

الجمهورية الجزائرية الديمقراطية الشعبية

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب - البليدة 1-

Université SAAD DAHLAB –BLIDA 1-



Faculté des Sciences de l'Ingénieur

Département des Sciences de l'Eau et de l'Environnement

MÉMOIRE

Projet de Fin d'Études

Pour l'obtention du diplôme de Master en Hydraulique

Option : Ressources Hydrauliques

Thème :

**Évaluation de l'impact des extrêmes sur les barrages en Algérie
par machine Learning. Cas de barrage Beni Haroun (Mila)**

Soutenu le 22/06/2025

Présenté par : Ouahab Ahmed AbdAldjalil

Devant le jury :

Mr. Benaaziza .A

MCA

Univ. Blida 1

Président

Mme. Benkaci .S

MCB

Univ. Blida 1

Examinatrice

Dr. Bouzeria .H

MCA

Univ. Blida 1

Promoteur

Promotion 2024/2024

Table des matières

Table des matières	1
Liste des figures.....	IV
Liste des tableaux.....	V
Liste des abréviations :	VI
Abstract:.....	VIII
المخلص:.....	VIII
Résumé :	IX
Introduction générale :	1
CHAPITRE I : REVUE DE LETIRATURE	2
1. Introduction :	3
2. Les barrages : rôle, types et vulnérabilité	3
2.1. Définitions et fonctions :	3
2.2. Les types des barrages :	3
2.3. Les barrages en Algérie :	3
3. Les extrêmes climatiques :	4
3.1. La sécheresse :	4
3.2 Les crues :	10
3.3 Impacts documentés :	12
4. Les barrages et les extrêmes climatiques :	13
4.1. Approches intégrées :	13
4.2 Limitations des approches classiques :	13
5. Machine Learning en hydrologie :	13
5.1 Principales techniques de Machine Learning appliquées à l'hydrologie :	13
5.2 Limites et perspectives :	14
6. Lacunes identifiées :	15
7. Conclusion :	16
CHAPITRE II :PRESENTATION DE LA ZONE D'ETUDE	17
1. Introduction :	18
2. Présentation de la région d'étude :	18
2.1 Localisation géographique :	18
2.2. Caractéristiques physiques :	20

2.3. Caractéristiques climatiques :	21
2.4. Morphologie du bassin :	24
3. Présentation et caractéristiques techniques du barrage de Beni Haroun :	30
3.1. Fonction et rôle du barrage :	30
3.2. Caractéristiques techniques générales :	31
3.3 Équipements hydrauliques et sécurité :	31
3.4 Fondations et géologie :	32
3.5 Historique et évolution :	32
3.6 Données hydrologiques du bassin :	32
4. conclusion :	33
CHAPITRE III :DONNEES ET METHODES.....	34
1. Introduction :	35
2. Données :	35
2.1. Sources et type de données utilisées :	35
2.2. Prétraitement des données :	36
3. Méthodes :	37
3.1. Modélisation prédictive par Machine Learning :	37
4. Conclusion :	41
CHAPITRE IV :RESULTATS ET DISCUSSION	42
1. Introduction :	43
2. Evaluation des performances :	43
2.2 Analyse de multicollinéarité : Variance Inflation Factor (VIF) et Tolérance.....	43
2.2 Rapport des classifications :	44
2.3 indicateurs numériques de performance (Models metrics) :	46
2.4 Matrices de confusion :	46
2.5 Importance des variables :	51
2.6 Courbes ROC et AUC :	53
3. Analyse temporelle des épisodes extrêmes et des tendances :	57
3.1. Cartes temporelles des épisodes extrêmes :	57
3.2. Analyse de tendance avec le test de Mann-Kendall :	69
4. Conclusion :	71
Conclusion générale :.....	72

Liste des figures

Figure 1 : carte de localisation géologique de bassin KR.....	19
Figure 2 : carte de répartition administratif de bassin.....	20
Figure 3: Variations interannuelles des précipitations, station de Beni Haroun(ANBT).....	22
Figure 4 : Carte d'occupation du sol / land cover (sentinel2).....	23
Figure 5 : carte de réseau hydrographique.....	29
Figure 6 : vue sur barrage BH (Evacuateur de crue) (ANBT).2017.....	30
Figure 7 : Retenue du barrage de Béni Haroun (ANBT).....	33
Figure 8 : matrice de confusion de Random Forest.....	47
Figure 9 : matrice de confusion de lightGBM.....	48
Figure 10 : matrice de confusion de CatBoost.....	49
Figure 11 : matrice de confusion de XGBoost.....	50
Figure 12 : courbe ROC avec indicateur AUC pour 'Random Forest'.....	53
Figure 14 : courbe ROC avec indicateur AUC pour 'CatBoost'.....	54
Figure 13 : courbe ROC avec indicateur AUC pour 'LightGBM'.....	54
Figure 13 : courbe ROC avec indicateur AUC pour 'XGBoost'.....	55
Figure 14 : carte temporelle des événements 'crue'.....	57
Figure 15 : carte temporelle des événements 'normale'.....	58
Figure 16 : carte temporelle des événements 'sécheresse'.....	59
Figure 17 : carte temporelle des événements 'crue' prédits par RF.....	60
Figure 19 : carte temporelle des événements 'crue' prédits par CatBoost.....	61
Figure 18 : carte temporelle des événements 'crue' prédits par LightGBM.....	61
Figure 20 : carte temporelle des événements 'crue' prédits par XGBoost.....	62
Figure 21 : carte temporelle des événements 'normale' prédits par RF.....	63
Figure 22 : carte temporelle des événements 'normale' prédits par LightGBM.....	63
Figure 24 : carte temporelle des événements 'normale' prédits par XGBoost.....	64
Figure 23 : carte temporelle des événements 'normale' prédits par CatBoost.....	64
Figure 25 : carte temporelle des événements 'sécheresse' prédits par RF.....	66
Figure 26 : carte temporelle des événements 'sécheresse' prédits par lightGBM.....	66
Figure 28 : carte temporelle des événements 'sécheresse' prédits par XGBoost.....	67
Figure 27 : carte temporelle des événements 'sécheresse' prédits par CatBoost.....	67
Figure 30 : tendance annuelle -normale.....	69
Figure 29 : tendance annuelle -crue.....	69
Figure 31 : tendance annule - sécheresse.....	70

Liste des tableaux

Tableau 1 : classifications selon SPI.....	7
Tableau 2 : classification selon PDSI.....	8
Tableau 3 : classification selon centiles.....	9
Tableau 4 : Températures moyennes mensuelles à la station Ain El Bey Constantine	22
Tableau 5 : Surfaces et périmètre par SBV.	24
Tableau 6 : Indices de compacité par sous bassin versant	25
Tableau 7 : longueur équivalente et largeur équivalente par sous bassin versant.....	26
Tableau 8 : les altitudes Caractéristiques par sous bassin versant	28
Tableau 9 : Longueur des oueds permanents et temporaires par sous-bassin versant	29
Tableau 10 : caractéristiques techniques de barrage de Beni Haroun (ANBT)	31
Tableau 11 : résultats de VIF et tolérance.....	43
Tableau 12 : indicateurs de performance par classe	44
Tableau 13 : performances globales des algos	46
Tableau 14 : explication pour les erreurs amener	51
Tableau 15 : Tableau comparatif de l'importance des variables.....	51
Tableau 16 : Résultats des AUC par classe	53
Tableau 17 : les résultats de test (Mann-Kendall).....	70

Liste des abréviations :

ML : Machine Learning

ANBT : Agence Nationale des Barrages et Transferts

ADE : Algérienne des Eaux

AEP : Alimentation en Eau Potable

AEI : Alimentation en Eau Industrielle

SPI : Standardized Precipitation Index

PDSI : Palmer Drought Severity Index

AVHRR : Advanced Very High Resolution Radiometer

NOAA : National Oceanic and Atmospheric Administration

NDVI : Normalized Difference Vegetation Index

VCI : Vegetation Condition Index

SVI : Standardized Vegetation Index

TCI : Temperature Condition Index

VHI : Vegetation Health Index

OMM : Organisation Météorologique Mondiale

HEC-HMS : Hydrologic Engineering Center - Hydrologic Modeling System

MIKE 11 : Modèle hydraulique unidimensionnel de DHI (Danish Hydraulic Institute)

LISFLOOD : Large-scale hydrological model developed by JRC/EC

RF : Random Forest

XAI : eXplainable Artificial Intelligence

ANRH : Agence Nationale des Ressources Hydrauliques

MNT : Modèle Numérique de Terrain

SBV : Sous-Bassin Versant

EVAP : Évaporation

VDF : Vidange de Fond

DEV : Déverse

P_j : Pluie journalière

P_m : Pluie mensuelle

ΔV : Variation de volume

GPU : Graphics Processing Unit

VIF : Variance Inflation Factor

ROC : Receiver Operating Characteristic

AUC: Area Under Curve

Abstract:

This thesis investigates the application of Machine Learning techniques (Random Forest, XGBoost, LightGBM, CatBoost) to model and predict extreme hydrological events, droughts and floods-based on time series data from the Beni Haroun dam in Algeria. Using daily historical records (2003-2025) provided by the ANBT, the study integrates hydrometeorological indicators (SPI) and derived variables (ΔV , outflows) to develop a three-class classification model (drought, normal, flood). Model performances are assessed using standard metrics (precision, recall, F1-score), and illustrated through temporal maps of predicted classes. The results reveal a high accuracy in detecting flood events, stable classification of normal periods, but limited precision in identifying drought episodes. This study highlights the practical potential of Machine Learning for risk assessment related to climatic extremes in Algerian hydraulic systems.

Keywords. Machine Learning, Floods, Drought, Classification, Beni Haroun Dam, Time Series, Hydrological Modeling, Temporal Maps

المخلص :

يهدف هذا البحث إلى دراسة تطبيق تقنيات التعلم الآلي (Random Forest، XGBoost، LightGBM، CatBoost) لنمذجة وتوقع الظواهر الهيدرولوجية القصوى، مثل الجفاف والفيضانات، بالاعتماد على سلسلة زمنية من بيانات سد بني هارون في الجزائر. بالاستناد إلى بيانات شهرية تاريخية للفترة 2003–2025، والمقدمة من الوكالة الوطنية للسدود والتحويلات (ANBT)، تم دمج مؤشرات هيدرولوجية (مثل SPI ومتغيرات مشتقة ΔV) لبناء نموذج تصنيف بثلاث فئات: جفاف، وضعية عادية، فيضان. تم تقييم أداء النماذج باستخدام مؤشرات دقيقة (الدقة، الاسترجاع، معامل F1)، وتم تمثيل النتائج من خلال خرائط زمنية للفئات المتوقعة. أظهرت النتائج قدرة عالية على اكتشاف الفيضانات، واستقرارًا جيدًا في تصنيف الحالات العادية، لكن بدقة أقل في التعرف على حالات الجفاف. يبرز هذا العمل الإمكانيات العملية لتقنيات التعلم الآلي في تقييم المخاطر المرتبطة بالتغيرات المناخية القصوى ضمن المنظومات المائية في الجزائر.

الكلمات المفتاحية. التعلم الآلي، الفيضانات، الجفاف، التصنيف، سد بني هارون، السلاسل الزمنية، النمذجة الهيدرولوجية، الخرائط الزمنية

Résumé :

Ce mémoire explore l'utilisation des techniques de Machine Learning (Random Forest, XGBoost, LightGBM, CatBoost) pour modéliser et prédire les événements hydrologiques extrêmes, sécheresses et crues à partir de séries temporelles issues du barrage de Beni Haroun en Algérie. En s'appuyant sur des données historiques journalières (2003–2025) fournies par l'ANBT, le travail combine des indicateurs hydrométéorologiques (SPI) et des variables dérivées (ΔV , défluent) pour construire un modèle de classification à trois classes (sécheresse, normale, crue). Les performances des modèles sont évaluées à l'aide des métriques classiques (précision, rappel, F1-score), et illustrées par des cartes temporelles des classes prédites. Les résultats montrent une forte capacité à détecter les crues, une bonne stabilité pour les épisodes normaux, mais une précision limitée sur les épisodes de sécheresse ce qui peut être justifié par la complexité de ce phénomène. Cette étude met en évidence l'intérêt opérationnel du Machine Learning pour l'évaluation des risques liés aux extrêmes climatiques dans les systèmes hydrauliques algériens.

Mots clé. Machine Learning , Crues, Sécheresse , Classification, Barrage de Beni Haroun, Séries temporelles, Modélisation hydrologique, Cartes temporelles

Introduction générale

L'Algérie, pays caractérisé par une forte variabilité climatique et une répartition inégale des ressources en eau, fait face à des défis croissants en matière de gestion hydrique, accentués par la fréquence accrue des événements extrêmes tels que les sécheresses prolongées et les crues soudaines. Dans ce contexte, les barrages jouent un rôle stratégique pour l'approvisionnement en eau potable, l'irrigation, et la protection contre les crues. Toutefois, la pression exercée par les changements climatiques met en évidence les limites des approches classiques d'évaluation des risques et souligne la nécessité d'outils plus performants et adaptatifs.

Ce mémoire s'inscrit dans cette perspective. Il vise à évaluer l'impact des extrêmes climatiques sur le fonctionnement des barrages, en combinant l'analyse de données historiques, des indices hydrométéorologiques, et des techniques avancées d'apprentissage automatique (Machine Learning). Le barrage de Beni Haroun, l'un des plus importants en Algérie, est pris comme cas d'étude

À travers une méthodologie intégrée, mobilisant à la fois des outils statistiques, des algorithmes de classification supervisée et des représentations temporelles, ce travail ambitionne d'apporter une contribution opérationnelle à la compréhension et à l'anticipation des épisodes hydrologiques extrêmes, en vue d'une gestion plus résiliente des ressources en eau.

CHAPITRE I : REVUE DE LA LETIRATURE

1. Introduction :

En Algérie, la gestion durable des ressources hydriques repose en grande partie sur un réseau de barrages répartis principalement au nord du pays. Ces infrastructures assurent l’approvisionnement en eau potable, l’irrigation, la protection contre les crues. Cependant, l’augmentation de la fréquence et de l’intensité des événements extrêmes tels que les sécheresses prolongées et les pluies torrentielles accentués par les changements climatiques actuels mettent à rude épreuve la protection de ces ouvrages.

Face à ces défis, les approches traditionnelles d’évaluation des risques montrent leurs limites, en particulier dans leur capacité à anticiper les événements exceptionnels ou à intégrer la dimension temporelle des vulnérabilités. Dans ce contexte, l’utilisation de Machine Learning (ML) émerge comme une solution puissante pour évaluer et modéliser les risques climatiques pesant sur les infrastructures hydrauliques.

Ce chapitre vise à établir les fondements théoriques et techniques nécessaires à la construction d’un cadre d’analyse innovant, intégrant les ML pour l’évaluation des impacts des extrêmes climatiques sur les barrages algériens, en particulier celui de Beni Haroun.

2. Les barrages : rôle, types et vulnérabilité

2.1. Définitions et fonctions :

Un barrage est un ouvrage hydraulique construit pour retenir, réguler ou détourner l’écoulement d’un cours d’eau, afin de constituer une réserve d’eau et/ou de produire de l’énergie. Les principaux usages sont l’irrigation, l’approvisionnement en eau potable, la régulation des crues, la production hydroélectrique et parfois la navigation ou les loisirs. [1]

2.2. Les types des barrages :

Les barrages se distinguent par :

- Leur type de structure : poids, voûte, contreforts, terre/enrochements ;
- Leur matériau : béton, matériaux granulaires, composites ;
- Leur hauteur et capacité : grands (≥ 100 m), moyens (15-100 m), petits (< 15 m). [1]

2.3. Les barrages en Algérie :

L’Algérie compte plus de 81 barrages opérationnels dont plusieurs de grande capacité comme Beni Haroun, et Koudiat Acredoune .

Les barrages sont gérés par l’Agence nationale des Barrages et Transferts (ANBT) et l’exploitation de l’eau potable de ces barrages est effectuée par l’Algérienne des Eaux (ADE)

Ces ouvrages sont stratégiques pour l’agriculture, l’urbanisation et la sécurité hydrique, mais sont exposés à des risques croissants liés à la variabilité du climat et à l’envasement . [2]

3. Les extrêmes climatiques :

3.1. La sécheresse :

3.1.1. Définition de la sécheresse :

La sécheresse est un phénomène naturel dû à un déficit des précipitations qui associé à d'autres facteurs climatiques (températures élevées, vents forts, hygrométrie faible, etc.), provoque s'il est prolongé dans le temps, une diminution du ruissellement des cours d'eau et de l'humidité des sols et un renouvellement insuffisant des aquifères. Ce phénomène est amplifié par l'action néfaste de l'homme sur l'environnement (déforestation, rejets des gaz nocifs, pression du cheptel sur les parcours, surexploitation des ressources naturelles « eaux et sols », etc.). Cela se répercute, négativement, parfois en de véritables catastrophes naturelles, sur l'alimentation des hommes et des cheptels et la production agricole. [3]

3.1.2. Les types de sécheresse:

La sécheresse de type météorologique :

Elle est définie par un déficit de précipitations sur une période de temps déterminée (les précipitations sont bien inférieures à la normale au cours d'une longue période). Le seuil choisi (exmp 90 % de la normale) et la durée de la période (par exemple six mois) varient d'un endroit à l'autre selon les besoins des utilisateurs ou les applications concernées. [4]

La sécheresse de type agricole :

La sécheresse agricole est plus définie par la présence dans le sol d'une quantité d'eau suffisante pour assurer la croissance des cultures et du fourrage que par l'écart des précipitations par rapport à la normale sur une période de temps déterminée. Il n'y a pas de relation directe entre la hauteur de précipitation et l'infiltration des précipitations dans le sol. Le taux d'infiltration varie en fonction des conditions d'humidité antérieures, de la pente de la nature du sol et de l'intensité des précipitations. Les sols ont aussi des caractéristiques variables: certains ont une faible capacité de rétention d'eau et sont donc plus sujets à type de sécheresse. [4]

La sécheresse de type hydrologique :

Elle correspond à une diminution des réserves d'eau dans les cours d'eau, les réservoirs de surface et la nappe phréatique (qui survient lorsqu'une longue sécheresse météorologique provoque une diminution sensible du niveau des eaux souterraines, des rivières, des fleuves et des lacs.). La sécheresse hydrologique est causée par un manque de précipitations accompagné d'une évaporation massive (des températures élevées). Cependant, des facteurs non météorologiques, comme la demande en eau, la disponibilité de réservoirs de surface et le forage de puits artésiens, en aggravent l'effet. [4]

3.1.3. Sécheresse en Algérie : contexte et vulnérabilités hydriques :

L'Algérie fait face à une pression croissante sur ses ressources en eau, exacerbée par des épisodes récurrents de sécheresse. Le pays ne dispose que de 13 % de sa superficie dans des zones bénéficiant de précipitations régulières, essentiellement situées dans la bande nord, entre l'Atlas tellien et la Méditerranée. Cette région, qui concentre la majorité de la population, des terres agricoles et des infrastructures, est particulièrement vulnérable aux variations climatiques.

La pluviométrie annuelle y présente de fortes disparités, allant de 400 mm dans l'Ouest à plus de 1 000 mm dans l'Est sur le littoral, et chute de 100 à 400 mm dans les Hautes Plaines steppiques et présahariennes. Ces zones intermédiaires, où prédominent les eaux souterraines et les systèmes endoréiques, sont également les plus sensibles à la désertification.

Selon les données de la Banque Mondiale, l'Algérie est passée sous le seuil de rareté hydrique (1 000 m³/hab/an) dès les années 1990. En 2020, la disponibilité par habitant ne dépassait pas 430 m³/an, bien en dessous des besoins estimés pour assurer la sécurité alimentaire, qui nécessiteraient entre 15 et 20 milliards de m³/an, dont 70 % destinés à l'agriculture.

Cette rareté structurelle est régulièrement aggravée par des sécheresses sévères, qui affectent à la fois la recharge des réservoirs de surface, les aquifères, et la productivité agricole. Par ailleurs, les vagues de chaleur de plus en plus fréquentes, telles que celles enregistrées en 2018 à In Salah et Adrar (avec des températures proches de 60 °C), exacerbent l'évaporation et réduisent l'efficacité des systèmes de stockage hydrique.

Dans ce contexte, la gestion des barrages devient un enjeu stratégique pour le pays. L'anticipation des épisodes de sécheresse et l'optimisation des ressources mobilisables appellent à des outils de prévision performants, capables de croiser données historiques, observations climatiques, et dynamiques spatiales.[5] [6]

3.1.4. Impacts de la sécheresse sur le secteur des ressources en eau :

- La réduction des apports hydriques vers les barrages et les lacs se traduit par une baisse notable des volumes stockés, aussi bien dans les réservoirs artificiels que dans les plans d'eau naturels. Cette diminution affecte directement la capacité de distribution de l'eau, notamment à destination des zones urbaines et des périmètres agricoles irrigués.
- Un déficit pluviométrique prolongé compromet le renouvellement des nappes phréatiques, ce qui entraîne une intensification donc une surexploitation, des ressources souterraines pour répondre à la demande.
- Le déséquilibre croissant entre l'offre en eau affaiblie par la diminution des précipitations et des apports de surface et la demande en constante augmentation, notamment pour les besoins agricoles et d'alimentation en eau potable (AEP/AEI), génère une pression accrue sur les nappes phréatiques, qu'elles soient renouvelables ou fossiles.

- La diminution des ressources contribue également à une détérioration de la qualité de l'eau. En effet, la baisse du débit des cours d'eau réduit la capacité de dilution des polluants, augmentant ainsi la concentration des contaminants dans les eaux de surface et souterraines.
- Enfin, les déséquilibres dans la distribution hydrique peuvent conduire à des situations de pénurie, nécessitant la mise en œuvre de mesures de gestion exceptionnelles, telles que des arbitrages entre usagers et le recours à des dispositifs de distribution mobiles [3]

3.1.5. Indicateurs de sécheresse :

1. Indices météorologiques :

Les indices météorologiques de sécheresse reposent principalement sur l'analyse des données de précipitations issues des stations météorologiques. Leur objectif est de caractériser les conditions de sécheresse en comparant les précipitations observées à des références historiques. Leur simplicité d'application, leur rapidité de calcul et leur accessibilité en font des outils largement utilisés.

a) Rapport à la normale :

Le rapport à la normale exprime l'écart entre les précipitations observées sur une période donnée et la moyenne historique correspondante, généralement calculée sur 30 ans. Il s'agit d'un indicateur exprimé en pourcentage, utilisé à des échelles spatiales locales ou régionales, pour des périodes variant d'un mois à une année [7] :

$$P_m = \left(\frac{P}{P_m} \right) * 100$$

Où :

- P : précipitations totales sur la période considérée (mm)
- P_m : moyenne historique des précipitations pour la même période (mm)

b) Indice Standardisé de Précipitation (SPI) :

Développé par McKee et al. (1993), le SPI est un indicateur statistique permettant de quantifier les anomalies pluviométriques (excès ou déficit) par rapport à la moyenne historique sur différentes échelles temporelles (de 1 à 24 mois). Il permet ainsi d'évaluer les effets des sécheresses sur différents compartiments du cycle hydrologique [8]:

$$SPI = \frac{(p - p_m)}{\delta_p}$$

Où :

- P : précipitations observées (mm)
- P_m: précipitations moyennes historiques (mm)
- δ_p: écart type historique des précipitations (mm)

La classification de l'intensité des sécheresses selon le SPI est la suivante :

Tableau 1 : classifications selon SPI

Valeur SPI	Classification
≥ 2	Humidité extrême
1.5 à 1.99	Humidité sévère
1.0 à 1.49	Humidité modérée
-0.99 à 0.99	Conditions normales
-1.0 à -1.49	Sécheresse modérée
-1.5 à -1.99	Sécheresse sévère
≤ -2	Sécheresse extrême

Une sécheresse est considérée en cours lorsque le SPI reste négatif et atteint une valeur inférieure ou égale à -1, et elle se termine lorsque le SPI redevient positif. La sévérité d'un épisode est quantifiée par la somme des SPI négatifs au cours de la période affectée.

c) Indice PDSI (Palmer Drought Severity Index):

Le PDSI, ou Indice de Sévérité de la Sécheresse de Palmer, évalue les anomalies d'humidité en fonction de l'offre et de la demande en eau, en tenant compte notamment de l'humidité du sol et des données de température. Il est généralement calculé sur une base hebdomadaire ou mensuelle et normalisé pour permettre des comparaisons interrégionales [9] [10]:

$$PDSI = X_{(t)} = 0.897X_{(t-1)} + \frac{Z_{(t)}}{3}$$

- $X_{(t-1)}$: PDSI de la période précédente « moisture Anomaly index »
- I : Mois de l'année.
- $Z_{(i)} = K(P - Pc)$
- K : facteur de poids (voir Alley,1984);
- P : précipitation actuelle (mm) ;
- Pc : précipitation CAFEC (mm) ;
- $Pc = \alpha_j PE + \beta_j PR + \gamma_j PRO + \delta_j PL$
- CAFEC: « Climatically Appropriate For Existing Conditions »
- $\alpha_j, \beta_j, \gamma_j, \delta_j$: coefficients climatiques mensuels ;
- j : Mois de l'année ;
- PE : Evapotranspiration Potentielle (mm) ;
- PR : Recharge du sol Potentielle (mm) ;
- PRO : Ruissellement Potentiel (mm) ;
- PL : Perte Potentielle dans le sol (mm).
- La classification du PDSI est la suivante :

Tableau 2 : classification selon PDSI

Valeur PDSI	Classification
≥ 4	Humidité extrême
3 à 3.99	Humidité sévère
2 à 2.99	Humidité modérée
1 à 1.99	Humidité légère
0.5 à 0.99	Début de période humide
-0.49 à 0.49	Conditions normales
-0.5 à -0.99	Sécheresse naissante
-1 à -1.99	Sécheresse légère
-2 à -2.99	Sécheresse modérée
-3 à -3.99	Sécheresse sévère
≤ -4	Sécheresse extrême

Le PDSI présente néanmoins certaines limites, notamment l'absence de prise en compte des ressources hydrologiques de surface (cours d'eau, réservoirs, etc.) et des changements dans l'usage de l'eau.

d) Centile :

le classement par centile divise la distribution des précipitations en cent intervalles égaux, chacun représentant 1 % de la fréquence historique. Cette méthode permet d'évaluer la fréquence des événements extrêmes avec plus de finesse que le simple rapport à la normale. [11]

Tableau 3 : classification selon centiles

Centiles	Classification
0–10	Bien en dessous de la normale
10–40	En dessous de la normale
40–60	Près de la normale
60–90	Au-dessus de la normale
90–100	Bien au-dessus de la normale

Cette méthode présente l'avantage de s'ajuster à la distribution réelle des données, sans dépendre de seuils absolus arbitraires. Elle est donc particulièrement adaptée à l'analyse des séries historiques incomplètes, aux comparaisons spatio-temporelles, ainsi qu'à la classification supervisée

2. Indicateurs de sécheresse issus de la télédétection :

La télédétection constitue un outil précieux pour le suivi de la sécheresse à grande échelle. Plusieurs indices dérivés des images satellites, notamment du capteur AVHRR à bord des satellites NOAA, permettent d'évaluer l'état de la végétation et les effets du stress hydrique.

Parmi les plus utilisés, on retrouve le NDVI (Normalized Difference Vegetation Index), indicateur classique du couvert végétal, ainsi que des indices dérivés comme le VCI (Vegetation Condition Index), le SVI (Standardized Vegetation Index) et le TCI (Temperature Condition Index), qui permettent de mieux évaluer les anomalies climatiques affectant la végétation. Ces indices peuvent être combinés pour former le VHI (Vegetation Health Index), souvent employé pour la surveillance intégrée des conditions de sécheresse à l'échelle régionale. [17]

Bien que ces indicateurs ne constituent pas l'axe principal de ce travail, leur mention reste pertinente dans le cadre d'une approche complémentaire de caractérisation des épisodes de sécheresse.

3.2 Les crues :

3.2.1. Définition de la crue :

Une crue est un phénomène hydrologique caractérisé par une élévation rapide et temporaire du niveau ou du débit d'un cours d'eau, causée par des apports exceptionnels d'eau dans le bassin versant. Elle résulte généralement de précipitations intenses, parfois combinées à la fonte des neiges ou à des ruptures de digues ou de barrages en amont. Lorsqu'une crue dépasse la capacité du lit mineur à contenir l'écoulement, elle peut provoquer des inondations, endommageant les infrastructures, les zones habitées et les terres agricoles. [12]

Selon l'Organisation météorologique mondiale (OMM), les crues constituent l'un des aléas naturels les plus fréquents et les plus destructeurs au niveau mondial.

3.2.2. Types des crues [12]:

On distingue plusieurs types de crues selon leurs mécanismes et origines hydrométéorologiques :

Crues pluviales :

Elles sont causées par des précipitations soudaines et abondantes, souvent concentrées sur un court laps de temps (quelques heures à quelques jours). Elles sont typiques des régions méditerranéennes, où la violence des orages peut provoquer une montée brutale des eaux. Ce type de crue est particulièrement fréquent en Algérie, notamment entre octobre et mars.

Crues de saturation :

Elles résultent de précipitations modérées mais prolongées sur des sols déjà saturés. L'infiltration devient impossible, et le ruissellement de surface augmente, alimentant les cours d'eau au-delà de leur capacité normale.

Crues par fonte nivale :

Dans les régions montagneuses, la fonte rapide des neiges au printemps peut entraîner des crues importantes, en particulier lorsqu'elle est accélérée par des pluies chaudes. Bien que ce type soit plus fréquent dans les zones tempérées et alpines, certaines régions du nord-est algérien peuvent y être exposées ponctuellement.

Crues torrentielles ou flash floods :

Ces crues éclair sont extrêmement rapides, survenant en quelques minutes à quelques heures après un événement pluviométrique intense. Elles sont typiques des petits bassins versants escarpés, comme ceux de Kabylie ou du Tell, où le ruissellement est intense et les temps de réponse très courts.

3.2.3. Crues en Algérie :

L'Algérie, en raison de sa situation climatique et géographique, est particulièrement exposée aux crues soudaines. Dans la région nord, les précipitations peuvent atteindre 100 mm ou plus en une seule journée, provoquant des montées brutales des cours d'eau. Les zones urbaines mal drainées, les versants déboisés et les sols imperméabilisés augmentent considérablement le ruissellement et donc le risque d'inondation. [13]

Plusieurs événements historiques ont marqué le pays :

- Bab El Oued (Alger, novembre 2001) : 712 mm de pluie en quelques heures, plus de 700 morts, destruction d'infrastructures, crue soudaine amplifiée par l'urbanisation non maîtrisée. [14]
- El Bayadh (2010), Tebessa (2012), Jijel (2019) : crues éclair dans des zones montagneuses provoquant pertes humaines et dégâts importants sur les réseaux routiers et hydrauliques. [15]
- Crues fréquentes de l'oued Rhumel (2003, 2008, 2009, 2015, 2024) dans la région de Constantine, influençant directement le niveau du barrage de Beni Haroun. [13]

3.2.4. Impacts des crues sur les infrastructures hydrauliques :

- Les crues, en particulier les crues extrêmes ou mal anticipées, peuvent avoir des conséquences majeures sur les barrages et les réseaux hydrauliques associés :
- Surcharge hydraulique : les débits entrants peuvent dépasser la capacité du déversoir, augmentant le risque de débordement (overtopping), l'une des premières causes de rupture de barrages.
- Érosion des talus et fondations : les vitesses d'écoulement élevées déstabilisent les berges, creusent les lits, et peuvent déchausser les assises des ouvrages.
- Apports sédimentaires massifs : les crues mobilisent de grandes quantités de matériaux solides (limons, graviers, boues), accélérant l'envasement des retenues et réduisant leur capacité utile.
- Blocage ou détérioration des ouvrages d'évacuation : la présence de débris (troncs, ordures, roches) peut obstruer les galeries de vidange, les conduites et les évacuateurs.
- Saturation des systèmes de régulation : l'augmentation brutale du volume stocké impose des manœuvres d'urgence (ouverture des vannes, décharges exceptionnelles), mettant à l'épreuve les équipements mécaniques.

3.2.5. Outils d'évaluation et d'alerte des crues :

L'identification et la prévision des crues nécessitent l'intégration de plusieurs types de données : [16]

- Données météorologiques : intensité des pluies, durée, fréquence ;
- Données hydrologiques : niveaux d'eau, débits, temps de réponse des bassins ;
- Modèles de simulation : HEC-HMS, MIKE 11, LISFLOOD ;

3.3 Impacts documentés :

L'évaluation des impacts des extrêmes climatiques sur les infrastructures hydrauliques se fonde non seulement sur des modèles théoriques, mais également sur l'analyse d'événements concrets ayant marqué certaines régions du monde et plus particulièrement l'Algérie. Ces événements illustrent de manière empirique la vulnérabilité des barrages face aux sécheresses prolongées ou aux crues soudaines.

a. Sécheresse et stress hydrique prolongé en Algérie :

Depuis plusieurs décennies, l'Algérie fait face à une succession de sécheresses sévères affectant la disponibilité et la qualité des ressources hydriques. En 2020, la quasi-totalité des retenues du nord du pays ont enregistré des taux de remplissage inférieurs à 35 %, avec des records historiquement bas dans les barrages de l'Ouest comme Fergoug et Beni Bahdel. Le barrage de Beni Haroun, bien que stratégique, a connu des fluctuations importantes de son volume stocké, compromettant temporairement l'irrigation dans les plaines de Mila et la distribution en eau potable vers Constantine. [17]

b. Crue exceptionnelle à Ghardaïa (2008) :

Le 1er octobre 2008, la ville de Ghardaïa a été frappée par une crue éclair résultant de fortes précipitations (plus de 60 mm en moins d'une heure). L'oued M'zab est entré en crue soudaine, avec un débit supérieur à 1 200 m³/s, entraînant la mort de 33 personnes et d'importants dégâts matériels. Les infrastructures hydrauliques, y compris les digues de protection et les canalisations d'adduction, ont été submergées ou arrachées, rendant le réseau d'eau potable temporairement inutilisable. [18]

c. Crue récente à Béchar (2024) :

En septembre 2024, la région de Béchar a connu des inondations importantes à la suite de pluies diluviennes localisées. L'oued Guir est entré en crue, inondant plusieurs quartiers et infrastructures. Le barrage de Djorf Torba a enregistré un apport brutal de 140 millions de m³, nécessitant des décharges d'urgence. Environ 260 familles ont été évacuées par la Protection civile. Cet événement met en lumière la vulnérabilité des bassins sahariens aux extrêmes soudains, amplifiés par des sols imperméables et une urbanisation mal contrôlée. [19]

d. Crise de Derna (Libye, 2023) :

Bien qu'externe à l'Algérie, la catastrophe de Derna en septembre 2023 constitue une alerte majeure pour l'Afrique du Nord. Suite au passage de la tempête Daniel, deux barrages anciens ont cédé sous la pression d'une pluie exceptionnelle (plus de 150 mm en 24 h), provoquant la mort de plus de 10 000 personnes. Cette tragédie souligne les risques liés au vieillissement des infrastructures, à l'absence de maintenance, et à la montée en intensité des extrêmes climatiques. [20]

4. Les barrages et les extrêmes climatiques :

4.1. Approches intégrées :

- Courbes de fragilité : modélisent la probabilité de défaillance en fonction d'une intensité d'aléa;
- Évaluation multicritère : combine facteurs géotechniques, hydrologiques et opérationnels;
- Analyses historiques et retour d'expérience : permettent de calibrer les modèles de risque en tenant compte des erreurs passées. [21]

4.2 Limitations des approches classiques :

Les modèles hydrauliques traditionnels nécessitent [22] :

- Des hypothèses simplificatrices ;
- Des données précises souvent indisponibles ;
- Un recalage difficile en présence d'aléas extrêmes non historiques.

5. Machine Learning en hydrologie :

L'application du Machine Learning (ML) en hydrologie connaît un essor rapide ces dernières années, grâce à la disponibilité croissante de jeux de données pluviométriques et hydrologiques à haute résolution et à l'amélioration des architectures algorithmiques (Random Forest, XGBoost, etc.). Ces approches 'data-driven' ont démontré leur supériorité par rapport aux modèles hydrologiques conceptuels ou physiquement basés sur de nombreuses tâches ; prévision de débits, estimation des sécheresses et même pour étudier le transport sédimentaire.

5.1 Principales techniques de Machine Learning appliquées à l'hydrologie :

L'application du Machine Learning (ML) dans l'analyse hydrologique permet de modéliser des phénomènes complexes comme les crues ou les sécheresses à partir de grandes quantités de données. Plusieurs algorithmes se distinguent par leur efficacité à trouver les relations entre les variables hydrologiques telles que les débits, les volumes stockés ou les indicateurs d'extrême avec les événements ou épisodes étudiés. Voici les techniques les plus pertinentes pour ce travail :

5.1.1. Random Forest (RF) :

Random Forest est un algorithme d'ensemble basé sur la construction de multiples arbres de décision. Il fonctionne par bootstrap (échantillonnage aléatoire avec remplacement) et sélection aléatoire de variables à chaque nœud, ce qui permet de réduire la variance et d'augmenter la robustesse du modèle. [23]

Avantages en hydrologie :

- Très adapté à la prédiction de séries temporelles (apports, volumes, SPI) avec peu de prétraitements.
- Réduction du surapprentissage grâce à l'usage d'arbres non corrélés.
- Excellents résultats dans la classification binaire (ex : crue simple = 1 / 0).

5.1.2. Gradient Boosting (XGBoost, LightGBM) :

Le gradient boosting consiste à construire des arbres de décision successifs, chaque arbre corrigeant les erreurs du précédent. XGBoost et LightGBM en sont des implémentations optimisées, très rapides et précises. [24]

Applications en hydrologie :

- Prédiction de débits journaliers ou sédiments avec faible erreur .
- Très bon compromis entre précision et interprétabilité.

5.2 Limites et perspectives :

Limites :

- Équifinalité et interprétabilité : les modèles ML, bien que performants, restent parfois opaques .
- Qualité et disponibilité des données : lacunes spatiales et temporelles dans de nombreux pays en développement, notamment pour les variables clés (débits, précipitation)
- Généralisation : les modèles entraînés dans une région climatiquement homogène peuvent mal transposer dans un autre contexte sans recalibration

Perspectives :

- Automatisation des workflows ML via plateformes cloud.
- Intégration de données IoT et de télédétection en quasi-temps réel.
- Hybridation avancée avec modèles physiques pour conjuguer rigueur hydrologique et performance prédictive. [25]

6. Lacunes identifiées :

Dans le contexte de l'évaluation des impacts des extrêmes climatiques sur un barrage, plusieurs insuffisances méthodologiques et opérationnelles persistent, tant au niveau global que spécifiquement en Algérie. Trois grands axes de lacunes émergent : la faible intégration des approches ML, le manque de cartographie spatiale prédictive pour la vulnérabilité des barrages, et des limitations liées aux données et à l'interprétabilité des modèles.

- Manque d'intégration de ML dans les évaluations hydrologiques : La littérature souligne que, malgré le potentiel conjoint du Machine Learning pour la modélisation hydrologique, les études intégrées restent rares et fragmentées
- Malgré que ce type de modélisation soit promise à améliorer la précision de la prévision des événements extrêmes, mais qu'elle se heurte à des défis de compatibilité des formats et de traitement des données hétérogènes
- Absence de cartographie prédictive spatiale des vulnérabilités pour les barrages en Algérie
Les études consacrées à l'Algérie se limitent principalement à la cartographie rétrospective des aléas (inondations, érosion) sans produire de cartes prévisionnelles de vulnérabilité des barrages .
- Limitations de données, d'interprétabilité et de transférabilité :
Le manque de données à haute résolution (satellites, stations hydrométriques, capteurs) freine la calibration de modèles ML robustes. également que l'interopérabilité des bases de données (formats, métadonnées) fait souvent obstacle à un flux de travail fluide par algorithmes d'apprentissage
- En outre, la complexité et l'opacité des modèles ML limitent leur adoption par les praticiens, qui réclament des outils explicatifs (XAI) pour valider les décisions de gestion et le fait que la généralisation des modèles développés dans un bassin ou un pays donné reste incertaine : sans recalibrage, les performances chutent lorsqu'on les applique à des contextes hydrologiquement divergents .

Pour combler ces lacunes il est nécessaire de développer un framework opérationnel alliant :

- Un pipeline de traitement ML standardisé et automatisé .
- Des bases de données hydrologiques enrichies et harmonisées.
- Des outils XAI ou des plateformes qui simplifie les tâches pour renforcer la confiance des utilisateurs .

Ce cadre permettra de passer d'une évaluation rétrospective à une vulnérabilité prédictive des barrages en Algérie, indispensable pour anticiper et atténuer efficacement les impacts des extrêmes climatiques.

7. Conclusion :

Ce chapitre a permis de poser les fondations théoriques indispensables à la compréhension des enjeux liés à l'évaluation des impacts des événements climatiques extrêmes sur les barrages en Algérie. Il a d'abord rappelé l'importance stratégique des barrages dans le système de gestion de l'eau, en soulignant leur vulnérabilité croissante face à la variabilité et à l'intensification des phénomènes extrêmes tels que les sécheresses prolongées et les crues soudaines . L'analyse détaillée de ces extrêmes, de leurs mécanismes et de leurs impacts directs sur les barrages met en évidence la nécessité d'outils d'alerte et d'aide à la décision plus performants. Les indicateurs classiques de suivi (SPI, PDSI, centiles, etc.) offrent un socle méthodologique robuste, mais montrent leurs limites en termes d'anticipation . Dans ce contexte, le recours au Machine Learning apparaît comme une alternative prometteuse, sa capacité à exploiter de grandes bases de données climatiques et hydrologiques et à modéliser des relations complexes non linéaires, et à prédire des événements rares, en fait un outil particulièrement adapté à la gestion proactive des risques sur les barrages.

Enfin, les limites recensées dans la littérature , notamment en matière de données, de transparence des modèles, et d'intégration spatiale , appellent à la mise en œuvre d'approches hybrides, combinant modélisation automatique et les indicateurs explicatifs. C'est dans cette logique que s'inscrit la suite de ce mémoire, avec un objectif de concevoir un cadre prédictif innovant appliqué au barrage de Beni Haroun.

CHAPITRE II : PRESENTATION DE LA ZONE D'ETUDE

1. Introduction :

La compréhension fine du contexte géographique, hydrologique et infrastructurel de la zone d'étude constitue une étape essentielle dans toute démarche d'évaluation des risques liés aux événements climatiques extrêmes. Dans le cadre de ce mémoire, l'analyse porte sur le barrage de Béni Haroun, situé dans le nord-est de l'Algérie, au sein du bassin versant de l'Oued Kébir. Ce choix est motivé par l'importance stratégique de cet ouvrage, qui représente la plus grande retenue d'eau du pays et assure l'alimentation en eau potable et agricole de plusieurs wilayas du nord-est algérien.

La région étudiée se distingue par une grande variabilité topographique et climatique, qui la rend particulièrement vulnérable aux aléas hydrologiques tels que les crues soudaines et les sécheresses prolongées. Ces phénomènes, exacerbés par le changement climatique, peuvent compromettre la sécurité des ouvrages hydrauliques, la stabilité des réserves en eau, et la continuité du service de distribution. Dès lors, il est indispensable de caractériser avec précision les composantes physiques et environnementales du territoire, ainsi que les caractéristiques techniques du barrage, pour pouvoir évaluer les risques associés et concevoir des modèles prédictifs fiables.

Ce chapitre est structuré en deux grandes parties. La première présente les principaux éléments du contexte régional : géographie, géologie, topographie, réseau hydrographique et climat. La seconde est consacrée à une description détaillée du barrage de Béni Haroun, de ses fonctionnalités, de ses dimensions et de son rôle dans le système national de gestion de l'eau. Ces éléments fourniront la base nécessaire à l'analyse des épisodes extrêmes dans les chapitres suivants

2. Présentation de la région d'étude :

2.1 Localisation géographique :

Le territoire algérien est subdivisé en cinq régions, selon le découpage hydrographique, qui sont gérées depuis 1996 par des agences dénommées Agence de Bassin Hydrologique (ABH) :

Oranais-Chott Chergui, Chellif-Zahrez, Algérois-Soummam-Hodna, Constantinois-Seybouse-Mellegue, Sahara.

Ces régions sont divisées en 17 bassins hydrographiques. La région d'étude concerne le bassin hydrographique de Kébir Rhummel, situé à la latitude 36° Nord et à la longitude 7° et délimiter comme suit :

- Au nord, par les deux bassins côtiers de Constantine Ouest et Centre.
- Au sud, par le bassin des Hauts Plateaux Constantinois.
- À l'ouest, par les bassins de l'Algérois-Hodna-Soummam.
- À l'est, par le bassin de la Seybouse. [26] [27]

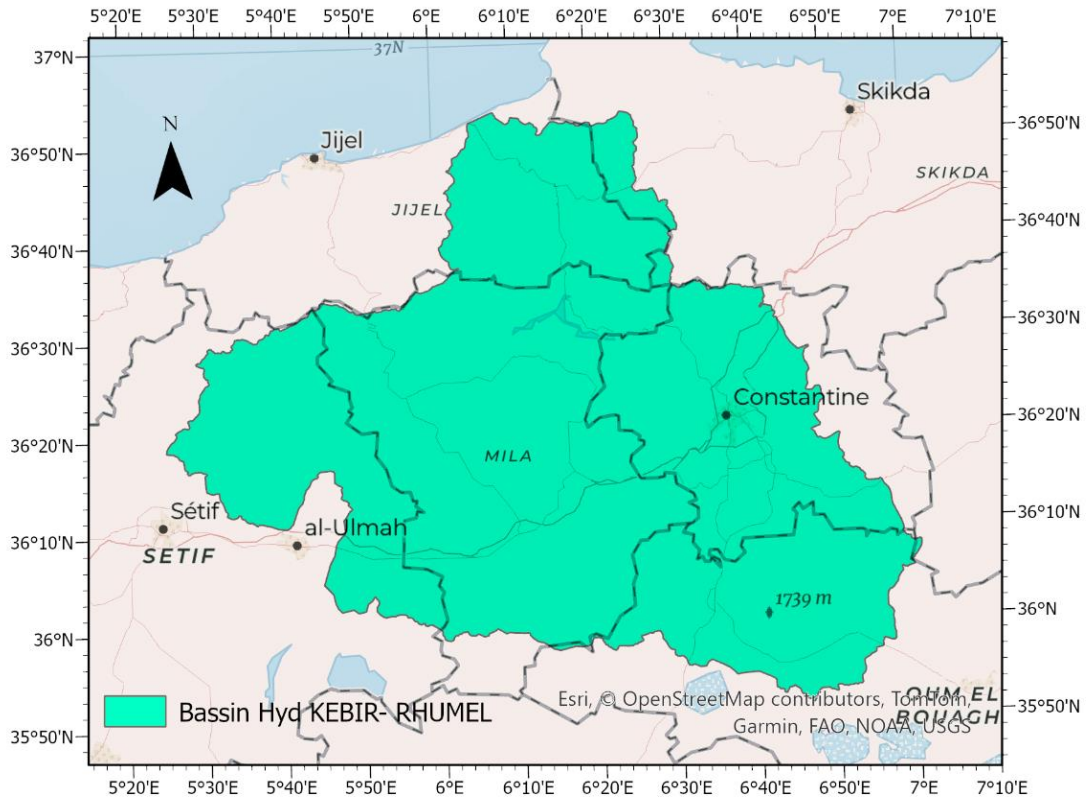
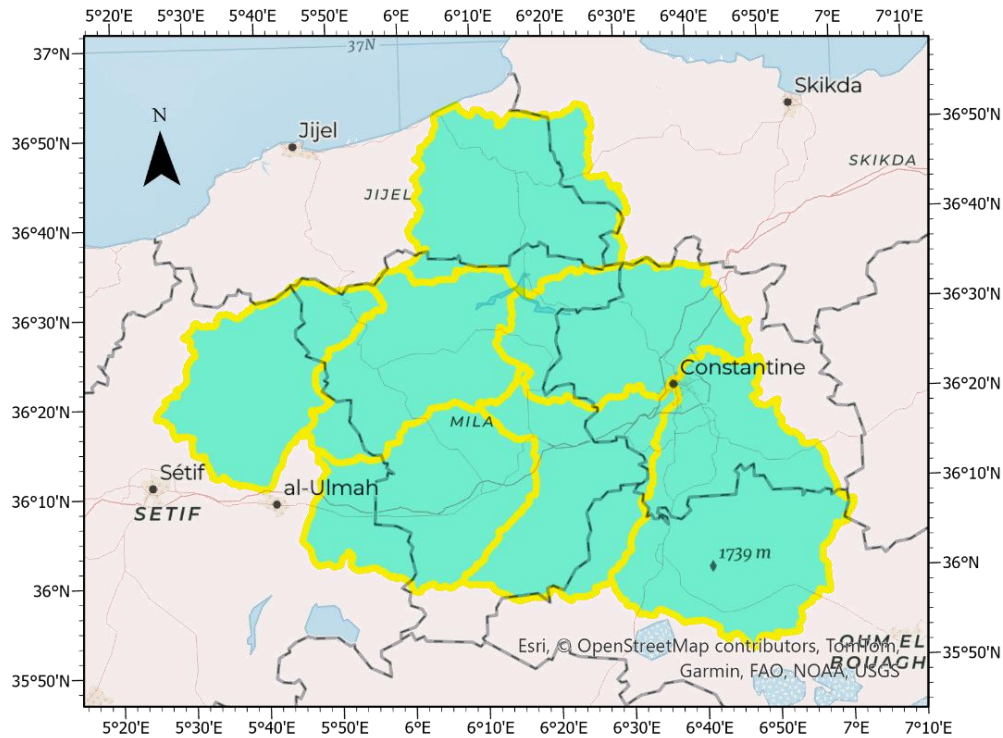


Figure 1 : carte de localisation géologique de bassin KR

Il est compte parmi les plus vastes et importants bassins hydrographiques d’Algérie. Couvrant une superficie de 8 815 km², il est subdivisé en sept sous-bassins hydrographiques.

Sur le plan administratif, le secteur d’étude couvre Partiellement six wilayas : Jijel, Mila, Skikda, Constantine, Setif, et Oum El Bouaghi



Legend

SBs KEBIR_RHUMEL

Figure 2 : carte de répartition administratif de bassin

L'oued Kébir-Rhumel constitue le principal cours d'eau qui draine le bassin, depuis la ville d'Ain M'lila, située dans les Hautes Plaines Constantinoises, jusqu'au réservoir du barrage de Beni Haroun. Plusieurs centres urbains majeurs se trouvent au sein de ce bassin versant, notamment Constantine et El Khroub à l'est, Mila au nord, Ain El Kebira à l'ouest, ainsi qu'Ain M'lila au sud. [27]

2.2. Caractéristiques physiques :

2.2.1. Géologie du bassin :

La géologie de la région est particulièrement complexe, marquée par une superposition d'unités structurales disposées du nord vers le sud.

Le substratum est dominé par des séries sédimentaires paléogènes (marnes, calcaires) et néogènes (grès, marnes argileuses), héritage de la phase alpine. Des plis et chevauchements, orientés est-ouest, témoignent d'un anticlinal du Tell, affectant l'amont du bassin et favorisant l'altération et l'épuration des roches en affleurement, source majeure de sédiments. [28]

2.2.2. Impact de la géologie sur l'hydrologie :

Le type de formations géologiques qui affleurent dans les bassins versants a une influence sur la répartition des écoulements superficiels. En effet, un bassin formé de matériaux très perméables avec une couverture végétale continue aura en général une densité de drainage faible assurant une meilleure infiltration des eaux superficielles.

Par ailleurs, un bassin formé de roches imperméables mais meubles et érodables, comme des marnes et des argiles, avec une végétation moyenne, présente souvent une densité de drainage élevée, ce qui favorise le ruissellement des eaux superficielles aux dépens de l'infiltration. [27]

2.2.3. Relief et topographie :

Ce site est situé dans une zone à relief accidenté, caractérisée par de forts contrastes topographiques, avec des massifs atteignant des altitudes élevées, ainsi que des ravins et talwegs profondément incisés. Le relief oscille entre 400 m et 1 200 m d'altitude. Deux grands ensembles se distinguent :

- Les monts telliens au nord, épine dorsale composée des chaînons du Babor (1 994 m) et des Bibans (1 722 m), qui forment une barrière orographique marquée ;
- Les Hauts-Plateaux au sud, ondulés, disséqués par de petites vallées et des cuvettes endoréiques, notamment le Chott El Hodna.

Les fortes pentes du Tell accélèrent le ruissellement et entraînent des crues soudaines lors d'épisodes pluvieux intenses.

2.3. Caractéristiques climatiques :

La connaissance des conditions climatiques de notre zone d'étude et de ses caractéristiques va nous permettre de mieux comprendre le comportement hydrologique des cours d'eaux et aussi l'intensité des extrêmes climatiques d'une manière générale, le climat de la région est de type continental, semiaride au Sud et subhumide au Nord avec un hiver pluvieux et froid et un été sec et chaud.

Le climat d'une région est influencé par de nombreux paramètres climatologiques qui sont :

- La température
- La précipitation
- L'humidité et le vent

2.3.1. La température :

C'est l'un des éléments les plus importants de la radiation solaire. Un certain nombre de phénomènes physiques tels-que l'évaporation, le dégel de l'eau, la gelée blanche. La dilatation et la concentration de certains matériaux : phénomène physiologique ou pathologique tel que la transpiration sont liées à la température de l'air. La température dont il s'agit est la température de l'air mesuré à l'ombre dans un abri météorologique. Les données de température se limitent à celles fournies à l'ANRH par la station météorologique de l'aéroport d'Ain El Bey à Constantine. La température moyenne saisonnière varie entre 10°C en hiver et 25°C en été. Par contre la moyenne annuelle et de l'ordre de 17.06 °C. [29]

Tableau 4 : Températures moyennes mensuelles à la station Ain El Bey Constantine

Mois	Sept	Oct	Nov	Déc	Jan	Fév	Mar	Avr	Mai	Jui	Juil	Août
Température °C	22.6	18.82	14.8	11.7	10.89	11.35	12.3	14.4	17.7	21.1	24.1	24.74

2.3.2 La précipitation :

La pluie est le facteur climatique le plus important qui conditionne l'écoulement saisonnier et influence directement le régime des cours d'eau et de toutes façons les extrêmes comme les crues intense et les sécheresses. La région d'étude est considérée parmi les régions les plus arrosées en Algérie . les précipitations sont relativement importantes, variables et irrégulières d'une année à une autre. [29]

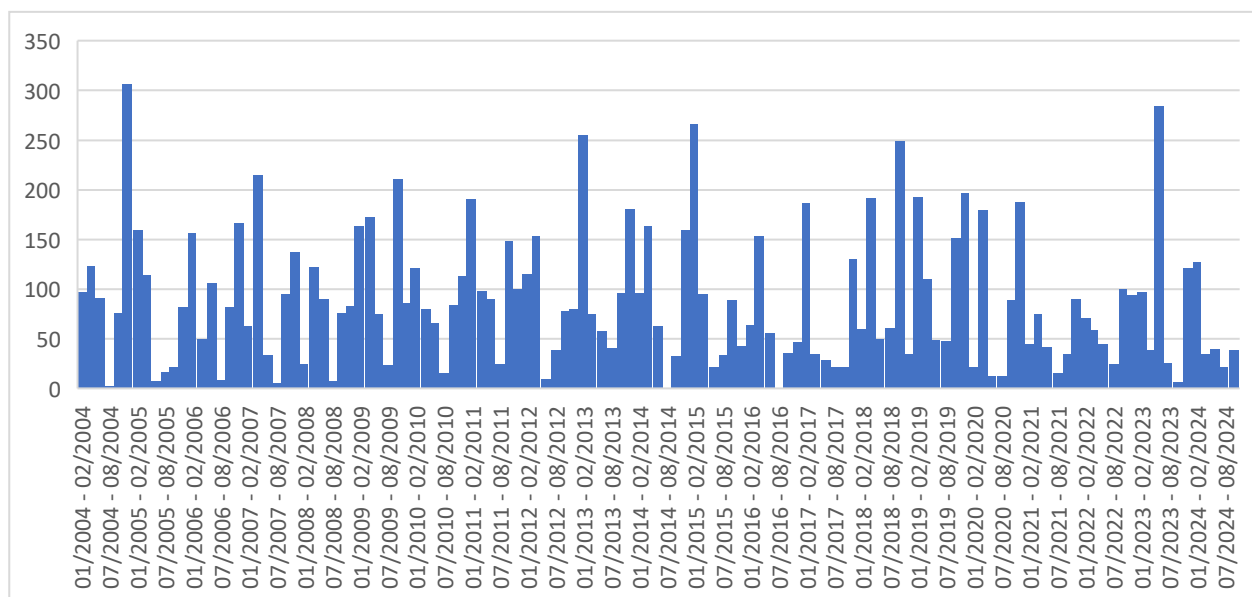


Figure 3: Variations interannuelles des précipitations, station de Beni Haroun(ANBT)

2.3.3. L'humidité et le vent :

L'humidité est plus élevée de l'année, s'étalent de juin jusqu'à septembre, le taux d'humidité moyen au cours de l'année est de 70%.

Le vent est un facteur important du climat il influe sur la température, l'humidité et l'évaporation. La direction, la fréquence et la vitesse des vents sont variables au cours de l'année. Les vents dominants dans la région sont ceux du Nord-Ouest avec une vitesse assez constante au cours de l'année, entre 1.9 et 2.5 m/s. [29]

2.3.4. Végétation du bassin :

Le couvert végétal de la région d'étude est peu développé. Il se limite essentiellement à des surfaces dominées par des herbacées, accompagnées de quelques eucalyptus, à l'exception du sous-bassin 1007 (en amont du barrage de Beni Haroun), qui présente une couverture végétale dense. Ce faible couvert végétal sur l'ensemble du bassin de Kebir-Rhumel a largement contribué à l'accentuation du phénomène d'érosion des sols. [27]

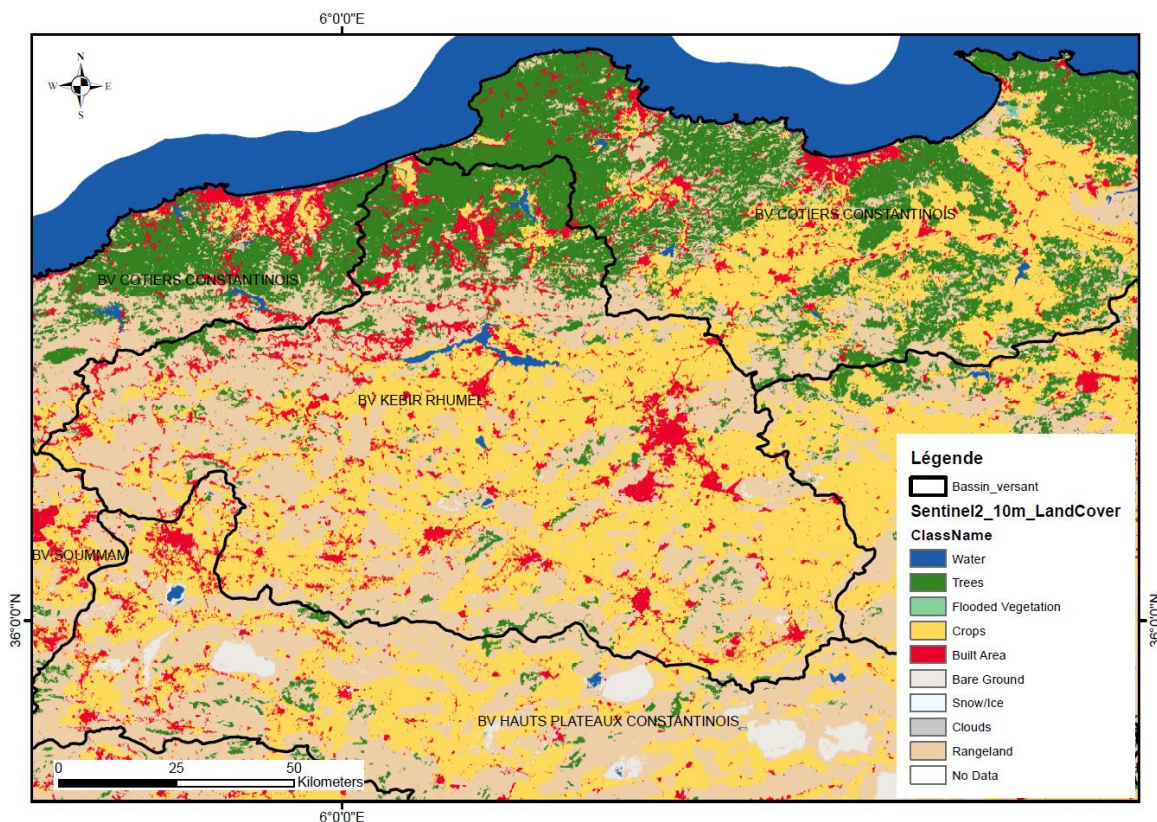


Figure 4 : Carte d'occupation du sol / land cover (sentinel2)

2.4. Morphologie du bassin :

Les caractéristiques morphométriques du bassin influencent fortement sa réponse hydrologique et en particulier le régime d'écoulement. Les géomorphologues américains tels que Horton, Leopold, Miller, Strahler, Schumm, etc., ont établi de nombreuses formules qui expriment généralement les relations statistiques observées entre les divers caractères morphologiques et physiques des lits et des bassins. Trois types de paramètres morphométriques permettent d'expliquer le comportement hydrologique des bassins [27] :

- Les paramètres de forme ;
- Les paramètres de relief ;
- Les paramètres morphométriques d'organisation du réseau hydrographique.

Plusieurs paramètres sont mesurés à l'aide de ArcGIS et bien sûr le Modèle Numérique de Terrain (MNT).

2.4.1. Paramètres de forme : [27]

a) Surface et périmètre :

Le bassin versant est l'aire de réception des précipitations qui alimentent un cours d'eau par écoulement. Il se caractérise par des indices géométriques qui se mesurent de façon directe et indirecte. La surface et le périmètre du bassin versant sont les premières caractéristiques et les plus importants et dans notre cas, ils sont calculés par le logiciel ArcGis pro

Tableau 5 : Surfaces et périmètre par SBV.

Bassin	CODE SBV	Nom	Superficie (km ²)	Périmètre (km)
KEBIR- RHUMEL	1001	Oued Dehamecha	1054.88	169.3
	1002	Oued Kebir- Endja	1116.36	175.5
	1003	Oued Rhumel Amont	1218.05	162.64
	1004	Oued Rhumel- Seguin	1162.84	200.76
	1005	Oued Bou Merzoug	1868.28	217.07
	1006	Oued Rhumel- Smendou	1082.04	170.2
	1007	Oued Kebir Maritime	1319.51	187.14

b) Indice de compacité de Gravelius :

Il existe différents indicateurs morphologiques pour caractériser l'environnement, et les bassins versants peuvent également être comparés entre eux. L'indice le plus couramment utilisé est celui de Gravius (1914). Il est défini comme le rapport de la circonférence d'un bassin à la circonférence d'un cercle de même surface. Il donne une idée de la géométrie du bassin. Le coefficient est obtenu par l'application de la formule :

$$KG = \frac{P}{2\sqrt{\pi A}} = 0.28 \frac{P}{\sqrt{A}}$$

Avec :

KG : indice de compacité de Gravelius ;

A : surface du bassin versant (km²) ;

P : périmètre du bassin versant (km),

Si :

$KG \leq 1 \rightarrow$ Bassin versant de forme circulaire.

$KG = 1.12 \rightarrow$ Bassin versant de forme carrée.

$KG > 1.12 \rightarrow$ Bassin versant de forme allongée.

Les valeurs de l'indice de compacité (tableau) montrent que tous les bassins sont allongés, leur indice de forme reste élevé (varie entre 1.3 à 1.65)

Tableau 6 : Indices de compacité par sous bassin versant

Bassin	CODE SBV	Nom	Superficie (km ²)	Périmètre (km)	Indice de compacité (kc)
KEBIR- RHUMEL	1001	Oued Dehamecha	1054.88	169.3	1.46
	1002	Oued Kebir- Endja	1116.36	175.5	1.47
	1003	Oued Rhumel Amont	1218.05	162.64	1.3
	1004	Oued Rhumel- Seguin	1162.84	200.76	1.65
	1005	Oued Bou Merzoug	1868.28	217.07	1.41
	1006	Oued Rhumel- Smendou	1082.04	170.2	1.45
	1007	Oued Kebir Maritime	1319.51	187.14	1.44

c) Rectangle équivalent :

Ce concept a été introduit pour permettre une comparaison facile des bassins versants les uns avec les autres du point de vue de l'impact des caractéristiques des bassins versants sur l'écoulement. En supposant que, dans les mêmes conditions climatiques, l'écoulement d'un bassin versant donné soit approximativement le même que celui d'un rectangle de même superficie, avec le même coefficient de Gravelius et la même distribution altimétrique, on sait que la distribution du sol, de la végétation et de la densité de drainage, dans la courbe horizontale différentes régions intermédiaires est respectée. Il s'agit donc d'une transformation purement géométrique dans laquelle le contour du bassin devient un rectangle de même périmètre, la courbe horizontale de la droite parallèle au petit côté du rectangle, et le débouché d'un des petits côtés du rectangle que nous appelons le rectangle équivalent (Roche, 1963). Soit L et I la longueur et la largeur du rectangle, et P et A le périmètre et l'aire du bassin versant,

On a :

$$L = \frac{K_G \sqrt{A}}{1.12} \cdot \left(1 + \sqrt{1 - \left(\frac{1.12}{K_G} \right)^2} \right)$$

Les dimensions d'un rectangle équivalent donnent une longueur égale à 4 fois la largeur. Les valeurs de la longueur équivalente et de la largeur équivalente sont présentées dans le tableau suivant :

Tableau 7 : longueur équivalente et largeur équivalente par sous bassin versant

Bassin	Code SBV	Nom	Superficie (km ²)	Périmètre (km)	Indice compacité (kc)	Longueur Equivalente (km)	Largeur Equivalente (km)
Kebir- Rhumel	1001	Oued Dehamecha	1054.88	169.3	1.46	69.47	15.19
	1002	Oued Kebir- Endja	1116.36	175.5	1.47	72.31	15.44
	1003	Oued Rhumel Amont	1218.05	162.64	1.3	61.52	19.8
	1004	Oued Rhumel- Seguin	1162.84	200.76	1.65	87.01	13.36
	1005	Oued Bou Merzoug	1868.28	217.07	1.41	87.08	21.46
	1006	Oued Rhumel- Smendou	1082.04	170.2	1.45	69.54	15.56
	1007	Oued Kebir Maritime	1319.51	187.14	1.44	76.27	17.3

2.4.2. Paramètres de relief : [27]

a) Altitudes Caractéristiques :

Altitude maximale et minimale :

Elles sont obtenues directement à partir de cartes topographiques. L'altitude maximale représente le point le plus haut du bassin, tandis que l'altitude minimale représente le point le plus bas du bassin. Ces deux paramètres déterminent l'amplitude altimétrique du bassin versant et servent au calcul de la pente.

Altitude moyenne :

Elle se déduit directement de la courbe hypsométrique ou encore par l'application de la formule suivant :

$$H_{moy} = \frac{\sum(h_i - A_i)}{A}$$

Avec :

H_{moy} : altitude moyenne du bassin (m)

A_i : aire comprise entre deux courbes de niveau (Km²)

h_i : altitude moyenne entre deux courbes de niveau (m)

A : superficie totale du bassin versant (Km²)

Altitude médiane :

L'altitude médiane correspond à l'altitude lue au point d'abscisse 50% de la surface totale du bassin, sur la courbe hypsométrique. Cette grandeur se rapproche de l'altitude moyenne si la pente de la courbe est régulière.

L'altitude H95% :

L'altitude H95% correspond à l'altitude lue au point d'abscisse 95% de la surface totale du bassin, sur la courbe hypsométrique.

L'altitude H5% :

L'altitude H5% correspond à l'altitude lue au point d'abscisse 5% de la surface totale du bassin, sur la courbe hypsométrique.

Les hauteurs et leurs moyennes sont représentées dans le tableau suivant :

Tableau 8 : les altitudes Caractéristiques par sous bassin versant

Bassin	CODE SBV	S (km ²)	P (km)	Alt Min (m)	Alt Max (m)	Alt Moy (m)	Alt médiane	Alt H95%	Alt H5%
KEBIR- RHUMEL	1001	1054.88	169.3	283	1660	890.79	973.67	629.95	1364.9
	1002	1116.36	175.5	178	1366	659.12	646.38	262.18	1088.64
	1003	1218.05	162.64	695	1281	903.16	994.2	842.28	1173.1
	1004	1162.84	200.76	558	1370	828.56	923.23	740.81	1156.09
	1005	1868.28	217.07	558	1722	853.78	946.69	737.88	1173.1
	1006	1082.04	170.2	178	1316	584.54	639.92	349.58	1156.09
	1007	1319.51	187.14	1	1455	453.26	491.53	147.22	1142.6
	Moy			350.14	1452.86	739.03	802.23		

D'autre paramètre de relief :

- La courbe hypsométrique
- Les indices des pentes
- Dénivelée spécifique

2.4.3. Réseau hydrographique : [31]

Le réseau hydrographique se définit comme l'ensemble des cours d'eau naturels ou artificiels, permanents ou temporaires, qui participent à l'écoulement.

Il peut prendre une multitude de forme en fonction de plusieurs facteurs qui peuvent avoir une influence directe sur le degré de drainage du réseau hydrographique. Citons la nature et la structure géologique du substratum, la pente des terrains et les conditions climatiques de la région.

Les paramètres importants qui régissent le régime hydrologique d'un cours d'eau sont : la densité de drainage, la fréquence des talwegs, le coefficient de torrentialité, le rapport de confluence, le rapport des longueurs et le temps de concentration.

a) Description du réseau hydrographique :

Le bassin versant du Kebir-Rhumel s'étend sur une superficie de 8821.96 km², il est subdivisé en sept sous bassins versants et porte le numéro 10 selon la nomenclature adoptée par l'ANRH. Ce bassin versant comprend trois grands oueds : Boumerzoug, Rhumel-Enja et Kebir.

Les longueurs des Oueds permanents et temporaires par SB sont mentionnées dans le tableau suivant :

Tableau 9 : Longueur des oueds permanents et temporaires par sous-bassin versant

Bassin	Code SB	Nom SOUS-BASSIN	Longue des oueds permanents Lp (km)	Longueur des oueds temporaires Lt (km)	Longueur total L (km)
KEBIR-RHUMEL	1001	Oued DEHAMECHA	162,56	680,54	843,10
	1002	Oued KEBIR- ENDJA	156,19	767,21	923,41
	1003	Oued RHUMEL AMONT	115,97	756,67	872,63
	1004	Oued RHUMEL- SEGUIN	156,58	801,56	958,14
	1005	Oued BOU MERZOUG	199,81	1297,49	1497,30
	1006	Oued RHUMEL- SMENDOU	179,00	809,60	988,61
	1007	Oued KEBIR MARITIME	242,53	977,37	1219,90

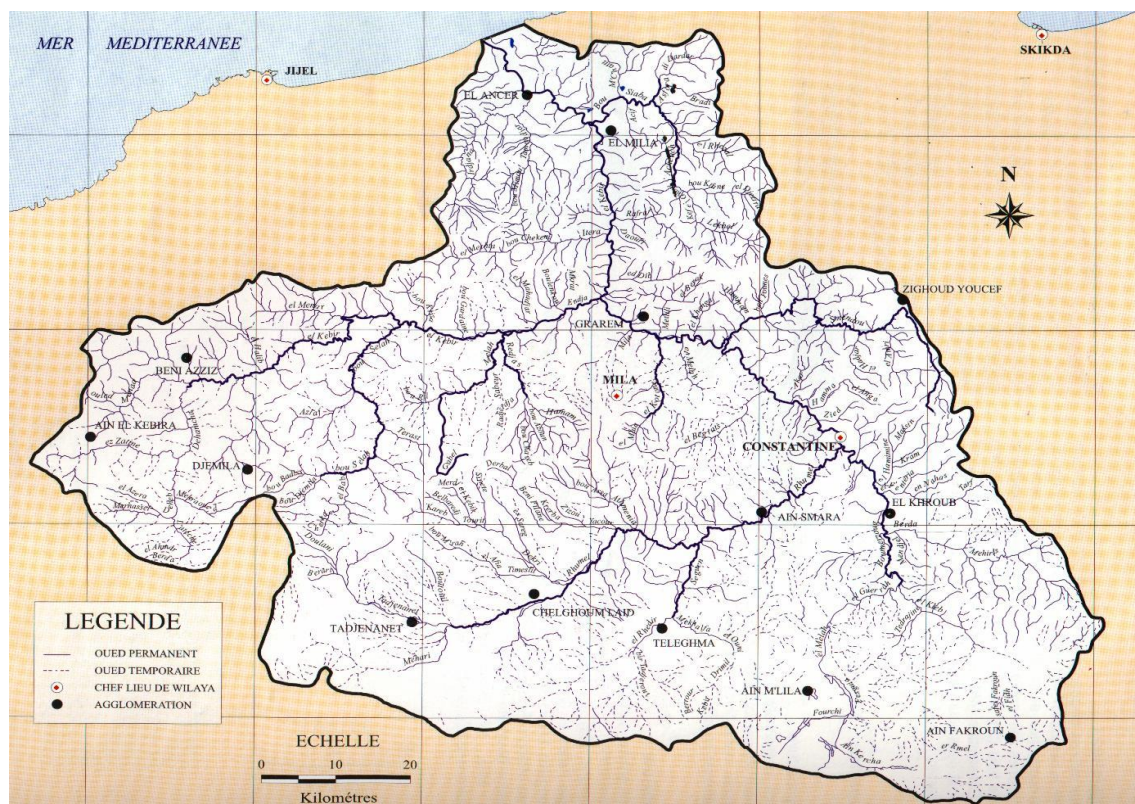


Figure 5 : carte de réseau hydrographique

3. Présentation et caractéristiques techniques du barrage de Beni Haroun :

Le barrage de Beni Haroun est l'un des ouvrages hydrauliques majeurs d'Algérie, tant par sa taille que par les fonctions stratégiques qu'il remplit. Situé sur l'oued Kébir, à environ 40 km au nord de Constantine et à 350 km à l'est d'Alger, il constitue un élément central du schéma national de mobilisation et de transfert des ressources en eau vers les hauts plateaux et les régions côtières de l'est algérien.



Figure 6 : vue sur barrage BH (Evacuateur de crue) (ANBT).2017

3.1. Fonction et rôle du barrage :

Mis en eau en 2003, le barrage de Beni Haroun a été conçu pour répondre à plusieurs objectifs prioritaires :

- L'alimentation en eau potable des grandes agglomérations de Mila, Constantine, Jijel, Oum El Bouaghi, Khenchela et Batna ;
- L'irrigation de plus de 30 000 hectares répartis dans les plaines de Téléghma, Chemora et Tafouna, via une dotation annuelle de 228 millions de m³ ;
- Le stockage stratégique pour le transfert inter-wilayas, dans le cadre d'un vaste projet piloté par l'ANBT .

Le barrage mobilise ainsi les ressources de l'oued Kébir et assure leur régulation saisonnière, en atténuant les effets des crues et des sécheresses. [33]

3.2. Caractéristiques techniques générales :

Tableau 10 : caractéristiques techniques de barrage de Beni Haroun (ANBT) [33]

Caractéristique	Valeur
Type d'ouvrage	Barrage-poids rectiligne en (BCR)
Année de construction	1996 – 2002
Mise en eau	2003
Hauteur sur fondation	118 m
Longueur en crête	710 m
Largeur en crête	8 m
Largeur maximale à la base	93 m
Niveau de crête	216,30 m
Côte des Plus Hautes Eaux (PHE)	214,80 m
Côte de retenue normale (RN)	200,00 m
Capacité initiale de stockage	997 hm ³
Capacité au levé 2004	996 hm ³
Apport moyen annuel	435 hm ³
Envasement moyen annuel	6 hm ³ /an
Volume de béton	1 800 000 m ³
Volume de terrassement	1 433 423 m ³
Volume d'injection	162 000 ml

3.3 Équipements hydrauliques et sécurité :

pour assurer une gestion fine des crues et des besoins en eau le barrage est équipé de :

Évacuateur de crue : seuil libre de 124 m de largeur situé au centre de l'ouvrage, prolongé par un coursier et un saut de ski, capable d'évacuer jusqu'à 13230 m³/s à PHE.

Vidanges de demi-fond : deux pertuis situés à la côte 140, débit maximal 670 m³/s.

Galeries de dérivation : deux galeries de 8 m de diamètre prévues pour des débits jusqu'à 3500 m³/s, utilisées durant la phase de construction.

Galeries de drainage : trois galeries principales aux cotes 100 m, 140 m et 175 m assurent la collecte des eaux d'infiltration et le suivi de la stabilité interne de l'ouvrage.

3.4 Fondations et géologie :

Le barrage repose sur un socle calcaire éocène très fracturé, incliné de manière complexe (pendage de 32° à 80° selon les rives), et recouvert localement par des marnes paléocènes et miocènes. Ce contexte géologique a nécessité la mise en œuvre :

- d'un rideau d'injection profond (jusqu'à 120 m) pour assurer l'étanchéité ;
- d'un système de drainage par forages inclinés (6 m d'espacement, inclinaison 15–25°) ;
- d'un voile de drainage dans le corps du barrage, à 7,5 m du parement amont.

Ces dispositifs assurent à la fois la stabilité du barrage et la maîtrise des pressions interstitielles.

3.5 Historique et évolution :

- 1996–2002 : réalisation de l'ouvrage par le bureau Tractebel et l'entreprise Dragados.
- 2003 : mise en eau, suivie de premières anomalies de sous-pressions détectées en rive gauche.
- 2004 : crue exceptionnelle du 4 janvier (~2000 m³/s), poussant à une réévaluation des conditions de sécurité.
- 2007–2014 : mise en place d'un panel d'experts, études géophysiques, renforcements par injections et améliorations continues du drainage (Trevi, Geomag, Hydrokarst...).

Aujourd'hui, le barrage fait l'objet d'une auscultation permanente, assurée par l'ANBT et les services techniques régionaux.

3.6 Données hydrologiques du bassin :

- Bassin versant total : 7725 km²
- Apports alimentés par :
- Oued Endja (ouest) – 2169 km², précipitations ~700 mm/an, topographie montagneuse (jusqu'à 1400 m)
- Oued Rhumel (est) – 5490 km², précipitations ~500 mm/an, terrain plus doux (jusqu'à 1300 m)
- Températures moyennes : de 5 °C en janvier à 25 °C en août
- Humidité relative : 50 % (été) à 70 % (hiver)
- période de crue : décembre à mai, avec des pics en janvier et mars

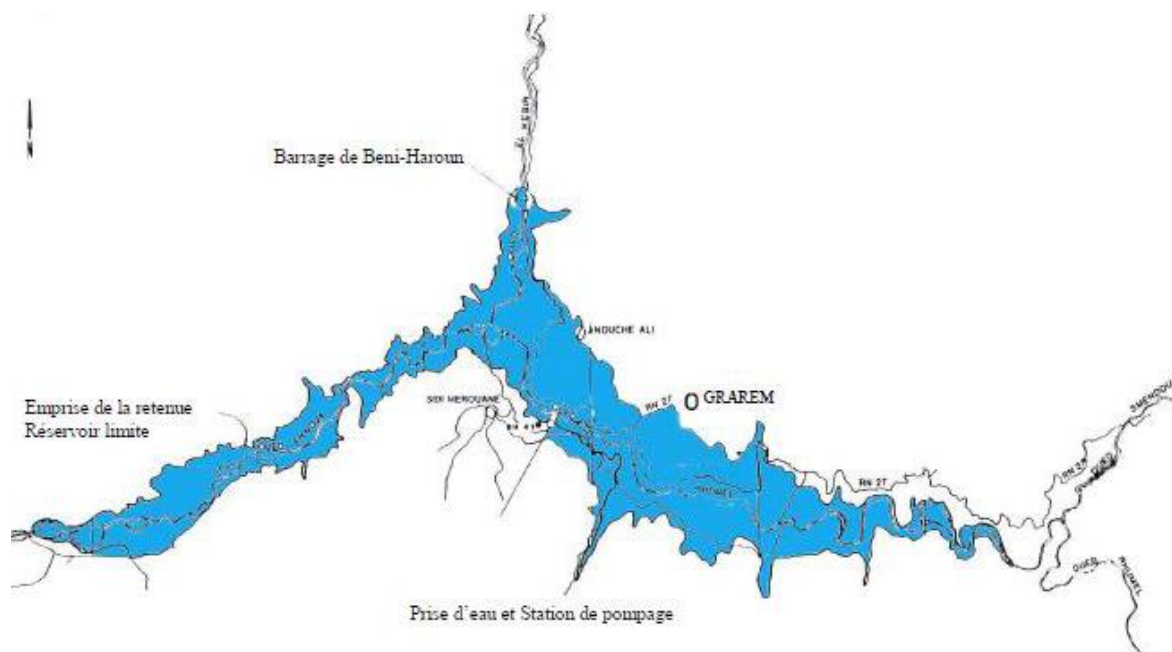


Figure 7 : Retenue du barrage de Béni Haroun (ANBT)

4. conclusion :

La présentation de la zone d'étude a permis de mieux cerner le cadre géographique, climatique, hydrologique et géologique . Ce contexte est essentiel pour comprendre les dynamiques environnementales et hydriques qui influencent directement le fonctionnement du système de régulation hydraulique. La variabilité climatique marquée, combinée à une topographie contrastée et à une forte dépendance aux apports hydriques, ce qui donne à cette région une sensibilité particulière face aux événements extrêmes tels que les sécheresses prolongées et les crues soudaines. Par ailleurs, les caractéristiques techniques du barrage étudié mettent en évidence son rôle stratégique dans la gestion des ressources en eau à l'échelle régionale. Cette contextualisation constitue ainsi un socle indispensable pour l'analyse des données historiques et l'évaluation des risques hydrologiques à travers les approches de modélisation développées dans les chapitres suivants.

CHAPITRE III : DONNEES ET METHODES

1. Introduction :

Ce chapitre présente en détail les données mobilisées ainsi que la démarche méthodologique suivie pour élaborer un modèle prédictif des événements hydrologiques extrêmes, notamment les crues et les sécheresses, dans le contexte du barrage de Beni Haroun. Il s'appuie sur l'exploitation de données hydrologiques, climatiques et opérationnelles fournies par l'Agence Nationale des Barrages et Transferts (ANBT).

Dans un premier temps, les sources de données sont décrites, suivies des étapes de prétraitement nécessaires à leur préparation. La deuxième partie est consacrée aux algorithmes de classification testés, aux critères retenus pour caractériser les événements extrêmes, ainsi qu'à la structuration des scripts d'implémentation.

L'ensemble de cette méthodologie, rigoureuse et reproductible, constitue le socle des analyses de performance et des représentations temporelles qui seront détaillées dans le chapitre suivant.

2. Données :

2.1. Sources et type de données utilisées :

Les données utilisées dans ce travail ont été fournies par la Direction du Contrôle, de la Maintenance et de l'Exploitation des Infrastructures (DCMEI), un département de l'Agence Nationale des Barrages et Transferts (ANBT), basée à Alger. Cette institution est chargée de la gestion technique et opérationnelle des barrages à l'échelle nationale.

Les informations recueillies concernent le barrage de Beni Haroun et son système en général. Elles se répartissent en deux grandes catégories :

Données techniques et descriptives :

- La monographie du barrage
- La fiche technique

Données d'exploitation journalières et mensuelles :

Ces données proviennent des bilans d'exploitation et comprennent les variables suivantes :

- | | |
|---|---|
| • Date | • Volume distribué pour l'alimentation en eau potable (AEP) |
| • Hauteur d'eau (cote) | • Fuites |
| • Superficie de la surface d'eau | • Apports (affluents) |
| • Volume stocké | • Pluie journalière (P _j) |
| • Taux d'évaporation (EVAP) | • Pluie mensuelle (P _m) |
| • Volume relâché par la vidange de fond (VDF) | |
| • Volume déversé (DEV) | |

2.2. Prétraitement des données :

Les séries temporelles fournies par l'ANBT étaient globalement bien organisées, vérifiées et prêtes à l'emploi. Elles ont été préparées et validées par les ingénieurs de l'ANBT dans un but opérationnel. Cela a permis de limiter les traitements préalables à un simple nettoyage manuel, ainsi qu'à une réorganisation et un ajustement des données en fonction des besoins spécifiques de notre étude.

Un examen minutieux des séries a été réalisé afin de détecter d'éventuelles incohérences, telles que des doublons, des valeurs non justifiées ou des variations anormales.

2.2.1. Vérification et nettoyage :

Le travail de prétraitement s'est limité à des interventions mineures :

- Suppression des lignes de synthèse annuelle présentes à la fin de chaque année dans les fichiers, afin de conserver uniquement les valeurs journalières;
- Vérification de l'absence de valeurs en double ou incohérentes ;
- Harmonisation de l'ordre chronologique et des noms de colonnes.

Aucune interpolation ou correction significative n'a été nécessaire, les données étant déjà cohérentes et complètes.

2.2.2. Calculs des dérivés et indicateurs

Pour enrichir l'analyse et préparer les données à la modélisation, plusieurs variables ont été dérivées à partir des séries principales, selon des formules simples et bien établies en hydrologie.

- **ΔV (variation journalière de volume):** Il est calculé, dans le cas d'une série journalière, comme la différence entre les apports d'eau quotidiens (affluents) et les pertes de volume (défluent). Cette valeur représente directement le bilan hydrique net du réservoir. Dans certains cas, elle peut également être estimée selon la formule suivante :

$$\Delta V = \text{Apports} - \text{Total defluent}$$

Où : Total défluent est défini comme la somme des pertes de volume, à savoir :

$$\text{Total défluent} = \text{Evaporation} + \text{Fuite} + \text{Déverse} + \text{Lâchers} + \text{Consommations (AEP/AEI)}$$

Cette variable permet de suivre la dynamique de vidange du réservoir et d'évaluer son comportement en période d'exploitation ou de stress hydrique.

- **Indice de Précipitation Standardisé (SPI) :** utilisé pour caractériser les anomalies de précipitations à différentes échelles temporelles. Deux méthodes ont été envisagées : [8]

- **Méthode rigoureuse:** ajustement d'une loi gamma sur les précipitations mensuelles agrégées, suivi d'une transformation en loi normale standard ;
- **Méthode simplifiée :** calcul du **score z** standardisé pour chaque mois :

$$SPI_i = \frac{P_i - \mu}{\sigma}$$

Où P_i est la précipitation du mois , μ la moyenne, et σ est l'écart-type des précipitations pour ce mois sur toute la période.

Cette méthode, bien qu'elle repose sur l'hypothèse d'une distribution normale, est fréquemment utilisée dans les analyses opérationnelles. C'est celle que nous avons utilisée pour calculer le SPI dans notre cas.

- **Indice de Sécheresse Standardisé (SDI) :** analogue au SPI, il est calculé sur les apports en eau ou les volumes stockés au lieu des précipitations. Le même principe de standardisation s'applique, avec :

$$SDI_i = \frac{V_i - \mu}{\sigma}$$

où V_i est soit le volume stocké, soit l'apport au mois , selon l'indicateur ciblé. Cet indice permet de quantifier les déficits ou excès hydriques par rapport au comportement moyen du barrage.

2.2.3. Standardisation des données pour la modélisation

Afin d'adapter les données aux exigences et pour être compatible avec tous les algorithmes de Machine Learning, notamment en ce qui concerne l'équilibre des échelles entre les différentes variables, une opération de standardisation a été appliquée. Cette méthode consiste à normaliser les valeurs de chaque variable numérique dans un intervalle [0 ; 1], selon la formule suivante :

$$X_{standardise} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Cette transformation conserve la distribution relative des données et permet d'éviter les biais d'interprétation dus aux différences d'échelle entre les variables climatiques (en mm) et hydrologiques (en hm³) et la grande variation des valeurs de ces variables.

3. Méthodes :

3.1. Modélisation prédictive par Machine Learning :

L'objectif de cette étape est de développer un modèle d'apprentissage automatique capable de prédire les situations extrêmes (crues ou sécheresses) susceptibles d'affecter un barrage et d'influencer la sécurité ainsi que la gestion du réservoir. Cette prédiction repose sur un ensemble

de variables explicatives d'ordre hydrologique et climatique. La modélisation adoptée utilise une approche de classification supervisée.

Contrairement à une approche basée sur un seul algorithme, ce travail s'appuie sur la comparaison de plusieurs modèles de machine learning, choisis pour leur pertinence dans le domaine hydrologique et leur usage fréquent dans la littérature scientifique.

3.1.1. Choix des algorithmes

Les modèles suivants ont été testés :

- Random Forest : algorithme robuste basé sur une combinaison d'arbres de décision. Il permet de gérer les relations non linéaires et de déterminer l'importance des variables.
- XGBoost (Extreme Gradient Boosting) : algorithme de boosting par gradient extrêmement efficace pour la modélisation de structures complexes.
- CatBoost : une alternative optimisée de boosting, performante sur les petits ensembles de données, avec une bonne gestion des variables catégorielles et des valeurs manquantes.
- LightGBM (Light Gradient Boosting Machine) : modèle de boosting très rapide et efficace, adapté aux grands ensembles de données et à la classification multiclassées.

3.1.2. Préparation des données

Dans le cadre d'une approche de Machine Learning, les données doivent être organisées selon deux catégories :

1. Les variables explicatives (Features) :

Il s'agit des variables d'entrée du modèle, c'est-à-dire des observations mesurées ou calculées susceptibles d'influencer l'occurrence d'un événement extrême. Les prédicteurs utilisés proviennent des données prétraitées et se répartissent comme suit :

Variables temporelles : année, mois

Variables climatiques : pluie journalière (P_j)

Variables hydrologiques : volume stocké, variation de volume (ΔV), total défluent, évaporation (EVAP)

Indices normalisés : SPI (Standardized Precipitation Index)

2. La variable cible (Target) :

Afin de permettre l'entraînement de modèle, il faut d'abord définir et transformer les données continues en classes cibles représentatives des situations hydrologiques extrêmes (crue ou sécheresse) observées dans la série temporelle. Ce travail de labellisation repose sur des critères hydrologiques simples et reproductibles.

Afin de classer automatiquement les situations hydrologiques une classe multicatégorie (C/N/S) a été introduite pour intégrer simultanément les deux types d'événements dans une seule variable à 3 modalités :

- $0 \rightarrow$ Sécheresse (S)
- $1 \rightarrow$ Situation normale (N)
- $2 \rightarrow$ Crue (C)

Cette classification repose sur une approche basée sur les centiles, appliquée aux variables principales de la série temporelle.

- Une situation est classée comme crue ($C = 2$) si les précipitations ou les apports dépassent le 90e centile de leur distribution respective.
- Elle est considérée comme sécheresse ($S = 0$) si l'une des conditions suivantes est remplie : précipitations, apports, variation de volume ou volume inférieurs au 10e centile.
- Toutes les autres situations sont considérées comme normales ($N = 1$).

Cette méthode de classification, plus robuste que des seuils fixes, permet de détecter les extrêmes hydrologiques en s'adaptant à la distribution réelle des variables du site étudié. Les valeurs seuils (P90, P10) ont été calculées sur la base de la série complète.

3.1.3. Environnement d'exécution :

L'ensemble des modèles de ML développés dans ce travail a été implémenté et exécuté dans un environnement Python à l'aide de la plateforme Google Colab. Ce choix s'explique par plusieurs avantages pratiques et techniques :

- Accessibilité : Google Colab est une interface gratuite en ligne, accessible depuis n'importe quel navigateur, sans besoin d'installation locale.
- Puissance de calcul : l'environnement offre l'accès à des ressources de calcul dans le cloud (processeurs performants, GPU si nécessaire), adaptées à l'entraînement de plusieurs modèles sur des volumes de données hydrologiques modérés.
- Écosystème riche : il permet l'utilisation directe des bibliothèques scientifiques standard de Python.

3.1.4. Structure générale des scripts de modélisation :

Pour assurer la comparabilité des résultats et la cohérence méthodologique, l'ensemble des algorithmes testés dans cette étude a été implémenté selon une structure de script commune, exécutée sur Google Colab en environnement Python. Chaque script suit une séquence standardisée comprenant les étapes suivantes :

1. Chargement des bibliothèques et des données :

La première étape consiste à importer les bibliothèques nécessaires (pandas, numpy, sklearn, matplotlib, seaborn, etc.) et à charger le fichier de données contenant les variables hydrologiques, climatiques et la classe cible.

2. Préparation des données :

- La conversion de la variable cible en format catégorique.
- La sélection des variables explicatives
- La vérification de la multicolinéarité par le calcul du VIF (Variance Inflation Factor) ; Avant l'entraînement, les variables prédictives ont été évaluées pour détecter la multicolinéarité à l'aide du facteur d'inflation de la variance (VIF) et des valeurs de tolérance. Les variables se situaient dans des seuils statistiques acceptables (VIF proche de la limite de 5).

3. Séparation du jeu de données : le dataset est divisé en deux sous-ensembles

- 80 % pour l'entraînement
- 20 % pour le test

La séparation est stratifiée pour respecter la distribution des classes.

4. Entraînement du modèle d'un des algorithmes choisis :

- RandomForestClassifier,
- LightGBM
- CatBoost
- XGBoost

5. Évaluation du modèle par :

- Le rapport de classification : précision, rappel, F1-score pour chaque classe.
- La matrice de confusion : texte et graphique (avec seaborn).
- L'analyse de l'importance des variables (feature importance), affichée sous forme de tableau et de diagramme en barres.
- La courbe ROC multiclass avec calcul de l'AUC (Area Under Curve) pour chaque classe via la méthode de binarisation des sorties.

6. Visualisation temporelle :

Une carte temporelle (heatmap) est générée pour visualiser l'évolution des événements extrêmes dans le temps pour chaque classe. Cette visualisation permet d'identifier les saisons dominantes pour les crues ou les sécheresses, et de mettre en évidence les périodes calmes ou critiques.

7. Analyse de tendance avec le test de Mann-Kendall

Pour chaque classe (Crue, Normale, Sécheresse), une série temporelle annuelle est extraite et soumise au test non paramétrique de Mann-Kendall, afin d'identifier l'existence d'une tendance significative à la hausse ou à la baisse dans le temps.

4. Conclusion :

Ce chapitre a permis d'établir les bases techniques nécessaires à la mise en œuvre d'une modélisation prédictive, en précisant les données exploitées, les variables dérivées, ainsi que la stratégie de classification supervisée des situations hydrologiques.

L'attention portée à la qualité des séries temporelles, à leur cohérence statistique, ainsi qu'aux transformations appliquées (SPI, SDI, ΔV , etc.) constitue un gage de fiabilité pour les prédictions issues des modèles.

En mobilisant différents algorithmes de Machine Learning dans un cadre expérimental rigoureux, cette étude vise à comparer objectivement plusieurs approches.

L'ensemble du processus, depuis la structuration des données jusqu'à la définition des classes cibles, est pensé pour fournir un outil d'aide à la décision aux gestionnaires de barrages, face aux défis posés par l'intensification des extrêmes climatiques.

CHAPITRE IV : RESULTATS ET DISCUSSIONS

1. Introduction :

Dans ce chapitre en va les résultats détaillés de la modélisation, de l'évaluation des performances des algorithmes de Machine Learning, et de l'analyse temporelle des événements hydrologiques extrêmes dans le contexte du barrage étudié. L'objectif principal est de tester la capacité des modèles à classer les situations hydrologiques en trois catégories : crue, normale et sécheresse, à partir d'un ensemble de variables climatiques et opérationnelles.

Les résultats obtenus sont analysés de manière critique, en s'appuyant à la fois sur la performance statistique des modèles et sur la cohérence physique des prédictions. Cette approche permet d'apporter un éclairage sur la pertinence des outils de Machine Learning pour l'anticipation et la gestion des risques liés aux événements hydrologiques extrêmes.

2. Evaluation des performances :

2.2 Analyse de multicollinéarité : Variance Inflation Factor (VIF) et Tolérance

L'analyse a été réalisée pour les quatre algorithmes testés, bien que ceux-ci soient relativement robustes à la multicollinéarité. Les résultats sont similaires entre modèles, ce qui indique une bonne cohérence structurelle des données.

Tableau 11 : résultats de VIF et tolérance

Variable	Mois	ΔV	P_j	Défluent	Année	Volume	EVAP	SPI	P_m
VIF	1.12	1.22	1.25	1.28	2.043	2.2	2.25	4.41	5.59
Tolérance	0.89	0.82	0.8	0.781	0.49	0.455	0.45	0.23	0.18

Les observations principales sont les suivantes :

Tous les VIF sont inférieurs à 5, le seuil généralement admis comme critique. Cela indique une absence de multicollinéarité significative entre les variables explicatives. Sauf pour les valeurs les plus élevées qui concernent les variables P_m (VIF = 5.59) et SPI (VIF = 4.41), qui restent toutefois dans une zone tolérable. La légère élévation du VIF pour P_m s'explique par sa forte corrélation avec d'autres variables climatiques, notamment le SPI. Cependant, cette variable constitue un facteur hydrologique central pour la détection des sécheresses, et son importance dans la modélisation justifie pleinement sa conservation. [34]

Les autres variables (ΔV , EVAP, Défluent, SPI, Pluie, Mois) présentent des VIF très faibles, souvent proches de 1, ce qui témoigne d'une indépendance statistique satisfaisante. Et pour les valeurs de tolérance (inverse du VIF) confirment ces résultats : toutes sont supérieures à 0.4, ce qui exclut les situations de forte interdépendance sauf aussi pour (SPI) et (P_m)

2.2 Rapport des classifications :

Le tableau ci-dessous présente un résumé des métriques obtenues pour chaque classe.

Tableau 12 : indicateurs de performance par classe

Classe	Algorithme	Précision	Rappel	F1-score	Support
Crue	Random Forest	0.95	0.97	0.96	268
	LightGBM	0.98	0.96	0.97	268
	CatBoost	0.94	0.97	0.95	268
	XGBoost	0.98	0.96	0.97	268
Normale	Random Forest	0.89	0.93	0.91	994
	LightGBM	0.88	0.96	0.92	994
	CatBoost	0.87	0.98	0.92	994
	XGBoost	0.89	0.95	0.92	994
Sécheresse	Random Forest	0.77	0.64	0.7	317
	LightGBM	0.83	0.63	0.71	317
	CatBoost	0.93	0.55	0.69	317
	XGBoost	0.82	0.63	0.72	317

a) Classe “Crue” :

Tous les modèles affichent des résultats très satisfaisants pour la détection des crues :

- Précision entre 0.94 (CatBoost) et 0.98 (LightGBM, XGBoost)
- Rappel constant à 0.96 ou 0.97 pour tous les modèles
- F1-score très élevé, atteignant 0.97 pour LightGBM et XGBoost

Interprétation :

Les crues sont des événements bien identifiés par l'ensemble des modèles. Cela s'explique par leur signature hydrologique claire, marquée par des pics soudains de précipitation (P_j), des hausses rapides de volume (ΔV), et des apports importants. Ces signaux nets facilitent leur détection par les algorithmes, même sans recalibrage complexe.

b) Classe “Normale” :

Les résultats pour les événements normaux sont également très solides :

- F1-score entre 0.91 (Random Forest) et 0.92 (LightGBM, CatBoost, XGBoost)
- Rappel particulièrement élevé pour CatBoost (0.98), ce qui montre une excellente capacité à ne pas confondre un mois normal avec une situation extrême.

Interprétation :

La classe "normale", largement dominante dans les données (support = 994), est bien apprise par les modèles. Sa fréquence élevée améliore l'apprentissage, ce qui garantit un bon équilibre entre la détection d'événements extrêmes et la reconnaissance fiable des conditions stables.

c) Classe “Sécheresse” :

C'est la classe la plus difficile à prédire, avec des résultats plus hétérogènes :

- Rappel faible pour tous les modèles, variant de 0.55 (CatBoost) à 0.64 (Random Forest, XGBoost)
- F1-score oscillant entre 0.69 (CatBoost) et 0.72 (XGBoost)
- CatBoost se distingue par une précision élevée (0.93) mais un rappel plus faible, suggérant qu'il prédit peu de sécheresses, mais avec justesse.

Interprétation :

La sécheresse étant un phénomène progressif et multiparamétrique (faibles apports, évaporation continue, SPI négatif), sa détection automatique est plus complexe. Les modèles ont tendance à confondre certains mois secs avec des conditions normales, en particulier lorsqu'aucun seuil fort n'est franchi.

Les performances des quatre algorithmes testés sont globalement élevées pour la détection des crues et des états normaux, mais varient davantage pour la détection des épisodes de sécheresse.

2.3 indicateurs numériques de performance (Models metrics) :

Le tableau ci-dessous présente les performances globales de chacun des quatre algorithmes appliqués à la modélisation des états hydrologiques : [34] [35]

Tableau 13 : performances globales des algorithmes

modélé	précision	rappel	f1-score	Exactitude
Random Forest	0.88	0.89	0.88	0.89
LightGBM	0.9	0.9	0.89	0.9
CatBoost	0.89	0.89	0.88	0.89
XGBoost	0.89	0.89	0.89	0.89

Le modèle LightGBM affiche la meilleure performance globale sur l'ensemble des indicateurs, avec une exactitude et un rappel de 90 %, indiquant une capacité légèrement supérieure à bien détecter les trois classes. XGBoost offre un excellent compromis, avec un F1-score parfait à 0.89, ce qui reflète une stabilité équilibrée entre précision et rappel. Random Forest et CatBoost présentent des scores très proches, avec une performance robuste mais légèrement en retrait.

Ces résultats montrent que tous les modèles sont performants et fiables pour la classification des états hydrologiques, avec des écarts relativement faibles entre eux. Toutefois, les algorithmes basés sur le boosting (LightGBM, XGBoost) semblent légèrement mieux adaptés au problème, notamment grâce à leur capacité à gérer les classes déséquilibrées et les relations complexes entre les variables climatiques et hydrologiques.

2.4 Matrices de confusion :

Les matrices de confusion permettent de visualiser en détail la nature des erreurs de classification pour chaque algorithme. Chaque ligne représente les classes réelles, tandis que chaque colonne représente les prédictions du modèle. Cela permet de comprendre non seulement combien de cas sont correctement classés, mais aussi où les erreurs se produisent le plus souvent, notamment entre les classes « crue », « normale » et « sécheresse ».

a) **Random Forest**

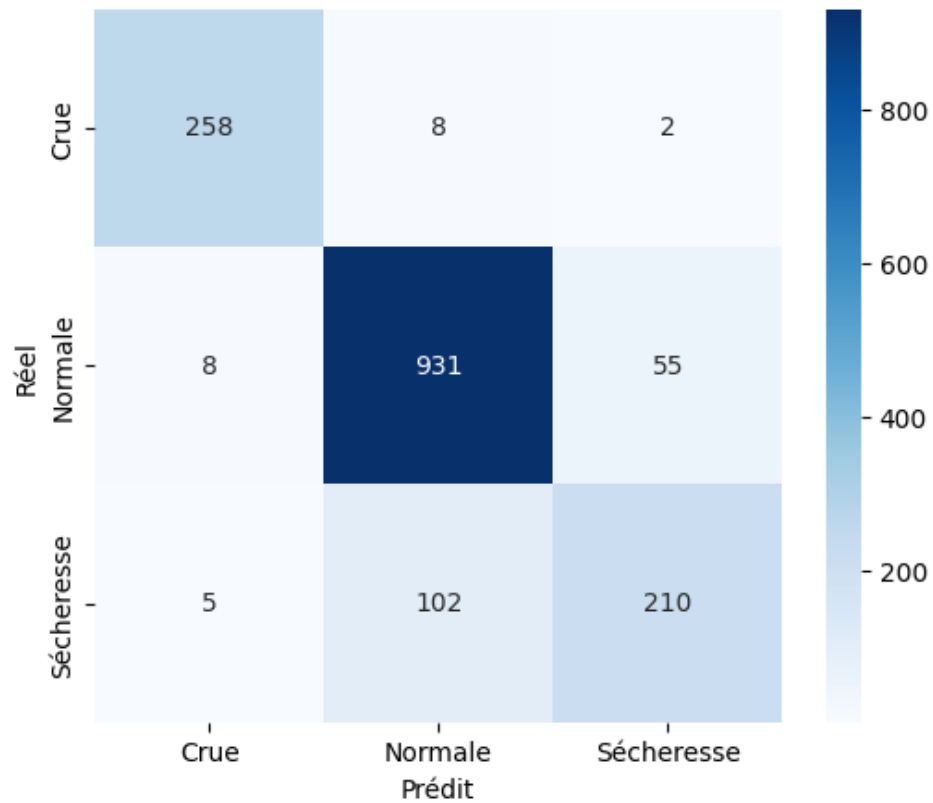


Figure 8 : matrice de confusion de Random Forest

- Cru : très bien détectée (258/268 correctes)
- Normale : 55 sécheresses confondues comme normales
- Sécheresse : erreurs fréquentes avec la classe normale (102 confusions)

b) LightGBM :

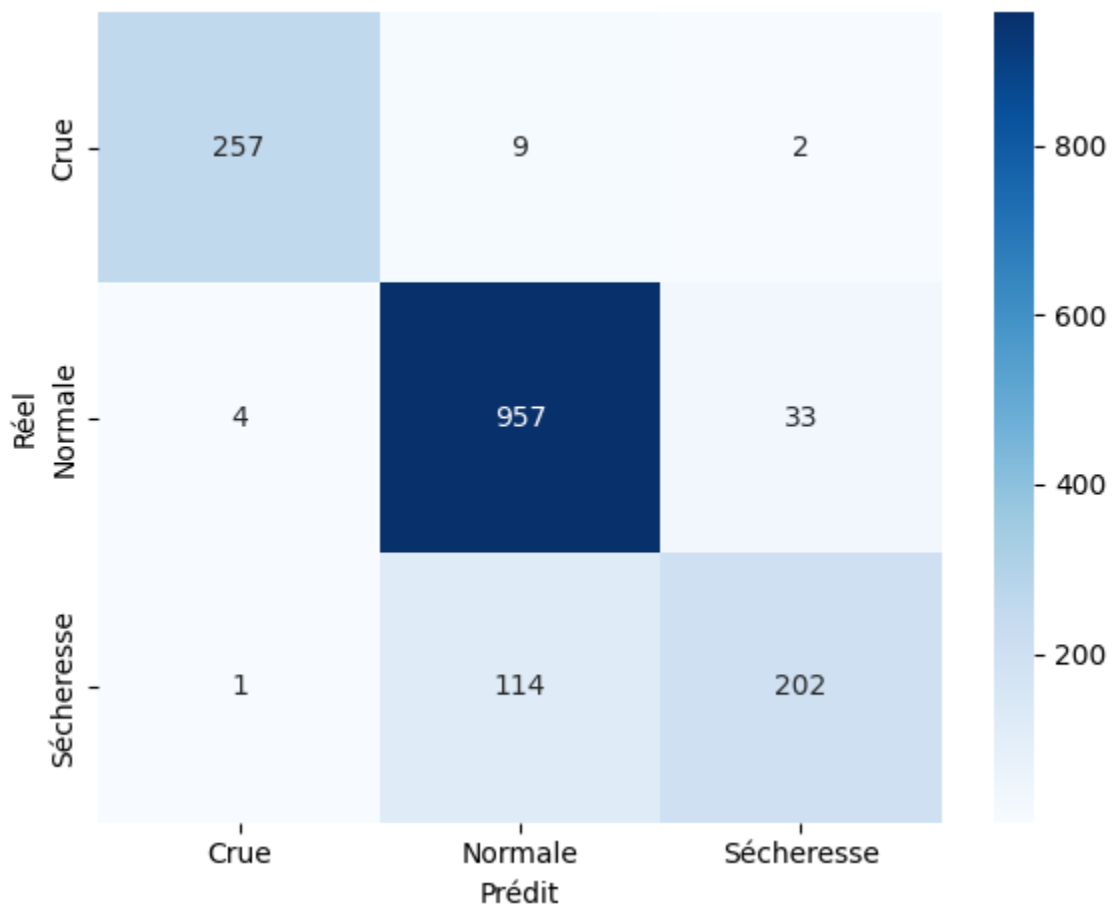


Figure 9 : matrice de confusion de lightGBM

- Crue : peu de confusion avec les autres classes
- Normale : un mieux rappel pour la classe « normale »
- Sécheresse : encore confondue avec normale (114 cas)

c) CatBoost :

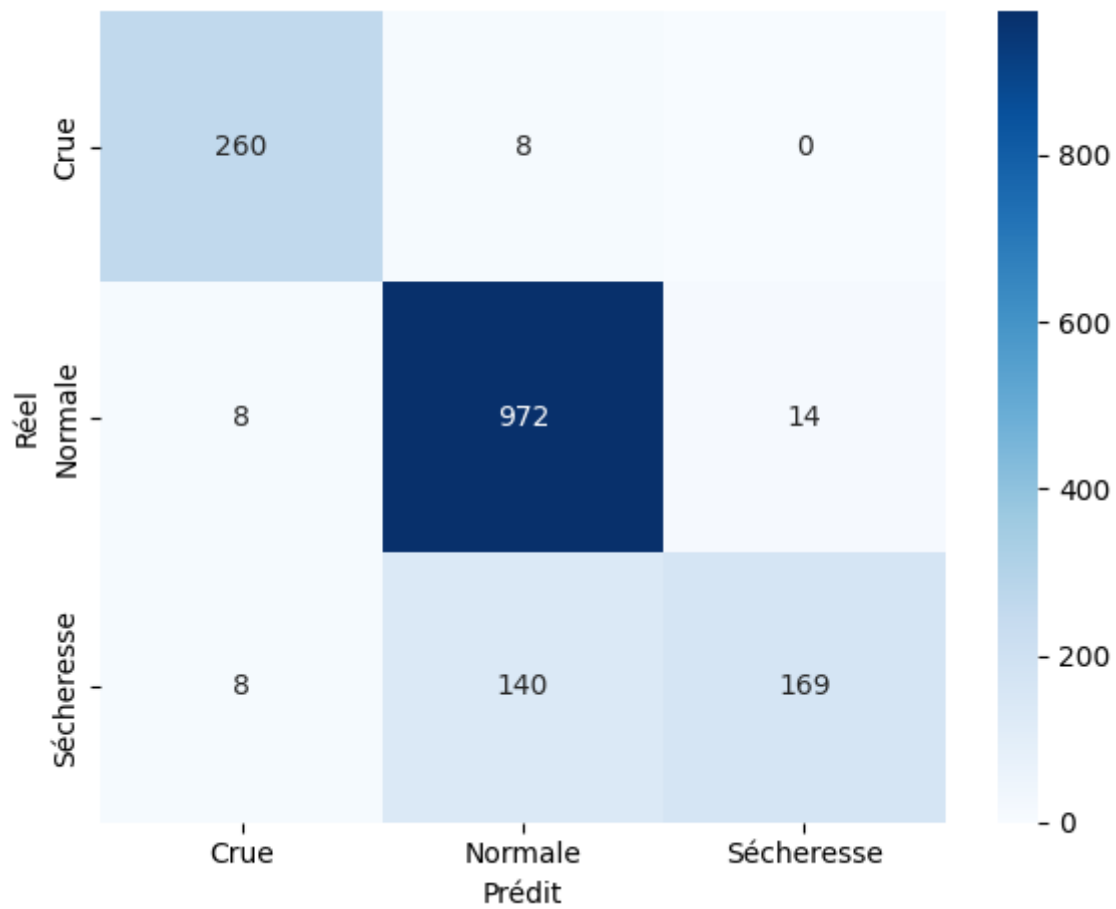


Figure 10 : matrice de confusion de CatBoost

- Crue : excellente détection des crues (aucune crue prédite comme sécheresse)
- Normale : très bon score sur les normales
- Sécheresse : forte confusion sécheresse → normale (140 cas)

d) XGBoost :

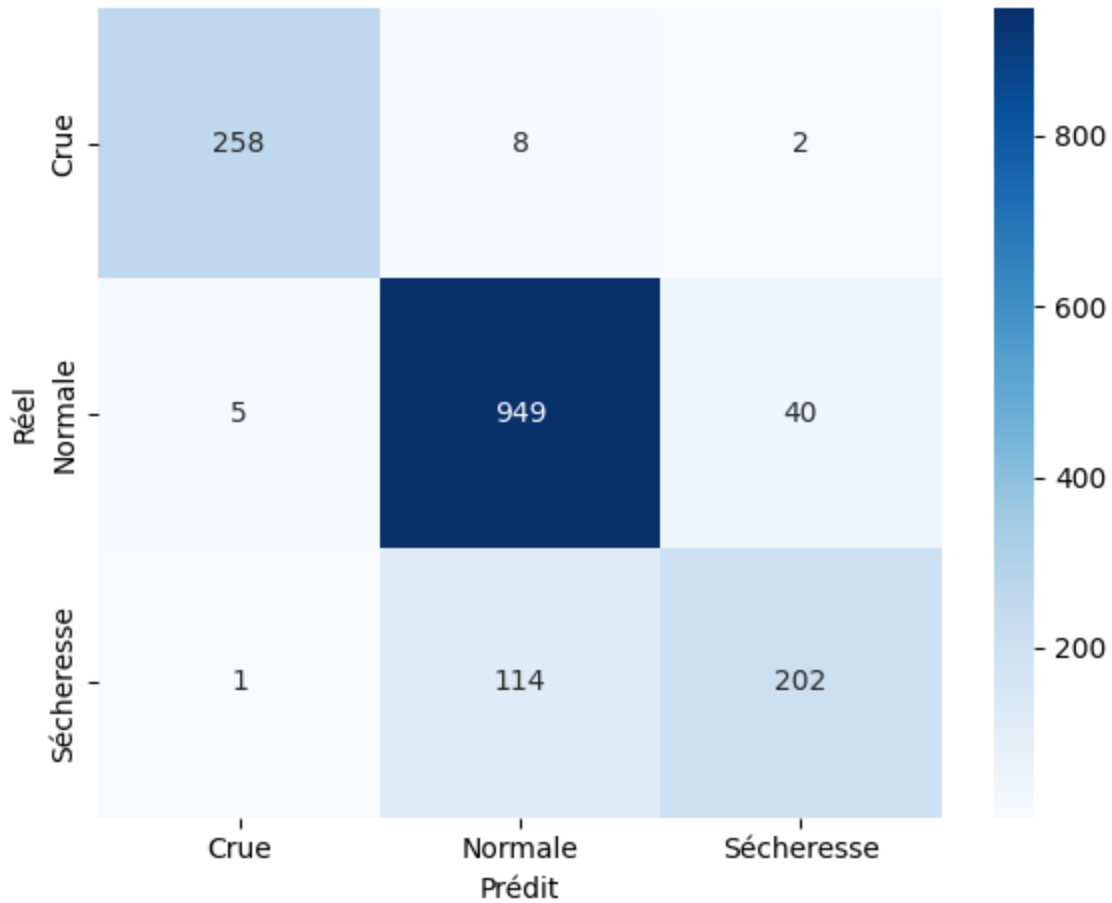


Figure 11 : matrice de confusion de XGBoost

- Similaire à LightGBM
- XGBoost a une bonne séparation des classes extrêmes
- La classe sécheresse souvent prédite comme normale, mais avec un meilleur équilibre que CatBoost

Tableau 14 : explication pour les erreurs amener

Erreurs typiques	Explication
Sécheresse / Normale	Phénomènes plus progressifs, moins marqués (difficiles à distinguer)
Crue / Normale	Très peu d'erreurs (modèles très fiables sur cette classe)
Normale / Sécheresse	Difficile de capter les sécheresses modérées proches de la normale

CatBoost se distingue par une très bonne détection des crues et normales, mais tendue à sous-détecter les sécheresses (rappel faible).

LightGBM et XGBoost offrent un meilleur équilibre, avec moins d'erreurs sècheresse / normale, ce qui justifie leur supériorité globale déjà observée dans les autres métriques.

2.5 Importance des variables :

L'analyse de l'importance des variables (feature importance) permet d'identifier quels paramètres influencent le plus la décision des modèles. Cela est particulièrement utile pour interpréter les prédictions, valider leur cohérence avec les processus hydrologiques, et guider la prise de décision en gestion des barrages.

le tableau ci-dessous présente les huit variables les plus importantes pour chaque algorithme , exprimées en pourcentage de contribution à la prédiction.

Tableau 15 : Tableau comparatif de l'importance des variables

Variable	RF (%)	LightGBM (%)	CatBoost (%)	XGBoost (%)
P_j	22.46	6.47	28.22	49.67
Défluent	18.21	14.37	26.44	15.93
Volume	17.43	12.53	14.64	11.77
EVAP	16.54	25.17	5.03	2.3
Année	7.66	8.71	4.02	2.59
ΔV	6.45	9.68	11.25	9.95
P_m	4.17	9.56	3.06	1.98
Mois	3.8	6.33	6.17	4.05
SPI	3.28	7.18	1.18	1.76

XGBoost attribue près de 50 % de l'importance à la pluie journalière maximale (P_j), ce qui confirme sa capacité à détecter efficacement les crues soudaines. CatBoost privilégie également (P_j) et (Défluent), qui représentent ensemble plus de 54 % de l'importance totale, illustrant sa forte sensibilité aux variables directement liées aux événements extrêmes. En revanche, LightGBM accorde davantage de poids à des variables plus régulières ou lentes à réagir, telles que EVAP (25.17 %), SPI, et P_m , ce qui peut expliquer ses performances équilibrées entre crue et sécheresse, en captant aussi les conditions climatiques prolongées.

Pour Random Forest, l'importance est répartie de manière plus équilibrée entre plusieurs variables : P_j , Défluent, Volume, EVAP et ΔV présentent toutes un poids relativement similaire. Cette approche plus généraliste pourrait expliquer sa robustesse globale, mais aussi des performances un peu plus modérées sur les classes minoritaires, notamment la sécheresse.

Malgré ces différences entre les modèles, quatre variables ressortent systématiquement comme importantes :

- Pluie journalière (P_j) – indicateur clé de crue
- Défluent – reflète l'activité hydraulique réelle du barrage
- Volume stocké – indicateur direct du niveau du réservoir
- ΔV – reflète les changements dynamiques rapides

Ces résultats sont cohérents sur le plan physique : les crues sont associées à des hausses brutales de pluie et de volume, tandis que la sécheresse dépend davantage de la précipitation, de l'évaporation et du déficit de réapprovisionnement.

2.6 Courbes ROC et AUC :

Les courbes ROC (Receiver Operating Characteristic) permettent d'évaluer la capacité des modèles à distinguer entre les différentes classes, indépendamment du seuil de classification choisi. Elles tracent le taux de vrais positifs (rappel) en fonction du taux de faux positifs ($1 - \text{spécificité}$) à tous les seuils possibles. [36]

Tableau 16 : Résultats des AUC par classe

Classe	Random Forest	LightGBM	CatBoost	XGBoost
Crue	1	1	1	1
Normale	0.94	0.96	0.95	0.95
Sécheresse	0.92	0.94	0.93	0.94

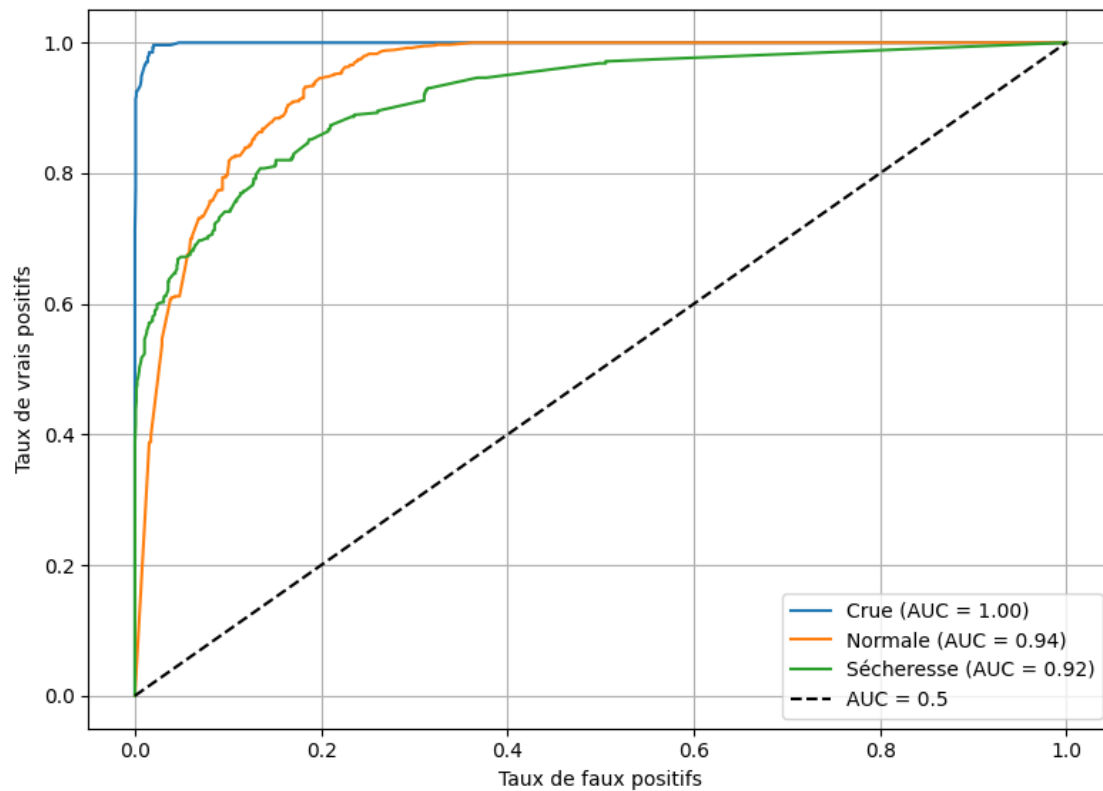


Figure 12 : courbe ROC avec indicateur AUC pour 'Random Forest'

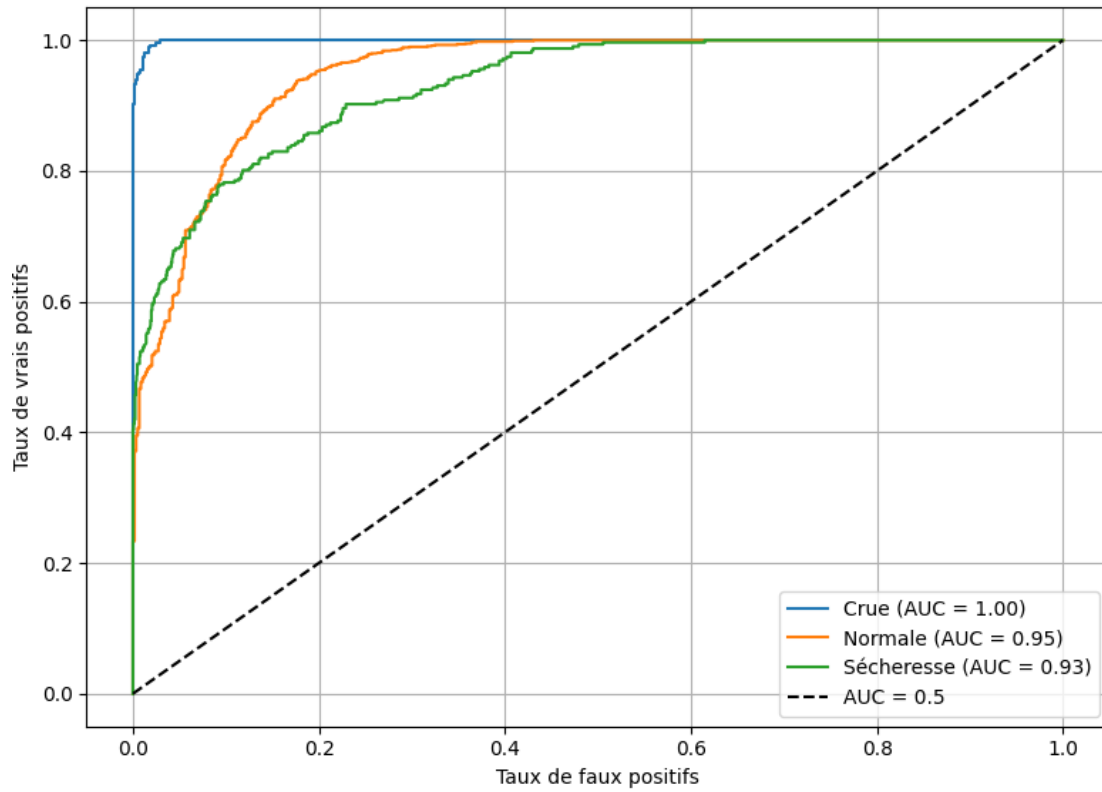


Figure 13 : courbe ROC avec indicateur AUC pour 'LightGBM'

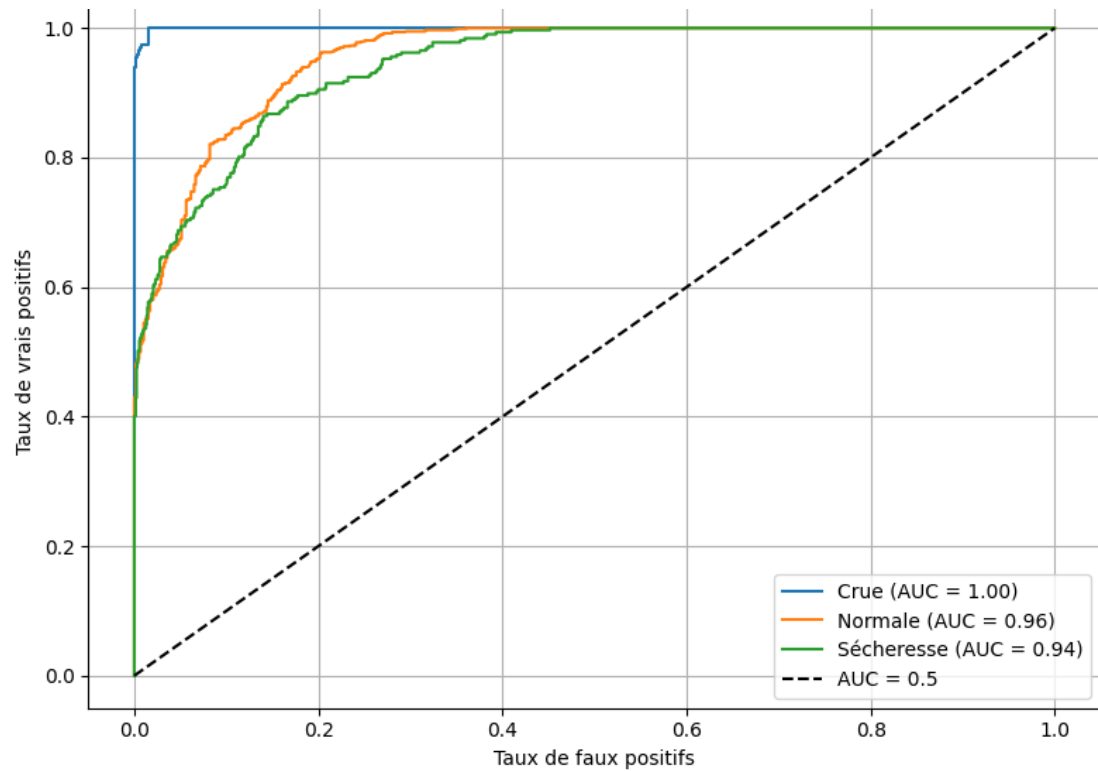


Figure 14 : courbe ROC avec indicateur AUC pour 'CatBoost'

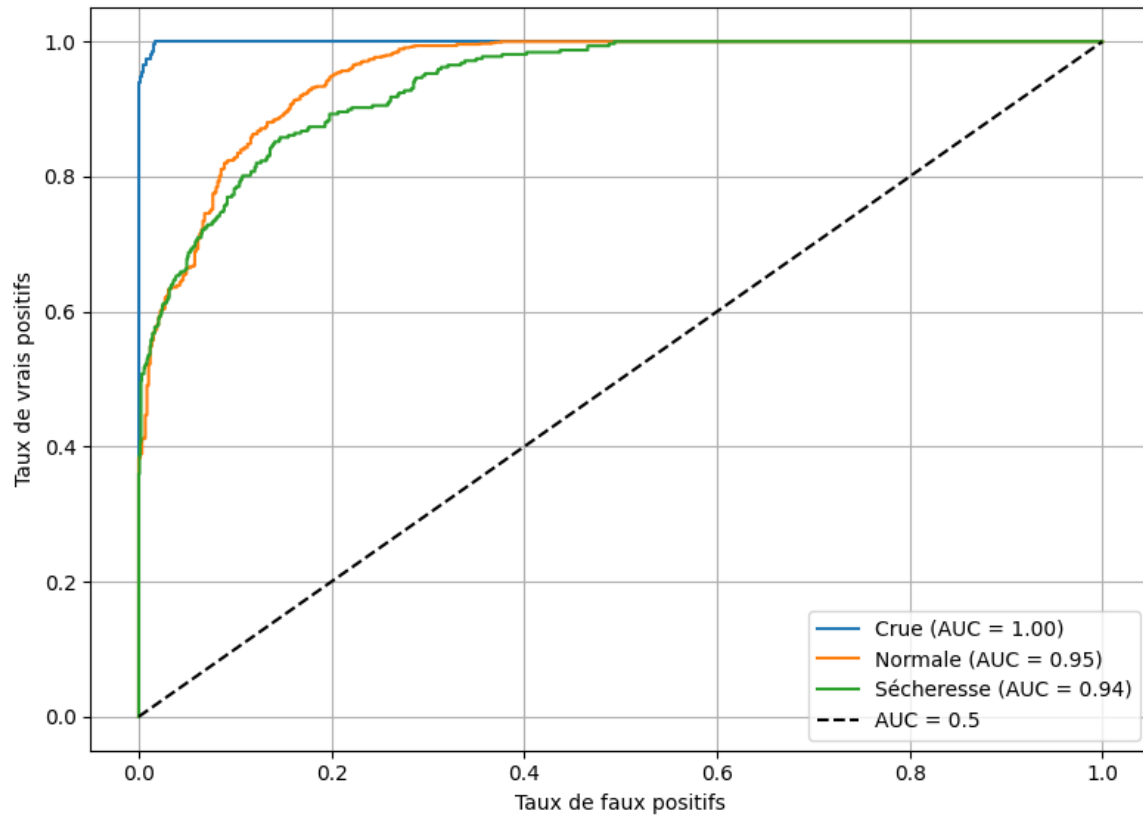


Figure 13 : courbe ROC avec indicateur AUC pour 'XGBoost'

L'indicateur associé à ces courbes est l'aire sous la courbe (AUC - Area Under Curve), qui résume la performance globale du modèle :

- Une AUC de 1.0 indique une séparation parfaite entre les classes,
- Une AUC de 0.5 équivaut à un modèle sans pouvoir de discrimination (tirage au sort).

<https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=fr>

Analyse et interprétation :

- Tous les modèles atteignent une AUC parfaite (1.00) pour la classe “Crue”, ce qui signifie qu’ils sont capables de discriminer relativement sans erreur les épisodes de crue des autres états hydrologiques. Cette performance s’explique par la signature forte des crues dans les données (pics de précipitations, afflux soudain, variations nettes de volume).
- Pour la classe “Normale”, les AUC varient entre 0.94 et 0.96, avec un meilleur score pour LightGBM (0.96). Cela montre que ce modèle parvient à bien distinguer les périodes calmes, même dans un contexte où la classe normale est fortement dominante.
- La classe “Sécheresse” obtient des AUC comprises entre 0.92 et 0.94, avec LightGBM et XGBoost en tête. Bien que les scores soient très bons, ils révèlent que la détection de la sécheresse reste légèrement plus difficile que celle des crues, ce qui est cohérent avec sa nature plus diffuse et progressive.

L’analyse des courbes ROC permet aussi de visualiser les compromis entre sensibilité et spécificité, ce qui est crucial pour ajuster un modèle dans un contexte de gestion du risque, où les faux positifs (fausses alertes) et faux négatifs (événements non détectés) peuvent avoir des implications opérationnelles différentes.

Les valeurs d’AUC confirment les tendances déjà observées dans les autres métriques :

- Les crues sont détectées de façon quasi parfaite par tous les modèles.
- LightGBM montre une capacité de séparation supérieure sur les trois classes, ce qui justifie sa place parmi les meilleurs candidats pour un système de prévision opérationnel.
- XGBoost se distingue également par son équilibre global entre précision, rappel et capacité de discrimination.

3. Analyse temporelle des épisodes extrêmes et des tendances :

3.1. Cartes temporelles des épisodes extrêmes :

3.1.1. Cartes temporelles des épisodes extrêmes pré définies :

Les cartes temporelles ci-dessous représentent l'évolution mensuelle des épisodes classés en crue, normale, et sécheresse sur la période d'étude (2003–2025). Ces cartes ont été générées à partir des variables cibles pré-définies, construites selon des critères hydrologiques . Elles ne dépendent pas des prédictions des modèles de machine learning, mais servent de référence pour l'analyse temporelle et l'évaluation visuelle de la dynamique des extrêmes.

a) Épisodes de crue :

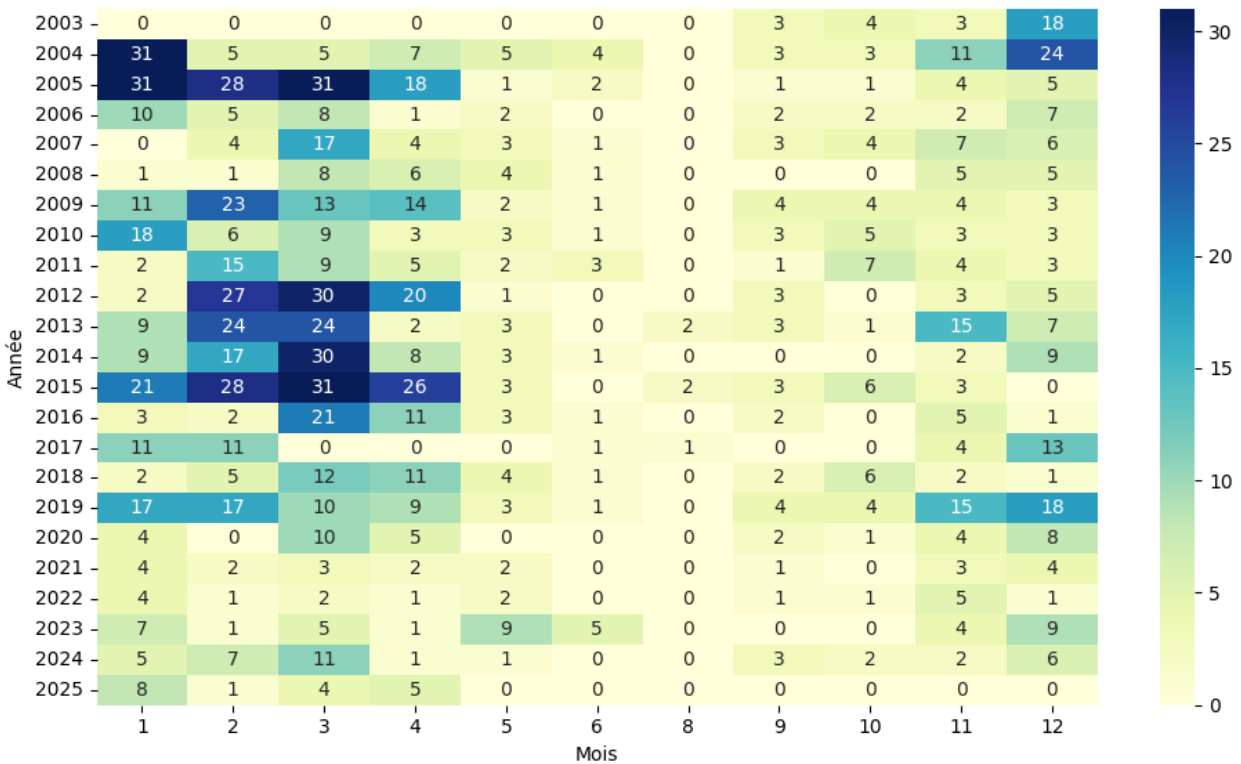


Figure 14 : carte temporelle des évènements 'crue'

La carte des crues montre une concentration marquée des épisodes intenses durant les mois de janvier à avril, ce qui correspond à la période de crue hivernale typique dans les régions méditerranéennes. On observe :

- Des pics d'occurrence en 2005, 2013, et 2015, avec des crues enregistrées presque tous les mois du premier semestre.
- Une réapparition plus fréquente à partir de 2019, ce qui pourrait suggérer une tendance à la reprise d'événements extrêmes.

- Des années comme 2003, 2006 ou 2020 avec très peu d'épisodes de crue détectés.

Ces résultats confirment que les crues, bien que soudaines, suivent une saisonnalité nette et peuvent être associées à des précipitations journalières extrêmes et des variations rapides de débits et de volumes stockés.

b) Épisodes normaux :

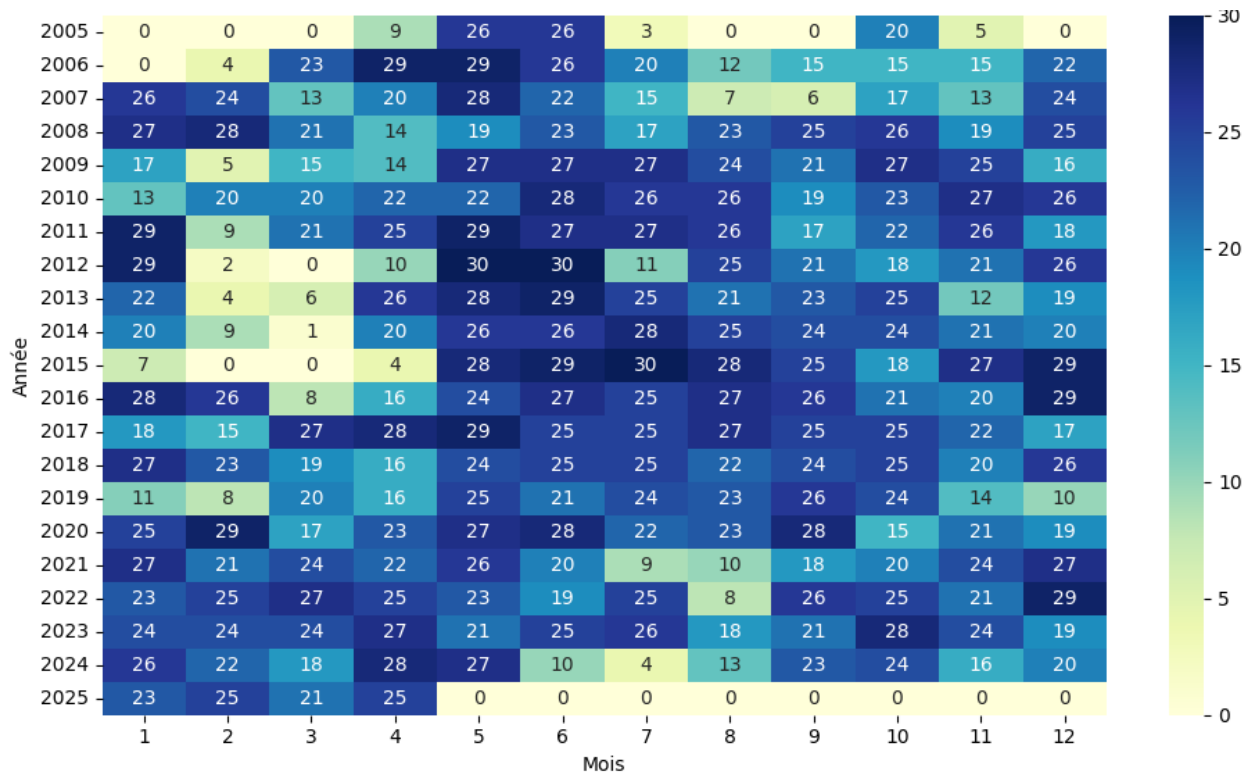


Figure 15 : carte temporelle des événements 'normale'

La classe normale, qui regroupe les mois sans événements extrêmes, montre une forte domination dans la plupart des années, notamment entre 2010 et 2020, où plus de 7 à 9 mois par an restent dans un état d'équilibre hydrologique. Cependant :

- On note une réduction de la stabilité normale sur certaines années extrêmes (ex. : 2005 ou 2023), où les épisodes de crue ou de sécheresse sont majoritaires.
- Ces séquences "normales" sont cruciales pour calibrer les modèles et détecter les basculements vers des régimes extrêmes.

c) Épisodes de sécheresse :

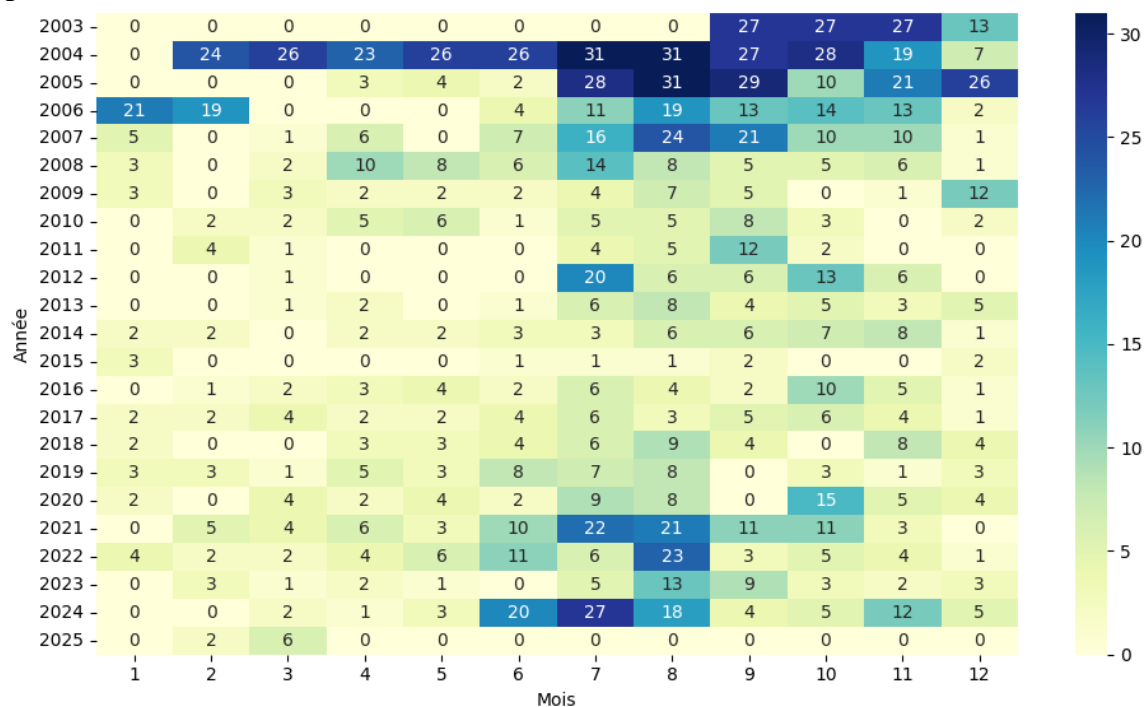


Figure 16 : carte temporelle des événements 'sécheresse'

Les sécheresses présentent une distribution plus diffuse, souvent répartie entre les mois d'été et d'automne :

- Les années 2004 à 2007 montrent des sécheresses quasi permanentes, notamment entre mai et novembre, avec des valeurs très élevées dans la carte.
- Une intensification visible à partir de 2020, avec des mois secs nombreux en 2022 et 2023.

Contrairement aux crues, les sécheresses ne montrent pas toujours une saisonnalité fixe, mais elles s'étendent souvent sur plusieurs mois, traduisant leur nature cumulative. Les performances élevées des modèles sur les classes crue et normale peuvent également s'expliquer par la clarté des signaux dans ces cartes, tandis que la classe sécheresse, plus progressive, reste plus difficile à modéliser avec précision.

Ces cartes offrent une visualisation synthétique de la variabilité hydrologique interannuelle et. Elles permettent :

- De repérer visuellement les années critiques, à fort risque de crue ou de sécheresse ;
- De mieux comprendre la dynamique saisonnière des événements extrêmes ;
- De justifier la nécessité d'une modélisation prédictive, en raison de la fréquence et de la complexité croissante de ces épisodes.

Bien que ces cartes basées sur des classes pré-définies constituent une référence précieuse pour analyser la dynamique saisonnière et interannuelle des événements hydrologiques, elles ne reflètent pas directement la capacité des modèles à reproduire ou anticiper ces épisodes.

C'est pourquoi, dans la section suivante, nous présentons les cartes temporelles construites à partir des prédictions des algorithmes de Machine Learning, afin d'évaluer dans quelle mesure ces derniers parviennent à restituer les rythmes hydrologiques observés. Cette comparaison permet de visualiser concrètement la portée opérationnelle des modèles dans un contexte de gestion prédictive des ressources en eau.

3.1.2. Cartes temporelles des épisodes extrêmes prédits :

Cette section présente les cartes générées à partir des prédictions des modèles de Machine Learning. Contrairement aux cartes de référence, celles-ci traduisent la manière dont chaque algorithme a interprété les états hydrologiques à partir des variables explicatives, sur l'ensemble de la série temporelle.

Les cartes seront présentées en termes des classes des événements pour mettre en évidence les différences de performance entre les modèles

1. Episodes prédites de crue :

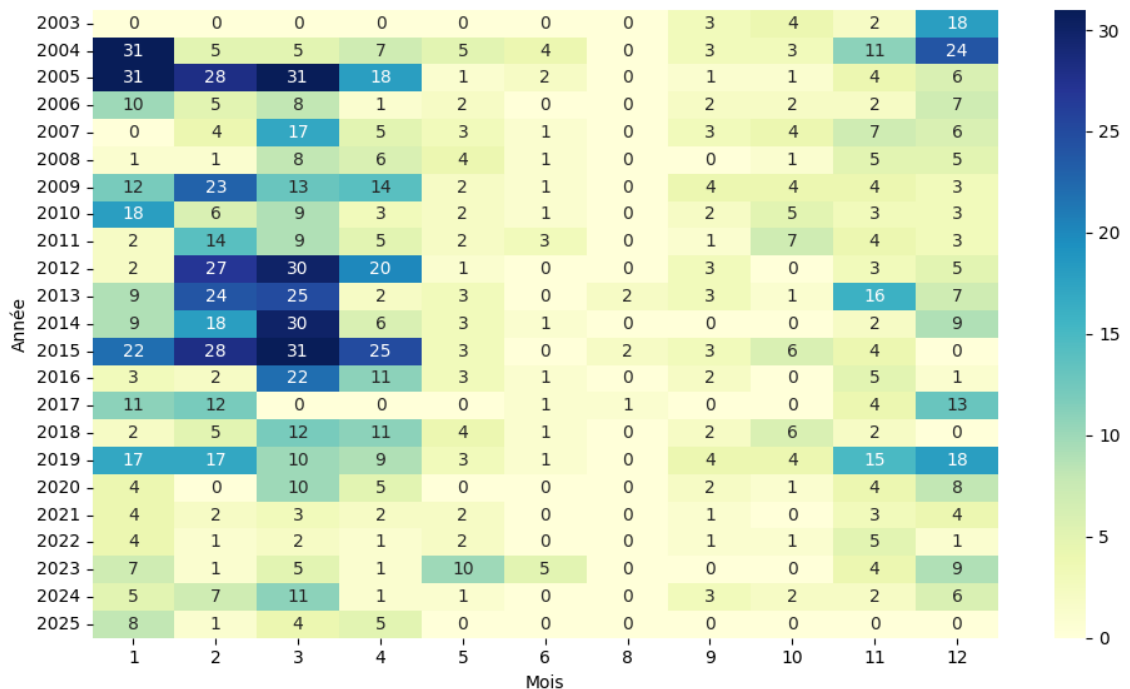


Figure 17 : carte temporelle des événements 'crue' prédits par RF

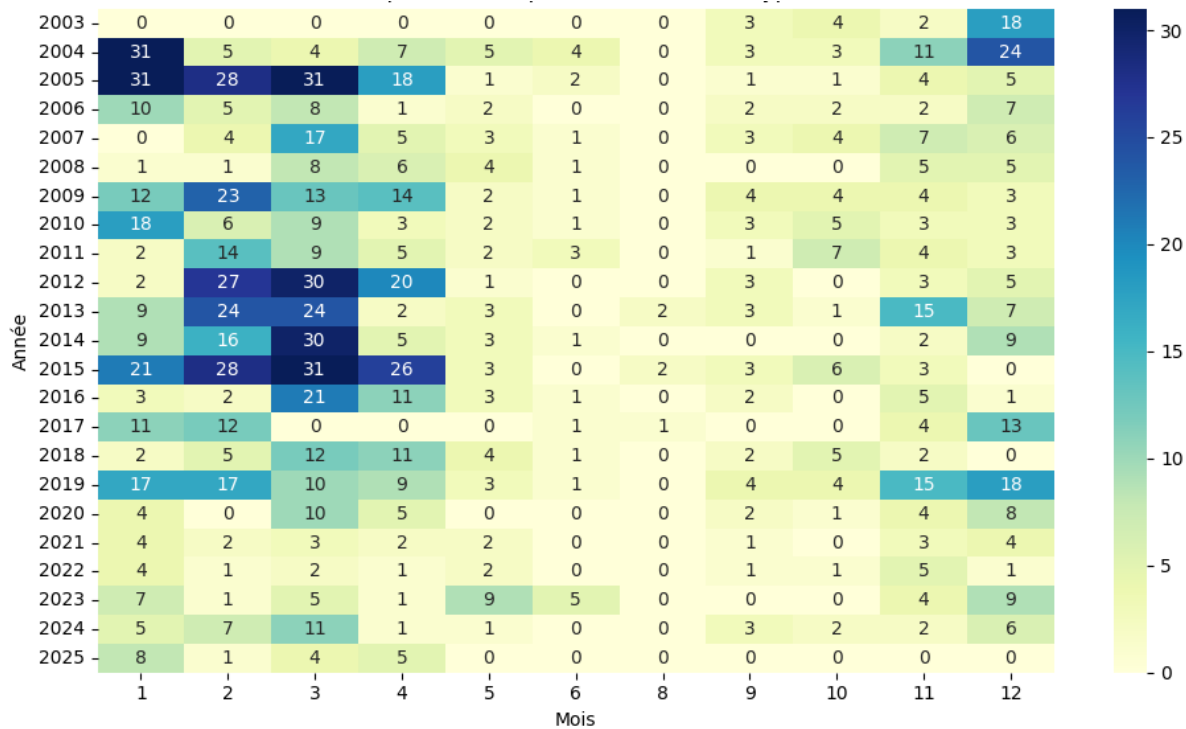


Figure 20 : carte temporelle des évènements 'crue' prédits par LightGBM

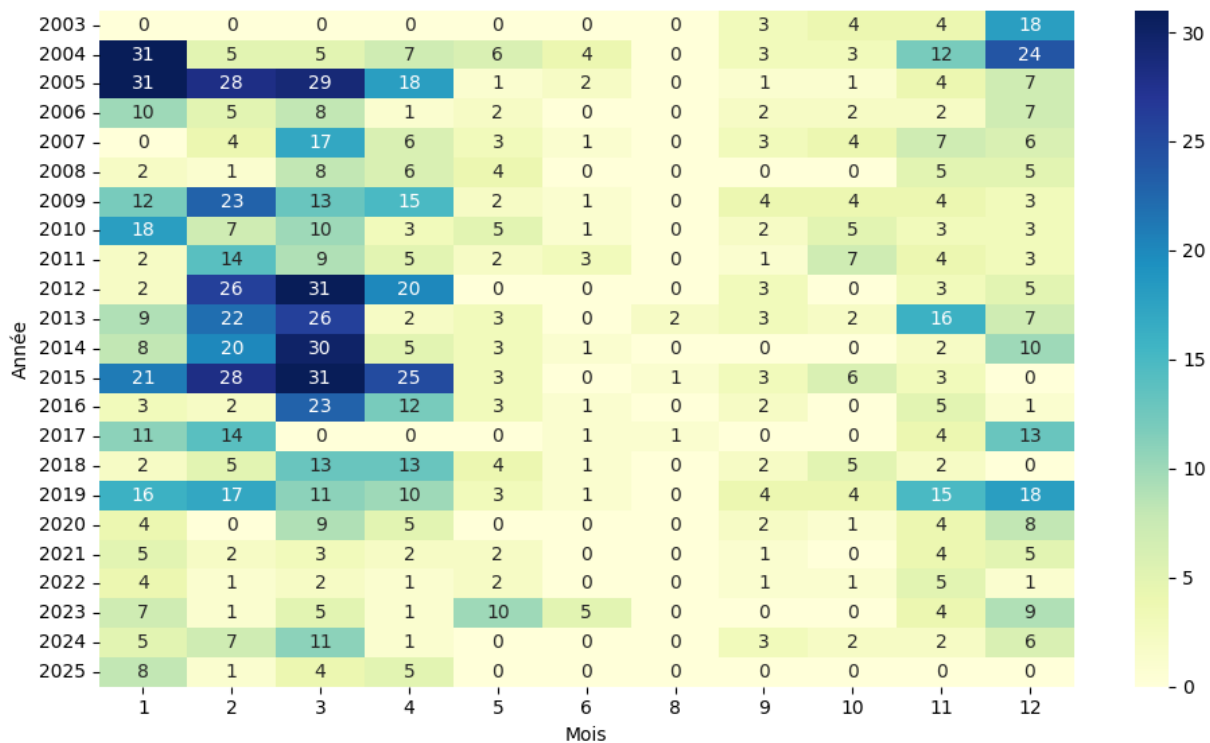


Figure 21 : carte temporelle des évènements 'crue' prédits par CatBoost

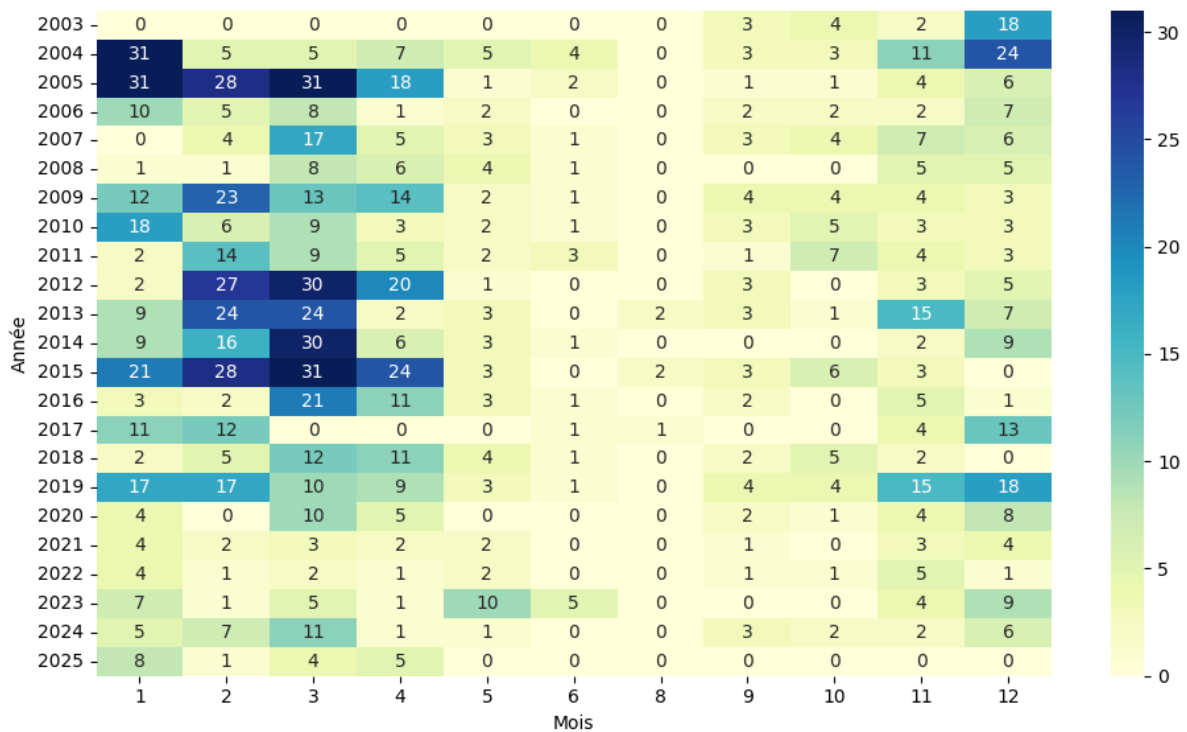


Figure 22 : carte temporelle des évènements 'crue' prédits par XGBoost

Les cartes temporelles issues des classes prédites par les différents algorithmes montrent une très forte similarité entre modèles. Quelle que soit la méthode utilisée les épisodes de crue identifiés correspondent presque exactement, aussi bien en termes de répartition mensuelle que d'occurrence annuelle.

De plus, les résultats obtenus à partir des prédictions sont largement conformes à ceux observés dans la carte des événements pré-définies, avec uniquement quelques écarts mineurs sur certains mois isolés. Ces divergences, ponctuelles et peu fréquentes, n'affectent pas l'interprétation générale.

Cette convergence entre les modèles confirme :

- la robustesse du signal hydrologique associé aux crues (fortes précipitations, hausse soudaine de volume) ;
- et la capacité des algorithmes à capturer fidèlement ces événements extrêmes, ce qui confirme les excellents scores de performance obtenus précédemment (F1-score ≥ 0.95 , AUC = 1.00).

Ces cartes prédictives confirment ainsi le potentiel opérationnel des modèles de Machine Learning dans la détection précoce des crues, en restituant fidèlement leur saisonnalité hivernale et leur variabilité interannuelle.

2. Episodes prédites normaux :

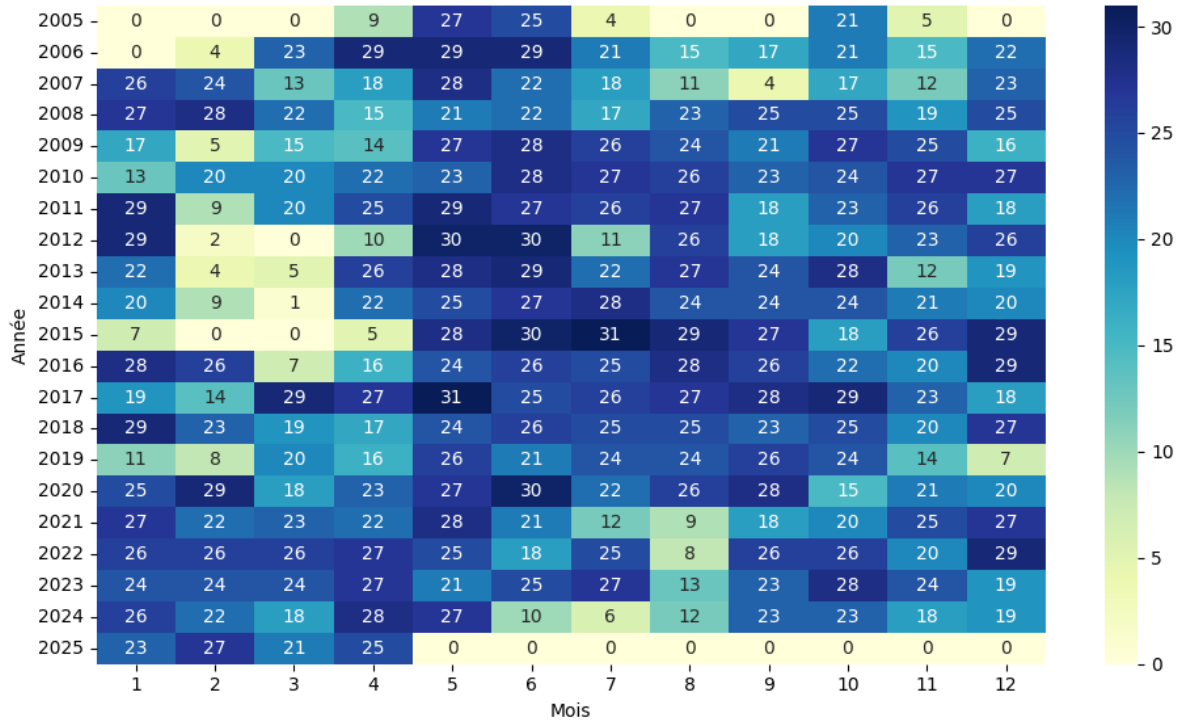


Figure 23 : carte temporelle des évènements 'normale' prédits par RF

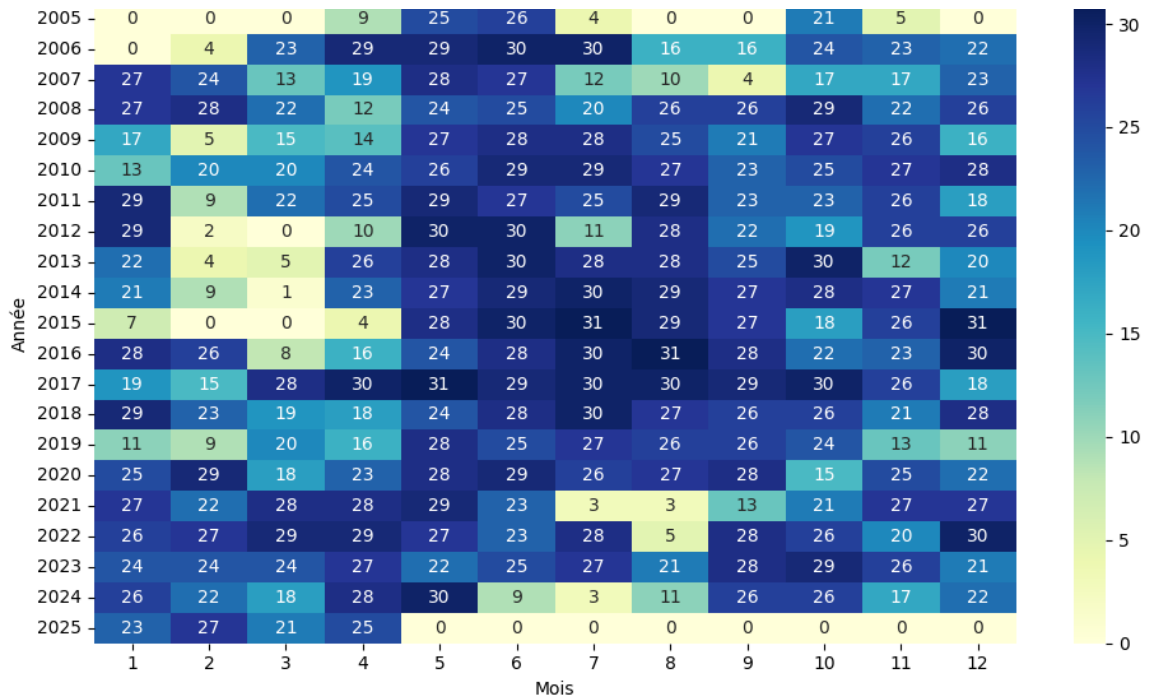


Figure 24 : carte temporelle des évènements 'normale' prédits par LightGBM

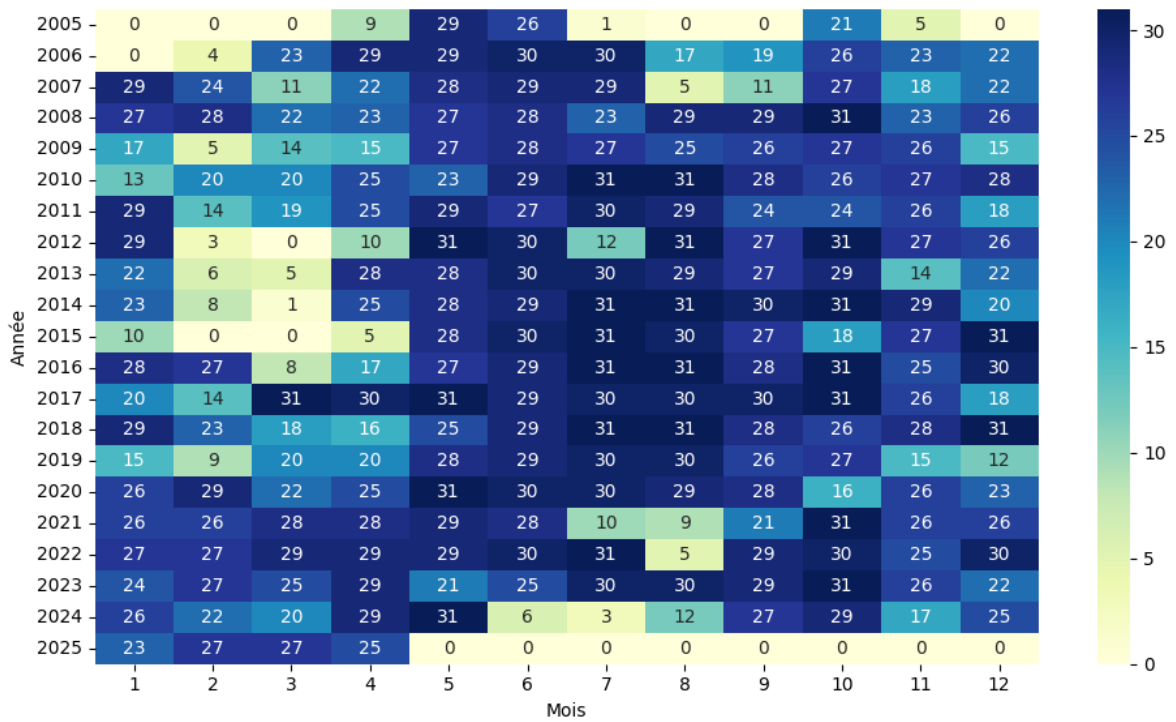


Figure 25 : carte temporelle des évènements 'normale' prédits par CatBoost

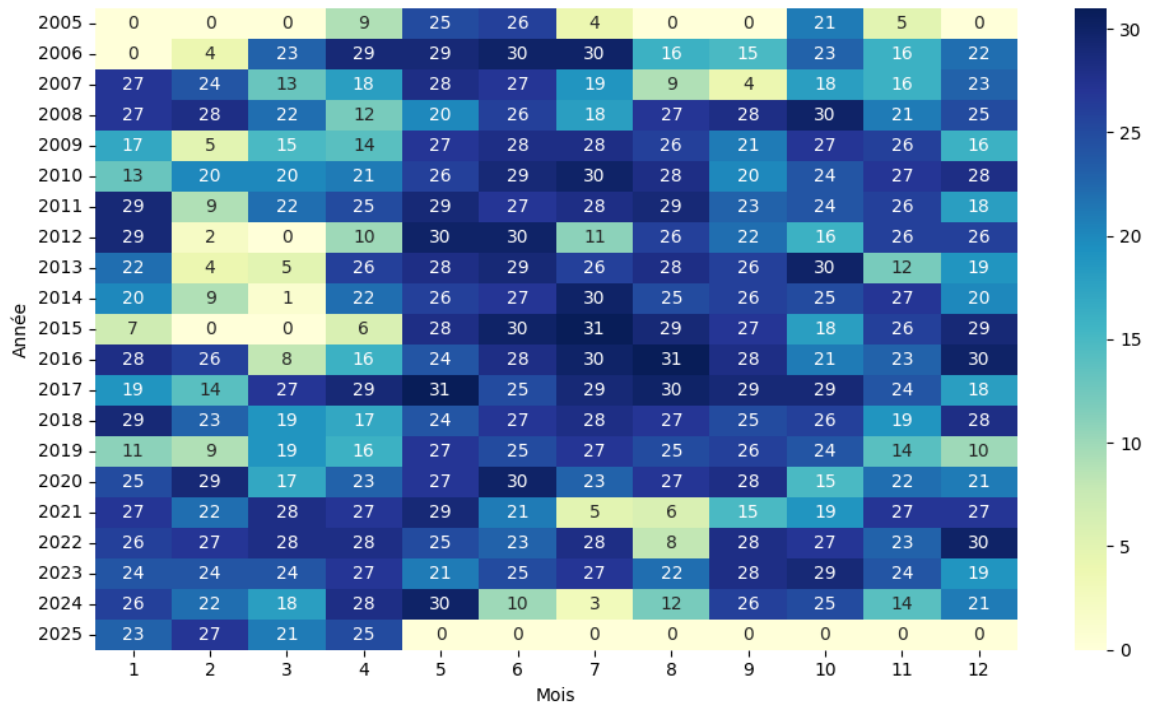


Figure 26 : carte temporelle des évènements 'normale' prédits par XGBoost

L'analyse des cartes temporelles issues des prédictions montre des différences plus marquées entre les algorithmes pour la classe "Normale", contrairement à la forte homogénéité observée dans la classe "Crue".

- Random Forest affiche une carte très proche de la carte de référence, avec un bon équilibre entre sous et sur-prédictions. Cette performance s'aligne parfaitement avec ses résultats statistiques : un F1-score de 0.91 et une précision de 0.89, indiquant une capacité robuste à identifier les périodes normales sans trop de confusion avec les extrêmes.
- LightGBM et XGBoost produisent des résultats assez similaires, marqués par une tendance à la sur-prédiction des mois "normaux", en particulier durant les mois intermédiaires (mois 6 à 11). Cela est confirmé par leurs rapports de classification qui montrent des rappels très élevés (0.96 pour LightGBM, 0.95 pour XGBoost), mais une précision légèrement inférieure (0.88 et 0.89 respectivement), traduisant cette tendance à inclure à tort des épisodes extrêmes dans la classe normale.
- CatBoost, en revanche, sur-prédit fortement la classe "Normale", notamment entre les mois d'avril et novembre. Bien qu'il affiche le rappel le plus élevé de tous les modèles (0.98), cela s'accompagne d'une précision relativement faible (0.87), confirmant que de nombreux épisodes extrêmes (en particulier des sécheresses modérées) sont mal classés comme normaux. Ce biais se traduit visuellement par une surestimation généralisée de l'état stable, ce qui affecte la fiabilité du modèle dans la détection des extrêmes.

Ces constats montrent que, malgré une classe bien représentée dans les données (support = 994), le compromis entre rappel et précision varie fortement selon l'algorithme, influençant directement la fidélité des cartes temporelles générées à partir des prédictions. Un haut rappel n'est pas toujours gage de fiabilité si la précision chute, comme le montre le cas de CatBoost. À l'inverse, Random Forest atteint un meilleur équilibre entre les deux métriques, ce qui se traduit par une restitution plus fidèle de la variabilité hydrologique observée.

3. Episodes prédites de sécheresse :

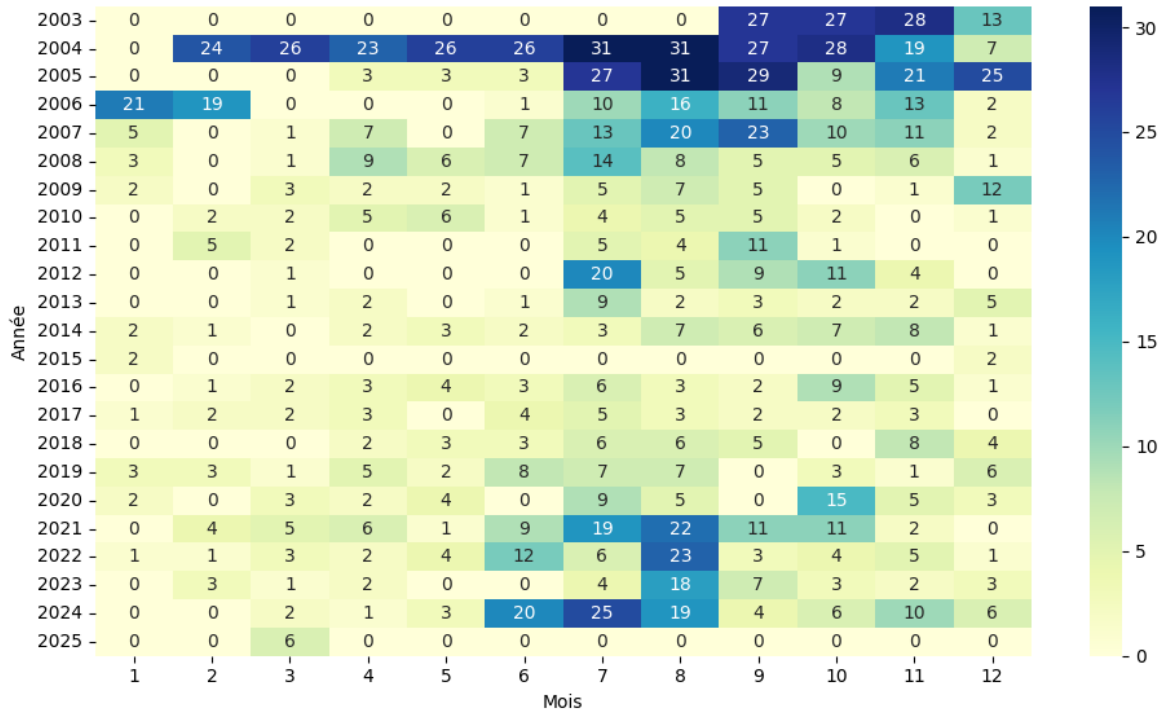


Figure 27 : carte temporelle des évènements 'sécheresse' prédits par RF

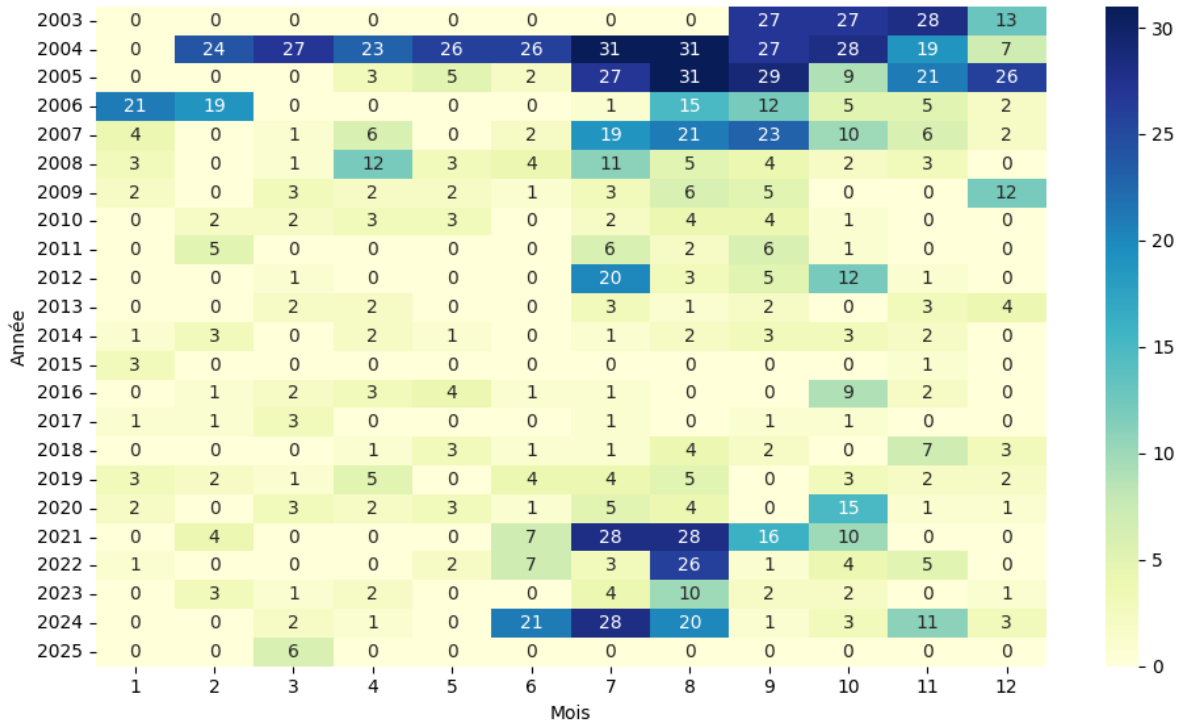


Figure 28 : carte temporelle des évènements 'sécheresse' prédits par lightGBM

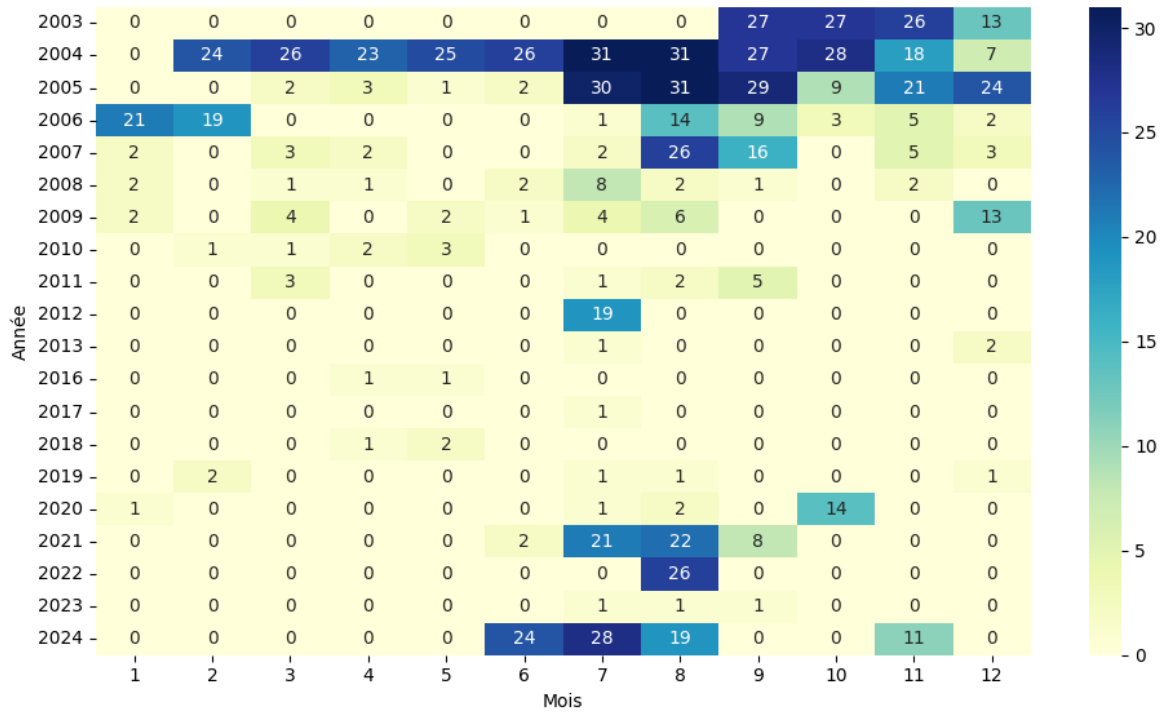


Figure 29 : carte temporelle des évènements 'sécheresse' prédits par CatBoost

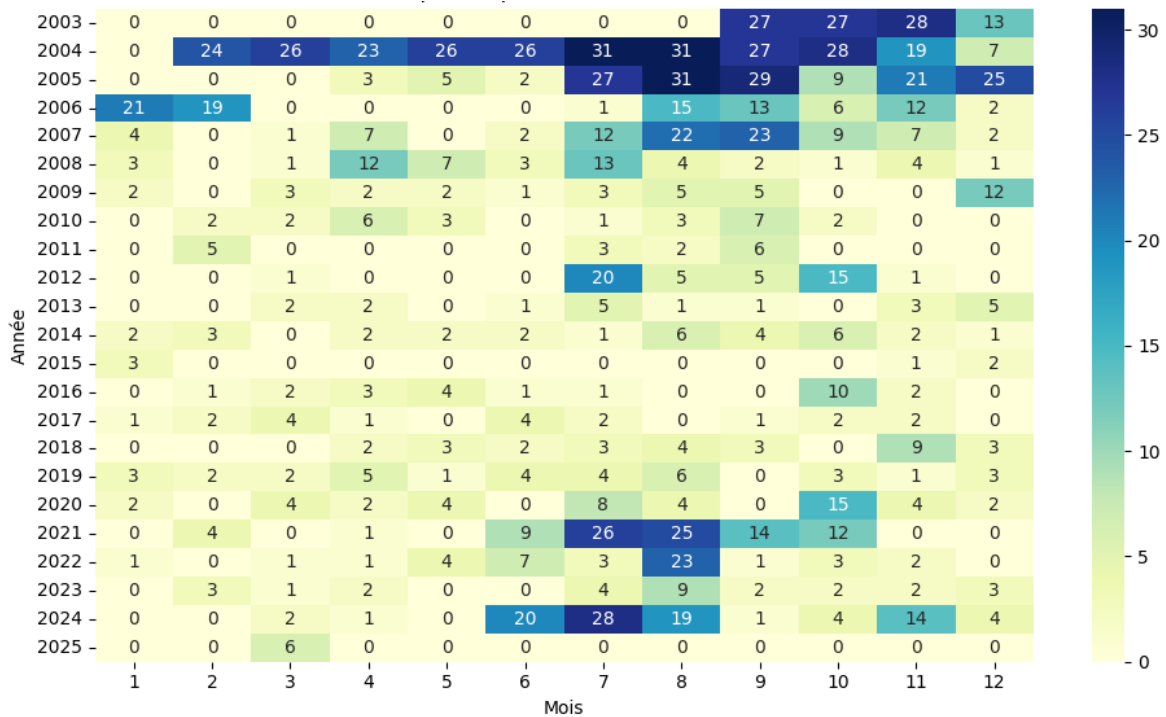


Figure 30 : carte temporelle des évènements 'sécheresse' prédits par XGBoost

La prédiction des épisodes de sécheresse reste le défi majeur pour tous les modèles. À la différence des crues, les sécheresses présentent une évolution lente, cumulative et souvent moins marquée dans les données, ce qui accroît la probabilité de confusion avec la classe “normale”, surtout en dehors des périodes d’extrême.

- Random Forest, bien qu’ayant généré une carte temporelle globalement proche de la référence de ou en peut dire subjectivement que c’est le meilleur modèle pour redicter les sécheresses, présente en réalité des performances numériques limitées sur cette classe : son rappel n’est que de 0.64 et son F1-score de 0.70, ce qui indique qu’un nombre de épisodes secs a été classé comme “normaux”. Sa précision (0.77) modérée montre qu’il fait également des erreurs dans l’autre sens. Cette combinaison suggère une tendance à sous-détecter les sécheresses modérées tout en maintenant un certain équilibre général. Il n’est donc pas le modèle le plus performant sur cette classe, malgré une carte visuellement acceptable.
- XGBoost et LightGBM montrent tous deux une légère sous-prédiction des sécheresses, perceptible sur les cartes entre les mois de juin et novembre. Cette tendance est bien reflétée dans leurs rappels respectifs (0.63), accompagnés d’un F1-score à peine meilleur que celui de RF (0.72 pour XGBoost, 0.71 pour LightGBM). Leur précision est plus élevée (0.82 et 0.83) par rapport à RF, et c’est ce qui indique que lorsqu’un épisode de sécheresse est identifié, elle est généralement bien classée. Toutefois, leur capacité à capturer l’ensemble des épisodes réels reste incomplète, d’où une sous-estimation de l’intensité ou de la fréquence des sécheresses sur certaines années.
- CatBoost, enfin, se distingue à nouveau comme le modèle le moins performant sur cette classe. Sa précision exceptionnelle (0.93) contraste fortement avec son rappel très faible (0.55), signe d’un modèle extrêmement conservateur. Il ne détecte que les cas de sécheresse les plus prononcés et ignore une large part des événements modérés, ce qui explique les nombreuses lacunes dans sa carte temporelle, notamment durant la seconde moitié de l’année hydrologique.

Ces résultats traduisent bien la difficulté commune à tous les modèles à appréhender les sécheresses de manière fine. Le compromis entre rappel et précision est particulièrement délicat pour cette classe. Aucun des algorithmes ne parvient à concilier une excellente couverture des événements secs (rappel) avec une précision suffisante, mais le XGBoost et LightGBM reste préférables dans ce cas, et c’est ce qui invite à renforcer et compléter la modélisation par des études approfondies au but de mieux comprendre le comportement des extrêmes spécialement la sécheresse

3.2. Analyse de tendance avec le test de Mann-Kendall :

Pour chaque classe , une série temporelle annuelle a été construite en calculant le nombre de mois appartenant à chaque catégorie par année. Ces séries ont ensuite été soumises au test non paramétrique de Mann-Kendall, reconnu pour sa robustesse dans la détection de tendances monotones dans des séries hydrologiques, sans exigence de normalité ni de linéarité.

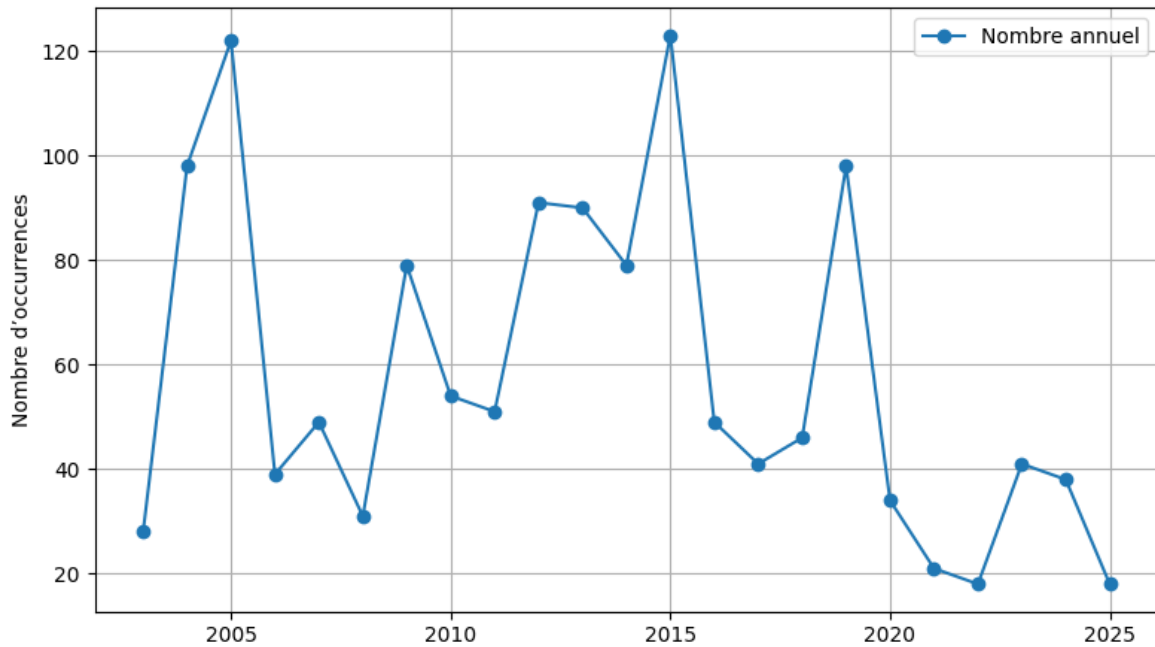


Figure 31 : tendance annuelle -crue

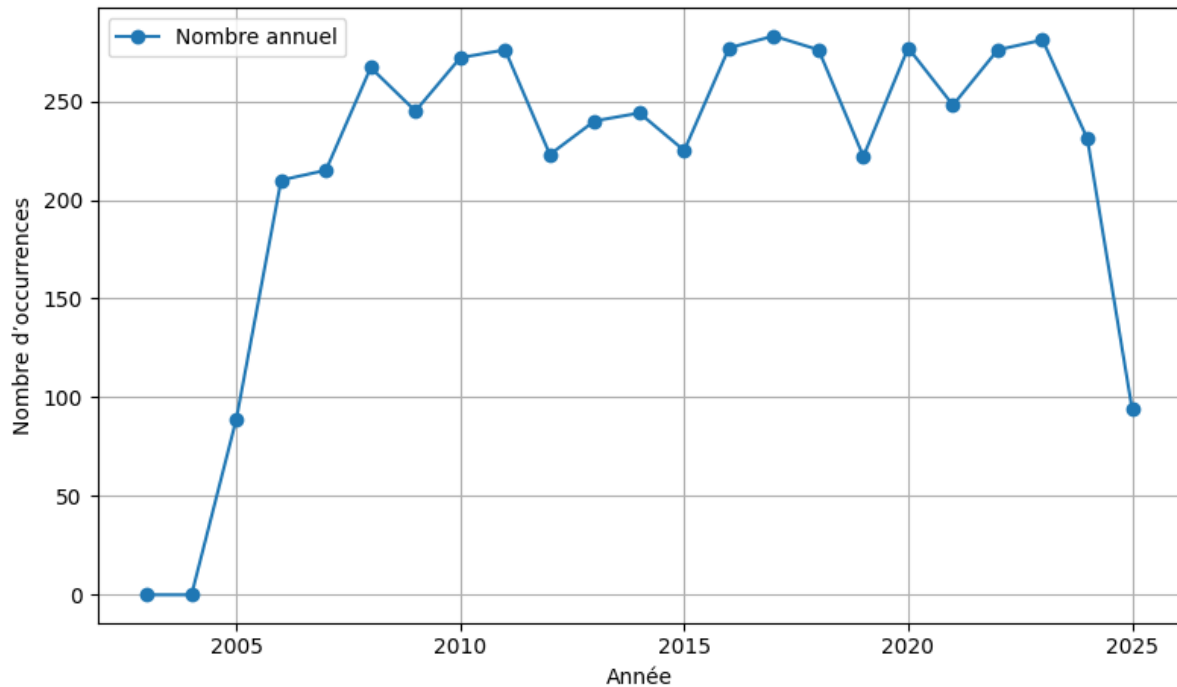


Figure 32 : tendance annuelle -normale

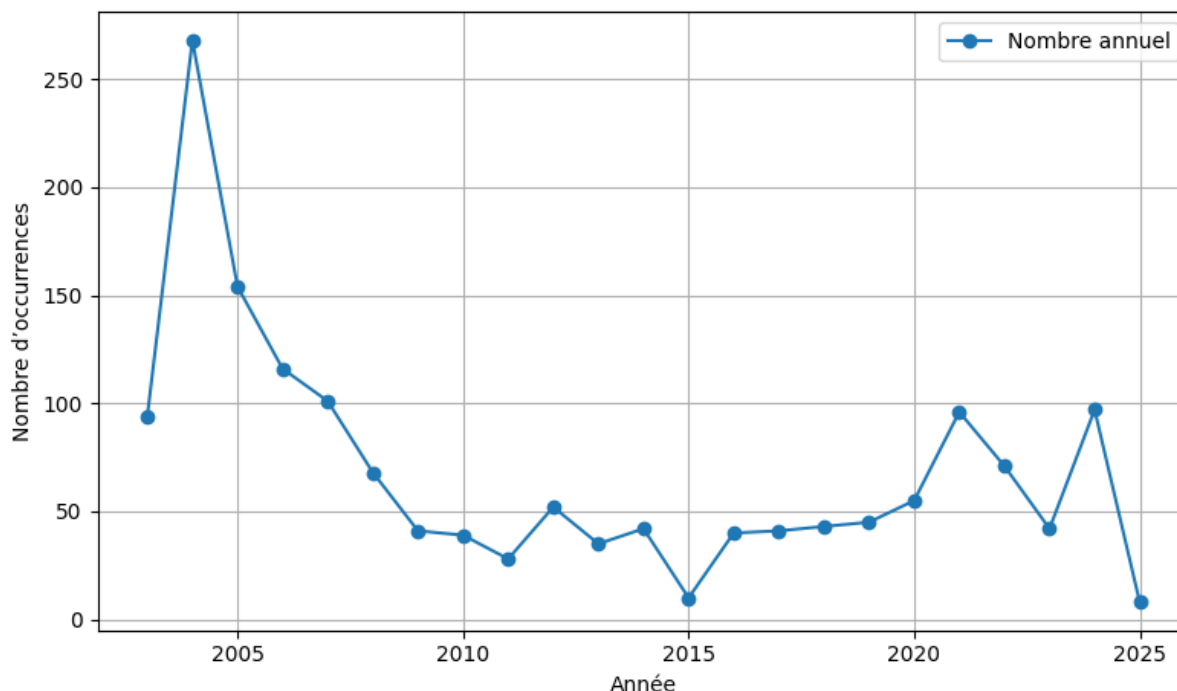


Figure 33 : tendance annule - sécheresse

Résultats du test de Mann-Kendall :

les résultats statistiques de test sont présentés dans le tableau suivant :

Tableau 17 : les résultats de test (Mann-Kendall)

Classe	Tendance détectée	p-value	Significativité (H_0 rejetée ?)
Sécheresse	Aucune tendance	0.1864	Non
Normale	Croissante	0.0119	Oui
Crue	Aucune tendance	0.0534	Non

Interprétation :

La classe “Normale” présente une tendance significative croissante ($p = 0.0119$), indiquant que le nombre de mois classés comme “normaux” augmente au fil du temps. Ce résultat est encourageant dans le contexte d'une gestion hydrique durable, car il pourrait refléter une stabilisation relative du régime hydrologique, ou du moins une atténuation temporaire des extrêmes.

La classe “Sécheresse” ne montre aucune tendance significative ($p = 0.1864$), ce qui signifie que, sur la période étudiée, aucune évolution claire dans la fréquence annuelle des événements secs ne peut être confirmée statistiquement. Cela n'exclut pas une variabilité importante d'une année à l'autre, mais suggère l'absence de dynamique monotone.

Pour la classe “Crue”, la p-value (0.0534) est très proche du seuil de significativité de 0.05, sans toutefois le franchir. Cela suggère qu’il existe un signal faible de variation, mais pas suffisamment fort pour affirmer une tendance significative. Cela est cohérent avec la nature irrégulière et localisée des crues, fortement influencées par des événements pluviométriques extrêmes ponctuels.

4. Conclusion :

Les analyses menées ont permis de démontrer la capacité des modèles de ML à classer efficacement les états hydrologiques, avec des performances particulièrement élevées pour la détection des crues et des situations normales. Les résultats montrent également que la classe “sécheresse” reste plus difficile à prédire, en raison de sa nature progressive et de sa variabilité spatiale et temporelle.

Les algorithmes de type boosting (LightGBM, XGBoost) ont montré un léger avantage sur les autres, en termes de précision globale et de capacité de généralisation. L’analyse de l’importance des variables confirme la pertinence des paramètres utilisés, notamment la pluie journalière maximale, les défluent et les variations de volume.

Les cartes temporelles et le test de Mann-Kendall ont mis en évidence des dynamiques interannuelles importantes. La seule tendance statistiquement significative détectée concerne une augmentation du nombre de mois "normaux", tandis que les sécheresses et crues ne présentent pas de tendance claire sur la période étudiée.

Ces résultats appuient la validité de l’approche de ML pour une application opérationnelle dans la gestion préventive des barrages, tout en soulignant l’importance de la mise à jour continue des données et des modèles pour suivre l’évolution des régimes hydrologiques .

Conclusion générale :

Ce mémoire a permis de démontrer la pertinence d'une approche combinée entre traitement des données hydrologiques, calcul d'indices climatiques et modélisation prédictive par Machine Learning, appliquée à la détection et à l'analyse des événements extrêmes influençant la gestion des barrages en Algérie.

L'exploitation des données du barrage de Beni Haroun a mis en évidence l'impact considérable des sécheresses et des crues sur les volumes stockés, les flux entrants et sortants, et les dynamiques de fonctionnement. La classification multicatégorie (sécheresse, normale, crue) a permis de structurer l'analyse en trois états hydrologiques distincts.

Les résultats des modèles (Random Forest, XGBoost, LightGBM, CatBoost) montrent une performance globalement satisfaisante pour les classes "crue" et "normale", mais des limites persistantes dans la détection des sécheresses modérées. Ces difficultés sont confirmées par l'analyse des cartes temporelles et des scores de performance, notamment sur les métriques de rappel et de précision.

Ce travail ouvre ainsi des perspectives de recherche et d'amélioration : intégration de données issues de la télédétection, élargissement du périmètre spatial à d'autres bassins, recours à des approches hybrides combinant modèles physiques et algorithmiques. À l'échelle opérationnelle, il souligne le potentiel des outils de Machine Learning pour soutenir la planification hydrique et renforcer la résilience des infrastructures face aux aléas climatiques.

Reference bibliographiques :

- [1]. Le Delliou, P. (2003). Les barrages: conception et maintenance. Presses Universitaires Lyon.
- [2]. <https://www.anbt.dz/liste-des-barrages/>
- [3]. Plan national sécheresse -Algérie lignes directrices en vue de son opérationnalisation Mohamed Safar-Zitoun (juin 2019) (Ministère de la Santé Publique)
- [4]. Organisation Météorologique Mondiale. Suivi de la sécheresse et alerte précoce: principes, progrès et enjeux futurs, OMM-N°1006, Genève Suisse,
- [5]. Merabti, A. et al. (2021). Spatiotemporal Characteristics and Trends of Meteorological Droughts in the Wadi Mina Basin, Northwest Algeria
- [6]. IMF Staff (2024). Country Report on Algeria, souligne que la disponibilité en eau par habitant (~250 m³/an) est moins de la moitié du seuil de rareté (500 m³)
- [7]. Projet de mise en place d'un système d'alerte précoce a la sécheresse dans trois pays de la rive sud de la méditerranée: Algérie, Maroc, et Tunisie « Calcul des indicateurs de sécheresse à partir des images NOAA/AVHRR » Mohammed LAYELMAM (septembre 2008)(Layelmam, 2008)
- [8]. McKee, T. B. et al. (1993). The relationship of drought frequency and duration to time scales. Conf. on Applied Climatology
- [9]. Palmer, W. C. (1965). Meteorological Drought. U.S. Weather Bureau Research Paper No. 45.
- [10]. Alley, W. M. (1984). The Palmer Drought Severity Index: Limitations and Assumptions. Journal of Climate and Applied Meteorology.
- [11]. Gibbs, W. J. & Maher, J. V. (1967). Rainfall deciles as drought indicators. Bureau of Meteorology Bulletin No. 48
- [12]. Remini B, Larhyss Journal, (2023), FLASH FLOODS IN ALGERIA
- [13]. Agueninie S, (2025),ANBT
- [14]. Gaume, E., et al. (2016). The Bab-El-Oued flash flood (Algeria) of November 2001: Analysis and lessons. Natural Hazards and Earth System Sciences
- [15]. Ziadi, S. R., & Keraghel, M. A. (2024). Flooding vulnerability in Algiers, Algeria: An analytic hierarchy process. Natural Hazards.
- [16]. Beven, K. (2012). Rainfall-Runoff Modelling

- [17]. Bouaziz, A., & Kallel, A. (2022). Reservoir storage changes and water supply deficit in Algeria under recent droughts. *Journal of Water Supply: Research and Technology*
- [18]. Benhassine, A., & Tamim, Y. (2011). Impact of flash floods on Ghardaïa (Algeria) – Case study of 1 October 2008
- [19]. Khélifa, F., & Bentchikou, Z. (2025). Extreme rainfall and flood events in Béchar region, Algeria, September 2024. *Journal of Arid Environments*
- [20]. Al-Madhagi, S., et al. (2024). Structural failure analysis of the Derna dams under Storm Daniel (2023). *International Journal of Disaster Risk Reduction*
- [21]. De Brito, M. M. et al. (2016). Multi-criteria decision analysis for infrastructure risk management: dams case study. *Environmental Earth Sciences*
- [22]. Beven, K. J. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system. *Hydrological Processes*,
- [23]. Tyrallis, H et al . (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources.
- [24]. Niazkar, M et al. (2024). Applications of XGBoost in water resources engineering: A systematic literature review. *Environmental Modelling & Software*.
- [25]. Bilolikar et al. (2023) soulignent l'importance de l'optimisation des hyperparamètres
- [26]. <https://www.anbt.dz>
- [27]. Mémoire, EVALUATION ET LA CARTOGRAPHIE DES RISQUES D'érosion par l'utilisation de sig et le modèle EPM cas de bassin versant du barrage de Ben Haroun, (université de M'sila).
- [28]. Article, Le Système « Beni-Haroun » (Oued Kebir-Rhumel, Algérie) : Aménagements Hydrauliques Et Contraintes Morpho-Géologiques, (AZZEDINE MEBARKI, CHAOUKI BENABBAS, FLORINA GRECU).
- [29]. Mémoire, Perte de capacité de stockage en eau dans quelques barrages de l'est Algérien (université 8 Mai 1945 de Guelma) Joins 2017
- [30]. A. Bechroune, Elaboration d'une base de données hydrologique du bassin versant de la Soummam via le modèle hydrologique HEC-HMS, Mémoire présenté pour l'obtention du diplôme de Magister, 2011.
- [31]. MOKHTARI, E. H. (2017). Impact de l'érosion hydrique sur l'envasement du barrage Ghrib. Thèse de doctorat en science, université Hassiba Ben Bouali Chlef.

- [32]. Barkat Kanza, 2016, Suivi de la qualité physico-chimique des eaux du Barrage Beni Haroun ,(Université de Des Frères Mentouri Constantine)
- [33]. fiche technique de barrage de Beni Haroun ,2017,(ANBT)
- [34]. Miles, J. (2005). Tolerance and variance inflation factor. Encyclopedia of statistics in behavioral science.
- [35]. <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=fr>
- [36]. <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=fr>