par la rétropropagation de l'erreur. Ce réseau est caractérisé par son modèle du neurone traitant les non-linéarités. Il peut comporter une ou plusieurs couches cachées et un plus grand nombre de connexions permettant de résoudre la majorité des problèmes.

Même avec les avantages des couches cachées et la performance de l'algorithme d'apprentissage, il reste plusieurs problèmes non réglés comme le choix du nombre de couches, le nombre de neurones par couche et le problème des minimums locaux où le réseau peut converger.

III.2.2.3 Réseau de neurones linéaire

Cette classe de réseaux diffère du perceptron car elle possède un neurone dont la fonction d'activation est linéaire. L'une des règles d'apprentissage permet d'effectuer une descente de gradient de l'erreur sur une mesure d'erreur quadratique. Les domaines d'application comprennent la commande, le contrôle, et le traitement du signal. L'avantage de ce réseau est qu'il converge sur un seul minimum si la solution existe, sinon l'ajout de couches n'a aucun

effet. Parmi ses inconvénients, il est limité à une couche de sortie et ne peut résoudre que les problèmes dont la relation entrées/sorties est linéaire.

III.2.2.4 Réseau Hopfield

Ce réseau est basé sur les principes de la physique statistique et il est fondamentalement une mémoire adressable par son contenu. Les neurones, basés sur le modèle étudié en, sont tous interconnectés. Plusieurs domaines d'application sont possibles, en particulier : les mémoires associatives et l'économie. La principale limitation est qu'il n'y a pas de couches cachées.

III.2.3 Structure d'un réseau de neurones artificiel

III.2.3.1 Structure d'un neurone

Le neurone est la cellule fondamentale d'un réseau de neurones artificiels. Par analogie avec le neurone biologique, le neurone doit être apte à accomplir les tâches suivantes : collecter, traiter les données qui viennent des neurones émetteurs et transmettre les messages aux autres neurones. La relation entre l'entrée et la sortie du neurone peut être donnée par l'équation suivante :

$$S_i = F(a) \quad \text{(III.1)}$$

$$a = \sum W(i, j). X(j) \quad \text{(III.2)}$$

Les variables N , S_i , F , X et W désignent respectivement le nombre d'entrées du réseau de neurones, le vecteur de sortie du réseau, la fonction d'activation, le vecteur des entrées du réseau de neurones et la matrice des poids. Nous présentons dans la Figure III.1 la structure d'un neurone simple. (Hocini 2016).

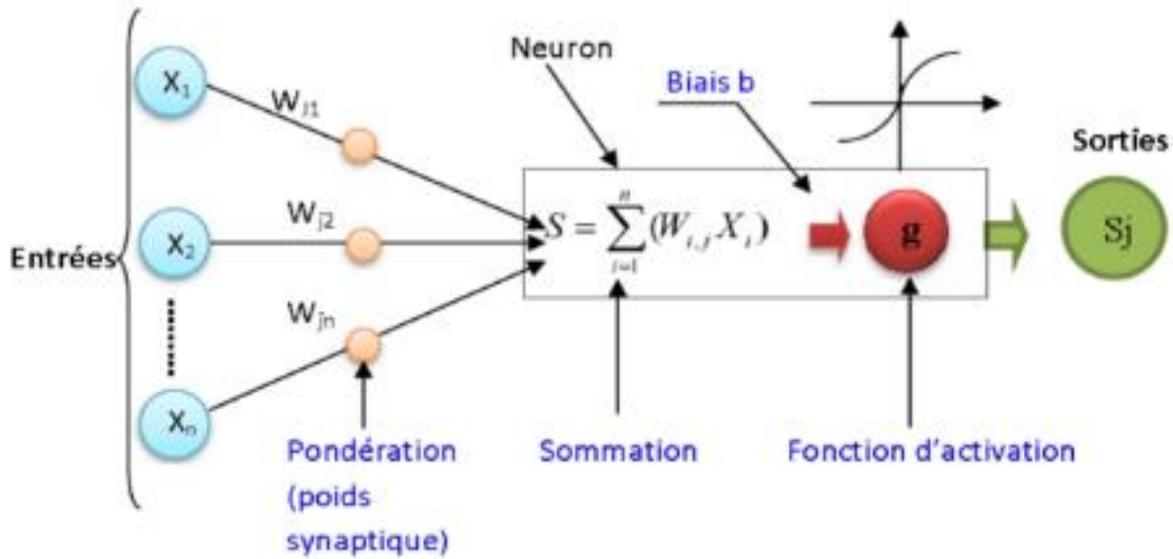


Figure III.1 : Structure d'un neurone simple (Google)

III.2.3.2 Structure d'un réseau de neurones à couches simples

Le réseau à couches est un réseau dont les neurones sont organisés en couches, la forme la plus simple est le réseau à une seule couche. Tous les signaux d'entrées sont propagés des nœuds d'entrée vers la couche de neurones de sortie. Le nombre de neurones d'entrée (nœuds) et de sortie est en général lié au problème à résoudre. Les entrées seront propagées à travers la matrice des poids W pour ensuite obtenir la réponse de sortie (Figure III.2). L'équation équivalente peut s'écrire sous la forme :

$$y(j) = \sum W(i, j) \cdot X(i) \quad (\text{III.3})$$

Où :

$X(i)$: Vecteur d'entrée.

$y(j)$: Vecteur de sortie.

$W(i,j)$: Poids du réseau de neurones.

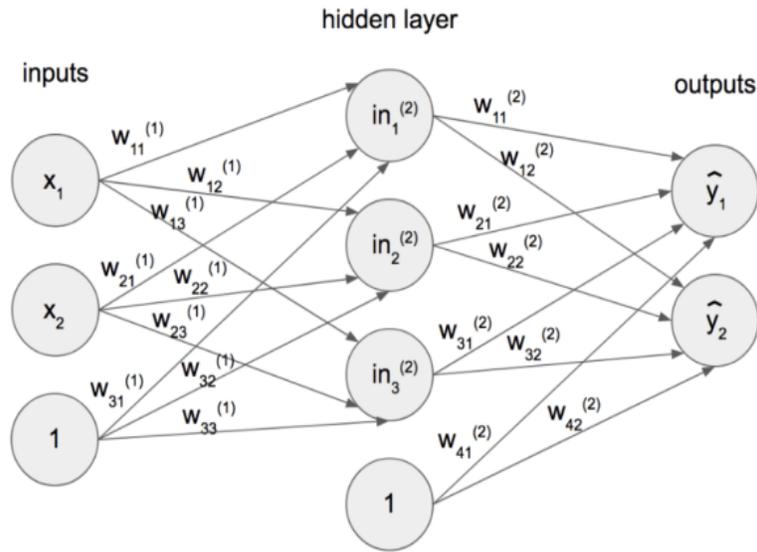


Figure III.2 : Réseau de neurones monocouche (Google)

III.2.3.3 Structure d'un réseau de neurones multicouches

Cette structure est caractérisée par une ou plusieurs couches. À chaque couche correspond une matrice de poids W , et un vecteur de seuils b , et on peut avoir aussi des fonctions d'activation différentes pour chaque couche. La fonction des couches cachées intervient entre les entrées x , et la couche de sortie y . Elle permet de résoudre des problèmes plus complexes que le réseau à couche simple. Les réseaux de neurones utilisant un apprentissage par rétropropagation sont constitués de plusieurs couches (multicouches) qui comportent chacune plusieurs neurones. Ces derniers sont reliés entre eux par des matrices de poids qui caractérisent le réseau. Dans ce type de réseau de neurones, chaque neurone a d'abord le rôle de sommer les différentes entrées pondérées par leurs poids respectifs, puis de passer la somme obtenue dans une fonction $F(\cdot)$ linéaire ou non-linéaire et finalement, de transmettre le résultat aux neurones de la couche suivante (Hocini 2016).

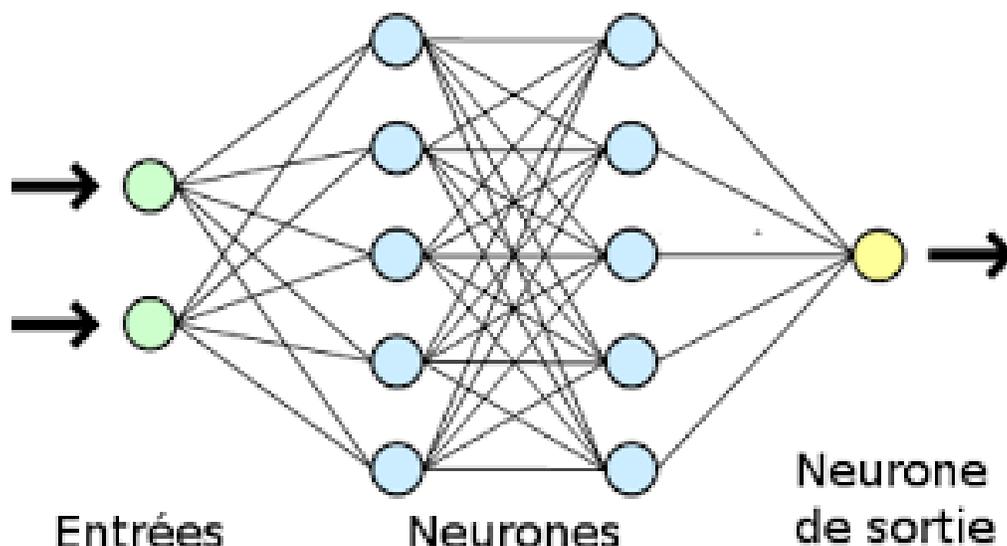


Figure III.3 : Réseau de neurones multicouches (Google)

III.2.4 Fonctions d'activation des RNA

La fonction d'activation calcule la valeur de l'état du neurone qui est transmise aux neurones avants. La fonction d'activation sert aussi à introduire une non-linéarité dans le fonctionnement du neurone. A la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle $[0, +1]$ ou $[-1, +1]$. Il existe de nombreuses formes possibles pour la fonction de transfert. (Si Moussa 2010).

On peut identifier trois principaux types de fonctions les plus connues : binaire à seuil, rampe avec saturation, et la sigmoïde.

III.2.4.1 Fonction binaire à seuil

La Figure III.4 présente la fonction d'activation qui a été utilisée. Le seuil introduit une non-linéarité dans le comportement du neurone, c'est le modèle tout ou rien.

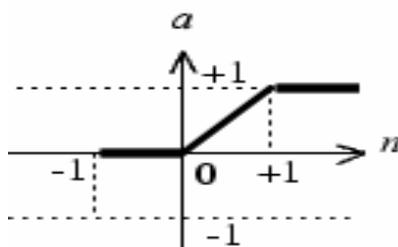


Figure III.4 : Fonction d'activation binaire à seuil (Si Moussa 2010)

III.2.4.2 Fonction à rampe avec saturation

Cette fonction représente un compromis entre la fonction linéaire et la fonction seuil : entre ses deux bornes, elle confère au neurone une combinaison linéaire de l'entrée. À la limite, la fonction linéaire est équivalente à la fonction seuil (Figure III.5).

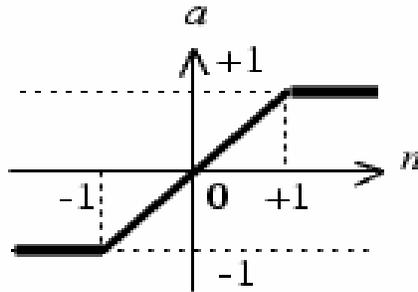


Figure III.5 : Fonction d'activation à rampe avec saturation (Si Moussa 2010)

III.2.4.3 Fonction sigmoïde

La fonction sigmoïde est une fonction continue qui maintient la sortie dans l'intervalle $[0,1]$ (Figure III.6). Son avantage principal est l'existence de sa dérivée en tout point. Elle est employée en général dans le perceptron multicouche.

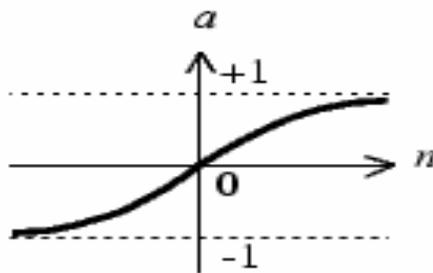


Figure III.6 : Fonction sigmoïde (Si Moussa 2010)

III.2.5 Modes d'apprentissage des réseaux de neurones artificiels

On peut définir l'apprentissage comme la capacité d'emmagasiner des informations qui peuvent être rappelées par la suite. Les connaissances d'un réseau connexionniste sont mémorisées dans les poids de connexions qui seront déterminés lors de l'apprentissage. Le but de l'apprentissage pour un réseau est de trouver un ensemble de poids synaptiques qui minimisent l'erreur entre la sortie du réseau et le résultat désiré. C'est la caractéristique principale des réseaux de neurones.

Les possibilités de reconnaissance des données dans le temps sont statiques ou temporelles. Un réseau de neurones est statique si l'ordre des données dans le temps ne change pas l'interprétation faite par celui-ci et il est temporel si cela change l'interprétation. Les possibilités de reconnaissance de données temporelles ajoutent la notion de mémoire aux réseaux de neurones. Cette mémoire est déterministe ou stochastique.

L'entraînement consiste à présenter, de façon répétitive, un ensemble de données pour permettre au réseau de neurones d'apprendre à résoudre ce problème en ajustant correctement ses poids. L'exécution est l'utilisation d'un réseau, avec tous ses paramètres et ses poids prédéterminés, pour la résolution d'un problème. Les caractéristiques qui vont constituer l'ensemble de données doivent être consistantes pour définir clairement les frontières des classes. Ces caractéristiques ont avantage à être limitées à celles qui sont vraiment nécessaires à la résolution du problème.

Les méthodes d'entraînement sont divisées en trois catégories : avec supervision, sans supervision et avec des poids prédéterminés. L'entraînement supervisé consiste à ajuster par interpolation, approximation ou optimisation les poids du réseau de neurones. Les réseaux de neurones supervisés dont l'ajustement des poids est obtenu par interpolation sont appelés à base de décision ("decision based neural network"). Avec l'entraînement non supervisé, le réseau de neurones s'ajuste uniquement en fonction de l'ensemble d'entrées. Il n'est pas nécessaire de connaître a priori les résultats et de constituer un ensemble de sorties désirées. Les réseaux de neurones à poids prédéterminés n'ont pas de phase d'apprentissage et la valeur des poids doit être déterminée par l'utilisateur. Une fonction d'énergie minimum est souvent utilisée pour calculer ces poids.

Pour l'apprentissage supervisé, on présente au réseau des entrées et on lui demande de modifier sa pondération de telle sorte que l'on retrouve la sortie correspondante. L'algorithme consiste, dans un premier temps à propager vers l'avant les entrées jusqu'à obtenir une sortie calculée par le réseau. La seconde étape compare la sortie calculée à la sortie réelle connue. On modifie alors les poids de telle sorte qu'à la prochaine itération, l'erreur commise entre la sortie calculée et connue soit minimisée. Malgré tout, il ne faut pas oublier que l'on a des couches cachées. On rétropropage alors l'erreur commise vers l'arrière jusqu'à la couche d'entrée tout en modifiant la pondération. On répète ce processus sur tous les exemples jusqu'à ce que l'on

obtienne une erreur de sortie considérée comme négligeable. Les poids optimaux seront alors appliqués sur le test pour vérifier la justesse et la précision des calculs. (Hocini 2016).

III.2.6 Sélection de la structure du réseau

Il est bien établi que la variation du nombre des neurones de(s) couche(s) cachée(s) a un effet significatif sur la capacité prédiction du réseau. Le nombre adéquat de neurones cachées dépend d'une façon complexe de :

- Les nombres de neurones de la couche d'entrée et de sortie ;
- La taille ou la dimension de l'échantillon d'apprentissage ;
- La complexité de la fonction ou la classification à faire apprendre au réseau ;
- L'architecture du réseau ;
- Le type de fonction d'activation des neurones cachés ;
- L'algorithme d'apprentissage ;

Dans la plupart des situations, il n'y a aucune façon de déterminer le nombre adéquat de neurones cachés sans faire l'apprentissage de plusieurs réseaux et d'évaluer ensuite l'erreur de généralisation de chaque réseau. Très peu de neurones cachés donnent des erreurs d'apprentissage et de généralisation top élevées en raison du sous apprentissage. Par contre un nombre est trop élevé de neurones cachés, donne une erreur d'apprentissage très faible mais une erreur de généralisation trop élevée en raison du sur apprentissage. (Si Moussa 2010).

III.2.7 Sélection de l'algorithme d'apprentissage

Plusieurs algorithmes d'apprentissage utilisent la procédure de rétro propagation La performance de chaque algorithme dépend du processus modélisé, de l'échantillon d'apprentissage et le mode d'apprentissage.

L'algorithme de rétro propagation du gradient comporte, donc, deux phases :

- La première phase "forward" concerne le signal fonction qui se propage vers l'avant du réseau, d'un neurone à un autre, après initialisation des poids synaptiques et des biais. Pendant cette phase, les poids synaptiques restent inchangés. Cette phase commence, donc, à la première couche cachée par le calcul du signal fonction de tous les neurones qui la constitue et se termine à la couche de sortie par le calcul de la fonction d'erreur de tous les neurones qui la constitue aussi.

- La deuxième phase "backward" concerne le signal erreur qui se propage, dans le sens contraire, couche par couche, à partir de la couche de sortie. Pour chaque neurone, le gradient local de l'erreur est calculé pour permettre la correction des poids synaptiques, liaison par liaison, selon une règle connue sous le nom de la règle Delta. Dans le cas d'un neurone de sortie, le gradient local est égal au signal d'erreur multiplié par la dérivée de la fonction d'activation appliquée à l'entrée nette de ce neurone. (Si Moussa 2010)

III.3 Arbres de Décision

L'arbre de décision est l'une des techniques les plus simples et intuitives dans l'apprentissage automatique, basée sur le paradigme diviser pour régner. L'idée de base derrière les arbres de décision est de diviser l'espace en parcelles, et d'adapter un modèle à une parcelle. Les arbres de décision ont un nombre de propriétés qui font d'eux un outil précieux, surtout quand il s'agit de faire l'analyse rapide d'un jeu de données ou d'élaborer un prototype de classifieur (Bennaceur, 2019) :

- Modèle white box : le résultat est facile à conceptualiser, à visualiser et à interpréter.
- Ils nécessitent peu de préparation de données (e.g. normalisation, etc.).
- Le coût d'utilisation des arbres est logarithmique (classification d'une nouvelle donnée très rapide).
- Ils sont capables d'utiliser des données catégorielles et continues.
- Ils sont capables de gérer des problèmes multi-classes.
- Ils ont un bon comportement par rapport aux valeurs extrêmes (Outliers).
- Ils gèrent bien les données manquantes.

III.3.1 Fonctionnement des Arbres de Décision

On considère d'abord le problème de classement. Chaque élément x de la base de données est représenté par un vecteur multidimensionnel (x_1, x_2, \dots, x_n) correspondant à l'ensemble de variables descriptives du point. Chaque nœud interne de l'arbre correspond à un test fait sur une des variables x_i :

- Variable catégorielle : génère une branche (un descendant) par valeur de l'attribut.
- Variable numérique : test par intervalles (tranches) de valeurs.

Les feuilles de l'arbre spécifient les classes. Une fois l'arbre construit, classer un nouvel candidat se fait par une descente dans l'arbre, de la racine vers une des feuilles (qui encode la

décision ou la classe). A chaque niveau de la descente on passe un nœud intermédiaire où une variable x_i est testée pour décider du chemin (ou sous-arbre) à choisir pour continuer la descente.

Au départ, les points de base de l'apprentissage sont tous placés dans le nœud racine. Cette variable est dite « variable cible ». La variable cible peut être catégorielle (problème de classement) ou valeur réelle (problème de régression). Chaque nœud est coupé (opération split) donnant naissance à plusieurs nœuds descendants. Un élément de la base d'apprentissage situé dans un nœud se retrouvera dans un seul de ses descendants.

L'arbre est construit par partition récursive de chaque nœud en fonction de la valeur de l'attribut testé à chaque itération (top-down induction). Le critère optimisé est l'homogénéité des descendants par rapport à la variable cible. La variable qui est testée dans un nœud sera celle qui maximise cette homogénéité. (Bennaceur, 2019).

Decision Tree Diagram

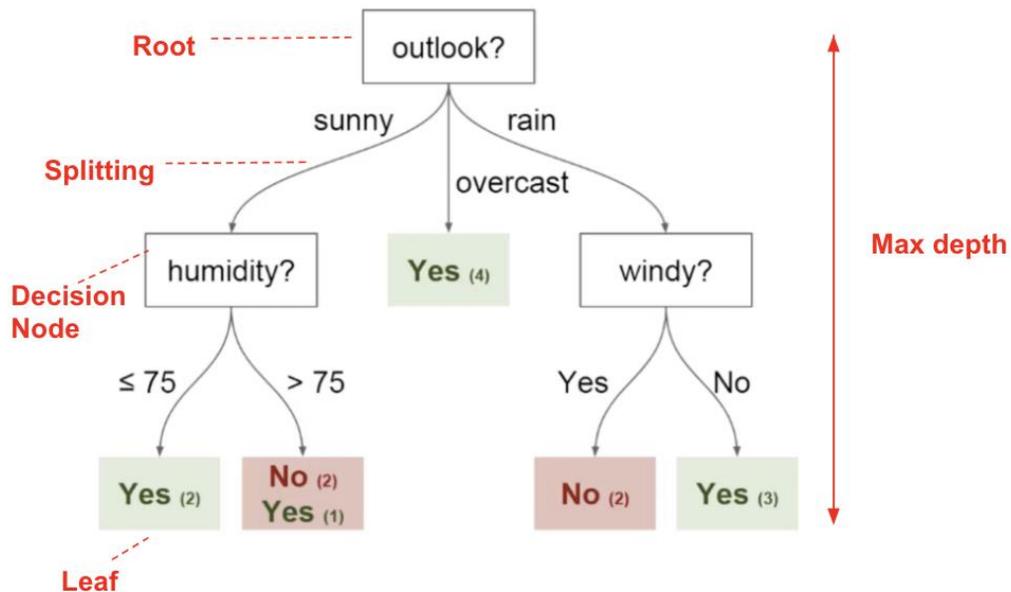


Figure III.7 : Principe de classification des arbres de décision (Google)

Le processus s'arrête quand les éléments d'un nœud ont la même valeur pour la variable cible (Homogénéité).

Il existe une implémentation alternative des arbres de décision, proposée par Breiman et al appelée CART, Les principales différences sont les suivantes :

- CART pose seulement de questions-test binaires (arbres binaires).

- Fonctionne aussi pour des attributs aux valeurs continues.
- CART cherche tous les attributs et tous les seuils pour trouver celui qui donne la meilleure homogénéité du découpage.

Quand un nœud S est coupé sur l'attribut j , seuil a_j , il donne naissance à deux descendants :

- Sous-nœud gauche S_g ($p_g \approx | S_g | / | S |$), qui contient tous les éléments qui ont une valeur de l'attribut $v_j \leq a_j$.
- Sous-nœud droit S_d ($p_d \approx | S_d | / | S |$), qui contient tous les éléments qui ont une valeur de l'attribut $v_j > a_j$.

Soit $I(S)$ une fonction qui mesure l'impureté de S par rapport à la classe cible. CART étudie le changement de l'impureté par rapport au seuil et pour tous les attributs :

$$\bullet E[I(S_g d)] = P_g I(S_g) + P_d I(S_d) \quad (\text{III.4})$$

$E[.]$ est l'opérateur de moyenne statistique.

$$\bullet \Delta I(s) = I(s) - E[I(S_g d)] = I(s) - P_g I(S_g) - P_d I(S_d) \quad (\text{III.5})$$

Le problème d'optimisation est le suivant :

$$\text{argmax}_{j; a_j} \Delta I(s)$$

CART choisit donc l'attribut et le seuil qui maximisent la décroissance de l'impureté du nœud par rapport à la cible. Pour un problème de régression on optimise le résidu quadratique moyen : minimise la variance moyenne des groupes.

Bien que les arbres de décision soient des outils assez robustes ils comportent tout de même des défauts :

- Parfois les arbres générés ne sont pas équilibrés (ce qui implique que le temps de parcours n'est plus logarithmique). Il est donc recommandé d'équilibrer la base de données avant la construction, pour éviter qu'une classe domine (en termes de nombre d'exemples d'apprentissage).
- Sur-apprentissage : parfois les arbres générés sont trop complexes et généralisent mal (solution : élagage, le contrôle de la profondeur de l'arbre et de la taille des feuilles).

- Ils sont instables : des changements légers dans les données produisent des arbres très différents. Les changements des nœuds proches de la racine affectent beaucoup l'arbre résultant. On dit que les arbres produisent des estimateurs de variance élevée.

Le besoin de répondre à ce troisième problème, qui n'admet pas de solution par optimisation algorithmique, a conduit aux approches de type Bagging et « Forêt aléatoires » (Bennaceur, 2019).

III.3.2 Random Forest

Le Bagging consiste à créer de nombreuses « copies » des données d'apprentissage (chaque copie est légèrement différente d'une autre), puis à appliquer un modèle à chaque copie pour obtenir plusieurs modèles dit « faibles », puis à les combiner. L'approche dite « bagging » est derrière l'algorithme d'apprentissage de forêt aléatoire. (Bennaceur, 2019).

L'algorithme « vanilla bagging » fonctionne comme suit. A partir d'un ensemble d'apprentissage, nous créons B échantillons aléatoires S_b (pour chaque $b = 1, \dots, B$) de l'ensemble d'apprentissage et construisons un modèle d'arbre de décision f en utilisant chaque échantillon S_b . Pour échantillonner S pour certains b , nous faisons l'échantillonnage avec remplacement. Cela signifie que nous commençons avec un ensemble vide, puis prenons au hasard un exemple dans l'ensemble d'apprentissage et plaçons sa copie exacte sur S_b en conservant l'exemple d'origine dans l'ensemble d'apprentissage initial. Nous continuons à choisir des exemples au hasard jusqu'à ce que $|S_b| = N$.

Après la période d'apprentissage, nous avons B arbres de décision. La prédiction pour un nouvel exemple x est obtenue sous forme de moyenne des prédictions B :

$$y \leftarrow \hat{f}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (\text{III.6})$$

L'algorithme des forêts aléatoire diffère de "vanilla bagging" d'une seule façon. Il utilise un algorithme d'apprentissage d'arborescence modifié qui inspecte, à chaque division du processus d'apprentissage, un sous-ensemble aléatoire de fonctions.

La raison est d'éviter la corrélation entre l'arbre : si une ou plusieurs paramètres d'entrée sont des prédicteurs très puissants pour la cible, elles seront sélectionnées pour fractionner des observations dans de nombreux arbres. Cela entraînerait de nombreux arbres corrélés dans notre

« forêt ». Les prédicteurs corrélés ne peuvent pas aider à améliorer la précision des prévisions. L'algorithme des forêts aléatoire est l'un des algorithmes d'apprentissage d'ensemble les plus largement utilisés. La raison en est qu'en utilisant plusieurs échantillons du jeu de données original, nous réduisons la variance du modèle final. (Bennaceur, 2019).

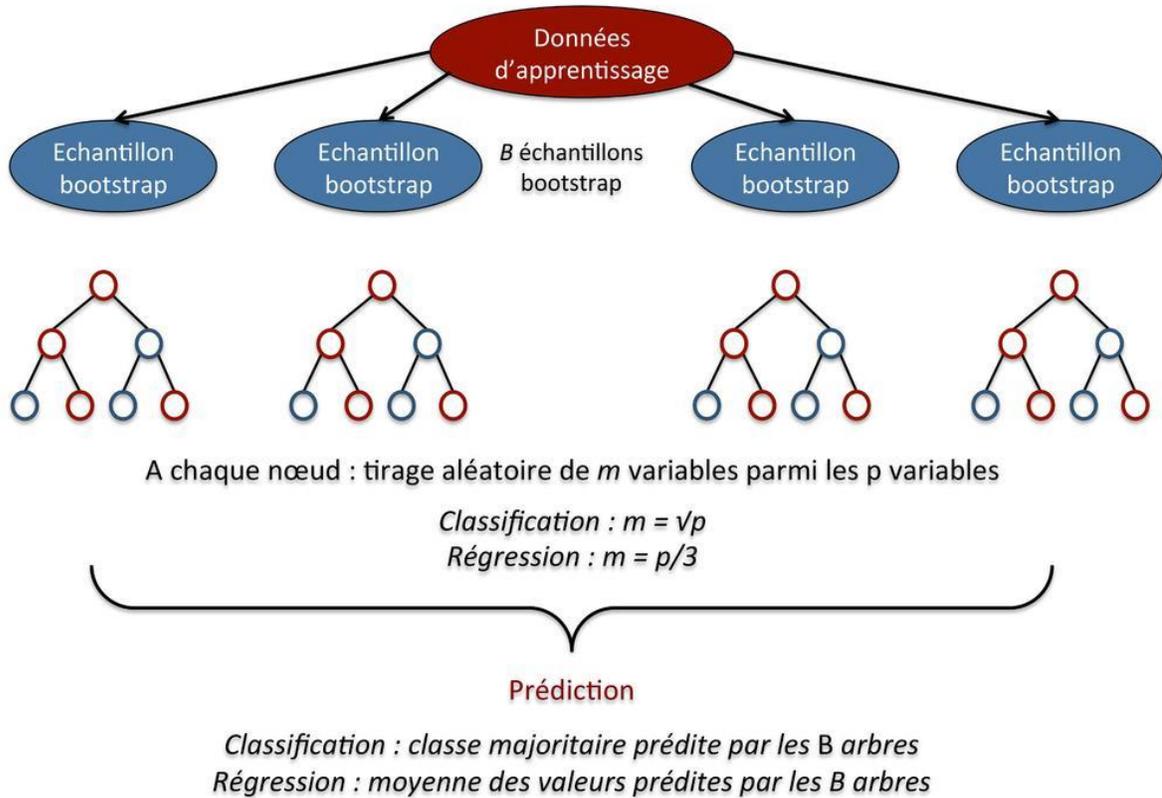


Figure III.8 : Principe de fonctionnement des forêts aléatoires (Google)

III.3.3 Gradient Boosting

Un autre algorithme d'apprentissage d'ensemble efficace est le "Gradient Boosting" (Bennaceur, 2019). Pour construire un modèle fort, nous commençons avec un modèle dit "constant" $f = f_0$ générée en utilisant l'algorithme des forêts aléatoire :

$$f = f_0(x) = \frac{1}{N} \sum_{i=1}^N y_i \quad (\text{III.7})$$

Puis on calcul la différence entre la valeur réelle et estimée, et on affect le résultat a une autre variable appelée résidus :

$$\hat{y}_i \leftarrow y_i - f(x_i) \tag{III.8}$$

La prochaine étape est de construire un autre modèle f_1 en utilisant le résidu comme sortie :

$$f = f_0 + \alpha f_1 \tag{III.9}$$

Puis on recalcule les résidus du modèle f_1 , et on les réutilise pour construire un autre modèle f_2 . Et on réitère le processus jusqu'à ce que les arbres de chaque modèle se combinent.

En calculant les résidus, nous estimons dans quelle mesure la sortie ou la variable à modélisé de chaque exemple d'entraînement est prédite par le modèle f . Nous formons ensuite un autre arbre pour corriger les erreurs du modèle actuel et ajoutons ce nouvel arbre au modèle existant avec un certain poids.

Par conséquent, chaque arbre supplémentaire ajouté au modèle corrige partiellement les erreurs commises par les arbres précédents jusqu'à ce que le nombre maximum d'arbres soit combinés.

Les trois principaux hyperparamètres sont le nombre d'arbres, le taux d'apprentissage et la profondeur des arbres. La profondeur des arbres influe également sur la vitesse d'entraînement et de prédiction.

On peut montrer que la formation sur les résidus optimise le modèle global f pour le critère d'erreur quadratique moyenne. La différence avec le bagging est que le boosting réduit le biais (ou sous-ajustement) au lieu de la variance, et cela peut conduire à un sur-apprentissage, cependant, en optimisant la profondeur et le nombre d'arbres, cela peut être largement évité (Bennaceur, 2019).

Chapitre IV :

Résultat et discussions

IV.1 Introduction

Dans ce chapitre, on a procédé à la modélisation du transport solide par les techniques de Machine Learning, pour cela, on a donc élaboré deux modèles débit liquide-débit solide : un par les réseaux de neurones artificiels et l'autre par le Gradient Boosting.

On a comparé entre les valeurs obtenues par les deux modèles appliqués sur les données de la Medjerda ensuite sur les données de Colorado, et on a fait une déduction par rapport à la qualité de nos données et à la puissance des deux méthodes.

Enfin, on a comparé et choisi un des deux modèles (RNA ou Gradient Boosting) en mettant en avant la relation débit liquide - débit solide. Le modèle retenu sera celui qui présentera le meilleur résultat, donc celui possédant un pouvoir prédictif élevé.

IV.2 Critères d'évaluation d'un modèle

Pour évaluer et discuter de la fiabilité et de la performance d'un modèle, plusieurs critères d'évaluation ont été développés, Ils peuvent être utilisés aussi bien en calibration qu'en validation.

IV.2.1 Critères graphiques

Pour notre cas, nous allons juger la performance du modèle par deux critères graphiques :

- **Courbe d'évolution des variables étudiées** : On compare, l'allure de la courbe débit solide observé vs temps, avec l'allure de la courbe débit solide mesuré vs temps, pour apprécier la différence entre les valeurs simulées et observées.
- **Courbe de corrélation (Q - Qplot)**: On trace le débit solide simulé en fonction du débit solide observé, si le nuage de point est bien réparti par rapport à la première bissectrice on peut conclure que notre modèle donne de bons résultats.

IV.2.2 Critères numériques

Une simple visualisation des résultats d'une simulation ne représente en aucun cas une évaluation complète d'un modèle. En effet, l'œil humain est incapable de voir toutes les différences qui peuvent exister entre deux courbes d'allure similaire. Ainsi, nous avons besoin en plus de l'analyse graphique de nous outiller de critères objectifs (dénommés aussi fonctions « objectif ») pour établir une comparaison entre les simulations et les observations, aussi pour hiérarchiser des simulations qui ont l'air équivalentes. Beaucoup de critères objectifs mathématiques existent pour évaluer la distance entre les observations et la simulation.

La performance de notre modèle est évaluée par le biais de quatre critères :

1. Le coefficient de détermination (R^2).
2. Le pourcentage de biais (PBIAS).
3. Le critère de Nash Sutcliffe (NSE).
4. RSR (Ratio of the root mean square error to the standard deviation of measured).

IV.2.2.1 Le coefficient de détermination (R^2)

Défini comme étant la valeur élevée au carré du coefficient de corrélation de Pearson. Donc il estime la dispersion entre la série des observées et des simulées, plus il sera important plus l'accord entre les deux séries de valeurs le sera aussi.

Il nous informe également sur le degré de corrélation entre deux variables à étudier (pour notre cas entre le débit liquide et le débit solide), si sa valeur est égale à « 0 » cela signifie qu'il n'y a aucune corrélation entre les deux variables.

La valeur maximale est de « 1 » et elle représente la valeur optimale. Le R^2 est donné par la relation :

$$R^2 = \frac{\sum_{t=1}^T (Q_0^t - \overline{Q_0}) * (Q_m^t - \overline{Q_m})}{\sqrt{\sum_{t=1}^T (Q_0^t - \overline{Q_0})^2} * \sqrt{\sum_{t=1}^T (Q_m^t - \overline{Q_m})^2}} \quad (IV.1)$$

L'inconvénient du R^2 est qu'il quantifie la dispersion des points "observés-simulés" et ne donne aucune information sur la qualité du modèle, donc même si les valeurs simulées sont sous-estimées ou surestimées mais que leur dispersion est identique à celle des observées, on aura un R^2 proche de « 1 ».

Q_0 : représente les valeurs observées.

Q_m : représente les valeurs simulées.

Q_0^t : représente les valeurs observées à l'instant t.

Q_m^t : représente les valeurs simulées à l'instant t.

$\overline{Q_0}$: La moyenne des valeurs observées.

$\overline{Q_m}$: La moyenne des valeurs simulées.

IV.2.2.2 Le pourcentage de biais (PBIAS)

Le pourcentage du biais (PBIAS) mesure la tendance moyenne des données simulées à être plus

grandes ou plus petites que leurs homologues observées. Le PBIAS prend sa valeur optimale quand il est égal à zéro.

Si le PBIAS <0 il indique que les valeurs simulées sont plus élevées que celle observés, ainsi on peut dire qu'il y'a une surestimation des valeurs.

Si le PBIAS >0 il indique que les valeurs simulées sont inférieures à celle observées, ainsi on peut dire qu'il y'a une sous-estimation des valeurs par le modèle.

$$PBIAS = \frac{\sum_{t=1}^T (Q_0^t - Q_m^t) * 100}{\sum_{t=1}^T Q_0^t}$$

(IV.2)

IV.2.2.3 Le critère de Nash Sutcliffe (NSE)

Utilisé en hydrologie, il fournit une indication sur la robustesse (pouvoir prédictif) des modèles utilisés. Ce qui nous permet de l'utiliser comme un outil d'aide à la décision. Le critère est le suivant :

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_0^t - Q_m^t)^2}{\sum_{t=1}^T (Q_0^t - \bar{Q}_0)^2} \quad (IV.3)$$

La valeur de NSE varie de « -∞ » à « 1 » et est généralement présentée en pourcentage.

- Une efficacité de NSE =1 correspond à une modélisation parfaite où les simulées = les observées.
- Une efficacité de NSE = 0 indique que les prédictions du modèle sont aussi précises que la moyenne des données observées.
- Une efficacité de NSE < 0 indique que les valeurs simulées conduisent à de moins bons résultats et que la moyenne observée est meilleur prédicateur que le modèle.

Le critère de Nash a pour inconvénient la surestimation de l'erreur des grandes valeurs (pics) contrairement aux petites valeurs qui sont négligées, c'est pour ça on utilise le Nash modifié qui est moins sensible aux extrêmes.

$$E = 1 - \frac{\sum_{t=1}^T |Q_0^t - Q_m^t|}{\sum_{t=1}^T |Q_0^t - \bar{Q}_0|} \quad (IV.4)$$

IV.2.2.4 L'erreur quadratique moyenne RMSE

Nous utilisons souvent la mesure de l'erreur quadratique moyenne comme pour mesurer l'écart entre la simulation et l'observation, en donnant notamment une idée de la dispersion entre les deux et cela dans l'unité des variables étudiées. Plus le RMSE sera faible et plus l'écart entre la simulation et les observations sera faible.

Le critère est le suivant :

$$RMSE = \sqrt{\frac{1}{n_0} \sum_{t=1}^T (Q_m^t - Q_0^t)^2} \quad (IV.5)$$

Le RMSE est généralement utilisé pour décrire l'erreur statistique. Bien que la valeur minimale

du RMSE retienne la performance du modèle, mais des études statistiques (établies par SINGH et al, 2004) l'ont normalisé et ont développé le RSR (ratio of the Root Mean Square Error to the standard deviation of measured). Il est calculé par le rapport entre l'erreur quadratique moyenne et l'écart type des données mesurées comme le montre l'équation :

$$RSR = \frac{RMSE}{STDEV_{OBS}} = \frac{\sqrt{\sum_{t=1}^T (Q_m^t - Q_0^t)^2}}{\sqrt{\sum_{t=1}^T (Q_0^t - \bar{Q}_0)^2}} \quad (IV.6)$$

Le RSR varie entre la valeur optimale qui est égale à zéro et un nombre très grand.

Le taux de performance des critères

Selon une étude sur l'évaluation de modèles pour la quantification systématique de la précision dans les simulations de bassins versants (Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations) les critères définis ci-dessus sont les meilleurs critères pour juger de la performance d'un modèle. Ainsi, ils seront tous retenus pour valider les modèles utilisés dans notre travail.

Un modèle est jugé satisfaisant si :

$$\left\{ \begin{array}{l} NSE \geq 50\% \\ RSR \leq 0.70 \\ R^2 \geq 50\% \\ PBIAS < \pm 55\% \end{array} \right\}$$

IV.3 Critique des données

Pour une période de mesure de 24 ans, on a récupéré une série de données de 1428 valeurs instantanées de débit liquide (Q_1 en m^3/s) et de concentration (C en g/l) qui sont présentées dans les deux figures suivantes :

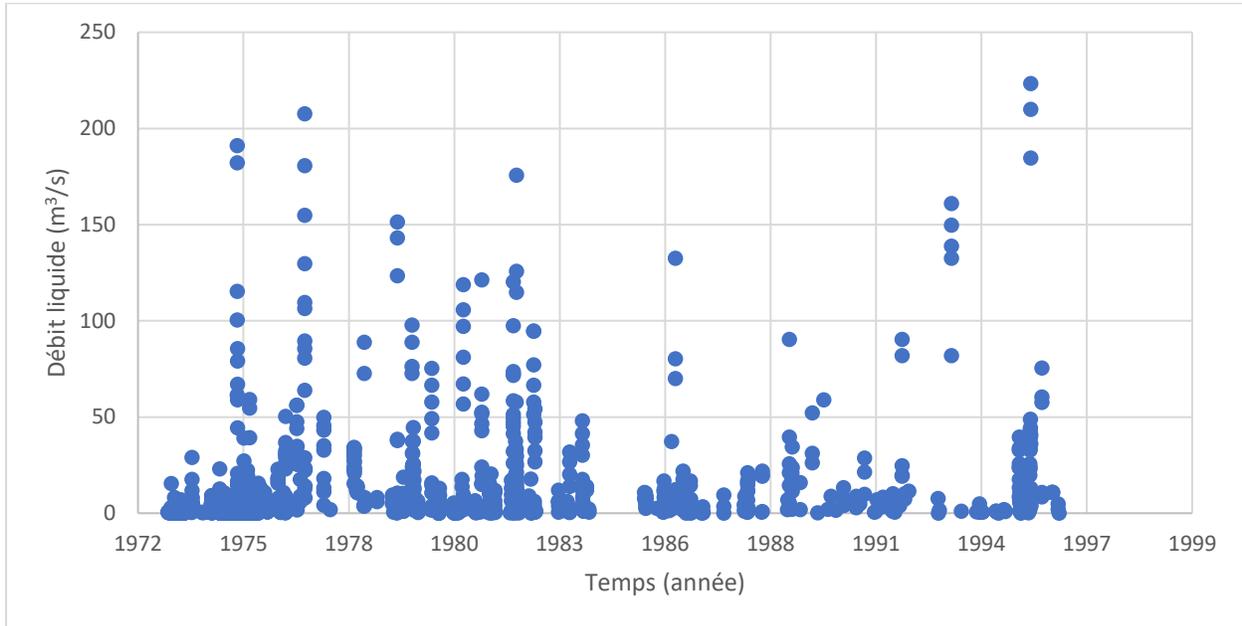


Figure IV.1 : Evolution du débit liquide observé au bassin de la Medjerda.

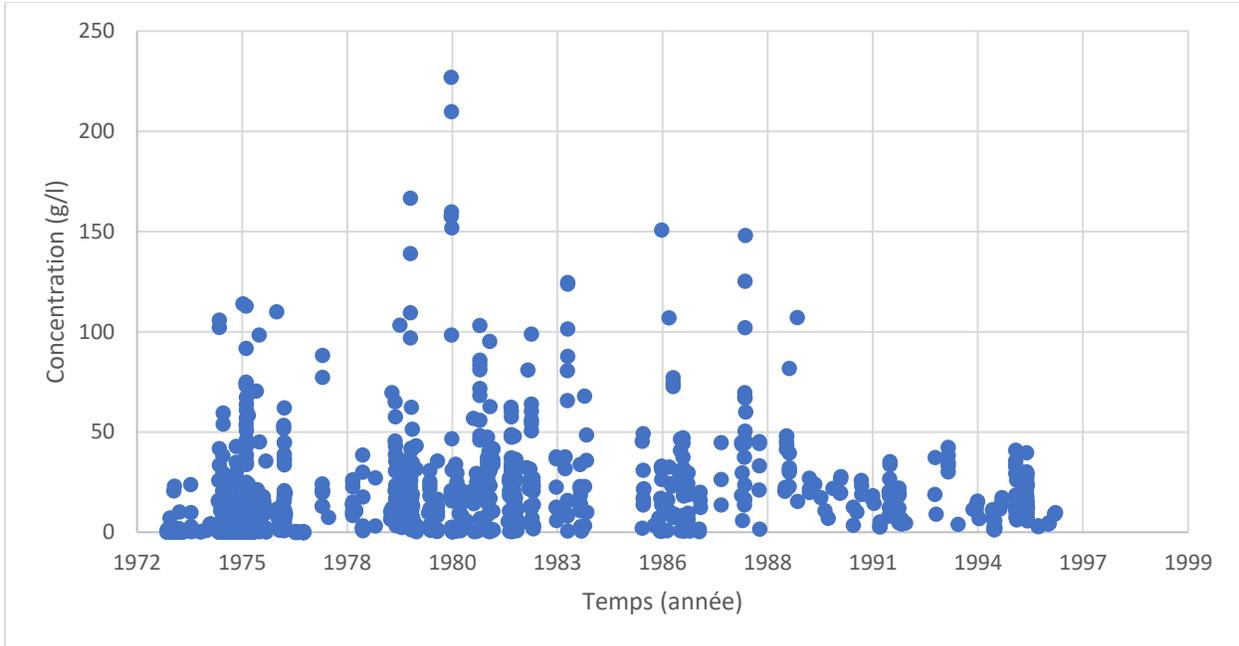


Figure IV.2 : Evolution de la concentration observée au bassin de la Medjerda.

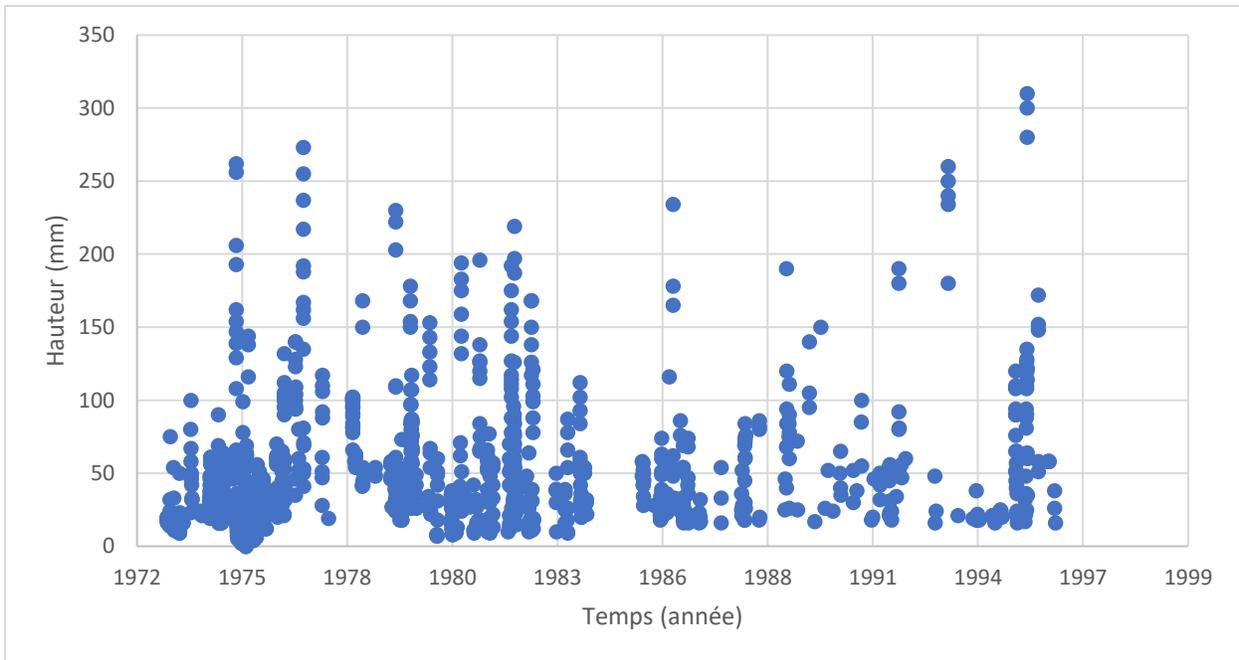


Figure IV.3 : Evolution de la hauteur observée du bassin de la Medjerda.

Pour avoir le débit solide, on fait une multiplication entre le débit liquide observé et la concentration. Les valeurs calculées sont représentées dans la figure suivante :

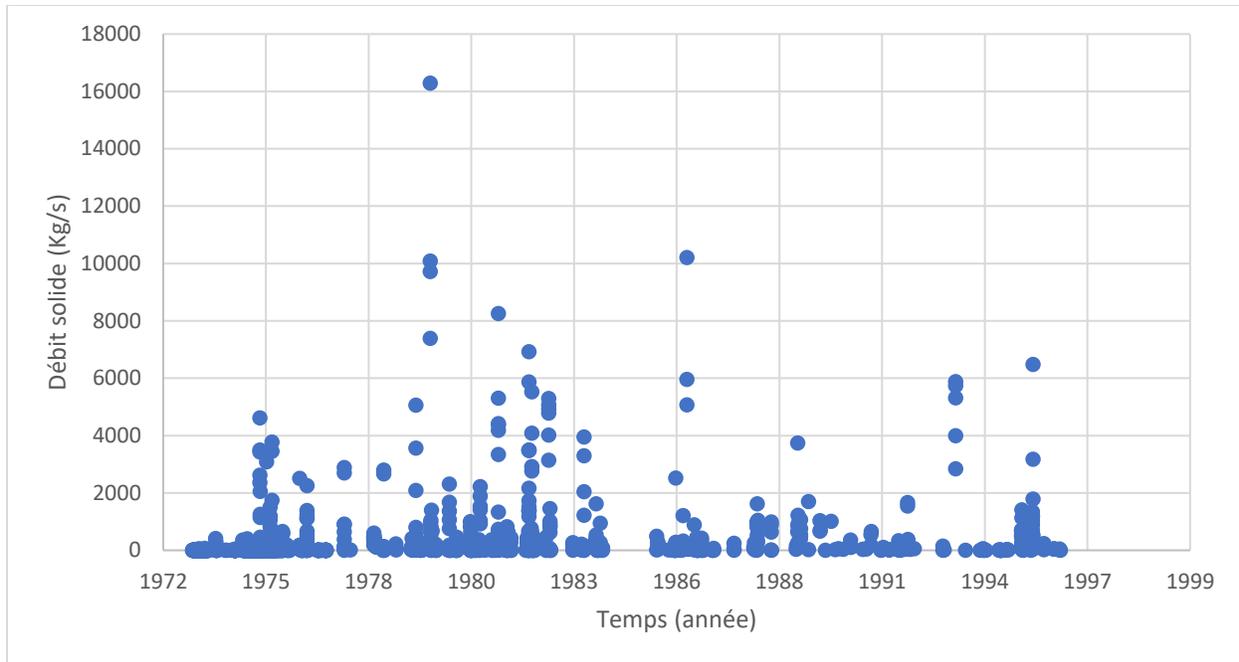


Figure IV.4 : Evolution de débit solide du bassin de la Medjerda.

On a travaillé avec les données journalières, c'est pour ça qu'on a calculé la moyenne des valeurs instantanées lorsqu'il y a plusieurs mesures dans la journée.

Pendant cette période de 24 ans, nous avons 924 données journalières sur 8453 jours qui représentent 11%, ce qui montre le manque de données flagrant à notre disposition.

IV.3.1 Traitement statistique des données

Plusieurs facteurs font que les données mesurées contiennent des valeurs erronées, celle-ci sont appelées valeurs aberrantes. Afin d'assurer l'adéquation de notre modèle il faut les éliminer, et pour cela les hydrologues proposent plusieurs tests statistiques. Pour notre cas nous avons utilisé le test de Grubbs à l'aide du logiciel XLSTAT et qui a donné les résultats suivants :

- Pour le débit solide, toutes les valeurs supérieures à 3600 Kg/s ou inférieures à 0.001Kg/s sont à rejeter.
- Pour le débit liquide, aucune valeur n'est douteuse.

IV.4 Application des modèles Machine Learning au bassin versant de la Medjerda

Dans tout ce qui suit, le choix des données pour les deux parties calage et validation se fait de la manière suivante :

- 70% de la série des observées pour le calage.

- 30% de la série des observées pour la validation.

Vu le nombre réduit de données on s'est proposé de tester cinq cas de figure afin de trouver le meilleur modèle. La différence entre eux c'est le nombre des entrées :

1^{er} cas : on suppose que le débit solide est en relation directe avec le débit liquide, sans prendre en considération l'influence des débits liquides et solides des jours précédents, c'est-à-dire que le débit solide mesuré au jour « j » n'est fonction que débit liquide du même jour.

2^{ème} cas : le débit solide mesuré au jour « j » est fonction du débit liquide du même jour et de celui du jour précédent (jour « j-1 »).

3^{ème} cas : le débit solide mesuré au jour « j » est fonction des débits liquide des jours « j », « j-1 » et « j-2 ».

4^{ème} cas : le débit solide mesuré au jour « j » est fonction des débits liquide des jours « j », « j-1 » et du débit solide du jour « j-1 ».

5^{ème} cas : le débit solide mesuré au jour « j » est fonction des débits liquide des jours « j », « j-1 », « j-2 » et des débits solides des jours « j-1 » et « j-2 ».

IV.4.1 Réseaux de Neurones Artificiels

On a conçu un réseau de neurones et on a fait plusieurs essais pour l'optimiser et ce pour chaque cas mentionné précédemment.

Le meilleur modèle obtenu correspond au modèle du 5^{ème} cas, avec 5 couches cachées et chaque couche contient 50 neurones, avec une fonction sigmoïde comme fonction d'activation.

Il faut noter que le nombre de données a diminué à 597 valeurs en associant chaque débit liquide aux débits solides des jours précédents à cause de la discontinuité et des lacunes présentes.

Les résultats obtenus sont résumés dans le tableau suivant :

Tableau IV.1 : Les critères d'évaluation numériques pour le RNA de Medjerda.

	Calage	Validation
R ²	0.20	0.25
RSR	0.89	0.86
PBIAS	-329.31	34.18
NASH	0.20	0.25

D'après le tableau IV.1, on remarque bien que les critères d'évaluations sont mauvais. Pour le calage du modèle, le Nash avec une valeur faible de 20%, un RSR dépassant la valeur

satisfaisante qui est 0.7, un critère de PBIAS qui surestime le modèle (les valeurs simulées sont plus élevées que celles observées).

En ce qui concerne la validation du modèle, les résultats obtenus ne sont pas satisfaisants aussi, le pouvoir prédictif du modèle est faible avec une valeur de 25%, un RSR de 86% et un coefficient de détermination égale à 0.25, par contre le PBIAS est dans les limites acceptables avec une sous-estimation des valeurs simulées.

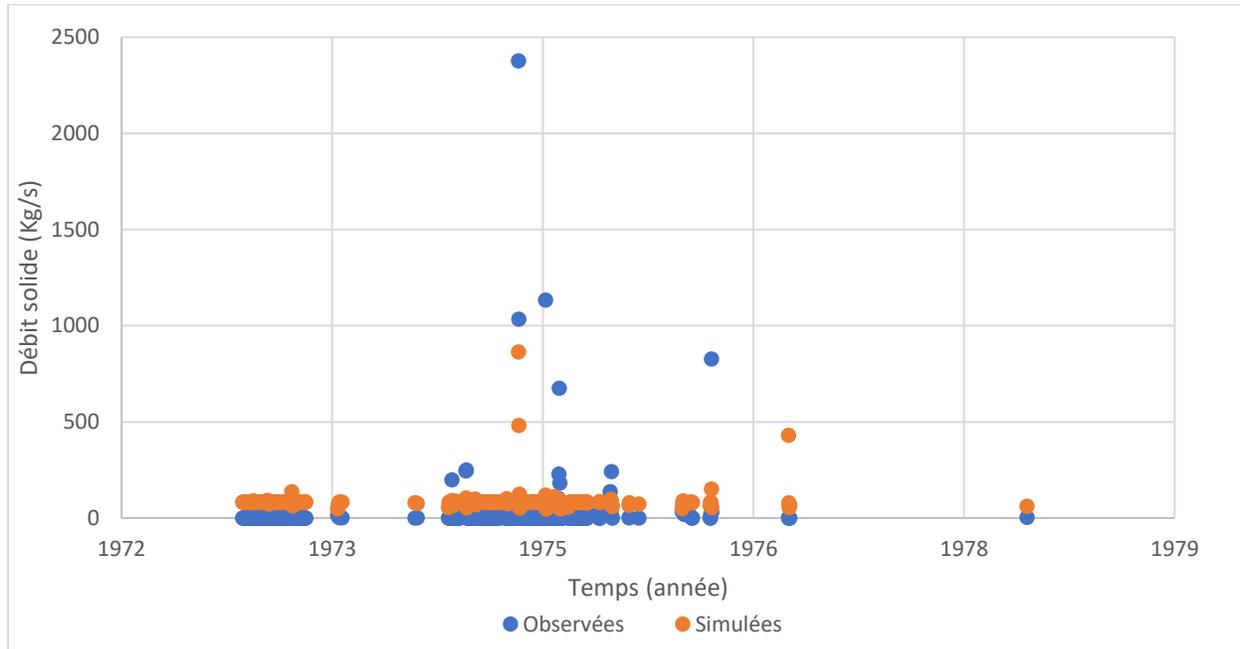


Figure IV.5 : Evolution des débits solides (mesuré et simulé) en période de calage (Medjerda, RNA). (450 valeurs)

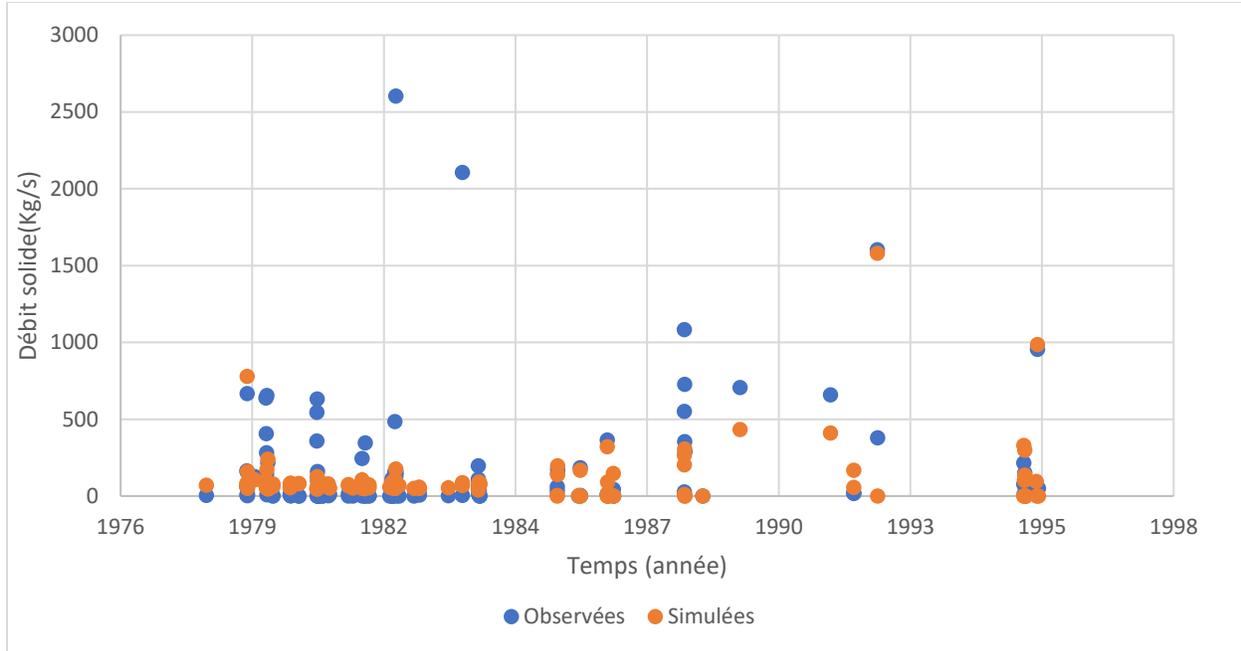


Figure IV.6 : Evolution des débits solides (mesuré et simulé) en période de validation (Medjerda, RNA). (147 valeurs)

Les figures IV.5 et IV.6 montrent que les débits simulés n'évoluent pas de la même manière que les débits observés et qu'ils n'arrivent à reproduire les pics.

Les résultats graphiques obtenus à partir des valeurs simulées-observées sont les suivants :

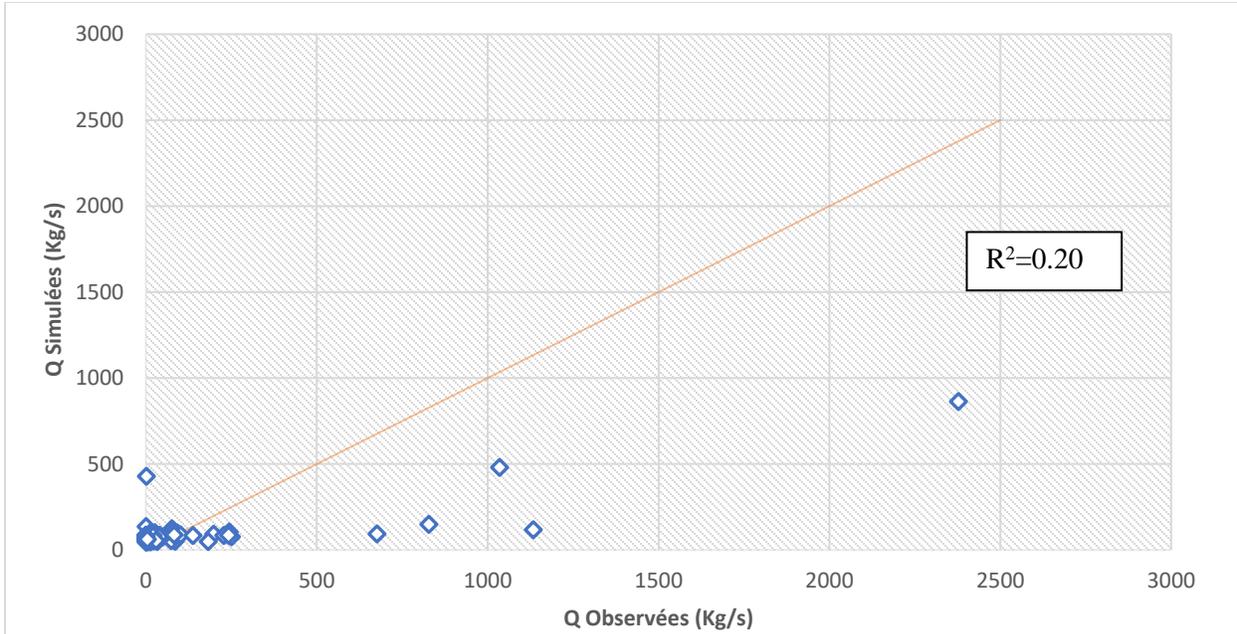


Figure IV.7 : Représentation des débits solides simulés-mesurés en période de calage (Medjerda, RNA).

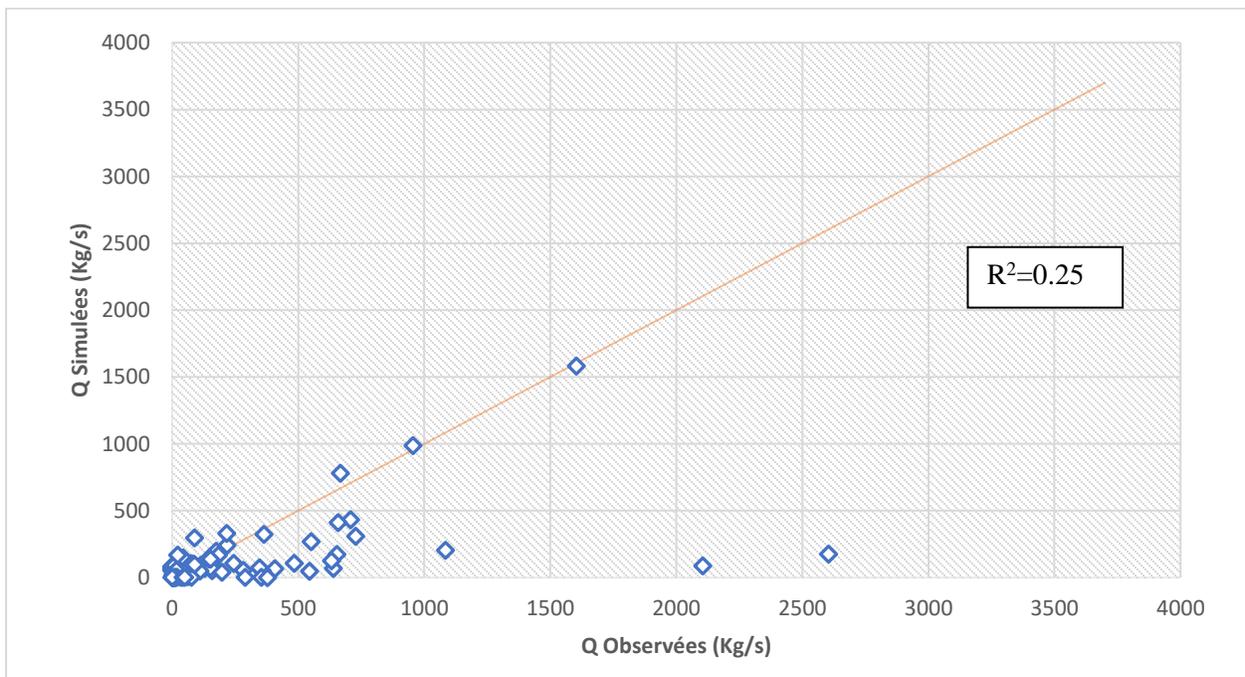


Figure IV.8 : Représentation des débits solides simulés-mesurés en période de validation (Medjerda, RNA).

Les figures IV.7 et IV.8 permettent clairement de conclure que la relation qui lie les débits solides simulés et observés n'est pas linéaire.

Les quatre critères n'étant pas satisfaits, on peut dire que ce modèle est à rejeter car il n'arrive pas à reproduire la relation qui lie le débit liquide au débit solide.

IV.4.2 Gradient Boosting

Les hyperparamètres optimisés sont :

- Le pourcentage du nombre d'inputs choisi, 083.
- Le taux d'apprentissage, 0,089.
- La profondeur maximale d'un arbre de décision, 4.
- Alpha, paramètre de la technique de régularisation L1, qui permet au modèle de ne pas diverger, ou d'être sujet à un sur-apprentissage, 1.
- Le nombre d'estimateurs utilisés, 63.

Tableau IV.2 : Les critères d'évaluation numériques pour le Gradient Boosting Medjerda.

	Calage	Validation
R ²	0.47	0.53
PBIAS	-83.47	21.59
RSR	0.73	0.68
NASH	0.47	0.53

En ce qui concerne le calage du modèle avec le Gradient Boosting, on remarque que les résultats se sont légèrement améliorés, mais restent quand même non satisfaisant. D'après le tableau IV.2, le pouvoir prédictif de Nash ne dépasse pas les 50%, le RSR est de 0.73 dépassant la valeur satisfaisante et un PBIAS dépassant les limites acceptables avec une valeur négative de -83.47 qui évalue la tendance du modèle à la surestimation.

Pour la partie validation, les résultats sont satisfaisants, on voit bien que le coefficient de détermination entre les valeurs simulées et celles mesurées est bonne avec un R² égale à 53%, un PBIAS dans l'intervalle des valeurs acceptables avec une valeur de 21.59 qui sous-estime le modèle, un RSR de 0.68 et un pouvoir prédictif de 53%.

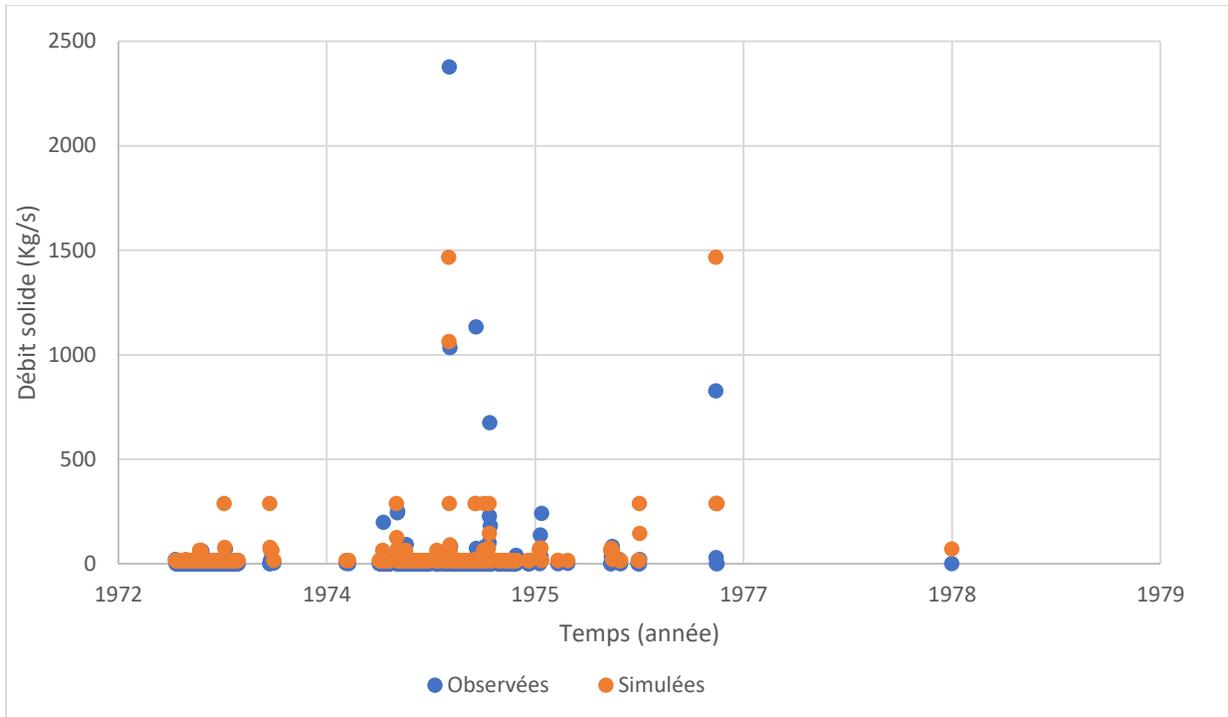


Figure IV.9 : Evolution des débits solides (mesuré et simulé) en période de calage (Medjerda, Gradient Boosting). (450 valeurs)

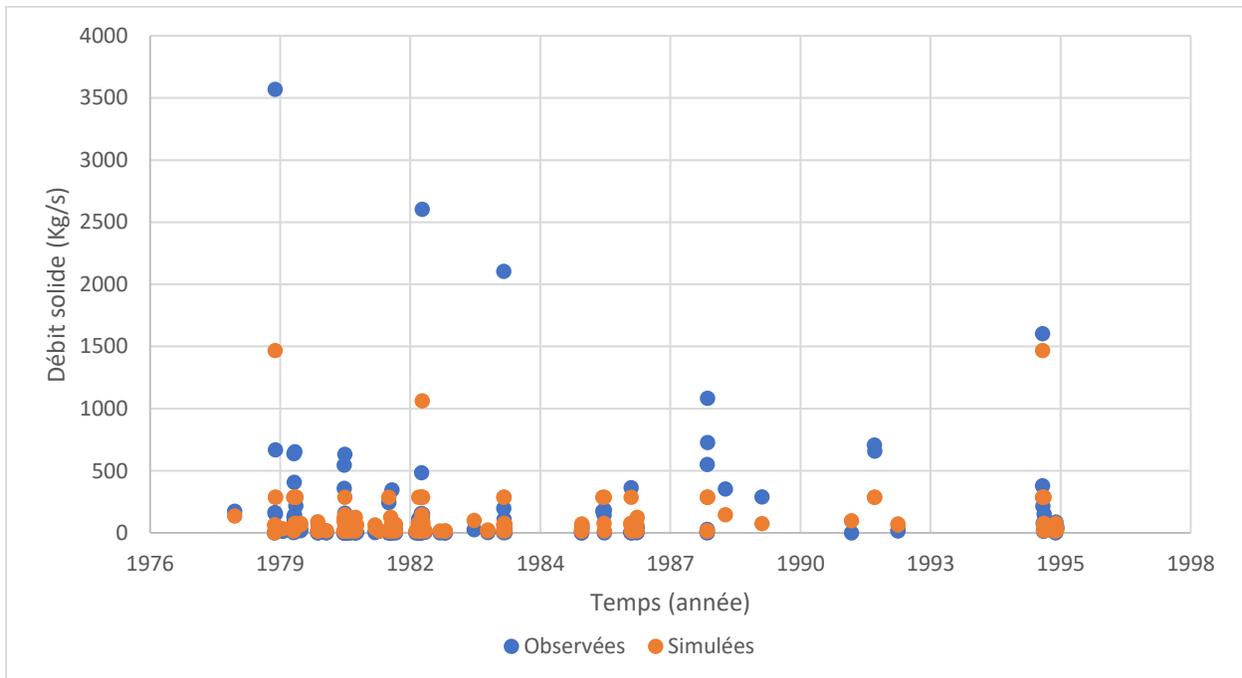


Figure IV.10 : Evolution des débits solides (mesuré et simulé) en période de validation (Medjerda, GB). (145 valeurs)

D'après les figures IV.9 et IV.10, on remarque que le modèle élaboré n'arrive pas à reproduire l'allure du phénomène étudié, et surtout les pics des valeurs observées.

Les graphes Q-Qplot entre les valeurs observées et simulées des débits solides sont donnés ci-dessous :

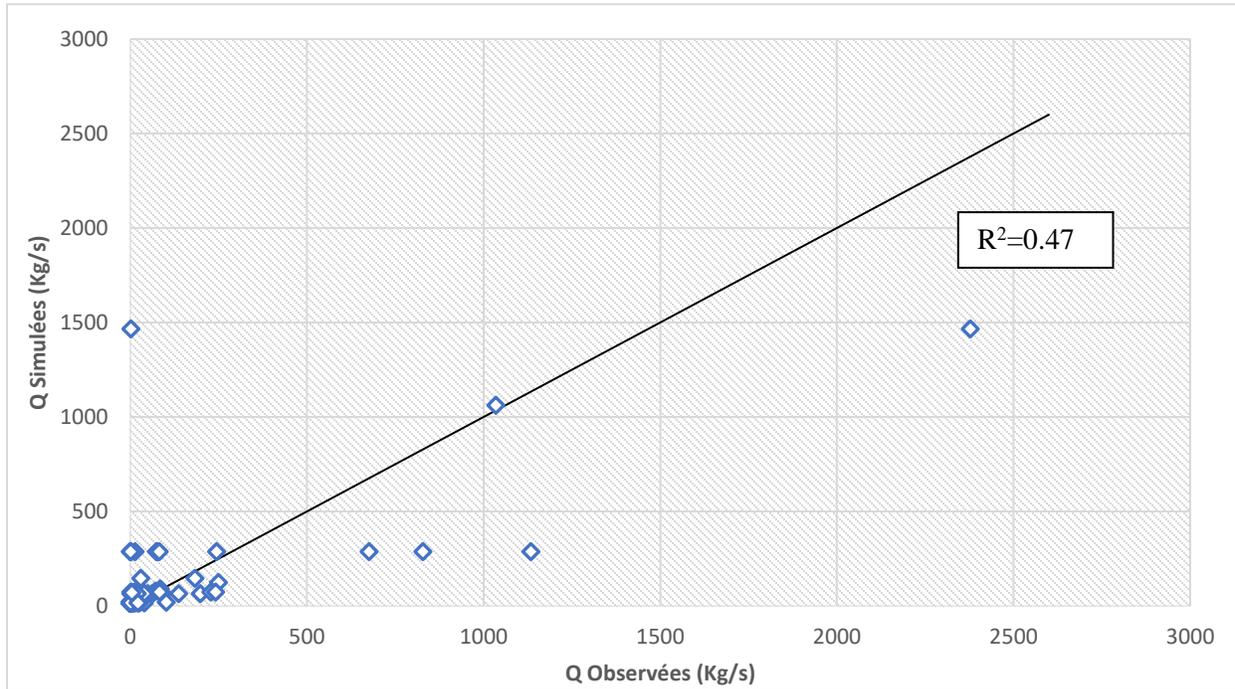


Figure IV.11 : Représentation des débits solides simulés et mesurés en période de calage (Medjerda, GB)

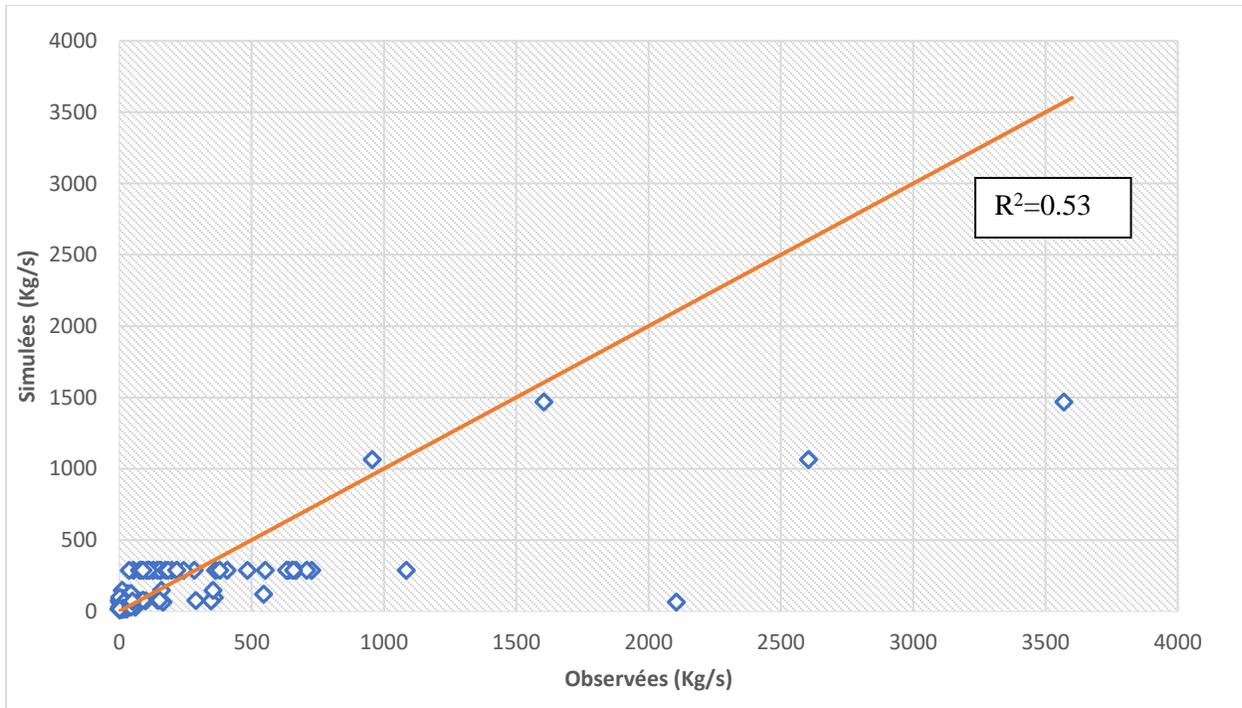


Figure IV.12 : Représentation des débits solides simulés et mesurés en période de validation (Medjerda, GB).

D'après les figures IV.11 et IV.12, la corrélation est bonne pour les faibles débits, mais pour les grandes valeurs, la corrélation est mauvaise avec un coefficient de détermination de R^2 de 53% en période de validation et 47% en période de calage.

IV.4.3 Interprétation des résultats

Les résultats obtenus par les deux modèles laissent à désirer surtout les réseaux de neurones qui ont donnés des résultats très médiocres.

Cela nous amène à dire que les nombreuses lacunes présentes dans la série sont la cause principale de ces mauvais résultats, et que les données recueillies de l'ANRH ne sont peut-être pas fiables.

Pour pouvoir confirmer cette hypothèse on se propose d'appliquer les mêmes modèles à une série de données beaucoup plus complète sans lacunes et plus fiable. On a choisi une station de mesures située aux Etats Unis dans le bassin du Colorado.

IV.5 Bassin de Colorado :

On a recueilli les données du site américain USGS :

https://waterdata.usgs.gov/nwis/inventory/?site_no=09251000&agency_cd=USGS

Voici le nom de la rivière et le code de la station de mesures :

USGS 09251000 YAMPA RIVER NEAR MAYBELL, CO.

Coordonnées : Latitude 40°30'10", Longitude 108°01'58"

Sa surface de drainage est de 777 ha.

On a récupéré une série de données débit liquide- débit solide contenant 1732 valeurs journalières sans aucune lacune du 01/01/1979 au 29/09/1983, elles sont représentées sur les figures suivantes :

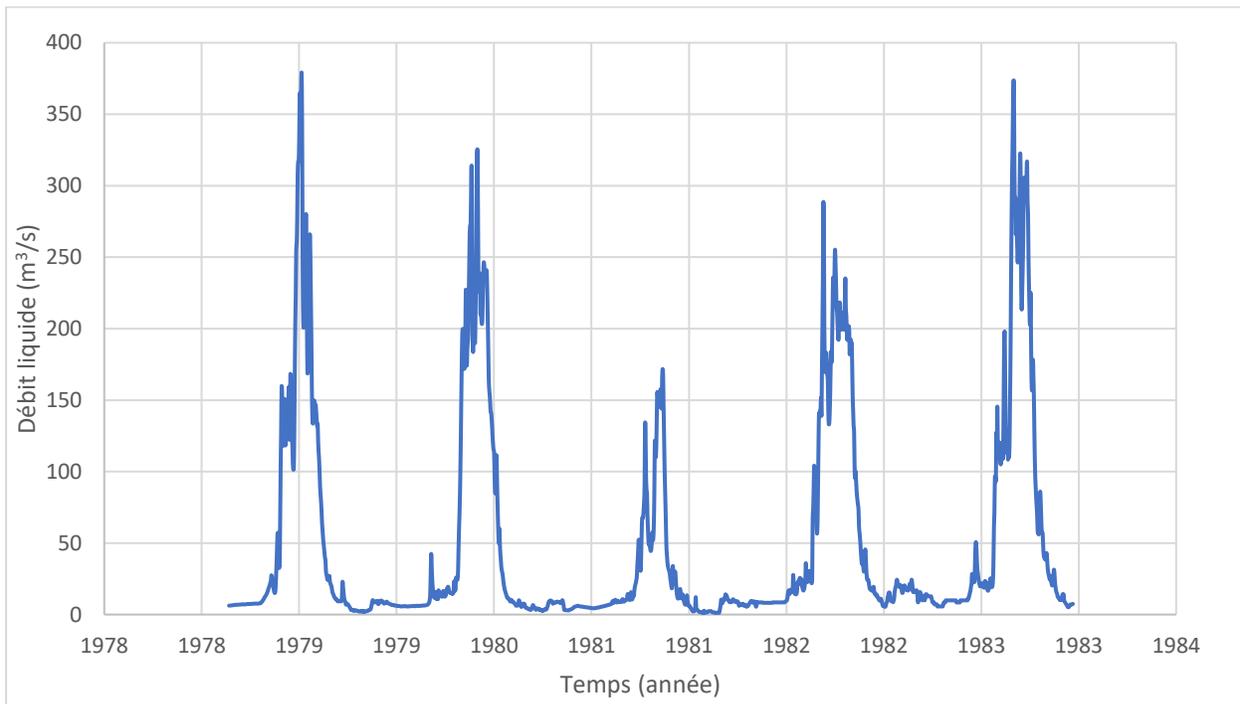


Figure IV.13 : Evolution du débit liquide observé du bassin de Colorado.

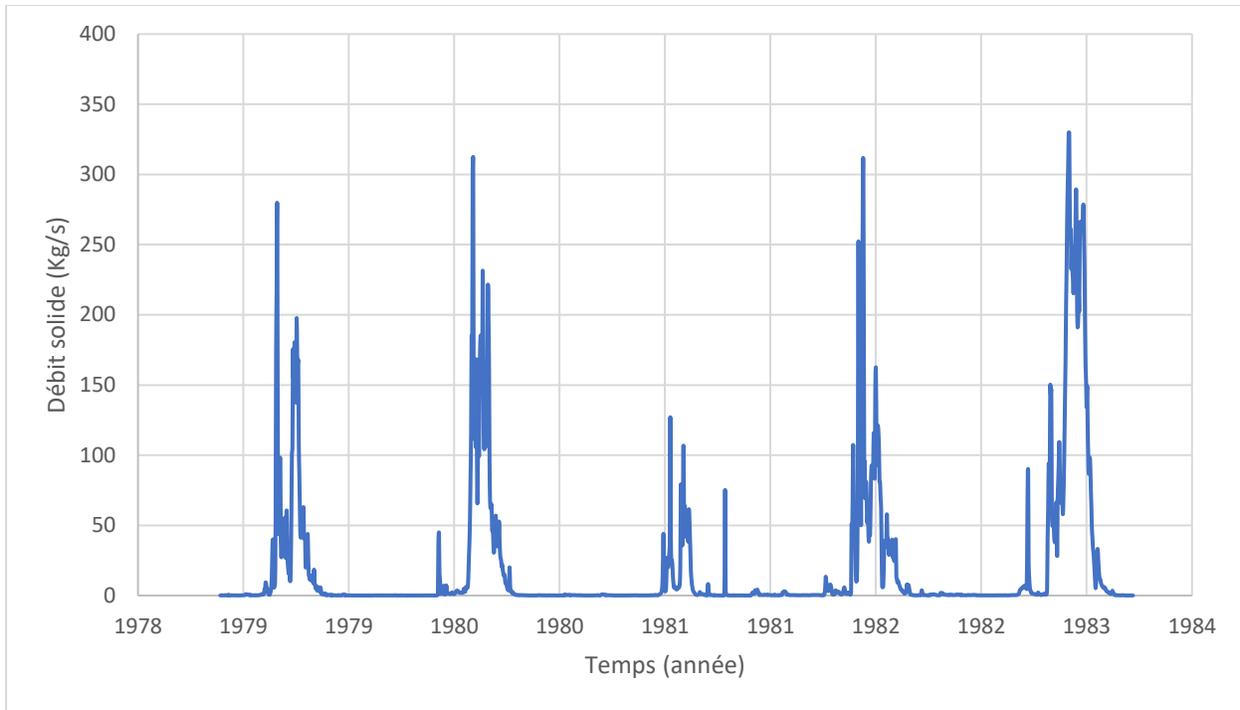


Figure IV.14 : Evolution du débit solide observé du bassin de Colorado.

IV.5.1 Application des modèles Machine Learning au bassin versant du Colorado

IV.5.1.1 Réseaux de Neurons Artificiels

On a conçu un réseau de neurones et on a fait plusieurs essais pour l'optimiser

Le meilleur modèle est le modèle constitué de 5 couches cachées et chaque couche contient 20 neurones, et on a pris en considération la fonction sigmoïde comme fonction d'activation.

Les résultats obtenus sont les suivants :

Tableau IV.3 : Les critères d'évaluation numériques pour le RNA Colorado.

	Calage	Validation
R²	0.96	0.97
RSR	0.21	0.18
NASH	0.96	0.97
PBIAS	1.43	-0.72

Les résultats obtenus pour le bassin de Colorado sont très bons. En effet, d'après le tableau IV.3, pour la partie calage, le R² avoisine les 96%, ce qui veut dire que la simulation donne des résultats presque parfaits, avec des valeurs simulées très proches de celles observées, une valeur

de RSR assez satisfaisante qui est égale à 0.21 (proche de 0 qui est la valeur optimale), un critère de PBIAS qui est égale à 1.43 (la valeur optimale c'est le 0).

Pour la partie validation, on remarque que les critères numériques sont aussi très bons, avec un Nash dépassant les 90%, un RSR de 0,18 et un PBIAS de -0.72.

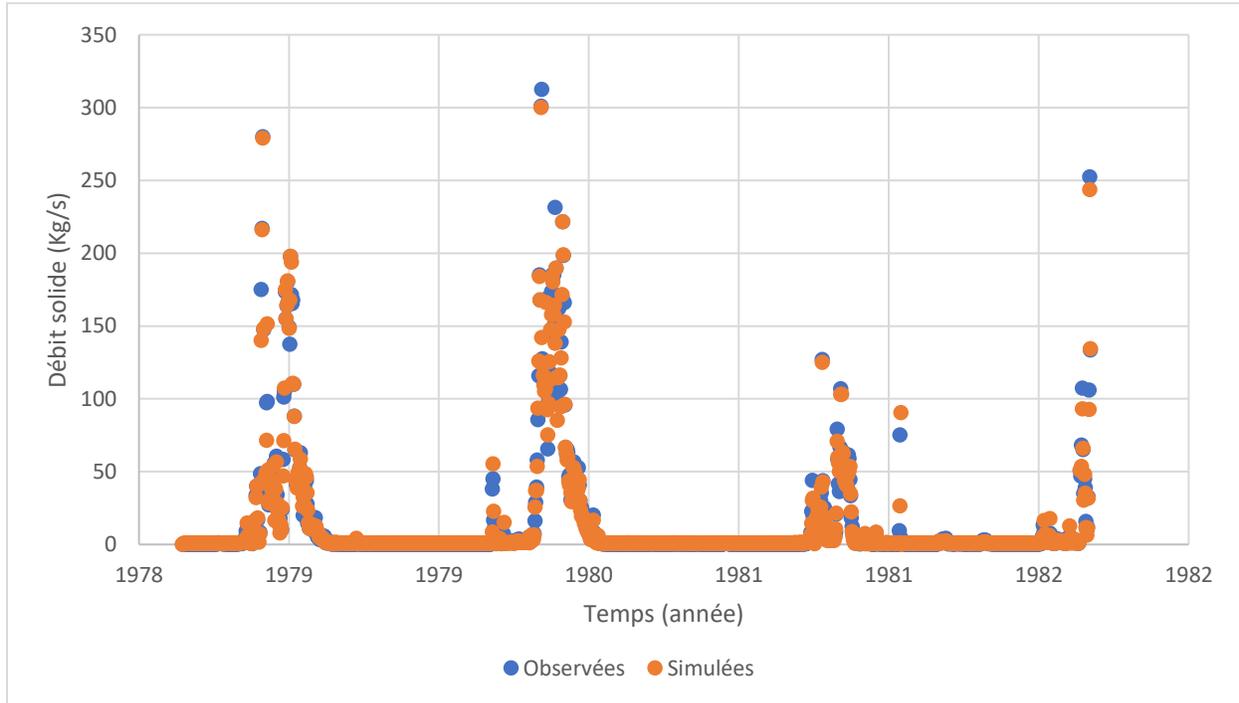


Figure IV.15 : Evolution des débits solides (mesuré et simulé) en période de calage (Colorado, RNA).

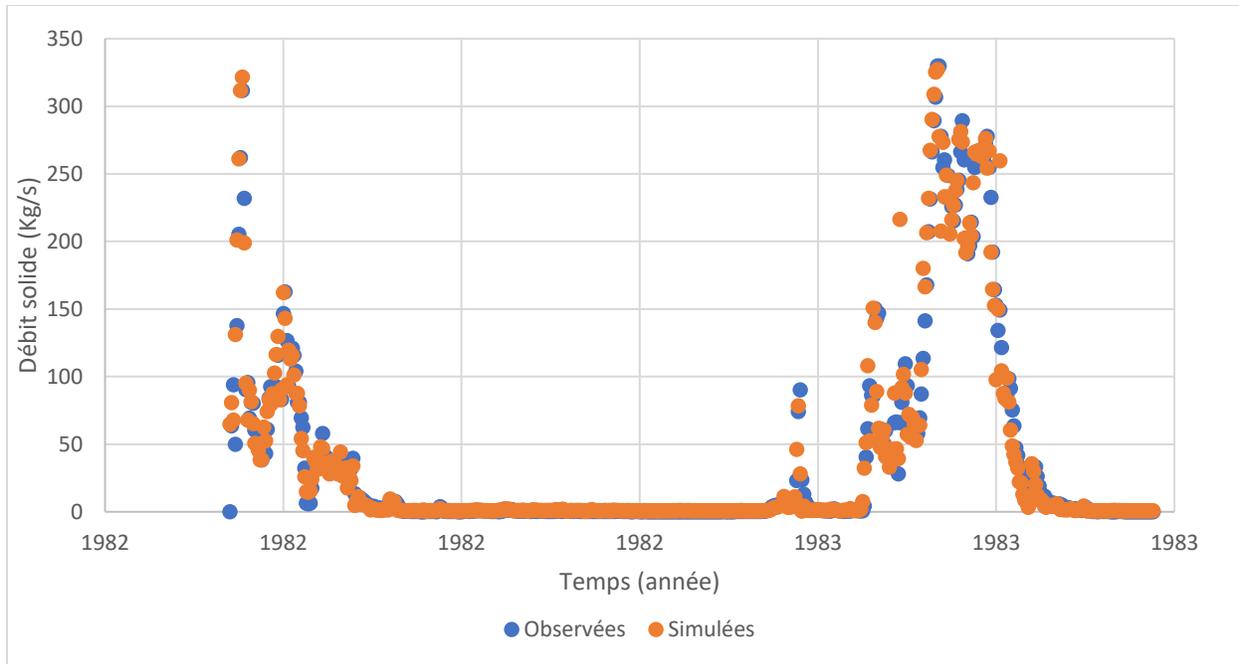


Figure IV.16 : Evolution des débits solides (mesuré et simulé) en période de validation (Colorado, RNA).

Comme on peut le voir sur les figures IV.15 et IV.16, le modèle élaboré reproduit bien l'allure du phénomène étudié, et arrive même à atteindre les pics des valeurs observées, que ce soit en calage ou en validation.

En outre les graphes Q-Qplot des figures IV.17 et IV.18, montrent une parfaite adéquation entre les valeurs simulées et observées que ce soit en calage ou en validation.

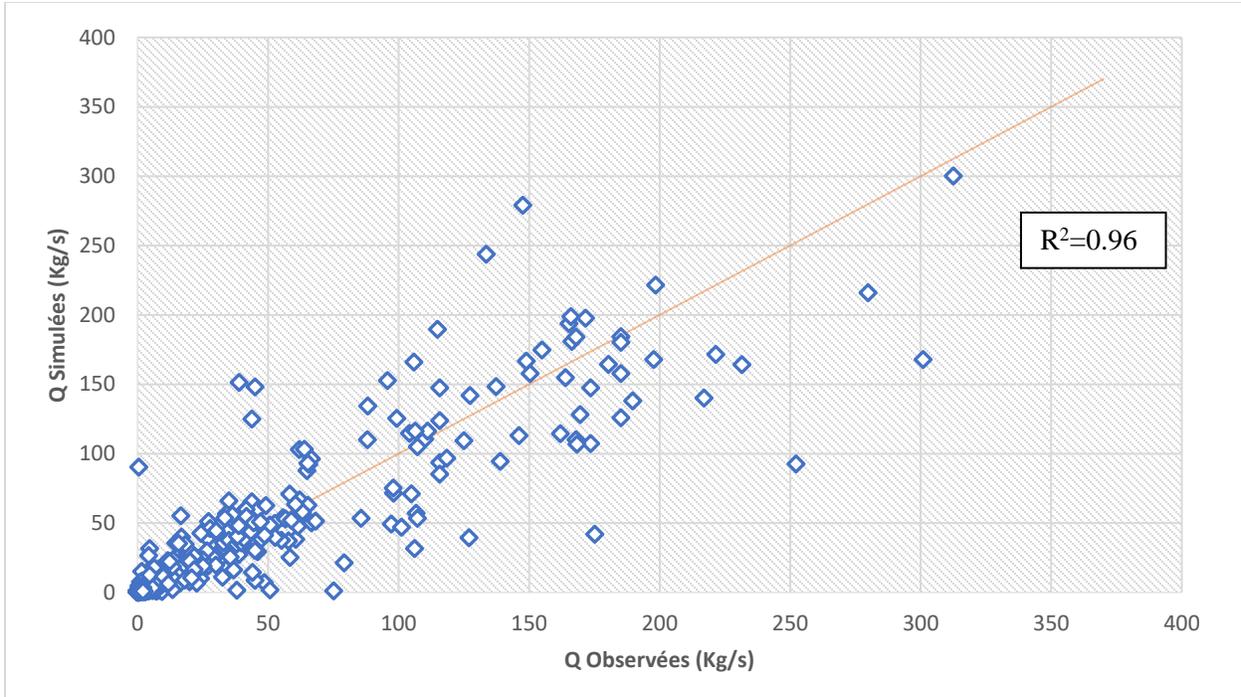


Figure IV.17 : Représentation des débits solides simulés et mesurés en période de calage (Colorado, RNA).

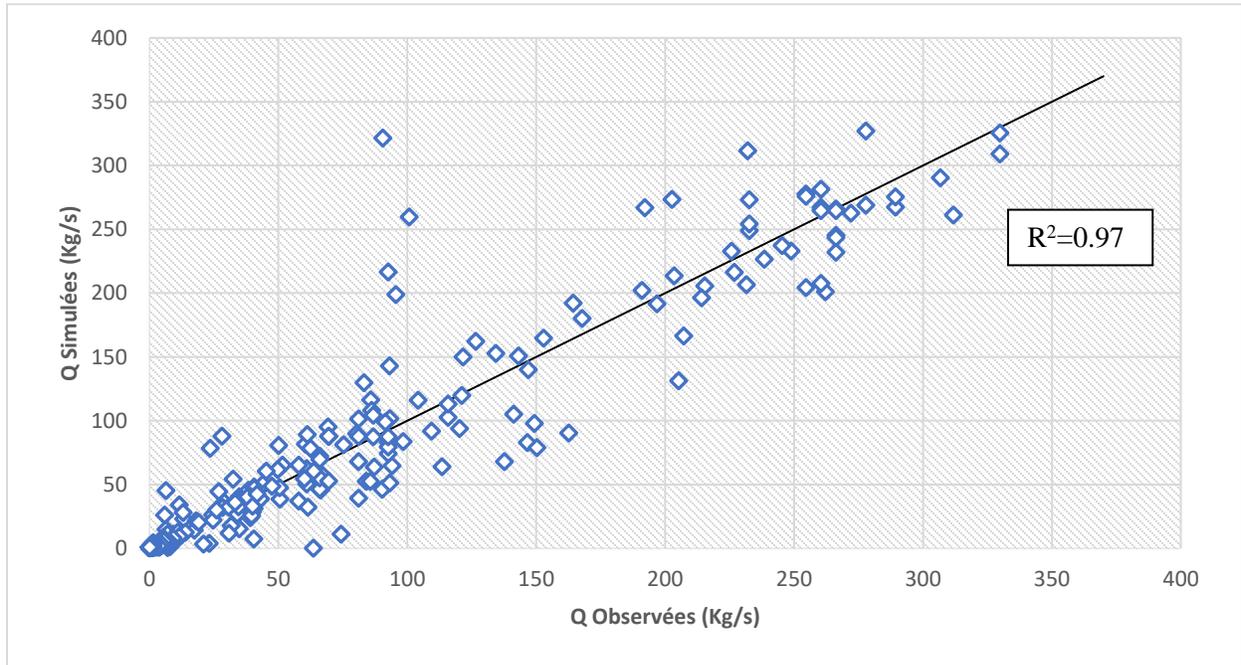


Figure IV.18 : Représentation des débits solides simulés et mesurés en période de validation (Colorado, RNA).

IV.5.1.2 Gradient Boosting

Les hyperparamètres optimisés sont :

- Le pourcentage du nombre d'inputs choisi, 0.707.
- Le taux d'apprentissage, 0,13.
- La profondeur maximale d'un arbre de décision, 2.
- Alpha, paramètre de la technique de régularisation L1, qui permet au modèle de ne pas diverger, ou d'être sujet à un sur-apprentissage, 3.
- Le nombre d'estimateurs utilisés, 100.

Tableau IV.4 : Les critères d'évaluation numériques pour le GB Colorado.

	Calage	Validation
R²	0.94	0.88
RSR	0.24	0.41
NASH	0.94	0.83
PBIAS	-0.50	16.16

Les résultats obtenus sont là aussi très bons comme le montre le tableau IV.4, tous les critères d'évaluation numériques sont très bons que ce soit en calage ou en validation, avec un pouvoir prédictif supérieur à 90% en période de calage et à 83% en période de validation, par contre le PBIAS est positif en période de validation avec une valeur de 16.16 ce qui tend à sous-estimer les valeurs.

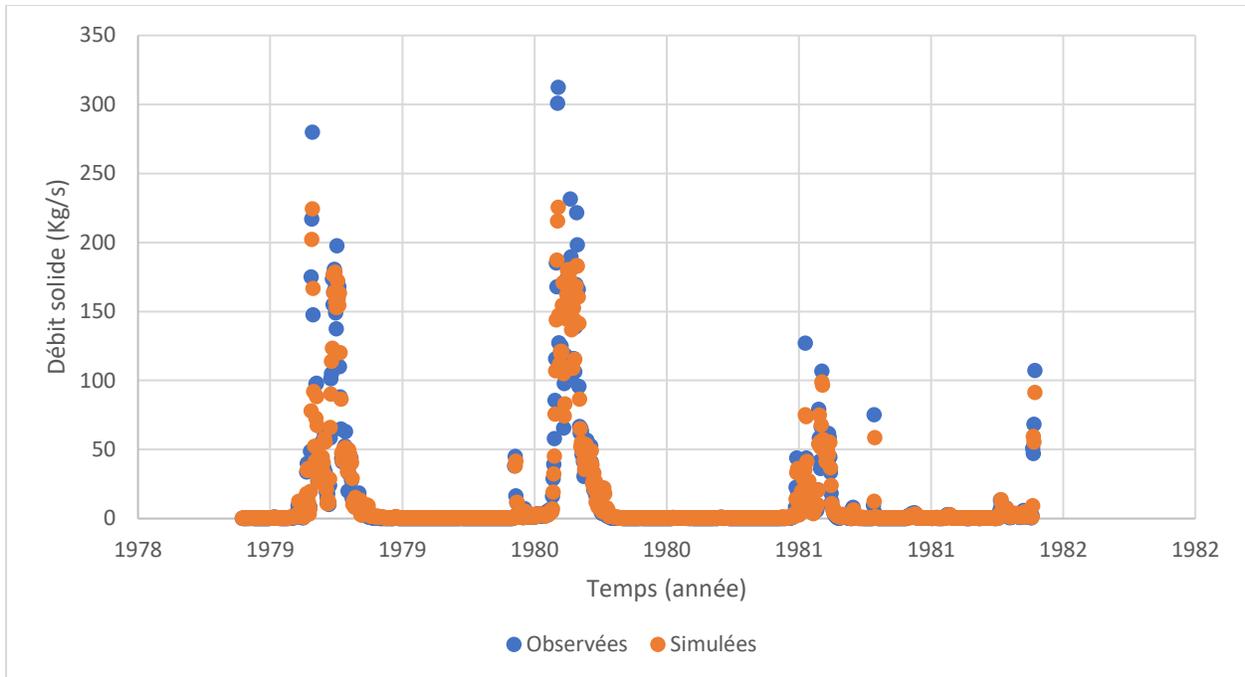


Figure IV.19 : Evolution des débits solides (mesuré et simulé) en période de calage (Colorado, GB)

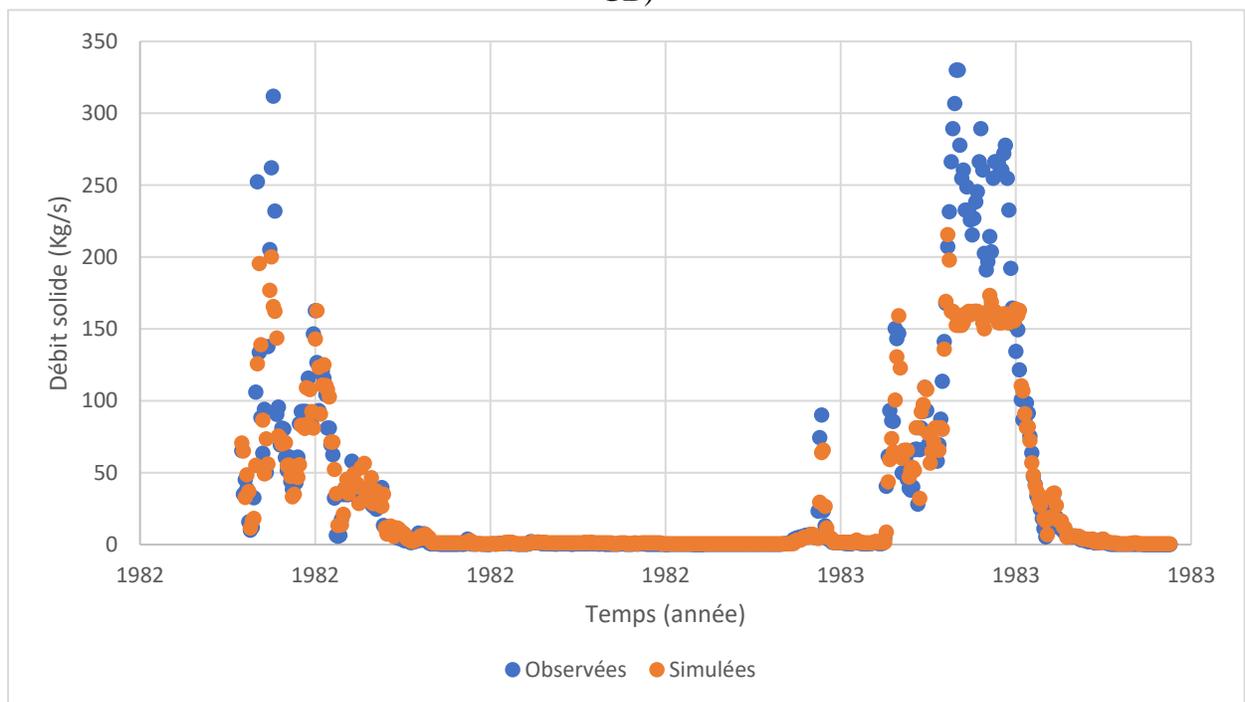


Figure IV.20 : Evolution des débits solides (mesuré et simulé) en période de validation (Colorado, GB).

La figure IV.19 explique bien les bons résultats obtenus vu que l'allure et les pics sont bien reproduits.

La figure IV.20 illustre ce qu'on a trouvé par les critères d'évaluation numériques, et on voit bien qu'il y a une sous-estimation des valeurs et les pics ne sont pas bien reproduits en période de validation.

Les graphes Q-Qplot entre les valeurs observées et simulées sont représentés comme suit :

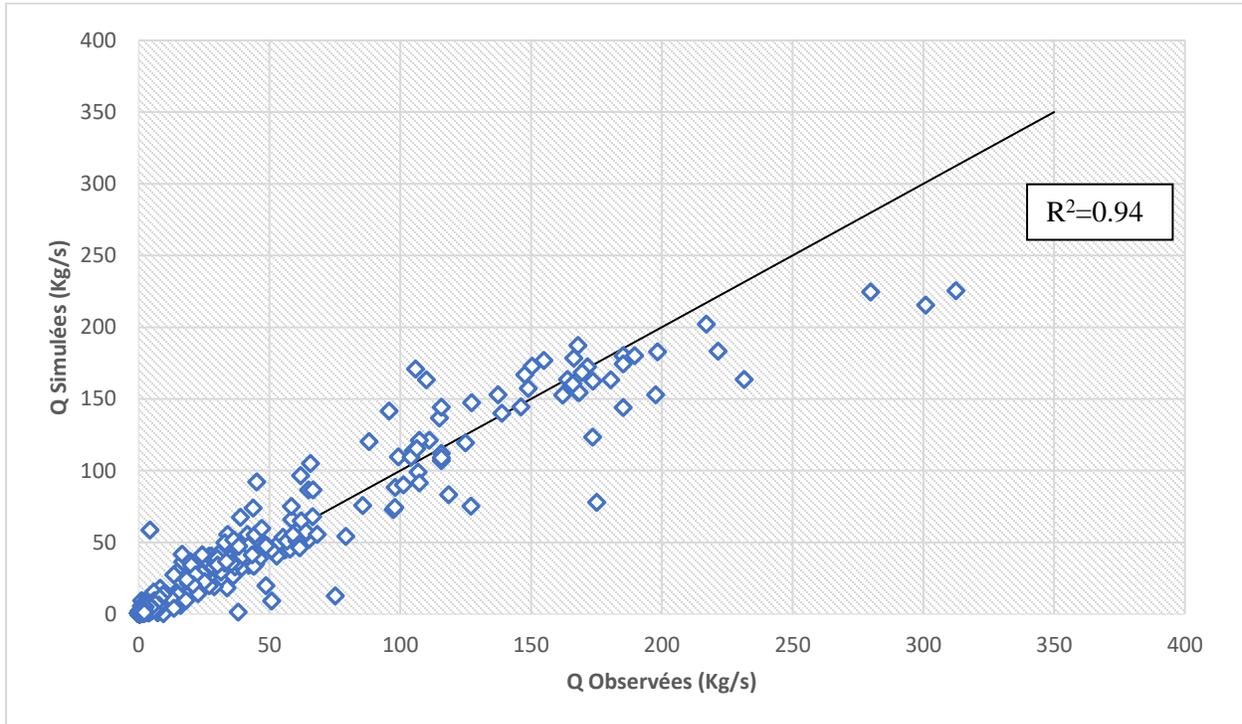


Figure IV.21 : Représentation des débits solides simulés et mesurés en période de calage (Colorado, GB).

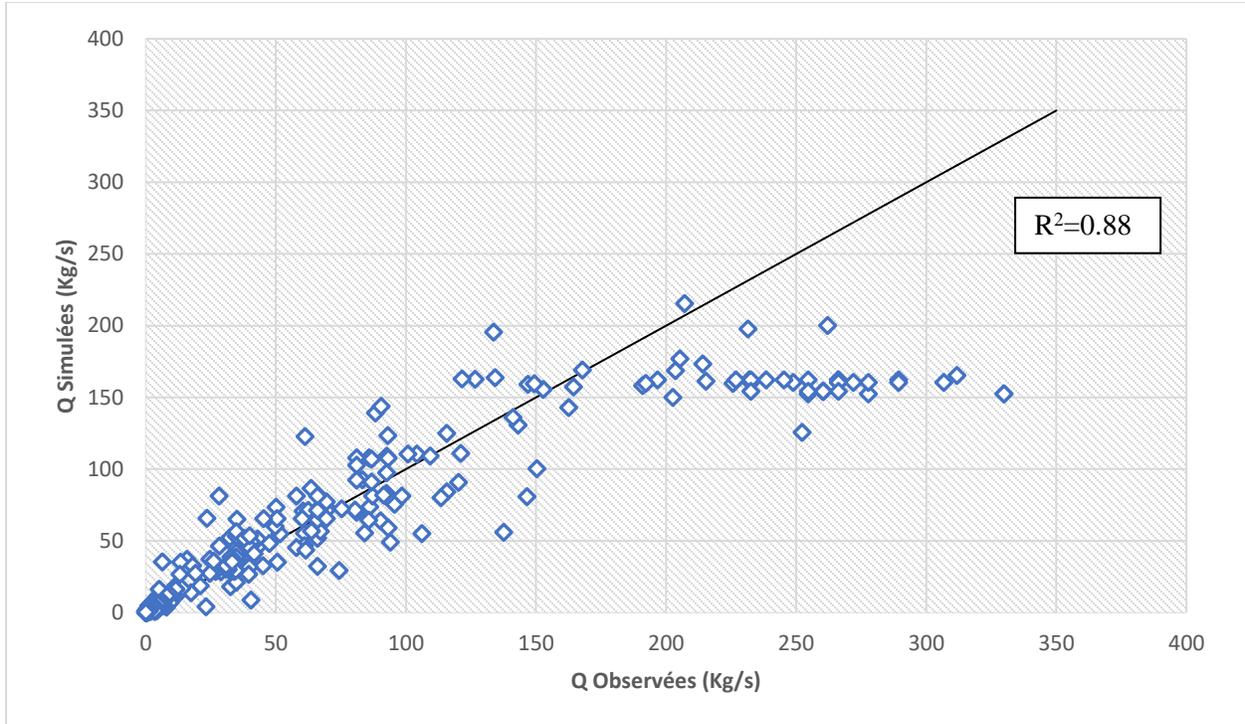


Figure IV.22 : Représentation des débits solides simulés et mesurés en période de validation (Colorado, GB).

Les figures IV.21 et IV.22 confirment les très bons résultats déjà trouvés qui existe entre les débits solides observé et simulé, ou l'on remarque que la majorité des points sont bien repartis de part et d'autres de la première bissectrice.

IV.6 Comparaison et conclusion

D'après les résultats obtenus précédemment par les deux méthodes et pour les deux séries de données on peut déduire que :

Les réseaux de neurones requièrent un grand nombre de données pour être réellement efficaces. Il est important que l'algorithme ait vu le plus de situations différentes possibles si l'on souhaite que les résultats produits soient cohérents. En plus du besoin important en quantité de données, celles-ci doivent également être de grande qualité. En effet, des données peu précises ou manquantes impliqueront des résultats erronés.

Ce résultat est bien confirmé aux vues de la différence flagrante entre les résultats obtenus sur les données de la Medjerda et celles de Colorado qui regroupent la qualité et la quantité.

Au final, on peut très bien dire que les réseaux de neurones ont un comportement mauvais en cas de faibles quantités de données et en cas de manque de fiabilité des données. Dans le cas contraire, ils présentent une bien meilleure souplesse, ce qui favorise fortement leurs utilisations.

L'algorithme XGBoost (Extreme Gradient Boosting) utilisé, est un algorithme ensembliste qui agrège des arbres. À chaque itération, le nouvel arbre apprend de l'erreur commise par l'arbre précédent. Ainsi, même si chaque arbre a un pouvoir prédictif faible, la règle de décision construite en sommant le résultat de chaque arbre est très fiable.

D'un point de vue opérationnel, cette approche dispose d'un avantage majeur : comme tout algorithme construit à partir d'arbres de décisions, il est capable de restituer les variables les plus discriminantes du modèle qu'il a construit. C'est important dans la mesure où l'on cherche à estimer de manière plus fine.

Dans notre cas pour les données de la Medjerda, même si on trouve qu'une seule ou quelques valeurs mesurées pendant une longue période de mesures voir même pendant des mois (problème de discontinuité) cela n'empêche que certaines de ces valeurs contiennent des informations pertinentes, donc l'algorithme par construction de modèles sur des échantillons de la population, favorise mieux ces paquets de données éparpillés (il les boostent d'où l'appellation : Boosting) et cela implique une bien meilleure gestion – par rapport aux autres algorithmes- du problème de discontinuité de données.

Il faut noter aussi que la gestion de données manquantes ou de zéros instrumentaux anormaux est prise en compte de façon originale dans l'implémentation de XGBoost. Celui-ci propose à chaque division une direction par défaut si une donnée est manquante. Pour pallier cette absence, le gradient est calculé sur les seules valeurs présentes. Cela peut encore plus expliquer les résultats obtenus par cette méthode qui sont meilleurs que ceux trouvés par RNA appliqués sur les données de Medjerda.

L'avantage qu'a pris cet algorithme sur les RNA au niveau des données de la Medjerda peut être aussi expliqué par sa capacité à prendre en compte l'interaction entre les différentes variables explicatives mises à sa disposition. En effet, parmi certaines variables qui seront fournies en entrée de l'algorithme (Q_1 , Q_{1-1} et Q_{1-2} ainsi que Q_{s-1} et Q_{s-2}), il peut exister des liens plus ou moins importants. Cette contrainte est très forte dans notre cas car le débit mesuré en une journée est fortement affecté par les débits des jours précédents. En effet, ces similitudes ont un impact direct sur le choix du modèle et la qualité de la modélisation.

On peut dire que le Gradient Boosting a un comportement assez bon en cas de faibles quantités de données et en cas manque de fiabilité des données donc une très bonne souplesse. Dans le cas contraire, ils donnent de meilleurs résultats mais il reste derrière le RNA qui lui, se montre plus robuste et plus précis quand il est question de nombres importants de données fiables.

Conclusion générale

Le bassin versant de la Medjerda se caractérise par l'importance des apports liquides hivernaux et printaniers suite à l'importance des précipitations surtout celles du mois de Mars, et par conséquent les volumes les plus importants en transport solide se concentrent durant la période allant de Décembre à Avril suite à la présence des conditions favorables.

Devant l'importance du phénomène du transport solide, qui est un phénomène de grande complexité, sa quantification précise est devenue une nécessité pour les projets de planification et de gestion. La non disponibilité des mesures fréquentes surtout celle de la concentration en matière en suspension au niveau des cours d'eau de cette région, fait appel à la modélisation afin d'établir une relation liant le débit solide au débit liquide.

Nous avons testé les deux modèles : réseaux de neurones et Gradient Boosting au niveau du bassin versant de la Medjerda. Une comparaison entre ces modèles s'est faite à base de critères d'évaluations qui attribue une appréciation sur le degré de relation entre les variables simulées et observées. Cependant, c'est la précision d'estimation du modèle qui juge son application. Dans notre cas les réseaux de neurones ont donné des résultats très médiocres comparés aux résultats du Gradient Boosting qui sont plus ou moins satisfaisant mais pas assez pour espérer pouvoir l'utiliser dans la prédiction à long terme. Par comparaison aux résultats fournis par ces deux méthodes appliquées aux données de Colorado, on peut clairement constater que nos données ne sont pas fiables et ne permettent pas de donner un jugement précis sur les modèles utilisés.

Ces modèles ne fournissent pas les explications concernant leurs résultats ce qui limite donc l'analyse des phénomènes existants. Cette limite est due à leur opacité qui empêche une analyse pertinente des solutions obtenues. Ils peuvent être assimilés à une boîte noire qui donne une réponse quand on lui fournit des données mais qui ne délivre pas toujours de justification simple à analyser. Les liens existants entre les variables du modèle ne sont pas toujours détectés. Ces systèmes ont donc un pouvoir explicatif limité.

Enfin, ce travail n'est qu'une contribution à la modélisation du phénomène du transport solide au niveau du bassin versant de la Medjerda par deux méthodes de Machine Learning très utilisées qui nécessite d'être complété en intervenant sur la qualité et la longueur de la série de données, et pour tirer des interprétations plus fondées afin d'aboutir à un modèle global représentant au mieux le processus du transport solide au niveau de ce bassin versant. Et comme

CONCLUSION GENERALE

les résultats et commentaires de cette étude sont basés sur seulement deux jeux de données, il serait donc intéressant de tester ces approches sur d'autres données notamment d'autres régions d'Algérie.

REFERENCES BIBLIOGRAPHIQUES

- 1- ABDEDDAIM Hakima. Contribution à l'étude de l'influence de la structure du réseau hydrographique sur le risque hydrologique « Cas des bassins de l'Est de l'Algérie ». Thèse présentée en vue de l'obtention du diplôme de Doctorat : sciences hydrauliques. Université Mohamed Khider – Biskra Faculté des Sciences et de la technologie Département : Hydraulique, 2018, 276 p.
- 2- Andriy Burkov. The Hundred-page Machine Learning Book, 2019.
- 3- BACHA Fella. Modélisation du transport solide par les modèles empiriques et neuro-flou à différents pas de temps. Projet de fin d'études pour l'obtention du diplôme d'Ingénieur d'Etat : hydraulique. Alger : Ecole Nationale Polytechnique, 2012, 117 p.
- 4- BENNACEUR Sofiane, Optimisation de la consommation énergétique du processus d'épuration par boues activées par le Gradient Boosting cas de la STEP Béni Messous. Projet de fin d'études pour l'obtention du diplôme d'Ingénieur d'Etat : hydraulique. Alger : Ecole Nationale Polytechnique, 2019, 84 p.
- 5- Josh Patterson & Adam Gibson. Deep Learning a practitioner's approach. First edition, SA,2017 [Revision 27/07/2017]. Format PDF. Disponible à l'adresse: <http://oreilly.com/safari>.
- 6- DERGHAL Zakarya. Modélisation du transport solide à l'échelle de la crue en utilisant le logiciel (HECRAS-HECHMS) (cas d'oued Abiod). Mémoire en vue de l'obtention du diplôme de master : hydraulique. Blida : Ecole Nationale Supérieure d'Hydraulique, 2016, 98 p.
- 7- DJABA Hichem. Ressources en eau et perspectives de gestion intégrée dans le bassin versant de Medjerda. Mémoire en vue de l'obtention du diplôme de magister : hydraulique. Annaba : Université Badji Mokhtar Faculté des Sciences de la Terre Département de Géologie, 2010, 155p.
- 8- ELAHCENE Omar. Etude de transport solide par charriage et en suspension dans le bassin versant d'Oueds Bellah. Thèse présentée en vue de l'obtention du diplôme de Doctorat : sciences agronomiques. Alger : Ecole Nationale Supérieure Agronomique ,2013, 109 p.
- 9- FANNY Arnaud. Nouvelles technologies pour mieux comprendre le fonctionnement sédimentaire des cours d'eau. Université Claude Bernard Lyon 1. 2016, 16 p.

- 10- GEAY Thomas. Mesure acoustique passive du transport par charriage dans les rivières. Sciences de la Terre. Université de Grenoble, 2013, 164 p.
- 11- GOUAIDIA, Laouar, Défaflia & Zenati. Caractérisation de la sécheresse météorologique au haut bassin d'Oued Mellegue Nord Est Algérien. Article 2018, 14 p.
- 12- HAMDI Aboubakeur Seddik. Estimation de l'érosion hydrique au Maghreb et l'application sur le bassin versant Oued Trouch. Mémoire en vue de l'obtention du diplôme de master : hydraulique. Blida : Ecole Nationale Supérieure d'Hydraulique, 2016, 100 p.
- 13- HOCINI Nabil & AMAMRA Tarik, Modélisation de la salinité des cours d'eau par les méthodes heuristiques et métaheuristiques : cas de l'oued Tafna, Mémoire en vue de l'obtention du diplôme de master : hydraulique. Alger : Ecole Nationale Polytechnique ,2016. 133 p.
- 14- LEFKIR Abdelouhab, Modélisation du transport solide par les modèles neuroflous. Thèse de doctorat : hydraulique. Alger : Ecole Nationale Polytechnique ,2007,150 p.
- 15- MAKKEB Zahia. Modélisation du transport solide par régression, Neuro Flou et Réseau de Neurones de l'Oued Isser. . Projet de fin d'études pour l'obtention du diplôme d'Ingénieur d'Etat : hydraulique. Alger : Ecole Nationale Polytechnique, 2016, 107 p.
- 16- OTTOU Paul. Méthodes d'apprentissage automatique appliquées au provisionnement ligne à ligne en assurance non-vie. Mémoire pour l'obtention du diplôme du Master Actuariat et l'admission à l'institut des Actuaires. Paris : Université Paris Dauphine, 2017, 91 p.
- 17- Plumb A. P., Rowe R. C., York P. and Brown M., 2005, Optimisation of the predictive ability of artificial neural network (ANN) models: A comparison of three ANN programs and four classes of training algorithm, Eur. J. Pharm. Sci., 25, 395-405.
- 18- RADJAH Soufyane. Modélisation du Transport solide par le modèle S.W.A.T (Application sur le Bassin versant des Issers). . Projet de fin d'études pour l'obtention du diplôme d'Ingénieur d'Etat : hydraulique. Alger : Ecole Nationale Polytechnique, 2017, 77 p.
- 19- SI MOUSSA Cherif, Modélisations des équilibres de phases à pressions élevées par les réseaux de neurones artificiels. Thèse de doctorat : génie chimique. Alger : Ecole Nationale Polytechnique ,2010,217 p.
- 20- STAMBOUL Mohammed Elamine. Analyses de l'état quantitatif et qualitatif des eaux superficielles et souterraines de la zone El Aouinet-Ouenza. Mémoire en vue de l'obtention du diplôme de magister : option Impact de changement climatique sur les ressources hydriques en Algérie. Tebessa : UNIVERSITE LARBI TEBESSI-TEBESSA ,2017, 270 p.

21- ZAABAR Narimane. Modélisation de la relation débit-MES par les modèles régressifs et les réseaux de neurones artificiels. Mémoire en vue de l'obtention du diplôme de master : hydraulique. Blida : Ecole Nationale Supérieure d'Hydraulique, 2016, 85 p.