



**HAL**  
open science

# Compréhension et caractérisation de l'intermittence du réseau hydrographique en Afrique : développements méthodologiques et applications hydrologiques

Patindé Axel Belemtougri

► **To cite this version:**

Patindé Axel Belemtougri. Compréhension et caractérisation de l'intermittence du réseau hydrographique en Afrique : développements méthodologiques et applications hydrologiques. Hydrologie. Sorbonne Université; Institut international d'ingénierie de l'eau et de l'environnement, 2022. Français. NNT : 2022SORUS318 . tel-03900431

**HAL Id: tel-03900431**

**<https://theses.hal.science/tel-03900431v1>**

Submitted on 15 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE EN COTUTELLE**

Pour obtenir le grade de

**DOCTEUR DE L'INSTITUT INTERNATIONAL D'INGENIERIE DE L'EAU  
ET DE L'ENVIRONNEMENT (2iE)**

École Doctorale Sciences et Technologies de l'Eau, l'Énergie et l'Environnement  
Spécialité : **Eau**

**DOCTEUR DE SORBONNE UNIVERSITE (SU)**

École Doctorale GRNE (*Géosciences, Ressources Naturelles et Environnement*) – ED 398  
Discipline : **Géosciences - Environnement**

Présentée et soutenue publiquement par

**Patindé Axel BELEMTOUGRI**

Le 23 septembre 2022

Réf : 2iE/2022-08

---

**Compréhension et caractérisation de l'intermittence du  
réseau hydrographique en Afrique : développements  
méthodologiques et applications hydrologiques**

---

**JURY**

M. Expédit Wilfrid VISSIN, <i>Professeur Titulaire</i> , UAC, Bénin	Président
M. Eric SAUQUET, <i>Directeur de Recherche</i> , HDR, INRAE, France	Rapporteur
M. Amani Michel KOUASSI, <i>Maître de Conférences (CAMES)</i> , INP-HB, Côte d'Ivoire	Rapporteur
Mme Valérie PLAGNES, <i>Professeur des Universités</i> , SU, France	Examinatrice
M. Gil MAHÉ, <i>Directeur de Recherche</i> , HDR, IRD, France	Examinateur
M. Dial NIANG, <i>Maître de Conférences (CAMES)</i> , 2iE, Burkina Faso	Examinateur
Mme Agnès DUCHARNE, <i>Directrice de Recherche</i> , CNRS-SU, France	Directrice de Thèse
M. Harouna KARAMBIRI, <i>Professeur Titulaire</i> , 2iE, Burkina Faso	Directeur de Thèse

*Laboratoire Eaux HydroSystèmes et Agriculture (LEHSA)*

*Milieux Environnementaux, Transferts et Interactions dans les hydrosystèmes et les Sols (METIS)*

# DÉDICACE

*Je dédie affectueusement ce travail à :*

*Mes très chers parents, **Raymond et Suzanne BELEMTUGRI,***

*Mes sœurs **Ornella et Sandrine.***

## REMERCIEMENTS

Pour paraphraser un proverbe « Mossi » qui dit « *on arrive à se frotter facilement et proprement l'entièreté du dos si l'on est aidé* », je ne saurais donc terminer cette thèse sans remercier tous ceux qui ont apporté leur contribution à la réalisation de ce projet.

J'aimerais tout d'abord remercier Agnès Ducharne et Harouna Karambiri, mes co-directeurs de thèse, pour m'avoir guidé et fait confiance pour mener à bien ce projet. Merci de m'avoir laissé la latitude d'investiguer certaines idées même si elles n'ont pas toutes abouti. Malgré leurs emplois du temps très chargés, ils n'ont ménagé aucun effort pour m'accorder assez de temps et d'énergie et pour cela, je leur suis très reconnaissant. J'ai pu au cours de ces années m'enrichir de leur analyse critique et rigueur scientifique et je les remercie pour cela.

Je tiens à remercier les membres du jury de thèse pour leurs commentaires et suggestions visant l'amélioration du document. Il s'agit notamment de M. Eric Sauquet et M. Amani Michel Kouassi qui ont accepté de rapporter cette thèse, M. Dial Niang, Mme Valérie Plagnes, M. Gil Mahé pour avoir accepté d'être examinateurs et M. Expédit Wilfrid Vissin d'en être président. Je saisis l'occasion pour les remercier encore tous pour leurs retours pertinents.

Je remercie également les membres de mon comité de thèse, dont M. Vazken Andréassian et M. Aymar Bossa, pour leur participation et différents retours sur le travail à chaque étape de la thèse.

Un grand merci aux partenaires financiers de cette thèse, la Banque Africaine de Développement (BAD) à travers l'institut 2iE dans le cadre du projet « *Nelson Mandela Institutes-African Institutions of Science and Technology (NMI-AIST)* » qui a financé principalement cette thèse et l'aide spécifique aux doctorants boursiers de Sorbonne Université (ASDB) qui a financé une partie de mes séjours en France notamment pendant la période de confinement en 2020 et une partie de la dernière mobilité de thèse en 2021.

Cette thèse en cotutelle s'est déroulée au sein de deux Universités (2iE et Sorbonne Université) avec respectivement deux laboratoires d'affiliations notamment LEHSA (Laboratoire Eaux HydroSystèmes et Agriculture) et METIS (Milieux Environnementaux, Transferts et Interactions dans les hydrosystèmes et les Sols).

À l'institut 2iE, j'aimerais remercier M. Yacouba Konaté, responsable du laboratoire LEHSA et M. Igor Ouédraogo responsable de l'école doctorale pour m'avoir permis de travailler dans des conditions correctes et faciliter les procédures administratives pour mes différentes mobilités.

Je remercie ici M. Fowé Tazen et M. Roland Yonaba pour leur disponibilité, les nombreux échanges qu'on a eus sur mes travaux, pour leur soutien scientifique, technique et pour la relecture de certains de mes documents.

J'adresse un remerciement à mes amis doctorants du bureau « Y » en commençant par les « *benguistes* » Patrick, Radegonde, Coulibaly, Bruno (*Lawrence*), Philomina, Inès, Karim, Moussa Diagne avec qui j'ai partagé plus de la moitié de mon temps de thèse, je les remercie tous pour avoir contribué à entretenir cette ambiance de travail si plaisante. À ces remerciements, j'associe, Inayatou, Cheick, Yossa, Bouraïma.

À Sorbonne Université, j'aimerais à présent remercier tous les membres du laboratoire METIS qui m'ont soutenu et accompagné au cours de ces années. Un merci à Sylvain Théry,

Marie Sylvestre, Aurélien Baro pour m'avoir dépanné de nombreuses fois en SIG. Également merci au directeur du laboratoire Jean Marie Mouchel ainsi que Bénédicte Sabatier et Valérie Girard pour m'avoir facilité les procédures administratives notamment pendant mes mobilités et m'avoir permis de travailler dans un cadre correct. Un super merci à Ludovic Oudin avec qui je discutais beaucoup de mon sujet de thèse et qui me donnait quelques suggestions et m'a aidé à écrire de nombreux scripts sur R. Je remercie Michel Meybeck avec qui je discutais souvent de ma thèse et recevais ses avis en retour. Merci également à Pierre Ribstein qui m'a aidé et guidé après mon Master pour la recherche.

J'adresse un merci amical à Mounir, Nouradine, Pedro, Gurpreet, Etienne, Diane, Pierre, Morgane, Elijah, Xingcheng, Thibault, à qui je souhaite une bonne fin de thèse et suite ; avec une pensée pour ceux qui ont déjà soutenu, dont Salma et Mohammed. J'adresse aussi un merci à Maxime et Alexandre.

Mes remerciements vont également à l'endroit de Camille Pinet de l'IGN-FI pour toutes les informations échangées sur les méthodes utilisées pour cartographier les cours d'eau intermittents dans le réseau hydrographique de quelques pays d'Afrique et qui a aidé à la recherche des données notamment d'occupation du sol sur le Burkina Faso concernant les années 1950 à 1990 même si ces données n'ont pas pu être exploitées à cause d'une certaine lisibilité des cartes. Je remercie Javier Perez Saez pour les nombreux échanges qu'on a eus sur son travail de prédiction de l'intermittence au Burkina Faso ainsi que les données et scripts qu'il a partagés avec moi.

Je remercie le BUMIGEB (Bureau des Mines et de la Géologie du Burkina Faso) pour la fourniture des données géologiques, l'ANAM-BF (Agence Nationale de la Météorologie du Burkina Faso) pour la fourniture des données météorologiques, les instituts géographiques du Burkina Faso (IGB), du Mali (IGM), du Bénin (IGN), de Madagascar (FTM) et de l'Afrique du Sud (DWS), pour la fourniture des données sur le réseau hydrographique. Je remercie également les fournisseurs des données de débit obtenues auprès de la DGRE (Direction Générale des Ressources en Eau), du SIEREM (Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation), et du GRDC (Global Runoff Data Center).

Je remercie chaleureusement mon ami Jedidja Ouedraogo pour son soutien et son aide lors de mes mobilités en France.

Un merci à la famille Poher et Dacheux chez qui j'ai logé à Éragny sur Oise toutes les fois où j'étais en mobilité en France. Je ne pourrai oublier la période où a débuté la covid 19 en France suivie du confinement général. Le fait d'avoir habité avec eux pendant cette période m'a permis de découvrir la vie familiale française et de facilement gérer la crise, car j'avais des gens avec qui discuter.

Je n'oublie pas mon ami Arsène et les amis du groupe d'étude « *SB* », ils sauront se reconnaître. Merci pour vos soutiens multiples et multiformes.

J'aimerais à présent remercier mes parents pour leur soutien sans faille qui m'a été d'un secours à chaque étape de mes études et particulièrement pendant la thèse. Recevez ici mes remerciements les plus sincères.

Dans ces remerciements non exhaustifs, je voudrais remercier tous ceux qui ont contribué d'une manière ou d'une autre à la réalisation de cette thèse. Qu'ils trouvent ici l'expression de ma profonde gratitude.

# EPIGRAPHE

*“If the earth is a mother then rivers are her veins.” — Amit Kalantri*

## RESUME

Les cours d'eau intermittents sont des cours d'eau qui cessent de couler à un moment donné dans l'espace et le temps. Ils constituent la majorité des cours du monde, se retrouvent sous tous les types de climats et sont particulièrement répandus dans les régions arides et semi-arides. Cependant, la compréhension de la variabilité spatio-temporelle des cours d'eau intermittents ainsi que leur localisation précise dans le réseau hydrographique reste limitée en raison d'un manque de données hydrométriques et hydrographiques et des contraintes associées aux méthodes traditionnelles de cartographie par image satellitaire (opacité de la végétation et des nuages, résolution insuffisante). Ces difficultés sont accentuées dans les régions à données limitées comme l'Afrique et cette thèse vise principalement à développer des approches méthodologiques qui s'appuient sur des modèles de forêts aléatoires (*Random Forest*) ainsi que des analyses statistiques pour caractériser la distribution spatiale des cours d'eau intermittents et mieux appréhender les mécanismes de contrôle de l'intermittence en Afrique afin de pallier les lacunes d'observations.

D'abord, cette thèse s'est focalisée à l'échelle régionale au Burkina Faso où 49 stations de jaugeage des cours d'eau avec au minimum quatre ans de données sur la période 1955-1985 ont été examinées. Le nombre moyen de mois à débit nul par an ( $\overline{Ndry}$ ) a été utilisé comme prédicteur pour définir quatre classes croissantes d'intermittence du débit à savoir : permanent (0-1 mois à débit nul), faiblement intermittent (2-4), fortement intermittent (5-7), éphémère (8-12). Une analyse en composante principale (ACP) réalisée sur les 49 stations de jaugeages a montré que, bien que l'ordre de Strahler et la précipitation moyenne annuelle influencent la répartition géographique des différentes classes d'intermittence au Burkina Faso, la perméabilité moyenne et la surface amont des bassins versants expliquent principalement cette répartition. Cette étude suggère que la saisonnalité de la précipitation au Burkina Faso rend également saisonnier l'écoulement des cours d'eau, quelle que soit la précipitation moyenne annuelle à moins que les processus hydrogéologiques ne soient impliqués, notamment à travers la contribution des aquifères au débit de base pendant les périodes d'étiage. L'approche de prédiction par *Random Forest* a permis d'estimer sur le réseau hydrographique LCS que 83% de la longueur totale des cours d'eau au Burkina Faso sont intermittents contre 98% dans le réseau hydrographique nationale de référence (IGB-BNDT). Cette différence de proportion s'explique en partie par le fait que le modèle est entraîné sur des stations localisées sur de grands cours d'eau et son extrapolation sur des petits cours d'eau sous-estime l'intermittence.

Ensuite, à l'échelle de l'Afrique, 1125 stations de jaugeage avec au moins 4 ans de données sont examinées sur la période 1958-1991. Plusieurs modèles *Random Forest* ont été entraînés à relier les classes d'intermittence observée au droit des stations de jaugeage aux valeurs caractéristiques de 15 variables environnementales clés identifiées. L'analyse de l'importance des variables suggère que l'indice d'aridité (P/ETP), la surface amont des bassins versants et l'évapotranspiration potentielle moyenne annuelle sont les facteurs de contrôle les plus importants de l'intermittence à l'échelle continentale de l'Afrique. Cela se confirme d'autant plus que la majorité des stations classées intermittentes dans l'échantillon

analysé ont des valeurs d'évapotranspiration potentielle moyenne annuelle qui sont supérieures à la précipitation moyenne annuelle (c.-à-d. indice d'aridité < 1). La perméabilité qui est identifiée au Burkina Faso comme la variable de contrôle la plus importante ne ressort pas à l'échelle de l'Afrique parmi les contrôles les plus importants. Cela s'expliquerait par les incertitudes associées aux données hydrogéologiques à large échelle, mais aussi par une influence moindre en général que celle des variables climatiques. Avec des performances modérées, le modèle prédit sur le réseau hydrographique LCS qu'en Afrique 44 % en longueur des cours d'eau sont permanents tandis que 56 % sont intermittents (9 % faiblement intermittents, 31 % fortement intermittents et 16 % éphémères). Les prédictions du modèle capturent en général la distribution spatiale de l'intermittence dans les réseaux hydrographiques nationaux de référence de l'Afrique du Sud, du Bénin, de Madagascar, du Mali et un peu moins au Burkina Faso. Cette étude met également en évidence que la classification de l'intermittence présente dans les réseaux hydrographiques nationaux de référence sous-estime en général l'intermittence.

Enfin, cette thèse s'est appesantie sur les difficultés des réseaux hydrographiques globaux et continentaux à reproduire la variabilité spatiale de la densité de drainage observée sur les réseaux hydrographiques nationaux de référence des différents pays en Afrique. Une méthode simple pour extraire les cours d'eau des MNT (Modèle Numérique de Terrain) à partir de surface contributive ( $A_{\min}$ ) variable spatialement a été développée. Une première relation établit entre  $A_{\min}$  et la densité de drainage observée au Burkina Faso et validée au Mali et à Madagascar a donné des performances satisfaisantes. Une seconde relation entre la densité de drainage observée et des variables environnementales clés sur l'ensemble du jeu de données (Burkina Faso, Mali, Madagascar) a donné également en calibration et validation des performances satisfaisantes. La combinaison de ces deux relations a permis de déduire des valeurs de  $A_{\min}$  à l'échelle de l'Afrique variables spatialement en fonction des caractéristiques du paysage. Ces valeurs de  $A_{\min}$  ont permis enfin d'extraire des cours d'eau à l'échelle de l'Afrique qui reflètent au mieux la variabilité spatiale de la densité de drainage observée dans les réseaux hydrographiques de référence.

**Mots-clés** : Hydrologie, Cours d'eau intermittents, Aridité, Perméabilité, *Random Forest*, Acp, Hydrographie, Densité de drainage, MNT, Surface contributive, Burkina Faso, Afrique.



## ABSTRACT

Intermittent rivers are rivers that stop flowing at some point in space and time. They constitute the majority of the world's rivers, occur in all types of climates, and are particularly prevalent in arid and semi-arid regions. However, the understanding of the spatio-temporal variability of intermittent rivers as well as their precise location in the hydrographic network remains limited due to a lack of hydrometric and hydrographic data and constraints associated with traditional satellite image mapping methods (vegetation and cloud opacity, insufficient resolution). These difficulties are accentuated in data-limited regions such as Africa and this thesis aims to develop methodological approaches that rely on Random Forest models and statistical analyses to characterize the spatial distribution of intermittent rivers and to better understand the controlling factors of intermittency in Africa in order to overcome the observation gaps.

Firstly, this thesis focused on the regional scale in Burkina Faso where 49 stream gauging stations with at least four years of data over the period 1955-1985 were examined. The mean number of months with zero flow per year ( $\overline{Ndry}$ ) was used as a predictor to define four increasing classes of flow intermittency namely: permanent (0-1 months with zero flow), weakly intermittent (2-4), highly intermittent (5-7) and ephemeral (8-12). A principal component analysis (PCA) performed on 49 gauging stations showed that, although the Strahler order and the average annual precipitation influence the geographical distribution of the different intermittency classes in Burkina Faso, the average permeability and the upstream catchment area mainly explain this distribution. This study suggests that the seasonality of precipitation in Burkina Faso also makes streamflow seasonal regardless of mean annual precipitation unless hydrogeological processes are involved, notably through the contribution of aquifers to baseflow during low-flow periods. The Random Forest model estimated that 83% of the total length of the rivers in Burkina Faso is intermittent compared to 98% in the national reference hydrographic databases (IGB-BNDT). This difference in proportion is partly explained by the fact that the model is trained on gauges stations located on large rivers and its extrapolation on small rivers underestimates the intermittence.

Secondly, at the African scale, 1125 gauging stations with at least 4 years of data are examined over the period 1958-1991. Several Random Forest models were trained to relate the classes of intermittency observed at the gauging stations to the characteristic values of the key environmental variables identified (15). Analysis of the importance of the variables suggests that the aridity index (P/ETP), upstream catchment area and mean annual potential evapotranspiration are the most important controlling factors of intermittency at the continental scale in Africa. This is further confirmed by the fact that the majority of gauges stations classified as intermittent in the sample analyzed have mean annual potential evapotranspiration values that are higher than the mean annual precipitation (i.e, aridity index  $< 1$ ). Permeability, which is identified in Burkina Faso as the most important controlling variable, does not emerge at the African scale as one of the most important controls. This could be explained by the uncertainties associated with large-scale hydrogeological data but also by a lesser influence of these variables compared to climatic variables at a larger scale.

With moderate performance, the model predicts that in Africa 44% of the river length is permanent while 56% is intermittent (9% weakly intermittent, 31% highly intermittent and 16% ephemeral). The model predictions generally capture the spatial distribution of intermittency in the national reference hydrographic database of South Africa, Benin, Madagascar, and Mali somewhat less so in Burkina Faso. This study also highlights that the classification of intermittency present in the national reference hydrographic database generally underestimates intermittency.

Finally, this thesis focused on the difficulties of global and continental river networks to reproduce the spatial variability of the observed drainage density in the national reference hydrographic database of different countries in Africa. A simple method for extracting streams from DEM (Digital Elevation Model) with a spatially variable contributing area ( $A_{min}$ ) was developed. A first relationship established between  $A_{min}$  and observed drainage density in Burkina Faso and validated in Mali and Madagascar gave satisfactory performances. A second relationship established between observed drainage density and key environmental variables on the whole dataset (Burkina Faso, Mali, Madagascar) also gave satisfactory performances in calibration and validation. The combination of these two steps relations allowed us to deduce spatially variable values of  $A_{min}$  at the African scale according to the characteristics of the landscape. These  $A_{min}$  values were finally used to extract rivers network in Africa that best reflect the spatial variability of the drainage density observed in the reference hydrographic databases.

**Keywords:** Hydrology, Intermittent rivers, Aridity, Permeability, Random Forest, PCA, Hydrography, Drainage density, DEM, Contributive area, Burkina Faso, Africa.

# DIFFUSION DES RESULTATS

## Articles

1. **Belemtougri, A.P.**, Ducharne, A., Tazen, F., Oudin, L., Karambiri, H., 2021a. Understanding key factors controlling the duration of river flow intermittency: Case of Burkina Faso in West Africa. *Journal of Hydrology: Regional Studies*, 37, 100908. <https://doi.org/10.1016/j.ejrh.2021.100908>
2. **Belemtougri, A.P.**, Ducharne, A., Karambiri, H., 2021b. Improvement of river network representation in Africa: application of an approach based on digital elevation data and environmental characteristics. *Proceedings of the International Association of Hydrological Sciences*, 384, 19–23. <https://doi.org/10.5194/piahs-384-19-2021>
3. Fovet, O., **Belemtougri, A.**, Boithias, L., Braud, I., Charlier, J., Cottet, M., Daudin, K., Dramais, G., Ducharne, A., Folton, N., Grippa, M., Hector, B., Kuppel, S., Le Coz, J., Legal, L., Martin, P., Moatar, F., Molénat, J., Probst, A., Riotte, J., Vidal, J., Vinatier, F., Datry, T., 2021. Intermittent rivers and ephemeral streams: Perspectives for critical zone science and research on socio-ecosystems. *Wiley Interdisciplinary Reviews: Water*, 8(4), e1523. <https://doi.org/10.1002/wat2.1523>

## Actes des colloques ou conférences

1. Yonaba, R., **Belemtougri, A.**, Tazen, F., Mounirou L., Koïta, M., Karambiri, H., Hamma, Y., 2021 : Assessing the accuracy of SM2RAIN (Soil Moisture to Rainfall) products in poorly gauged countries : the case of Burkina Faso in the West African Sahel, **IAHS-AISH Scientific Assembly 2022**, Online, 29 May – 3 June 2022, IAHS2022-263. <https://doi.org/10.13140/RG.2.2.22517.86249>. **Communication orale.**
2. **Belemtougri, A.P.**, Ducharne, A., Karambiri, H., 2021. Improvement of river network representation in Africa: Application of an approach based on digital elevation data and environmental characteristics. **4<sup>e</sup> conférence internationale sur « l'hydrologie des grands fleuves africains »**, Cotonou, Bénin, novembre 2021. **Communication orale.**
3. **Belemtougri, A.P.**, Ducharne, A., Tazen, F., Oudin, L., Karambiri, H., 2021. Understanding key factors controlling the duration of river flow intermittency: Case of Burkina Faso in West Africa, **1<sup>re</sup> conférence Ozcar Tereno, Strasbourg, France, 5 - 7 Octobre 2021.** **Communication orale.**

4. **Belemtougri, A., Ducharne, A., and Karambiri, H.:** Influence of Lithology, Climate and Topography on the duration of flow intermittence in Burkina Faso., **EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-3909, <https://doi.org/10.5194/egusphere-egu2020-3909>.** **Communication orale.**
5. **Belemtougri, A., Ducharne, A., Karambiri, H., 2020.** Vers une caractérisation exhaustive du réseau hydrographique et de son intermittence en Afrique : développements méthodologiques et applications hydrologiques. **Journée des doctorants de l'ED GRNE. Communication orale.**
6. **Belemtougri, A., Ducharne, A., Karambiri, H., 2020.** Comprendre les facteurs clés qui contrôlent la durée de l'intermittence du débit des rivières : *cas du Burkina Faso en Afrique de l'Ouest. 8<sup>e</sup> édition des Doctoriales 2iE, Ouagadougou-Burkina Faso.* **Communication orale.**
7. **Belemtougri, A., Ducharne, A., Karambiri, H., 2019.** Vers une caractérisation exhaustive du réseau hydrographique et de son intermittence en Afrique : développements méthodologiques et applications hydrologiques. *7<sup>e</sup> édition des Doctoriales 2iE, Ouagadougou-Burkina Faso.* **Poster.**

# TABLE DES MATIERES

<b>DÉDICACE</b> .....	<b>i</b>
<b>REMERCIEMENTS</b> .....	<b>ii</b>
<b>EPIGRAPHE</b> .....	<b>iv</b>
<b>RESUME</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>DIFFUSION DES RESULTATS</b> .....	<b>ix</b>
<b>TABLE DES MATIERES</b> .....	<b>xi</b>
<b>LISTE DES ACRONYMES ET ABREVIATIONS</b> .....	<b>xiv</b>
<b>INTRODUCTION GENERALE</b> .....	<b>1</b>
<b>1- Contexte scientifique et problématique de la recherche</b> .....	<b>2</b>
<b>2- Questions de recherches et objectifs de la thèse</b> .....	<b>9</b>
<b>3- Structuration du manuscrit de la thèse</b> .....	<b>10</b>
<b>Chapitre I - Description, analyse préliminaire des données de l'étude et méthodes utilisées</b> .....	<b>13</b>
<b>I - 1 Données hydrographiques</b> .....	<b>14</b>
I - 1 - 1 Description et comparaison en Afrique des données hydrographiques globales collectées .....	<b>14</b>
I - 1 - 2 Présentation des données hydrographiques de référence collectées en Afrique .....	<b>25</b>
I - 1 - 3 Comparaison des données hydrographiques de référence avec celles d'études globales et locales.....	<b>31</b>
<b>I - 2 Sources et description des données hydro-environnementales</b> .....	<b>41</b>
I - 2 - 1 Données de débits.....	<b>41</b>
I - 2 - 2 Données hydro-climatiques .....	<b>44</b>
I - 2 - 3 Données topographiques .....	<b>47</b>
I - 2 - 4 Données géologiques et propriétés des sols .....	<b>50</b>
I - 2 - 5 Données de pressions anthropiques.....	<b>53</b>
<b>I - 3 Prétraitements des données et caractéristique des bassins versants</b> .....	<b>54</b>
<b>I - 4 Mise en œuvre d'outils d'apprentissage automatique : cas de Random Forest</b> ....	<b>57</b>
I - 4 - 1 Notions générales sur Random Forest.....	<b>57</b>
I - 4 - 2 Paramètres d'une Random Forest.....	<b>61</b>
I - 4 - 3 Méthode de rééchantillonnage.....	<b>63</b>

I - 4 - 4	Mesure de l'importance des variables et sélections des variables prédictives.	65
I - 4 - 5	Critères d'évaluation de la précision du modèle .....	68
<b>I - 5</b>	<b>Conclusion .....</b>	<b>71</b>
<b>Chapitre II - Comprendre les facteurs clés contrôlant la durée de l'intermittence du débit des rivières : cas du Burkina Faso en Afrique de l'Ouest..... 73</b>		
<b>II - 1</b>	<b>Résumé .....</b>	<b>74</b>
<b>II - 2</b>	<b>Introduction .....</b>	<b>75</b>
<b>II - 3</b>	<b>Data and Methods .....</b>	<b>78</b>
II - 3 - 1	Study Area.....	78
II - 3 - 2	Hydrological and environmental data .....	80
II - 3 - 3	Flow index and intermittency classification .....	83
II - 3 - 4	Assessment of the potential factors of intermittency .....	84
<b>II - 4</b>	<b>Results .....</b>	<b>87</b>
II - 4 - 1	Flow intermittency classes .....	87
II - 4 - 2	Principal Component Analysis (PCA) .....	92
<b>II - 5</b>	<b>Discussion.....</b>	<b>94</b>
II - 5 - 1	Comparison with previous studies .....	94
II - 5 - 2	Link between intermittency and catchment hydrology .....	96
II - 5 - 3	Limitation of the study and potential applications .....	99
<b>II - 6</b>	<b>Conclusion.....</b>	<b>100</b>
<b>II - 7</b>	<b>Supplementary to the journal article .....</b>	<b>102</b>
<b>II - 8</b>	<b>Analyses complémentaires sur la prédiction de l'intermittence.....</b>	<b>114</b>
<b>Chapitre III - Caractérisation et prédiction spatiale de l'intermittence des cours d'eau à l'échelle continentale de l'Afrique ..... 124</b>		
<b>III - 1</b>	<b>Résumé.....</b>	<b>125</b>
<b>III - 2</b>	<b>Introduction .....</b>	<b>126</b>
<b>III - 3</b>	<b>Données et méthodes .....</b>	<b>128</b>
III - 3 - 1	Sélection des stations hydrométriques .....	128
III - 3 - 2	Classification de l'intermittence des cours d'eau .....	135
III - 3 - 3	Réseau hydrographique et variables prédictives hydro-environnementales.....	137
III - 3 - 4	Développement du modèle Random Forest.....	142
<b>III - 4</b>	<b>Résultats .....</b>	<b>145</b>
III - 4 - 1	Classes d'intermittence des cours d'eau échantillonnés .....	145
III - 4 - 2	Performances du modèle Random Forest et ordre d'importance des variables .....	149
III - 4 - 3	Distribution spatiale de l'intermittence prédite à l'échelle de l'Afrique ....	157

III - 4 - 4 Comparaisons de la distribution spatiale de l'intermittence à l'échelle de l'Afrique .....	167
<b>III - 5 Discussion .....</b>	<b>179</b>
<b>III - 6 Conclusion .....</b>	<b>185</b>
<b>Chapitre IV - Développement d'une méthode pour extraire les cours d'eau à partir d'aires contributives variables en Afrique .....</b>	<b>187</b>
<b>IV - 1 Résumé.....</b>	<b>188</b>
<b>IV - 2 Introduction.....</b>	<b>189</b>
<b>IV - 3 Matériel et Méthodes .....</b>	<b>191</b>
IV - 3 - 1 Zones d'étude .....	191
IV - 3 - 2 Description des données .....	192
IV - 3 - 3 Méthodes .....	193
<b>IV - 4 Résultats et Discussion .....</b>	<b>197</b>
<b>IV - 5 Conclusion .....</b>	<b>205</b>
<b>CONCLUSIONS GENERALES ET PERSPECTIVES .....</b>	<b>206</b>
<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>212</b>
<b>LISTE DES FIGURES .....</b>	<b>239</b>
<b>LISTE DES TABLEAUX.....</b>	<b>248</b>
<b>ANNEXES .....</b>	<b>250</b>
<b>Annexe - 1 - Représentation spatiale des variables hydro-environnementales investiguées.....</b>	<b>251</b>
<b>Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso.....</b>	<b>255</b>
<b>Annexe - 3 - Éléments supplémentaires sur la prédiction de l'intermittence des cours d'eau en Afrique .....</b>	<b>272</b>

## LISTE DES ACRONYMES ET ABREVIATIONS

<b>2iE</b>	: Institut International d'Ingénierie de l'Eau et de l'Environnement
<b>ACP</b>	: Analyse en Composante Principale
<b>ADHI</b>	: African Database of Hydrometric Indices
<b>ANAM-BF</b>	: Agence Nationale de la Météorologie du Burkina Faso
<b>AWRD</b>	: African Water Resource Database
<b>BAD</b>	: Banque Africaine de Développement
<b>BNDT</b>	: Base Nationale de Données Topographiques
<b>BUMIGEB</b>	: Bureau des Mines et de la Géologie du Burkina Faso
<b>CRU</b>	: Climate Research Unit
<b>DEM</b>	: Digital Elevation Model
<b>DGRE</b>	: Direction Générale des Ressources en Eau
<b>EGU</b>	: European Geosciences Union
<b>EPSG</b>	: European Petroleum Survey Group
<b>ESRI</b>	: Environmental Systems Research Institute
<b>FAO</b>	: Food and Agriculture Organisation
<b>FRIEND</b>	: Flow Regimes from International Experimental and Network Data
<b>FTM</b>	: Foiben Taosarintanin'i Madagasikara
<b>GIRES</b>	: Global Intermittent Rivers and Ephemeral Streams
<b>GLHYMPS</b>	: Global Hydrogeology MaPS
<b>GLiM</b>	: Global Lithological Map
<b>GRAND</b>	: Global Reservoir and Dam Database
<b>GRDC</b>	: Global Runoff Data Center
<b>GRIN</b>	: Global River Network
<b>GRNE</b>	: Géosciences Ressources Naturelles et Environnement
<b>GSIM</b>	: Global Streamflow Indices and Metadata
<b>GTOPO30</b>	: Global 30 arc-second elevation
<b>HydroSHEDS</b>	: Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales
<b>IGB</b>	: Institut Géographique du Burkina Faso
<b>IGM</b>	: Institut Géographique du Mali
<b>IGN-FI</b>	: Institut Géographique National France International



<b>IRD</b>	: Institut de Recherche et de Développement
<b>LAI</b>	: Leaf Area Index
<b>LCS</b>	: Lithology Climate and Slope
<b>LEHSA</b>	: Laboratoire Eaux HydroSystèmes et Agriculture
<b>MDA</b>	: Mean Decrease Accuracy
<b>MDG</b>	: Mean Decrease Gini
<b>MERIT</b>	: Multi-Error-Removed Improved-Terrain
<b>METIS</b>	: Milieux Environnementaux, Transferts et Interactions dans les hydrosystèmes et les Sols
<b>MNT</b>	: Modèle Numérique de Terrain
<b>NASA</b>	: National Aeronautics and Space Administration
<b>NDVI</b>	: Normalized Difference Vegetation Index (indice de végétation normalisé)
<b>NHD</b>	: National Hydrography Dataset
<b>ORSTOM</b>	: Office de Recherche Scientifique et Technique d'Outre Mer
<b>PCA</b>	: Principal Components Analysis
<b>RF</b>	: Random Forest
<b>RMSE</b>	: Root Mean Square Error (erreur quadratique moyenne)
<b>SIEREM</b>	: Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation
<b>SIG</b>	: Système d'Information Géographique
<b>SRTM</b>	: Shuttle Radar Topography Mission
<b>SU</b>	: Sorbonne Université
<b>UNEP</b>	: United Nation Environment Programme
<b>USGS</b>	: United States Geological Survey
<b>UTM</b>	: Universal Transverse Mercator
<b>WGS84</b>	: World Geodetic System 1984

# **INTRODUCTION GENERALE**

## 1- Contexte scientifique et problématique de la recherche

Les cours d'eau représentent une faible portion de la ressource en eau existante sur terre, soit 0,00012% (De Marsily, 2009). Leur importance, inversement à cette proportion, est immense. Ils sont les principales sources d'eau douce pour satisfaire de multiples besoins fondamentaux de la société tels que l'approvisionnement en eau pour l'agriculture, la consommation industrielle et domestique (Hoekstra et Mekonnen, 2012). La représentation précise et la connaissance des caractéristiques des cours d'eau constituent dès lors des informations essentielles pour une meilleure quantification et gestion des ressources en eau (Heine *et al.*, 2004).

Le réseau hydrographique représente l'ensemble hiérarchisé et structuré des cours d'eau qui assurent le drainage superficiel ou souterrain, permanent ou temporaire, d'un bassin versant, d'une région donnée ou d'un pays. Le réseau hydrographique de référence ou base hydrographique de référence est le terme adopté dans ce manuscrit pour désigner le réseau hydrographique établi par les services hydrologiques et instituts géographiques d'un pays donné.

Au sein du réseau hydrographique, les cours d'eau peuvent être classés comme intermittents ou permanents selon la variabilité spatio-temporelle de leur écoulement de surface (**Figure i-1**). Les cours d'eau permanents peuvent être définis comme des cours d'eau qui ont un écoulement continu tout au long de l'année tandis que les cours d'eau intermittents cessent de couler à un moment donné dans l'espace et le temps (Acuña *et al.*, 2014 ; González-Ferreras et Barquín, 2017). D'autres définitions de l'intermittence existent, elles dépendent en général du nombre de jours ou mois à débit nul par an pour être considérés comme intermittents ou permanents et en fonction de cette définition de l'intermittence, les résultats de classification diffèrent (Fritz *et al.*, 2020). Dans la littérature, les cours d'eau intermittents prennent diverses appellations comme temporaire, éphémère, irrégulier, non permanent, épisodique, saisonnier, et interrompu (Uys et O'Keeffe, 1997).

Les cours d'eau intermittents sont présents sur tous les continents et sous tous les climats (Larned *et al.*, 2010). Ils représenteraient la majorité des cours d'eau à l'échelle mondiale (Datry *et al.*, 2014) et sont susceptibles d'augmenter à l'avenir dans certaines régions en raison des perturbations anthropiques et du changement et de la variabilité climatique (Chiu *et al.*, 2017 ; De Girolamo *et al.*, 2017 ; Pumo *et al.*, 2016).



**Figure i-1:** Le Massili à Gonsé - Burkina Faso : Différentes périodes d'un cours d'eau intermittent, **a)** période sèche (13/04/2019) ; **b)** période humide (04/08/2019) – Coord. Géog. (-1,39 ; 12,48°). Crédit : Axel Patindé Belemtougri.

L'identification préalable du caractère intermittent ou permanent des cours d'eau est essentielle pour une meilleure évaluation de l'état écologique des écosystèmes fluviaux (Makarewicz, 2020 ; Merritt *et al.*, 2021). En effet, la durée de la phase sèche est critique pour la survie des écosystèmes et le fonctionnement hydrologique des cours d'eau (Sánchez-Montoya *et al.*, 2018 ; Vidal-Abarca *et al.*, 2020). La caractérisation de l'intermittence des cours d'eau fournit également une information spatio-temporelle sur l'intermittence du débit des cours d'eau et sert d'aide à la décision pour une mobilisation efficace de la ressource en eau de surface.

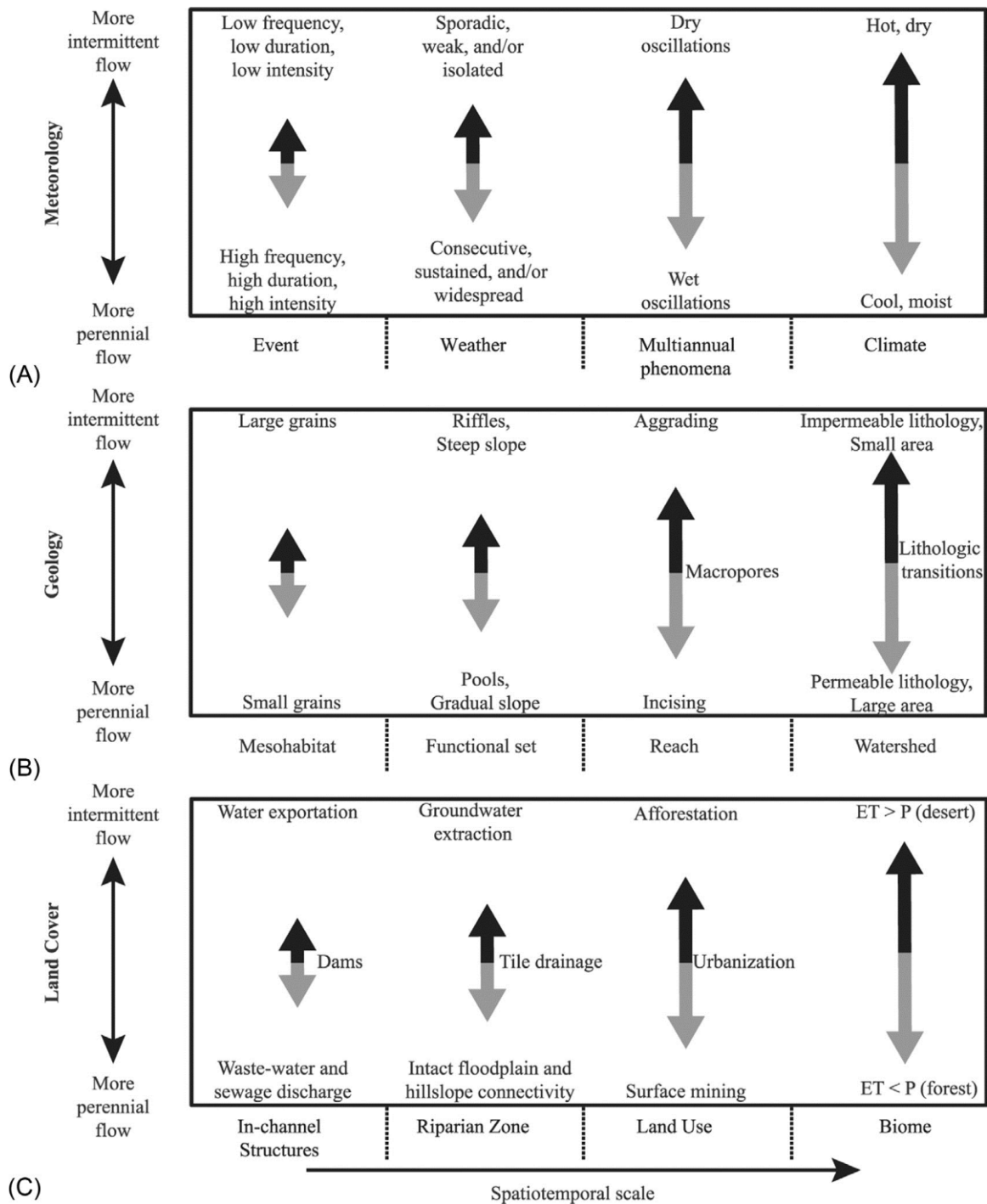
Les causes de l'assèchement de ces cours d'eau (**Figure i-2**) peuvent être naturelles (par exemple, le climat, la géologie ou la topographie) ou anthropiques (par exemple, les barrages, l'utilisation des terres) (Borg Galea *et al.*, 2019 ; Boulton *et al.*, 2017 ; Yonaba *et al.*, 2021a). La compréhension du contrôle de ces variables environnementales (climatiques, physiographiques, anthropiques, etc...) sur l'intermittence du débit des cours d'eau est importante pour la gestion des ressources en eau, en particulier dans les régions arides et semi-arides qui constituent en Afrique la majorité du climat (57,2%) (Peel *et al.*, 2007).

Dans les régions arides et semi-arides, l'évapotranspiration potentielle est souvent largement supérieure à la précipitation en moyenne annuelle, ce qui crée une limite sur la disponibilité de l'eau (Budyko *et al.*, 1974), favorisant généralement l'intermittence des cours d'eau. Dans ces régions, les cours d'eau intermittents ont tendance à être localisés à tous les ordres du réseau hydrographique (Acuña *et al.*, 2017). Dans les régions plus humides où la demande évaporative est inférieure à la précipitation, la majorité des cours d'eau a tendance à être permanents et les cours d'eau intermittents sont généralement localisés à l'amont du réseau

hydrographique (cours d'eau de tête de bassin) (Acuña *et al.*, 2017 ; Beaufort *et al.*, 2018). Cette dualité entre l'évapotranspiration potentielle et la précipitation est souvent résumée par l'indice d'aridité qui est le rapport entre l'évapotranspiration potentielle et la précipitation (Dey et Mishra, 2017 ; Greve *et al.*, 2019).

La **Figure i-2** (issue de Costigan *et al.*, 2016) résume l'influence de nombreux facteurs de contrôle de l'intermittence des cours d'eau. Ainsi, la fréquence, la durée, l'intensité et la variation interannuelle des précipitations ont un effet sur l'intermittence ou la permanence des cours d'eau. Les cours d'eau ayant des bassins versants avec une forte perméabilité du sol et un stock d'eau souterraine important sont généralement moins sensibles à l'intermittence, en particulier pendant les périodes d'étiage où la contribution des aquifères au débit de base peut être très importante (Lovill *et al.*, 2018 ; Wirth *et al.*, 2020). La pente du cours d'eau et la surface amont drainée influencent l'intermittence ou la permanence des tronçons de cours d'eau avec les fortes pentes et faible surface amont qui ont tendance à favoriser l'intermittence en facilitant l'évacuation rapide de l'écoulement en aval (Prancevic et Kirchner, 2019 ; Warix *et al.*, 2021).

L'influence anthropique est aussi un des facteurs de contrôle de l'intermittence et peut se manifester sous de nombreuses formes comme le pompage direct des cours d'eau, ou des nappes phréatiques et la construction de barrages (Fowe *et al.*, 2015 ; Zimmer *et al.*, 2020). Lorsque le pompage des eaux souterraines pour l'irrigation, l'approvisionnement en eau potable ou autres activités est suffisamment important et prolongé, il peut faire drastiquement baisser le niveau des aquifères, et si ces aquifères sont une source d'eau primaire pour soutenir l'écoulement des cours d'eau, le pompage peut effectivement contribuer à assécher les cours d'eau (Goodrich *et al.*, 2018 ; Kustu *et al.*, 2010). La construction de grands barrages peut entraîner une réduction très importante du débit en aval (Cooney et Kwak, 2013), ce qui contribuerait à l'intermittence des cours d'eau ou au contraire, les rejets artificiels d'eau peuvent contribuer à maintenir une certaine permanence de l'écoulement (Smakhtin, 2001) notamment dans le cas des barrages hydro-électriques.



**Figure i-2:** Échelle spatio-temporelle (axe des abscisses) des contrôles météorologiques (A), géologiques (B) et de l'occupation des sols (C) sur le régime d'écoulement, avec les facteurs favorisant l'intermittence (flèches noires supérieures) ou l'écoulement pérenne (flèches grises inférieures) le long des axes des ordonnées, l'axe des y. Cette figure est extraite de Costigan et al. (2016).

Des études récentes à l'échelle globale et continentale (aux États-Unis, en Europe et en Australie) ont montré que les facteurs de contrôle les plus importants de l'intermittence des cours d'eau sont liés au climat et en particulier à l'indice d'aridité (Addor et al., 2018 ; Hammond et al., 2021 ; Sauquet et al., 2021). À l'échelle locale du bassin versant, Bourke et al. (2021) et González-Ferreras et Barquín (2017) ont trouvé respectivement dans la péninsule

de Dampier dans le nord-ouest de l'Australie et dans le nord de l'Espagne que la perméabilité et la surface amont du bassin versant étaient les facteurs de contrôle les plus importants de l'intermittence des cours d'eau. Ces études montrent que l'interaction entre les variables environnementales est un jeu complexe qui contrôle l'intermittence ou la permanence des cours d'eau à différentes échelles spatiales. Par conséquent, il est nécessaire d'améliorer notre compréhension sur les facteurs de contrôle de l'intermittence ainsi que la distribution spatiale de l'intermittence.

Des informations précises sur la distribution spatiale des cours d'eau intermittents permettent par exemple d'évaluer leur rôle dans le cycle biogéochimique (Raymond *et al.*, 2013) ainsi que la disponibilité spatiale et temporelle en eau de surface (Fadul *et al.*, 2021 ; Petheram *et al.*, 2008). Les bases hydrographiques nationales de référence contiennent des informations hydrographiques sur la répartition spatiale des cours d'eau ainsi que, pour certaines d'entre elles, leur caractère intermittent ou permanent. Ces informations sont couramment issues de la digitalisation d'images satellitaires à haute résolution, des cartes topographiques, des photographies aériennes, de la photo-interprétation (Bertoldi *et al.*, 2012 ; Bird *et al.*, 2010 ; Jaud *et al.*, 2013 ; Kostianoy *et al.*, 2020 ; Scherer *et al.*, 2012) d'enquêtes de terrain et de la connaissance des résidents (Fritz *et al.*, 2013 ; Turner et Richter, 2011). Cependant, l'acquisition d'informations par ces méthodes est très coûteuse et reste difficilement applicable sur de très vastes zones (Fovet *et al.*, 2021).

Sur les cartes topographiques, les cours d'eau permanents sont souvent matérialisés par des lignes « bleues » continues, et les cours d'eau intermittents par des « pointillés ». Malheureusement, ces informations ne sont pas régulièrement mises à jour et la comparaison des informations sur l'intermittence disponible dans les bases de données nationales avec les observations réelles du débit des cours d'eau montre souvent une divergence (Hafen *et al.*, 2020). De plus, les très petits cours d'eau ne sont souvent pas représentés dans le réseau hydrographique (Schneider *et al.*, 2017) et en fonction du type de climat les cours d'eau sont souvent classés par défaut. Dans des milieux arides et semi-arides comme au Burkina Faso, les cours d'eau du réseau hydrographique sont supposés intermittents par défaut, sauf si un écoulement pérenne est observé (Fovet *et al.*, 2021). Enfin, l'accessibilité à ces données hydrographiques auprès des services hydrologiques et instituts géographiques nationaux reste particulièrement difficile en Afrique en raison du coût requis et de la réticence de certains pays à partager leurs données.

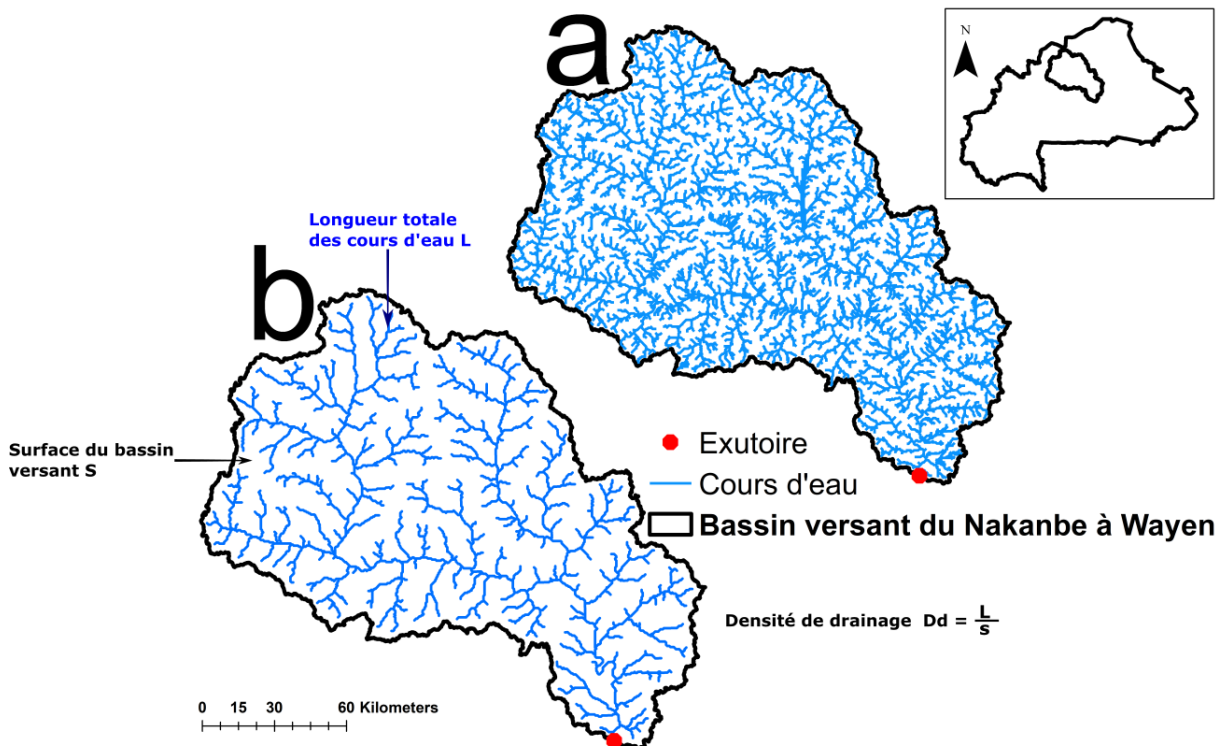
La période de "non débit" ou "débit nul" est une composante importante du régime hydrologique des cours d'eau intermittents. Une grande partie des travaux de recherche menés sur ces cours d'eau s'est appuyée sur le moment, la durée et la fréquence des périodes de débit nul comme indices hydrologiques pour classer (Hammond *et al.*, 2021 ; Kaplan *et al.*, 2019 ; Zipper *et al.*, 2021) et prédire la dynamique spatiale des cours d'eau intermittents (Jaeger *et al.*, 2019 ; Snelder *et al.*, 2013 ; Yu *et al.*, 2018). Sur de très vastes zones, la modélisation statistique peut être utilisée pour établir des relations entre ces caractéristiques du débit des cours d'eau (extraites des mesures des stations de jaugeage) et les variables environnementales (climat, géologie, topographie, occupation du sol, etc.), pour tenter d'estimer l'importance de chaque variable et de régionaliser les observations locales de l'intermittence du débit à travers les tronçons de rivière non jaugeés (Jaeger *et al.*, 2019). Dans ce cadre, les méthodes statistiques d'apprentissage automatique telles que les forêts aléatoires (*Random Forest*) sont couramment utilisées. Les forêts aléatoires sont une combinaison de classifieurs individuels appelés arbres de décision construits à partir des données d'apprentissage (Breiman *et al.*, 1984) et pour laquelle une prédiction finale est calculée en moyennant la prédiction de tous les arbres. Ces méthodes ont montré leur efficacité pour identifier les cours d'eau intermittents et ont permis par exemple d'estimer leur proportion en France entre 20 et 39 % de la longueur du réseau hydrographique (Snelder *et al.*, 2013). La première cartographie à l'échelle globale de la distribution spatiale des cours d'eau intermittents a été réalisée (Messenger *et al.*, 2021), permettant d'estimer leur proportion entre 51 à 60 % de la longueur totale des cours d'eau présents dans le réseau hydrographique hydroRIVERS (Linke *et al.*, 2019), mais ce réseau hydrographique manque de précision sur les petits cours d'eau (Amatulli *et al.*, 2022 ; Schneider *et al.*, 2017), et cette cartographie se limite à une classification binaire de l'intermittence (intermittent et permanent) avec les cours d'eau intermittents qui sont définis comme des cours d'eau qui ont au moins un jour sans écoulement par an (en moyenne sur au moins dix ans) et inversement pour les cours d'eau permanents.

Pour tenir compte des petits cours d'eau et caractériser l'intermittence sur de vastes zones, les méthodes d'extrapolation statistique peuvent être appliquées en exploitant l'information hydrographique extraite des modèles numériques de terrain (MNT) à haute résolution spatiale, désormais disponible à l'échelle globale (30 m, 90 m, 500 m, 1 km) (Lehner *et al.*, 2008). En effet, grâce aux Systèmes d'Information Géographique (SIG), il est possible de déduire de la pente des MNT, les directions d'écoulement de l'eau, étant donné que l'eau suit les pentes les



plus raides de l'amont vers l'aval, et ces informations permettent également de délimiter les bassins versants. Les MNT fournissent donc une occasion d'avoir, sur de vastes zones, des détails sur les cours d'eau, mais sous condition d'avoir des informations suffisantes sur l'endroit exact où débutent les cours d'eau. Cette extraction passe généralement par la définition d'une surface contributive minimale ( $A_{min}$ ) à partir de laquelle on estime passer d'un ruissellement de surface à une vraie rivière (présence d'un écoulement concentré dans les thalwegs) (Jenson et Domingue, 1988 ; O'Callaghan et Mark, 1984).

Ce seuil défini par l'utilisateur, permet de séparer les pixels du MNT qui drainent une surface suffisamment grande pour être considéré comme des cours d'eau (cours d'eau extraits) des pixels non cours d'eau. Il est généralement choisi assez arbitrairement et appliqué de manière constante sur de grandes zones. Ainsi, le choix d'un seuil  $A_{min}$  faible augmente la longueur des cours d'eau extraits et inversement (**Figure i-3**). Le choix de la valeur de cette surface contributive ( $A_{min}$ ) va donc conditionner la longueur des cours d'eau et influencer la densité de drainage (Dd) qui peut être définie comme le rapport entre la longueur totale des cours d'eau d'un bassin versant et sa surface (Horton, 1932, 1945).



**Figure i-3:** Cours d'eau du bassin versant du Nakanbé à Wayen (Burkina Faso) extraits à partir d'un MNT (500m x 500m) issu de la base HydroSHEDS. **a)**  $A_{min} > 2,5 \text{ km}^2$  ; **b)**  $A_{min} > 25 \text{ km}^2$ .

La répartition spatiale des cours d'eau ainsi générés dépend alors uniquement de la topographie, négligeant le contrôle des variables climatiques, géologiques et l'occupation des

sols alors qu'il peut par endroits être très important (Hooshyar *et al.*, 2019 ; Papageorgaki et Nalbantis, 2017 ; Sangireddy *et al.*, 2016). Une conséquence majeure est d'aboutir à une répartition spatiale très uniforme de la densité de drainage, ce qui ne correspond pas à la réalité sur le terrain (Passalacqua *et al.*, 2010 ; Schneider *et al.*, 2017 ; Strobl et Forte, 2007).

Pour s'affranchir de cette limite et obtenir un réseau hydrographique plus réaliste à l'échelle globale, des études récentes (Lin *et al.*, 2021 ; Schneider *et al.*, 2017) font dépendre spatialement les valeurs de  $A_{\min}$  en fonction de variables environnementales permettant donc d'extraire des cours d'eau dont la densité de drainage varie spatialement. Malgré des performances très satisfaisantes dans d'autres régions, ces réseaux hydrographiques appréhendent partiellement la densité de drainage observée sur les réseaux hydrographiques de référence en Afrique.

## **2- Questions de recherches et objectifs de la thèse**

Cette thèse se focalise en Afrique où il y a peu d'études consacrées à l'intermittence des cours d'eau certainement à cause des difficultés d'accès et de la disponibilité des données hydrographiques et hydrométriques, malgré d'importants problèmes d'accès à l'eau sur une grande partie du continent (Papa *et al.*, 2022). Cette thèse vise principalement une meilleure compréhension des processus qui contrôlent la permanence ou l'intermittence des cours d'eau à différentes échelles et s'inscrit dans une volonté de proposer une approche méthodologique de caractérisation des cours d'eau, en termes d'intermittence du débit, sur de grands domaines jusqu'à l'échelle du continent, afin de pallier les lacunes d'observations. Elle propose également une approche méthodologique pour améliorer la cartographie du réseau hydrographique total en Afrique.

Les questions de recherche principales qui motivent cette thèse sont les suivantes :

1. Quelles sont les variables environnementales les plus pertinentes pour expliquer la distribution spatiale de l'intermittence des cours d'eau à l'échelle régionale et continentale ?
2. Dans quelle mesure ces variables environnementales clés peuvent-elles être utilisées pour régionaliser ou extrapoler les observations locales d'intermittence tout le long du réseau hydrographique dans le but de cartographier la distribution spatiale des cours d'eau intermittents ?

3. Peut-on améliorer le réalisme des cours d'eau extraits (permanents + intermittents) à partir de MNT en se basant sur des informations hydrographiques de référence ainsi que de variables climatiques, physiographiques et géologiques ?

Les objectifs spécifiques (OS) se déclinent comme suit :

- **OS 1-)** Identifier les facteurs de contrôle de l'intermittence du débit des cours d'eau à l'échelle régionale : cas du Burkina Faso en Afrique de l'Ouest.
- **OS 2-)** Identifier les facteurs de contrôle de l'intermittence du débit et caractériser la distribution spatiale de l'intermittence du débit des cours d'eau à l'échelle continentale de l'Afrique.
- **OS 3-)** Proposer une nouvelle approche méthodologique de caractérisation du réseau hydrographique en Afrique qui se base sur une relation établie entre les surfaces contributives minimales ( $A_{min}$ ) et des variables environnementales pour extraire les cours d'eau à partir de MNT.

### **3- Structuration du manuscrit de la thèse**

Après cette **introduction générale**, le manuscrit de thèse se poursuit par quatre chapitres, complétés d'annexes :

- Le **premier chapitre** commence par une présentation de l'ensemble des données hydrographiques, climatiques, physiographiques et anthropiques collectées à l'échelle continentale comme régionale pour appuyer le travail. Il décrit les méthodes d'acquisition et de calculs de chacune de ces données. Le réseau hydrographique global qui représente au mieux les cours d'eau par rapport aux réseaux de référence de quelques pays (Burkina Faso, Mali, Bénin, Madagascar) est sélectionné en fin de compte comme support pour la cartographie des cours d'eau intermittents en Afrique. Une présentation détaillée des concepts et outils statistiques nécessaires à la compréhension des chapitres suivants est faite notamment sur les Systèmes d'Information Géographique (SIG) et les méthodes d'apprentissage automatique basées sur les forêts aléatoires « *Random Forest (RF)* ».

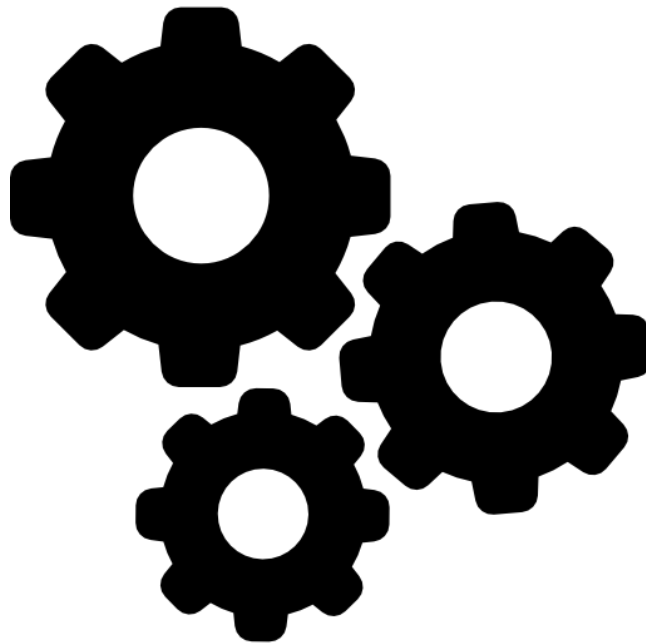
- Le **deuxième chapitre**, publié sous forme d'article scientifique dans *Journal of Hydrology: Regional Studies* (Belemtougri *et al.*, 2021a), identifie le rôle de variables environnementales qui expliquent le mieux la variation géographique du régime d'intermittence des cours d'eau et en particulier la durée de l'intermittence du débit. Cette étude est menée au Burkina Faso où les différents contextes climatiques sont représentatifs de l'Afrique de l'Ouest et font du pays un modèle réduit intéressant pour étudier et comprendre les facteurs clés qui contrôlent l'intermittence du débit dans cette région. À l'aide des facteurs clés identifiés et en préparation de l'application du modèle (*RF*) en Afrique, une prédiction spatiale de l'intermittence des cours d'eau au Burkina Faso est proposée en complément dans ce chapitre.
- Le **troisième chapitre** prolonge la recherche menée au chapitre II en identifiant les variables environnementales clés ainsi que leur importance dans le contrôle de l'intermittence des cours d'eau à l'échelle continentale de l'Afrique. Pour ce faire, plusieurs modèles (*RF*) sont utilisés pour établir des relations entre les variables environnementales clés et les informations sur l'intermittence extraites au droit d'un grand nombre de stations de jaugeage qui sont réparties spatialement sur le continent. Les meilleurs modèles en validation sont sélectionnés et utilisés en extrapolation pour prédire l'intermittence sur tout le linéaire hydrographique. L'influence individuelle des variables les plus importantes de l'intermittence est discutée et comparée aux résultats des études antérieures. Les proportions en longueur des différentes classes d'intermittence prédites sont évaluées et comparées aux informations présentes dans les réseaux hydrographiques globaux et de référence de quelques pays en Afrique (Burkina Faso, Bénin, Mali, Madagascar, Afrique du Sud).
- Dans le **quatrième et dernier chapitre**, la précision de l'information locale (réseau hydrographique de référence) collectée auprès des instituts géographiques et services hydrologiques a été utilisée pour tester une nouvelle manière d'extraire le réseau hydrographique. Le principe est de procéder en deux étapes, en établissant d'abord une relation entre les surfaces contributives minimales ( $A_{\min}$ ) et la densité de drainage observée ( $Dd_{\text{obs}}$ ), puis des relations entre la densité de drainage observée et divers facteurs explicatifs possibles (portant sur le climat, la lithologie, la pente et l'altitude). La combinaison de ces deux types de relations devrait permettre d'obtenir des valeurs de  $A_{\min}$  variables spatialement en fonction de ces facteurs explicatifs permettant ainsi d'extraire un réseau hydrographique réaliste à l'échelle de l'Afrique. Un gain

d'efficacité est espéré par la genericité de la première étape, comme indiqué dans les résultats du chapitre IV, publiés dans les *Proceedings de l'IAHS* (Belemtougri *et al.*, 2021b), ce qui permettrait de réduire les analyses spécifiques d'une région donnée à la seule étape 2.

- La **conclusion générale** fait un bilan des points importants abordés ainsi que la contribution apportée par cette étude dans la compréhension du fonctionnement des cours d'eau intermittents tout en suggérant quelques applications possibles. Des recommandations sont formulées à l'endroit des gestionnaires de bassins versants. Des discussions sont également abordées sur les limites de l'étude et des perspectives de recherche sont envisagées à la suite de cette thèse.

## **Chapitre I - Description, analyse préliminaire des données de l'étude et méthodes utilisées**

Ce chapitre vise à présenter les différentes données collectées, et à mieux expliciter la problématique de la représentation des cours d'eau et de leur intermittence à travers une comparaison des différentes bases de données hydrographiques à l'échelle de l'Afrique et dans quelques pays. Cette comparaison permettra de sélectionner la base de données hydrographiques la plus adaptée comme support pour la cartographie de l'intermittence des cours d'eau au **chapitre III**.



## I - 1 Données hydrographiques

### I - 1 - 1 Description et comparaison en Afrique des données hydrographiques globales collectées

Plusieurs bases de données en accès libre existent pour décrire les réseaux hydrographiques et sont principalement extraites des MNT. Les principales bases de données hydrographiques globales et continentales qui couvrent l'Afrique ainsi que leur description sont résumées dans le **Tableau I-1**.

Pour caractériser spatialement et comparer les différents réseaux hydrographiques, la densité de drainage (Dd) a été calculée en utilisant des mailles carrées pour lesquelles la longueur cumulée des cours d'eau est divisée par la surface de la maille. Une analyse basée sur des mailles a été effectuée, car elle permet de calculer la densité de drainage spatialement avec le même support géographique (Oguchi, 1997 ; Schneider *et al.*, 2017). Dans son étude sur les variables contrôlant la variation de la densité de drainage en tenant compte de la taille du maillage (1, 4, 16 et 64 km<sup>2</sup>), Luoto (2007) a trouvé que ce sont les relations établies sur le maillage de 64 km<sup>2</sup> qui expliquent le mieux la variance de la densité de drainage; ce qui conforte le choix de la taille du maillage 5' (environ 8 km x 8 km) dans cette thèse pour calculer les densités de drainage.

Il est important de préciser que les termes mailles, pixels, cellules, utilisés dans cette thèse sont indifférenciés et font référence à une représentation de taille carrée.

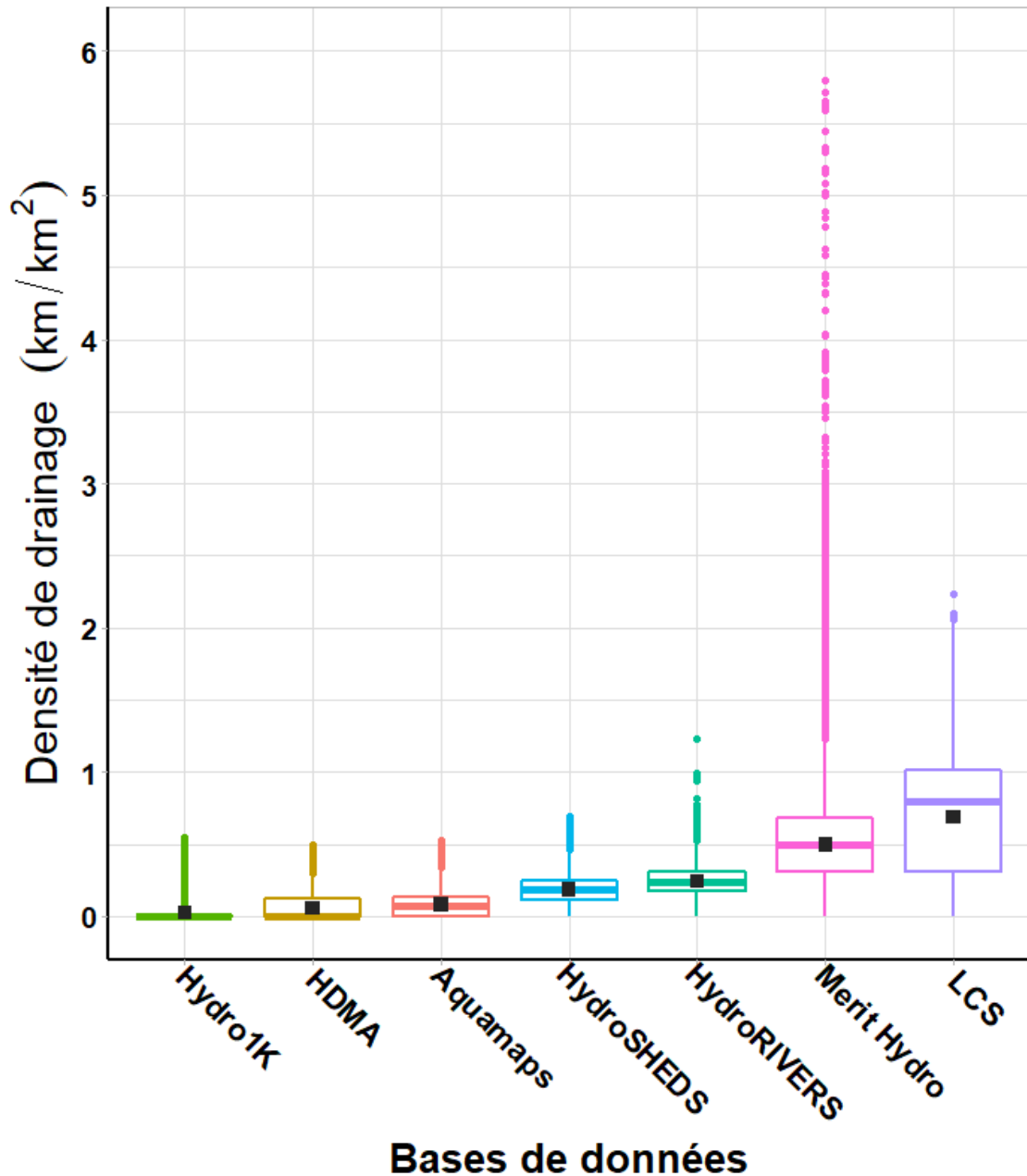
**Tableau I-1:** Description des bases de données hydrographiques principales couvrant l'Afrique, les densités de drainages ( $Dd$ ) ainsi que l'écart-type ( $Sd$ ) sont calculés dans des mailles de 5' à l'échelle de l'Afrique. Les bases de données sont classées par ordre chronologique croissantes.

Bases de données	Description	Autres informations
<b>Digital Chart of the World (DCW)</b>	Aussi connue sous le nom de « VMAP0 », elle a été mise en place initialement par l'agence de cartographie de la défense américaine, elle est une base de données à l'échelle globale conçue pour répondre aux besoins des pilotes et des équipages aériens lors de la navigation. Elle prend en compte plusieurs thématiques, dont l'hydrographie, et fournit donc des informations sur les rivières et lacs. La dernière mise à jour date de 1992 et sa précision est connue pour varier considérablement d'une région à une autre (Lehner <i>et al.</i> , 2006). La version d'origine n'est plus distribuée officiellement, mais une version très simplifiée du jeu de données d'origine est actuellement distribuée par (ESRI, 1993).	Échelle : 1 / 1 000 000 Dernière mise à jour en 1992
Source	<a href="https://en.wikipedia.org/wiki/Digital_Chart_of_the_World">https://en.wikipedia.org/wiki/Digital_Chart_of_the_World</a>	
<b>HYDRO1K</b> (U.S. Geological Survey, 2000)	HYDRO1K fournit des informations hydrographiques et est extraite à partir du MNT GTOPO30 de l'USGS avec une aire contributive minimale $A_{min}$ <b>constante de 1000 km<sup>2</sup></b> .	Résolution MNT : 1 km Couches vectorielle et raster <b><math>Dd_{moy} = 0,03 \text{ km/km}^2</math></b> <b><math>Sd = 0,05 \text{ km/km}^2</math></b>
Source	<a href="http://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-hydro1k">www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-hydro1k</a>	
<b>HydroSHEDS</b> (Lehner <i>et al.</i> , 2008 ; Lehner et Grill, 2013)	Les données MNT de la base HydroSHEDS sont dérivées principalement de données brutes SRTM (Farr et Kobrick, 2000). HydroSHEDS est l'une des bases de données les plus avancées actuellement pour les applications hydrologiques. Elle fournit des MNT à différentes résolutions et conditionnés pour des applications hydrologiques. Elle propose aussi un réseau hydrographique extrait à partir d'un MNT avec une $A_{min}$ <b>constante de 25 km<sup>2</sup></b> . Elle couvre toutes les parties du globe à l'exception des terres dont la latitude est supérieure à 60°N.	Résolution MNT : 500 m Couche vectorielle <b><math>Dd_{moy} = 0,18 \text{ km/km}^2</math></b> <b><math>Sd = 0,10 \text{ km/km}^2</math></b>
	Récemment, une mise à jour de la base HydroSHEDS appelée HydroRIVERS est disponible à l'échelle globale, où les cours d'eau ont été définis avec un seuil $A_{min}$ <b>constant de 10 km<sup>2</sup> ou par endroits avec un seuil de débit moyen à long terme dépassant 0,1 m<sup>3</sup>/s</b> . Les données de débits à long terme (1971-2000) proviennent d'une descente d'échelle des données de débits à 0,5 degré issus du modèle WaterGAP (Döll <i>et al.</i> , 2003).	Résolution MNT : 500 m Couche vectorielle <b><math>Dd_{moy} = 0,24 \text{ km/km}^2</math></b> <b><math>Sd = 0,10 \text{ km/km}^2</math></b>
Source	<a href="http://www.hydrosheds.org/">www.hydrosheds.org/</a>	

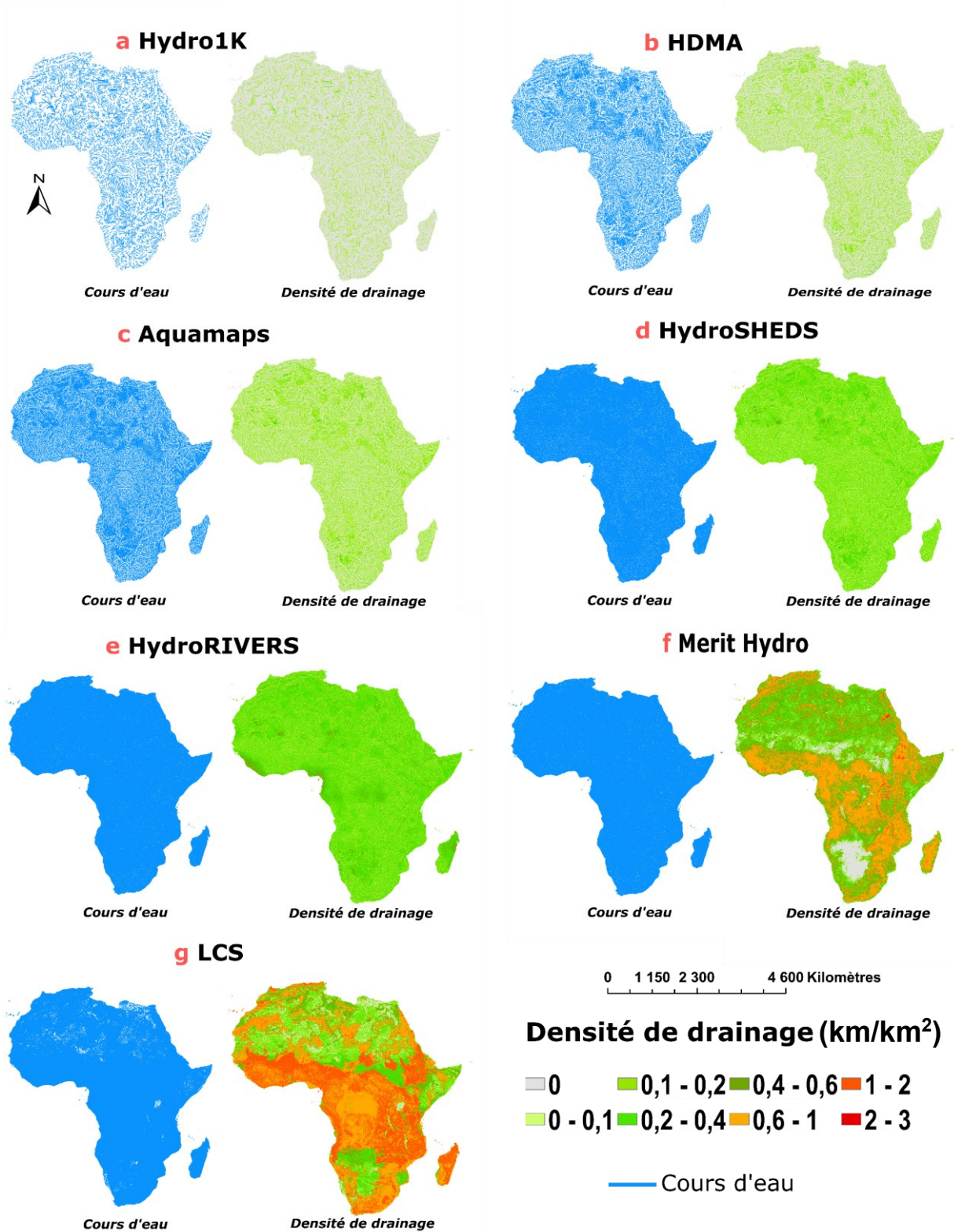


<p><b>FAO (AQUAMAP)</b> (FAO, 2014)</p>	<p>Elle propose un réseau hydrographique extrait à partir d'un MNT d'HydroSHEDS avec une <b>A<sub>min</sub> constante de 100 km<sup>2</sup></b> et fournit en plus des informations sur les cours d'eau intermittents et permanents.</p>	<p>Résolution MNT : 500 m Couche vectorielle <b>Dd<sub>moy</sub> = 0,08 km/km<sup>2</sup></b> <b>Sd = 0,08 km/km<sup>2</sup></b></p>
<p>Source</p>	<p><a href="http://www.fao.org/nr/water/aquamaps/">www.fao.org/nr/water/aquamaps/</a></p>	
<p><b>Hydrologic Derivatives for Modeling and Analysis (HDMA)</b> (Verdin Kristine, 2017)</p>	<p>À la suite de la base HYDRO1K développée par l'USGS, la base HDMA fournit divers types de données allant de la superficie des bassins à l'hydrographie. Le modèle numérique de base résulte de l'union de plusieurs autres bases notamment Hydrosheds pour les zones &lt; 60°N et GMTED 2010 (Danielson et Gesch, 2011) pour les zones &gt; 60°N. Le réseau hydrographique est extrait avec une aire contributive <b>A<sub>min</sub> constante minimum de 250 km<sup>2</sup></b>.</p>	<p>Résolution MNT : 90m Couche vectorielle <b>Dd<sub>moy</sub> = 0,06 km/km<sup>2</sup></b> <b>Sd = 0,08 km/km<sup>2</sup></b></p>
<p>Source</p>	<p><a href="https://www.sciencebase.gov/catalog/item/5925ce50e4b0b7ff9fb3cb0a">https://www.sciencebase.gov/catalog/item/5925ce50e4b0b7ff9fb3cb0a</a></p>	
<p><b>LCS (Lithology, Climate and Slope)</b> (Schneider <i>et al.</i>, 2017)</p>	<p>LCS est une méthode statistique développée par Schneider <i>et al.</i> (2017) à partir d'un MNT issu de la base HydroSHEDS. Les auteurs ont calibré des aires contributives minimales pour correspondre localement à des densités de drainage observées (à partir des réseaux hydrographiques nationaux de référence de la France et de l'Australie) en distinguant plusieurs classes de précipitation, de pente et de lithologie avant d'extraire un réseau à l'échelle globale du nom de « LCS ». Le réseau hydrographique est donc extrait à partir <b>d'aires contributives variables spatialement (A<sub>min</sub>)</b> et fonction de la lithologie, du climat et de la pente du milieu.</p>	<p>Résolution MNT : 500 m Couche vectorielle et raster <b>Dd<sub>moy</sub> = 0,69 km/km<sup>2</sup></b> <b>Sd = 0,38 km/km<sup>2</sup></b></p>
<p>Source</p>	<p><a href="http://www.metis.upmc.fr/en/node/211">www.metis.upmc.fr/en/node/211</a></p>	
<p><b>MERIT Hydro-Vector</b> (Lin <i>et al.</i>, 2021)</p>	<p>MERIT Hydro est une nouvelle base de données hydrographique globale développée par Lin <i>et al.</i> (2021) à partir du MNT MERIT (<i>Multi-Error-Removed Improved Terrain</i>) à 90 m (Yamazaki <i>et al.</i>, 2017). Elle propose une représentation des cours d'eau en tenant compte de la variabilité spatiale de leur densité de drainage. Les cours d'eau sont initialement extraits à l'échelle globale à partir d'une aire contributive <b>A<sub>min</sub> constante de 1 km<sup>2</sup></b>. Ils sont par la suite raccourcis (d'amont vers l'aval) sur chacun des 156 571 bassins versants considérés à la surface du globe, jusqu'à ce que la densité de drainage résultante corresponde à une <b>densité de drainage</b> estimée (par machine learning), à partir d'un entraînement aux USA (dans des 'Hydrologic Unit Codes (HUC-10)') d'une surface moyenne de 470 km<sup>2</sup>.</p>	<p>Résolution MNT : 90 m Couche vectorielle <b>Dd<sub>moy</sub> = 0,5 km/km<sup>2</sup></b> <b>Sd = 0,27 km/km<sup>2</sup></b></p>
<p>Source</p>	<p><a href="https://doi.org/10.6084/m9.figshare.c.5052635">https://doi.org/10.6084/m9.figshare.c.5052635</a></p>	

La représentation des cours d'eau et des densités de drainage (Dd) résultant de ces bases de données hydrographiques en Afrique (**Figure I-1, Figure I-2**) montre des densités de drainage plus variables pour les réseaux Merit Hydro et LCS par rapport aux autres réseaux, ce qui est lié à la prise en compte d'une aire contributive variable spatialement.



**Figure I-1:** Distribution des densités de drainage ( $\text{km}/\text{km}^2$ ) calculées dans les mailles de 5' (environ  $8\text{km} \times 8\text{km}$ ) par base de données collectées. Dans chaque boîte à moustache, le trait horizontal inférieur est le premier quartile qui représente la valeur correspondant à 25 % de la distribution des données tandis que les valeurs moyennes sont matérialisées par un carré au centre de chaque boîte à moustache, le second trait horizontal est la médiane qui est le deuxième quartile et qui représente la valeur correspondant à 50 % de la distribution des données. Le trait horizontal supérieur est le troisième quartile et représente la valeur correspondant à 75 % de la distribution des données, enfin les valeurs extrêmes sont matérialisées par des points.



**Figure I-2:** Représentation en Afrique, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) des bases de données. Les cartes sont affichées dans la projection WGS84, mais les différents calculs sont faits dans la projection Africa Albers Equal Area conic (code ESRI : 102022).

En Afrique, seules quelques bases de données hydrographiques et études proposent une représentation du régime d'intermittence des cours d'eau. Nous résumons dans les lignes suivantes, les sources et les approches utilisées et représentons en **Figure I-3** la distribution spatiale des cours d'eau intermittents / permanents issue de ces bases et études.

- **AQUAMAPS (FAO, 2014) :**

Le réseau AWRD<sup>1</sup> contient une classification du régime d'écoulement (intermittent, permanent) des cours d'eau en Afrique développée par la FAO dans les années 1990. Pour classer les cours d'eau de la base AQUAMAPS, une calibration du régime d'intermittence de ce réseau AWRD est réalisée en analysant pour chaque ordre de Strahler, la distribution fréquentielle des classes intermittentes et permanentes en fonction de l'indice d'aridité (défini comme le rapport entre la précipitation moyenne annuelle et l'évapotranspiration potentielle moyenne annuelle). Ainsi, pour chaque ordre de Strahler, des seuils d'indice d'aridité calibré permettent de définir si le cours est permanent ou intermittent (**Tableau I-2**).

**Tableau I-2 :** Classification du régime d'intermittence basée sur l'ordre de Strahler et l'indice d'aridité (FAO, 2014). Le terme P pour « Permanent » et le terme I pour « Intermittent ». On peut lire par exemple que tous les cours d'eau d'ordre de Strahler 1 ayant un indice d'aridité inférieur à 0,35 sont classés intermittents et inversement pour les cours d'eau permanents.

Indice d'aridité	Ordre de Strahler								
	1	2	3	4	5	6	7	8	9
0,35	P	P	P	P	P	P	P	P	P
0,3	I	P	P	P	P	P	P	P	P
0,24	I	I	P	P	P	P	P	P	P
0,17	I	I	I	P	P	P	P	P	P
0,09	I	I	I	I	P	P	P	P	P
0	I	I	I	I	I	P	P	P	P

La classification développée<sup>2</sup> a ensuite été appliquée pour classer les cours d'eau du réseau hydrographique AQUAMAPS en Afrique, en utilisant la carte globale de l'aridité<sup>3</sup> de la FAO. Par la suite, cette méthode a été généralisée à l'échelle globale pour classer le régime d'intermittence des cours d'eau. Cette classification de l'intermittence des cours d'eau est donc uniquement basée sur le climat et la taille des cours d'eau. Dans cette base, la proportion estimée des cours d'eau intermittents à l'échelle globale est de 34% selon le tableau 1 de Schneider *et al.* (2017).

<sup>1</sup> <http://www.fao.org/3/a0907e/a0907e00.htm>

<sup>2</sup> <https://storage.googleapis.com/fao-maps-catalog-data/geonetwork/aquamaps/AquamapsRiverDatadescription.pdf>

<sup>3</sup> <https://data.apps.fao.org/map/catalog/static/search?keyword=aridity>

- **LCS (Schneider et al., 2017):**

L'étude propose une méthode de premier ordre pour l'identification des cours d'eau intermittents en se basant sur le réseau LCS qui tient compte de la densité de drainage variable des cours d'eau. Cette méthode considère les rivières qui sont localisées sous un certain seuil de précipitation comme intermittentes. Celles qui sont localisées au-delà de ce seuil sont classées comme permanentes ou pérennes. Le calibrage de ce seuil est fait en Australie à partir du réseau hydrographique de la FAO (AQUAMAPS) dans lequel 69% des rivières de l'Australie sont classées intermittentes. Le seuil de précipitation retenu est celui dont les isohyètes chevauchent le mieux les limites des réseaux intermittents de la FAO. Le seuil calibré est de 500 mm/an. L'extrapolation de cette calibration permet d'estimer à au moins 29% la proportion des cours d'eau intermittents à l'échelle globale.

- **Merit Hydro (Lin et al., 2021) :**

Cette étude se focalise sur une estimation des cours d'eau à l'échelle globale en tenant compte de leur densité de drainage variable. Les auteurs proposent une estimation de la proportion des cours d'eau intermittents ( $f_i$ ) par bassin versant à partir de ce réseau. La méthode utilisée se base sur un apprentissage par "Machine learning (*Gradient boosting tree*, GBT)" des informations sur le régime d'intermittence présent dans la base hydrographique de référence des États-Unis (NHDplusV2)<sup>4</sup> avant d'être extrapolée à l'échelle globale. Les variables explicatives de l'intermittence des cours d'eau prises en compte sont les mêmes que celles utilisées pour estimer la densité de drainage des cours d'eau dans l'étude et sont essentiellement : indice d'aridité, altitude, ruissellement moyen, profondeur de la nappe, pourcentage de sable/limon/argile, indice foliaire de végétation (LAI), fraction d'urbanisation, conductivité hydraulique et porosité. Cette estimation de l'intermittence, comme le précisent les auteurs, est très incertaine, du fait de l'incertitude sur les classes d'intermittences présente dans le jeu de donnée initiale (NHDplusV2) et car elle n'a pas été validée ailleurs (en dehors des États-Unis) en raison d'un manque de données de comparaison. De ce fait, les données sur l'intermittence ne sont pas publiées avec le jeu de donnée sur les cours d'eau. À travers une demande adressée aux auteurs, nous avons pu nous procurer ces estimations qui sont représentées à la **Figure I-3.e**.

---

<sup>4</sup> <https://nhdplus.com/NHDPlus/>

- **GIRES (Messenger *et al.*, 2021) :**

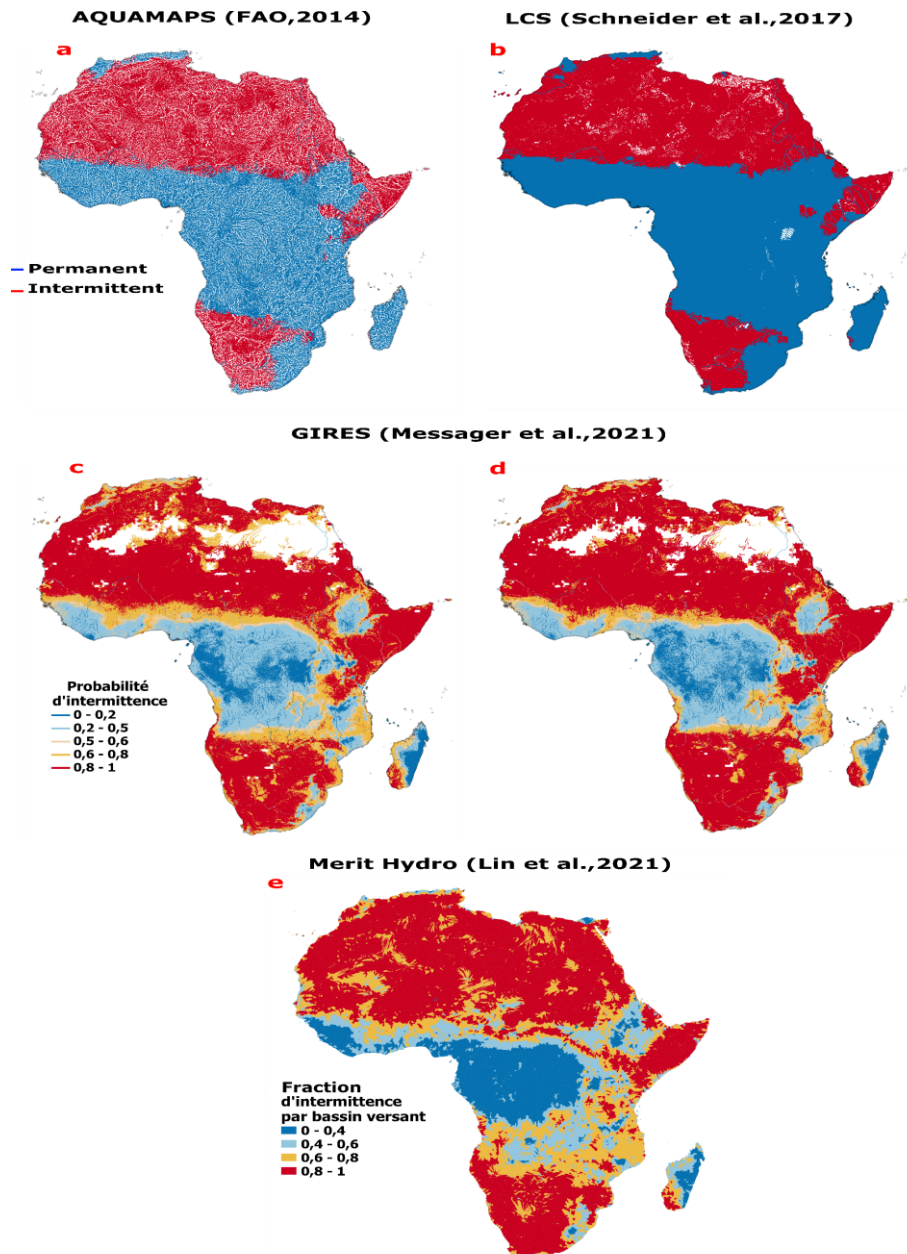
GIRES (*Global Intermittent Rivers and Ephemeral Streams*) est une nouvelle base de données globale qui caractérise la distribution naturelle (sans influence anthropique) des cours d'eau intermittents et permanents. Cette analyse s'appuie sur des stations hydrométriques non influencées et constitue une amélioration notable par rapport aux analyses précédentes en fournissant deux classifications de l'intermittence (i) tous les cours d'eau qui ont plus d'un jour sans écoulement par an en moyenne sur au moins 10 ans et inversement pour les cours d'eau permanents ; (ii) tous les cours d'eau qui ont plus d'un mois (30 jours) sans écoulement par an en moyenne sur au moins 10 ans et inversement pour les cours d'eau permanents. À l'aide des stations de jaugeages collectées (5615 avec une chronique moyenne de plus de 34 ans) dans diverses bases de données à travers le globe, et en s'appuyant sur diverses variables climatiques, géologiques, hydrologiques et d'occupation des sols actuellement disponibles à l'échelle mondiale, les auteurs ont développé des modèles de forêt aléatoire (*Random Forest*) pour prédire la distribution globale des rivières intermittentes sur le réseau hydrographique HydroRIVERS qui est extrait à partir d'un MNT de 500 m de résolution issue de la base HydroSHEDS (Lehner *et al.*, 2008), avec une  $A_{\min}$  constante de 10 km<sup>2</sup> ou par endroits avec un seuil de débit moyen à long terme dépassant 0,1 m<sup>3</sup>/s. Les prédictions du modèle ont été comparées à la classification de l'intermittence présente dans les réseaux hydrographiques de référence de quatre pays (États-Unis, Brésil, Argentine et Australie) et à l'intermittence modélisée en France par une étude précédente (Snelder *et al.*, 2013), en sélectionnant uniquement les cours d'eau avec des seuils d'initiation qui correspondent à celui du réseau HydroRIVERS. Dans la base de référence nationale aux États-Unis, la proportion en longueur des cours d'eau intermittents est estimée à 19% tandis que le modèle prédit plus de cours d'eau intermittents (51% et 36% respectivement pour la classification de l'intermittence basée sur le débit nul  $\geq 1$  jour en moyenne par an et débit nul  $\geq 1$  mois en moyenne par an). Les auteurs attribuent cette surestimation à une tendance du modèle à surestimer l'intermittence dans l'est des États-Unis et une sous-estimation de l'intermittence dans les cours d'eau de taille moyenne à grand dans le réseau hydrographique de référence. En Australie, la proportion en longueur des cours d'eau intermittents est estimée à 91% dans la base de référence tandis que le modèle prédit 95% (débit nul  $\geq 1$  jour en moyenne par an) et 95% (débit nul  $\geq 1$  mois en moyenne par an). En France, la proportion en longueur des cours d'eau intermittents est estimée à 17% dans l'étude de Snelder *et al.* (2013) en tenant compte des cours d'eau correspondants aux seuils d'initiation du réseau HydroRIVERS, tandis que le

modèle prédit 14 % des cours d'eau comme intermittents. Les auteurs attribuent cette différence à la définition de l'intermittence qui est différente dans les deux études. En Argentine et au Brésil, les réseaux hydrographiques de référence n'ont pas d'informations sur la surface amont drainée ni sur le débit moyen à long terme des cours d'eau ce qui ne permet pas une comparaison quantitative avec les prédictions faites sur le réseau HydroRIVERS. Néanmoins, les auteurs montrent que la prédiction de l'intermittence dans ces pays coïncide avec la configuration spatiale de l'intermittence présente dans les réseaux hydrographiques de référence de ces pays.

Ces comparaisons sont faites dans des pays qui sont bien couverts par les stations de jaugeage tandis qu'en Afrique, aucune donnée de validation n'est prise en compte et un certain nombre de données hydrométriques et facteurs anthropiques sont occultés par cette étude. De la **Figure I-3.a** et **Figure I-3.b** résultant respectivement des estimations de la (FAO, 2014) et de la base LCS (Schneider *et al.*, 2017), on voit que les configurations spatiales de l'intermittence sont très similaires pour les deux bases qui lient essentiellement l'intermittence en fonction du climat. On remarque que les **Figure I-3.c** et **Figure I-3.d** de la base GIRES qui sont estimées à partir de pas de temps différents (journaliers et mensuel) sont presque identiques. Quelques inconsistances s'observent sur la **Figure I-3.e** avec quelques bassins estimés avec une proportion d'intermittence  $< 60\%$  dans la partie désertique de l'Afrique (Sahara).

Une définition et quantification des cours d'eau intermittents en Afrique (**Tableau I-3**), à partir des cartes de la **Figure I-3** sont proposées. Les estimations de Lin *et al.* (2021) étant des proportions d'intermittence moyennes dans des bassins versants, nous les avons rastérisés à la résolution de 5' pour avoir des proportions d'intermittence dans des mailles de 5'. Dans l'étude de Messenger *et al.* (2021), tous les cours d'eau prédits avec une probabilité d'être intermittente  $\geq 0,5$  ont été classés comme étant intermittents (et permanents sinon), comme préconisés par les auteurs. Pour les études de la FAO, (2014), Messenger *et al.* (2021) et Schneider *et al.* (2017), des pourcentages d'intermittence sont calculés par mailles de 5' (environ 8 km x 8 km) en faisant la somme des longueurs des cours d'eau classés intermittents sur la longueur totale des cours d'eau dans la maille. Les pourcentages moyens d'intermittence en Afrique résultent de la moyenne des pourcentages d'intermittence dans les mailles de 5' à l'échelle de l'Afrique. Les pourcentages moyens d'intermittence sont plus importants dans les études de Lin *et al.* (2021) et Messenger *et al.* (2021) par rapport aux études de la FAO (2014) et Schneider *et al.* (2017) qui lient essentiellement l'intermittence au

climat. On constate également que les définitions de l'intermittence en fonction des études sont très différentes et certaines études ne définissent pas clairement l'intermittence (FAO, 2014 ; Schneider *et al.*, 2017). La problématique de la définition de l'intermittence peut être une des sources de cette différence dans les pourcentages moyens d'intermittence calculés en Afrique et constitue un écueil dans la comparaison des études sur l'intermittence.



**Figure I-3** : Distribution spatiale de l'intermittence estimée en Afrique par les bases de données et études **a)** AQUAMAPS (FAO, 2014), **b)** LCS (Schneider et al., 2017), **c)** GIRES (Messenger et al., 2021) estimation mensuelle avec les cours d'eau intermittents définis comme des cours d'eau s'arrêtant de s'écouler en moyenne un mois par an et inversement pour ceux définis permanents, **d)** GIRES (Messenger et al., 2021) estimation journalière avec les cours d'eau intermittents définis comme des cours d'eau s'arrêtant de s'écouler en moyenne un jour par an et inversement pour ceux définis permanents, **e)** Merit-hydro (Lin et al., 2021) proportion des cours d'eau intermittents (fi) par bassin versant.



**Tableau I-3 :** Définition et quantification de l'intermittence dans quelques études qui proposent une représentation du régime d'intermittence des cours d'eau en Afrique.

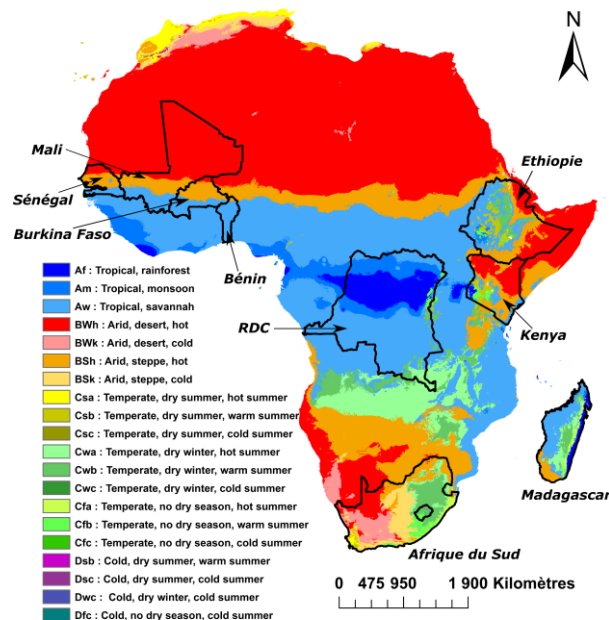
Études	Base de données hydrographiques	Densité de drainage moyenne (km/km <sup>2</sup> )	Définition de l'intermittence	Pourcentage moyen d'intermittence (%)
(FAO, 2014)	AQUAMAPS	0,08	Aucune définition de l'intermittence n'est proposée explicitement, ce qui laisse supposer que les auteurs considèrent les cours d'eau intermittents comme des cours d'eau qui s'arrêtent de s'écouler à un moment donné dans l'année, mais sans plus de précision sur le nombre de jour ou mois à début nul à partir duquel le cours d'eau est considéré intermittent.	52
(Schneider <i>et al.</i> , 2017)	LCS	0,69		48
GIRES-Classe 1 (Messenger <i>et al.</i> , 2021)	HydroSHEDS/ HydroRIVERS	0,24	Les cours d'eau intermittents sont définis comme des cours d'eau s'arrêtant de s'écouler en moyenne au moins un jour par an et inversement pour ceux définis permanents.	76
GIRES-Classe 2 (Messenger <i>et al.</i> , 2021)	HydroSHEDS/ HydroRIVERS	0,24	Les cours d'eau intermittents sont définis comme des cours d'eau s'arrêtant de s'écouler en moyenne au moins un mois par an et inversement pour ceux définis permanents.	74
(Lin <i>et al.</i> , 2021)	Merit-Hydro	0,5	L'étude de Lin <i>et al.</i> (2021) se base sur une calibration des informations d'intermittence présentes dans la base NHDplusV2. Dans la base NHDplusV2, les cours d'eau intermittents sont définis comme des cours d'eau qui ne s'écoulent qu'une partie de l'année, avec une durée de l'écoulement qui est supérieure à la saison des pluies et la fonte des neiges tandis que les cours d'eau éphémères ne contiennent de l'eau qu'après la pluie ou la fonte des neiges (Jaeger <i>et al.</i> , 2021). Cette définition ne donne pas plus de précision sur le nombre de jour ou mois à début nul à partir duquel le cours d'eau est considéré intermittent. Les cours d'eau intermittents considérés donc dans l'étude de Lin <i>et al.</i> (2021) sont une combinaison des cours d'eau intermittents et éphémères identifiés dans la base NHDplusV2.	73

De fortes incertitudes existent donc sur l'ensemble de ces estimations qui n'ont pas été validées en Afrique, mais elles fournissent une information de premier ordre de

l'intermittence à l'échelle globale et donc à l'échelle de l'Afrique. Bien que la période, le jeu de données investigué et la définition de l'intermittence ne soient pas le même, ces informations peuvent être utilisées comme éléments de comparaison avec de futures estimations notamment celles qui seront faites dans cette thèse au **chapitre III** et qui prendront en compte des données de validation (réseau hydrographique de référence) présentes dans quelques pays en Afrique en plus d'un lot additionnel de stations hydrométriques (Base SIEREM) et un tri dans les variables environnementales pour limiter l'autocorrélation.

## I - 1 - 2 Présentation des données hydrographiques de référence collectées en Afrique

Quelques pays d'Afrique ont été ciblés pour la collecte des données hydrographiques de référence auprès des instituts géographiques<sup>5</sup>. Il s'agit du Burkina Faso, du Sénégal, du Mali, du Bénin, de la République Démocratique du Congo (RDC), de l'Éthiopie, du Kenya, de l'Afrique du Sud et de Madagascar. Le choix de ces pays s'est fait sur la base de la diversité de climats<sup>6</sup> qu'ils présentent (**Figure I-4**).



**Figure I-4** : Carte de la répartition des types de climats (période 1980-2016) selon la classification de Köppen-Geiger (Beck et al., 2018 ; Peel et al., 2007). Les contours des pays initialement ciblés pour la collecte des données hydrographiques sont matérialisés en noir.

<sup>5</sup> <https://www.fao.org/3/w7238f/w7238f06.htm>

<sup>6</sup> <http://eduterre.ens-lyon.fr/thematiques/climat/climats-de-la-terre>

Sur l'ensemble de l'Afrique, ces pays représentent plus de 90% de la diversité de climats selon la classification du climat de Peel *et al.* (2007). Malgré des demandes répétées auprès de certains de ces instituts géographiques, nous n'avons pas eu de réponse de tous. Les données obtenues du Kenya ont été écartées à cause d'une absence trop importante de cours d'eau et d'un manque de description. Les données finalement retenues ont été acquises auprès des instituts géographiques du Burkina Faso (IGB), du Mali (IGM), du Bénin (IGN), de Madagascar (FTM) et de l'Afrique du Sud (DWS). La description ci-dessous de ces bases est faite en fonction des informations présentes dans les fiches descriptives des données et d'informations obtenues auprès de l'IGN-FI et des instituts géographiques nationaux.

- **Réseau hydrographique du Burkina Faso :**

La Base Nationale de Données Topographiques (BNDT) version 1.0 se présente comme une base de données de SIG (Système d'Information Géographique) distribuée par l'Institut Géographique du Burkina (IGB) dont les études ont été menées entre 2011 et 2015. Cette base de données a été produite par l'Institut Géographique du Burkina Faso (IGB) en collaboration avec l'Institut Géographique National France Internationale (IGN-FI) lors du projet<sup>7</sup> « *PC200K Burkina Faso* » de production cartographique à l'échelle de 1 : 200 000 couvrant l'ensemble du pays. Ce jeu de données regroupe 10 couches thématiques dont l'hydrographie, et les cours d'eau identifiés sont ceux principalement digitalisés d'images satellitaires (notamment RapidEye<sup>8</sup> à 5 mètres de résolution). Les cours d'eau de largeur supérieure à 50 mètres sont numérisés sous la forme de surfaces d'eau (polygones). Pour les lacs et les réservoirs uniquement, les contours ont été dessinés sur des images satellites prises en saison sèche donc avec des niveaux variables (niveau haut en octobre vs niveau très bas en mai). Les cours d'eau seraient classés permanents / intermittents en fonction de ce qui est visible sur l'image satellitaire, la connaissance du terrain (tous les cours d'eau sont à priori considérés intermittents et des investigations sont menées uniquement sur les grands cours d'eau du pays) et d'informations fournies par la population locale et la DGRE. Une représentation de l'intermittence est proposée en **Figure I-5.a** à l'échelle du pays.

---

<sup>7</sup> <https://www.ignfi.fr/fr/portfolio-item/cartographie-burkina-faso/>

<sup>8</sup> <https://fr.wikipedia.org/wiki/RapidEye>

- **Réseau hydrographique du Mali :**

Le réseau hydrographique de référence du Mali est distribué par l'Institut Géographique du Mali (IGM) à l'échelle de 1 : 200 000. Les études<sup>9</sup> pour établir cette base de données ont été menées entre 2012 et 2016 en collaboration avec l'Institut Géographique National France Internationale (IGN-FI). Une part essentielle des cours d'eau a été récupérée à partir des anciennes cartes IGN des années 1960 qui avaient été produites à partir de prises de vues aériennes en stéréo à l'échelle du 1 : 40 000. Les données d'images satellitaires exploitées sont celles de SPOT 6 avec 1,5 m de résolution<sup>10</sup> (autour de l'année 2013). Comme au Burkina Faso, l'intermittence est renseignée en fonction de ce qui est visible sur les images satellitaires. Une représentation de l'intermittence est proposée en **Figure I-5.b** à l'échelle du pays.

- **Réseau hydrographique du Bénin :**

Le réseau hydrographique de référence du Bénin est distribué par l'Institut Géographique National (IGN) à l'échelle de 1 : 50 000. Les études ont été menées entre 2014 et 2018 en collaboration avec l'Institut Géographique National France Internationale (IGN-FI) dans le cadre<sup>11</sup> du « *Projet d'Appui à la Préservation et au Développement des Forêts Galeries et Production de Cartographie de base numérique (PAPDFGC)* ». Elle regroupe 10 thèmes dont l'hydrographie avec les cours d'eau principalement digitalisés de prises de vues aériennes à une résolution proche de 40 cm. Les prises de vues aériennes ont été acquises en saison sèche et les cours d'eau ont été saisis sur la base de ce qui était visible sur les photographies aériennes. De même, l'information sur l'intermittence est renseignée aussi en fonction de ce qui est visible sur les images satellitaires. Une représentation de l'intermittence est proposée en **Figure I-5.c** à l'échelle du pays.

- **Réseau hydrographique de Madagascar :**

La BD200 est une base de données géographiques créée lors de la réalisation des cartes d'Inventaire Ecologique et Forestier National (IEFN) de Madagascar. Les objets géographiques de référence qui servent de fond pour les cartes IEFN sont obtenus de la numérisation des cartes à l'échelle 1 : 200 000 de Madagascar, elles-mêmes issues de la généralisation des cartes à 1 : 100 000. Ce jeu de données regroupe 4 couches thématiques,

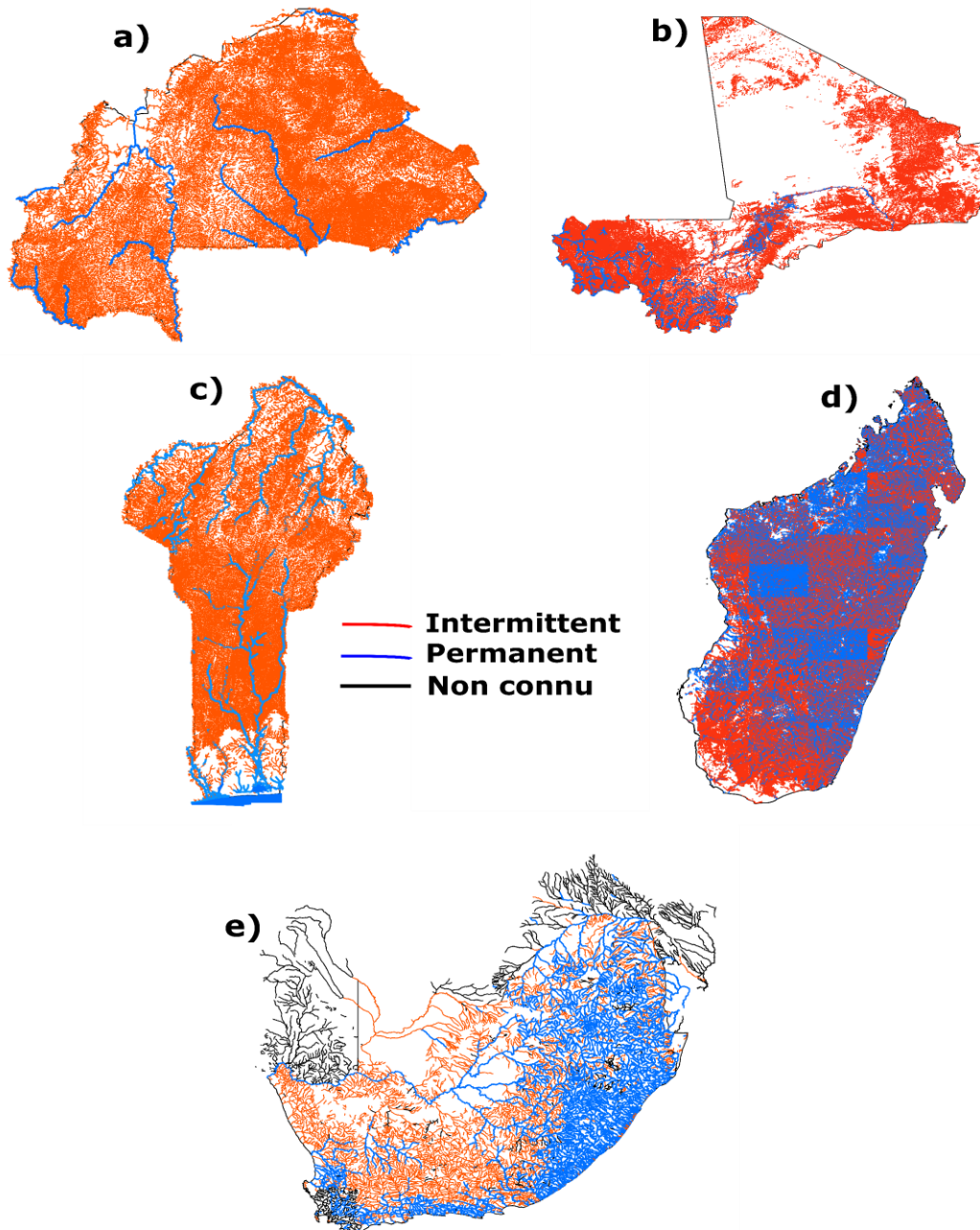
---

<sup>9</sup> <https://www.ignfi.fr/fr/portfolio-item/cartographie-mali/>

<sup>10</sup> [https://fr.wikipedia.org/wiki/SPOT\\_\(satellite\)](https://fr.wikipedia.org/wiki/SPOT_(satellite))

<sup>11</sup> <https://www.ignfi.fr/fr/portfolio-item/cartographie-benin/>

dont l'hydrographie. Ces entités thématiques sont regroupées dans le thème "Occupation du sol", qui est le résultat du traitement d'images satellitaires<sup>12</sup> Landsat 5 à la résolution de 30 m prises entre 1999 et 2000. Aucune information n'a été fournie sur la méthodologie de cartographie des cours d'eau intermittents présents dans cette base de données. Une représentation de l'intermittence est proposée en **Figure I-5.d** à l'échelle du pays.



**Figure I-5** : Représentation spatiale de l'intermittence dans les bases hydrographiques de référence du **a)** Burkina Faso, **b)** Mali, **c)** Bénin, **d)** Madagascar, **e)** Afrique du Sud.

<sup>12</sup> <https://landsat.gsfc.nasa.gov/>

- **Réseau hydrographique de l'Afrique du Sud :**

Un réseau hydrographique de référence de l'Afrique du Sud à l'échelle de 1 : 500 000 du « *Resource Quality Information Services (RQIS)* » est disponible en téléchargement libre sur le site du « *Departement water and sanitation* »<sup>13</sup>. Il est basé à l'origine sur une carte des cours d'eau de 1994 provenant du « *Chief Directorate of National Geo-spatial Information* ». Cette version est mise à disposition "telle quelle" pour la commodité de ceux qui ont besoin d'une couverture hydrographique à moyenne échelle pour l'affichage ou la modélisation. Pour la classification en termes d'intermittence, nous n'avons pas pu collecter des informations sur la méthodologie utilisée. Une représentation de l'intermittence est proposée en **Figure I-5.e** à l'échelle du pays.

De manière générale, dans les fiches descriptives des données, une des méthodologies décrites pour caractériser l'intermittence présente dans ces bases de données de référence est la digitalisation de la présence ou absence d'eau dans les cours d'eau à partir des images satellitaires, mais l'application de cette méthodologie aux petits cours d'eau qui sont généralement majoritaires dans les bases de données reste fastidieuse sur de larges domaines sans oublier les difficultés liées à l'opacité de la végétation et des nuages qui empêchent souvent l'identification des cours d'eau par image satellitaire (Turner et Richter, 2011). Dans ce sens, en milieu aride et semi-aride comme au Burkina Faso par exemple (**Figure I-5.a**), à l'exception des grands cours d'eau qui sont classés permanents, tous les cours d'eau sont par défaut intermittents (Fovet *et al.*, 2021). Ce même constat s'applique au Mali (**Figure I-5.b**) et au Bénin (**Figure I-5.c**) où seuls les grands cours d'eau sont classés permanents, le reste des cours d'eau est par défaut classé intermittent. À Madagascar, on remarque sur la **Figure I-5.d** un certain motif rectangulaire à différents endroits de la carte ce qui est peut-être lié aux dalles d'images satellites analysées. En Afrique du Sud, on constate une dépendance de la configuration spatiale de l'intermittence (**Figure I-5.e**) avec celle du climat (**Figure I-4**).

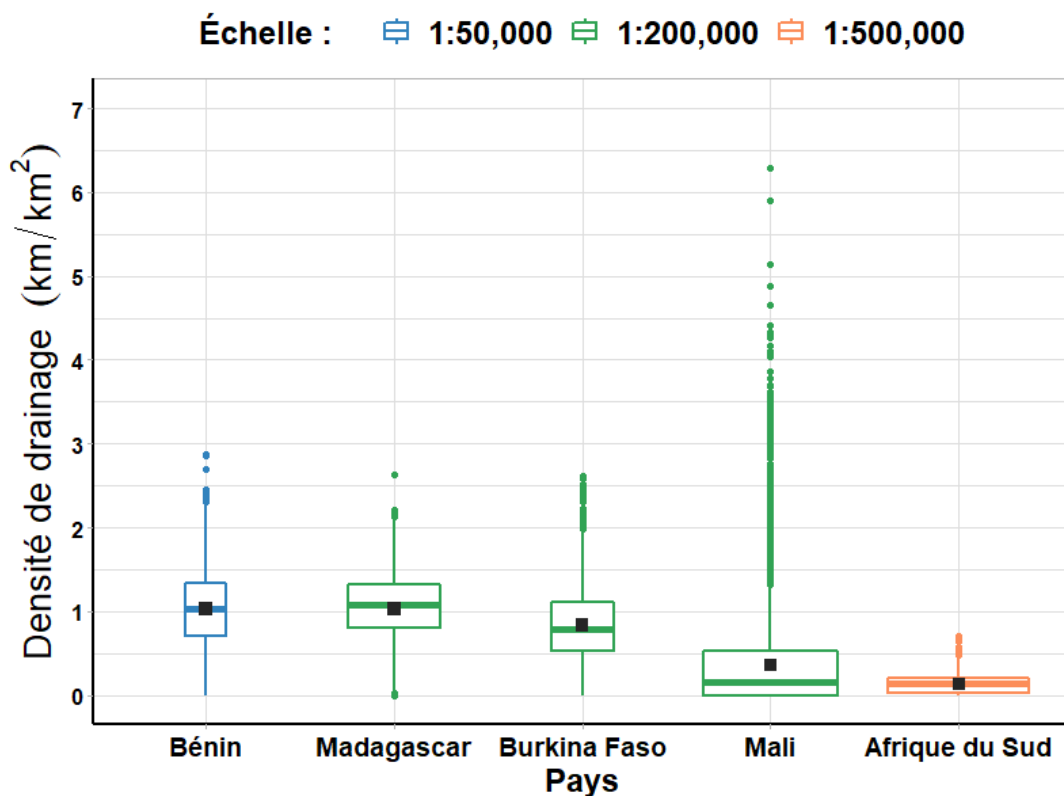
En matière de définitions de l'intermittence des cours d'eau, on peut dire que l'information présente dans les bases de données hydrographiques de référence ainsi que dans les études de la FAO (2014); Lin *et al.* (2021) et Schneider *et al.* (2017) manque de précision sur le nombre de jour ou mois à débit nul à partir duquel le cours d'eau est considéré intermittent. Ce manque d'une précision quantitative dans la définition de l'intermittence est un obstacle à la comparaison des études.

---

<sup>13</sup> [https://www.dws.gov.za/iwqs/gis\\_data/river/rivs500k.aspx](https://www.dws.gov.za/iwqs/gis_data/river/rivs500k.aspx)

Malgré ces limites, les réseaux hydrographiques nationaux de référence constituent la source d'information la plus courante utilisée par la plupart des agences et institutions gouvernementales pour déterminer la distribution spatiale de l'intermittence, et constituent donc des éléments de référence utile pour la validation de la carte d'intermittence à l'échelle de l'Afrique qui sera faite au **chapitre III**.

En termes de densité de drainage, on observe que les plus faibles densités de drainage sont observées au Mali et en Afrique du Sud (**Figure I-6**) avec quelques très fortes densités de drainage dans certaines parties du Mali. Les faibles densités de drainage observées en Afrique du Sud sont certainement liées à l'échelle de la base cartographique qui est plus petite. Il faudra noter ici que plus le dénominateur de l'échelle (1/x) est petit plus le niveau de détail est important. Dans ce sens l'échelle 1 : 200 000 est plus grande que 1 : 500 000 et donc avec un niveau de détail plus important.



**Figure I-6 :** Variabilités de la densité de drainage observées ( $Dd_{obs}$ ) par pays où les données hydrographiques de références ont été collectées. La largeur des boîtes à moustaches varie en fonction du nombre d'observations. Les densités de drainage ont été calculées dans des mailles de 5' (environ 8km x 8km) par base de données collectées.

### I - 1 - 3 Comparaison des données hydrographiques de référence avec celles d'études globales et locales

Dans l'optique de mieux cerner la précision des bases de données globales (**Tableau I-1**) qui représentent les cours d'eau en Afrique (en termes de densité de drainage), une comparaison est faite avec les bases de données hydrographiques de référence issues du Burkina Faso, du Mali, du Bénin et de Madagascar. Le but est d'identifier le réseau hydrographique global qui représente au mieux les cours d'eau tels que fournis par les réseaux de référence et de l'utiliser au final comme support pour la cartographie des cours d'eau intermittents en Afrique.

Les paramètres statistiques utilisés pour comparer la densité de drainage des réseaux sont la moyenne, l'écart-type, la corrélation de Pearson et la RMSE (racine carrée de l'erreur quadratique moyenne) qui se calculent respectivement selon les équations (I-1),(I-2),(I-3),(I-4) suivantes :

$$Moyenne = \frac{1}{n} \sum_{i=1}^n x_i \quad (I-1)$$

$$Ecart\ type = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (I-2)$$

$$Correlation\ de\ Pearson = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (I-3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (I-4)$$

Où  $x_i$  représente la densité de drainage d'un réseau hydrographique donné dans une maille  $i$ ,  $\bar{x}$  la moyenne,  $y_i$  la densité de drainage du réseau de référence dans une maille  $i$ ,  $\bar{y}$  la moyenne des densités de drainage de l'ensemble des mailles et  $n$  le nombre de mailles de l'échantillon.

Il faudra noter que la RMSE prend des valeurs entre 0 et  $+\infty$  et est d'autant plus proche de zéro que les deux réseaux ont des densités de drainage similaires, le coefficient de corrélation



prend des valeurs entre 0 et 1 avec des valeurs qui tendent vers 0 indiquant une absence de corrélation, ici en ce qui concerne la distribution géographique de la densité de drainage.

Une comparaison des densités de drainage des bases de données globales synthétisées dans des mailles de 5' a montré une densité de drainage plus variable avec les réseaux Merit hydro et LCS (**Figure I-1, Figure I-2**). On constate une mauvaise restitution de la variabilité spatiale de la densité de drainage par les bases de données globales par rapport aux réseaux hydrographiques de référence (**Figure I-7, Figure I-8, Figure I-9, Figure I-10**) avec des coefficients de corrélation spatiale qui sont assez faibles dans l'ensemble (**Tableau I-4**). Les corrélations les plus fortes dans chacun des pays (Burkina Faso, Mali, Madagascar et Bénin) sont obtenues avec les réseaux Merit hydro et LCS qui sont extraits avec des aires contributives variables.

Le réseau Merit Hydro et LCS sont respectivement à une échelle approximative de 1 : 100 000 et 1 : 50 000, donc supérieure à celle de la plupart des bases de référence collectées à l'exception du réseau hydrographique de référence du Bénin qui a une échelle de 1 : 50 000. Malgré cela, on observe des performances toujours faibles de ces réseaux par rapport à celui du Bénin, Merit Hydro montre une meilleure variabilité (**Figure I-7.f**) par rapport à LCS (**Figure I-7.g**) en représentant mieux les faibles densités de drainage au sud du pays (**Figure I-7.h**). De façon générale, en termes de distribution spatiale de la densité de drainage, Merit hydro montre une meilleure variabilité que LCS, mais souvent avec des valeurs de densités de drainage mal positionnées comme au Burkina Faso (**Figure I-7**) dans les zones de forte/faible densité de drainage. Une des causes de cette différence avec les réseaux de référence pourrait être liée à l'information géologique prise en compte dans les bases hydrographiques globales (Lin *et al.*, 2021 ; Schneider *et al.*, 2017). Comme illustrée au **chapitre II** et **chapitre III**, l'information géologique est un des facteurs de contrôle important de l'intermittence et de la densité de drainage, mais dont la cartographie à l'échelle globale n'est pas toujours assez détaillée.

Cependant, parmi l'ensemble des réseaux hydrographiques globaux investigués, le réseau LCS est celui qui s'approche le mieux des bases nationales hydrographiques de référence, en termes de distribution spatiale de la densité de drainage (**Figure I-7.h, Figure I-8.h, Figure I-9.h, Figure I-10.h**) et de paramètres statistiques (moyenne, corrélation et RMSE) (**Tableau I-4**). Au regard de ces considérations, le réseau LCS sera utilisé comme support pour la cartographie de l'intermittence des cours d'eau à l'échelle de l'Afrique au **chapitre III**. Les

ordres de Strahler sont fournis comme attribut du réseau LCS. L'ordre de Strahler permet de classer les cours d'eau par types (Strahler, 1957), en fonction de leur nombre d'affluents. La classification se fait d'amont vers l'aval et plus il est élevé, plus le nombre d'affluents est censé être important. Il décrit donc la nature dendritique des cours d'eau et est couramment utilisé pour caractériser la taille des cours d'eau sur de grands domaines (Downing *et al.*, 2012).

Une comparaison de la classification de l'intermittence présente dans les études (FAO, 2014 ; Lin *et al.*, 2021 ; Messenger *et al.*, 2021 ; Perez-Saez *et al.*, 2017 ; Schneider *et al.*, 2017) avec la classification de l'intermittence présente dans les réseaux hydrographiques de référence a été effectuée dans le but de quantifier la part d'intermittence dans chaque étude par rapport aux réseaux de référence et d'apprécier la configuration spatiale de l'intermittence. La méthode utilisée pour calculer le pourcentage moyen d'intermittence dans chacune des études (**Tableau I-5**) est la même que celle utilisée dans le **Tableau I-3** avec les paramètres statistiques (écart-type, coefficient de corrélation, RMSE) qui sont calculés avec les formulations (I-2),(I-3),(I-4).

Les études de la FAO (2014) et de Schneider *et al.* (2017) qui sont basées uniquement sur des relations simplistes entre le climat et l'ordre de Strahler pour classifier l'intermittence des cours d'eau sous-estiment largement l'intermittence (pourcentage moyen d'intermittence très faible) par rapport au réseau de référence à l'exception de l'Afrique du Sud (**Tableau I-5**, **Figure I-11**). Dans l'étude de Perez-Saez *et al.* (2017) au Burkina Faso (**Figure I-11.k**) dont les facteurs de contrôle les plus importants de l'intermittence sont associés au débit moyen annuel, à la pente et à l'indice d'aridité, les pourcentages moyens d'intermittence (75,4%) ne sont pas trop différents de celles de la base de référence (98%). Par ailleurs, l'intermittence prédite par cette étude présente une configuration spatiale qui est très dépendante du gradient nord-sud de la précipitation moyenne annuelle et donc différente de celle que présente la base de référence. Pour les études de Lin *et al.* (2021) et Messenger *et al.* (2021), les proportions d'intermittence ne sont pas trop différentes de celles des bases de référence en général, à l'exception de l'Afrique du Sud où la proportion moyenne d'intermittence calculée est supérieure à celle de la base de référence (**Tableau I-5**). L'un des éléments de cette différence est lié à la définition de l'intermittence qui est différente entre les études comme précédemment évoquées dans les sections précédentes.

**Tableau I-4** : Performances des bases de données par rapport au réseau hydrographique de référence du Bénin, Burkina Faso, Mali, Madagascar. Les paramètres s'appliquent pour les différents réseaux sur les densités de drainage (km/km<sup>2</sup>) synthétisées dans les mailles de 5' (environ 8 km x 8 km). Les valeurs de corrélation et Rmse affichées sont celles entre le réseau de référence et les réseaux comparés.

Pays (Échelle)	Réseau hydrographique	Paramètres statistiques			
		Moyenne (km/km <sup>2</sup> )	Écart-type (km/km <sup>2</sup> )	Corrélation de Pearson (-)	RMSE (km/km <sup>2</sup> )
<b>Bénin</b> (1 : 50 000)	Hydro1K	0,02	0,05	0,16	1,1
	HDMA	0,05	0,08	0,21	1,08
	Aquamaps	0,07	0,08	0,23	1,06
	HydroSHEDS	0,16	0,09	0,23	0,98
	HydroRIVERS	0,21	0,09	0,23	0,94
	Merit Hydro	0,58	0,16	0,33	0,63
	LCS	1,06	0,13	0,25	0,49
	IGN Bénin (référence)	1,03	0,47	1	0
<b>Burkina Faso</b> (1 : 200 000)	Hydro1K	0,02	0,04	0	0,93
	HDMA	0,05	0,07	0,01	0,9
	Aquamaps	0,07	0,07	0,02	0,89
	HydrSHEDS	0,16	0,09	0,01	0,81
	HydroRIVERS	0,20	0,09	0,00	0,78
	Merit Hydro	0,51	0,17	0,02	0,56
	LCS	1,2	0,14	0,12	0,53
	IGB (référence)	0,84	0,43	1	0
<b>Mali</b> (1 : 200 000)	Hydro1K	0,03	0,05	0	0,63
	HDMA	0,06	0,08	0	0,61
	Aquamaps	0,08	0,08	0	0,60
	HydroSHEDS	0,18	0,1	0	0,57
	HydroRIVERS	0,22	0,1	0,10	0,56
	Merit Hydro	0,37	0,20	0,17	0,53
	LCS	0,56	0,41	0,43	0,54
	IGM (référence)	0,36	0,52	1	0
<b>Madagascar</b> (1 : 200 000)	Hydro1K	0,02	0,05	0	1
	HDMA	0,05	0,08	0	1
	Aquamaps	0,08	0,08	0	1
	HydroSHEDS	0,17	0,1	0	0,95
	HydroRIVERS	0,30	0,11	0,22	0,83
	Merit Hydro	0,64	0,24	0,34	0,55
	LCS	0,97	0,19	0,46	0,36
	FTM (référence)	1	0,40	1	0

**Tableau I-5 :** Paramètres statistiques de comparaison de l'intermittence estimée en Afrique par des études par rapport au réseau de référence du Bénin, Burkina Faso, Mali, Madagascar et Afrique du Sud. Les paramètres s'appliquent pour les différents réseaux sur le pourcentage d'intermittence (%) synthétisé dans les mailles de 5' (environ 8 km x 8 km).

Pays (Échelle)	Bases de données / Études	Paramètres statistiques			
		Pourcentage moyen interm. (%)	Écart- type (%)	Coefficient de corrélacion de Pearson (-)	RMSE (%)
<b>Bénin</b> (1 : 50 000)	Aquamaps (FAO, 2014)	0,36	5	0,02	92
	LCS (Schneider <i>et al.</i> , 2017)	0	0	*	94
	Gires 1 (Messenger <i>et al.</i> , 2021)	98	11,6	0,42	16
	Gires 2 (Messenger <i>et al.</i> , 2021)	97	13	0,41	16
	Merit-Hydro (Lin <i>et al.</i> , 2021)	69	12	0,04	31
	IGN Bénin (référence)	93	16	1	0
<b>Burkina Faso</b> (1 : 200 000)	Aquamaps	22	40	0,15	85
	LCS	10,5	30	0,05	93
	Gires-Classe 1	100	0	*	7,6
	Gires-Classe 2	100	0	*	7,6
	(Perez-Saez <i>et al.</i> , 2017)	75,4	40	0,08	47
	Merit-Hydro	85	11	0,05	18
IGB (référence)	98	7,4	1	0	
<b>Mali</b> (1 : 200 000)	Aquamaps	68	46	0,38	50
	LCS	61	48	0,26	57
	Gires-Classe 1	99	8	0,4	15
	Gires-Classe 2	98	10	0,45	14
	Merit-Hydro	89	13	0,34	17
	IGM (référence)	94	15	1	0
<b>Madagascar</b> (1 : 200 000)	Aquamaps	0,75	8	0,1	55,6
	LCS	2	15	0,24	58
	Gires-Classe 1	44	47	0,27	48
	Gires-Classe 2	42	47	0,29	48
	Merit-Hydro	0,55	21	0,35	24
	FTM (référence)	55	22	1	0
<b>Afrique du Sud</b> (1 : 500 000)	Aquamaps	47	49	0,67	39
	LCS	50	49	0,56	46
	Gires-Classe 1	91	26	0,34	58
	Gires-Classe 2	88	30	0,40	55
	Merit-Hydro	85	17	0,53	51
	DWS (référence)	54	47	1	0

\* Impossibilité de calculer le coefficient de corrélacion avec le réseau de référence en raison d'un écart-type nul du pourcentage d'intermittence dans la zone de l'étude considérée ce qui signifie concrètement qu'on a une valeur constante sur la zone de l'étude.

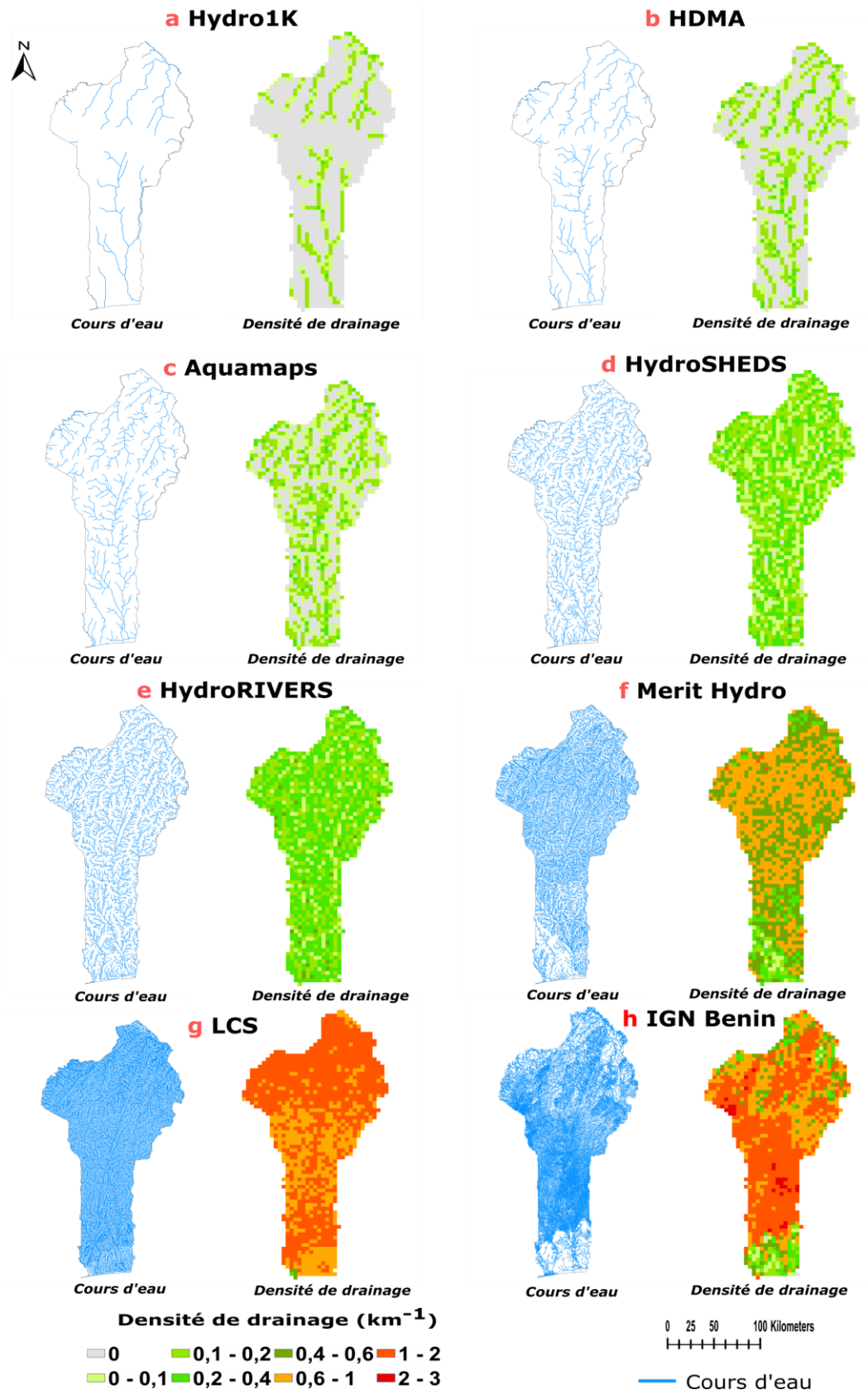
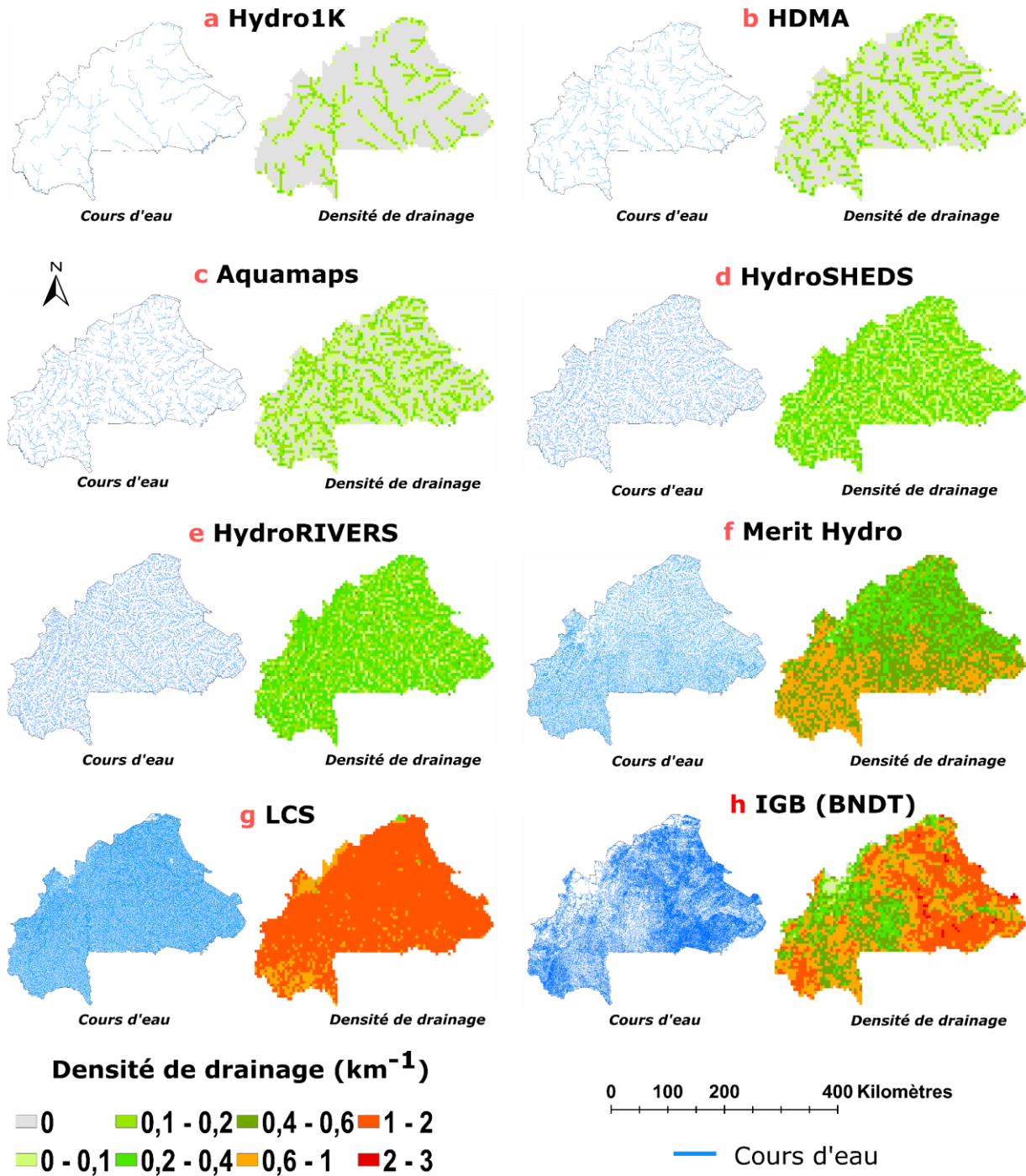


Figure I-7 : Représentations au Bénin, des cours d'eau (à gauche) et densités de drainage (à droite) correspondantes de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Bénin sont faits dans la projection WGS84 UTM zone 31 N (code EPSG : 32631).



**Figure I-8 :** Représentations au Burkina Faso, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Burkina Faso sont faits dans la projection WGS84 UTM zone 30 N (code EPSG : 32630).

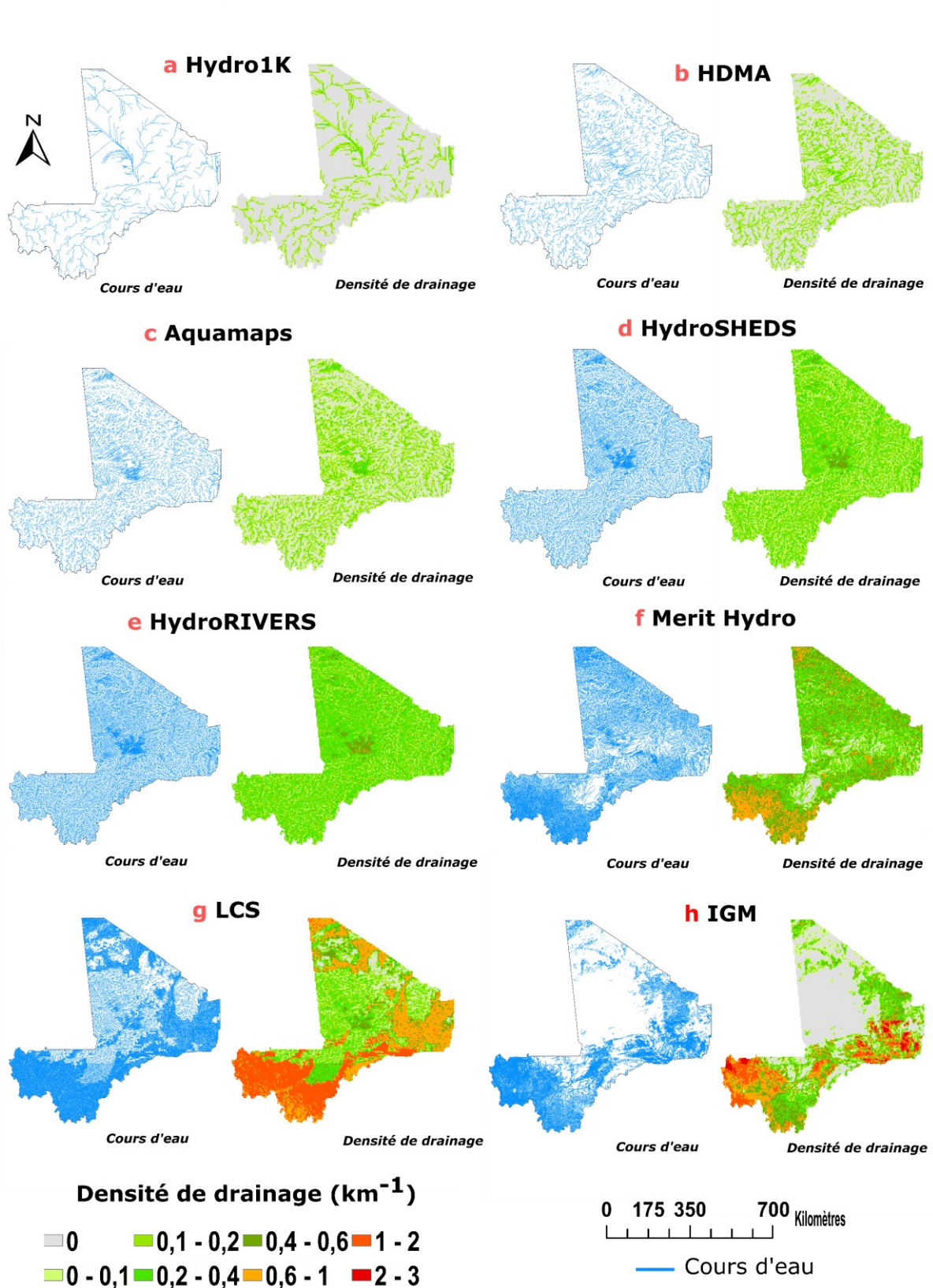
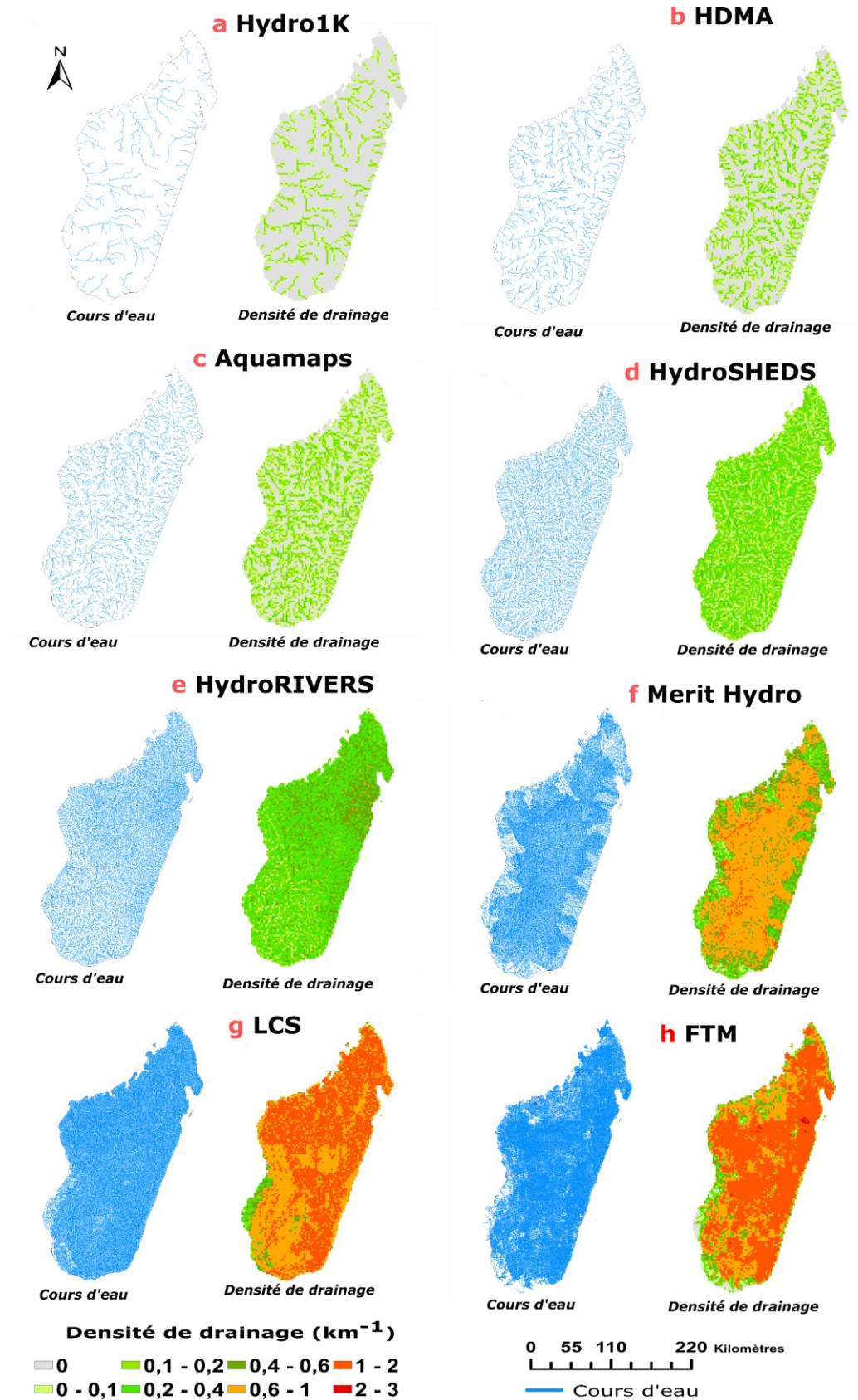
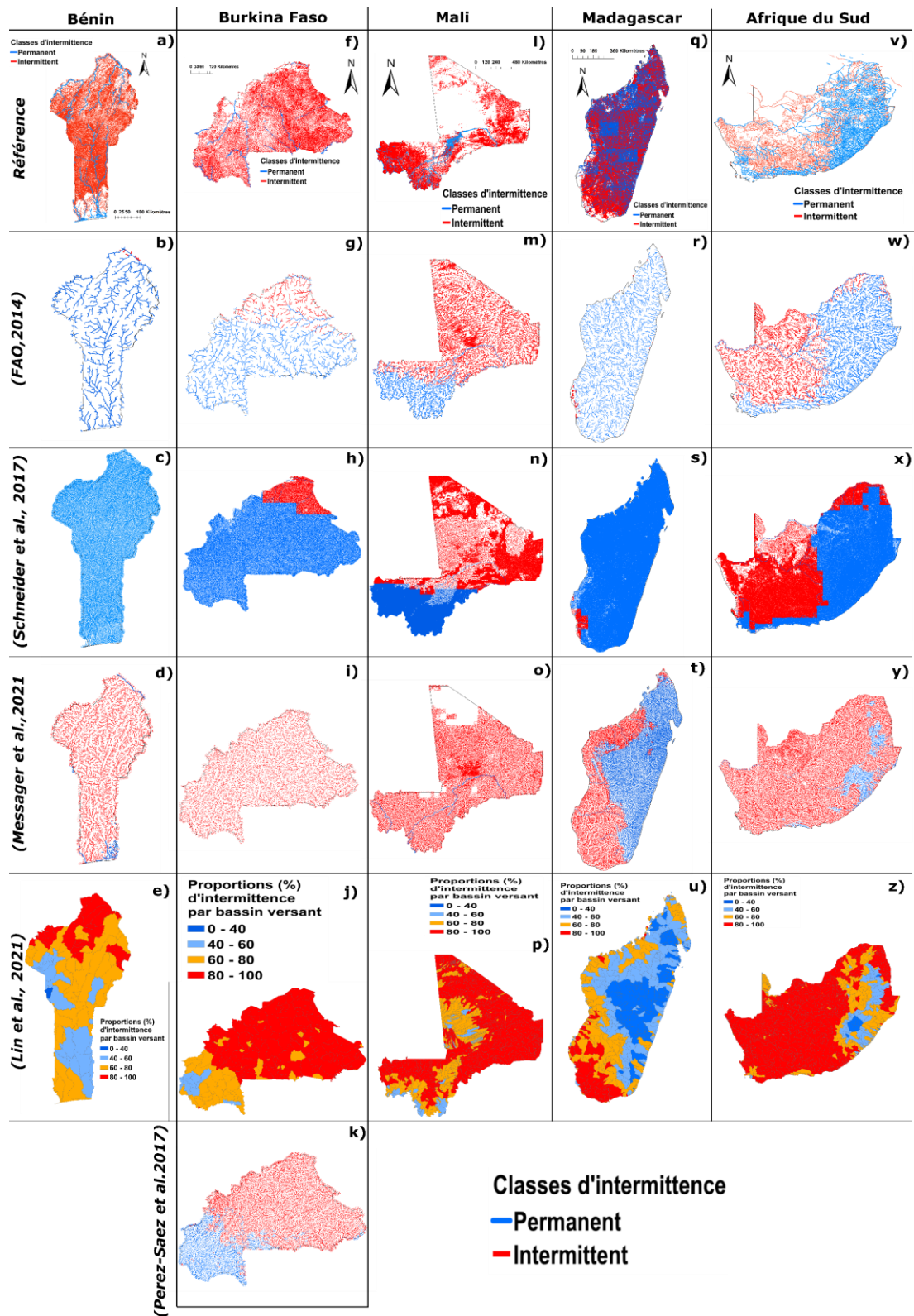


Figure I-9 : Représentations au Mali, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Mali sont faits dans la projection WGS84 UTM zone 30 N (code EPSG : 32630).



**Figure I-10 :** Représentations à Madagascar, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, mais les différents calculs sont faits dans la projection Tananarive / Laborde Grid (code EPSG : 29701).





**Figure I-11** : Comparaison de la distribution spatiale des cartes nationales de l'intermittence du Bénin, Burkina Faso, Mali, Madagascar (Référence) avec celles des études (FAO, 2014 ; Lin et al., 2021 ; Messenger et al., 2021 ; Perez-Saez et al., 2017 ; Schneider et al., 2017). Dans l'étude de Perez-Saez et al. (2017), les cours d'eau classés « permanent, permanent low variability et permanent high variability » et ceux « mid ephemeral, strongly ephemeral » ont été regroupés respectivement dans cette représentation en « cours d'eau permanent » et « cours d'eau intermittent ».

## I - 2 Sources et description des données hydro-environnementales

Un certain nombre de variables hydro-environnementales (informations climatiques, hydrologiques, physiographiques et géologiques) sont investiguées dans cette thèse pour comprendre et caractériser spatialement l'intermittence des cours d'eau. Les variables choisies varient en fonction des chapitres, mais dans chaque chapitre, une description détaillée est fournie sur la fenêtre temporelle utilisée ainsi que l'échelle spatiale à laquelle elle est utilisée. La représentation spatiale de ces variables est proposée au Burkina Faso (section II-7, Figure II-A-10) et à l'échelle de l'Afrique (Annexe - 1).

### I - 2 - 1 Données de débits

#### I- 2 - 1 - 1 Données de débits ponctuels

Les données de débits ponctuels (débits mesurés à un point d'un cours d'eau) utilisées dans cette thèse proviennent de trois bases de données notamment les bases SIEREM, et GRDC (continentales) et DGRE (au Burkina Faso) :

- **Le SIEREM** (*Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation*) est une base de données<sup>14</sup> qui fournit des informations hydro-environnementales collectées à travers des projets ainsi que des résultats scientifiques majeurs résultant de l'exploitation de ces données (Boyer *et al.*, 2006). Ces données sont héritées principalement de l'ORSTOM (Office de la Recherche Scientifique et Technique Outre-Mer), aujourd'hui IRD (Institut de Recherche pour le Développement). Les réseaux de mesures hydrométriques en Afrique ont connu une extension importante depuis les années 1945 avec la création de l'ORSTOM qui a installé de nombreuses stations hydrométriques et en particulier sur des petits bassins expérimentaux. Dans les années 1970, la gestion de ces réseaux de mesures est progressivement prise en main par les pays eux-mêmes à travers leurs services nationaux (Mahé, 2006). À la suite de cette reprise, les mesures sur les petits bassins versants ont été abandonnées par manque de moyens, si bien que la densité spatiale des réseaux hydrométriques s'est effondrée, se limitant maintenant aux bassins

---

<sup>14</sup> <http://www.hydrosciences.fr/SIEREM/>

versants présentant des enjeux économiques plus importants (fort potentiel hydro-électrique, agricole, etc.) (Nka Nnomo, 2016). Les données issues de la base SIEREM sont donc des données à majorité historiques aux pas de temps journaliers et mensuels.

- **Le GRDC** (*Global Runoff Data Center*) est un centre international de collecte de données hydrométriques créé en 1988. Il est implanté à l'institut fédéral Allemand d'hydrologie (BfG) à Koblenz, en Allemagne et il fonctionne sous les auspices de l'OMM (Organisation Mondiale de Météorologie). L'échange de données et d'informations hydrologiques à l'échelle mondiale est la principale raison d'être du GRDC. Il relie ainsi les services hydrologiques et météorologiques nationaux, qui sont les principaux fournisseurs de données et métadonnées sur les débits des cours d'eau, avec la communauté des chercheurs scientifiques qui utilisent cette collection de données unique. Il faut noter que tous les pays ne font pas remonter toutes les informations dont ils disposent au GRDC si bien que les données sur certaines zones sont plutôt historiques. Une plateforme en ligne<sup>15</sup> permet désormais de sélectionner et trier les données recherchées à l'échelle globale sans avoir à transmettre un formulaire de demande de données au gestionnaire de cette base de données. Les données fournies sont au pas de temps mensuels ou journaliers.
- **La DGRE** (*Direction Générale des Ressources en Eau*) du Burkina Faso est l'entité en charge de la collecte et de la gestion des stations hydrométriques sur l'ensemble du territoire burkinabè. Elle dispose d'un certain nombre de stations hydrométriques distribuées à travers le pays (**Figure I-12**). Pour cette étude les données de débits fournies par la DGRE sont les données les plus récentes (1955-2017). Ce jeu de données a été complété au Burkina Faso par les stations non redondantes disponibles dans les bases de données SIEREM et GRDC.

Il y a une certaine réticence des services hydrologiques nationaux à partager ces informations hydrométriques ce qui reste une entrave au développement de la recherche (Paturel *et al.*, 2010). Du fait des questions traitées dans les chapitres suivants, le manuscrit utilise des échantillons de stations hydrométriques différents selon les chapitres de thèse. Le nombre de stations et les différents traitements effectués à ces jeux de données de débits collectées à l'échelle de l'Afrique sont détaillés dans les différents chapitres concernés.

---

<sup>15</sup> <https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Home>

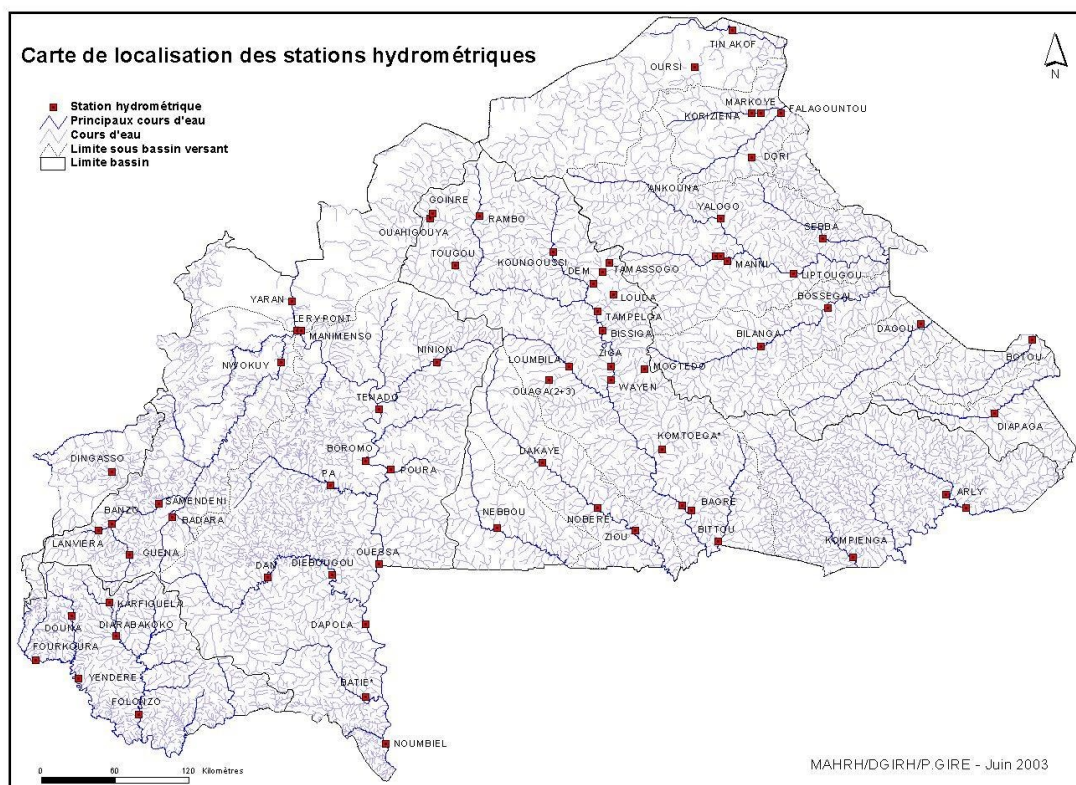


Figure I-12. Noms et localisations des stations hydrométriques au Burkina Faso (juin 2003). Source (extrait de l'annuaire hydrologique national 2011 du Burkina Faso, DGRE).

### I- 2 - 1 - 2 Données de débits spatialisés

Les données de débits spatialisés sont des données de débits répartis dans l'espace et le temps et généralement obtenues à l'aide de modèles entraînés à relier les débits ponctuels observés à des variables environnementales avant d'être utilisés pour extrapoler spatialement les débits. C'est la démarche employée (ici par réseau de neurones artificiels) pour produire la base de données globales FLO1K (Barbarossa *et al.*, 2018), qui fournit des données spatialisées sur le débit des cours d'eau à la résolution de 30 secondes d'arc (environ 1 km). Les estimations de débits ( $m^3/s$ ) minimum, moyen et maximal annuelles sur la période 1960-2015 sont disponibles à l'échelle globale. Pour ce faire, les observations mensuelles de débit au niveau de 6600 stations de jaugeage dans le monde sont prises en compte dans l'étude. Pour chaque station de jaugeage, les débits annuels minimums et maximaux correspondent aux débits mensuels moyens les plus bas et les plus élevés pour une année donnée. Les indices de débit sont liés à des variables prédictives (topographie, climat, géologie) par des modèles de réseau de neurones artificiels. L'évaluation des modèles de réseaux de neurones sur une partie du jeu de données initialement réservés pour la validation (10% des stations, soit 660 stations)

indique une bonne concordance avec un coefficient de détermination  $R^2$  de 0,83, 0,90 et 0,91 respectivement pour les débits minimums, maximums et moyens annuels sur la chronique 1960-2015. Les relations établies sont ensuite utilisées pour extrapoler spatialement les indices de débits à la résolution de 1 km. En général, la qualité des données spatialisées dépendra de la densité spatiale des stations de mesure. Une représentation spatiale du débit moyen annuel à l'échelle de l'Afrique est proposée en annexe **Figure A1-1.a**.

## I - 2 - 2 Données hydro-climatiques

### I- 2 - 2 - 1 Données de précipitations ponctuelles

Les données de précipitations ponctuelles sont fournies par l'Agence Nationale de la Météorologie (ANAM) du Burkina Faso. Les données collectées auprès de cette structure sont essentiellement des données de précipitations journalières (1954 – 2017) sur 3 stations représentatives des 3 zones climatiques (zone sahélienne, soudano-sahélienne, soudanienne) du Burkina Faso respectivement pour les stations de Dori, Ouagadougou aéroport, et Bobo-Dioulasso. Ces données sont utilisées au **chapitre II** pour analyser l'évolution de l'indice standardisé des précipitations sur la chronique de 1954 à 2017.

### I- 2 - 2 - 2 Données hydro-climatiques spatialisées

TerraClimate (Abatzoglou *et al.*, 2018) est une base de données<sup>16</sup> à haute résolution spatiale ( $1/24^\circ$ ,  $\sim 4$  km) du climat mensuel à l'échelle globale et disponible sur la période 1958 à 2017. Trois ensembles de données climatiques globales grillées ont été utilisés pour développer TerraClimate : WorldClim (Fick et Hijmans, 2017), CRU Ts4.0 (Harris *et al.*, 2020), JRA-55 (Kobayashi *et al.*, 2015). WorldClim2.0 est une base de données climatiques mondiales développées sur la base des données des stations météorologiques. Les normales climatiques mensuelles pour la période 1970-2000 ont été développées pour la température maximale et minimale, les précipitations, le rayonnement solaire, la pression de vapeur et la vitesse du vent à quatre résolutions spatiales différentes. Les données à la résolution de 2,5 minutes d'arc ( $1/24^\circ$ ,  $\sim 4$  km) ont été retenues pour le développement de la base TerraClimate. Les données CRU Ts4.0 fournissent les moyennes mensuelles de la température maximale et minimale, de la pression de vapeur et des précipitations cumulées à

---

<sup>16</sup> <http://www.climatologylab.org/terraclimate.html>

la résolution spatiale de 0,5° de 1901 à 2015 à l'échelle globale et ont été développées sur la base des données des stations météorologiques. Le JRA-55 est le produit de réanalyse moderne avec la plus grande profondeur temporelle, fournissant des données spatiales (1,25°) et temporelles complètes (3 h) couvrant la période 1958 jusqu'à aujourd'hui. Pour les ensembles de données CRU Ts4.0 et JRA-55, les anomalies mensuelles pour toutes les variables ont été calculées par rapport à une période de référence de 31 ans (1970-2000) définie par WorldClim v2.0. Une interpolation climatologique assistée a été utilisée pour superposer les anomalies climatiques mensuelles de CRU Ts4.0 et JRA-55 avec les normales climatiques mensuelles de WorldClim afin d'estimer les séries chronologiques mensuelles de 1958-2015. L'interpolation climatologique assistée (Willmott et Robeson, 1995) est une approche simple de réduction d'échelle spatiale qui utilise l'interpolation bilinéaire des anomalies temporelles d'un ensemble de données à plus haute résolution temporelle et à plus basse résolution spatiale vers un ensemble de données à plus basse résolution temporelle et à plus haute résolution spatiale.

L'évapotranspiration de référence mensuelle (ET<sub>o</sub>) est calculée en utilisant l'approche de Penman Montith (Allen *et al.*, 1998). Un modèle de bilan hydrique unidimensionnel modifié Thornthwaite-Mather (Willmott *et al.*, 1985) a été utilisé en prenant en compte l'évapotranspiration de référence, la précipitation, la température et des données sur la capacité de stockage de l'eau du sol extractible à la résolution de 0,5° provenant de Wang-Erlandsson *et al.* (2016) dans le but de calculer le bilan hydrique mensuel de 1958 à 2015. Les ensembles de données mensuelles résultant du modèle de bilan hydrique sont l'écoulement total, le déficit hydrique climatique, l'évapotranspiration réelle, ainsi que l'équivalent en eau de la neige et l'humidité du sol.

Les aspects spatio-temporels de TerraClimate ont été validés en utilisant la température annuelle, la précipitation, l'évapotranspiration potentielle annuelle et l'écoulement annuel calculé à partir des données de stations provenant d'une variété de réseaux dont le *Global Historical Climate Network* (GHCN), *Snowpack Telemetry* (SNOTEL), *Remote Automated Weather Stations* (RAWS) et FLUXNET. Cependant, les auteurs notent que parmi les stations prises en compte dans la base GHCN, un certain nombre ont été utilisées dans le développement de la base CRU et WorldClim ce qui rend la validation avec ces jeux de données pas complètement indépendants. En validation, les ensembles de données TerraClimate ont montré une amélioration notable de l'erreur absolue moyenne globale (MAE), une corrélation de Pearson ( $\rho$ ) et un réalisme spatial accru par rapport aux ensembles

de données à résolution plus grossière comme CRU (**Tableau I-6**). Ces bonnes performances à haute résolution spatiale ont justifié l'utilisation de cette base climatique dans cette thèse.

**Tableau I-6 :** Comparaison des paramètres statistiques de validation pour TerraClimate et CRU Ts4.0 pour la température moyenne annuelle (Tas), les précipitations cumulées (Pr), l'évapotranspiration de référence (ET<sub>0</sub>) et le débit cumulé (Q). La valeur médiane des statistiques de validation pour le coefficient de corrélation de Pearson ( $\rho$ ) et l'erreur absolue moyenne (MAE) est indiquée. L'erreur absolue moyenne relative est indiquée pour Pr, ET<sub>0</sub> et Q, l'erreur absolue moyenne étant indiquée entre parenthèses. Cette illustration est extraite de (Abatzoglou et al., 2018).

Variable	# Stations	Network	TerraClimate		CRU Ts4.0	
			$\rho$	MAE	$\rho$	MAE
Tas	3,230	GHCN	0.95	0.32 °C	0.95	0.53 °C
Pr	6,102	GHCN	0.90	9.1% (62.9 mm)	0.89	11% (76 mm)
Tas	1,292	SNOTEL+RAWS	0.90	0.84 °C	0.90	1.21 °C
Pr	503	SNOTEL	0.78	30% (251 mm)	0.78	41% (316 mm)
ET <sub>0</sub>	50	FLUXNET	0.77	8.3% (57 mm)	0.76	8.5% (59 mm)
Q	587	GRDC	0.80	36.0% (101 mm)	0.76	42.0% (118 mm)

Les variables climatiques issues de cette base et utilisées dans cette thèse se résument à la précipitation (P), l'évapotranspiration potentielle (ETP), l'évapotranspiration réelle (ETR), la température minimale (Tmin) et la température maximale (Tmax). Ces variables sont extraites au pas de temps mensuel et agrégées par la suite au pas de temps annuel.

Les variables dérivées pour cette thèse à partir de ces variables directement fournies par la base TerraClimate sont l'indice d'aridité (rapport entre la précipitation et l'évapotranspiration potentielle annuelle), la température moyenne (moyenne entre la température minimale et maximale), le nombre moyen de mois pluvieux par an. Le nombre moyen de mois pluvieux par an est la moyenne sur la période investiguée du nombre de mois pour laquelle la précipitation mensuelle > 15 mm, soit un seuil > à 0,5 mm/jour en moyenne pour être considéré comme un jour pluvieux (Kranjac-Berisavljevic et al., 2014). Dans les chapitres où ces données sont utilisées, un tableau récapitulatif est fourni sur les différents traitements et échelles temporelles auxquels elles sont agrégées. L'ensemble de ces variables est représenté à l'échelle de l'Afrique en annexe, **Figure A1-1** et **Figure A1-2**.

### I- 2 - 2 - 3 Données sur la recharge des nappes

L'information sur la recharge provient de l'étude de MacDonald et al. (2021). Cette étude a quantifié spatialement la moyenne à long terme de la recharge des eaux souterraines à l'échelle de l'Afrique sur la période 1970-2019 à partir de 134 mesures de terrain qui ont été extrapolées spatialement à l'aide de méthodes statistiques (modèle linéaire mixte), en

négligeant la recharge par infiltration directe des grands cours d'eau, des lacs ou de l'irrigation bien que celle-ci peut être importante notamment dans les milieux semi-arides (Rusagara *et al.*, 2022). La variable la plus importante du modèle est la précipitation moyenne annuelle issue de la base CRU Ts4.0 (Harris *et al.*, 2020) et le volume de la recharge annuelle estimée pour toute l'Afrique est de 1500 km<sup>3</sup> et se caractérise par une forte variabilité entre les aquifères sédimentaires à fort stockage/faible recharge en l'Afrique du Nord, et les aquifères de roche cristalline altérée à faible stockage et à forte recharge dans la majeure partie de l'Afrique tropicale. Le jeu de données est distribué à la résolution de 0,28° (31 km x 31 km). Une représentation spatiale de la recharge moyenne annuelle des nappes à l'échelle de l'Afrique est proposée en annexe **Figure A1-1.e**.

#### I- 2 - 2 - 4 Teneur en eau du sol

Les données sur la teneur en eau<sup>17</sup> du sol proviennent de l'étude de Trabucco et Zomer (2019). Un modèle global spatialement distribué sur le bilan en eau du sol a été implémenté sur les données WorldClim et CGIAR-CSI PET en utilisant ArcAML (ESRI) comme outil de modélisation. Ce modèle utilise des valeurs spatialement distribuées (moyenne sur la période 1950-2000) de précipitations mensuelles et d'évapotranspiration potentielle mensuelle et renvoie en sorties des valeurs mensuelles spatialement distribuées de l'évapotranspiration réelle, du ruissellement et de l'humidité du sol et dont les équations de bilan sont détaillées dans Trabucco et Zomer (2010). Le SWC\_fr définit la teneur en eau disponible pour les processus d'évapotranspiration (en pourcentage de l'humidité maximale en eau du sol). Les valeurs moyennes annuelles des teneurs en eau sont utilisées à la résolution de 1 km et sont à une résolution plus fine que celles des teneurs en eau proposée dans la base TerraClimate (~ 4 km). Une représentation spatiale de la teneur en eau du sol annuelle moyenne à l'échelle de l'Afrique est proposée en annexe **Figure A1-1.f**.

#### I - 2 - 3 Données topographiques

##### I- 2 - 3 - 1 Altitudes et pentes du terrain

L'altitude désigne la hauteur verticale entre un point et le plus souvent le niveau de la mer. À large échelle, l'information sur l'altitude est généralement obtenue à l'aide de MNT (Modèle

---

<sup>17</sup> [https://figshare.com/articles/dataset/Global\\_High-Resolution\\_Soil-Water\\_Balance/7707605/3](https://figshare.com/articles/dataset/Global_High-Resolution_Soil-Water_Balance/7707605/3)



Numérique de Terrain) qui est une représentation numérique des altitudes du terrain naturel. Les données d'altitude utilisées proviennent de la base hydroSHEDS<sup>18</sup> (Lehner *et al.*, 2008) et les résolutions utilisées dans cette étude sont de 15 et 3 secondes d'arc (environ 500 m et 90 à l'équateur). Les données de MNT présentes dans cette base ont été acquises lors de la mission SRTM (Shuttle Radar Topography Mission) de la NASA en 2000 (Farr et Kobrick, 2000) et pour laquelle le MNT a été corrigé des erreurs d'altitudes.

La pente du terrain naturel est la tangente de l'angle formé entre deux points du terrain naturel par rapport à l'horizontale. Elle est un indice qui permet d'avoir une appréciation de la morphologie du terrain et permet la description de l'écoulement potentiel de l'eau vers les cours d'eau sous l'action de la gravité. À l'aide des logiciels SIG et en utilisant un MNT, la pente entre deux cellules adjacentes est calculée comme le rapport de leur différence d'élévation sur la distance horizontale entre leurs centres.

Une représentation spatiale de la pente et de l'altitude et à l'échelle de l'Afrique est respectivement proposée en annexe **Figure A1-3.a,b**.

### I- 2 - 3 - 2 Topographic wetness index

L'indice topographique ou *Topographic wetness index* (Twi) est un indice couramment utilisé dans l'analyse hydrologique (Mattivi *et al.*, 2019) et qui décrit la tendance d'une zone à accumuler l'eau sous l'effet de la topographie. Aussi connu sous le nom de *compound topographic index* (Cti), l'indice topographique (Twi) est le rapport logarithmique de la surface de drainage amont spécifique ( $a$  en m) sur la pente locale ( $\tan(\beta)$ , sans dimension) pour chaque point de l'espace :

$$Twi = \ln \left( \frac{a}{\tan(\beta)} \right) \quad (I-5)$$

Les valeurs élevées de cet indice se retrouvent généralement dans des régions plates drainant une grande surface amont, ce qui correspond à une forte propension à la saturation tandis que les zones à pentes raides et/ou faibles surfaces amont sont associées à de faibles valeurs de Twi. L'indice a montré de bonnes corrélations avec la distribution réelle de l'humidité du sol dans de nombreux petits bassins versants, bien qu'il présente certaines limites dues à la dépendance aux changements saisonniers et à d'autres variables non topographiques comme

---

<sup>18</sup> [www.hydrosheds.org/](http://www.hydrosheds.org/)

les propriétés du sol, l'utilisation des terres, la végétation, l'évapotranspiration (Tootchifatidehi, 2019). Les données de  $Twi$ <sup>19</sup> utilisées dans cette thèse proviennent de l'étude de Hengl (2018) où les valeurs ont été calculées à partir du MNT Merit à la résolution de 500 m. Une représentation spatiale de l'indice *Topographic wetness index* à l'échelle de l'Afrique est proposée en annexe **Figure A1-3.c**.

### I- 2 - 3 - 3 Topographic climate index

Le *Topographic climate index* (Tci) est une amélioration de l'indice ( $Twi$ ) pour tenir compte de l'effet du climat, en plus de l'effet de la topographie, dans la propension d'une zone à accumuler de l'eau. En s'inspirant de Merot *et al.* (2003), il se définit selon l'équation suivante :

$$Tci = \ln\left(\frac{a * Pe}{\tan(\beta)}\right) = Twi + \ln(Pe) \quad (I-6)$$

Où  $a$  est la surface amont drainée spécifique (m),  $\tan(\beta)$  la pente locale (en radians),  $Twi$  ( $\ln(m)$ ), et  $Pe$  est la moyenne annuelle de la pluie efficace (m). Celle-ci est d'abord calculée au pas de temps mensuel selon l'équation (I-7) :

$$P_{m,y}^e = \max(0, P_{m,y} - ETP_{m,y}) \quad (I-7)$$

où  $P_{m,y}$  et  $P_{m,y}^e$  sont les précipitations mensuelles et mensuelles effectives (en m) pour le mois  $m$  et l'année  $y$ , et  $ETP_{m,y}$  est l'évapotranspiration mensuelle (en m) pour le mois  $m$  et l'année  $y$ .  $Pe$  est ensuite calculé comme la somme des 12 moyennes pluriannuelles des précipitations mensuelles effectives. Les variables climatiques requises sont extraites des données météorologiques mensuelles du TerraClimate (Abatzoglou *et al.*, 2018), désagrégées à la résolution de 500 m qui est celle du  $Twi$  de l'étude (Hengl, 2018). Une représentation spatiale de l'indice *Topographic climate index* à l'échelle de l'Afrique est proposée en annexe **Figure A1-3.d**

<sup>19</sup> <https://zenodo.org/record/1447210#.YdRfUWjP2M8>

## I - 2 - 4 Données géologiques et propriétés des sols

### I- 2 - 4 - 1 Perméabilité

À l'échelle de l'Afrique, les données de perméabilité<sup>20</sup> utilisées sont issues de la base GLHYMPS2.0 (GLobal HYdrogeology MaPS 2.0) (Huscroft *et al.*, 2018). C'est une mise à jour de la version GLHYMPS (Gleeson *et al.*, 2014) qui fournit des données de perméabilité et porosité à l'échelle globale. Ces données de perméabilité s'appuient sur la cartographie mondiale de la lithologie (Hartmann et Moosdorf, 2012) également utilisée pour contraindre le réseau hydrographique LCS. Cette étude améliore les données GLHYMPS existantes en représentant mieux les sédiments non consolidés et profite donc des avancées récentes dans la cartographie des sédiments non consolidés de la carte globale GUM "Global Unconsolidated Map" (Börker *et al.*, 2018) et de la carte de la profondeur du substratum rocheux « *Depth to bedrock* » dans SoilGrids (Shangguan *et al.*, 2017). La perméabilité a été attribuée aux polygones du GUM en fonction de la distribution des tailles de grains et des types de sédiments. Cette nouvelle base de données fournit une carte globale de perméabilité superficielle (de l'ordre de 100 m). Il convient de noter que les données sur la perméabilité superficielle ont deux attributs : avec et sans l'effet du permafrost. Nous avons utilisé celle sans l'effet du permafrost puisque celui-ci n'est pas présent en Afrique. Une représentation spatiale de la perméabilité à l'échelle de l'Afrique est proposée en annexe **Figure A1-4.f**.

Au Burkina Faso, la disponibilité d'une carte géologique détaillée du BUMIGEB (Bureau des Mines et de la Géologie du Burkina Faso) qui est une version mise à jour de la carte géologique de Castaing *et al.* (2003) à l'échelle de 1 : 1 000 000 a permis de mieux représenter les unités lithologiques, de les classer selon la nomenclature de Hartmann et Moosdorf (2012) et de déduire des valeurs de perméabilité *a priori* plus réalistes au **chapitre II**. Ces valeurs de perméabilité attribuées à chaque type de lithologie sont basées sur la gamme de perméabilité par type de sédiment ou de roche issue du chapitre 6 de Maidment (1993) et en tenant compte de la variation spatiale de la densité de drainage observée à partir du réseau hydrographique de référence BNDT. Les cartes relatives à la lithologie de Hartmann et Moosdorf (2012), à la carte géologique du Burkina Faso et la carte de perméabilité dérivée sont respectivement présentées en (**Figure II-A-3, Figure II-A-6**).

---

<sup>20</sup> <https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/TTJNIU>

## I- 2 - 4 - 2 Stocks d'eau souterraine

L'information sur le stockage souterrain dans les aquifères<sup>21</sup> est obtenue de MacDonald *et al.* (2012). Cette étude présente les estimations à l'échelle du continent africain du stockage des aquifères et des rendements potentiels des forages en se basant sur un examen approfondi des cartes, publications et données disponibles pour estimer les paramètres hydrogéologiques. Pour estimer le stockage des eaux souterraines, l'épaisseur de l'aquifère saturé a été multipliée par la porosité effective. Le stockage total des eaux souterraines en Afrique est estimé à 0,66 million de km<sup>3</sup> avec une inégale répartition, les plus importants stockages se trouvant dans les grands aquifères sédimentaires des pays d'Afrique du Nord, en Libye, en Algérie, en Égypte et au Soudan. Les données sont distribuées sous la forme de couches de polygones et converties dans le cadre de cette thèse en données raster à la résolution par défaut conseillée par les auteurs qui est de 0,26° x 0,26°, soit approximativement 30 km x 30 km. Une représentation spatiale des stocks d'eau souterraine à l'échelle de l'Afrique est proposée en annexe **Figure A1-4.d**.

## I- 2 - 4 - 3 Profondeur de la nappe

La profondeur de la nappe phréatique ou « *water table depth* » (Wtd) peut influencer les processus hydrologiques de multiples façons en contribuant par exemple à l'écoulement de base des rivières et des lacs. Fan *et al.* (2013) ont estimé la profondeur de la nappe phréatique à l'échelle globale (environ 1 km de résolution spatiale) à partir d'une modélisation hydrogéologique en exploitant des données extraites d'archives gouvernementales et de la littérature. À cet effet, plus d'un million de données piézométriques sont collectées et analysées, dont seulement 431 observations en Afrique. Le modèle suppose un écoulement stationnaire, et les flux d'eau latéraux sont calculés en utilisant la loi de Darcy et l'approximation de Dupuit-Forchheimer pour un écoulement en 2D. La calibration porte principalement sur le paramètre qui contrôle la décroissance exponentielle de la conductivité hydraulique avec la profondeur et ce taux de décroissance dépend de la pente du terrain naturel. La pente est extraite des données d'altitudes à 30 secondes d'arc (issue de la base HydroSHEDS pour les latitudes < 60° N et ASTER/NASA-JPL pour les latitudes > 60° N). Pour estimer la transmissivité de subsurface, les conductivités hydrauliques du sol ont été

---

<sup>21</sup> <https://www2.bgs.ac.uk/groundwater/international/africanGroundwater/mapsDownload.html>

dérivées des cartes pédologiques de l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO) (à la résolution de 5 minutes d'arc) et des cartes pédologiques du ministère américain de l'Agriculture (USDA) pour les États-Unis (à la résolution 30 secondes d'arc). La recharge a été modélisée à la résolution de 0,5° à l'aide du modèle WaterGAP (Döll et Fiedler, 2008) sur la période (1979-2007) à base des données climatiques du CRU. À l'échelle globale, le niveau de la mer est le facteur le plus important de la profondeur de la nappe ; la profondeur de la nappe est faible le long des côtes maritimes de tous les continents et sous tous les climats. À l'échelle régionale, le climat émerge comme le facteur de contrôle le plus important de la profondeur de la nappe. Dans les latitudes moyennes et faibles, les zones de profondeur de la nappe élevée correspondent aux régions de faible recharge. Une représentation spatiale de la profondeur de la nappe à l'échelle de l'Afrique est proposée en annexe **Figure A1-4.e**.

#### I- 2 - 4 - 4 NDVI

Le NDVI (*Normalized Difference Vegetation Index*) ou indice de végétation par différence normalisée est un indice qui permet de quantifier la densité de végétation active à partir de la différence entre la réflectance dans la bande visible du rouge (R) et du proche infrarouge (PIR), telle que mesuré par des capteurs embarqués sur des drones ou des satellites :

$$NDVI = \frac{PIR - R}{PIR + R} \quad (I-8)$$

Les valeurs de cet indice oscillent entre -1 et 1, les valeurs négatives correspondant aux surfaces autres que les couverts végétaux, comme la neige, l'eau ou les nuages pour lesquels la réflectance dans le rouge est supérieure à celle du proche infrarouge. Pour les sols nus, les réflectances étant à peu près du même ordre de grandeur dans le rouge et le proche infrarouge, le NDVI présente des valeurs proches de 0. Les formations végétales ont quant à elles, des valeurs de NDVI positives, les valeurs les plus élevées correspondant aux couverts les plus denses. Cet indice donne donc des informations indirectes sur l'occupation du sol (végétation, eau, terre dégradée).

Les données de NDVI analysées dans cette thèse (GIMMS NDVI 3g)<sup>22</sup> proviennent de l'étude de Pinzon et Tucker (2014). Il s'agit d'une série de données à la résolution spatiale de 8 km, produite à partir des mesures du radiomètre avancé à très haute résolution AVHRR (*Advanced*

---

<sup>22</sup> <https://climatedataguide.ucar.edu/climate-data/ndvi-normalized-difference-vegetation-index-3rd-generation-nasagfsc-gimms>

*Very High Resolution Radiometer*) depuis 1981 jusqu'à aujourd'hui. Une représentation spatiale des valeurs de NDVI à l'échelle de l'Afrique est proposée en annexe **Figure A1-4.c**.

#### I- 2 - 4 - 5 Texture du sol (argile, limon et sable)

La texture du sol est définie en fonction des pourcentages en masse d'argile, limon et de sable contenus dans le sol (dont la somme fait 100%) et est un bon indicateur des propriétés hydrauliques du sol (Tafasca, 2020). Les données sur les pourcentages en sable, limon et argile utilisées dans cette thèse proviennent de la base de données SoilGrids<sup>23</sup> à la résolution spatiale de 250 m et aux profondeurs allant de 0 à 200 cm (0 cm, 5 cm, 15 cm, 30 cm, 60 cm, 100 cm et 200 cm). Les données de cette base (avec plus de 7 paramètres de sol disponibles) sont générées à partir de l'étude de Hengl *et al.* (2017) qui se base sur les données observées de 150 000 profils de sols à l'échelle globale entraînées par un ensemble de modèles d'apprentissage automatique (*Random Forest* voir section I-4, gradient boosting, régression logistique) et de nombreuses variables environnementales prédictives (158). Les auteurs soulignent une amélioration de la variance du modèle par rapport à la version précédente de SoilGrids à 1 km dont les relations étaient basées sur des modèles linéaires (60% à 230 % d'amélioration en fonction des paramètres).

Nous avons choisi dans le cadre de cette étude de prendre pour chaque type du sol (argile, limon, sable) les valeurs de pourcentage des prédictions intégrées sur l'épaisseur (0-30 cm) du profil de sol (Lin *et al.*, 2020). Une représentation spatiale des valeurs de NDVI à l'échelle de l'Afrique est proposée en annexe **Figure A1-4.c**.

#### I - 2 - 5 Données de pressions anthropiques

En fonction de la capacité de stockage des barrages à l'amont, les cours d'eau peuvent être influencés et devenir plus ou moins intermittents (Chiu *et al.*, 2017). Les données sur l'influence anthropique prises en compte dans cette thèse concernent les barrages, grâce à la base de données GRAND (*Global Reservoir and Dam Database*) (Lehner *et al.*, 2011) mise à jour en 2019 (version 1.3). Cette base de données a harmonisé et rassemblé les données existantes sur les barrages et réservoirs (en supprimant les doublons) en une base de données unique, géographiquement explicite. La compilation d'informations sur les barrages et

---

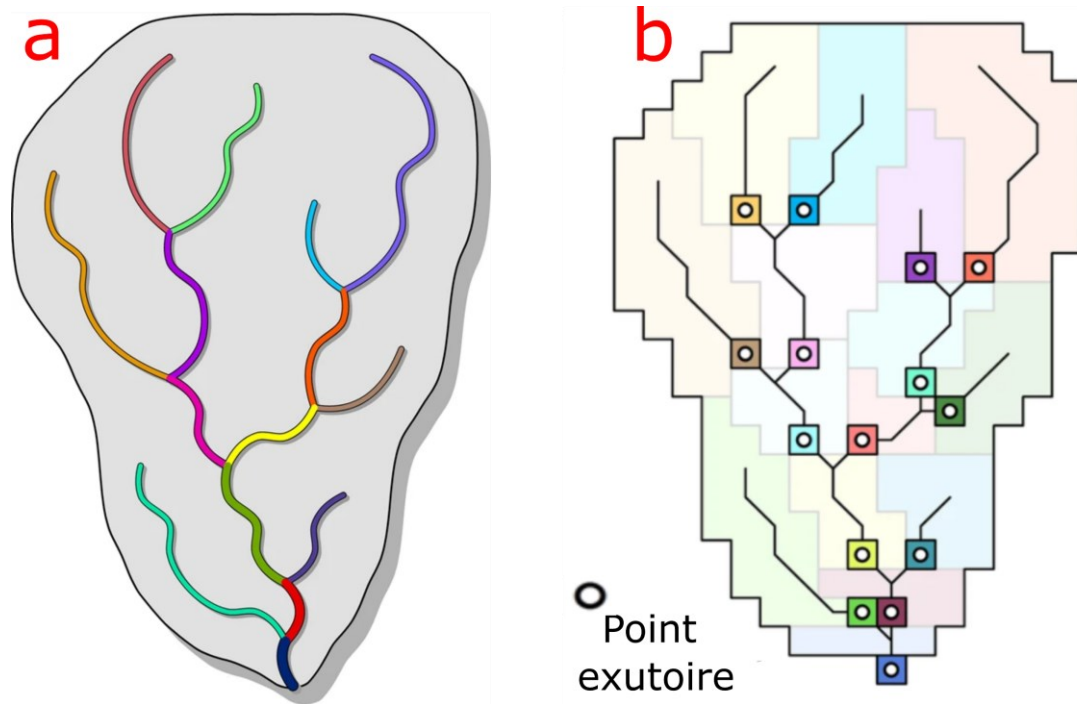
<sup>23</sup> [www.soilgrids.org](http://www.soilgrids.org)

réservoirs s'est faite à partir de diverses sources, dont les archives nationales, certaines bases de données privées et études, dont Messenger *et al.* (2016) par exemple, ce qui a permis de géoréférencer l'emplacement de plus de 6862 barrages à l'échelle globale. Bien que la base GRAND comporte des données sur des petits barrages, elle contient principalement des réservoirs ayant une capacité de stockage de plus de 0,1 km<sup>3</sup>. L'information principale extraite de cette base concerne la capacité de stockage des barrages ainsi que leur emplacement à l'échelle de l'Afrique. En Afrique, 681 barrages sont répertoriés sur une période de 1896 à 2017 pour une capacité de stockage entre 0,1 à 204 800 km<sup>3</sup>. Pour la majorité des barrages, la fonction principale est l'irrigation, l'approvisionnement en eau et l'hydro-électricité. Il faut noter que les dérivations / pompes sur les cours d'eau pour diverses activités peuvent contribuer à accroître l'intermittence (Benejam *et al.*, 2010) tandis que le rejet d'effluents par des installations industrielles peut contribuer à maintenir une certaine permanence de l'écoulement (Bischel *et al.*, 2013), mais ces informations sont difficiles à obtenir, surtout à l'échelle continentale. Une représentation spatiale de la capacité de stockage des barrages à l'échelle de l'Afrique est proposée en annexe **Figure A1-3.e**.

## I - 3 Prétraitements des données et caractéristique des bassins versants

Pour caractériser spatialement l'intermittence des cours d'eau, le réseau hydrographique est souvent considéré en tronçons de cours d'eau (Linke *et al.*, 2019 ; Snelder *et al.*, 2013) qui sont définis comme un segment ou portion ou section de rivière situés entre deux affluents ou entre un affluent et le début ou la fin d'une rivière (**Figure I-13.a**). Le bassin versant amont se rapporte à toute la zone en amont qui est hydrologiquement connectée à ce tronçon. Pour extraire les caractéristiques amont associées à chaque tronçon de cours d'eau, deux échelles de calcul sont considérées :

- i. Tout le bassin versant amont associé au tronçon de cours d'eau.
- ii. Le point exutoire associé à chaque tronçon de cours d'eau (**Figure I-13.b**). Le point exutoire est notamment adapté pour extraire des variables hydrologiques comme le débit ou l'ordre de Strahler (FLO1K par exemple, section **I-2-1-2**) qui est extrait ponctuellement, mais caractérise le bassin versant amont.



**Figure I-13** : Aperçus sur les tronçons de rivières, chaque tronçon est représenté par une ligne d'une couleur différente (a), les points exutoires des tronçons de rivières (b). Les figures (a), (b) sont extraites de Linke et al. (2019).

L'outil ArcMap 10.6.1 a été utilisé pour le traitement des données (projection, rééchantillonnage, découpage, superposition). Pour faciliter la manipulation des données, elles ont été projetées dans le système de projection globale WGS84 (code EPSG : 4326). Le réseau hydrographique LCS est dérivé du MNT hydroSHEDS à la résolution de 500 m et la grille d'écoulement « *Flow direction* » et d'accumulation « *Flow accumulation* » exploitée durant cette thèse sont également issues de la même base à la résolution de 500 m. Dans le but d'avoir des données qui se superposent correctement aux grilles de LCS et aux grilles d'accumulation et d'écoulement qui sont nécessaires pour extraire les statistiques des bassins versants amont, les données ont été rééchantillonnées c'est-à-dire agrégées ou désagrégées à la résolution de 15 secondes d'arc (environ 500 m à l'équateur). Dans notre cas, à l'exception des données sur la teneur en eau (section I-2-2-4) qui sont à la résolution de 250 m, l'ensemble des données avait une résolution plus grande que 15 secondes d'arc ; elles ont donc été désagrégées, les cellules de grande taille ont simplement été divisées en plus petites tout en conservant donc leur valeur d'origine. Les données de teneur en eau ont, elles, été agrégées (par moyenne) à la résolution de 15 secondes d'arc. Les données distribuées sous forme de couches de polygones ont aussi été converties en rasters de 15 secondes d'arc pour faciliter les traitements. C'est le cas des données de perméabilité (Huscroft *et al.*, 2018).



Le réseau LCS est fourni avec les tronçons de cours d'eau qui sont déjà identifiés. Pour dériver les points exutoires associés à chaque tronçon de cours d'eau, les grilles d'écoulement et d'accumulation ont été exploitées dans l'outil Archydro<sup>24</sup> 10.6 qui est une extension de l'outil ArcMap.

Pour calculer les caractéristiques du bassin versant amont associé à chaque tronçon de cours d'eau notamment la moyenne (qui est l'opérateur principalement utilisé dans cette thèse), les valeurs de toutes les cellules amont de la variable sont accumulées le long de la grille d'écoulement communément appelée « *Flow direction* » puis divisées par le nombre de cellules amont du bassin versant (Linke *et al.*, 2019).

Dans notre cas, le domaine spatial (Afrique) étant grand, le système de projection globale WGS84 a été utilisé, or dans cette projection, la surface des cellules varie en fonction de la latitude (Budic *et al.*, 2016 ; Verdin Kristine, 2017). Cette variation de la taille des cellules doit être prise en compte dans le calcul des statistiques de bassin versant amont. Une correction a été effectuée pour tenir compte de la distorsion latitudinale dans la taille des cellules due à la projection géographique appliquée. La surface de chaque cellule de 15 secondes d'arc (1°/240) varie en fonction de la latitude suivant la formule :

$$w_i \cong \left( \frac{2\pi R}{360} * \frac{1}{240} \right)^2 * \cos\alpha_i = 0,215 * \cos\alpha_i \quad (\text{I-9})$$

avec  $w_i$  en km<sup>2</sup> et R = 6378,137 km (rayon de la terre). Comme prévu, la surface  $w_i$  est maximale à l'équateur (latitude = 0°,  $w_i = 0,215$  km<sup>2</sup>) avec une valeur minimale  $w_i = 0,169$  km<sup>2</sup> en Afrique notamment pour les cellules à la pointe de l'Afrique du Sud.

Ainsi, pour chaque bassin versant amont associé à un tronçon de cours d'eau, des moyennes pondérées par la surface de la cellule sont calculées.

---

<sup>24</sup> <http://downloads.esri.com/archydro/archydro/>

## I - 4 Mise en œuvre d'outils d'apprentissage automatique : cas de Random Forest

### I - 4 - 1 Notions générales sur Random Forest

*Random Forest* ou forêt aléatoire (RF) est une méthode statistique non paramétrique d'apprentissage automatique introduite par Breiman (2001). Le but de *Random Forest* est de prédire une ou plusieurs variables d'intérêt données (variables de réponse) à partir de variables explicatives. Les variables explicatives sont des variables prédictives qui ont toutes un lien avec une variable d'intérêt donnée qu'on cherche à expliquer ; les variables explicatives sont utilisées pour prédire cette variable d'intérêt, généralement à partir d'un modèle. *Random Forest* ne nécessite pas d'hypothèse sur la distribution normale des données, ce qui permet d'utiliser différents types de variables explicatives en entrée. Cette méthode statistique se base sur la construction d'un ensemble d'arbres décisionnels (**Figure I-14**) qui constitueront au final la « Forêt ». Les arbres décisionnels sont aussi appelés arbre de classification et de régression (CART) et sont individuellement créés en partitionnant (découpant) plusieurs fois les données des variables explicatives en différents sous-espaces (Breiman *et al.*, 1984) de telle sorte que le résultat dans ces espaces soit le plus homogène que possible (**Figure I-15.a**). Cette approche de découpage des valeurs des variables explicatives pour construire un arbre est appelée « *Recursive partitioning* » (Therneau et Atkinson, 1997). L'arbre est dit de « *Classification* » quand les variables à prédire sont catégorielles c'est-à-dire sous forme de classes par exemple « *Oui ou Non* » (2 catégories). Il est dit de régression quand les variables à prédire sont continues par exemple « *les valeurs moyennes de températures d'un bassin versant* ». Les différentes parties d'un arbre peuvent être résumées comme suit :

- **Root Node** : C'est le tout premier nœud, il désigne l'ensemble de la population et est divisé en deux ou plusieurs nœuds de décision en fonction des valeurs caractéristiques d'une variable explicative.
- **Decision Node** : La division d'un nœud en un plus grand nombre de sous-nœuds est appelée nœud de décision.
- **Leaf Node** : ce sont les nœuds terminaux, car ils ne peuvent pas se diviser davantage.
- **Branch ou sub-tree** : une subdivision d'un arbre complet.

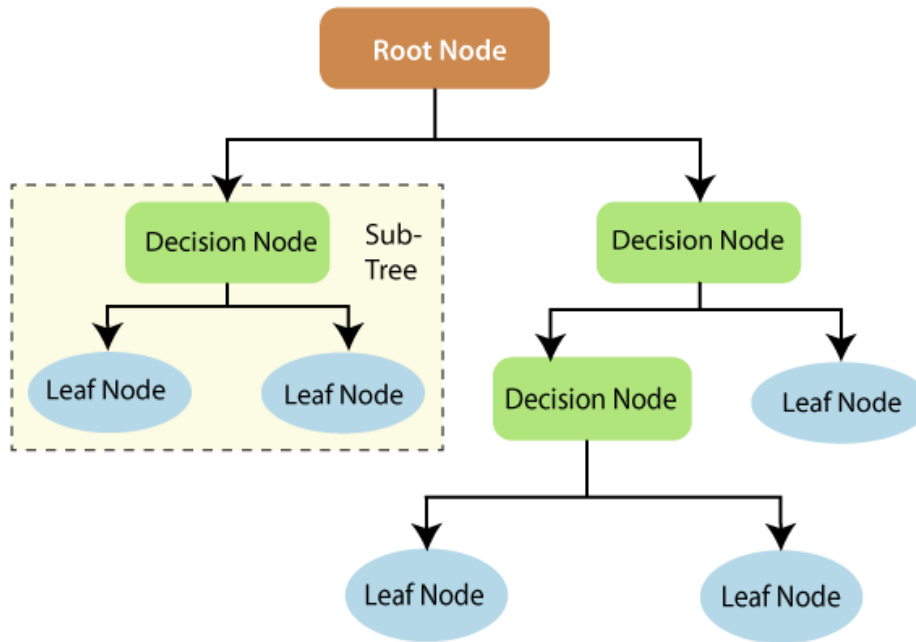


Figure I-14 : Schéma d'un arbre de décision<sup>25</sup>.

Par exemple, considérons l'application d'un arbre de décision à un ensemble de données où la variable à prédire est catégorielle, à savoir la variété d'une fleur parmi 3 possibilités (**setosa**, **versicolor** et **virginica**). La prédiction s'appuie sur deux variables explicatives, la largeur et la longueur du sépale d'une fleur (**Sepal.Width**, **Sepal.Length**). Le découpage des données se traduit par la construction de régions rectangulaires qui séparent ici 5 jeux d'observations, ou triplets Variété, *Sepal.Width*, *Sepal.Length* (Figure I-15.a). La classe ayant la plus grande proportion (majoritaire) dans chaque région est la classe prédite dans l'arbre de classification correspondant (Figure I-15.b).

<sup>25</sup> <https://laptrinhx.com/decision-trees-in-machine-learning-1041658267/>

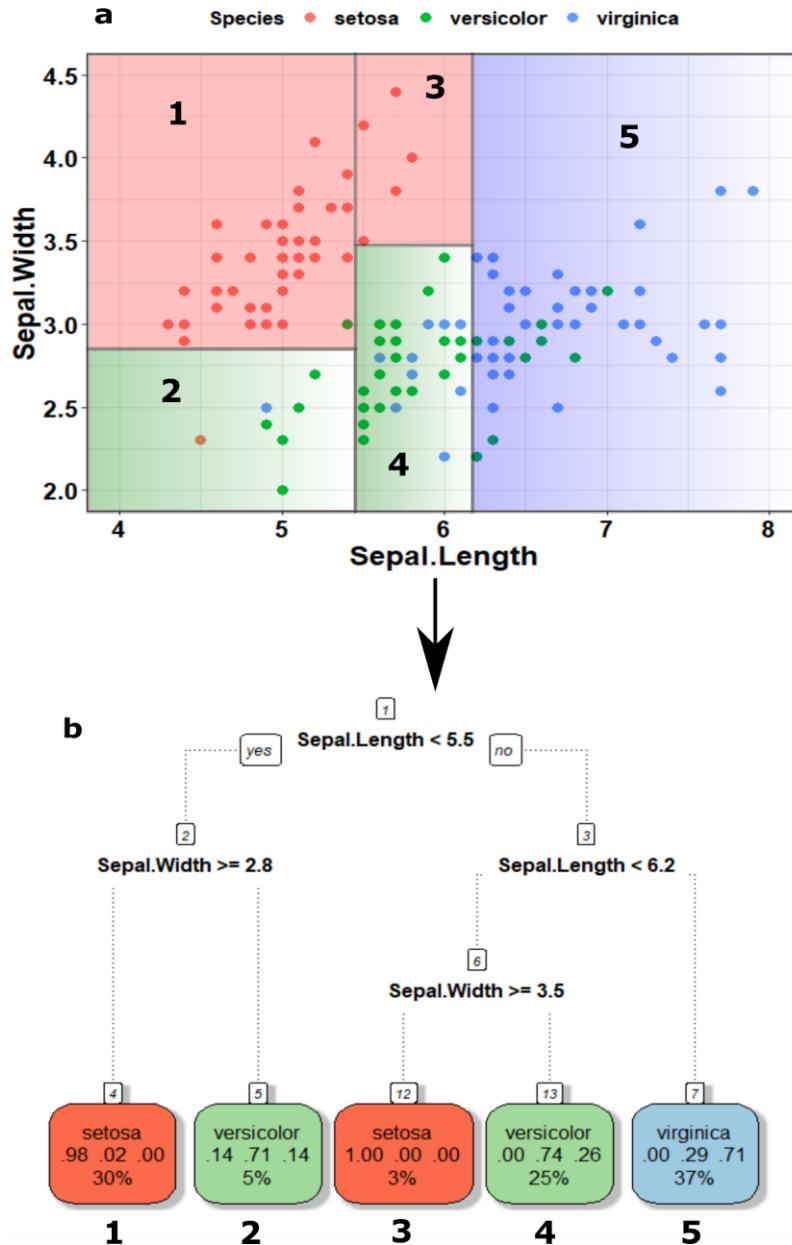
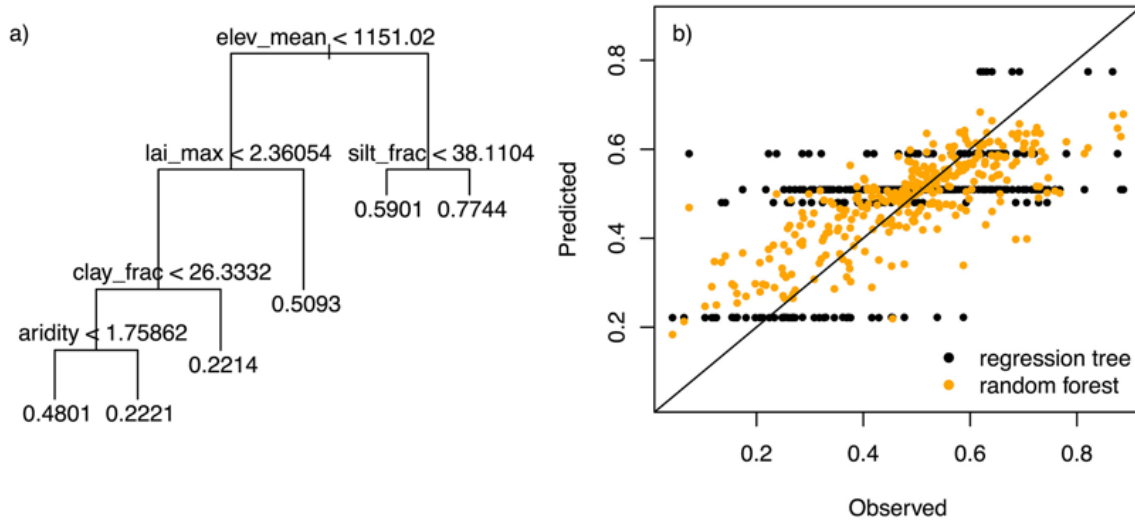


Figure I-15 : (a) Découpage des données d'entrée en différents sous-espaces (type de fleur en fonction des variables Sepal.Length et Sepal.Width), (b) arbre de décision résultant du découpage.

À chaque division de l'arbre, une variable et un seuil sont sélectionnés en phase d'apprentissage. Pour un arbre de régression, la valeur prédite est donnée par la moyenne des observations d'entraînement qui appartiennent à la même branche (Figure I-16.a). En revanche, pour un arbre de classification, la valeur prédite appartient à la classe d'observations la plus fréquente (majoritaire) dans la branche à laquelle elle appartient. En conséquence, les prédictions d'un seul arbre de régression sont des valeurs discrètes limitées. La Figure I-16.b illustre bien ce cas où les valeurs de prédictions d'un seul arbre de régression sont matérialisées par l'alignement horizontal des points (en noir) comparés aux prédictions effectuées par RF (en points orange) avec des valeurs de prédictions plus variables.

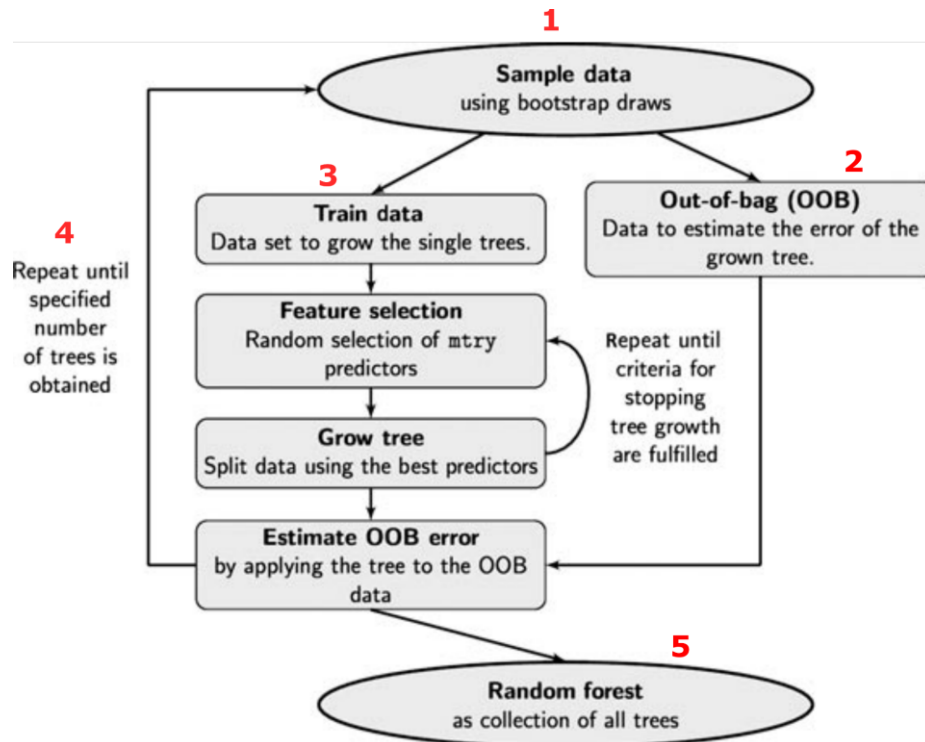


**Figure I-16** : (a) Exemple d'un arbre de régression entraîné pour prédire l'indice de débit de base. (b) Comparaison des observations de l'indice de débit de base avec les prédictions de l'arbre de régression présenté en (a) et d'une forêt aléatoire. La figure est extraite de Addor *et al.* (2018).

Si les arbres de décision sont intuitifs et faciles à interpréter, ils manquent généralement de robustesse (Addor *et al.*, 2018), c'est-à-dire qu'un petit changement dans les données peut entraîner un grand changement dans l'arbre final estimé (Ziegler et König, 2014). Pour pallier à cette limite, on s'appuie généralement sur l'agrégation d'un grand nombre d'arbres de décision, ce qui définit une forêt aléatoire ou *Random Forest*, pour produire un ensemble de prédictions avec des performances considérablement améliorées (Breiman, 2001), mais avec le désavantage d'être difficile à interpréter. Ces très bonnes performances prédictives sur un ensemble de données, qu'il soit large ou réduit, ont contribué à l'utilisation de cette méthode dans de nombreux domaines scientifiques (Capitaine, 2020 ; Cutler *et al.*, 2007) bien que Nearing *et al.* (2021) trouvent que son apport en hydrologie est encore sous-exploité. Il est bien adapté à la prévision hydrologique / hydrogéologique (Lange et Sippel, 2020 ; Saadi *et al.*, 2019 ; Yang *et al.*, 2019) en raison de (1) sa capacité à traiter de nombreuses variables explicatives ou prédictives avec des relations potentiellement non linéaires, (2) du risque relativement faible d'ajustement excessif à un sous-ensemble des données de l'échantillon (surapprentissage), et de (3) la facilité d'interprétation de l'importance de chaque variable d'entrée (Tyrallis *et al.*, 2019 ; Zipper *et al.*, 2021). Nous l'avons retenue, car il a été démontré qu'elle atteignait des performances élevées en prédiction par rapport à d'autres approches, notamment d'autres modèles d'apprentissage automatique, comme la régression logistique, les « *gradient boosting machines* », ou les SVM (« *support vector machines* ») (Wainer, 2016).

## I - 4 - 2 Paramètres d'une Random Forest

Les RF ont un certain nombre de paramètres qui doivent être définis par l'utilisateur (connus sous le nom d'hyperparamètres dans la littérature) et dont les plus importants sont (i) le nombre d'arbres à construire (*ntree*), (ii) le nombre de variables à sélectionner à chaque nœud pour le découpage des arbres (*mtry*). La **Figure I-17** illustre le fonctionnement d'une RF, dont l'algorithme peut se résumer comme suit, à la fois pour la classification et la régression.



**Figure I-17** : Fonctionnement de l'algorithme de Random Forest. La figure est adaptée de Boulesteix et al. (2012).

Cet algorithme s'appuie sur une boucle sur les  $B$  arbres qui composeront la forêt (*ntree*). Pour construire chaque arbre  $b$  de cet ensemble, on réalise les opérations suivantes :

1. Construire un échantillon *bootstrap*<sup>26</sup> (par tirage avec remise) de taille  $N$  égale aux nombres d'observations initiales.
2. Réserver  $N/3$  des données bootstrappées (*Out of bag data*) pour estimer l'erreur de l'arbre qui sera construit.
3. Construire un arbre de régression ou classification  $T_b$  en utilisant le reste ( $2N/3$ ) des données bootstrappées, en répétant récursivement les étapes suivantes pour chaque nœud de l'arbre :

<sup>26</sup> [https://fr.wikipedia.org/wiki/Bootstrap\\_\(statistiques\)](https://fr.wikipedia.org/wiki/Bootstrap_(statistiques)) : Un *bootstrap* consiste à créer de « nouveaux échantillons », mais uniquement par tirage avec remise, à partir de l'échantillon initial.

- i. Sélectionner aléatoirement  $mtry$  variables parmi les  $p$  variables explicatives investiguées ;
- ii. Choisir les meilleures variables parmi les  $mtry$  pour diviser les données de chaque nœud en deux sous nœuds. Les meilleures variables choisies sont celles qui minimisent l'impureté de Gini<sup>27</sup> dans le cas de la classification et la variance dans le cas de régression ;
- iii. Estimer l'erreur de l'arbre construit OOB «*Out Of Bag error*», en appliquant l'arbre pour prédire la variable de réponse dans les données réservées à l'étape 2 (*Out of bag data*).

Lorsque tous les arbres ont été construits, une erreur OOB de la RF est finalement calculée en faisant la moyenne de l'ensemble des erreurs individuelles calculées au sein des arbres. Il faut noter que le nombre de nœuds dans un arbre est contrôlé par le paramètre «*nodesize*» qui fixe le nombre d'observations minimales à chaque nœud de l'arbre et les valeurs par défaut sont couramment utilisées et sont respectivement de 1 en classification et 5 en régression. Une fois que les  $B$  arbres ont été construits, la forêt est constituée de l'ensemble de ces arbres. **Les étapes 1 et 3.i sont celles qui introduisent la composante "aléatoire" dans *Random Forest* et qui font que tous les arbres  $B$  qui composent la forêt sont tous différents.**

Dans une RF, la prédiction finale en sortie est la moyenne des prédictions de l'ensemble des arbres quand on est en régression, et la classe majoritaire des prédictions de l'ensemble des arbres quand on est en classification (Belgiu et Drăguț, 2016). Dans cette thèse, à chaque fois que nous avons utilisé une RF pour prédire une variable de réponse liée à l'intermittence, nous l'avons construite en régression et en classification afin de choisir l'option qui donne les meilleures performances.

Pour ajuster les hyperparamètres du modèle RF, la valeur par défaut de  $mtry$  est  $\frac{p}{3}$  en régression, par opposition à  $\sqrt{p}$  pour la classification, où  $p$  est le nombre de variables explicatives investiguées, le nombre d'arbres à construire ( $n_{tree}$ ) par défaut est fixé à 500 (Liaw et Wiener, 2002). Il est bien connu que dans la plupart des cas, RF fonctionne raisonnablement bien avec les valeurs par défaut des hyperparamètres (Biau et Scornet, 2016). Néanmoins, l'ajustement des paramètres peut améliorer les performances de RF (Probst *et al.*, 2019). Ces ajustements peuvent se faire en faisant varier ces hyperparamètres et en sélectionnant celles permettant d'avoir les meilleures performances du modèle et cette

---

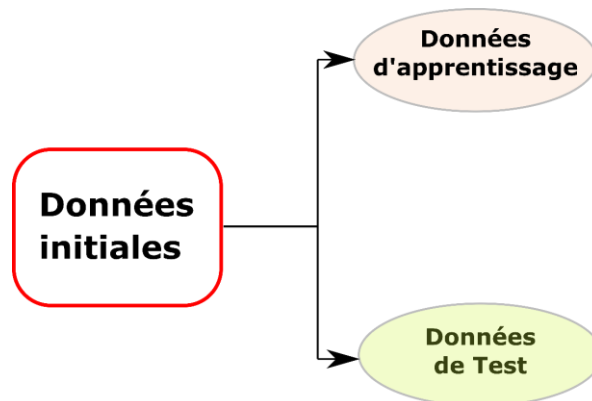
<sup>27</sup> <https://www.ibm.com/docs/fr/cognos-analytics/11.1.0?topic=terms-gini-impurity-measure>

technique est connue sous le nom de « *grid search* » dans la littérature. Des études ont montré que la précision de la classification est moins sensible à *n*tree qu'au paramètre *mtry* et que l'optimisation de ce dernier ne peut entraîner qu'une légère amélioration des performances (Belgiu et Drăgut, 2016 ; Kuhn et Johnson, 2013).

### I - 4 - 3 Méthode de rééchantillonnage

Dans le but de mieux estimer les performances du modèle, l'échantillon initial est très souvent séparé en données d'entraînement « *Training dataset* » et en données de validation « *Test dataset* » selon la **Figure I-18** avec les objectifs suivants :

- **Données de calibration / entraînement / apprentissage** : ces données sont utilisées pour développer le modèle RF, ajuster les paramètres, comparer les modèles et choisir un modèle final ;
- **Données de validation / test** : le modèle entraîné ou calibré est appliqué sur ces données de validation pour évaluer ces performances.



**Figure I-18** : Répartition des données en ensembles d'apprentissages et de tests.

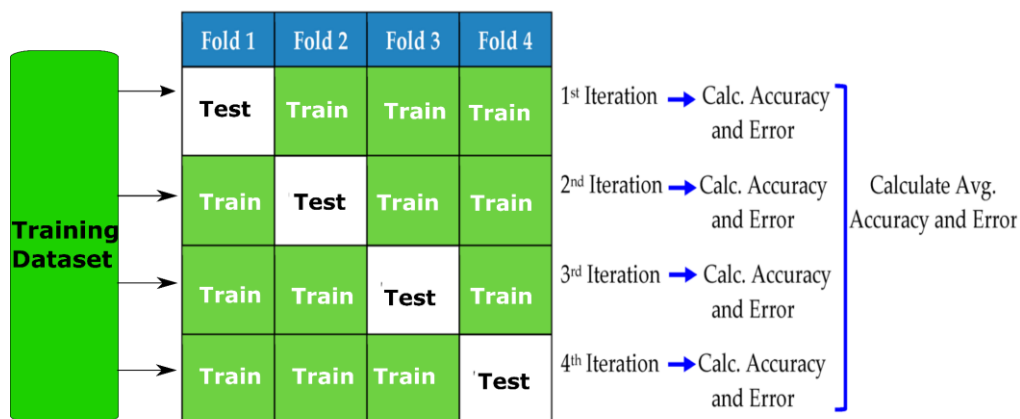
Les proportions fréquemment utilisées pour entraînement / validation des données sont respectivement de l'ordre 60% - 40% ; 70% - 30% ; 80% - 20% et même 90% - 10% (Hengl *et al.*, 2017). Dans les faits, ce seuil de répartition est laissé à l'appréciation de l'utilisateur en fonction du problème à résoudre et de la taille de l'échantillon initial.

Les deux façons les plus courantes de répartir les données en entraînement / validation sont la répartition aléatoire et la répartition stratifiée. Dans la répartition aléatoire, les données sont sélectionnées en ne tenant pas compte de leurs attributs tels que la distribution de la variable de réponse (variables à prédire). Dans la répartition stratifiée, les données peuvent être sélectionnées de façon à ce que les données d'entraînement et validation aient des



distributions similaires (proportion) de la variable de réponse. Avec RF cela est couramment utilisé en classification où les classes peuvent avoir des proportions très différentes (exemple 80% des observations sont classées permanentes vs 20 % qui sont classés intermittentes). La répartition stratifiée permet alors d'avoir ces mêmes proportions de classe dans le jeu de données en calibration et en validation ce qui évite de créer des biais dans la répartition des données.

Les performances du modèle sur un seul jeu de validation peuvent être très variables et peu fiables, si la taille de l'échantillon initial est faible (Molinaro *et al.*, 2005). Les méthodes d'échantillonnage telles que « *k-fold cross validation* » fournissent une alternative en permettant d'ajuster de manière répétée un modèle à des parties des données d'entraînement et de tester ses performances sur d'autres parties des données d'entraînement. La validation croisée *k-fold* est une méthode de rééchantillonnage qui divise aléatoirement les données d'apprentissage en  $k$  groupes ou blocs (*k folds*) de taille approximativement égale. Le modèle est entraîné sur les  $k-1$  blocs et ses performances sont évaluées sur le seul bloc restant. Cette procédure est répétée  $k$  fois en interchangeant les blocs de validation de telle sorte à ce que tous les blocs passent une fois en validation et à chaque bloc de validation l'erreur du modèle est calculée. Ainsi, une estimation de l'erreur du modèle est calculée en faisant la moyenne des  $k$  erreurs sur les blocs de test, ce qui nous fournit une approximation de l'erreur à laquelle nous pouvons nous attendre (**Figure I-19**). Les valeurs de  $k$  couramment utilisées sont  $k = 5$  et  $k = 10$ , car empiriquement ces valeurs sont connues pour estimer des erreurs du modèle qui n'ont pas un grand biais et une forte variance. D'un point de vue pratique, plus  $k$  est défini grand, plus le temps de calcul augmente.



**Figure I-19** : Illustration d'une validation croisée à  $k$  blocs avec  $k = 4$ . Cette figure est adaptée de Ali *et al.* (2017).

Dans certains cas en classification, un déséquilibre se produit lorsqu'une ou plusieurs classes à prédire ont des proportions très faibles dans les données d'entraînement par rapport aux autres classes (Hoens et Chawla, 2013). Deux approches généralement utilisées pour traiter ce problème sont le sous-échantillonnage et le suréchantillonnage des données d'apprentissage. Le suréchantillonnage désigne toute technique qui simule des points de données supplémentaires (sur les classes minoritaires) pour améliorer l'équilibre entre les classes, tandis que le sous-échantillonnage désigne toute technique qui réduit les données de l'échantillon initial (notamment sur la classe majoritaire) pour améliorer l'équilibre entre les classes. Il existe plusieurs techniques pour suréchantillonner et sous-échantillonner les données de sorte à avoir des proportions à peu près égales des classes. Dans cette étude, pour les cas de données déséquilibrées en classification, nous avons appliqué une technique de sous échantillonnage qui consiste à échantillonner de façon stratifiée les différentes classes dans la construction des arbres au sein de RF (Kuhn et Johnson, 2013, chapitre 16) de sorte à obtenir à peu près le même nombre de valeurs par classe et une technique de suréchantillonnage appelée SMOTE (*Synthetic Minority Over-sampling TEchnique*) (Chawla *et al.*, 2002) qui consiste à créer des données synthétiques pour les classes minoritaires à l'aide des données initiales. Ainsi, pour chaque point de données, la méthode recherche parmi les  $k$  voisins les plus proches et crée un nouveau point de données dans le segment de ligne qui relie le point original et chacun de ses voisins (Khorshidi et Aickelin, 2020). Cette dernière technique est fréquemment utilisée dans le milieu médical pour traiter les cas de données déséquilibrées (Zeng *et al.*, 2016).

#### I - 4 - 4 Mesure de l'importance des variables et sélections des variables prédictives

Dans la mise en place du modèle RF, il est possible d'évaluer l'ordre d'importance des variables environnementales dans la prédiction de la variable de réponse. L'importance est évaluée à travers deux mesures fondamentalement différentes : l'importance de Gini et l'importance par permutation (Schwarz *et al.*, 2010). Dans l'approche d'importance par permutation, en classification, une variable est importante si elle a un effet important sur « l'accuracy » de la classification. L'accuracy correspond à la proportion de classes correctement classifiées. La métrique utilisée pour apprécier cet effet est la décroissance moyenne de l'accuracy « *Mean decrease accuracy (%)* » qui correspond au pourcentage moyen perdu en « accuracy » si la variable en question n'est pas prise en compte dans

l'entraînement du modèle. En régression la métrique utilisée est le pourcentage d'augmentation de l'erreur quadratique moyenne « *percent increase in mean square error (%incMSE)* ». L'approche pour calculer ces paramètres pour une variable  $x_j$  est la suivante :

1. Construire un arbre  $T_b$  comme décrit ci-dessus (**Figure I-17**) et déterminer l'*accuracy* ( $A_b$ ) de prédiction dans les échantillons OOB data.
2. Permuter aléatoirement les valeurs de la variable prédictive  $x_j$ , rompant ainsi l'association de la caractéristique  $x_j$  avec la variable de réponse.
3. Déterminer à nouveau l'*accuracy* ( $A_b^*$ ) de prédiction dans les échantillons OOB, et calculer la différence d'*accuracy* de prédiction entre les échantillons *bootstrap* originaux et l'échantillon *bootstrap* après permutation de la variable  $x_j$ .

Cette différence moyennée sur tous les arbres  $T_b$ , appelée MDA (*Mean Decrease Accuracy*), est utilisée comme une mesure de l'importance de la variable  $x_j$  :

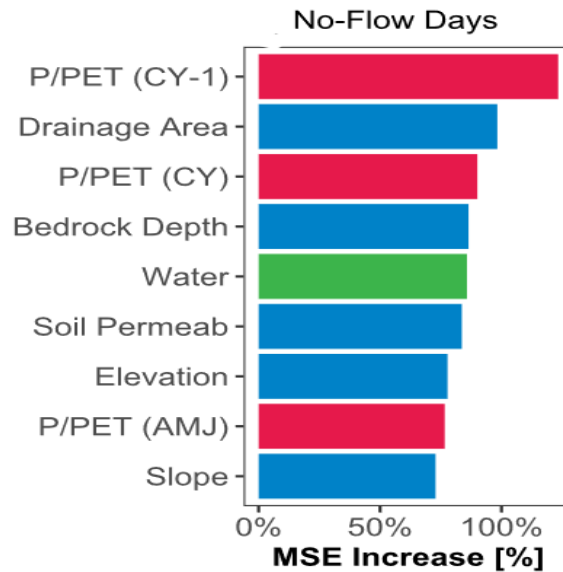
$$MDA = \frac{1}{B} \sum_{b=1}^B A_b - A_b^* \quad (I-10)$$

avec  $B = ntree$  qui est le nombre d'arbres à construire.

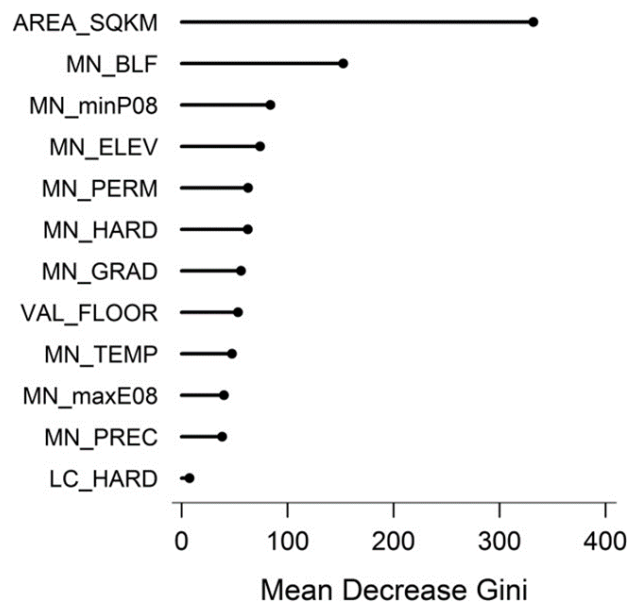
Cette procédure est répétée pour toutes les variables investiguées comme variables explicatives dans RF afin de pouvoir estimer l'ordre d'importance des variables.

La même approche est utilisée en régression. Le classement des variables les plus importantes au moins importantes se fait par ordre décroissant du MDA en classification et (*% incMSE*) en régression (**Figure I-20**).

En classification, l'importance des variables dans le modèle est aussi très souvent appréciée à travers l'indice de Gini : plus le nœud d'un arbre est pur, plus l'indice de Gini est petit, ce qui indique qu'un nœud contient des observations provenant principalement d'une seule classe. La métrique MDG (*Mean Decrease Gini*) rapporte la diminution totale en moyenne sur tous les arbres des impuretés des nœuds résultant du fractionnement à partir de chaque variable explicative (Breiman, 2001). Au final, une valeur élevée de cet indice MDG indique une grande importance de la variable dans la classification (Hastie *et al.*, 2009), comme illustrée en **Figure I-21**.



**Figure I-20** : Exemple de la représentation de l'importance des variables (% incMSE) dans un modèle random Forest en régression, calibré pour prédire le nombre de jours moyens sans débit. Cette figure est extraite de Zipper et al. (2021).



**Figure I-21** : Exemple de la représentation de l'importance des variables (Mean Decrease Gini) dans un modèle Random Forest calibré pour prédire l'intermittence des cours d'eau (González-Ferreras et Barquín, 2017).

Dans la littérature, le MDG est couramment préféré à MDA dans la mesure de l'importance des variables en classification, car les classifications basées sur le MDA sont instables aux petites perturbations de l'ensemble de données et que les classifications basées sur le MDG fournissent des résultats plus robustes (Calle et Urrea, 2011).

Des graphiques de dépendance partielle<sup>28</sup> peuvent être utilisés pour évaluer individuellement la relation entre les variables explicatives les plus importantes avec la variable de réponse.

<sup>28</sup> <https://christophm.github.io/interpretable-ml-book/pdp.html>

Bien que ces graphiques ne soient pas une représentation parfaite des effets produits par chaque variable, ils fournissent une base utile pour comprendre la relation entre la réponse et chaque prédicteur (variables hydro-environnementales) (Álvarez-Cabria *et al.*, 2016 ; Apley et Zhu, 2020).

Les variables à variance nulle ou presque nulle, c'est-à-dire les variables qui ont des valeurs presque similaires, fournissent peu d'information utile à un modèle. En outre, elles peuvent poser des problèmes lors du rééchantillonnage, car il y a une forte probabilité qu'un échantillon donné ne contienne qu'une seule valeur unique (la valeur dominante) pour cette caractéristique. Il est donc conseillé d'éliminer les variables à faible variance dans la mise en place du modèle (Kuhn et Johnson, 2019).

Dans l'estimation de l'importance des variables, une simplification des variables est cruciale, car l'interprétation de l'importance des variables (facteurs de contrôle les plus importants qui expliquent l'intermittence par exemple) peut être biaisée en utilisant des variables trop corrélées (Boulesteix *et al.*, 2012 ; James *et al.*, 2013). Dans ce sens, Strobl *et al.* (2007); Amaratunga *et al.* (2008) ont montré que les corrélations fortes entre les prédicteurs peuvent avoir un impact significatif dans l'évaluation de l'importance des variables dans RF. Une des approches courantes consiste à observer à travers une matrice de corrélation les coefficients de corrélations entre les variables et pour toutes paires de variables très corrélées (selon un seuil de corrélation couramment défini entre 0,7 et 0,9), une seule variable est retenue selon sa relation avec les autres variables où la catégorie (topographie, climat, géologie) à laquelle elle appartient (González-Ferreras et Barquín, 2017).

Des outils de sélection automatique des variables explicatives non redondantes existent notamment « *Boruta* » (Kursa et Rudnicki, 2010), mais ne permettent pas une interprétation physique plus facile, car très souvent en boîte noire. D'un point de vue pratique, un modèle comportant de très nombreuses variables environnementales intercorrélées devient souvent plus difficile à interpréter et rallonge les temps de calcul et les techniques décrites ci-dessus devraient permettre d'aboutir à la mise en place d'un modèle parcimonieux.

#### I - 4 - 5 Critères d'évaluation de la précision du modèle

Les critères de performance du modèle en régression sont principalement basés sur le  $R^2$  (mesure de la proportion de variances expliquées par le modèle) et la RMSE (racine carrée de

l'erreur quadratique moyenne) dont la formulation est déjà présentée en équation (I-4). La proportion de variation expliquée par le modèle est dérivée comme suit (I-11) :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (I-11)$$

où  $n$  est le nombre d'observations en validation croisée,  $y_i$  la valeur de l'observation  $i$ ,  $\hat{y}_i$  la valeur prédite correspondante et  $\bar{y}$  la moyenne des observations. Un coefficient de détermination proche de 1 indique un modèle parfait, c'est-à-dire que 100% de la variation a été expliquée par le modèle. Il varie entre 0 et 1.

Afin d'évaluer le modèle en classification et donc par rapport aux prédictions sur les classes, une série de métriques issues de la matrice de confusion (tableau de contingence) a été utilisée. La matrice de confusion résume les performances de classification et fournit une représentation visuelle des classes réelles par rapport aux classes prédites par le modèle (**Tableau I-7**). Les cellules de la diagonale (en vert) indiquent les cas où les classes sont correctement prédites, tandis que les cellules hors de la diagonale (orange) illustrent le nombre d'erreurs pour chaque cas possible. Dans la matrice de confusion ci-dessous, chaque colonne représente la classe prédite tandis que chaque ligne représente la classe réelle ou observée.

**Tableau I-7** : Exemple de matrice de confusion pour une classification multiclasse. Le calcul des faux positifs (FP), faux négatif (FN), précision, recall de chaque classe ainsi que l'accuracy globale sont détaillés.

Matrice de confusion		Prédit				Faux Négatif (FN)	Recall
		Classe 1	Classe 2	Classe 3	Classe 4		
Obs e r v é	Classe 1	A	B	C	D	B + C + D	A/(B + C + D)
	Classe 2	E	F	G	H	E + G + H	F/(E + G + H)
	Classe 3	I	J	K	L	I + J + L	K/(I + J + L)
	Classe 4	M	N	O	P	M + N + O	P/(M + N + O)
	Faux positif (FP)	E + I + M	B + J + N	C + G + O	D + H + L	Accuracy globale = A + F + K + P / (somme des boîtes oranges et verts)	
Precision	A/(E + I + M)	F/(B + J + N)	K/(C + G + O)	P/(D + H + L)			

Vrais positifs (TP) ; 
  Vrais négatifs (TN) ; 
  Cas mal classés ; 
  Faux positifs (FP) ; 
  Faux négatifs (FN)

L'*accuracy* globale (**Tableau I-7**) (proportion des classes correctement identifiées) a été utilisée comme critère de performance et permet d'apprécier globalement les proportions de stations de jaugeage correctement prédites, mais ne permet pas de discriminer entre les différents types d'erreurs de classification et présente certains inconvénients dans le cas de données déséquilibrées où l'*accuracy* reflète plus celle de la classe majoritaire (Grandini *et al.*, 2020) ce qui nous a amenés à prendre en plus d'autres critères d'évaluation tels que le *recall*, la précision et le score F1.

En prenant la classe 1 comme exemple, la précision répond à la question : quelle proportion de l'échantillon prédit comme classe 1 est réellement de la classe 1 ? Au contraire, le *recall* aussi appelé sensibilité ou *true positive rate* répond à la question quelle proportion de l'échantillon observé comme classe 1 est correctement prédite ? Le score F1 est la moyenne harmonique de la précision et du *recall*. Les valeurs de toutes ces métriques oscillent entre 0 et 1 avec les meilleurs scores qui tendent vers 1. Ces métriques sont couramment définies pour les problèmes de classification binaire<sup>29</sup> où le résultat de la classe est soit "*Classe 1*", soit "*Classe 2*". Comme nous avons identifié 4 classes d'intermittence au **chapitre II** et au **chapitre III** et que nous traitons d'un problème multiclasse, nous avons donc calculé ces métriques en calculant TN, TP, FP, et FN de chaque classe séparément (Ali *et al.*, 2017). Les formules de ces critères de performance sont résumées par les équations (I-12), (I-13), (I-14) suivantes :

$$Score\ F1 = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (I-12)$$

$$Recall = \frac{TP}{TP + FN} \quad (I-13)$$

$$Precision = \frac{TP}{TP + FP} \quad (I-14)$$

Les classes d'intermittence définies dans cette étude (**chapitre II** et **chapitre III**) sont ordinales c'est-à-dire qu'il existe un ordre inhérent entre les classes et les erreurs de prédiction sont moins importantes si elles sont proches de la classe prédite. Pour tenir compte de ces

<sup>29</sup> [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)

types d'erreurs et comparer les différents modèles de RF explorés, nous nous sommes basés sur le Kappa pondéré (Cohen, 1968 ; De la Torre *et al.*, 2018) comme critère, car il tient compte de l'ordre inhérent entre les différentes classes par rapport à l'*accuracy* et au Kappa qui donnent le même poids à ces différentes erreurs.

Le Kappa pondéré<sup>30</sup> est une pondération du critère de Kappa par une matrice de coût (poids). La définition de la matrice de coût est faite par l'utilisateur en fonction du problème traité. Au **chapitre II** (section **II-8**) et **chapitre III** (section **III-3-4**) des détails sont fournis sur le choix des valeurs de la matrice de coût. Le Kappa pondéré se définit selon l'équation (I-15) :

$$K_w = 1 - \frac{\sum_{i,j=1}^n w_{i,j} P_{i,j}}{\sum_{i,j=1}^n w_{i,j} P_i Q_j} \quad (\text{I-15})$$

où  $w_{i,j}$  est le coût pour chaque paire  $i,j$  dans la matrice de coût ( $w$ ),  $P_{i,j}$  sont les valeurs observées dans la matrice des fréquences relatives ( $P$ ) pour chaque paire  $i,j$ . La matrice des fréquences relatives est obtenue en divisant les valeurs de chaque paire  $i,j$  de la matrice de confusion ( $O$ ) par le nombre total d'observations  $N$ ,  $P_i$  et  $Q_j$  sont respectivement les totaux marginaux des lignes et colonnes dans la matrice des fréquences relatives ( $P$ ) et se calcule selon l'équation :

$$P_i = \sum_{j=1}^n P_{i,j} \quad (\text{I-16})$$

$$Q_j = \sum_{i=1}^n P_{i,j} \quad (\text{I-17})$$

Les valeurs du Kappa pondéré varient entre -1 et 1 et s'interprètent comme le Kappa avec les valeurs qui tendent vers 1 qui traduisent une meilleure concordance entre les classes prédites et celles observées tandis qu'une valeur égale 0 traduit le fait que l'accord n'est pas meilleur que celui qui serait obtenu par hasard et une valeur  $< 0$  traduit le fait que l'accord est inférieur à l'accord attendu par hasard.

## I - 5 Conclusion

Dans ce **chapitre I**, la description et l'analyse préliminaire des données hydrographiques ont permis de sélectionner le réseau hydrographique LCS comme support pour la cartographie de

<sup>30</sup> <http://www.real-statistics.com/reliability/interrater-reliability/weighted-cohens-kappa/>



l'intermittence des cours d'eau à l'échelle de l'Afrique. Il est ressorti de l'analyse préliminaire des données que l'absence de définition ou des définitions différentes de l'intermittente est un des écueils à la comparaison des études sur l'intermittence. Le modèle *Random Forest* sera utilisé pour faire de la prédiction spatiale de l'intermittence le long du réseau hydrographique en Afrique et sera exploré en régression et en classification afin de choisir celui qui donne les meilleures performances. Pour traiter des données déséquilibrées, un sous-échantillonnage interne dans RF et la méthode de suréchantillonnage SMOTE sera exploré. Les critères de performances du modèle se baseront en régression sur le coefficient de détermination  $R^2$ , la RMSE et en classification sur une série de métriques issues de la matrice de confusion dont l'*accuracy* globale, la précision, le *recall*, le score F1. Pour tenir compte du degré d'erreur en fonction de l'ordre inhérent entre les classes, le critère de Kappa pondéré sera utilisé pour départager les différents modèles de RF explorés.

## Chapitre II - Comprendre les facteurs clés contrôlant la durée de l’intermittence du débit des rivières : cas du Burkina Faso en Afrique de l’Ouest



Le Nakanbé à Wayen - Burkina Faso : Différentes périodes d’un cours d’eau intermittent, **a)** période humide (BOYER Jean-François, 07/10/2002) ; **b)** période sèche (GUILLIOD Maurice, 27/03/2007) - source SIEREM<sup>31</sup>.

Dans ce chapitre, on s’est intéressé à la compréhension de la distribution spatiale de l’intermittence des cours d’eau jaugés au Burkina Faso ainsi que des facteurs de contrôle associés. Ce pays a été choisi à cause de la facilité d’accès aux données notamment hydrométriques, géologiques, hydrographiques, climatiques. La période investiguée va de 1955-1985 avec un nombre maximum d’observations de jaugeages disponibles (49). Une partie de ce chapitre a fait l’objet d’un article publié dans *Journal of Hydrology Regional Studies (EJRH)* (section **II-2** à section **II-6**, page 78 – 109) :

Belemtougri, A.P., Ducharne, A., Tazen, F., Oudin, L., Karambiri, H., 2021. Understanding key factors controlling the duration of river flow intermittency: Case of Burkina Faso in West Africa. *Journal of Hydrology: Regional Studies*, 37, 100908. <https://doi.org/10.1016/j.ejrh.2021.100908>

Les éléments complémentaires à l’article de journal sont présentés en section **II-7**, pages 109-122. Par ailleurs, à l’aide des variables clés identifiées dans ce chapitre, une prédiction de la distribution spatiale de l’intermittence au Burkina Faso est proposée en section **II-8**, pages 122-130 par la méthode *Random Forest* en préparation de son application à l’échelle de toute l’Afrique en **chapitre III**.

<sup>31</sup> <http://www.hydrosciences.fr/sierem/>

## II - 1 Résumé

Les rivières intermittentes cessent de couler ou s'assèchent complètement à divers moments et endroits ; ils représentent plus de 50% de tous les cours d'eau du monde. Cette étude se focalise principalement sur le Burkina Faso en Afrique de l'Ouest et vise à identifier les variables environnementales qui expliquent le mieux les variations géographiques du régime d'intermittence du débit, et en particulier la durée de l'intermittence des cours d'eau. Les données de débit de 49 stations de jaugeage ont été prises en compte, principalement localisées sur les grands cours d'eau. Une période de pré-développement (1955-1985) a été choisie afin que les régimes des rivières soient faiblement altérés tout en maximisant le nombre de stations disponible. La période 1955-1985 peut être divisée en deux sous-périodes : une période normale (1955-1969) suivie d'une période de sécheresse (1970-1985). Le nombre moyen de mois secs ( $\overline{Ndry}$ ) a été utilisé comme prédicteur pour définir quatre classes croissantes d'intermittence du débit, pour lesquelles les variables environnementales explicatives potentielles ont été évaluées sur la base d'une analyse de corrélation et d'une analyse en composantes principales (ACP). Les deux premières composantes (PC) représentent 82% de la variance totale. La PC1 (52%) est liée à l'ordre de Strahler, à la précipitation moyenne annuelle et à la surface des bassins versants. La plupart des bassins versants avec des classes d'intermittence similaires se regroupent selon l'axe PC2 (30%) qui est principalement lié à la perméabilité des bassins versants. Malgré l'importance de l'ordre de Strahler et de la précipitation annuelle dans la discrimination des rivières intermittentes, cette étude suggère que la perméabilité et la surface des bassins versants expliquent la variation géographique des classes d'intermittence du débit au Burkina Faso. Nous avons également montré que si les variations spatiales de la précipitation contribuent aux configurations d'intermittence observées, les contrastes à long terme de précipitation entre les sous-périodes normales et sèches semblent avoir une faible influence sur les classes d'intermittence, ce qui suggère une résilience importante des rivières dans le pays. Cette faible influence des précipitations peut être due à leur forte saisonnalité dans l'ensemble du pays, ce qui rend la rivière intermittente à moins que les propriétés du bassin versant ne l'empêchent, lorsque l'écoulement des eaux souterraines est important. À l'aide des variables clés identifiées (perméabilité, surface amont, précipitation moyenne annuelle et ordre de Strahler) nous proposons en analyse complémentaire (en section II-8) une cartographie de l'intermittence à l'échelle du Burkina Faso.

## II - 2 Introduction

Intermittent rivers are rivers that stop flowing or go dry at some points in space and time, while perennial rivers, under normal conditions, sustain a continuous flow throughout the year (Acuña *et al.*, 2014 ; González-Ferreras et Barquín, 2017). Most hydrological studies have long focused on perennial rivers rather than intermittent rivers (Meerveld *et al.*, 2020). Still, the latter also play an important role and deserve the same attention (Leigh *et al.*, 2016). Intermittent rivers may account for more than 50% of all the world’s rivers (Skoulikidis *et al.*, 2017) and make up a large proportion of headwater rivers (Ward *et al.*, 2020). In recent years, they have been the subject of growing interest, especially in their ecological functioning (Datry *et al.*, 2011 ; Wilding *et al.*, 2018). They are involved in the upstream-to-downstream transport of sediment and organic matter required for feeding fish and other aquatic species (Reid et Ziemer, 1994 ; Stubbington *et al.*, 2019), regulation of floods (Morin *et al.*, 2009), water quality processes (Dodds and Oakes, 2008), and in the share of greenhouse gas emissions (Datry *et al.*, 2018).

Faced with climate change challenges, many studies predict an increase in the number of intermittent rivers in many parts of the world for the coming years (De Wit, 2006 ; Jaeger *et al.*, 2014). In West Africa, Senegal and Mauritania may experience an increase in intermittency, while intermittency may be relieved further east (Döll and Schmied, 2012), as a result of contrasting precipitation changes between the western and eastern parts of Sahel over the 21st century (Gaetani *et al.*, 2020). These climatic trends could be modified by intense anthropogenic pressure on water resources, especially in arid and semi-arid environments of West Africa, leading to an increased threat to biodiversity and water supply.

In arid and semi-arid regions, most rivers are intermittent and are often the only significant water source for consumptive use (Jacobson et Jacobson, 2013 ; Wekesa *et al.*, 2020). In these regions, they are considered the primary groundwater recharge sources (Koïta *et al.*, 2017 ; Shanafield and Cook, 2014). People heavily depend on them, for instance for local irrigation (Acuña *et al.*, 2014 ; Haile Ghebremariam et van Steenbergen, 2006). The hydrological classification of rivers or catchments according to the characteristics of their flow regime is therefore critical, especially in these regions where the well-being of populations depends on the ability to manage water scarcity for food production (Castelli *et al.*, 2018 ; Gebrehiwot *et al.*, 2011 ; Zouré *et al.*, 2019).

In Africa and West Africa in particular, hydrological studies are hampered by data scarcity and poor gauging, especially about streamflow discharge (Rabba *et al.*, 2018). This is a serious methodological challenge and may explain the low number of studies (Berhanu *et al.*, 2015 ; Brousseau *et al.*, 2010 ; Gebrehiwot *et al.*, 2011 ; Perez-Saez *et al.*, 2017 ; Rivers-Moore *et al.*, 2020 ; Schacht, 2019), which focused on the classification of the hydrological regime of catchments in Africa.

A range of hydrological indices has been used to classify catchments according to their flow regimes. The main components described by these indices are the magnitude, timing, duration, rate-of-change, and frequency of flow events (Poff et Ward, 1989 ; Richter *et al.*, 1996). As Smakhtin (2001) described, intermittent and ephemeral streams are characterized by naturally extended periods of zero flow, which may generally be perceived as the ‘lower bound’ of low-flow in hydrology. The indices related to the duration and frequency of flow (e.g., mean, median, percent, and frequency of the number of days/months with flow or zero flow per year/season) are the most important to assess flow intermittency and are commonly used in the literature (Costigan *et al.*, 2017 ; Larned *et al.*, 2010).

Among these indices, the duration of zero-flow is of particular relevance as an indicator of water availability. It is also an essential characteristic for river ecosystems; and is considered a significant indicator of the potential aquatic species distribution and persistence that may inhabit a given environment (Datry *et al.*, 2012 ; Leigh et Datry, 2017 ; Vardakas *et al.*, 2017). For example, Snelder *et al.* (2013) used hydrological indices to assess flow intermittency in France based solely on the mean annual frequency of zero-flow periods and the mean duration of zero-flow periods.

In the literature, intermittent streams are referred to by various names such as temporary, ephemeral, irregular, non-permanent, non-perennial, episodic, seasonal, interrupted; Such profusion of terms can be confusing, as noted by Uys and O’Keeffe, (1997). In this way, Busch *et al.*, (2020) suggest using only the terms “non-perennial, intermittent, and ephemeral” to characterize these streams. The classification of streamflow intermittency in databases such as the NHD (National Hydrography Dataset) in the USA mainly focuses on field surveys and interviews with residents, but *in situ* flow observations often disagree with this classification (Hafen *et al.*, 2020). The quantitative criteria used to classify the streams vary among authors. Hedman and Osterkamp (1982), for example, classified streams as perennial, intermittent, and ephemeral when flowing water is present annually more than 80%, 10-80%, and less than

10% respectively. Reynolds *et al.* (2015) categorized rivers as strongly intermittent when the mean number of zero-flow days/year >20, weakly intermittent when the mean number of zero-flow days/year < 20, and perennial when the mean number of zero-flow days/year equals zero. Yu *et al.* (2018) defined perennial streams as those that ceased flowing < 10% of the year (0–1 month), weakly intermittent as those that ceased flowing 10–70% of the year (2–8 months), and strongly intermittent as those that ceased flowing >70% of the year (>8 months).

Despite a handful of studies conducted in several regions regarding the identification and classification of intermittent rivers, an adequate understanding of the drivers controlling intermittency is still lacking (Costigan *et al.*, 2015 ; Tooth, 2000). We can cite several studies throughout the world that focused on identifying flow indices and potential environmental variables that best explain the different aspects of the flow regime (Alcázar et Palau, 2010 ; Assani *et al.*, 2006 ; Bejarano *et al.*, 2010 ; D’Ambrosio *et al.*, 2017 ; Gallart *et al.*, 2012 ; Gebrehiwot *et al.*, 2011 ; Kennard *et al.*, 2010 ; Lane *et al.*, 2017 ; Moliere *et al.*, 2009 ; Oueslati *et al.*, 2015). However, few studies have focused on assessing the main drivers of flow intermittency (Godsey and Kirchner, 2014). In their study, Assani *et al.*(2006) assessed the factors influencing the spatial variability of annual minimum flow characteristics in the southern part of the St. Lawrence River watershed in Quebec. They showed that the spatial difference of annual minimum flow characteristics between the north and south shore is mainly due to climatic factors, particularly annual precipitation. Reynolds *et al.* (2015) found that the most influential landscape drivers of flow intermittency and low flows are precipitation, snowfall, potential evapotranspiration, soil type, and drainage area. According to Trancoso *et al.* (2017), the controlling factors of zero flow ratio (proportion of zero flows in the full flow record) across different regions were aridity index and precipitation. However, they also pointed out that precipitation does not seem to impact the zero-flow ratio in some tropical areas, which are more sensitive to groundwater recharge and baseflow processes. Overall, it appears that the spatial and temporal variation of streamflow intermittency results from the interaction between several environmental variables.

At the global and regional scale, climatic variables, especially the aridity index, are likely to be the first-order control of river flow intermittency (Costigan *et al.*, 2017 ; Trancoso *et al.*, 2017). However, river flow intermittency is also controlled by processes acting at smaller scales than climate drivers (Snelder *et al.*, 2013). Local factors such as geology, topography, and land use/land cover are important variables that also control the river intermittency and

operate at different spatial and temporal scales (Buttle *et al.*, 2012 ; Costigan *et al.*, 2016 ; Kennard *et al.*, 2010 ; Williams, 2006).

In West-Africa, there have been very few studies specifically dedicated to assessing the controlling factors of flow intermittency. Understanding the potential drivers of river intermittency should lead to better water resources management through effective water allocations following climate induced pressures on the intermittent flow regime.

This study aimed to investigate the role of climate, lithology, and topographic controls on geographical variations in the hydrological regime of rivers in Burkina Faso, with a particular focus on the duration of intermittency (periods with zero flow). To achieve this objective, the following research goals were carried out: (i) define a classification framework applied to river gauging observations in order to identify patterns in flow intermittency classes; (ii) evaluate the role of lithology, topography, and climate on flow intermittency classes through multivariate analysis; (iii) discuss their interactions and understand the spatial variability of flow intermittency.

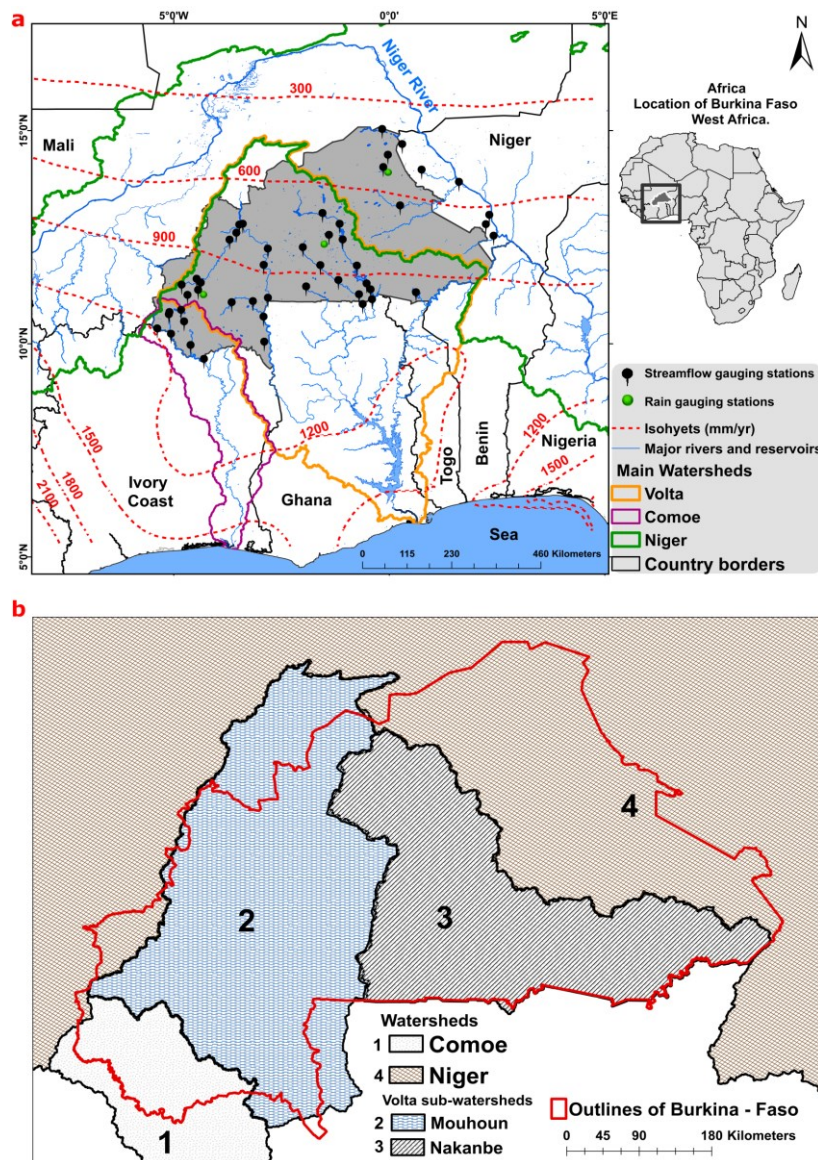
## II - 3 Data and Methods

### II - 3 - 1 Study Area

Located in central West Africa, Burkina Faso is a landlocked country covering an area of 274,200 km<sup>2</sup>. The altitude over half of the country ranges between 250 and 350 m above the sea level, whereas the highest values do not exceed 750 m, making the topography relatively flat. The country area is divided into three principal watersheds: Comoe (17 590 km<sup>2</sup>), Niger (83 642 km<sup>2</sup>), and Volta (172 968 km<sup>2</sup>) (**Figure II-1.a**). The Volta in Burkina Faso is subdivided into two sub-watersheds (Mouhoun and Nakanbe) (**Figure II-1.b**). The study area is characterized by two alternating and contrasting seasons influenced by the movement of the Inter-Tropical Convergence Zone (ITCZ), where the humid wind coming from the Atlantic Ocean in the South (Monsoon) meets the warm, dry wind coming from the Sahara in the North (Harmattan) (Ibrahim, 2012). Of the two seasons, there is a long dry season with almost no rainfall from October to May and a short rainy season from June to September. Rainfall is consequently highly seasonal, with a peak generally occurring in August (PANA, 2007).

There is a climatic South-North gradient of decreasing precipitation and increasing temperature and aridity across the country. Three climatic zones are commonly identified in

the country based on the average annual precipitation (Jalloh *et al.*, 2013 ; PANA, 2007) : the Sahelian zone in the north (average annual rainfall less than 600 mm), the Sudano-Sahelian zone in the central region (average annual rainfall between 600 and 900 mm), and the Sudanian zone in the south (average annual rainfall between 900 and 1200 mm) (**Figure II-1.a**). These various climatic settings make the country a representative regional case study to investigate and understand the factors that control flow intermittency.



**Figure II-1** : Hydroclimatic characteristics of the study area. Burkina Faso is shown in dark grey. **a)** The main catchments’ boundaries are derived from the HydroSHEDS database (Lehner *et al.*, 2008). The main rivers and water reservoirs (light blue lines) are derived from the Digital Chart of the World (DCW) (ESRI, 1993). Isohyets (red dashed lines) represent average annual rainfall (mm/yr) calculated from Climatic Research Unit (Harris *et al.*, 2020) data at 0.5° spatial resolution over the period 1950-2018. The black and green dots give respectively the location of 49 streamflow gauges and 3 rain gauge stations provided by the hydrological and meteorological services; some gauges are superimposed in the representation due to their spatial proximity. **b)** Focus on the different watersheds and sub-watersheds of Volta (Mouhoun and Nakanbe) in Burkina Faso.



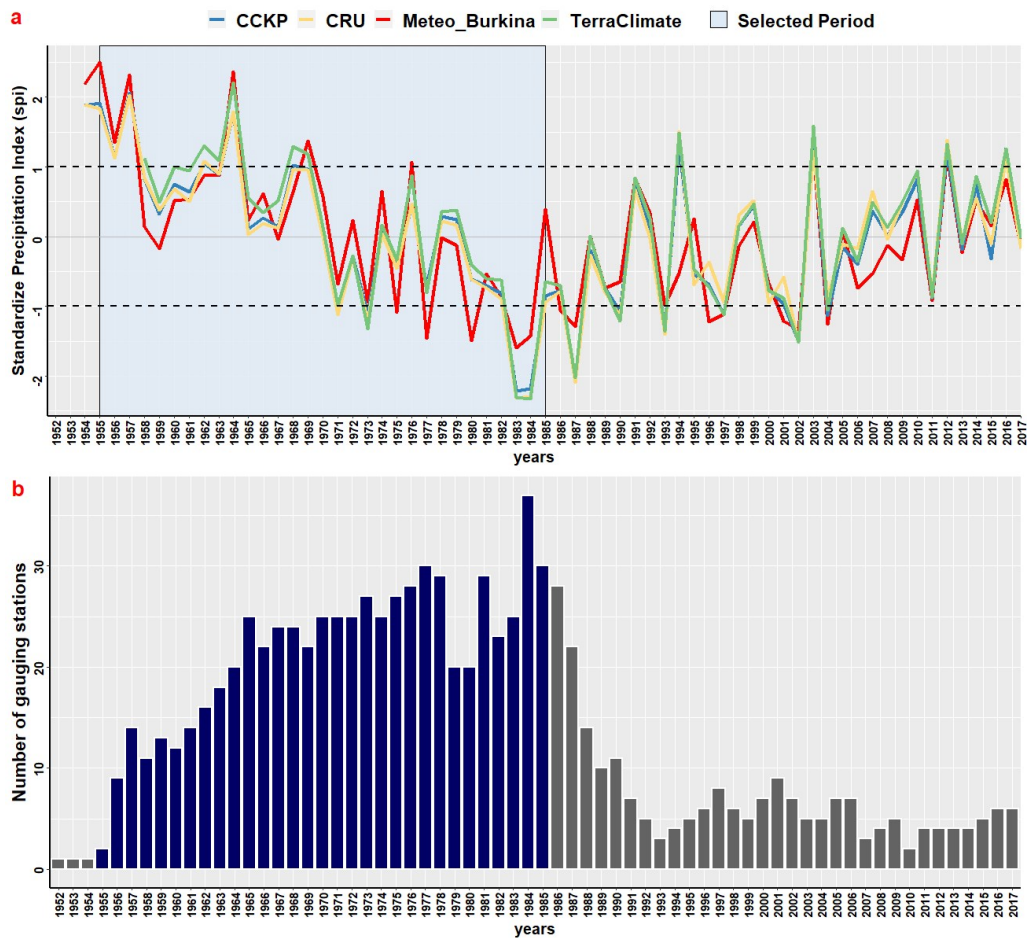
## II - 3 - 2 Hydrological and environmental data

Discharge data from 36-gauged stations, 16 stations with daily data, and 19 stations with monthly data were provided by the DGRE (*Direction Générale des Ressources en Eau*) in Burkina Faso over the period (1955-2017). In order to have more gauges, some global and regional hydrological databases were used to supplement the gauging stations obtained from DGRE and then provide additional information on gauging stations located in neighboring countries: Niger, Ghana, and Ivory Coast. To this end, we used monthly data from the Global Runoff Data Center (GRDC, 2019) for 45 gauged stations over the period (1951-1991), as well as daily data from the SIEREM (*Système d’Informations Environnementales sur les Ressources en Eau et leur Modélisation*) database (Boyer *et al.*, 2006) for 80 gauging stations over the period (1952-2006).

The complete dataset was checked and processed according to the following four criteria: (i) gauged stations with less than 5 daily missing discharge records values per month were considered and averaged to monthly values; (ii) for each gauge station, only the years with complete data (12 months) were considered; (iii) for duplicate gauged stations featured in different databases, only the one with the most extended record period was retained; (iv) finally, gauged stations with at least four years of complete monthly data were considered. Monthly time steps are often used to characterize river flow regimes (Gallart *et al.*, 2012 ; Yu *et al.*, 2018), particularly in data-scarce areas (Perez-Saez *et al.*, 2017). A set of 58 gauged stations over the period 1952 -2017 was finally constituted for further analysis.

A period of drought and desertification strongly affected West Africa in the 1970s and 1980s (**Figure II-2.a**). Thus, many studies pointed out a change in rainfall patterns from the year 1970 with a recovery around the 1990s (Barbé and Lebel, 1997 ; Ouedraogo *et al.*, 2002). Over the 58 gauges stations, the maximum number of available gauging stations per year is observed between 1955 and 1990 (**Figure II-2.b**). From 1985, the number of available gauging stations had sharply decreased, mostly restricted to large rivers. To avoid the anthropic influence of numerous dams constructed since the 1980s (Cecchi *et al.*, 2007), a pre-development period (1955-1985) was considered so that the river regimes are considered weakly altered. The period 1955-1985 can be subdivided into two sub-periods: a normal period (1955-1969) followed by a drought period (1970-1985). Over the period 1955-1985, 49 gauging stations with at least four years of complete monthly data were considered for this

study. Discharge data have an average availability of 14 years, and the majority of gauges (61.22 %) have at least ten years of data (**Figure II-A-1.a**).



**Figure II-2:** Normal and drought period from 1954 to 2017. **a)** The Standardized Precipitation Index (SPI) was calculated considering three rain gauge stations located on the country’s different climatic zones. The Dori rain gauge station (World Meteorological Organization Code: 1200010000) is located in the Sahelian zone; the Ouagadougou rain gauge station (WMO Code: 1200000100) is located in the Sudano-Sahelian zone; and the Bobo-Dioulasso rain gauge station (WMO Code: 1200004000) in the Sudanian zone. The SPI is calculated using the average annual precipitation of the three selected stations. The National Meteorology Agency (ANAM-BF) provided the meteorological data (Météo\_Burkina) in Burkina Faso. Other databases are considered, and precipitation values are extracted at the locations of the synoptic stations. CCKP data are taken from the Climate Change Knowledge Platform, <http://sdwebx.worldbank.org/climateportal/>; Climatic Research Unit (CRU) (Harris et al., 2014); TerraClimate (Abatzoglou et al., 2018). **b)** Distribution of the number of gauging stations available per year.

The river network used in this study was derived from the BNDT database in Burkina Faso (IGB, 2012) at a scale of 1 : 200 000. The Strahler’s order of the river network (Strahler, 1957) was determined using the RivEx tool (Hornby, 2020) under ESRI ArcGIS software (version 10.6.1). Catchments delineated from each gauging station were derived from HydroSHEDS (Lehner et al., 2008) digital elevation model (DEM) with 3’’ (approximately 90 m) resolution. The process was implemented in the R language (R Core Team, 2021) with an interface to GRASS GIS. The delineated catchment areas were compared with the

information of catchment areas provided in the GRDC and DGRE databases. Thus, 43 catchments are compared and showed similar areas ( $R^2 = 0.99$ ,  $p\text{-value} < 0.001$ , **Figure II-A-1.b**). The catchments studied cover 87% of the country’s surface area, of which more than 50% have a surface area between 1000-10000 km<sup>2</sup> (**Figure II-A-2**).

The permeability estimate in the study area was derived from lithological information (rock type). In Burkina Faso, the information on the main lithological formations, an updated version (2018) of Castaing *et al.* (2003) map, was acquired from the BUMIGEB service (*Bureau des Mines et de la Géologie du Burkina Faso*) at a scale of 1 : 1 000 000. To take into account the lithological information of some catchments that have a part of their area outside Burkina Faso, we used the GLiM lithological layer (Hartmann and Moosdorf, 2012), which is the most recent global layer available at the scale of 1 : 3 750 000. However, to ensure the nomenclature compatibility between the two databases, for each lithology type defined at the country scale in the BUMIGEB database, correspondence and reclassification criteria with the GLiM database were defined. This allowed taking advantage of the lithological precision of the geological map of BUMIGEB (scale 1: 1 000 000) compared to those of the global GLiM database (scale 1 : 3 750 000) (**Figure II-A-3**). Then, a permeability map was derived from the lithological map by assigning a single permeability value for each lithology type. As detailed in the supplementary (section II-7), the value of each lithology type is based from the permeability range by sediment or rock type (Chapter 6) of (Maidment, 1993) and on the fact that for a given climate and slope, prior research has found higher drainage densities in areas with lower permeability (Schneider *et al.*, 2017). For each catchment, the average permeability is obtained by weighting the permeabilities obtained for each lithology type by the area occupied by each lithology type in the catchment.

To investigate the impact of dams on river intermittency, we used information on available dams from 1947-1985 (section II-7), as extracted from the GRanD database (Lehner *et al.*, 2011). Over the period 1947-1985, the two most important dams are: the Lery dam (250 Mm<sup>3</sup>), built in 1976 for agricultural purposes in the Mouhoun watershed, with 4 studied gauges downstream of the dam; and the Kompienga dam (2000 Mm<sup>3</sup>) in the Nakanbe watershed, built in 1984 for hydroelectric purposes, with only one studied station on the river, far upstream of the dam (**Figure II-A-8**). A significant proportion of small dams (< 1 Mm<sup>3</sup>) are not listed in the GRanD database. However, all small dams together account for less than 4 % of national storage capacity (Cecchi *et al.*, 2009) and represent about 60.7 % of all dams in 1986 (Piquemal, 1991).

### II - 3 - 3 Flow index and intermittency classification

Generally, the flow duration is the primary descriptor of different stream classification (Fritz *et al.*, 2020 ; Kaplan *et al.*, 2019). Given the monthly time step chosen and for the sake of parsimony, the mean number of dry months per year ( $\overline{Ndry}$ ) which represents the duration of intermittency is the flow index selected in this study to assess intermittency and is widely used in the literature. A dry month is defined as a month for which all flows are equal to zero.  $\overline{Ndry}$  is a key metric relevant to river ecology (Vadher *et al.*, 2018) and is calculated according to equation (II-1) :

$$\overline{Ndry} = \frac{\sum_i^n Ndry(i)}{n} \quad (II-1)$$

where  $Ndry(i)$  is the number of dry months in a given year  $i$  and  $n$  is the number of considered years.

In this study, based on terminologies established by Uys and O’Keeffe (1997), rivers were categorized into four classes according to their mean number of dry months per year ( $\overline{Ndry}$ ) (Tableau II-1).

**Tableau II-1** :Flow intermittency classes and their description.

Flow intermittency classes	Description	$\overline{Ndry}$ (month/year)	Proportion of flow intermittency classes (%)
Perennial (P)	Sustaining a flow more than 90 % of the time.	(0-1)	26.53
Weakly intermittent (WI)	Characterized by stagnant and isolated pools of water during the dry season <sup>a</sup> .	(2-4)	16.33
Highly intermittent (HI)	Marked by a lack of surface water during the dry season.	(5-7)	26.53
Ephemeral (E)	Usually, small headwater streams that flow in direct response to precipitation events.	(8-12)	30.61

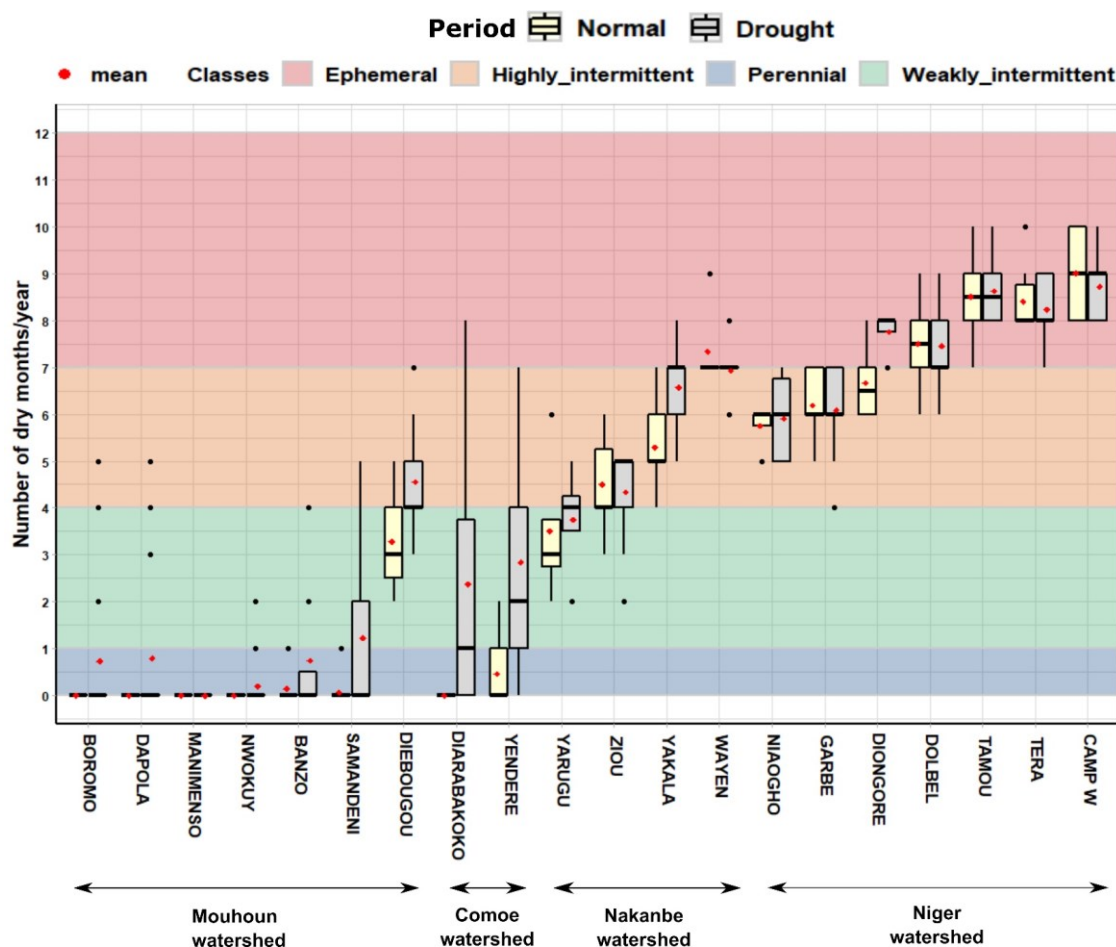
<sup>a</sup> The dry season is characterized by very little or no rain from October to May.

To get a long enough data record to characterize the spatial patterns of river flow intermittency in the study area, with a consistent number of gauges throughout the period, the monthly time step was preferred to the daily time step, and the flow intermittency classes obtained when comparing the common gauges (5) with daily and monthly time step data over the study period are similar (section II-7, **Tableau II-A-2**).

### II - 3 - 4 Assessment of the potential factors of intermittency

A range of variables describing environmental attributes (**Tableau II-2**) was considered as potential explanatory factors of intermittency. These variables were selected from the literature for their possible relevance in assessing the perennial or intermittent behavior of river streams (D’Ambrosio *et al.*, 2017 ; Kennard *et al.*, 2010 ; Moliere *et al.*, 2009 ; Snelder *et al.*, 2013) and also based on the availability and temporal resolution of the data. These include topography (5 variables), climate (5 variables), and lithology/bedrock permeability (1 variable). Land use/land cover (LULC) data were not included because no dataset could be found over the studied period (1955-1985). The oldest dataset of the national land use database, called BDOT (Base de Données d’Occupation des Terres), corresponds to the year 1992, and was shown to be of poor quality by (Cecchi *et al.*, 2009).

Anthropogenic factors might also influence the timing, magnitude, and duration of river flow (Hughes, 2005 ; Skoulikidis, 2009). The evolution of flow intermittency classes for the 20-gauge stations common to the normal (1955-1969) and drought (1970-1985) sub-periods shows that only 30% of these gauges exhibit a shift of flow class from the normal to the drought period. Besides, this shift is modest as it consists of only one intermittency class (**Figure II-3**, section II-7). This analysis indicates a weak influence of both dams and drought on the evolution of intermittency classes, and led us to neglect dams as potential explanatory factors.



**Figure II-3** : Boxplot of the number of dry months per year during the normal and drought periods (1955-1969 and 1970-1985, respectively) at the 20 common gauging stations. The outliers of the boxplots are displayed in black dots and the gauges are ordered by watershed.

All environmental variables were averaged for the catchments upstream of the gauging stations except for Strahler order, which is estimated at the gauging station location. The set of environmental variables was further reduced to avoid multicollinearity (by minimizing statistical redundancy among variables) through a Pearson correlation matrix. At this stage, for pairs of variables with a correlation above the threshold  $|r| > 0.75$  (considered to indicate a high level of correlation), only one was retained (). The retained variable is the one that is also correlated with multiple others (i.e., the variable with the highest absolute mean correlation with the other variables) and therefore allows a set of highly correlated variables to be reduced. It must be noted that evaporation related variables, including the aridity index, although often considered in similar studies, were not kept in this process since they show a strong correlation with precipitation (section II-7). A final reduced set of four significant environmental variables (independent variables) included in further analyses are presented in **Tableau II-2** (in bold).

**Tableau II-2** : Set of environmental variables considered in this study as potential flow intermittency controls. In bold, the independent variables selected for further analyses.

Type	Environmental variable	Description (units)	Data source, reference, and time scale
<b>Topography</b>	<b>Area</b>	<b>Catchment area (km<sup>2</sup>)</b>	<b>3” Digital Elevation Model (DEM) from hydroSHEDS (Lehner <i>et al.</i>, 2008)</b>
	Slope	Catchment average slope (°)	3” Digital Elevation Model
	Elevation	Catchment average elevation (m)	3” Digital Elevation Model
	Topo_wi	Catchment average of Topography wetness index (-)	Global DEM derivatives at 1 km based on the MERIT DEM (Hengl, 2018)
	<b>Strahler</b>	<b>River Strahler order of gauging station location (-)</b>	<b>Derived from national hydrographic database IGB (BNDT)</b>
<b>Climate</b>	<b>P</b>	<b>Catchment average annual precipitation (mm)</b>	<b>TerraClimate (Abatzoglou <i>et al.</i>, 2018) at the resolution of 4kmx4km and 1958-1985 period.</b>
	PET	Catchment average annual potential evapotranspiration (mm)	TerraClimate (Abatzoglou <i>et al.</i> , 2018) at the resolution of 4km x 4km and 1958-1985 period.
	AET	Catchment average annual actual evapotranspiration (mm)	TerraClimate (Abatzoglou <i>et al.</i> , 2018) at the resolution of 4km x 4km and 1958-1985 period.
	Aridity (PET/P)	Catchment average of aridity index (-)	TerraClimate (Abatzoglou <i>et al.</i> , 2018) at the resolution of 4km x 4km and 1958-1985 period.
	Tmoy	Catchment monthly average of mean air temperature (°C)	TerraClimate (Abatzoglou <i>et al.</i> , 2018) at the resolution of 4km x 4km and 1958-1985 period.
<b>Lithology</b>	<b>K</b>	<b>Catchment average permeability (m<sup>2</sup>)</b>	<b>Estimated from the lithological map (see Figure II-A-6)</b>

The principal component analysis (PCA) is a multivariate statistical method for reducing large dataset dimensionality, increasing interpretability while minimizing information loss (Jolliffe et Cadima, 2016). In this study, it was used to: (i) explore the relationship between significant

environmental variables (to show how they correlated to or deviated from each other); (ii) identify patterns in catchments with similar flow intermittency classes; (iii) analyze the relationship between the duration of intermittency ( $\overline{Ndry}$ ) and significant environmental variable. PCA was performed using XLSTAT (Addinsoft, 2015) based on the set of 49 catchments and four (4) environmental variables (in bold in Table 2). Before performing the PCA, the input variables were standardized to z-score values so that all variables have the same weight (van den Berg *et al.*, 2006). The number of meaningful Principal Components (PCs) was determined using Kaiser’s criterion (Kaiser, 1960) by selecting all components with eigenvalues greater than 1 (Assani *et al.*, 2006). The contribution of environmental variables to the explanation of a PC was considered significant if their loading value was at least 25%.

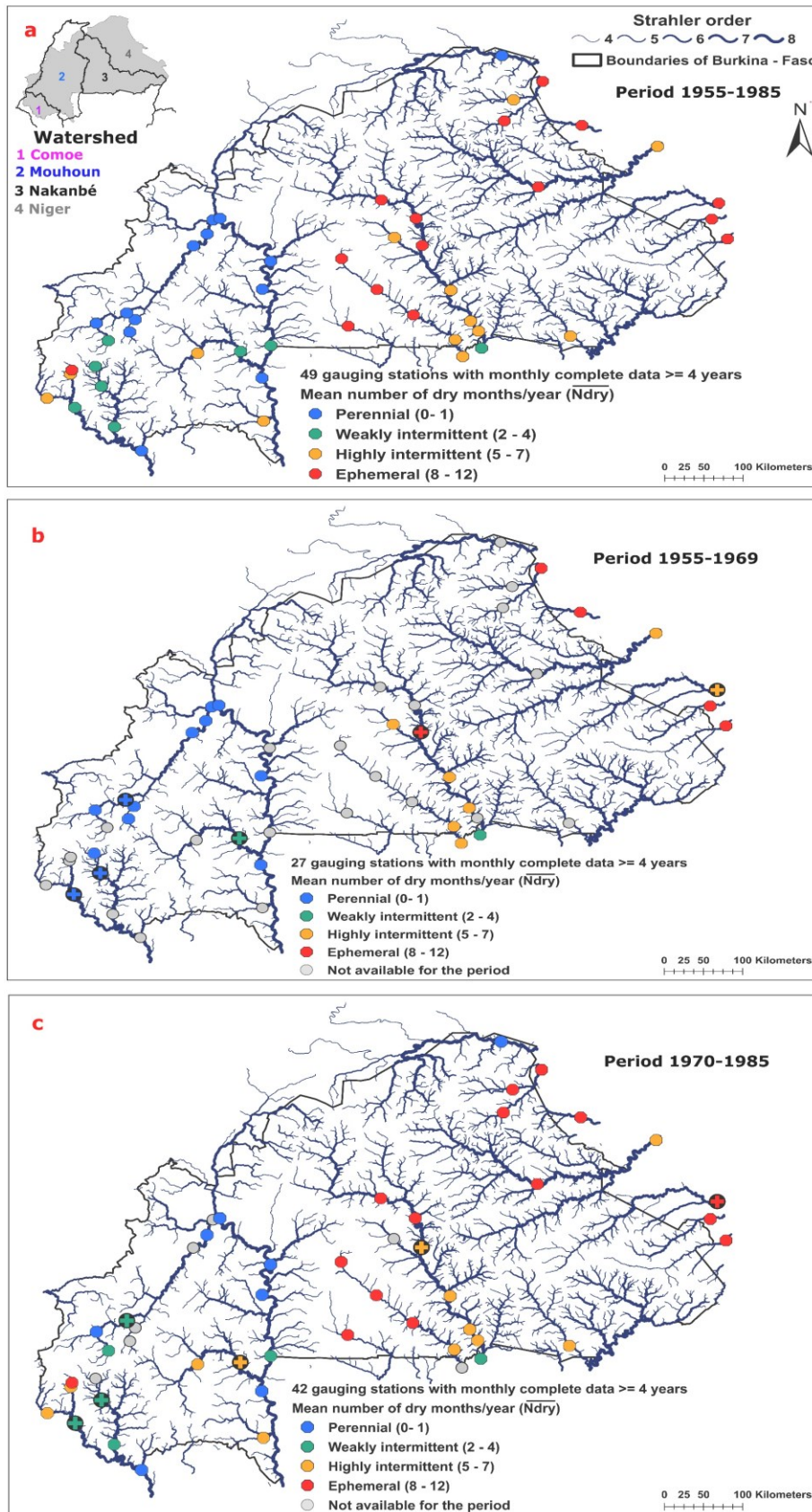
The duration of intermittency  $\overline{Ndry}$  was introduced as a supplementary quantitative variable once the PCA has been constructed. This supplementary quantitative variable is not included in the calculation of distances between individuals. It should facilitate the interpretation of the results and help detect the environmental variables that have the most significant impact on intermittency duration.

## II - 4 Results

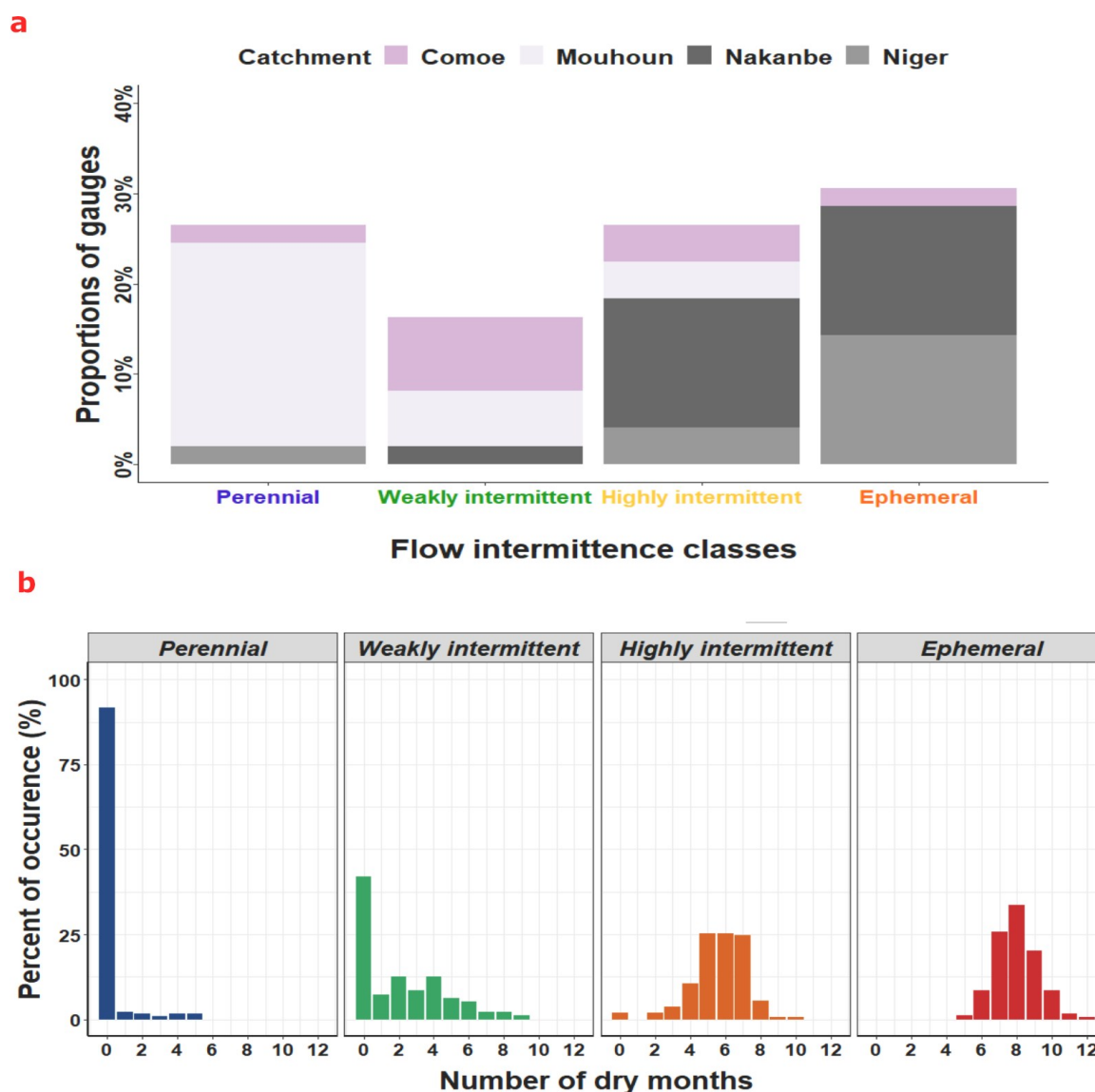
### II - 4 - 1 Flow intermittency classes

All studied stations fall on medium to large rivers, with a Strahler order between 4 and 8 (**Figure II-4.a**). Four flow intermittency classes were considered to categorize rivers in terms of intermittency (section **II-3-3**). It appears that 26.5 % of gauge stations are perennial, 16 % are weakly intermittent, 26.5 % are highly intermittent, and 30% are ephemeral (**Tableau II-1, Figure II-4.a**). The perennial and weakly intermittent gauge stations are mostly located in the south-western part of the country (Mouhoun and Comoe basins, **Figure II-4.a, Figure II-5.a**), where intermittent gauges are only found in upstream reaches, with medium Strahler orders (4-5). In contrast, the eastern part of the country (Nakanbe and Niger basins) shows a huge majority of highly intermittent and ephemeral gauges, whichever the stream order (**Figure II-4.a, Figure II-5.a**). **Figure II-5.b** shows that the classification based on the mean number of dry months per year remains effective at the annual timescale, since the number of dry months per individual year markedly increases from the perennial to the ephemeral class.





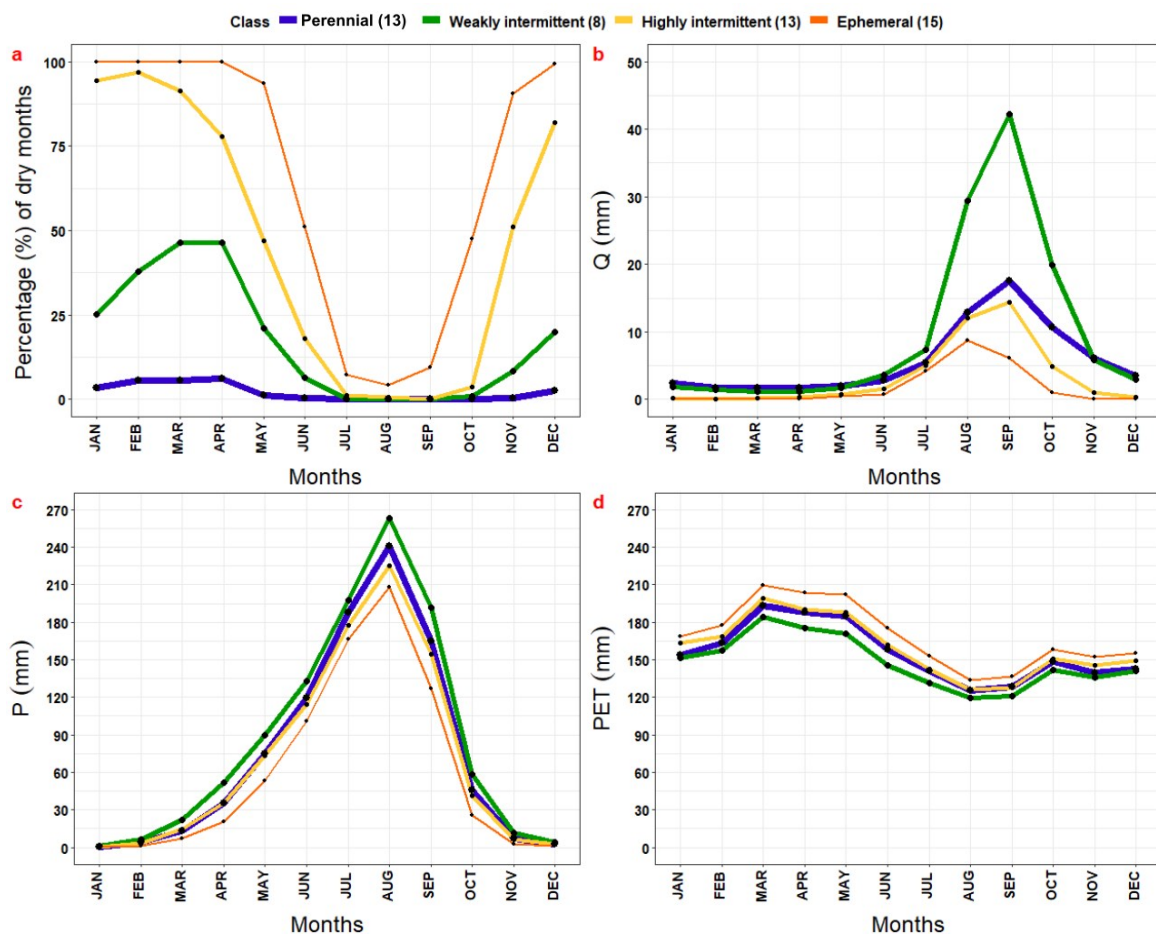
**Figure II-4** : Geographical variation of flow intermittency classes based on the mean number of dry months per year: **a**) over 1955-1985; **b**) over 1955-1969 (normal sub-period); **c**) over 1970-1985 (drought sub-period). In **b** and **c**, gauging stations that have a different intermittency class between both sub-periods are evidenced by crosses inside the circles. River line thickness is proportional to Strahler’s order.



**Figure II-5 :** **a)** Proportion of gauging stations within large basins and flow intermittency classes over 1955-1985. **b)** Distribution of the number of dry months for each individual year of each station belonging to an intermittency class over 1955-1985.

Among the 49 gauging stations studied over the period (1955-1985), 20 gauging stations are shared between the normal (1955-1969) and drought (1970-1985) periods. Among these 20 stations, only six have a different intermittency class over the two periods, with a shift of only one class towards the ephemeral side in response to the drought. In particular, three perennial gauges stations on medium rivers located in the south-west of the country become weakly intermittent (**Figure II-4.b, Figure II-4.c**). Overall, the intermittency patterns depicted by the three sets of stations over the full, normal, and dry periods (**Figure II-4**) are very similar, and we preferred to work on the full set (1955-1985) to better sample the spatial variability of the potential intermittency drivers.

The analysis of the percentage of dry months (**Figure II-6.a**), the average of monthly discharge (**Figure II-6.b**), and the average of monthly precipitation and potential evapotranspiration regime (**Figure II-6.c, Figure II-6.d**) of the country shows a marked distinction of values between classes. A difference can be noted in the dry month percentage per-flow class timing, duration, and magnitude. The ephemeral class is characterized by a lower percentage of dry months and a peak discharge in August (**Figure II-6.a, Figure II-6.b**), when the maximum monthly average precipitation is observed for all classes **Figure II-6.c**). This synchronicity of peak flow and peak precipitation is suggestive of a strong dependence of the ephemeral rivers on the annual precipitation regime.

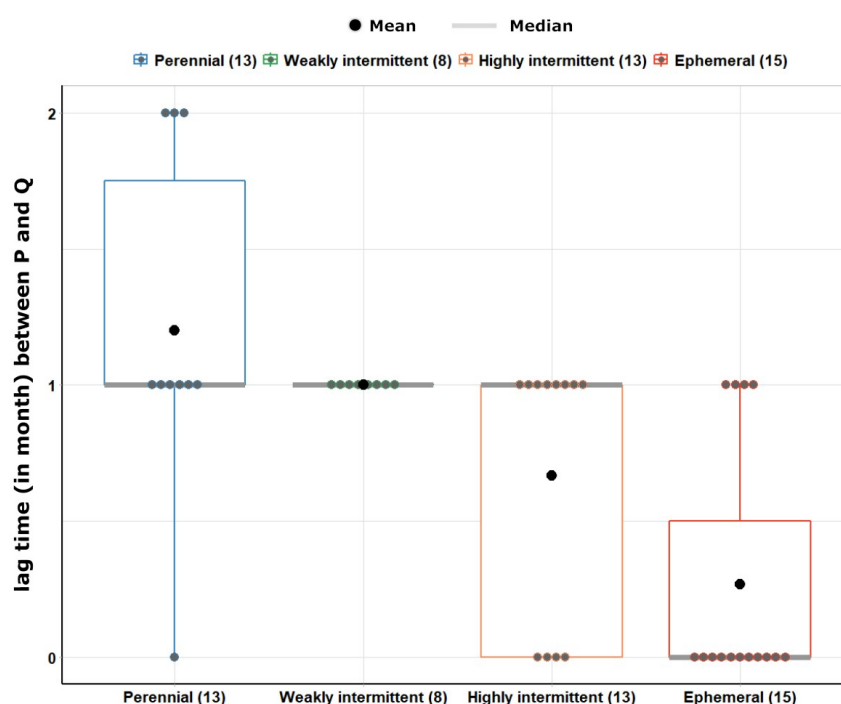


**Figure II-6 :** Percentage of dry months per class (a), the long-term average of monthly discharge per class (b), average precipitation regime pattern (c), average potential evapotranspiration regime pattern (d). The width of the graphic lines is proportional to the mean catchment area per class. The considered period is 1955-1985.

In contrast, for the three other classes, peak flow occurs in September, i.e., one month later than the maximum monthly precipitation. The highly intermittent, weakly intermittent, and perennial classes are also characterized throughout the year by a lower percentage of dry months than the ephemeral class. The lag time can be defined as the time between the center of mass of the precipitation excess and the peak discharge (Askew, 1970 ; Fang *et al.*, 2005).

This higher time lag between peak precipitation and peak flow than in the ephemeral class is probably explained by contrasting hydrological processes in the catchments of the different classes, especially given the large similarity of precipitation timing and intensity in all four classes.

The lag time depends on the catchment properties such as catchment surface area, land use/land cover, geology, soil type, and slope (Dingman, 2015). Generally calculated with a finer time step than monthly, a lag time can be determined in each catchment by analyzing the cross-correlation between the monthly average precipitation and the monthly average discharge (Dettinger and Diaz, 2000). In this framework, the lag time is the time lag (here in months), maximizing the correlation coefficient between the lagged monthly precipitation and monthly discharge. **Figure II-7** shows a decrease in the mean value of the lag time from the perennial class to the ephemeral class, and confirms that the most intermittent classes (highly intermittent and ephemeral) have shorter lag times than the other classes (weakly intermittent and perennial).



**Figure II-7** : Boxplot of estimated lag time value according to the intermittency classes over 1955-1985. The lag time (in month) is estimated for each catchment through a cross-correlation between the average monthly precipitation (P) and the average monthly discharge (Q).

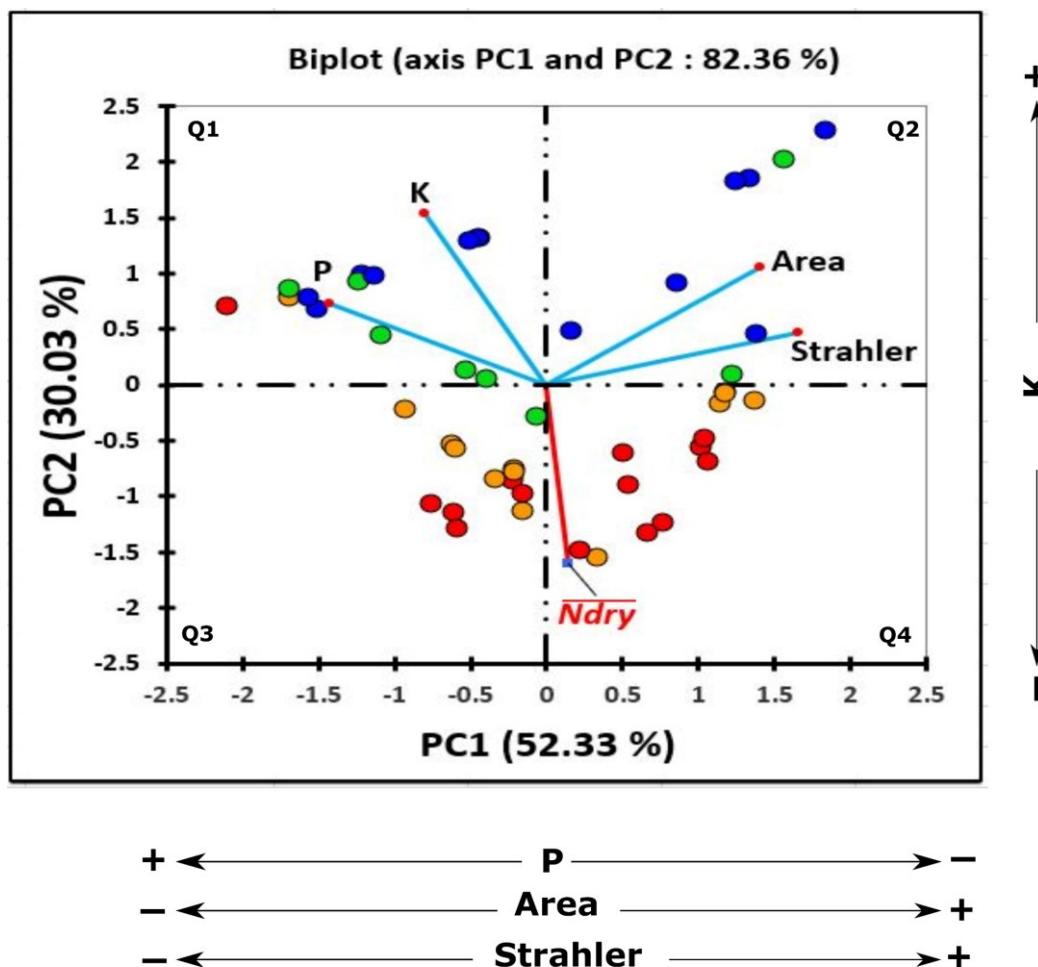
This shorter lag time for the ephemeral and highly intermittent class can be related to several catchment characteristics, such as catchment area and permeability. Indeed, the thickness of the plotted lines in **Figure II-6** is proportional to the average catchment area per class,

revealing its decrease from the perennial class to the ephemeral class. Most ephemeral and highly intermittent rivers are located in catchments of low permeability (**Figure II-A-4**), which may result in lower storage of water and a high runoff coefficient in these catchments (Pfister *et al.*, 2017).

**Figure II-6.b** also shows that highly intermittent and ephemeral classes produce discharge from early May to late November (up to December for a highly intermittent class). However, the perennial and weakly intermittent classes display a significant flow discharge throughout the year, probably due to groundwater contributions to streamflow (**Figure II-A-4**, **Figure II-A-6**). This is consistent with Buttle *et al.* (2012), who observe that the short duration response of ephemeral streamflow to precipitation inputs can be expected to produce relatively steep falling limbs of the hydrograph compared to perennial basins where groundwater inputs support flow. Differences in storage properties (permeability), contributing area, Strahler order, between catchment drained by perennial versus ephemeral streams, may also manifest in catchment response. The maximum periods of potential evapotranspiration (**Figure II-6.d**) coincide with low precipitation periods (**Figure II-6.c**), especially from December to May. The highest values of the percentage of dry months and the lowest discharge values are also observed from December to May. This period is favorable for the intermittency of rivers.

## II - 4 - 2 Principal Component Analysis (PCA)

The PCA carried out with the most significant environmental variables (P, K, Area, and Strahler order) showed that only the first two components have an eigenvalue greater than 1 (**Tableau II-A-3**). These first two components account for 82 % of the total variance (**Figure II-8**). The first component (PC1) accounts for 52 % of the total variance and is related to catchment topography and climate variables. PC1 axis is mainly explained by Strahler order (37%), followed by precipitation (28%) and catchment surface area (27%) (**Figure II-A-11**). The second component (PC2) accounts for 30 % of the total variance and is dominated by catchment permeability and catchment surface area, which contribute to the variance of the PC2 axis at 55% and 27%, respectively (**Figure II-A-11**).



**Figure II-8** : Biplot of the first two axes resulting from the PCA ( $n = 49$  catchments, over 1955-1985) for significant environmental variables and the flow index ( $\overline{Ndry}$ ) introduced as a supplementary variable. Q1 to Q4 designates quadrants 1 to 4. The catchments are classified as perennial (blue dots), weakly intermittent classes (green dots), strongly intermittent classes (orange dots), and ephemeral classes (red dots). Legend: P - average precipitation (mm/yr), Area - catchment area (km<sup>2</sup>), K - permeability (m<sup>2</sup>), Strahler - Strahler order (-). Some catchments are superimposed in the representation due to their spatial proximity.

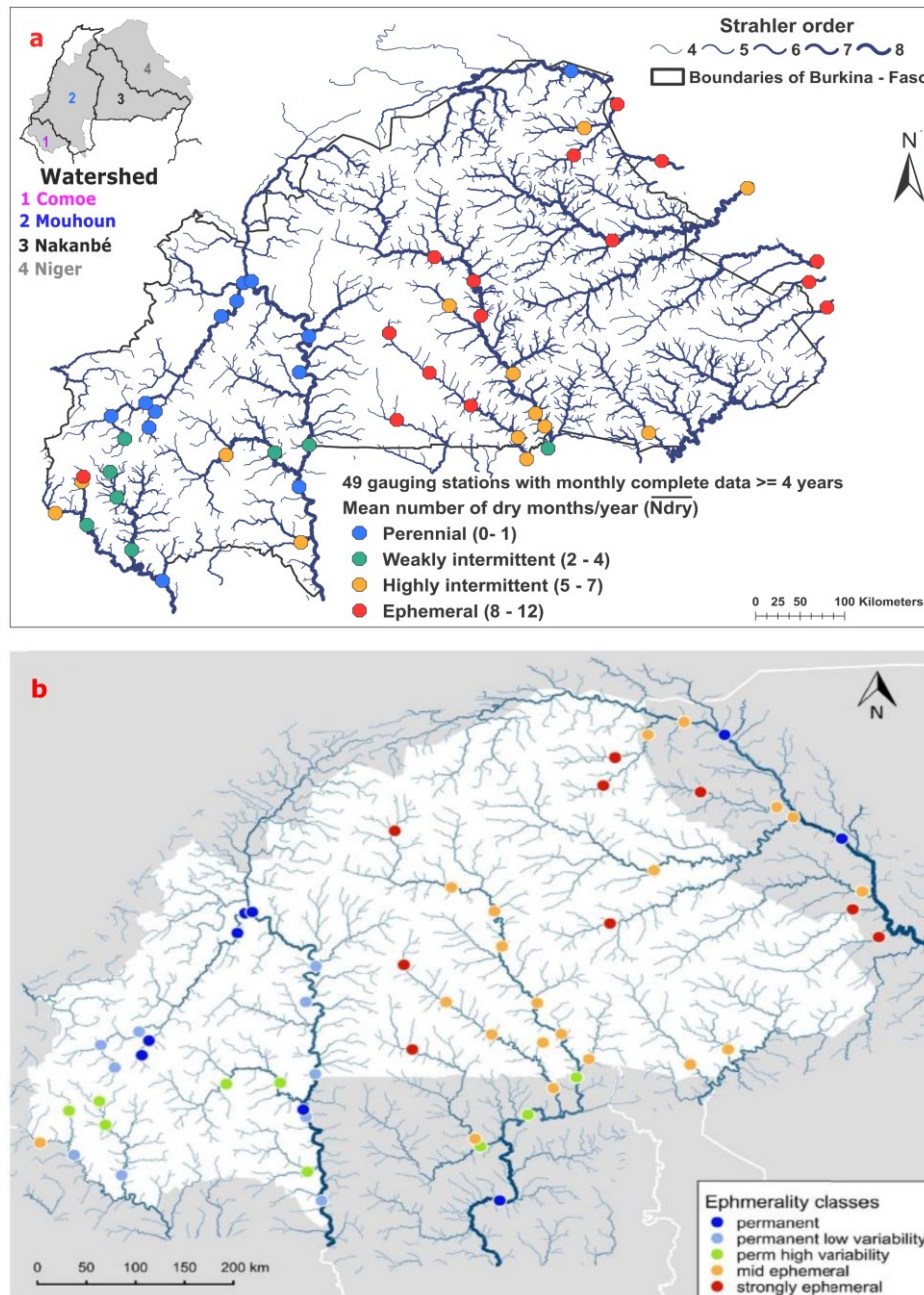
The flow index  $\overline{Ndry}$  added to PCA as a supplementary variable shows a high and negative correlation with permeability ( $r = -0.75$ ) (**Figure II-8**), PC2 ( $r = -0.85$ ) and PC1 ( $r = 0.08$ ) (**Figure II-8**, **Tableau II-A-4**). The PC2 axis is more related to flow intermittency classes than the PC1 axis, suggesting that permeability and catchment area are the most critical environmental variables in discriminating flow intermittency classes in Burkina Faso. **Figure II-8** also shows the positioning of the 49 catchments with their intermittency classes in the plane formed by the first two components (PC1, PC2). Catchments with higher values of Strahler order, surface area, permeability, or annual precipitation values are more likely to be classified into the perennial or weakly intermittent classes (**Figure II-8**, **Figure II-A-12**). On the other hand, highly intermittent and ephemeral classes correlate with lower values of Strahler order, surface area, permeability, or annual precipitation. One interesting result is an

apparent clustering of catchments located on the first two quadrants (above the first axis) compared to those located on the third and fourth quadrants (**Figure II-8**). Thus, the catchments with positive PC2 values (thus high permeability) are more likely to be perennial or weakly intermittent. Two exceptions, the Douna and Niofila catchments, display high permeability and very high rainfall values (resp. 1163 and 1166 mm/yr) but belong to the highly intermittent and ephemeral classes. Some studies have shown that streams located upstream of the river network are likely to have flow intermittently due to their size (Beaufort *et al.*, 2018 ; Wohl, 2017). Conversely, catchments with negative PC2 values (low permeability) are more likely to be highly intermittent or ephemeral.

## II - 5 Discussion

### II - 5 - 1 Comparison with previous studies

In the present study, there is a consistent geographical variation in the intermittency classes with stations that tend to be perennial, mostly located in the south-western part of the country. A previous study of Perez-Saez *et al.* (2017) focuses on modeling the dynamics of waterborne diseases in Burkina Faso based on hydrological classification and stream ephemerality prediction. An unsupervised clustering approach based on the number of zero flow months per year and for each gauge was used to determine the number of intermittency classes. Comparing the results obtained with those of Perez-Saez *et al.* (2017) on flow intermittency classes in Burkina Faso reveals a strong consistency, although the number of classes defined and classification methods remain different (**Figure II-9**). In general, no more than one intermittency class difference is observed between the two maps (**Figure II-9**). The significant differences lie in the number of gauged stations taken into account. Perez-Saez *et al.* (2017) only consider gauges stations (58) with at least three years of data, whereas in the present study, the selected stations (49) have at least four years of data. The period investigated is also different: 1955-1985 in this study 1960-1990 in Perez-Saez *et al.* (2017). These numerous elements listed above are probably at the origin of the discrepancies between **Figure II-9.a** and **Figure II-9.b**.



**Figure II-9** : Comparisons of intermittency classes in Burkina Faso: **a)** The present study over 1955-1985, **b)** Adapted from (Perez-Saez et al., 2017).

In Perez-Saez *et al.* (2017) study, the variables selected as input for ephemerality controls in the statistical model (Gradient Boosting Tree) were limited to aridity index, annual discharge, and remote-sensing based vegetation characteristics. However, no hydrogeological variable was considered, which constrained the conclusion that the most important driving variable of ephemerality was mean annual discharge.

Outside of Burkina Faso, Hammond *et al.* (2021) showed that climate, particularly the aridity index, is the dominant controlling factor of stream intermittency at the continental scale of the



USA for both human-impacted and non-impacted gauges stations. The effect of physiographic variables such as permeability and catchment area emerges more at a smaller scale. Similar conclusions can be drawn from the large-scale study of Sauquet *et al.* (2021), encompassing Australia, Europe, and the USA. The aridity index emerges as the climatic variable that best discriminates the different flow intermittency classes, although the control of physiographic variables on flow intermittency classes was not investigated. In the present study, we did not keep the aridity index as an explanatory factor, as it is strongly correlated to precipitation (**Figure II-A-13**), so we only kept the latter to reduce the number of autocorrelated variables (see section **II-3-4**).

## II - 5 - 2 Link between intermittency and catchment hydrology

One of the main results of the present study is that the geographical variation of intermittency classes in Burkina Faso shows a significant spatial coherence. The perennial and weakly intermittent classes are mostly located in the Comoe and Mouhoun catchments. In contrast, the highly intermittent and ephemeral classes are mainly located in the Nakanbe and Niger catchments (**Figure II-4.a**, **Figure II-5.a**). The upper Mouhoun catchment is entirely based on a sedimentary zone, and 86 water sources are identified; 77 sources are perennial (Dakoure, 2003), indicating a strong groundwater contribution to baseflows. Among these sources, there is a famous source called “*Nasso-Guinguette*”, which is the most important water source in West Africa (Huneau *et al.*, 2011). The Comoe catchment is partly sedimentary (about 20% of its surface area), and more than 100 perennial sources were identified (Dakoure, 2003). Many other studies showed the importance of spring flow rates in the sedimentary zone as contributing to the Mouhoun and Comoe rivers’ baseflow and their tributaries (Kouanda *et al.*, 2018 ; Koussoube, 2010 ; Ouédraogo, 1994 ; Sauret, 2013 ; Tirogo *et al.*, 2016). Kouanda *et al.* (2018) estimated at 45% the contribution of the average groundwater discharges to the total flow of Samendeni river (a tributary of Mouhoun river) during the rainy season in 2017. During the dry season, the Mouhoun and Comoe rivers are essentially fed by substantial inflows of baseflow (Pavelic *et al.*, 2012), explaining their high probability of being perennial or weakly intermittent.

Nakanbe and Niger catchments in Burkina Faso are mainly located on metamorphic rocks. Therefore, some studies (Mahé, 2009 ; Yameogo, 1988) have shown that the aquifer in the Nakanbe catchment is a fractured aquifer and very deep to be connected to the river. These

distinct hydrogeological conditions between the Mouhoun and Comoe catchments compared to the Nakanbe and Niger catchments could explain this observed geographical variation in flow intermittency classes.

The distribution of intermittent and perennial streams among the gauged rivers is not strongly dependent on the average annual precipitation because, as shown by the large proportion of perennial streams found in arid and semi-arid climates (Sahelian and Sudano-Sahelian zones) (Figure II-10). There, the precipitation effect is overruled by the mean catchment permeability, the catchment area, and Strahler order. This result is consistent with those obtained by Carlier *et al.* (2018) in the Swiss Plateau and Prealpes, who suggested that catchments with relatively permeable geological units have a high buffering potential on precipitation, and therefore, a significant water storage dynamic.

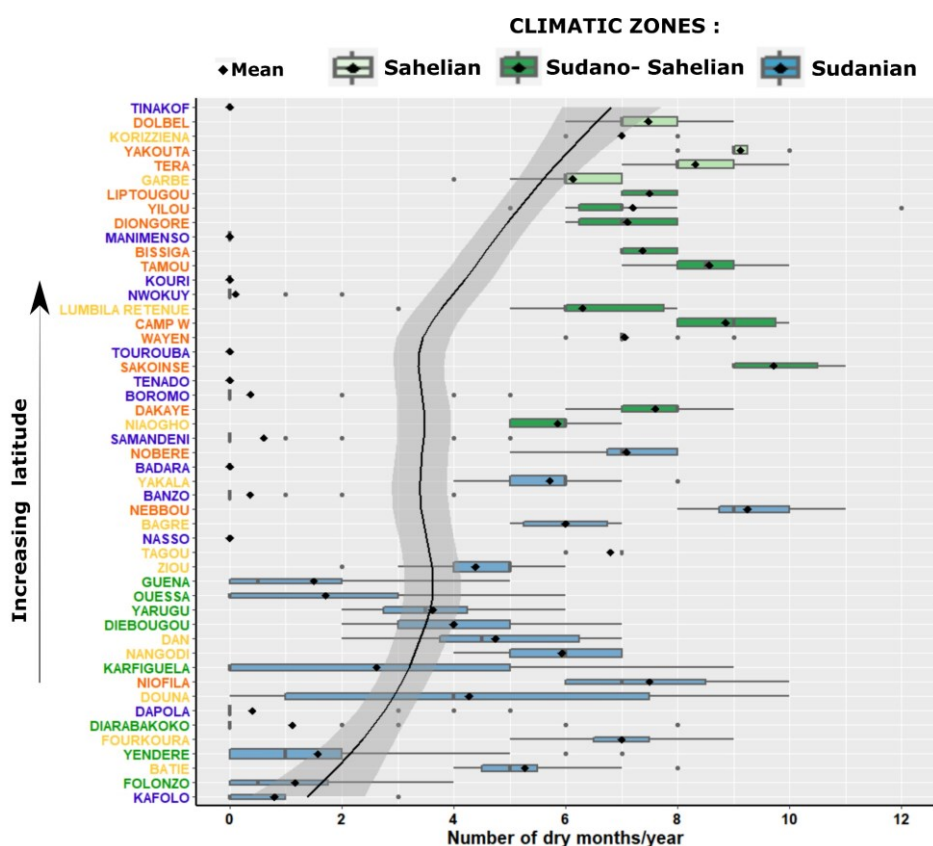


Figure II-10 : The observed number of dry months per gauging stations over 1955-1985. The middle line represents the median in each boxplot, whereas the black dot represents the average value. The boxplot thickness is proportional to the number of observations. The gauged stations are ordered according to the increasing latitude. The trend curve (black line), with a 95% confidence interval (grey zone), is obtained by local weighted least-squares smoothing of the raw counts of dry months. According to the climatic zones, the boxplots' color is defined from the Sudanian zone in blue to the Sahelian zone in light green. The color of gauging station names depends on their intermittency class: perennial (blue), weakly intermittent (green), highly intermittent (orange), ephemeral (red).

Our results also suggest that the catchments belonging to the ephemeral class have a shorter lag time than other classes, which may be due to their size (relatively small catchment area) and low permeability compared to other classes.

Among the non-anthropogenic variables identified in this study as intermittency controls, only precipitation is likely to change in the future, the other variables being relatively “static”. The comparison between the normal and drought sub-periods shows little sensitivity of intermittency classes to substantial variations of mean annual precipitation (**Figure II-3**). **Figure II-10** also shows that mean annual precipitation is not the primary factor explaining the distribution of perennial streams in the country, which means that rivers display a certain resilience against persistent droughts, as also evidenced by the similar spatial distribution of intermittency between the normal and dry sub-periods (**Figure II-3, Figure II-4**). The intermittency classes for 8 gauges stations common to the period 1955-1985 and the period 1986-2017 (with at least 4 years of data for each period) also shows little variation with 2 gauges stations that shift, one from weakly intermittent to perennial and the other from highly intermittent to weakly intermittent (**Figure II-A-13**), which can be explained by the recovery from the drought period (Lebel and Ali, 2009). This low sensitivity of rivers to annual rainfall amounts may be related to the very high seasonality of rainfall in the country: by default, river flow is seasonal too, whichever annual mean precipitation, unless particular catchment processes, here groundwater flow, make it perennial. This implies that the evolutions of the rainy season length and groundwater recharge processes may be more important for the evolution of intermittency in Burkina Faso than the ones of annual mean precipitation.

Future climate projections in West Africa suggest an increase of precipitation in the central and eastern parts, and a decrease in the western parts, with no clear indication about a possible shift in the length of the rainy season (Biasutti, 2019 ; Gaetani *et al.*, 2020). Although uncertainties increase when we zoom in, the above studies may imply a significant increase of precipitation in the eastern part of Burkina Faso. In this area of low permeability, this may not affect a lot the evolution of intermittency, which does not depend strongly on total precipitation over the 1955-1985 period and also does not show significant changes at current gauges analyzed. But some changes may not be detectable at the monthly time step. Nevertheless, this contrasts with the results of (Döll and Schmied, 2012), who predicted a shift of river flow regimes from intermittent to perennial in this eastern part of West Africa due to increasing precipitation under climate change. However, the increase of precipitation

may lead to increased floods, as already observed in the Sahelian part of the Niger basin (Casse *et al.*, 2016).

### II - 5 - 3 Limitation of the study and potential applications

In this present study, it is crucial to be aware that many shortcomings remain. Firstly, land use/land cover (LULC) were not considered as potential explanatory variables due to LULC data’s unavailability over the study period. Other studies also neglected their potential influence (Beaufort *et al.*, 2019 ; Snelder *et al.*, 2013), while Kaplan *et al.* (2020) found they were not dominant explanatory variables of no-flow metrics in a small catchment of Luxembourg. Dams, which are a form of LULC alteration, were also overlooked in our study, as we have shown that their influence on flow intermittency classes seems to be weak during the 1955-1985 study period (section II-7), but their effects may not necessarily be reflected at the monthly time scale. In Burkina Faso, the primary function of dams is to store water to make up for shortages during the dry season. Their effect might be felt at the beginning of the rainy season, when downstream flow is reduced, but once they are full, water flows through the spillway until the end of the rainy season. Only stations located downstream of huge dams could be influenced beyond the beginning of the rainy season, and become either more intermittent due to storage, or less intermittent if the dams are used for hydropower.

Additionally, the reclassification of geological information from the BUMIGEB database to the GliM’s nomenclature has undoubtedly led to some loss of information because certain types of lithologies have been grouped into one. Nevertheless, the use in Burkina Faso of the BUMIGEB lithological map at the scale of 1 : 1,000,000 to the complement of the GliM map at the scale of 1 : 3,750,000 is a better alternative to the single-use of GliM map (Hartmann et Moosdorf, 2012), which has a very coarse scale and lacks precision compared to the local BUMIGEB map. The permeability map is derived from this final lithology map, using permeability values within a range depending on rock type and observed drainage density. These assumptions potentially introduce uncertainty to the estimation of permeability. Although the estimated permeability map shows some similarities with the British Geological Survey’s (BGS) aquifer type/productivity map (section II-7), where areas mapped as having high aquifer productivity also generally exhibit high permeability, the estimated permeability remains uncertain because it is derived from coarse geologic maps.

In arid and semi-arid regions where intermittent streams are assumed to be predominant (Gallo *et al.*, 2020), stations tend to be located on large streams (Zimmer *et al.*, 2020), so our study is not representative of the majority of rivers in the country, smaller and assumed to be intermittent by the local authorities (Fovet *et al.*, 2021). The lack of hydrological data makes it impossible to estimate intermittency indices in ungauged catchments. Data loggers with sensors are technologies to measure the absence or presence of water continuously and are less expensive than traditional gauging stations to survey flow intermittency (Bhamjee *et al.*, 2016 ; Kaplan *et al.*, 2019 ; Peirce and Lindsay, 2015). They can help to better monitor headwater streams in numerous ungauged areas of West Africa. These new data can make it possible to study river flow intermittency at much smaller scales.

As a perspective, the identified flow intermittency controls in this study can notably be used in flow regionalization processes or as predictive variables to produce intermittency maps as performed by Snelder *et al.* (2013). However, particular care should be taken when extrapolating these control factors outside the study area, where other factors may be dominant.

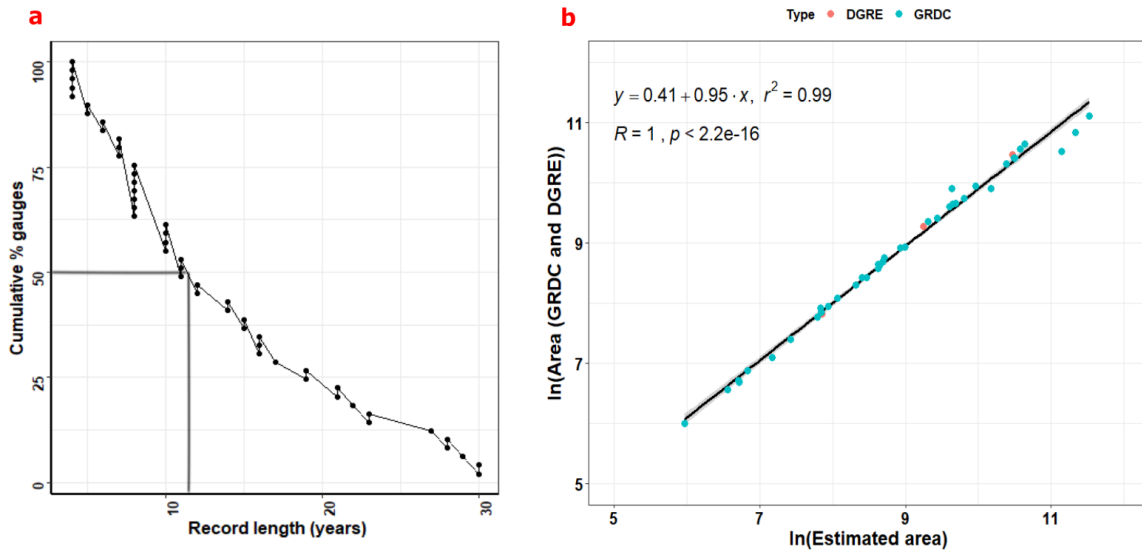
## II - 6 Conclusion

This study highlighted the control of several environmental variables on flow intermittency in Burkina Faso selected among the components that are climate, lithology, and topography. Significant differences are observed in the geographic variation of flow intermittency classes, with perennial and weakly intermittent gauges stations mostly located on Comoe, and Mouhoun catchments; meanwhile, highly intermittent and ephemeral stations are mostly located on Nakanbe and Niger catchment. The main feature discriminating these river catchments comes from their lithology and permeability. This is consistent with the finding that rivers classified as ephemeral in Burkina Faso have shorter lag times than other flow intermittency classes. Despite the importance of annual precipitation and Strahler order in discriminating intermittent rivers, this study suggests that catchment permeability and surface area explain the geographical variation of flow intermittency class in Burkina Faso. We also showed that, if the spatial variations of precipitation contribute to the patterns of flow intermittency, the long-term contrasts of precipitation between the normal and dry sub-periods may have a weak influence on intermittency classes, which could indicate a substantial resilience of rivers in the country.

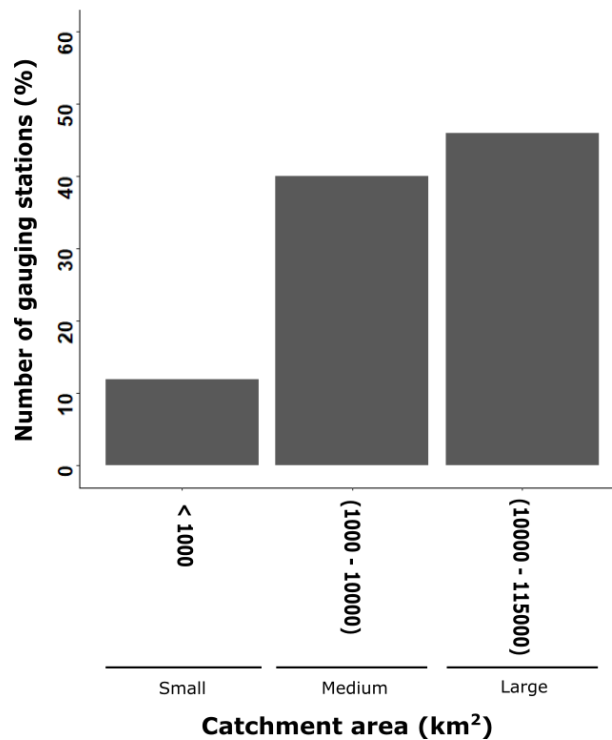
Further studies are needed to understand better the other factors and hydrological processes controlling intermittency in arid and semi-arid environments. In Sahel, since the 1990s, there is some evidence of recovery from the drought (Lebel and Ali, 2009), large dams have been built, and land conversion for agriculture has increased (Yonaba *et al.*, 2021b). With the increasing availability of data, notably from remote-sensing, the influence of these anthropogenic factors could now be considered, although the number of river gauging stations has decreased. Other environmental variables could be also investigated, such as the dry season length or the Standardized Precipitation-Evapotranspiration Index (Vicente-Serrano *et al.*, 2010), which was strongly related to the annual and seasonal zero-flow day occurrence in many stations of Europe and Mediterranean countries outside Europe (Tramblay *et al.*, 2021b). Eventually, the identified control variables could be used to develop predictive models, to provide a comprehensive spatial distribution of river intermittency throughout the country, retrospectively and in the future.

## II - 7 Supplementary to the journal article

### Gauges record length and catchments area

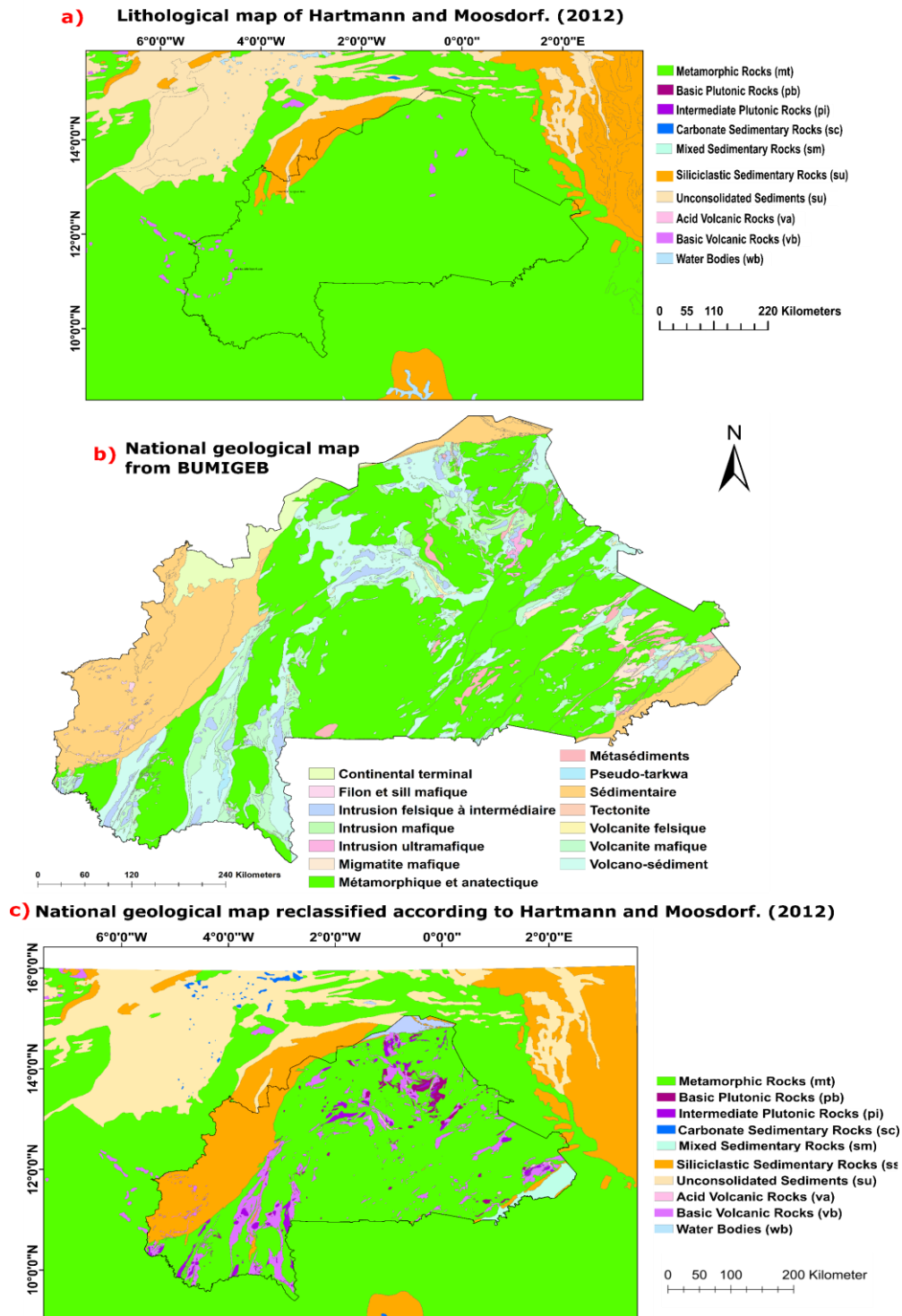


**Figure II-A-1 :** a) Cumulative distribution of gauging stations in relation to the length of records. The majority of gauges (>50 %) have at least ten years of data. b) Comparing delineated catchment areas with catchment area information in the GRDC and DGRE databases (43/49 catchments). The consistency is satisfactory ( $r^2 = 0.99$ ).



**Figure II-A-2 :** Distribution of the 49 gauging stations studied against drainage area. The majority of catchments are large.

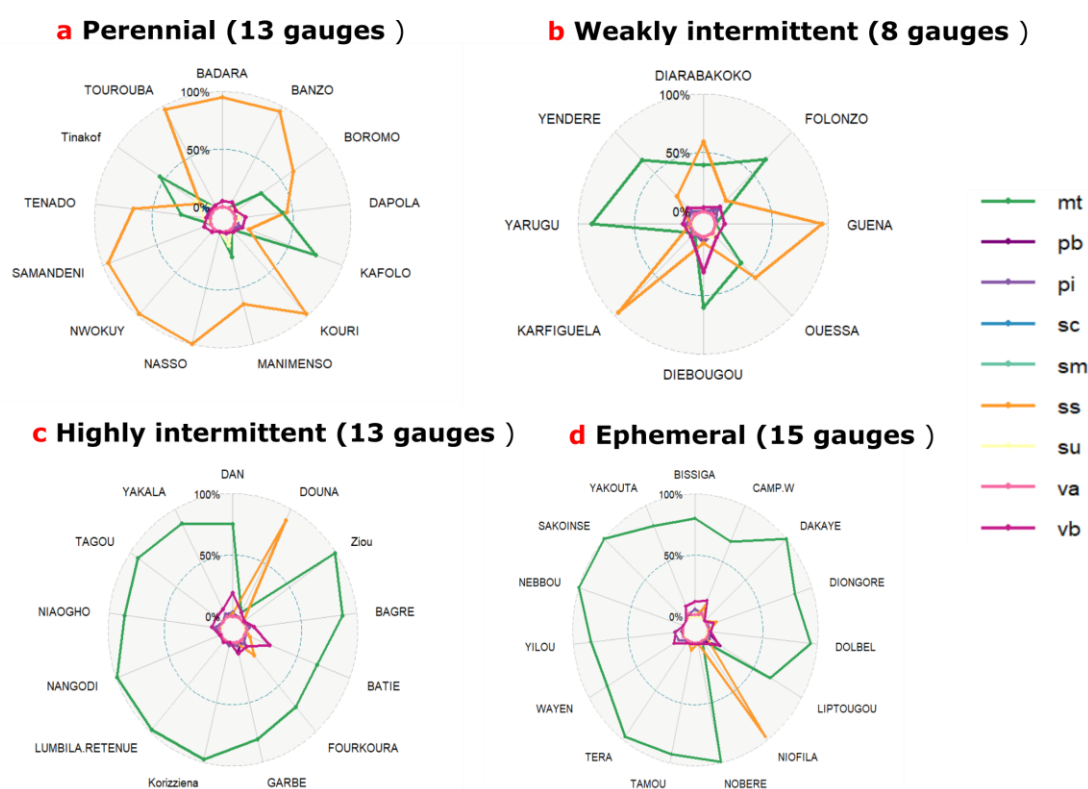
**Lithology and permeability datasets**



**Figure II-A-3** : Map of lithological formations of Burkina Faso and neighbouring countries. **a)** Lithological map of Hartmann and Moosdorf (2012). **b)** National lithological map. **c)** National lithological map reclassified according to Hartmann and Moosdorf (2012). In **c)** Outside the borders of Burkina Faso, lithology is represented by Hartmann and Moosdorf (2012) map at 1/3 750 000 scales. In Burkina Faso, lithological formations are represented by a national lithological map at the scale of 1/1 000 000 reclassified according to (Hartmann et Moosdorf, 2012) nomenclature to have nomenclature compatibility between the two databases. This allowed considering the catchment areas with parts outside of Burkina Faso.



Using the lithological information in Burkina Faso and neighboring countries (**Figure II-A-3.c**), it appears that 67% of the gauging stations classified as perennial and weakly intermittent have more than 50% of catchment areas over siliciclastic sedimentary rocks (high permeability) (**Figure II-A-4.a** , **Figure II-A-4.b**). In contrast, 93 % of the gauging stations classified as highly intermittent and ephemeral have a large part of their catchments located on metamorphic rocks (low permeability), except Douna and Niofila headwater catchments (**Figure II-A-4.c** , **Figure II-A-4.d**). The two gauging stations also appeared as exceptions in the PCA (main text **Figure II-8**). Specifically, these stations are distinguished by their high precipitation (over 1100 mm/yr) and their high percentage of siliciclastic sedimentary rocks values (over 90%) even though they are highly intermittent and ephemeral.



**Figure II-A-4** : Percentage of area (%) occupied by the different types of lithologies per catchment. Legend: mt (metamorphic rocks), pb (basic plutonic rocks), pi (intermediate plutonic rocks), sc (carbonate sedimentary rocks), sm (mixed sedimentary rocks), ss (siliciclastic sedimentary rocks), su (unconsolidated sediments), va (acid volcanic rocks), vb (basic volcanic rocks).

To obtain a permeability map (one continuous variable) that is as easily exploitable as a lithology map (several categorical variables) in the PCA, each lithology type (**Figure II-A-3.c**) has been associated with a permeability value. Permeability range by sediment or rock type is taken from chapter 6 of Maidment (1993). For each lithology type (**Figure II-A-3.c**), the detail of the different rocks or sediments composing it is used to deduce the average permeabilities (**Tableau II-A-1**). In parallel, average drainage densities by type of

lithology are used to detect roughly in order, the most permeable to the least permeable lithologies (**Figure II-A-5**).

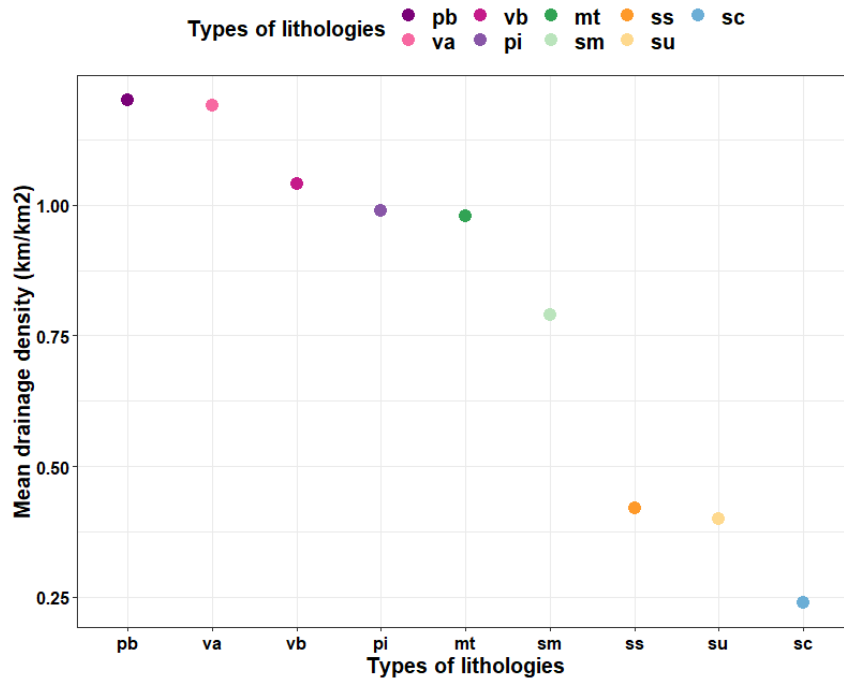
The river network organization is often indicative of geological formations' properties (Dobos *et al.*, 2005 ; Mardhel *et al.*, 2020). Gardiner (1995) showed that higher drainage densities are generally associated with impermeable rocks. Drainage density is generally inversely related to the hydraulic conductivity or permeability of the underlying soil (Vogt *et al.*, 2003). In its study on the hydrogeology of fissured formations, Yameogo (1988), during the geological analysis of its study area in Burkina Faso, found that high drainage densities are associated with impermeable rocks and vis-versa. This leads him to conclude that the hydrographic network can help draw the geological contours.

**Tableau II-A-1** : Permeability estimated values for all lithological units in the study area.

Lithologies	Description	Permeability range derived from chapter 6 of (Maidment, 1993)	Permeability (m <sup>2</sup> ) average range (from 3rd column)	Permeability (m <sup>2</sup> ) average by type of lithology (from 4rd column)
Plutonic (pb, pi) and Volcanic (va, vb)	Fractured igneous and metamorphic rocks	[10 <sup>-17</sup> – 10 <sup>-13</sup> ]	[5 * 10 <sup>-14</sup> ]	[2.5 * 10 <sup>-14</sup> ]
	Unfractured igneous and metamorphic rocks	[10 <sup>-21</sup> – 10 <sup>-17</sup> ]	[5 * 10 <sup>-18</sup> ]	
Metamorphic rocks (mt)	Fractured igneous and metamorphic rocks	[10 <sup>-17</sup> – 10 <sup>-13</sup> ]	[5 * 10 <sup>-14</sup> ]	[5 * 10 <sup>-14</sup> ]
Mixed sedimentary (sm)	Sandstone	[10 <sup>-15</sup> – 10 <sup>-12</sup> ]	[5 * 10 <sup>-13</sup> ]	[3.5 * 10 <sup>-13</sup> ]
	Silt	[10 <sup>-16</sup> – 10 <sup>-12</sup> ]	[5 * 10 <sup>-13</sup> ]	
	Carbonate mud	[10 <sup>-15</sup> – 10 <sup>-13</sup> ]	[5 * 10 <sup>-14</sup> ]	
Siliciclastic sedimentary (ss)	Sandstone	[10 <sup>-15</sup> – 10 <sup>-12</sup> ]	[5 * 10 <sup>-13</sup> ]	[5 * 10 <sup>-13</sup> ]
	Silt	[10 <sup>-16</sup> – 10 <sup>-12</sup> ]	[5 * 10 <sup>-13</sup> ]	
Unconsolidated sediments (su)	Fine to coarse sands	[10 <sup>-14</sup> – 10 <sup>-9</sup> ]	[5 * 10 <sup>-10</sup> ]	[8 * 10 <sup>-13</sup> ]*
Carbonate sedimentary (sc)	Dolomie, limestone	[10 <sup>-16</sup> – 10 <sup>-12</sup> ]	[5 * 10 <sup>-13</sup> ]	[10 <sup>-12</sup> ]*

\*Values selected regarding drainage density **Figure II-A-5**.

The average permeability by type of lithology is obtained by averaging the permeability of the rocks composing it. For the (sc) lithology layer, the permeability value was chosen to be lower than that of the (su) layer by referring (**Figure II-A-5**). Drainage density is low in (sc) than in (su) and thus a potentially higher permeability in (sc) than in (su).



**Figure II-A-5** : Mean drainage density by type of lithology in Burkina-Faso (see lithology code description in Figure II-A-3.c).

The estimated permeability values (**Tableau II-A-1**) are inversely related to the observed drainage density (**Figure II-A-5**). The estimated permeability map (**Figure II-A-6**) is almost consistent with the British Geological Survey's aquifer type/productivity map (BGS) (**Figure II-A-7**), with areas mapped as having high aquifer productivity generally have higher permeability.

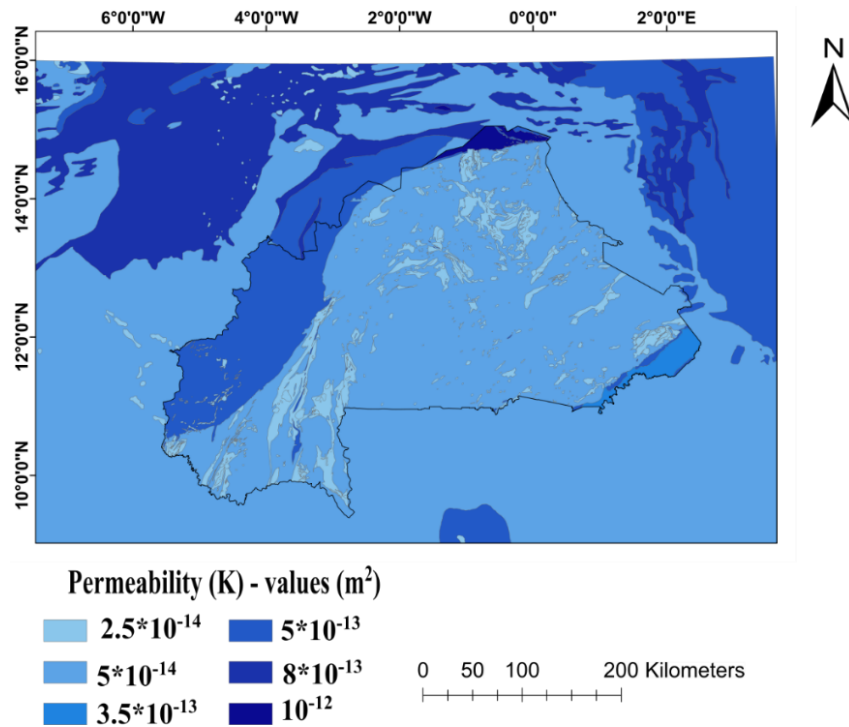


Figure II-A-6 : Estimated permeability map. For each catchment, the average permeability is obtained by weighting the permeabilities obtained for each lithology type (Tableau II-A-1) by the area occupied by each lithology type in the catchment.

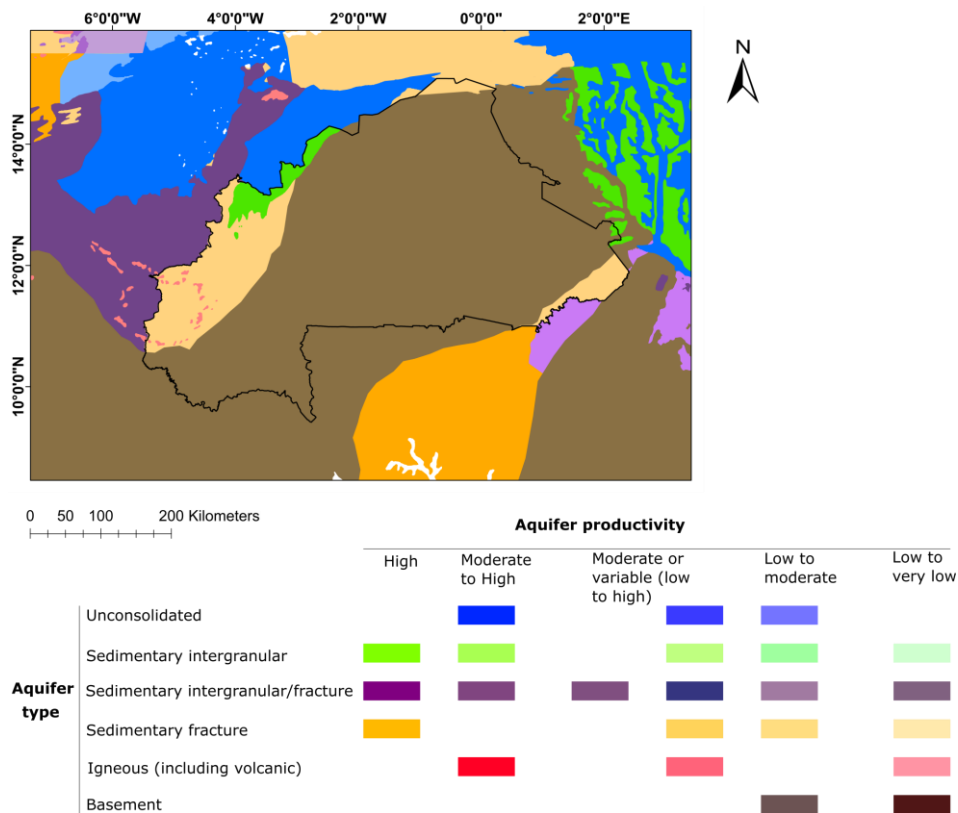
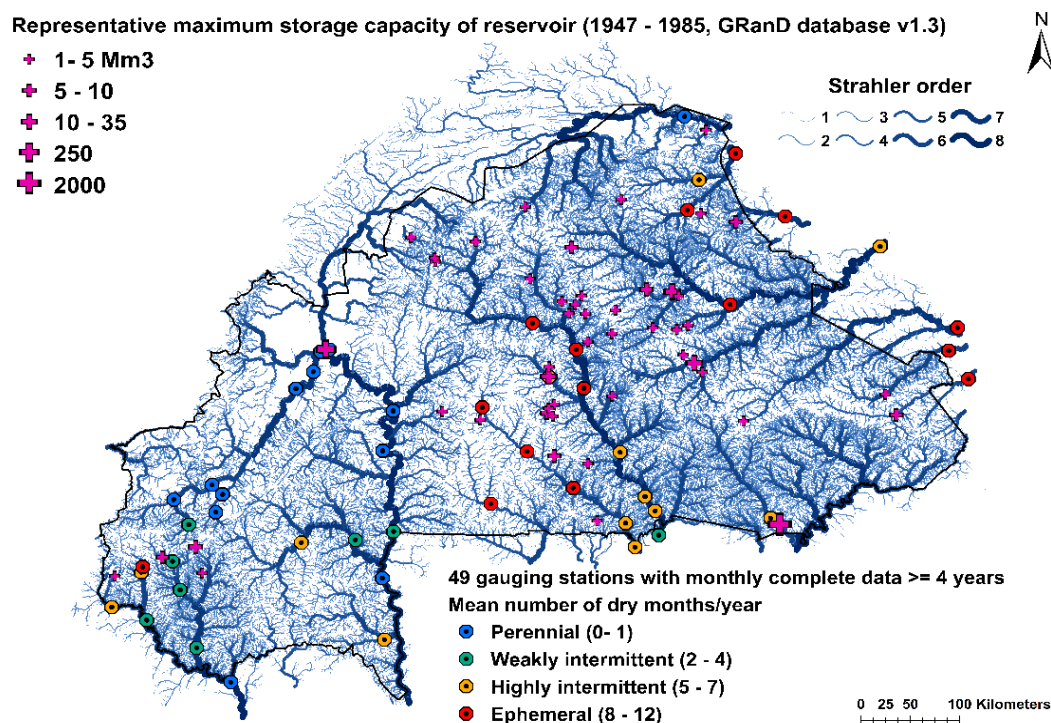


Figure II-A-7 : Hydrogeology maps showing the combination of aquifer type and aquifer productivity. The data are downloaded from Africa Groundwater Atlas at <https://www2.bgs.ac.uk/africagroundwateratlas/downloadGIS.html>

**Assessing dams influence**

In our study, 49 gauging stations are studied over the period (1955-1985), among which 20 stations are shared between the normal period (1955-1969) and the drought period (1970-1985) with at least 4 years in each period. The drought period (1970-1985) was the one when important dams were built to secure water for people and their livestock (Cecchi *et al.*, 2009). Of these 20 gauging stations, only six exhibits a shift of flow class from the normal (1955-1969) to the dry period (1970-1985), and this shift is modest as it consists of only one intermittency class (main text **Figure II-3** and **Figure II-4**).

Dams are mainly located in the central part of the country (**Figure II-A-8**; Cecchi *et al.*, 2009), particularly in the Nakanbe and Niger watersheds. However, 4 of the 6 stations that have changed flow classes between 1955-1970 and 1970-1985 are located in the south-west of the country and upstream of the watersheds, particularly in the Mouhoun (2 gauges) and Comoe (2). The remaining two gauges are located in the Nakanbe (1) and Niger (1) watersheds. On the Mouhoun, no station located downstream of the Lery dam have changed class. Based on this analysis, we assume a very low impact of dams during this period. The change of flow intermittence class for the 6 stations is mainly due to the persistent drought.



**Figure II-A-8** : Geographical variation of flow intermittency classes based on the mean number of dry months per year over 1955-1985. In this figure, the dams in operation over the period 1947-1985 (n=49) are represented by crosses (purple color); their size varies according to capacity (Mm<sup>3</sup>). The data are from the GRanD database (Lehner *et al.*, 2011).

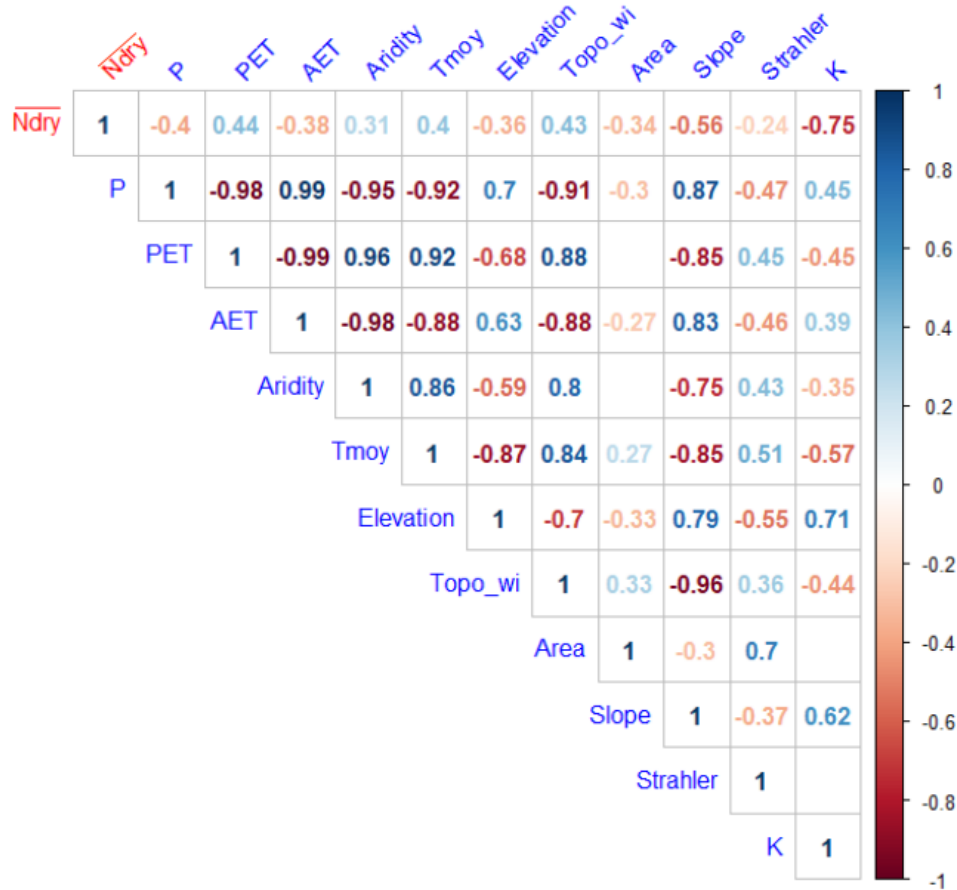
**Classification based on monthly versus daily time step data**

In our study, in addition to the monthly data collected, we have collected the following daily data: at 16 DGRE flow gauges, and at 80 flow gauges from the SIEREM database (see main text for more detail). After pre-processing these data (in terms of missing data, duplicate gauged stations, available years, finally selected period), only 5 flow gauges with daily time step data could be retained. Using data from these 5 gauges, we calculated the mean number of dry days (i.e., days with zero flow) per year for each gauge over the data record (**Tableau II-A-2**). We compared it with a similar estimate based on the monthly data, obtained by multiplying the resulting mean number of dry months per year (as used throughout our study) by 30. The values in the number of dry days are very similar, although slightly lower when estimated from daily records. In all five cases, the same flow class is found by using either daily or monthly discharge data.

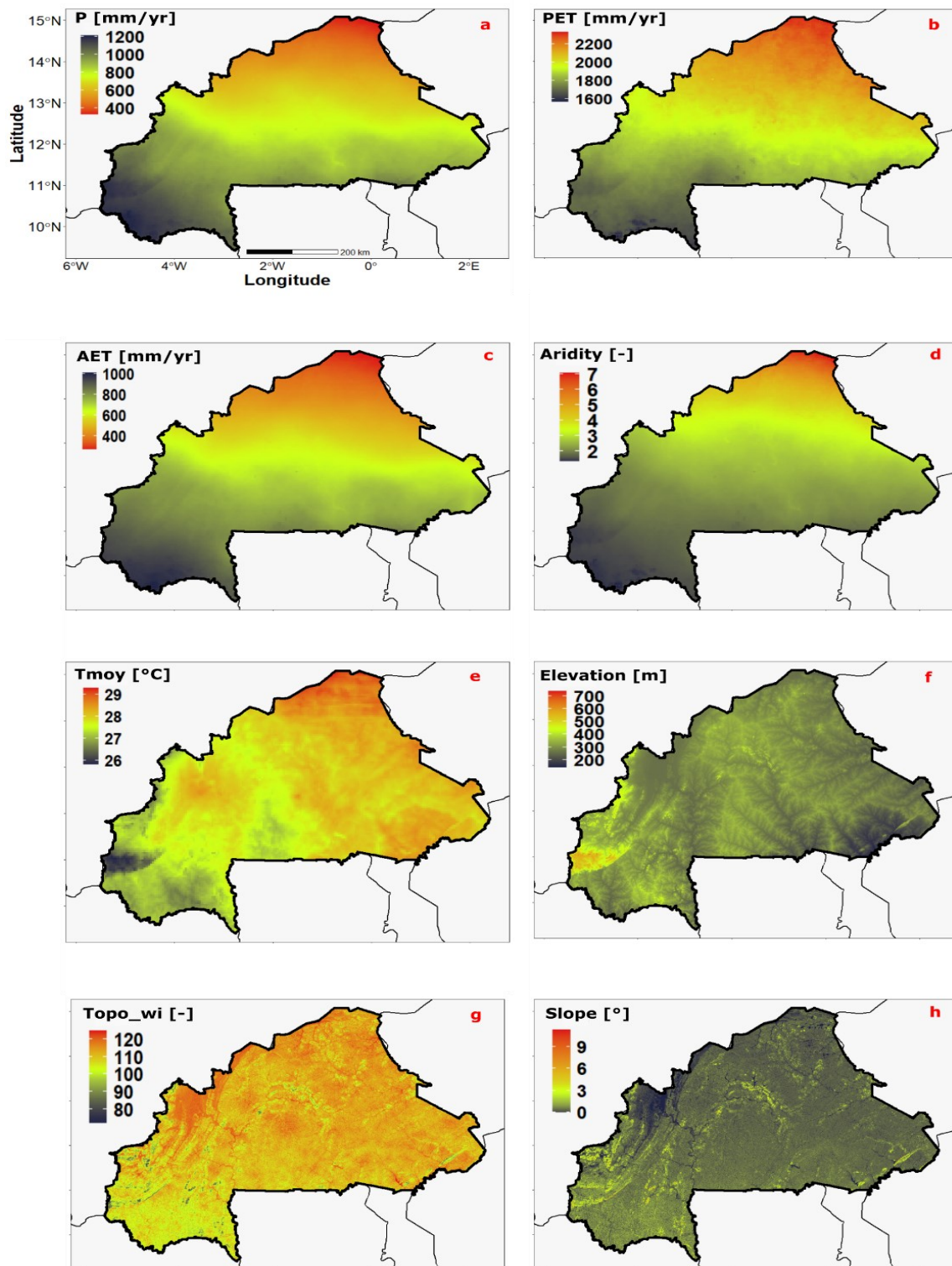
*Tableau II-A-2 : Classification in terms of intermittency of gauges based on monthly versus daily data.*

<b>Name of gauge station (1)</b>	<b>Mean number of dry months/year (2)</b>	<b>Mean number of dry days/year estimated from 2<sup>nd</sup> column (3)</b>	<b>Mean number of dry days/year (4)</b>	<b>Error of monthly estimate (in months) (5) = ((3) - (4)) / 30</b>	<b>Flow class (6)</b>
<b>TOUROUBA</b>	0	0	0	0	Perennial
<b>LUMBILA RETENUE</b>	7	210	220	-0.33	Highly intermittent
<b>TENADO</b>	0	0	0	0	Perennial
<b>FOLONZO</b>	2	60	38	0.73	Weakly intermittent
<b>KAFOLO</b>	1	30	29	0.03	Perennial

**Analysis of correlation between potential explanatory variables of intermittence and their spatial representation**



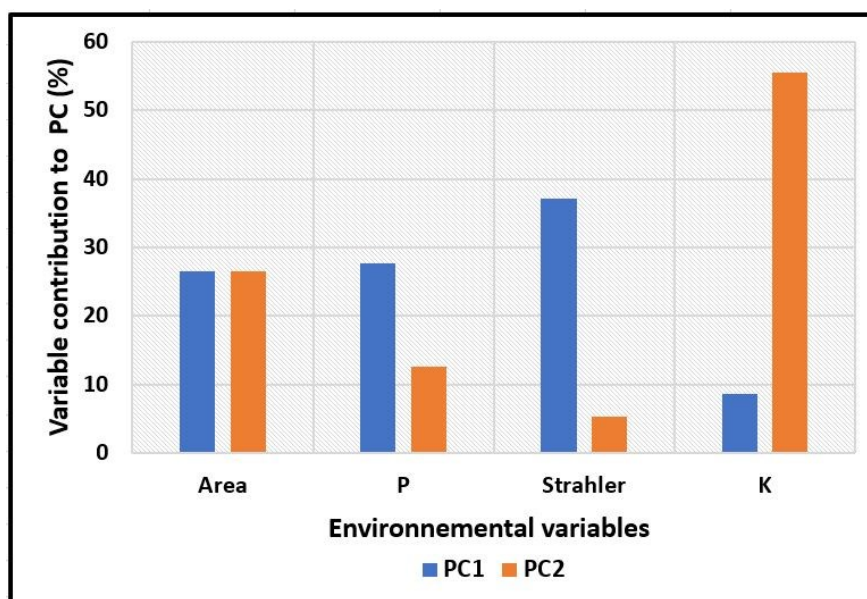
**Figure II-A-9** : Pearson correlation matrix of the 11-initial set of variables. Legend: P (Precipitation), PET (Potential evapotranspiration), AET (Actual evapotranspiration), Aridity (Aridity index), Tmoy (Mean air temperature), Elevation, Topo\_wi (Topography wetness index), Area, Slope, Strahler (Strahler order), K (Permeability). The empty tiles refer to non-significant correlation values ( $p$ -value > 5%), whereas the displayed values are correlation coefficients found to be significant at the 5% level ( $p$ -value < 0.05).



**Figure II-A-10** : Spatial representation of potential environmental variables of intermittence (see variables code description in ). Note that the Strahler order and the permeability (K) are respectively represented in **Figure II-4** in the main text and in **Figure II-A-8**.



**Principal component analysis results (PCA)**



**Figure II-A-11** : Significant environmental variables contribution to principal component 1 (PC1) and principal component 2 (PC2).

**Tableau II-A-3** : Eigenvalues for the four principal components. Only the first two components have an eigenvalue greater than 1. They are retained for interpretation.

	PC1	PC2	PC3	PC4
Eigenvalue	<b>2.093</b>	<b>1.201</b>	0.451	0.255
Variability (%)	52.331	30.029	11.274	6.366
Cumulative %	52.331	82.360	93.634	100.000

**Tableau II-A-4** : Correlations between variables and factors.  $\overline{Ndry}$  added to PCA as a supplementary variable shows a high negative correlation with PC2.

	PC1	PC2	PC3	PC4
Area	0.746	0.565	0.109	-0.336
P	-0.761	0.390	0.517	0.028
Strahler	0.881	0.252	0.174	0.360
K	-0.425	0.816	-0.376	0.108
$\overline{Ndry}$	0.077	<b>-0.851</b>	-0.008	-0.252

**Boxplot of the significant environmental variable per-flow intermittence classes**

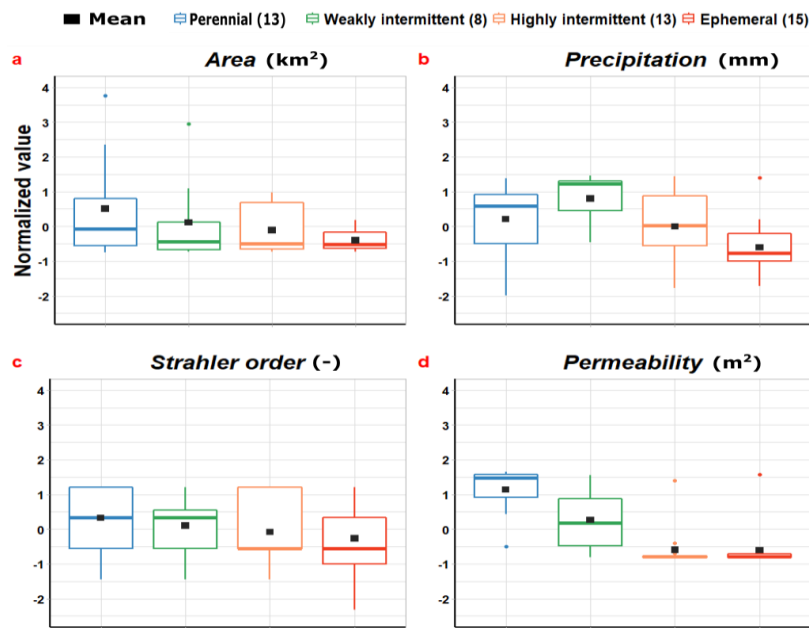


Figure II-A-12 : Boxplots of normalized data. The boxplots show the distribution values of a) Area, b) Precipitation, c) Strahler order, and d) Permeability according to the different flow intermittent class.

**Comparison of intermittency classes between the period 1955-1985 and 1986 - 2017**

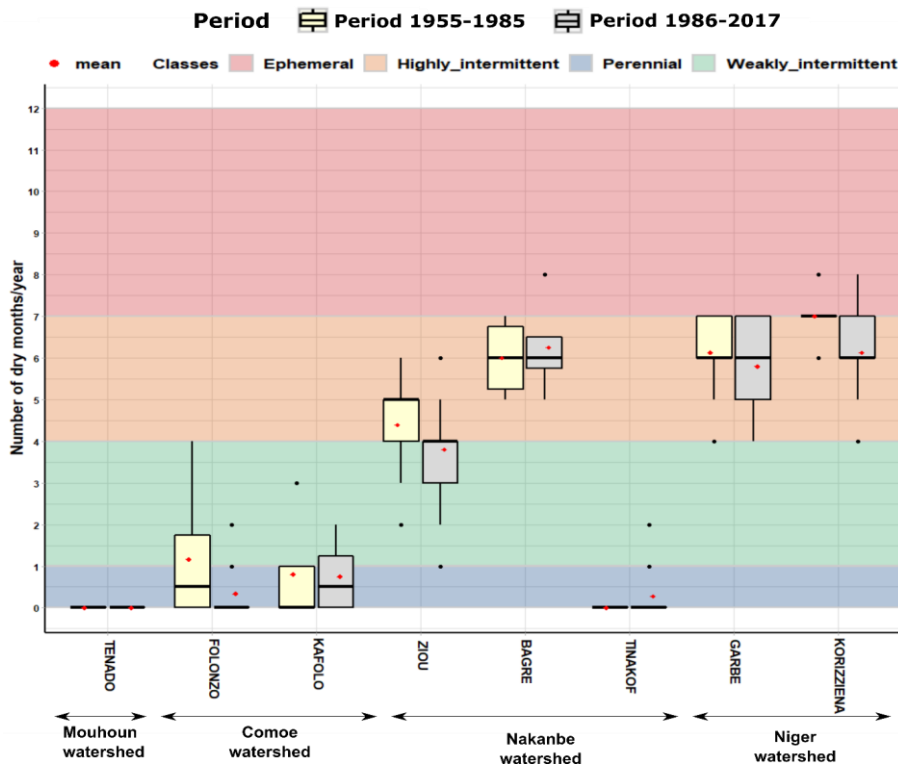


Figure II-A-13 : Boxplot of the number of dry months per year during the periods (1955-1985 and 1986-2017, respectively) at the 8 common gauging stations. The outliers of the boxplots are displayed in black dots, and the gauges are ordered by watershed.

## II - 8 Analyses complémentaires sur la prédiction de l'intermittence

À la suite de l'identification des facteurs de contrôle de l'intermittence au Burkina Faso (perméabilité, surface amont des bassins versants, précipitation, ordre de Strahler) par la méthode de l'ACP, qui ne permet pas de construire un modèle prédictif, cette section présente la construction de modèles *Random Forest* pour la prédiction spatiale de l'intermittence au Burkina Faso et en préparation de son application à l'échelle continentale de l'Afrique en **chapitre III**.

### *Prédiction spatiale de l'intermittence des cours d'eau au Burkina Faso*

Dans ce qui précède, le réseau hydrographique de référence du Burkina Faso a été utilisé pour estimer les ordres de Strahler ainsi que pour la représentation des cours d'eau. Pour la prédiction spatiale de l'intermittence au Burkina Faso, le réseau LCS a été pris en compte, car le réseau BNDT n'étant pas dérivé d'un MNT, il est difficile d'extraire les bassins versants amont associés aux tronçons de rivières. À l'exception des informations sur l'ordre de Strahler, les sources de données où ces variables sont extraites restent les mêmes (**Tableau II-2**) avec les informations sur la perméabilité dérivée de la couche lithologique du BUMIGEB. Plusieurs modèles de forêt aléatoires (*Random Forest*) ont été construits pour lier les variables explicatives (perméabilité, surface amont des bassins versants, précipitation, ordre de Strahler) aux différentes classes d'intermittence observées au droit des stations de jaugeage (49 sur la période 1955-1985) dans l'optique d'extrapoler ces classes d'intermittence à tous les tronçons de cours d'eau non jaugés. Pour chaque tronçon de cours d'eau du réseau LCS, des valeurs moyennes amont par bassin versant de ces variables sont calculées. Nous avons exclu un certain nombre de cours d'eau situé le long du fleuve Niger, car notre zone d'étude ne couvre pas la totalité de leur bassin versant en amont, ce qui empêche le calcul correct des variables associées.

Deux modèles de prédictions ont été explorés à savoir un modèle *Random Forest* en classification (M1-LCS-BF) et un autre en régression (M2-LCS-BF). Dans le modèle en classification nous avons entraîné le modèle pour prédire directement les différentes classes d'intermittence et dans le second pour prédire le nombre moyen de mois sec que nous avons converti ensuite en classe d'intermittence selon les critères définis dans le **Tableau II-1**. En nous basant sur les critères de performance des deux modèles en validation (*accuracy*, *recall*,

*precision*, score F1, Kappa, Kappa pondéré,  $R^2$  et RMSE) nous avons choisi celui qui donnait les meilleures performances.

Pour entraîner le modèle nous avons réparti les 49 stations de jaugeage en données de calibration (41) et de validation (8) soit une répartition 80%, 20% en tenant compte d'avoir des proportions des classes d'intermittence en calibration et validation similaire à celle des données initiales (répartition stratifiée), cela permet d'éviter un biais dans la calibration et la validation du modèle.

Pour ajuster les hyperparamètres du modèle, il est souvent recommandé de fixer le nombre d'arbres à entraîner par le modèle (*n<sub>tree</sub>*) en fonction de la puissance de calcul de l'ordinateur (Probst et Boulesteix, 2017). La valeur de *n<sub>tree</sub>* a été fixée à 800 en référence à des études récentes qui utilisaient RF pour la prédiction de l'intermittence des cours d'eau (Messenger *et al.*, 2021 ; Price *et al.*, 2021). Pour obtenir le paramètre (*m<sub>try</sub>*), nous avons procédé à l'optimisation des fonctions critères en faisant varier ce paramètre de 1 à 4 avec un pas de variation de 1 effectuant donc 4 itérations dans lesquelles le modèle est entraîné et évalué en interne à travers une validation croisée à 10 blocs (voir **chapitre I** section **I-4-3** pour plus de détails). Dans le cas de la classification, le modèle est entraîné pour optimiser l'*accuracy* (proportion des classes correctement classifiées sur le total) ainsi que la valeur du Kappa. Dans le cas de la régression, le modèle est entraîné pour optimiser la RMSE (*Root mean square error*) et le  $R^2$ . Les valeurs optimales de (*m<sub>try</sub>*) ainsi ajustées en classification et régression sont respectivement de 3 et 4.

La calibration du modèle RF en utilisant les hyperparamètres définis ci-dessus a donné en classification une *accuracy* globale de 63% et une valeur de Kappa de 0,46 qui peut être jugée assez bonne (Landis et Koch, 1977). En validation, les valeurs obtenues sont du même ordre respectivement de 63% et 51% (**Tableau II-A-5**) et sont assez bonnes. En régression, les critères de performance sur le nombre moyen de mois secs prédits par an, sont : en calibration, une RMSE de 1,38 (en nombre de mois secs, ce qui implique une erreur moyenne de l'ordre d'un mois entre les valeurs observées et prédites) et un  $R^2$  de 0,85 ; et en validation, une RMSE de 1,4 et un  $R^2$  de 0,82 (**Tableau II-A-5**) ce qui est bon en termes de performances. En convertissant les valeurs du nombre moyen de mois secs prédits en classe d'intermittence, il ressort de la matrice de confusion correspondante une *accuracy* de 0,88 et un Kappa de 0,82 ce qui est très bon comme critère de performance. En raison de ces performances en validation, l'*accuracy* et le Kappa sont supérieurs à celle du modèle RF en classification, le modèle RF en régression semble d'emblée meilleur. Les critères de

performances des modèles ainsi que les matrices de confusion en calibration et en validation sont présentés dans (Tableau II-A-5, Figure II-A-14).

Tableau II-A-5 : Critères de performances obtenus en calibration et validation pour les différents modèles de RF explorés pour prédire l'intermittence au Burkina Faso.

RF en classification		RF en régression	
<b>Calibration (41 stations)</b>			
ntree	800	ntree	800
mtry	3	mtry	4
Accuracy	0,63	RMSE	1,38
Kappa	0,46	R <sup>2</sup>	0,85
<b>Validation (8 stations)</b>			
Accuracy	0,63	Accuracy	0,88
		R <sup>2</sup>	0,82
Kappa	0,51	Kappa	0,82
		RMSE	1,4

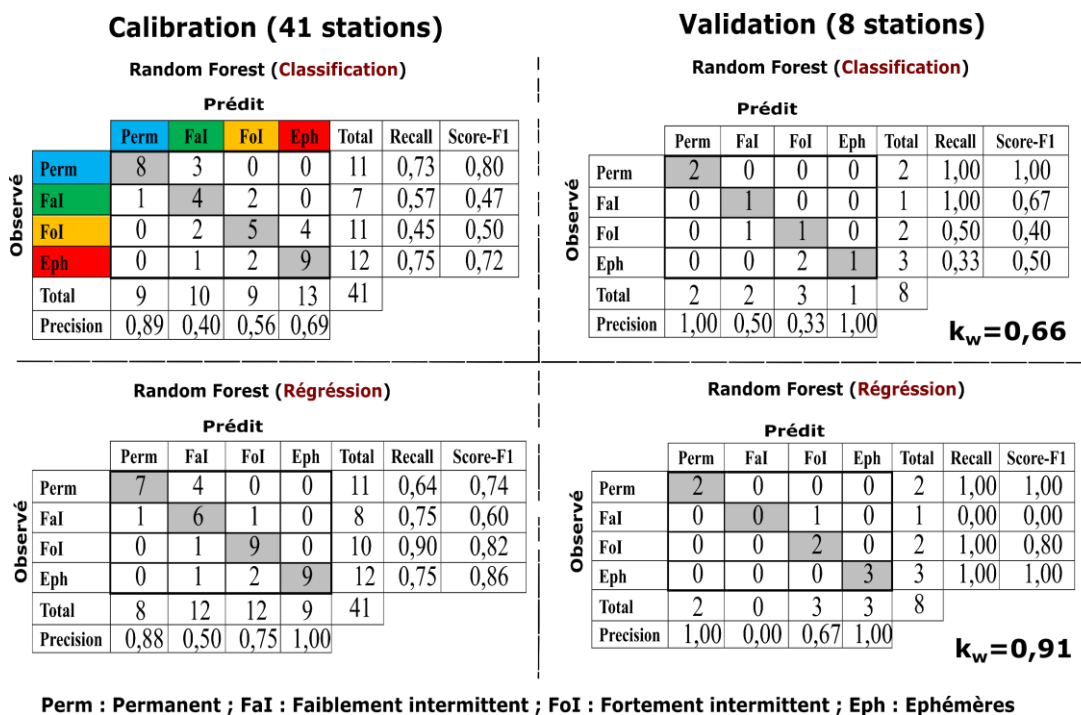


Figure II-A-14 : Matrice de confusion en calibration et validation pour les différents modèles testés au Burkina Faso. Notons que les cellules de la diagonale indiquent les cas où les classes sont correctement prédites. En régression les valeurs prédites sont numériques (nombre moyen de mois sec) et ont été converties en classe d'intermittence selon le Tableau II-1 pour générer les matrices de confusion correspondantes.

En se basant sur le critère de Kappa pondéré (Kw), le modèle finalement adapté pour la prédiction a été choisi, car ce critère tient compte de l'ordre inhérent entre les différentes

classes. Les quatre classes d'intermittence définies dans le **chapitre II** sont ordinales c'est-à-dire qu'il existe un ordre inhérent entre les classes (permanent, faiblement intermittent, fortement intermittent, éphémère), les erreurs de prédiction sont moins importantes si elles sont proches de la classe prédite (par exemple l'erreur sur une classe actuellement permanente qui est prédite faiblement intermittente est moindre que si elle est prédite éphémère). Pour tenir compte de ces types d'erreurs, nous avons défini une matrice de coût et pour laquelle le coût de l'éloignement d'une classe prédite par rapport à sa classe réelle est défini par la différence entre les valeurs moyennes du nombre de mois secs des classes concernées (**Tableau II-A-6**). Par exemple, la valeur moyenne du nombre de mois secs de la classe faiblement intermittente (2-4) est de 3 mois et celle permanente (0-1) est de 0,5 mois. Ainsi, la différence entre les valeurs moyennes du nombre de mois secs de la classe faiblement intermittente et celle permanente est de 2,5 mois tandis qu'entre celle éphémère et permanente, elle est de 9,5 mois. Le Kappa pondéré est une pondération du Kappa par cette matrice de coût. L'interprétation des valeurs du Kappa pondéré se fait comme celle du Kappa. La valeur du critère de Kappa pondéré est de 0,66 pour le modèle RF en classification et 0,91 pour le modèle RF en régression et peut être considérée comme un accord fort entre les valeurs observées et prédites pour le modèle en classification et comme un accord presque parfait pour le modèle en régression (Landis et Koch, 1977). Le modèle RF en régression (M2-LCS-BF) qui a montré des performances très satisfaisantes en validation a été finalement choisi pour la prédiction spatiale de l'intermittence au Burkina Faso.

**Tableau II-A-6** : Matrice de coût utilisée comme pondération pour le calcul du critère de Kappa.

		Prédit			
		Perm	FaI	FoI	Eph
Observé	Perm	0	2,5	5,5	9,5
	FaI	2,5	0	3	7
	FoI	5,5	3	0	4
	Eph	9,5	7	4	0

À l'aide de l'approche *Random Forest*, nous avons également évalué l'ordre d'importance des variables environnementales dans la prédiction de l'intermittence au Burkina Faso (**Figure II-A-15**) à travers une approche par permutation des valeurs qui permet de mesurer le pourcentage d'augmentation de l'erreur quadratique moyenne (*%IncMSE*) si une variable donnée est supprimée (voir section I-4-4 pour plus de détails). Comme mise en évidence par l'ACP (section II-4-2), la perméabilité apparaît ici par cette autre méthode (**Figure II-A-15**) comme la variable la plus importante dans la discrimination des classes d'intermittence au

Burkina Faso suivie par la précipitation moyenne annuelle et la surface amont qui sont presque à égalité et enfin l'ordre de Strahler.

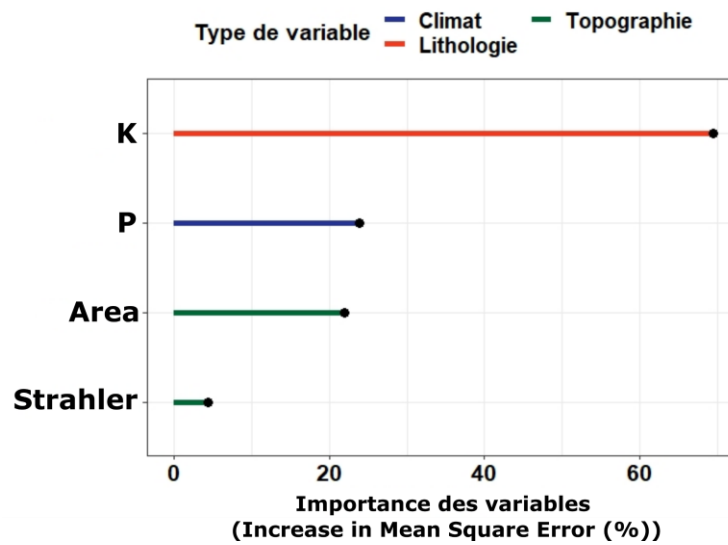


Figure II-A-15 : Ordre d'importance des variables dans le modèle Random Forest en régression. K (Perméabilité moyenne amont du bassin), P (Précipitation moyenne annuelle amont du bassin), Area (Surface amont du bassin), Strahler (Ordre de Strahler).

La Figure II-A-16 présente la distribution spatiale des différentes classes d'intermittence prédite par le modèle au Burkina Faso et aux voisinages environnants.

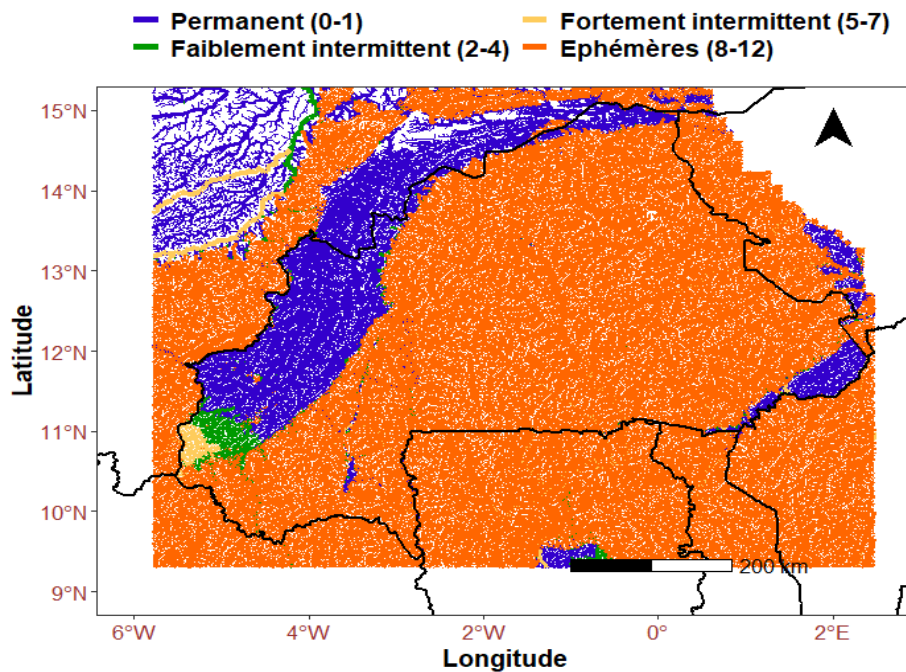
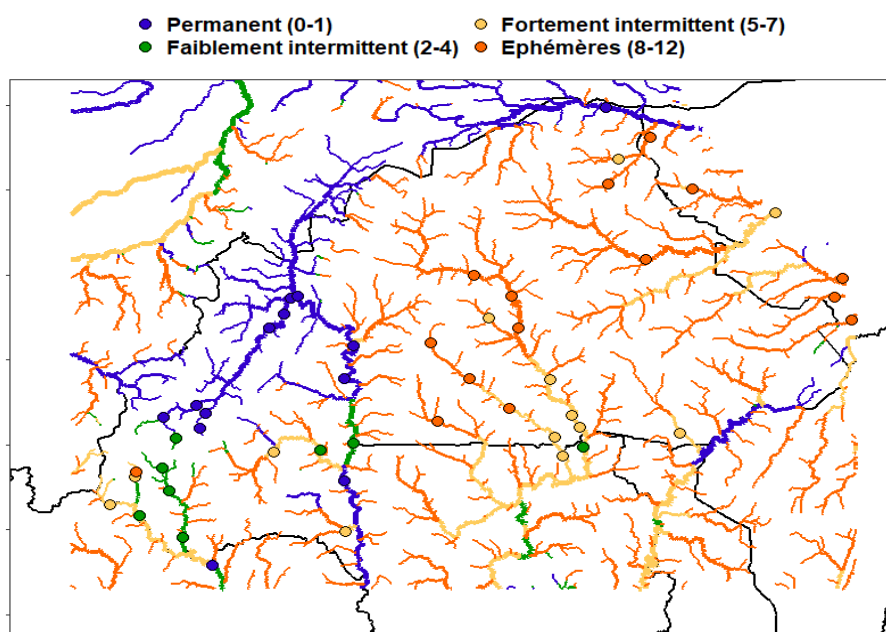


Figure II-A-16 : Répartition spatiale des classes d'intermittences prédites au Burkina Faso sur le réseau LCS (ordre de Strahler 1-9) par le modèle RF en régression.

Les ordres de Strahler des cours d'eau varient entre 1 et 9. Au Burkina Faso, en termes de proportion en longueur sur les ordres de Strahler 1-9, les cours d'eau permanents constituent 17% des cours d'eau prédits et intermittents 83% (faiblement intermittents (2%), fortement intermittents (1%), éphémères (80%)). Les prédictions effectuées sur les cours d'eau d'ordre de Strahler 1 à 4 de la **Figure II-A-16** sont fortement incertaines, car l'ordre de Strahler des cours d'eau sur lequel sont situées les stations de jaugeage varie entre 5 et 9. En termes de proportion en longueur sur les ordres de Strahler 5-9 (**Figure II-A-17**), les cours d'eau permanents constituent 25% des cours d'eau prédits et intermittents 75% (faiblement intermittents (4%), fortement intermittents (12%), éphémères (59%)).



*Figure II-A-17 : Répartition spatiale des classes d'intermittences prédites au Burkina Faso sur le réseau LCS (ordre de Strahler 5-9) par le modèle RF en régression.*

Sur ces figures, on peut dire que les cours d'eau permanents sont initiés dans des zones plutôt perméables que le reste des classes et que les cours d'eau éphémères drainent en général une superficie de bassin amont moins importante que celle fortement intermittente et faiblement intermittente de même que pour l'ordre de Strahler. Pour la précipitation moyenne annuelle qui croît du Nord au Sud (600 à 1200 mm), son influence sur la distribution spatiale des classes d'intermittence n'est pas très perceptible.

En comparaison avec la classification de l'intermittence présente dans la base hydrographique de référence de l'IGB dans laquelle on calcule en moyenne 98% des cours d'eau comme intermittents, le modèle M2-LCS-BF prédit en moyenne 83% des cours d'eau comme



intermittents (en regroupant les classes faiblement intermittentes, fortement intermittentes, et éphémères) et donc plus de cours d'eau permanents que le réseau de référence (**Tableau II-A-7**). Comparativement aux études précédentes (**Figure II-A-18**), celles de Lin *et al.* (2021) et Perez-Saez *et al.* (2017) prédisent également des proportions moins importantes de cours d'eau intermittents respectivement de 84% et 75,4% et la configuration spatiale des cours d'eau permanents et intermittents est très dépendante de la précipitation moyenne annuelle (**Figure II-A-10.a**) qui est plus importante dans la partie sud-ouest du pays. Pour les deux classifications de l'intermittence (à pas de temps journalier comme mensuel) de Messenger *et al.* (2021), tous les cours d'eau du Burkina Faso sont prédits intermittents avec une proportion moyenne de 100 %. La prédiction la plus corrélée au réseau de référence est M2-LCS-BF (**Tableau II-A-7**, **Tableau I-5**) avec un coefficient de corrélation de 0,16 et qui est très faible.

**Tableau II-A-7** : Performances du modèle M2-LCS par rapport au réseau hydrographique de référence du Burkina Faso. Les paramètres s'appliquent pour les différents réseaux sur les densités de drainage (km/km<sup>2</sup>) synthétisées dans les mailles de 5' (environ 8 km x 8 km).

Pays (Échelle)	Bases de données / Études	Paramètres statistiques			
		Pourcentage moyen interm. (%)	Écart-type (%)	Coefficient de corrélation de Pearson (-)	RMSE (%)
<b>Burkina Faso</b> <b>(1 : 200 000)</b>	M2-LCS-BF	81	37	0,16	39
	IGB (référence)	98	7,4	1	0

Le fait que le modèle M2-LCS-BF prédit plus de cours permanents semble nettement porté par les zones de forte perméabilité (**Figure II-A-6**). Cela pourrait en partie s'expliquer par le fait que la calibration du modèle à partir de stations situées sur de grands cours d'eau (ordre de Strahler de 5 à 9) donne de mauvais résultats en extrapolation sur les plus petits cours d'eau (ordre de Strahler 1 à 4) qui constituent la majorité des cours d'eau dans un réseau hydrographique (Wohl, 2017).

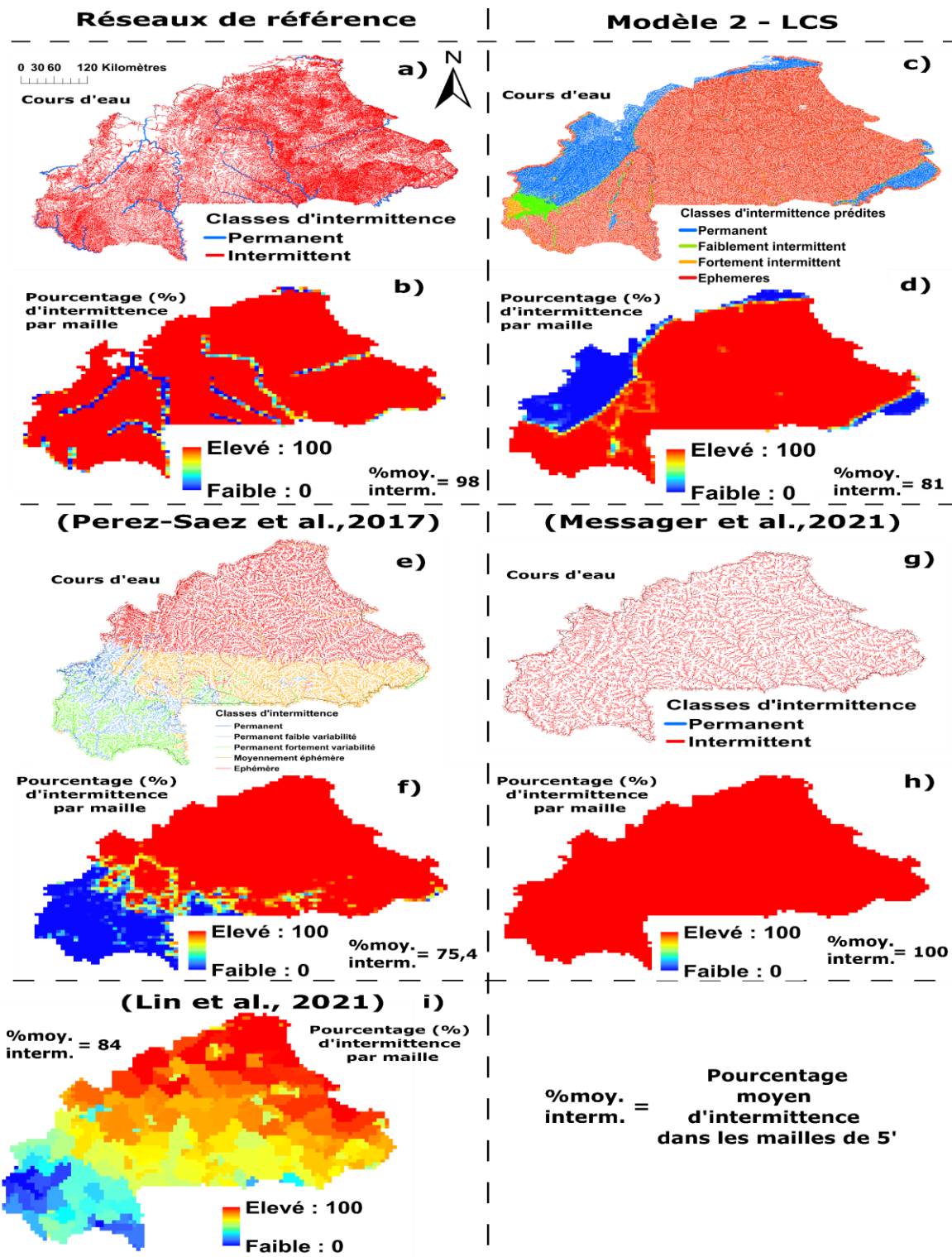
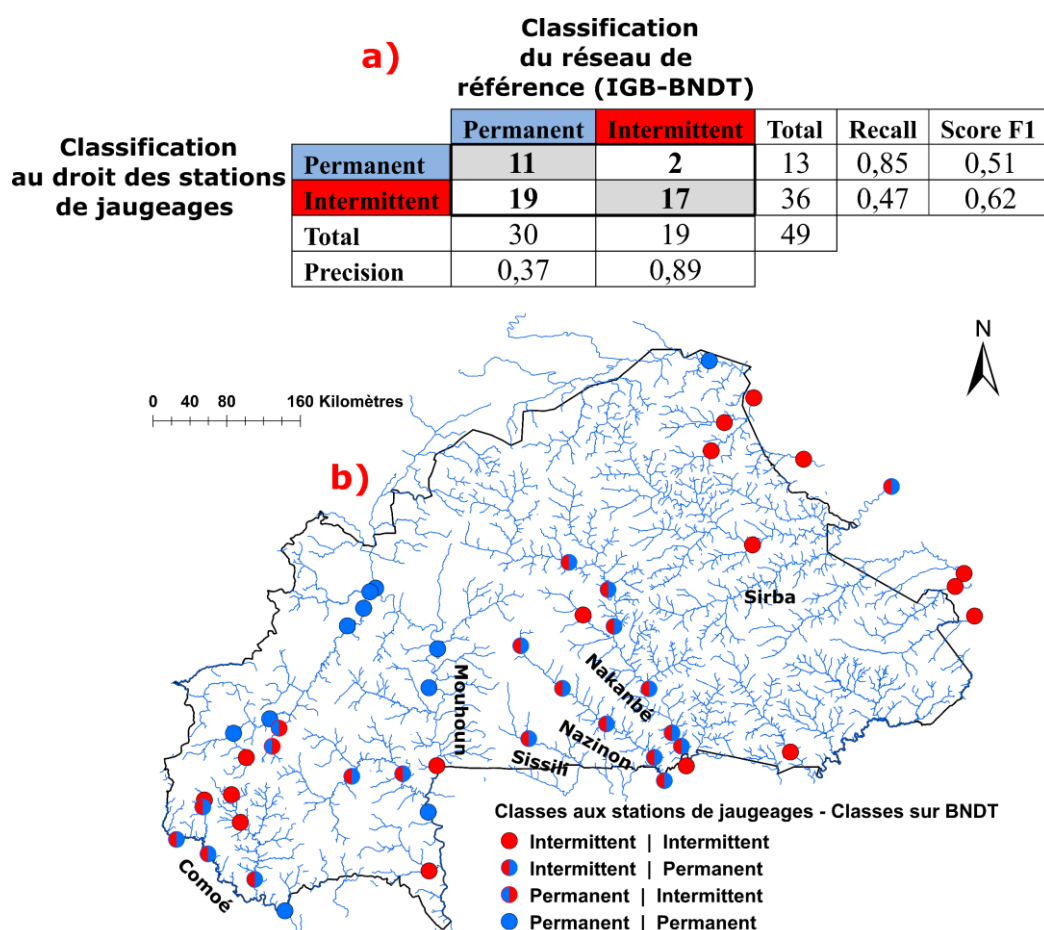


Figure II-A-18 : Comparaison de la distribution spatiale de l'intermittence prédite par M2-LCS-BF et les études (Lin et al., 2021 ; Messenger et al., 2021 ; Perez-Saez et al., 2017) à la carte nationale de l'intermittence (IGB-BNDT) du Burkina Faso.

Par ailleurs, la comparaison de la classification de l'intermittence observée au droit des stations de jaugeage (en regroupant les classes faiblement intermittentes, fortement intermittentes, et éphémères en une seule classe intermittente) avec la classification de

l'intermittence présente dans le réseau hydrographique de référence du Burkina Faso aux emplacements des stations de jaugeage montre de fortes divergences. De la matrice de confusion résultante (**Figure II-A-19.a**), on calcule une *accuracy* de 57,1% ce qui montre un accord sur un peu plus de la moitié des stations, mais aussi une part importante de désaccord (42,9 %). Le *recall* le plus faible (0,47) est obtenu pour la classe intermittente ce qui signifie que seulement 47% des stations intermittentes sont correctement classées intermittentes dans le tronçon correspondant dans la base BNDT, le reste est classé permanent. La distribution spatiale des accords et désaccords entre la classe d'intermittence observée au droit des stations de jaugeage et celle observée au niveau de la BNDT montre que les cours d'eau de la Sissili, du Nazinon, du Nakanbé et une partie de la Comoé qui sont classées permanentes dans la base BNDT sont intermittents selon les observations des stations de jaugeage (**Figure II-A-19.b**).



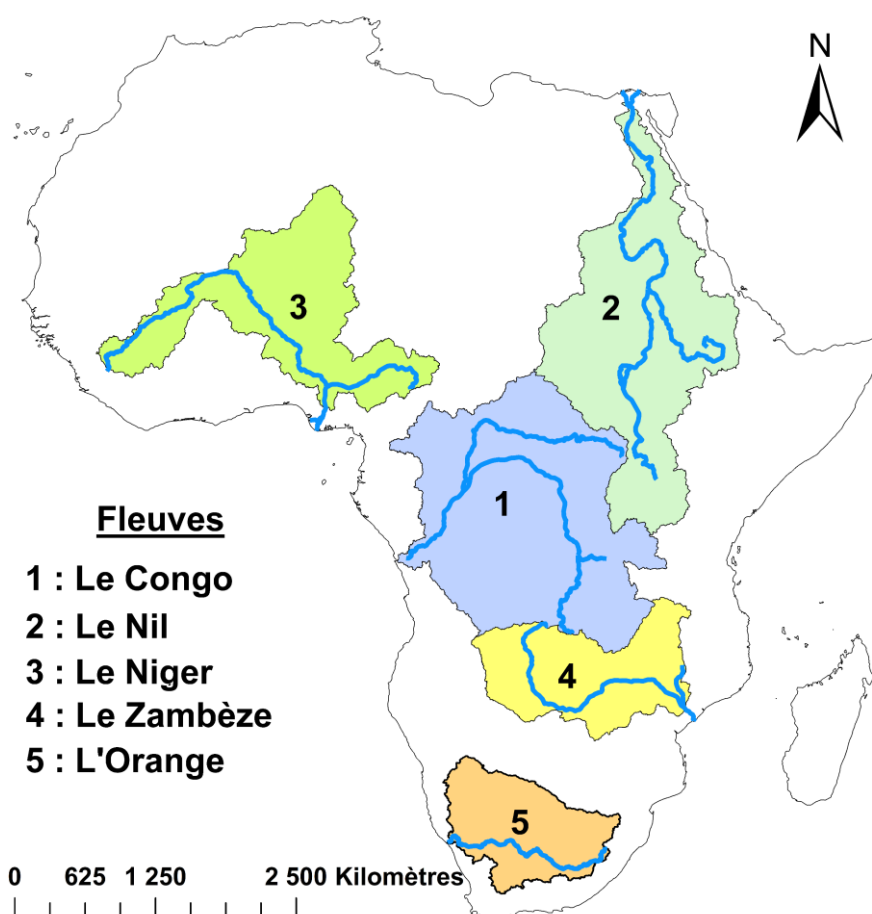
**Figure II-A-19 :** Comparaison de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence BNDT du Burkina Faso aux emplacements des stations de jaugeages. **a)** matrice de confusion, **b)** distribution spatiale des classes d'intermittence, la première moitié du cercle correspond à la classification calculée aux stations et la seconde à la classification observée au niveau du réseau BNDT.

En récapitulatif, l'approche *Random Forest* développée dans cette section a permis de confirmer en plus de la méthode de l'ACP, l'influence significative de la perméabilité sur les

variations géographiques des différentes classes d'intermittence au Burkina Faso. Les analyses effectuées dans cette section montrent que bien que les classifications de l'intermittence présente dans les réseaux hydrographiques de référence soient une base utile pour la comparaison avec nos résultats de prédiction, l'intermittence observée au droit des stations de jaugeage divergent très souvent avec la classification présente dans les réseaux hydrographiques de référence, le même constat est fait aux États-Unis par Hafen *et al.* (2020). L'approche *Random Forest* développée dans cette section sera généralisée dans le **chapitre III** suivant dans le but d'identifier les facteurs de contrôle de l'intermittence et prédire la distribution spatiale des cours d'eau intermittents à l'échelle continentale de l'Afrique.

## Chapitre III - Caractérisation et prédiction spatiale de l'intermittence des cours d'eau à l'échelle continentale de l'Afrique

Le chapitre précédent s'est intéressé à l'échelle régionale, au Burkina Faso, pour identifier les facteurs de contrôle les plus importants de l'intermittence des cours d'eau, et proposer une prédiction spatiale de l'intermittence dans la section II-8 au chapitre II. Le chapitre III vise à généraliser cette démarche à l'échelle continentale de l'Afrique, sur la période 1958-1991, et développe une approche basée sur les forêts aléatoires (*Random Forest*) pour identifier les facteurs de contrôle de l'intermittence et prédire spatialement l'intermittence du débit des cours d'eau pour au moins 15 millions de tronçons de cours d'eau présents dans la base de données LCS (Schneider *et al.*, 2017).



Aperçu sur les fleuves en taille de bassin les plus importants d'Afrique<sup>32</sup> (Meybeck et Ragu, 1997).

<sup>32</sup> Sources des données : <http://riverbasins.wateractionhub.org/>

## III - 1 Résumé

Ce chapitre présente le développement d'une méthode pour la prédiction et la cartographie de la distribution spatiale des cours d'eau intermittents en Afrique. Nous avons travaillé à l'échelle des tronçons de rivières du réseau hydrographique et des bassins versants associés à ces tronçons afin d'aboutir à une information spatialisée la plus fine possible. Des modèles de forêts aléatoires (*Random Forest*) ont été entraînés pour établir des liens entre les données de débits ponctuels (sur la période 1958-1991) avec diverses variables climatiques, géologiques, hydrologiques, de quantification d'influence anthropique et de propriétés des sols. Le modèle retenu s'est avéré fiable et efficace en validation pour la classification des classes d'intermittence avec une précision globale (*global accuracy*) de 69%, un critère de Kappa et Kappa pondéré respectivement de 0,37 et 0,42. Il ressort des bassins versants examinés que l'indice d'aridité, la surface amont et l'évapotranspiration potentielle moyenne annuelle sont les facteurs de contrôle les plus importants de l'intermittence des cours d'eau à l'échelle continentale de l'Afrique. La majorité des cours d'eau intermittents étant localisés dans les zones où l'évapotranspiration potentielle moyenne annuelle est supérieure à la précipitation moyenne annuelle (c.-à-d. aridité  $[P/ETP] < 1$ ). Cette étude a aussi permis de prédire à l'échelle continentale que 44 % en longueur des cours d'eau sont permanents tandis que 56 % sont intermittents (9 % faiblement intermittents, 31 % fortement intermittents, 16 % éphémères). Ces nouvelles estimations contrastent avec celles d'études récentes (Lin *et al.*, 2021 ; Messenger *et al.*, 2021) qui ont prédit respectivement qu'entre 69% et 71% de la longueur des cours d'eau sont intermittents en Afrique. Cette différence pourrait être liée à la densité de drainage constante et à l'échelle du réseau hydrographique considéré dans ces études antérieures. Les prédictions du modèle affichent une bonne ressemblance spatiale avec les bases de référence nationales au Mali et en Afrique du Sud (respectivement une corrélation  $r = 0,43$  et  $r = 0,56$ ), une assez bonne ressemblance spatiale au Bénin et à Madagascar (respectivement  $r = 0,41$  et  $r = 0,39$ ), mais une ressemblance spatiale mitigée au Burkina Faso ( $r = 0,1$ ) en grande partie à cause d'erreurs dans la classification de l'intermittence dans la base de référence sur certains grands cours d'eau. Cette étude souligne le fait que le nombre de stations permettant de mesurer les débits des cours d'eau intermittents est généralement très faible par rapport aux stations permanentes. Les cours d'eau de tête de bassin sont en général naturellement sujets à l'intermittence du débit (Beaufort *et al.*, 2018). Ainsi, un nombre plus important de stations de mesure installées sur des petits cours d'eau

plus à l'amont du réseau hydrographique devrait permettre de mieux prédire les cours d'eau intermittents et comprendre les mécanismes de contrôle de l'intermittence. Cette classification et cartographie de l'intermittence des cours d'eau avec un niveau de détails sur quatre classes d'intermittence fournissent une base qui pourra servir dans la mise en œuvre des politiques de gestion de l'eau et de la biodiversité en Afrique.

## III - 2 Introduction

La caractérisation de la distribution spatiale des cours d'eau intermittents suscite un intérêt croissant dans le monde entier, de même que la compréhension de la manière dont ils varient spatialement en fonction des facteurs environnementaux (Costigan *et al.*, 2017). Traditionnellement, les méthodes utilisées pour cartographier les cours d'eau intermittents vont des campagnes de terrain (qui sont laborieuses sur de grands domaines) à la numérisation d'images satellitaires (Kaplan *et al.*, 2019 ; Kostianoy *et al.*, 2020 ; Turner et Richter, 2011) dont l'opacité de la végétation et l'application sur de vastes zones peuvent être des facteurs limitants.

Les méthodes statistiques de régionalisation de l'information locale de l'intermittence (notamment à travers les stations de jaugeage) sont couramment utilisées et restent à ce jour des méthodes qui ont donné des résultats satisfaisants dans la prédiction spatiale des cours d'eau intermittents (Sando et Blasch, 2015). Dans ce cadre, Lin *et al.* (2021) ont proposé une estimation de la fraction  $f_i$  des cours d'eau intermittents (longueur des cours d'eau intermittents sur la longueur totale des cours d'eau par bassin versant) à l'échelle globale, à partir d'une calibration basée sur la classification des cours d'eau intermittents dans le réseau hydrographique de référence des États-Unis (NHDPlusV2). Comme le soulignent les auteurs, cette estimation est sujette à de très fortes erreurs, car elle n'a pas été validée en dehors des États-Unis, et, car s'ajoute l'incertitude sur le calcul de la fraction ( $f_i$ ) à partir des données hydrographiques (NHDPlusV2) pour lesquelles la classification de l'intermittence ne coïncide pas très souvent avec les observations de débits (Hafen *et al.*, 2020). Messenger *et al.* (2021) proposent une cartographie à l'échelle globale de la probabilité d'occurrence des cours d'eau intermittents (cours d'eau ayant en moyenne au moins un jour de débit nul par an) en se basant sur une extrapolation des observations d'intermittence du débit le long du réseau hydrographique HydroRIVERS (Lehner et Grill, 2013). Ces deux études ont également identifié les facteurs de contrôle de l'intermittence à l'échelle globale.

Spécifiquement en Afrique, très peu d'études se sont penchées sur la compréhension des facteurs de contrôle de l'intermittence des cours d'eau ainsi que la cartographie de la distribution spatiale de l'intermittence (Belemtougri *et al.*, 2021a). Cela peut être dû à la difficulté d'accès aux données de débits ainsi que le faible nombre de stations hydrométriques (Hannah *et al.*, 2011).

Ce chapitre propose, pour ce continent, une classification et une cartographie de l'intermittence plus détaillée (basée sur quatre classes d'intermittence) que dans les études globales récentes qui fournissent une classification et cartographie binaire de l'intermittence. Pour ce faire, un réseau hydrographique qui tient compte de la variabilité spatiale de la densité de drainage dans la représentation des cours d'eau (Schneider *et al.*, 2017) a été utilisé. Un jeu d'observations de débits, issues principalement des bases GRDC et SIEREM (Boyer *et al.*, 2006), a été collecté. De nombreuses variables caractérisant plusieurs composantes environnementales, telles que le climat, l'hydrologie, la géologie, la topographie, les propriétés du sol et les influences anthropiques, ont été investiguées comme facteurs potentiels de contrôle de l'intermittence. Plusieurs modèles de classification et régression de type *Random Forest* (RF) (Breiman, 2001) ont été explorés pour identifier les facteurs de contrôle principaux et caractériser la distribution spatiale des cours d'eau intermittents. Le modèle RF a déjà montré son utilité dans la régionalisation des indices hydrologiques (Addor *et al.*, 2018 ; Tyrallis *et al.*, 2019) et a été utilisé dans de nombreuses études pour prédire la distribution spatiale des cours d'eau intermittents de l'échelle d'un bassin versant à l'échelle globale (González-Ferreras et Barquín, 2017 ; Reynolds *et al.*, 2015). Une présentation très détaillée du modèle de *Random Forest* est faite au **chapitre I**, et il est appliqué ensuite pour prédire l'intermittence au Burkina Faso en **chapitre II**. Ce **chapitre III** vise donc à :

- i) fournir une classification des observations de jaugeage en termes d'intermittence du débit en Afrique ;
- ii) déterminer les facteurs de contrôle / les déterminants les plus importants expliquant la distribution spatiale des différentes classes d'intermittence ;
- iii) prédire l'étendue spatiale des différentes classes d'intermittence sur l'ensemble du linéaire hydrographique.

Cette étude devrait fournir des informations précieuses pour la conservation des cours d'eau, la modélisation hydrologique et les études environnementales.



## III - 3 Données et méthodes

### III - 3 - 1 Sélection des stations hydrométriques

#### III- 3 - 1 - 1 Sélection basée sur la disponibilité des données de débit

Comme détaillé plus bas, la sélection des stations de jaugeage pour cette étude s'est basée sur de nombreux critères, à savoir la longueur de la chronique de données, l'unicité des stations de jaugeage (pas de doublons), la période couverte, ainsi que la qualité des données. Le jeu de données de débits utilisés dans ce chapitre provient majoritairement de deux bases de données : la base GRDC (*Global Runoff Data Center* ; GRDC, 2019) et la base SIEREM (*Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation* ; Boyer *et al.*, 2006). Ce jeu de données a été complété avec celles de quelques stations collectées auprès de la DGRE (*Direction Générale des Ressources en Eau*) au Burkina Faso et qui n'existent pas dans ces bases.

Le jeu de données initialement collectées est composé de 2714 stations dont 1571 stations à pas de temps mensuel de la base GRDC, 1135 stations (dont 1019 à pas de temps journalier) de la base SIEREM (dont beaucoup sont dupliquées avec la base GRDC), et 8 stations à pas de temps mensuel de la DGRE. En Afrique, la densité des stations hydrométriques et pluviométriques reste faible par rapport aux autres continents. À cela s'ajoute la réticence des différents pays à partager ces informations hydro-pluviométriques (Dixon *et al.*, 2020 ; Fekete et Vörösmarty, 2007). Afin d'obtenir des chroniques suffisamment longues pour caractériser la distribution spatiale de l'intermittence des cours d'eau en Afrique, le pas de temps mensuel a été préféré au pas de temps journalier, comme dans le **chapitre II** au Burkina Faso.

Pour y parvenir, le jeu de données initiales a d'abord suivi le processus de sélection et traitement suivant :

- 1 - Toutes les stations disposant de données de débits journaliers qui ont moins de 5 jours de données manquantes par mois (soit moins de 14 %) ont été moyennées en données mensuelles de débit, autrement elles n'ont pas été prises en compte ;
- 2 - Pour chaque station, seules les années complètes (ayant des valeurs mensuelles pour les 12 mois) sont conservées ;

- 3 - Seules les stations avec au minimum 4 ans de données sont conservées pour rester cohérent avec l'étude faite au **chapitre II** ;
- 4 - Plusieurs de ces stations sont dupliquées dans les différentes bases, il est donc important de les identifier. Les listes des stations ont été fusionnées et les stations ayant le même nom ont été indexées et inspectées minutieusement. Pour les stations ayant le même nom, la station retenue est celle qui possède la chronique de données la plus longue, pas nécessairement des années consécutives. Une deuxième inspection, visuelle et par tri de coordonnées géographiques, est faite par pays et les stations très proches (distances de moins de 5 km) sont épinglées. Celles retenues parmi ces doublons sont encore celles qui ont la chronique la plus longue. Enfin, toutes les stations identifiées dans les bases de données comme appartenant à un pays donné et qui se retrouvent par leurs coordonnées géographiques dans un autre pays ont été simplement supprimées.

Ainsi, à l'issue de ce prétraitement, un ensemble de 1513 stations a été jugé conforme aux exigences susmentionnées (1085 GRDC, 420 SIEREM, 8 DGRE). Les années renseignées à ces stations couvrent la période 1871-2016 (**Figure III-1**), les données plus anciennes provenant de la station du Nil à Assouan en Égypte.

La **Figure III-1** montre que la période offrant la plus forte disponibilité de stations par an s'étale entre 1958-1991 (34 ans), avec un maximum dans les années 1970, et une baisse drastique des stations de jaugeage à partir des années 1970 et qui se poursuit après 1990, comme observé également sur la **Figure II-2.b** du **chapitre II** ainsi que dans certaines études précédentes (Crochemore *et al.*, 2020 ; Dieulin *et al.*, 2019 ; Tramblay *et al.*, 2021a). À partir des années 1990, les stations se limitent sans doute aux grandes rivières. En conséquence, nous avons encore réduit notre sélection en ne conservant que les stations de jaugeage ayant au minimum 4 ans de données sur la période 1958-1991, ce qui permet de considérer au moins 400 stations par an tout au long de la période à investiguer (**Figure III-1**). Cette sélection comprend 1437 stations avec au moins 4 ans de données (1047 GRDC, 385 SIEREM, 5 DGRE).

Le choix d'une période d'étude 1958-1991 permet aussi d'exploiter pleinement le jeu de données climatiques à haute résolution TerraClimate (Abatzoglou *et al.*, 2018) déjà utilisé dans le chapitre précédent et disponible sur la période 1958-2017. Enfin, restreindre l'étude avant 1991 permet de réduire l'influence anthropique et donc de traiter des écoulements moins perturbés par les aménagements éventuels depuis 1991.

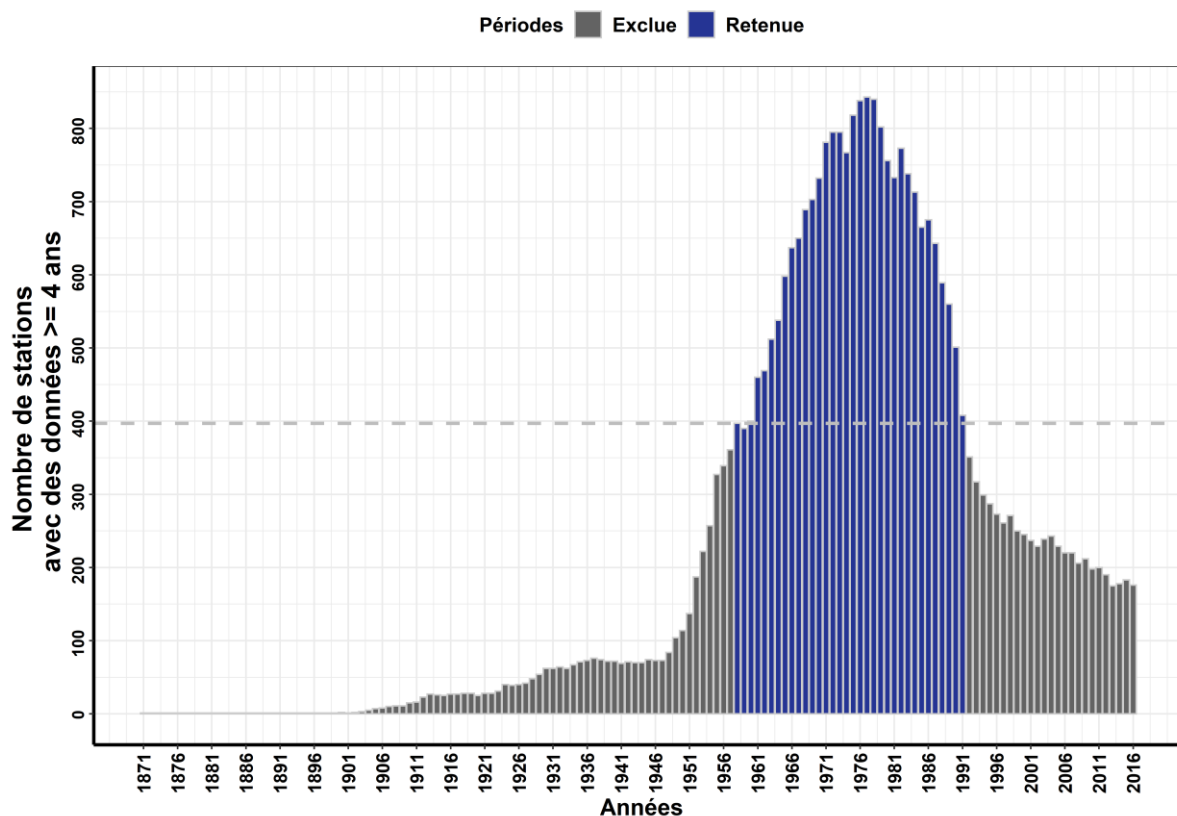


Figure III-1 : Distribution du nombre de stations disponibles par an entre 1871 et 2016. La période d'étude finalement sélectionnée 1958-1991 est matérialisée en bleu.

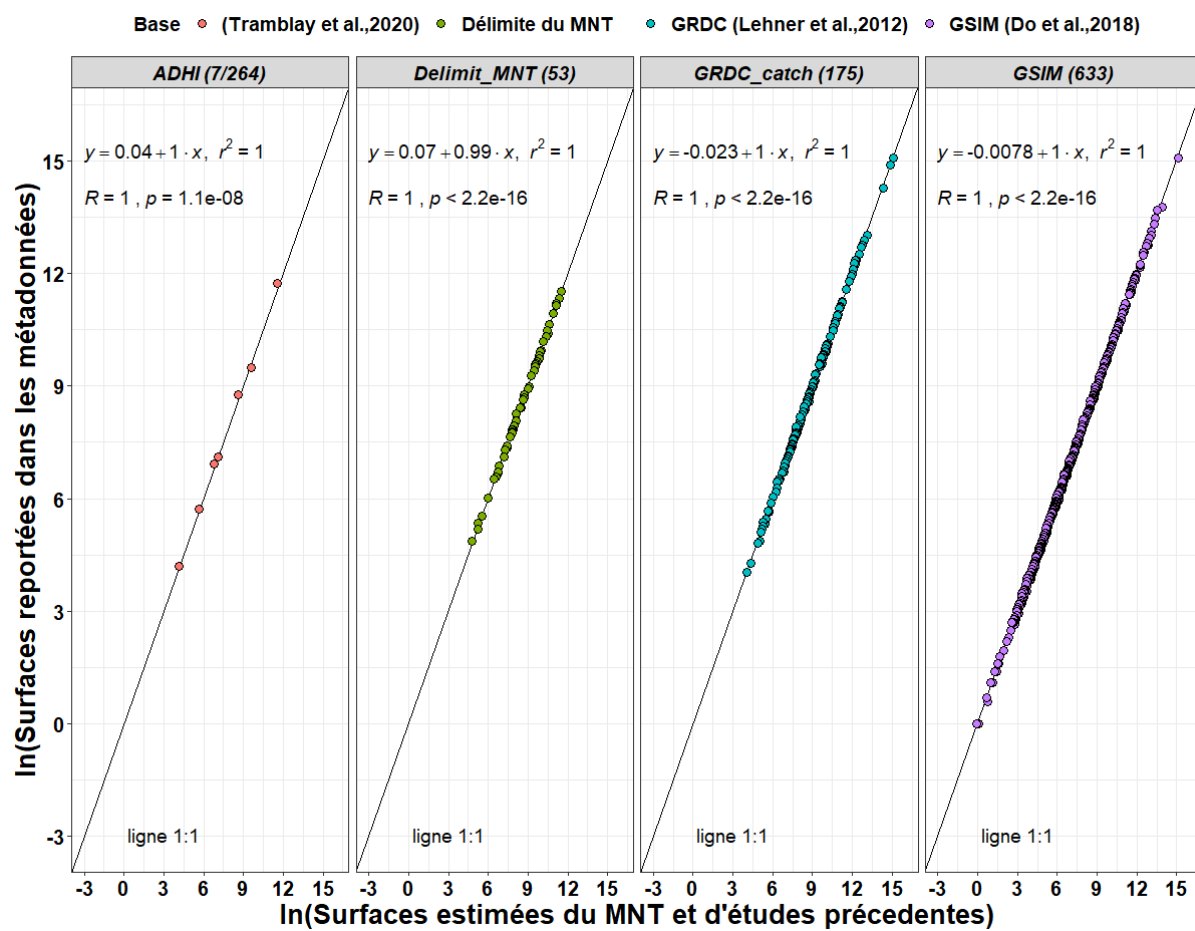
### III- 3 - 1 - 2 Sélection basée sur la connaissance du bassin versant amont

Le contour des bassins versants amont des stations est nécessaire pour quantifier l'influence de nombreux facteurs de contrôle de l'intermittence (section III-3-3). Pour la plupart des 1437 stations présélectionnées sur la période 1958-1991, ces contours ont été récupérés de différentes bases de données issues d'études récentes, ou alors délimitées à partir de MNT. Ces contours de bassins sont de préférence récupérés d'études récentes, car des analyses approfondies ont déjà été menées station par station pour mieux repositionner les stations sur les cours d'eau afin de délimiter les bassins versants correspondants. Les bassins versants de la majorité des stations issues de GRDC sont récupérés de la base GSIM (*Global Streamflow Indices and Metadata Archive* ; Do et al., 2018) et de l'étude « *GRDC report number 41* » (Lehner, 2012) que nous appelons dans ce chapitre base « *GRDC\_catch* ». La majorité des bassins versants de la base SIEREM sont récupérés de la base ADHI (*African Database of Hydrometric Indices* ; Trambly et al., 2021a).

Généralement, les bassins versants des stations qui ont une chronique de débits inférieure à 10 ans ne sont pas délimités par aucune étude. Les contours de bassins versants de ces stations ont été délimités à partir de MNT. Une procédure automatique a été adoptée pour délimiter les contours de bassins versants à partir du MNT HydroSHEDS (Lehner *et al.*, 2008) à 15'' de résolution (environ 500 x 500 m à l'équateur), car c'est le MNT dont dérive le réseau hydrographique LCS également utilisé dans cette étude. À cet effet, tous les pixels rivières de la grille d'accumulation dans un rayon de 5 km (10 pixels) autour d'une station donnée ont été utilisés pour délimiter des bassins versants. Celui dont la superficie est la plus proche de la superficie du bassin versant de la station selon les métadonnées d'origine a été retenu comme bassin de la station, déplacée au « *pixel exutoire* » de ce bassin versant par MNT. Malgré un contrôle minutieux des coordonnées des stations, il est à noter que la délimitation des bassins versants par MNT est souvent sujette à erreur notamment dans les zones plates (Al-Muqdadi et Merkel, 2011).

Une partie des stations étudiées n'ont cependant pas de métadonnée sur la surface des bassins amont, ce qui rend impossible la comparaison avec les valeurs estimées d'études et de MNT. Sur les 1437 stations disponibles, les surfaces de 1006 stations ont pu être comparées avec les informations de surface présentes dans les métadonnées d'origine, ce qui a amené à supprimer 138 stations avec une erreur sur la surface du bassin versant extrait de plus de 15 %. Sur les 431 stations dont les surfaces de bassin versant n'ont pas pu être validées, 257 contours de bassin ont été récupérés de la base ADHI (Tramblay *et al.*, 2021a) où les auteurs ont pu vérifier que les informations renseignées dans les archives de l'IRD (ex ORSTOM) ou de la connaissance du terrain ne diffèrent pas de plus de 10% de la surface estimée dans l'étude, si bien que ces stations ont été incluses dans notre étude. Les 174 stations restantes ont été supprimées.

À partir des 2714 stations initialement collectées, les différentes étapes de sélection ont permis de retenir un jeu final de 1125 stations non dupliquées sur la période 1958-1991, avec au minimum 4 ans de données mensuelles complètes et des surfaces de bassin versant qui ne diffèrent pas de plus de 15% des informations renseignées dans les métadonnées d'origine (**Figure III-2**). À l'issue de ces processus, les coordonnées des stations ont été mises à jour en tenant compte des informations présentes dans les études utilisées ou des délimitations dans lesquelles les bassins versants ont été extraits (« *pixel exutoire* »).

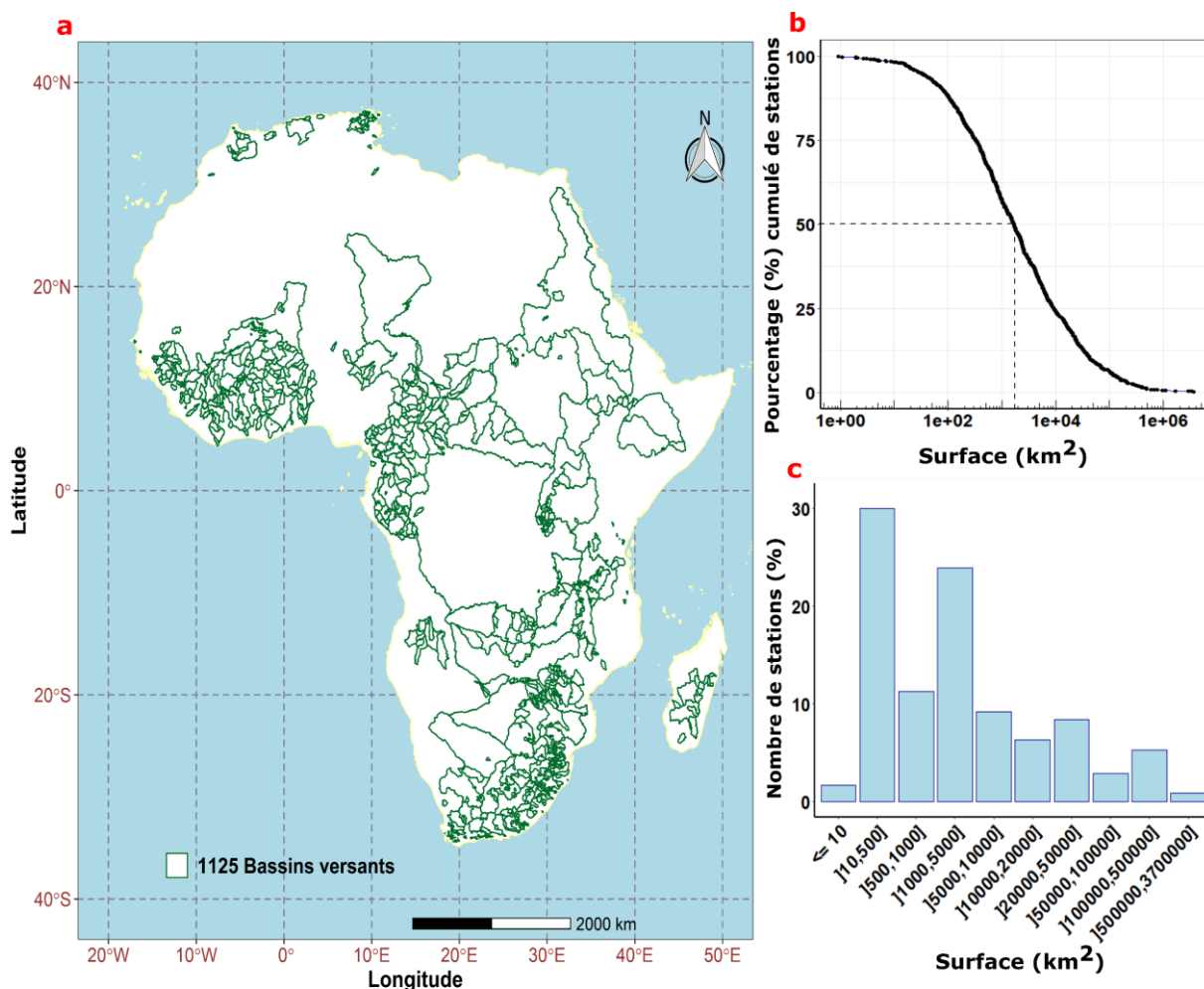


**Figure III-2 :** Comparaisons des surfaces des bassins versants récupérés d'études ou délimités par MNT avec celles reportées dans les métadonnées d'origine avec une différence de surface de moins de 15 %. Dans ce graphique, chaque colonne représente une étude ou donnée délimitée par MNT. En dernière colonne on peut lire par exemple que les surfaces de 633 bassins récupérés de l'étude de Do et al. (2018) ont pu être comparées avec les informations renseignées dans les métadonnées.

La **Figure III-3.a** montre la répartition des contours de bassins versants que contrôlent les 1125 stations retenues à l'échelle de l'Afrique. La taille des bassins versants varie de 1 à 3 634 864 km<sup>2</sup> avec une médiane et moyenne de 1665 km<sup>2</sup> et 35 495 km<sup>2</sup> respectivement. Parmi ces bassins, 50 % ont une surface inférieure à 1803 km<sup>2</sup> (**Figure III-3.b**) et environ 40% des bassins versants peuvent être classés comme bassin versant de petite taille (< 1000 km<sup>2</sup>) (**Figure III-3.c**). Une bonne partie des bassins versants de surface inférieure à 10 km<sup>2</sup> est localisée en Afrique australe notamment en Afrique du Sud. Le bassin de 3 634 864 km<sup>2</sup> correspond au fleuve Congo à Brazzaville.

Les stations issues de la base de données GRDC sont réparties spatialement un peu partout en Afrique avec une forte densité en Afrique australe notamment en Afrique du Sud tandis que les stations issues de SIEREM se concentrent plutôt en Afrique de l'Ouest, du Centre et du Nord et particulièrement dans les pays francophones (**Figure III-4.a**). Ces stations du

SIEREM complètent bien le manque de données GRDC dans les pays comme la Tunisie, la Guinée, le Bénin, la Côte d'Ivoire, le Gabon et le Rwanda. Il faut cependant souligner l'absence de station dans les grands déserts africains (Sahara, Namibie) et une faible densité de stations dans quelques autres pays comme la République Démocratique du Congo ou le Nigéria, hormis sur le cours principal des fleuves Congo et Niger.



**Figure III-3 :** Répartition spatiale des bassins versants étudiés (1125) à l'échelle de l'Afrique sur la période 1958-1991 (a), pourcentage cumulé des stations en fonction de la surface amont des bassins versants (b), distribution des stations étudiées en fonction de différentes classes de surface (c).

Sur la période étudiée (1958-1991), les stations ayant les chroniques les plus longues se localisent en Afrique australe (principalement l'Afrique du Sud) (**Figure III-4.b**) et sont à majorité issues de la base GRDC. En Afrique de l'Ouest, les stations disposant des plus longues chroniques sont localisées au Mali et au Burkina Faso respectivement sur le fleuve Niger et le fleuve Mouhoun. Au final, plus de la moitié des stations prises en compte ont au moins 15 ans de données (**Figure III-5**).

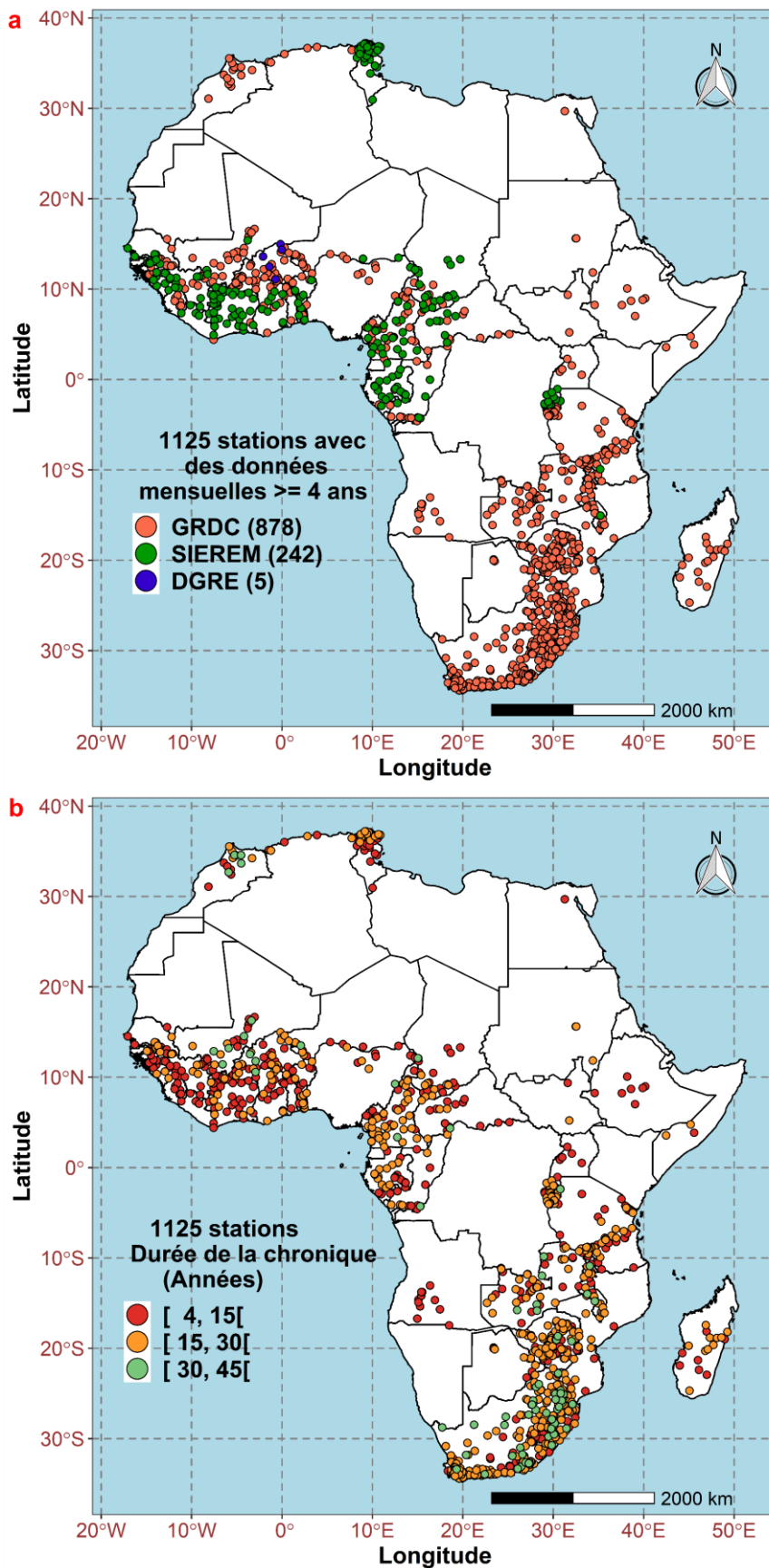


Figure III-4 : Répartition des stations de jaugeage par base de données sur la période 1958-1991 (a), et par durée de chronique (b).

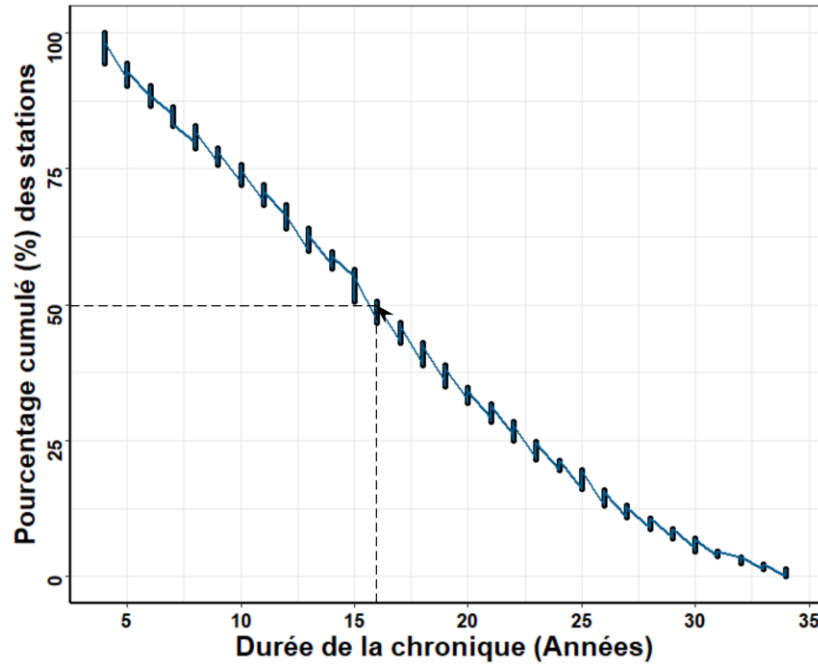


Figure III-5 : Pourcentage cumulé des stations en fonction de la durée de la chronique sur la période 1958-1991.

### III - 3 - 2 Classification de l'intermittence des cours d'eau

La moyenne du nombre de mois secs  $\overline{Ndry}$  a été utilisée pour regrouper les stations finales considérées (1125) en quatre classes d'intermittence (**Tableau III-1**) comme dans le **chapitre II**. Il se calcule selon l'équation (III-1) suivante :

$$\overline{Ndry} = \frac{\sum_i^n Ndry(i)}{n} \quad (III-1)$$

où  $Ndry(i)$  est le nombre de mois secs dans une année donnée  $i$  et  $n$  est le nombre d'années considérées. Un déséquilibre s'observe à la vue des proportions des différentes classes d'intermittence avec la classe permanente largement dominante par rapport au reste des classes (**Tableau III-1**).

Pour apprécier la saisonnalité de l'intermittence (mois moyen d'apparition de l'intermittence) dans l'année à l'échelle de l'Afrique, nous nous sommes appuyés sur des métriques de statistiques directionnelles (Burn, 1997 ; Sauquet *et al.*, 2021 ; Trambly *et al.*, 2021a) afin d'éviter les ruptures entre les derniers mois de l'année et les premiers mois de l'année suivante dans le calcul de la moyenne.



Tableau III-1 : Les classes d'intermittence des cours d'eau et leur description.

Classes d'intermittence	Description	$\overline{Ndry}$ (mois/an)	Proportion des classes d'intermittence (%)
<b>Permanent</b>	S'écoule plus de 90 % du temps.	(0-1)	70,7
<b>Faiblement intermittent</b>	Caractérisé par des tronçons de cours d'eau stagnants et isolés pendant la saison sèche.	(2-4)	12,5
<b>Fortement intermittent</b>	Marqué par une absence d'eau pendant la saison sèche.	(5-7)	9,6
<b>Éphémères</b>	Généralement, il s'agit de petits cours d'eau de tête de bassin qui s'écoulent en réponse directe aux précipitations.	(8-12)	7,2

Les mois d'occurrence de l'intermittence par station ont donc été convertis en valeurs angulaires (le mois de décembre correspondant par exemple à  $2\pi$ ) pour calculer le mois d'occurrence ( $\theta$ ) (équation (III-2)) :

$$\theta_i = (mois)_i * \left(\frac{2\pi}{12}\right) \quad (III-2)$$

avec  $\theta_i$  la valeur angulaire en radians pour le mois  $i$  à débit nul. Pour une station donnée, tous les mois à débits nuls peuvent être convertis en valeur angulaire  $\theta_i$ . Les coordonnées moyennes  $\bar{x}$  et  $\bar{y}$  de la date moyenne d'occurrence peuvent être calculées par les formulations de l'équation (III-3) et équation (III-4) :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \quad (III-3)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \quad (III-4)$$

L'angle moyen (date moyenne) d'occurrence de l'intermittence  $\bar{\theta}$  pour une station donnée peut alors être obtenue par l'équation (III-5) suivante :

$$\bar{\theta} = \arctan\left(\frac{\bar{x}}{\bar{y}}\right) \quad (III-5)$$

Une mesure de la variabilité de l'occurrence de l'intermittence ( $r$ ) autour du mois moyen d'occurrence de l'intermittence peut se calculer selon l'équation (III-6) :

$$r = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (\text{III-6})$$

avec  $0 < r \leq 1$  et  $r$  proche de 1 implique une faible variation et une forte concentration des données autour du mois moyen d'occurrence, et  $r$  proche de 0 indique une forte variation et une grande dispersion autour du mois moyen d'occurrence ( $\bar{\theta}$ ).

### III - 3 - 3 Réseau hydrographique et variables prédictives hydro-environnementales

Dans l'optique de cartographier l'intermittence des cours d'eau à l'échelle de l'Afrique nous sommes appuyés sur une représentation numérique du réseau hydrographique (LCS) (Schneider *et al.*, 2017), dérivée du MNT HydroSHEDS à la résolution de 500 m et dont les cours d'eau sont initiés à partir d'aires contributives variables (cf. **chapitre I**). Ce réseau hydrographique est composé de 15 500 000 tronçons de rivières avec une longueur moyenne de 1,33 km et une longueur totale en Afrique de 20 682 677 km. Un tronçon de cours d'eau est défini comme un segment ou portion ou section de rivière situé entre deux affluents ou entre un affluent et le début ou la fin d'une rivière (section **I-3**).

Un total de 23 variables environnementales réparties entre l'hydrologie, le climat, la topographie, l'influence anthropique, la géologie et les propriétés du sol ont été investiguées comme facteur de contrôle potentiel de l'intermittence des cours d'eau (**Tableau III-2**). Elles ont été sélectionnées en se basant sur une revue de la littérature (Beck *et al.*, 2017 ; González-Ferreras et Barquín, 2017 ; Hammond *et al.*, 2021 ; Messenger *et al.*, 2021 ; Olden *et al.*, 2021 ; Trambly *et al.*, 2021b ; Zipper *et al.*, 2021). Ce choix s'est aussi basé sur la disponibilité temporelle, la qualité (données manquantes dans certains endroits) et la résolution des données disponibles. Toutes ces données ont donc été inspectées pour identifier toute lacune spatiale et pour vérifier que tous les ensembles de données couvrent bien l'étendue de l'Afrique. Ainsi, la variable SPEI (Standardized Precipitation-Evapotranspiration Index ; Vicente-Serrano *et al.*, 2010) qui avait initialement été sélectionnée a été supprimée finalement à cause d'un manque de données assez important sur les bordures des côtes africaines. L'indice d'aridité qui est investigué comme contrôle de l'intermittence est défini comme le rapport (ETP/P) au **chapitre II**. À l'échelle de l'Afrique l'opérationnalité de cette définition posait problème notamment dans les zones désertiques où P est largement inférieure à ETP ce qui conduit à une plage de variation trop importante par rapport aux autres zones. Pour restreindre cette plage de variation, dans ce **chapitre III** l'indice d'aridité est

défini comme le ratio entre la précipitation et l'évapotranspiration (Messenger *et al.*, 2021 ; Sauquet *et al.*, 2021 ; Zipper *et al.*, 2021). Une description détaillée des bases de données associées à l'ensemble de ces variables environnementales est proposée en (section I-2).

Pour chacun des 1125 bassins versants amont comme des 15 500 000 bassins versants amont associés respectivement aux stations de jaugeage et aux tronçons de rivières, des valeurs amont caractéristiques de ces variables environnementales sont calculées selon les spécifications définies dans la deuxième colonne du **Tableau III-2**.

**Tableau III-2** : Liste des variables hydro-environnementales (23) examinées comme contrôle de l'intermittence. En gras, les variables indépendantes sélectionnées pour les analyses ultérieures.

Abréviation des variables hydro-environnementales	Description (Unité) - Échelle de calcul	Valeurs moyennes (min, max) sur les 1125 stations	Source des données, référence et résolution spatio-temporelle
<b>Topographie (n = 6)</b>			
<b>Surf</b>	<b>Surface du bassin versant (km<sup>2</sup>)</b>	<b>35 495 (1 ; 3 634 864)</b>	<b>MNT HydroSHEDS à 15'' (500 m x 500 m)</b> (Lehner <i>et al.</i> , 2008)
Order	Ordre de Strahler de la rivière (-) au point exutoire	6 (1 ; 11)	Dérivé du réseau hydrographique LCS (Schneider <i>et al.</i> , 2017)
<b>Pente</b>	<b>Pente moyenne du bassin versant (°)</b>	<b>3 (0,03 ; 16,4)</b>	<b>MNT HydroSHEDS à 15'' (500 m x 500 m)</b> (Lehner <i>et al.</i> , 2008)
<b>Alti</b>	<b>Altitude moyenne du bassin versant (m)</b>	<b>954 (32 ; 2950)</b>	<b>MNT HydroSHEDS à 15'' (500 m x 500 m)</b> (Lehner <i>et al.</i> , 2008)
<b>Topowi</b>	<b>Topographic wetness index au point exutoire ln(m)</b>	<b>96 (63 ; 123,6)</b>	<b>MNT MERIT hydro</b> , (Hengl, 2018) à 500 m
<b>Tci</b>	<b>Topographic climate index au point exutoire ln(m)</b>	<b>3,8 (0,01 ; 6,7)</b>	<b>Calculé du MNT MERIT hydro</b> (Hengl, 2018) à 500 m <b>et de TerraClimate</b> (Abatzoglou <i>et al.</i> , 2018) à 4km, période 1958-1991, rééchantillonner à 500 m
<b>Climat (n = 8)</b>			
<b>P</b>	<b>Précipitation annuelle moyenne du bassin versant (mm)</b>	<b>1018 (58 ; 2911)</b>	<b>TerraClimate</b> (Abatzoglou <i>et al.</i> , 2018) à 4 km, période 1958-1991

<b>Etp</b>	<b>Évapotranspiration potentielle annuelle moyenne du bassin versant (mm)</b>	<b>1428 (886 ; 2323)</b>	<b>TerraClimate (Abatzoglou et al., 2018) à 4 km, période 1958-1991</b>
Etr	Évapotranspiration réelle annuelle moyenne du bassin versant (mm)	753 (56 ; 1254)	TerraClimate (Abatzoglou et al., 2018) à 4 km, période 1958-1991
<b>Arid</b>	<b>Indice d'aridité (P/Etp) moyen du bassin versant (-)</b>	<b>0,76 (0,03 ; 2,5)</b>	<b>TerraClimate (Abatzoglou et al., 2018) à 4 km, période 1958-1991</b>
<b>Tmoy</b>	<b>Température de l'air moyenne annuelle du bassin versant (°C)</b>	<b>19,7 (6 ; 29)</b>	<b>TerraClimate (Abatzoglou et al., 2018) à 4 km, période 1958-1991</b>
<b>Nwet</b>	<b>Nombre moyen de mois pluvieux par an (-)</b>	<b>8 (2 ; 12)</b>	<b>TerraClimate (Abatzoglou et al., 2018) à 4 km, période 1958-1991</b>
Rech	Recharge annuelle moyenne du bassin versant (mm)	74,5 (1,61 ; 277,8)	Recharge annuelle moyenne (MacDonald et al., 2021) à 0,28 ° (31 km), période 1970-2019
Swc	Teneur en eau du sol annuelle moyenne du bassin versant (%)	51 (3 ; 95)	Teneur en eau du sol (Trabucco et Zomer, 2010) à 1 km, période 1950 – 2000
<b>Géologie et propriétés des sols (n = 7)</b>			
Clay	Teneur en argile du sol moyenne du bassin versant (%)	27,6 (0,13 ; 46)	Soil grids (Hengl et al., 2017) à la résolution de 250 m
Silt	Teneur en silt du sol moyenne du bassin versant (%)	19,4 (1,5 ; 37,6)	Soil grids (Hengl et al., 2017) à la résolution de 250 m
<b>Sand</b>	<b>Teneur en sable du sol moyenne du bassin versant (%)</b>	<b>49 (5 ; 98)</b>	<b>Soil grids (Hengl et al., 2017) à la résolution de 250 m</b>
Ndvi	Indice de végétation normalisé moyen du bassin versant (-)	0,5, (0,12 ; 0,8)	NDVI 3g (Pinzon et Tucker, 2014) à 8 km, période 1981-1991
<b>Grwstor</b>	<b>Stockage des aquifères moyens du bassin versant (mm)</b>	<b>6523 (475 ; 50000)</b>	<b>Stockage souterrain des aquifères (MacDonald et al., 2012) à 30 km</b>

<b>Wtd</b>	<b>Profondeur de la nappe moyenne du bassin versant (m)</b>	<b>38 (7 ; 160)</b>	<b>Profondeur de la nappe (Fan <i>et al.</i>, 2013) à 1 km</b>
<b>Perm</b>	<b>Log<sub>10</sub> de la perméabilité moyenne du bassin versant log<sub>10</sub>(m<sup>2</sup>) - amont</b>	<b>-1362 (-1649 ; -1051)</b>	<b>Perméabilité (Huscroft <i>et al.</i>, 2018) couche polygone</b>
<b><i>Influence anthropique (n = 1)</i></b>			
<b>Damstor</b>	<b>Capacité de stockage maximum des barrages dans le bassin versant (Mm<sup>3</sup>)</b>	<b>991 (0 ; 315 491)</b>	<b>GRanD v1.1 (Lehner <i>et al.</i>, 2011) à l'échelle d'environ 1 : 1 million, avec les barrages sélectionnés sur les années ≤ 1991</b>
<b><i>Hydrologie (n = 1)</i></b>			
<b>Qspec</b>	<b>Débit spécifique moyen annuel du bassin versant (l/s/km<sup>2</sup>), résulte du débit moyen annuel (m<sup>3</sup>/s) rapporté à la surface du bassin versant amont (km<sup>2</sup>)</b>	<b>0,45 (0,0003 ; 24)</b>	<b>FLO1K (Barbarossa <i>et al.</i>, 2018) à 1 km, période 1960 – 1991</b>

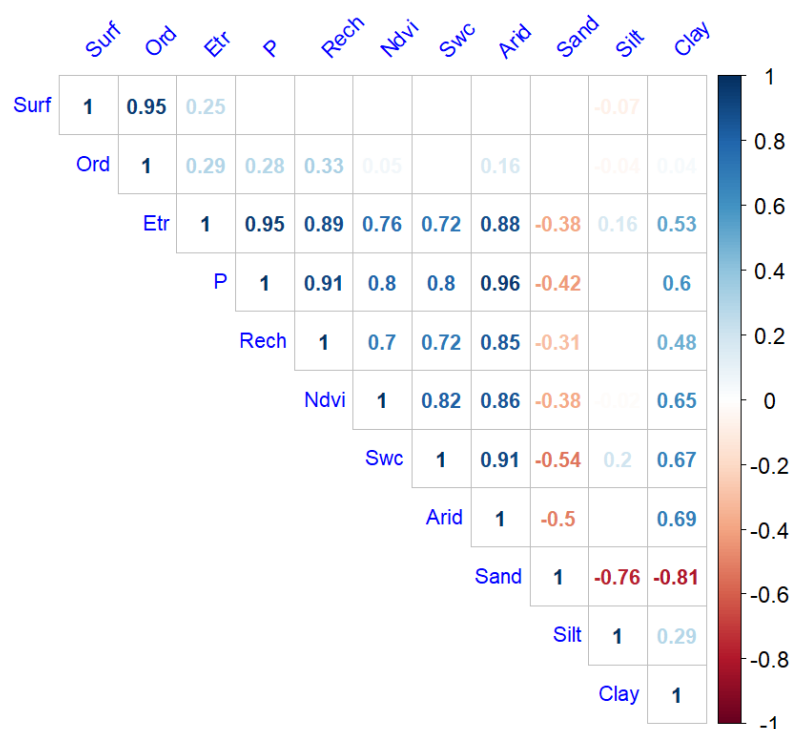
À la suite de la sélection de ces variables, nous avons contrôlé pour chacune d'elles les valeurs au niveau des bassins versants amont associés aux stations (1125) comme au niveau des bassins versants amont associés aux tronçons de rivières (15 500 000) afin de détecter les variables dont la variance est proche de zéro<sup>33</sup> (en raison d'un potentiel prédictif minimal d'une variable donnée lorsque ses valeurs sont constantes ou presque constantes sur l'ensemble d'un jeu d'observation), mais dans le jeu de données investiguées aucune des variables n'est concernée.

La multicolinéarité des variables (corrélation forte entre les variables qui implique une redondance de l'information) n'affecte pas significativement les performances de prédiction du modèle *Random Forest*, mais l'ordre d'importance des variables ainsi que les graphiques de dépendances partielles (analyse de sensibilité) des variables peuvent être biaisés (Boulesteix *et al.*, 2012). Pour pallier à cette éventualité, une matrice de corrélation basée sur la corrélation de rang de Spearman (qui s'adapte à la non-linéarité et aux données non normales ; Kumar *et al.*, 2018) a été générée. Pour cela, une valeur seuil de 0,75 au-delà de

<sup>33</sup> <https://tgmstat.wordpress.com/2014/03/06/near-zero-variance-predictors/#ref1>

laquelle la corrélation est jugée très forte (Beck *et al.*, 2017 ; Yu *et al.*, 2019) a été fixée. Les variables non redondantes ont été sélectionnées selon les critères suivants :

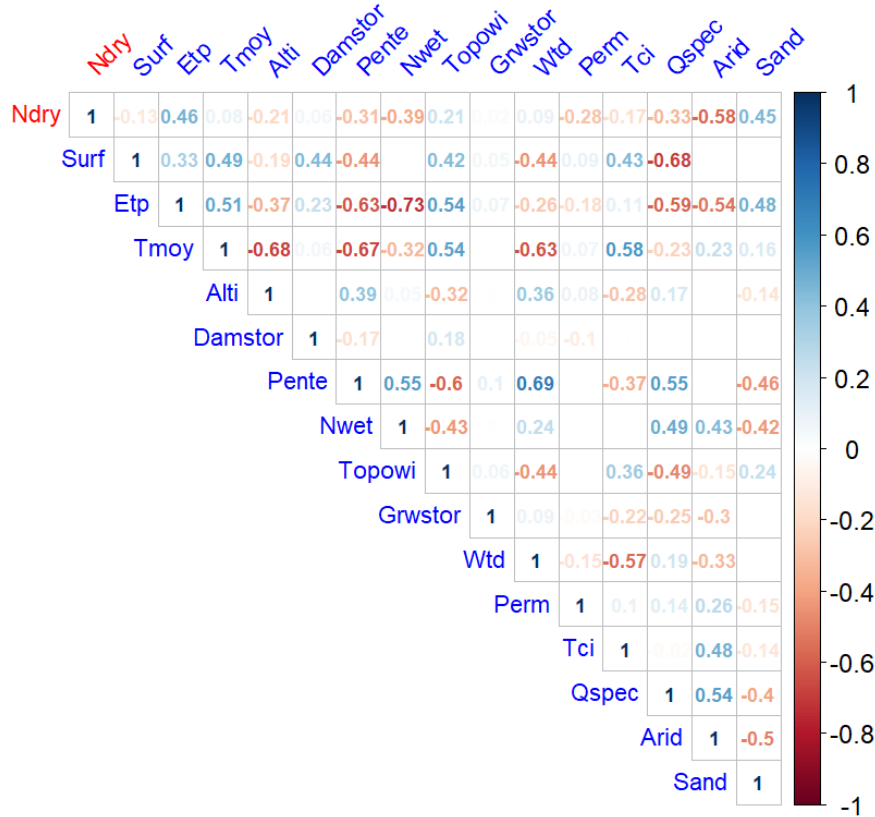
- 1 - Toutes les variables, dont la corrélation de Spearman  $|r| \leq 0,75$  ont été automatiquement sélectionnées pour la suite des analyses ;
- 2 - Le reste des variables fortement corrélées est représenté en **Figure III-6**. Dans ce lot, pour toutes les paires de variables avec un seuil  $|r| > 0,75$ , la variable retenue est celle qui est le plus souvent identifiée comme ayant un fort contrôle sur l'intermittence des cours d'eau dans la littérature, et qui facilite l'interprétation du modèle. Ainsi, parmi les variables topographiques par exemple, la surface des bassins versants est fortement corrélée à l'ordre de Strahler (0,95), et nous avons préféré garder la surface des bassins. Pour les variables climatiques, l'indice d'aridité a été retenu au détriment des cinq variables (Etr, P, Rech, Ndvi, Swc) pour lesquelles les corrélations sont très fortes. Le pourcentage moyen en sable qui est très corrélé aux pourcentages de limon (Silt) et argile (Clay) a finalement été sélectionné.



**Figure III-6** : Matrice de corrélation de Spearman des variables (11) dont la corrélation est supérieure à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ).

La matrice de corrélation des 15 variables retenues confirme leur non redondance (**Figure III-7**). Ces 15 variables apparaissent en gras dans le **Tableau III-2**. Leur distribution

géographique en Afrique au niveau des bassins versants que contrôlent les 1125 stations de jaugeage est proposée en **Annexe-3-1**.



**Figure III-7** : Matrice de corrélation de Spearman des variables (15) dont la corrélation est inférieure ou égale à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ).

### III - 3 - 4 Développement du modèle Random Forest

Une des approches pour la prédiction spatiale consiste à utiliser des modèles prédictifs pour extrapoler les observations locales (*via* les stations de jaugeage) du comportement des cours d'eau à tous les tronçons de cours d'eau (McManamay et DeRolph, 2019). Toutes les stations de jaugeage sélectionnées (1125) ont été positionnées sur des tronçons de cours d'eau donc sur 1125 tronçons de cours d'eau. Un modèle prédictif (RF) a été construit pour lier les variables hydro-environnementales sélectionnées (**Figure III-7**, **Tableau III-2**), aux différentes classes d'intermittence observées au droit des stations de jaugeage (**Tableau III-1**). Le modèle ainsi calibré est appliqué en extrapolation sur les valeurs caractéristiques des bassins versants amont associés aux tronçons de cours d'eau LCS (comme définie dans la colonne 2 du **Tableau III-2**) dans l'optique de prédire les classes d'intermittence. La **Figure**

III-8 résume la méthodologie développée dans ce chapitre, du traitement des données d'entrées à la prédiction du modèle.

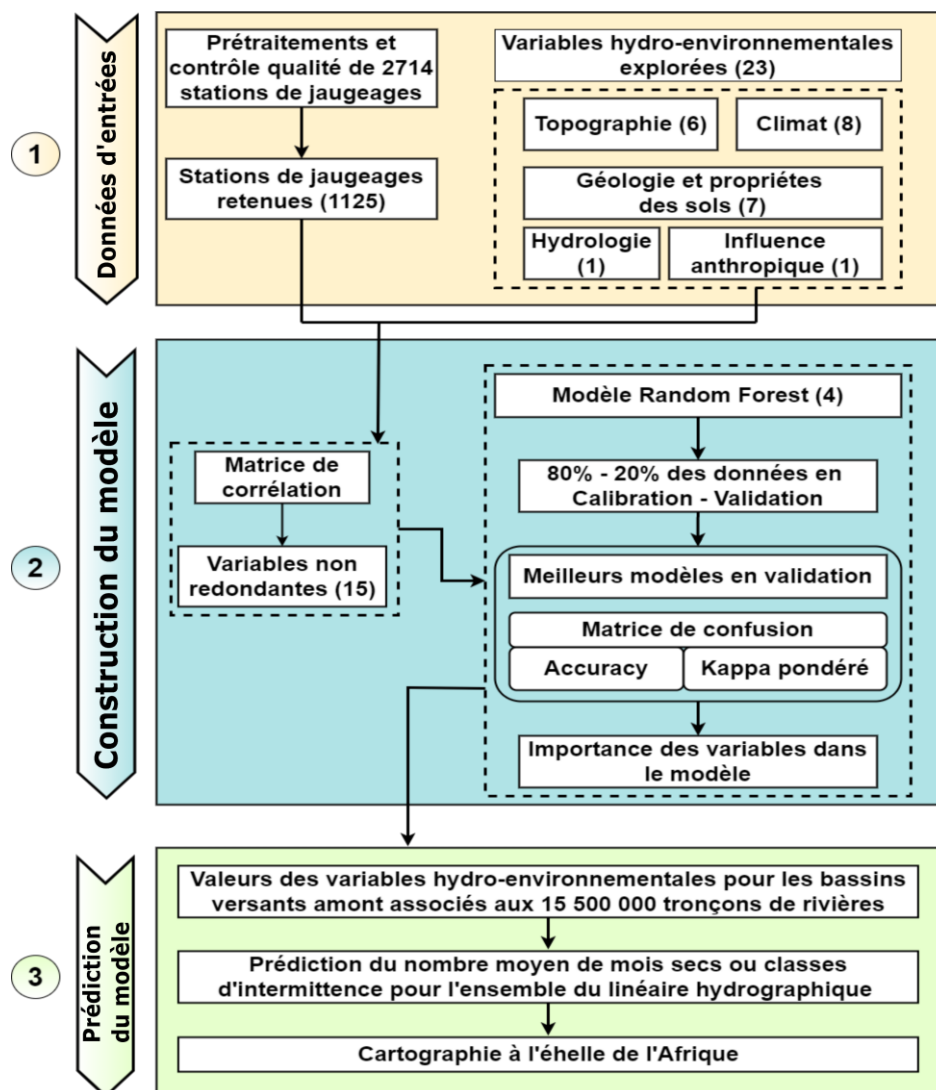


Figure III-8 : Aperçu global de l'approche méthodologique développée dans cette étude.

Le modèle prédictif *Random Forest* (Breiman, 2001) a été implémenté dans le logiciel R (R Core Team, 2021) à l'aide du package 'caret' (Kuhn *et al.*, 2020). Pour entraîner le modèle, les 1125 stations de jaugeage ont été réparties en données de calibration (901) et de validation (224), soit une répartition d'environ 80% et 20% (Beaufort *et al.*, 2019 ; Gupta *et al.*, 2021) en prenant soin d'avoir des proportions des différentes classes d'intermittence similaires à celles observées dans les données initiales, que ce soit en calibration et en validation. Cela permet d'éviter un biais dans la calibration et la validation du modèle. Une visualisation de la localisation des stations de calibration et validation est proposée en **Annexe-3-2**.

Parmi les 1125 stations retenues dans ce chapitre, la proportion de la classe permanente (70,7 %) est environ deux fois plus importante que les autres classes prises toutes ensemble, ce qui



crée un large déséquilibre dans les données (**Tableau III-1**). Ainsi, entraîner un modèle sur un jeu de donnée pareil sans prétraitement pourrait favoriser la classe majoritaire dans la prédiction. Pour traiter ce cas de données déséquilibrées et obtenir un modèle *à priori* robuste, quatre modèles de prédiction ont été explorés dans ce chapitre, dont trois modèles *Random Forest* en classification et un autre en régression (**Tableau III-3**).

**Tableau III-3** : Récapitulatif des différents modèles *Random Forest* explorés

Modèle	Type	Techniques de gestion des données
Modèle 1	Classification	Normal (RF_Class)
Modèle 2	Classification	Sous échantillonnage interne (RF_SI)
Modèle 3	Classification	Sur échantillonnage des données initiales (RF_SMOTE)
Modèle 4	Régression	Normal (RF_Regr)

Dans les trois modèles de classification, le modèle est entraîné pour prédire directement les différentes classes d'intermittence, alors que le modèle en régression vise à prédire des valeurs continues (nombre moyen de mois sec de 0 à 12). Les quatre modèles peuvent être décrits de la sorte :

- 1 - Le premier modèle est un modèle de classification ou aucun traitement sur l'échantillonnage des données n'est effectué ;
- 2 - Le second modèle utilise une technique de sous échantillonnage interne, qui consiste à échantillonner à peu près le même nombre de valeurs par classe dans la construction de chaque arbre au sein du modèle RF (Kuhn et Johnson, 2013 , chapitre 16), de sorte à limiter le déséquilibre initial des quatre classes ;
- 3 - Le troisième modèle utilise une technique de suréchantillonnage appelée SMOTE (*Synthetic Minority Over-sampling Technique* ; Chawla *et al.*, 2002 ; Hoens et Chawla, 2013) qui consiste à créer des données synthétiques interpolées pour les classes minoritaires à l'aide des données initiales. De nouvelles données synthétiques sont ainsi générées uniquement pour les classes faiblement intermittentes, fortement intermittentes et éphémères jusqu'à ce que le nombre d'observations de chacune de ces classes avoisine celui de la classe permanente. Ce travail a été mené grâce au package R '*smotefamily*' ;
- 4 - Le quatrième et dernier modèle est un modèle RF utilisé en régression pour prédire le nombre moyen de mois secs ( $\overline{Ndry}$ ) qui a été converti ensuite en classe d'intermittence selon les critères définis dans le **Tableau III-1**. Les valeurs du  $\overline{Ndry}$  étant des valeurs continues (0 à 12), l'effet du déséquilibre dans les données se fait généralement moins ressentir en régression.

Les critères de performance en validation (*accuracy*, *recall*, *precision*, score F1, Kappa, Kappa pondéré) définis en section I-4-5 ont été utilisés pour sélectionner les modèles qui donnent les meilleures performances. La matrice de coût associé au calcul du critère de Kappa pondéré est présentée en Annexe-3-2.

Comme expliquée dans le chapitre I (section I-4-2), la mise en œuvre du modèle RF nécessite de choisir deux paramètres principaux : le nombre d'arbres élémentaires à construire pour former la forêt (*ntree*), et le nombre de variables prédictives choisies aléatoirement pour diviser les données à chaque nœud des arbres (*mtry*). Bien que les performances du modèle puissent être optimisées pour les paramètres *ntree* et *mtry*, le choix a été fait d'optimiser uniquement le paramètre *mtry*. La même approche que celle définie en section II - 8 dans l'application du modèle RF au Burkina Faso a été utilisée. Pour chacun des modèles de RF exploré, la même approche a été suivie.

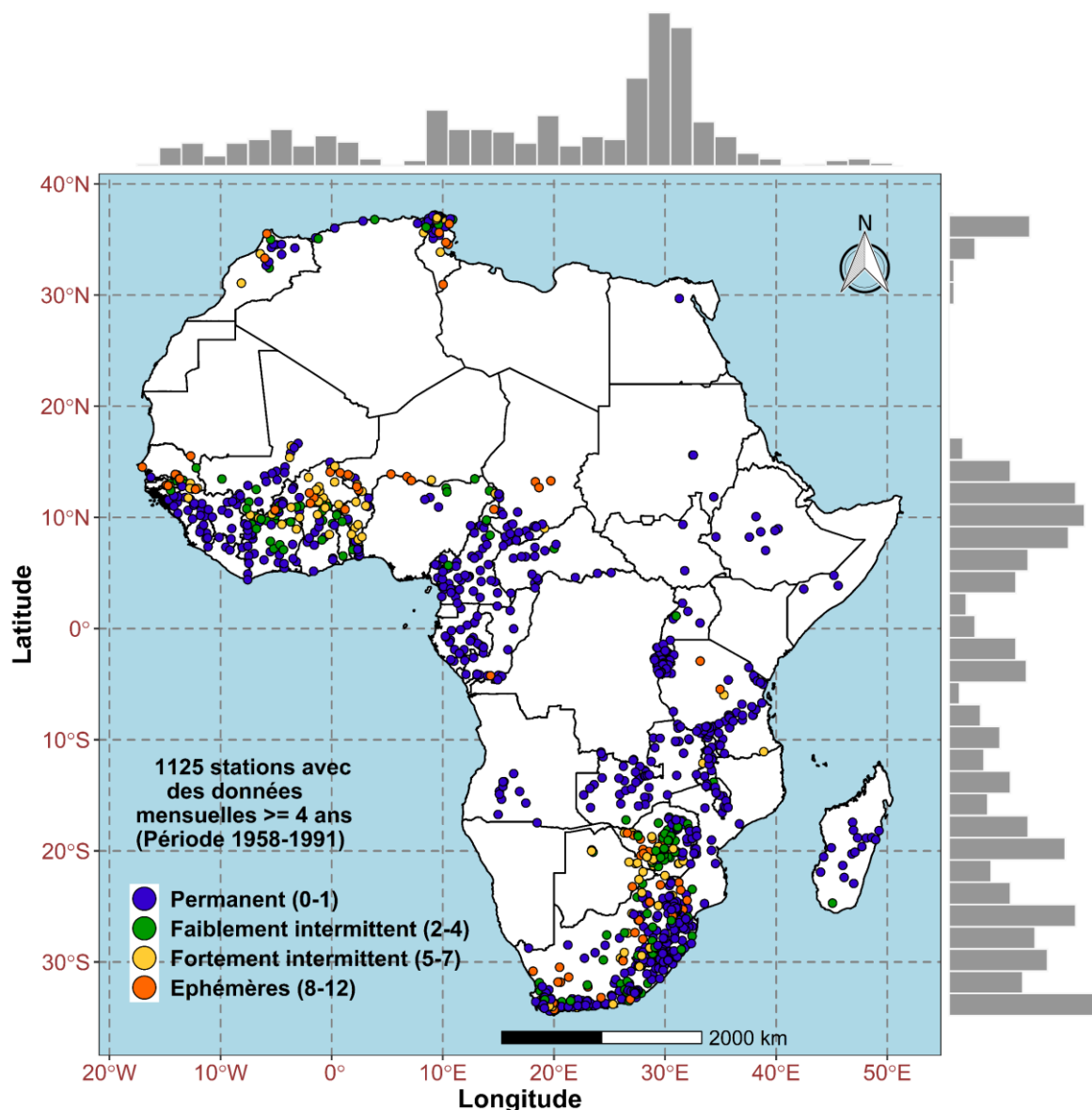
Pour apprécier l'ordre d'importance des variables dans la prédiction de l'intermittence par le modèle RF nous nous sommes appuyés sur l'indice MDG (*Mean Decrease Gini*) et le %incMSE (*percent increase in mean square error*), respectivement pour la classification et la régression. L'analyse des dépendances partielles (une sorte d'analyse de sensibilité) a été effectuée pour les variables les plus importantes à l'aide du package R '*iml*' (Molnar *et al.*, 2018). Il permet d'évaluer individuellement le changement moyen de la variable de réponse par rapport à une variable donnée, illustrent les décisions du modèle et permettent de voir comment le modèle fonctionne sur l'ensemble de données de calibration.

## III - 4 Résultats

### III - 4 - 1 Classes d'intermittence des cours d'eau échantillonnés

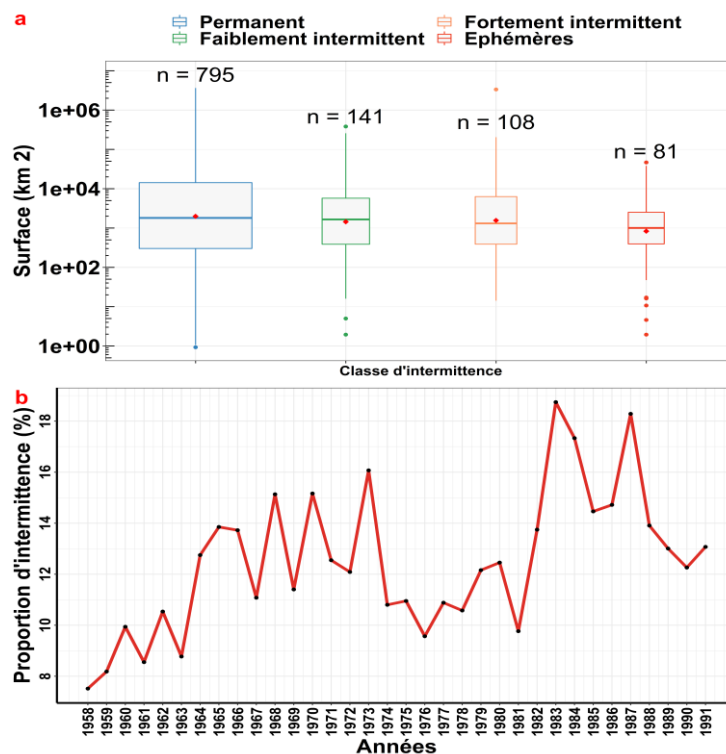
La Figure III-9 présente la distribution spatiale de l'intermittence du débit aux 1125 stations de jaugeage retenues à l'échelle de l'Afrique. Comme déjà notée en Tableau III-1, une large majorité des stations dans le jeu de donnée sont classées permanentes (70,7%), alors que 12,5% sont faiblement intermittentes, 9,6% sont fortement intermittentes et 7,2% sont éphémères. De cette figure, on constate une forte coïncidence des stations intermittentes avec les types de climats arides et semi-arides de l'Afrique (Figure I-4, Figure A1-2.a). Cette coïncidence est d'autant plus prononcée sur toute l'Afrique avec la classe *Bsh* (climat de steppe sec et chaud, Figure I-4, Figure III-9) et en Afrique du Nord avec la classe *BWh*

(climat désertique chaud et sec) et *BWk* (climat désertique froid et sec). Les plus fortes densités de stations intermittentes dans le jeu de données se retrouvent majoritairement localisées en Afrique de l'Ouest et Australe. Dans le jeu de données, en termes d'intermittence, les pays les plus concernés en Afrique de l'Ouest sont le Burkina Faso, le Niger et les parties nord de la Côte d'Ivoire, du Ghana, du Togo, du Bénin et en Afrique australe principalement le Zimbabwe et l'Afrique du Sud.



**Figure III-9 :** Variation géographique des classes d'intermittence du débit (sur la base du nombre moyen de mois secs par an) pour 1125 stations avec des données mensuelles  $\geq 4$  ans. Les histogrammes en marge de la figure sur l'axe x et l'axe y représentent respectivement la distribution des stations selon la longitude et la latitude.

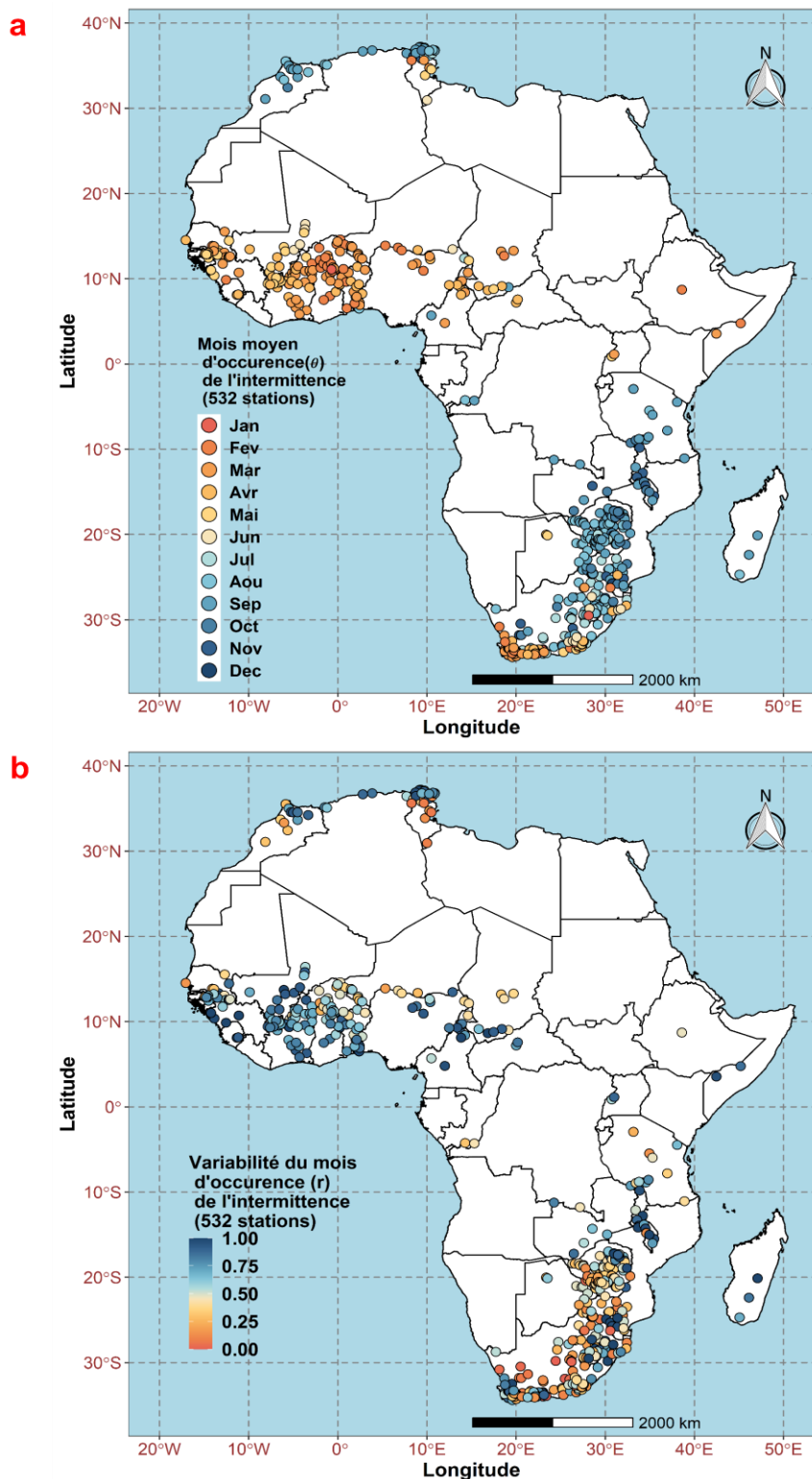
Les classes d'intermittence peuvent varier fortement dans l'espace même pour des stations très voisines (**Figure III-9**) ce qui pourrait être dû à l'influence des caractéristiques locales (géologie, propriétés des sols, etc.). La surface des bassins versants amont des stations prises en compte (**Figure III-10.a**) montre que la moyenne de la surface des bassins versants en fonction des différentes classes d'intermittence est à peu près la même avec des valeurs moyennes autour de 1000 km<sup>2</sup> bien qu'une part importante des petits bassins versants soit classée permanente. Sur la période investiguée (1958-1991), la proportion d'intermittence varie entre 7,5% (1958) et 18 % (1983) avec une nette tendance à l'augmentation (**Figure III-10.b**). Cette augmentation peut être influencée par la disponibilité des stations par an, mais aussi par la variabilité climatique, par exemple les épisodes de sécheresse qui ont touché l'Afrique de l'Ouest entre les années 1970 et 1990. Au Burkina Faso, cette dernière n'a cependant pas eu d'influence majeure sur la distribution géographique des classes d'intermittence, comme mise en évidence au **chapitre II**.



**Figure III-10** : Répartition de la surface amont des stations de jaugeage en fonction de leur classe d'intermittence (a), évolution de la proportion d'intermittence sur la période 1958-1991 (b).

Sur les 1125 stations, 532 (soit 47%) comportent au moins un mois sec (mois à débit nul) dans leur chronique de données. Il ressort de l'analyse de la saisonnalité de ces stations (532) à travers le mois moyen d'occurrence de l'intermittence ( $\bar{\theta}$ ) (voir section III-3-2) que les stations peuvent se répartir en deux groupes : un premier groupe de stations dont l'occurrence

de l'intermittence se fait durant les six premiers mois de l'année et un second groupe dont l'occurrence intervient durant les six derniers mois de l'année (**Figure III-11.a**).



**Figure III-11** : Mois moyen d'occurrence de l'intermittence (**a**) ainsi que la variabilité du mois d'occurrence de l'intermittence (**b**) à l'échelle de l'Afrique pour 532 stations comportant au moins un mois sec sur la période 1958-1991.

Le premier groupe se localise en Afrique de l'Ouest et à la pointe de l'Afrique du Sud tandis que le second groupe se localise en Afrique du Nord et une grande partie de l'Afrique Australe. L'analyse de la **Figure III-11.b** montre en général une faible valeur du coefficient de la variabilité du mois d'occurrence ( $r$ ) en Afrique australe et une forte valeur en Afrique de l'Ouest ce qui correspond respectivement à une forte variabilité du mois moyen d'occurrence de l'intermittence en Afrique australe et une faible variabilité en Afrique de l'Ouest. Ce contraste en termes de saisonnalité de l'intermittence entre ces deux groupes pourrait être lié à la saisonnalité des précipitations bien qu'à l'échelle locale, certaines variables environnementales peuvent en être la cause.

### III - 4 - 2 Performances du modèle Random Forest et ordre d'importance des variables

Les modèles RF sont évalués sur les critères de performances en validation (sur 224 stations) et en se basant en particulier sur le critère de Kappa pondéré ( $K_w$ ), car les valeurs d'*accuracy* et de Kappa dans le cas de données déséquilibrées comme le nôtre reflètent plus celles de la classe majoritaire (classe permanente). Les critères de performances des différents modèles explorés sont récapitulés dans le **Tableau III-4** et les matrices de confusion en validation sont représentées en **Figure III-12**. Le critère de Kappa pondéré tient compte de l'ordre inhérent entre les différentes classes en permettant d'affecter des poids moins importants aux erreurs de prédiction quand elles sont proches de la classe prédite et inversement pour les erreurs de prédictions éloignées de la classe observée tandis que le critère de Kappa donne le même poids aux différentes erreurs de prédictions. Le critère de Kappa pondéré varie entre -1 et 1 et s'interprète comme le Kappa avec les valeurs qui tendent vers 1 qui traduisent une meilleure concordance entre les classes prédites et celles observées. Le critère de Kappa pondéré pour les différents modèles explorés n'est pas trop différent. Les modèles présentant les meilleures performances sont par ordre les modèles 3, 4, 2 et 1 avec respectivement des valeurs de Kappa pondéré de 0,42, 0,42, 0,41 et 0,39 (**Tableau III-4**). Les modèles 1 et 4 sont en quelques sortes des modèles de référence puisqu'aucun traitement concernant le déséquilibre des données n'est expérimenté dans ces modèles. Dans la littérature, des valeurs de Kappa pondéré entre 0,41 et 0,60 sont considérées comme assez bonnes (Landis et Koch, 1977) et seuls les modèles 2, 3 et 4 ont des valeurs supérieures ou égales à 0,41. Bien que nous ayons exploré des techniques visant à équilibrer les classes, les gains de performances ne sont pas

très importants entre les modèles de référence (modèles 1 et 4) et les modèles ajustés (modèles 2 et 3).

**Tableau III-4** : Critères de performance obtenus en calibration et validation pour les différents modèles de RF explorés. Les critères de performances en calibration du modèle 3 ne sont pas comparables aux modèles 1 et 2 parce que ce n'est pas le même échantillon.

Modèle 1 (RF_Class)		Modèle 2 (RF_SI)		Modèle 3 (RF_SMOTE)		Modèle 4 (RF_Regr)	
Calibration (901 stations)				Calibration (901 stations + 1494 données interpolées)		Calibration (901 stations)	
n <sub>tree</sub>	800	n <sub>tree</sub>	800	n <sub>tree</sub>	800	n <sub>tree</sub>	800
m <sub>try</sub>	3	m <sub>try</sub>	2	m <sub>try</sub>	5	m <sub>try</sub>	6
Accuracy	0,76	Accuracy	0,75	Accuracy	0,90	RMSE (mois)	2
Kappa	0,42	Kappa	0,36	Kappa	0,87	R <sup>2</sup>	0,50
Validation (224 stations)							
Accuracy	0,75	Accuracy	0,66	Accuracy	0,69	Accuracy	0,71
						R <sup>2</sup>	0,45
Kappa	0,4	Kappa	0,35	Kappa	0,37	RMSE (mois)	2,15
						Kappa	0,37
Kappa pondéré (K <sub>w</sub> )	0,39	Kappa pondéré (K <sub>w</sub> )	0,41	Kappa pondéré (K <sub>w</sub> )	0,42	Kappa pondéré (K <sub>w</sub> )	0,42

Dans l'ensemble, en validation, tous les modèles prédisent très bien la classe permanente avec des valeurs de *precision*, *recall* et score F1 supérieures à 0,75 dans tous les cas (**Figure III-12**). Le score F1 est une moyenne harmonique entre le *recall* et la *precision* et tient simultanément compte de la proportion de la classe correctement prédite parmi les valeurs observées ainsi que dans les valeurs prédites comme appartenant à la classe, la proportion correctement prédite. Pour les classes faiblement intermittentes, fortement intermittentes et éphémères, les meilleures valeurs de score F1 sont obtenues respectivement avec le modèle 2 (0,37) (**Figure III-12.b**), modèle 4 (0,36) (**Figure III-12.d**), modèle 1 (0,37) (**Figure III-12.a**), ce qui correspond à des scores passables. Les résultats dans la suite de ce chapitre sont présentés uniquement pour les modèles 3 et 4 qui affichent les meilleurs scores du Kappa pondéré (0,42) (**Tableau III-4**).

**Validation (224 stations)**

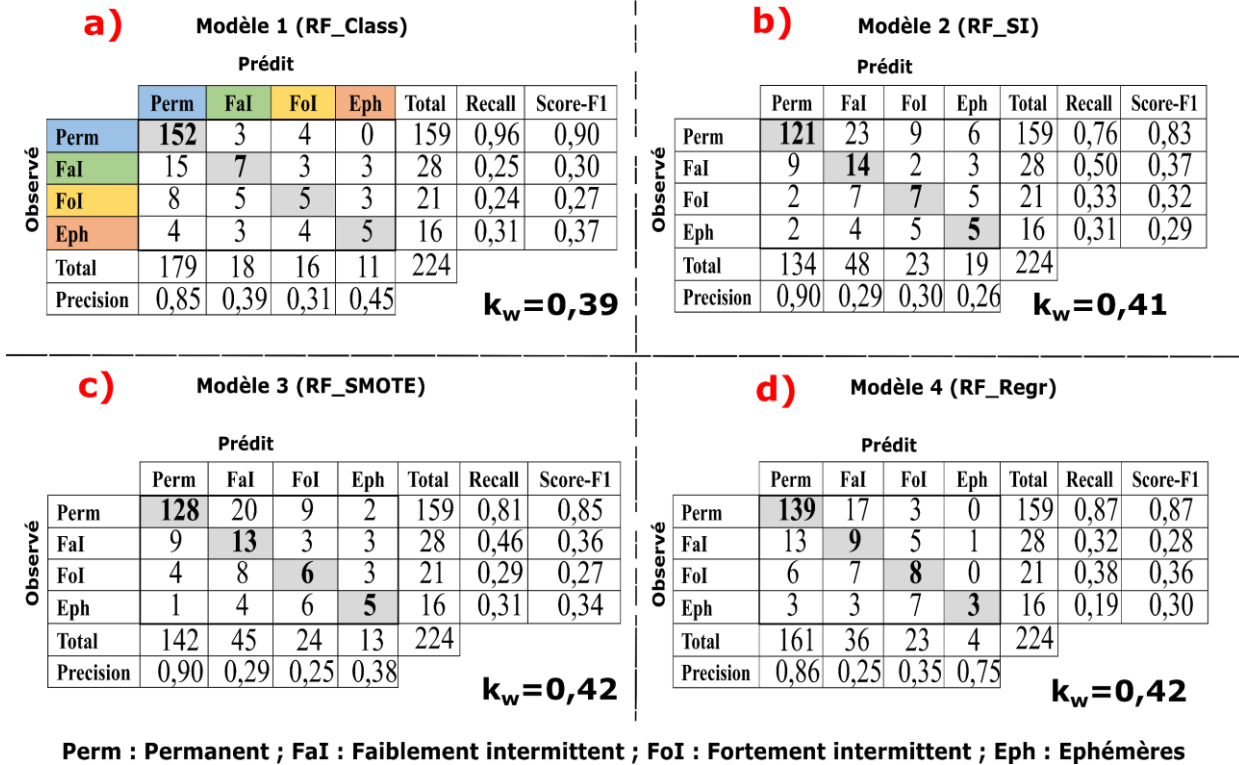


Figure III-12 : Matrice de confusion en validation pour les différents modèles testés. Notons que les cellules de la diagonale indiquent les cas où les classes sont correctement prédites.

Les variables hydro-environnementales les plus importantes dans les modèles 3 et 4 par ordre d'importance sont représentées en Figure III-13. Pour le modèle 3 (Figure III-13.a) à l'échelle des bassins versants, les cinq variables les plus importantes par ordre sont : (1) l'indice d'aridité (P/Etp), (2) la surface amont des bassins versants (Surf), (3) l'évapotranspiration potentielle (Etp), (4) le pourcentage de sable (%) et (5) le nombre moyen de mois pluvieux par an (Nwet). Ces variables sont à majorité dominées par le climat (3/5) avec une valeur de réduction de l'impureté des nœuds « Mean decrease Gini » très importante pour l'indice d'aridité par rapport aux autres variables ce qui se traduit concrètement dans le modèle par un meilleur découpage des données permettant de caractériser les classes d'intermittence à l'aide de cette variable. L'influence anthropique à travers la capacité de stockage amont des barrages (Damstor, Figure A1-3.e) ressort comme la variable la moins importante dans le modèle 3. À travers la variable Damstor, nous avons cherché à évaluer si les stations qui ont à l'amont un certain nombre de barrages et dont la capacité de stockage total est importante sont plus susceptibles d'être permanent ; étant donné que les barrages peuvent contribuer à la maintenance de débits pendant la saison sèche. Sur les 901 stations utilisées pour entraîner le modèle, seulement 24% (soit 216 stations) présentent une valeur de capacité de stockage total amont supérieur à 0. De ces 24% de stations (Figure A3-5), on note



qu'en général les stations qui ont une capacité de stockage total des barrages à l'amont important sont plus susceptible d'être permanents. Par contre 76% des stations de calibration analysées n'ont pas de barrages à l'amont (Damstor=0) ce qui peut expliquer que globalement la variable Damstor ne ressorte pas comme l'une des variables les plus importantes pour expliquer l'intermittence dans le modèle3.

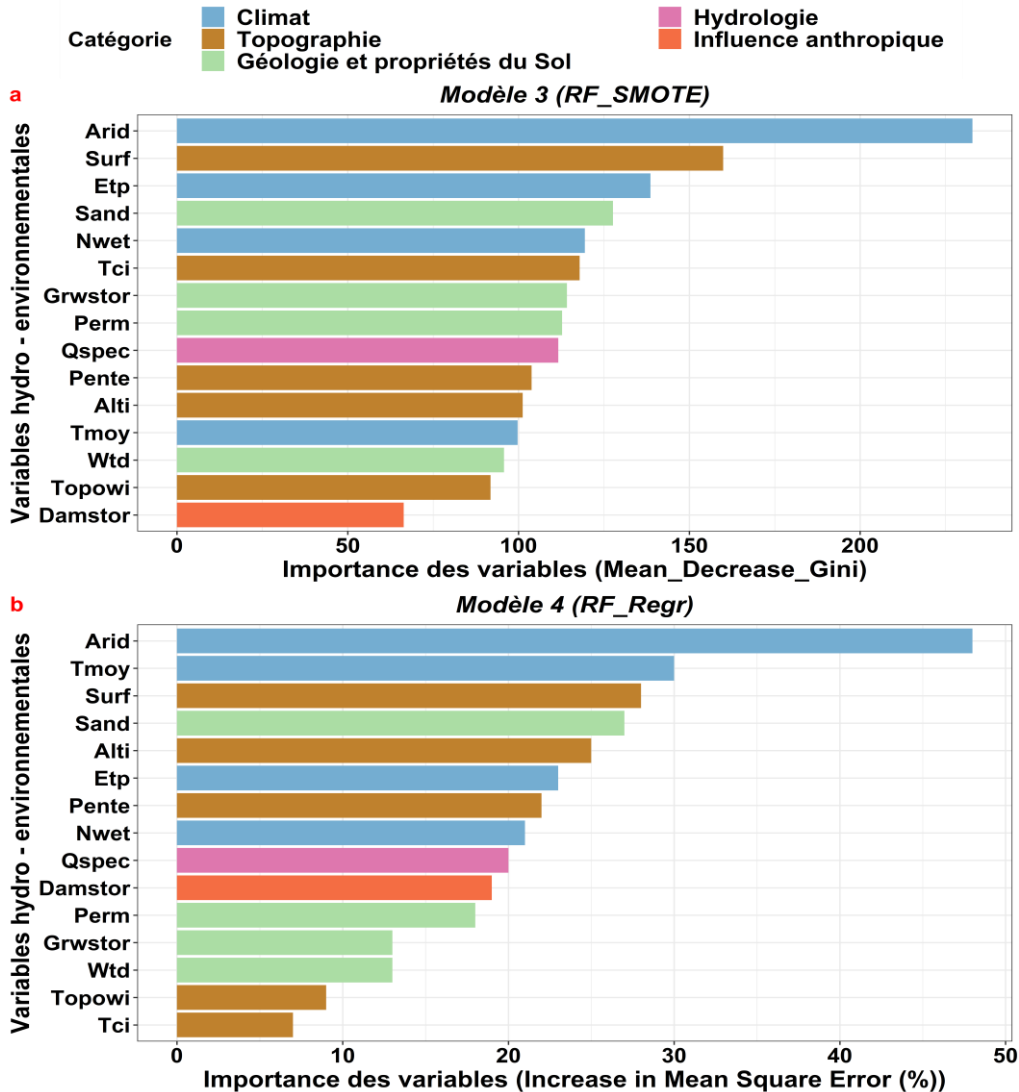


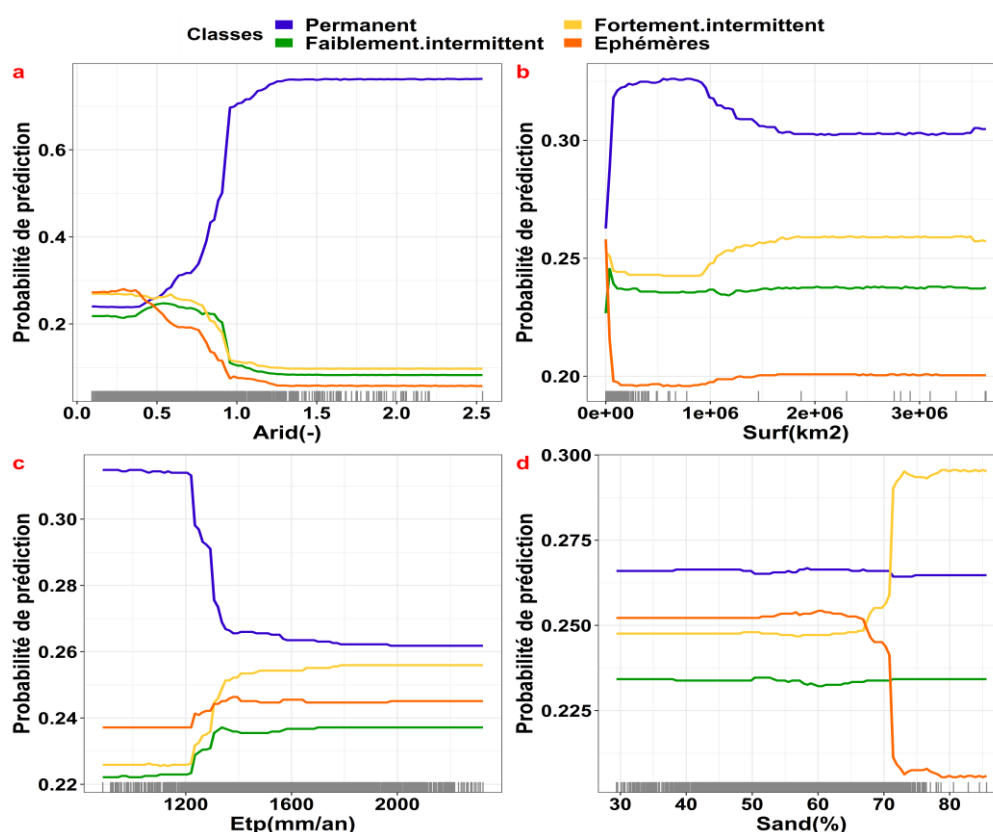
Figure III-13 : Ordre d'importance des variables dans le modèle a) modèle 3, l'importance est appréciée par la décroissance moyenne de l'indice de Gini dans le modèle, b) modèle 4, l'importance est appréciée par l'augmentation de l'erreur quadratique moyenne (MSE) si la variable est supprimée du modèle. Les noms complets des variables sont définis dans le Tableau III-2.

Pour le modèle 4 (Figure III-13.b), à l'échelle des bassins versants, les cinq variables les plus importantes par ordre sont : (1) l'indice d'aridité (P/Etp), (2) la température moyenne annuelle de l'air (Tmoy), (3) la surface amont des bassins versants (Surf), (4) le pourcentage de sable (Sand) et (5) l'altitude moyenne (Alti). Ces variables se répartissent majoritairement entre la topographie et le climat. La Figure III-13.b peut être lue comme suit : si l'indice d'aridité est

supprimé du modèle cela résulte en une augmentation de l'erreur quadratique moyenne de (%IncMSE) de 43 %. La variable la moins importante est l'indice (Tci).

Les variables les plus importantes communes aux deux modèles sont : l'indice d'aridité, la surface amont des bassins versants et le pourcentage de sable tandis que la température moyenne qui est importante dans le modèle 4 ne l'est pas dans le modèle 3. Cela pourrait s'expliquer par le fait que Tmoy et Etp sont liées d'une certaine manière (même si cela n'apparaît pas vraiment dans la matrice de corrélation). Les variables hydrogéologiques (Perm, Grwstor, Wtd) sont moins importantes dans le modèle 4 que dans le modèle 3.

Pour comprendre l'effet individuel ou la sensibilité des variables les plus importantes dans chaque modèle par rapport à l'intermittence (classes d'intermittence associées pour le modèle 3 et nombre de mois secs pour le modèle 4), une analyse des graphes de dépendance partielle est effectuée pour les quatre variables les plus importantes dans le modèle 3 et modèle 4 respectivement représentés en **Figure III-14** et **Figure III-15**.



**Figure III-14** : Graphiques de dépendance partielle pour les quatre variables les plus influentes du **modèle 3** en fonction des différentes classes d'intermittence, les graphiques sont classés par ordre d'importance des variables dans le modèle a) Arid, b) Surf, c) Etp, d) Sand, voir la description complète du code des variables au **Tableau III-2**. Le "tapis" en bas (couleur grise) montre la distribution des valeurs prises par les stations (901) pour chaque variable représentée.

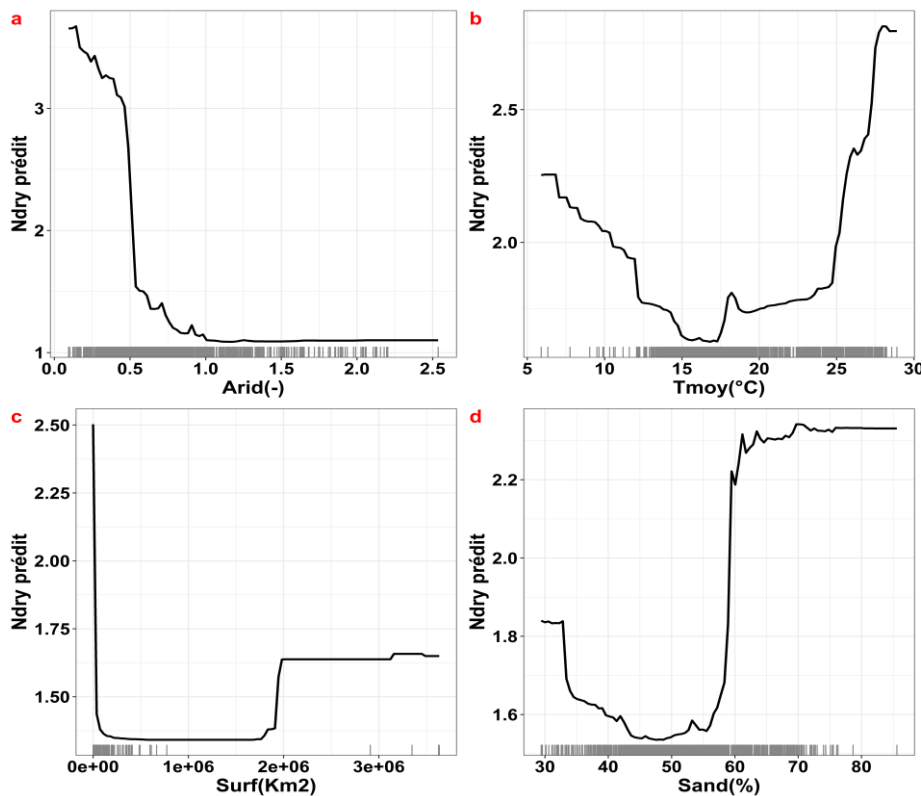
Dans ces figures (**Figure III-14** , **Figure III-15**), le "tapis" en bas (couleur grise) montre la distribution des valeurs prises par les stations (901) pour chaque variable représentée. De façon générale, les figures montrent une relation non linéaire entre les variables environnementales et la variable de réponse.

La **Figure III-14.a** nous montre que plus l'indice d'aridité (P/Etp) augmente, plus la probabilité de prédiction de la classe permanente est grande ce qui entraîne en conséquence une probabilité très faible de prédiction des autres classes d'intermittence. Cette réponse est raisonnable puisqu'un indice d'aridité inférieur à 1 indique que la précipitation moyenne annuelle est inférieure à l'évapotranspiration potentielle annuelle, ce qui favorise les classes les plus intermittentes. Une large part des stations de calibration (plus de 80%) ont des surfaces de bassins < 500 000 km<sup>2</sup> et dans cette plage, la probabilité de prédiction de la classe permanente croît monotonement en fonction de la surface du bassin, ce qui est l'inverse pour les autres classes (**Figure III-14.b**). Cet effet se poursuit jusqu'au seuil de 1 000 000 km<sup>2</sup> à partir duquel on a une légère décroissance de la probabilité de prédiction de la classe permanente due au fait qu'on a qu'une dizaine de stations entre 1 000 000 et 3 634 864 km<sup>2</sup> comme le confirme le "tapis" au bas de la **Figure III-14.b**. L'évolution croissante de l'évapotranspiration potentielle annuelle s'accompagne d'une baisse de la probabilité de prédiction de la classe permanente et inversement pour les autres classes (**Figure III-14.c**).

Finalement, les probabilités de prédiction des classes d'intermittence faible (permanent et faiblement intermittent) sont relativement indépendantes du pourcentage moyen en sable des bassins versants (**Figure III-14.d**). Au contraire, pour les classes les plus intermittentes (fortement intermittent et éphémères), les fortes valeurs du pourcentage en sable (à partir d'un seuil de 68%) augmentent les chances de prédictions de la classe fortement intermittente et inversement pour la classe éphémère. Cela s'expliquerait par le fait que dans l'échantillon de calibration (901) la majorité des stations qui ont un pourcentage moyen en sable supérieur à 68 % sont classées fortement intermittentes.

Les graphes de dépendance partielles représentent l'effet marginal de chaque variable explicative par rapport à la variable de réponse ( $\overline{Ndry}$ ), c'est la raison pour laquelle les valeurs de  $\overline{Ndry}$  dans la **Figure III-15** en régression n'excède pas plus de 4 mois secs car c'est l'effet combiné de ces graphes qui permet une prédiction finale de la valeur de  $\overline{Ndry}$ . Au niveau du modèle 4, l'augmentation de l'indice d'aridité (P/Etp >1) se traduit par une baisse de la prédiction du nombre moyen de mois secs  $\overline{Ndry}$  (**Figure III-15.a**). L'augmentation de

la température moyenne de l'air (**Figure III-15.b**), entre 5 et 18°C se traduit par une baisse du  $\overline{Ndry}$  prédit (il y a peu de stations dans cette plage comme le montre le 'tapis' au bas de la figure) puis une augmentation pour toutes valeurs supérieures à 18°C. Sur la plage entre 0-500 000 km<sup>2</sup> où se concentre la majorité des bassins, une augmentation de la surface se traduit par une baisse du  $\overline{Ndry}$  prédit (**Figure III-15.c**), après ce seuil (500 000 km<sup>2</sup>) la valeur du  $\overline{Ndry}$  prédit reste relativement faible et insensible à l'augmentation de la surface ce qui se justifie par le fait qu'on n'a plus trop de stations après ce seuil comme le confirme le 'tapis' au bas de la figure. La courbe d'évolution du pourcentage moyen en sable évolue comme celle de la température, pour des valeurs en pourcentage de sable < 50% on a une légère décroissance des valeurs prédites du  $\overline{Ndry}$ , et pour des valeurs > au seuil de 50%, on a une croissance des valeurs prédites du  $\overline{Ndry}$  (**Figure III-15.d**).

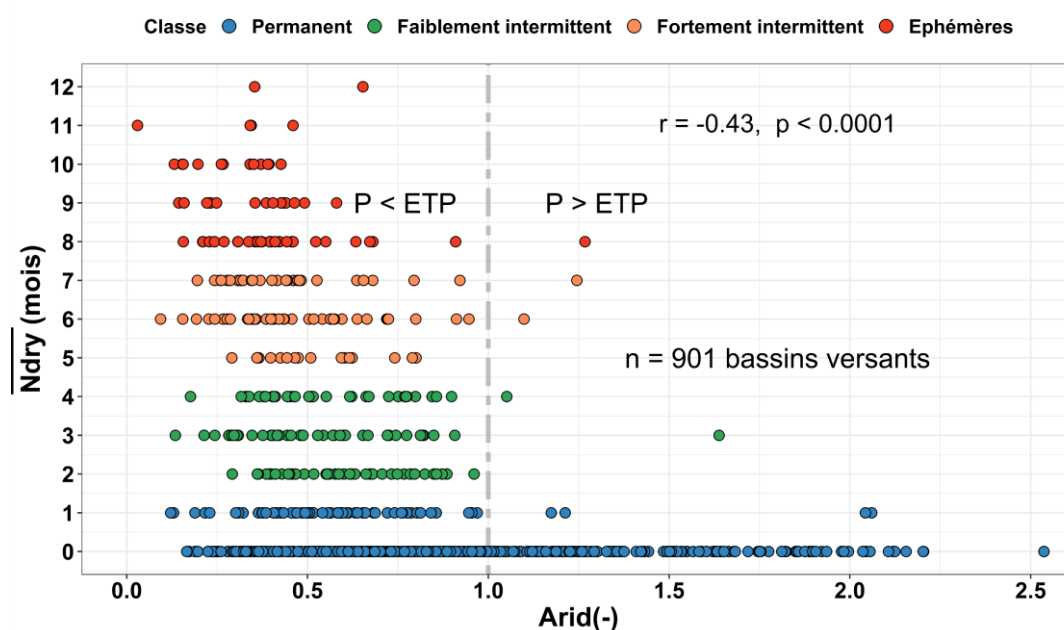


**Figure III-15 :** Graphiques de dépendance partielle pour les quatre variables les plus influentes du modèle 4, les graphiques sont classés par ordre d'importance des variables dans le **modèle 4 a) Arid, b) Surf, c) Etp, d) Sand**, voir la description complète du code des variables au **Tableau III-2**. Le "tapis" en bas (couleur grise) montre la distribution des valeurs prises par les stations (901) pour chaque variable représentée.

Des valeurs d'indice d'aridité ( $P/Etp > 1$ ) favorisent la permanence des cours d'eau (**Figure III-14.a, Figure III-15.a**) tandis que des valeurs élevées d'évapotranspiration moyenne annuelle et de température moyenne annuelle favorisent l'intermittence (**Figure III-14.c, Figure III-15.b**). Le débit des rivières tend à devenir plus permanent avec l'augmentation de la surface amont drainée ce qui se reflète dans les prédictions des modèles (**Figure III-14.b,**

**Figure III-15.c).** Enfin, des pourcentages de sable très élevés dans un bassin favoriserait une forte infiltration et une probable réduction du ruissellement, ce qui peut favoriser l'intermittence des cours d'eau dans des zones désertiques, mais inversement, une percolation vers une nappe phréatique peu profonde pourrait favoriser une permanence des cours d'eau, bien que ces types de relations ne soient pas clairement visibles en **Figure III-14.c**.

Les graphes de dépendance partielle (**Figure III-14**, **Figure III-15**) sont issus respectivement de l'interprétation de l'apprentissage des modèles 3 et 4 qui montrent des relations non linéaires en général ; l'analyse suivante (**Figure III-16**) examine sur les 901 stations de calibration, la relation linéaire entre l'indice d'aridité et la moyenne du nombre de mois secs ( $\overline{Ndry}$ ) pour les différentes classes d'intermittence. L'analyse entre l'indice d'aridité et  $\overline{Ndry}$  montre une assez bonne corrélation (**Figure III-16**). On remarque une structuration des classes d'intermittence à partir d'un seuil d'indice d'aridité égal à 1. En effet, plus de 90% des stations intermittentes se situent sous un seuil d'indice d'aridité inférieur à 1, ce qui implique qu'à ces endroits, la précipitation moyenne annuelle est inférieure à l'évapotranspiration potentielle moyenne annuelle, ce qui créerait un déficit hydrique favorisant naturellement l'intermittence des cours d'eau (**Figure III-16**). Par contre, on retrouve des cours d'eau permanents sur toutes les plages d'indice d'aridité ce qui est sans doute lié à l'effet combiné des autres variables (**Figure III-13**), mais dont la relation ne semble pas linéaire comme montré par les graphes de dépendances partielles en **Figure III-14** et **Figure III-15**.



**Figure III-16 :** Relation linéaire simple entre la moyenne du nombre de mois secs et l'indice d'aridité (Arid) pour les 901 stations de calibration.

### III - 4 - 3 Distribution spatiale de l'intermittence prédite à l'échelle de l'Afrique

Quinze variables environnementales ont été retenues comme pertinentes et prises en compte dans le modèle RF pour déterminer les différentes classes d'intermittence des 15 500 000 tronçons de rivières. Les cours d'eau en tête des bassins versants constituent en général la majorité des cours d'eau au sein du réseau hydrographique (Nadeau et Rains, 2007). Plusieurs études (Acuña *et al.*, 2014 ; Datry *et al.*, 2014) ont montré que de nombreux bras de cours d'eau en tête de bassin versant sont de petite taille et représenteraient une part très importante du réseau hydrographique (Brooks et Colburn, 2011 ; Fritz *et al.*, 2013). Cela s'applique également au réseau LCS où 78 % des cours d'eau drainent moins de 20 km<sup>2</sup> de surface. La surface amont que drainent l'ensemble des cours d'eau LCS varie entre 0,34 km<sup>2</sup> et 3 711 080 km<sup>2</sup>. L'échantillon de calibration (901 stations) couvre une gamme similaire de surfaces amont, qui varient entre 0,93 et 3 634 865 km<sup>2</sup>, mais avec une distribution très différente de celle que drainent les cours d'eau LCS (**Figure III-17.a**). En effet, seulement 4 % des stations de calibration drainent 20 km<sup>2</sup> de surface. Cela signifie que la prédiction (extrapolation) faite sur les petits cours n'est pas représentative des données de calibration et donc très incertaine compte tenu du fait qu'on dispose de très peu ou pas de stations sur ces cours d'eau. Ainsi, dans la présentation des résultats de prédiction, trois cartes de l'intermittence sont proposées pour chacun des modèles 3 et 4. Il s'agit des cartes de :

1. la prédiction sur tous les cours d'eau à l'échelle de l'Afrique, mais avec des incertitudes sur les petits cours d'eau qui constituent la majorité du réseau hydrographique ;
2. la prédiction sur tous les cours d'eau drainant au moins 100 km<sup>2</sup> (soit plus de 80% de la surface que drainent les stations de calibration) ;
3. la prédiction sur tous les cours d'eau drainant au moins 6000 km<sup>2</sup> où seuls les grands cours d'eau du continent sont mis en évidence.

La comparaison de la distribution des autres variables au niveau des bassins versants amont des tronçons de rivière (n = 15 500 000) ainsi que des stations de calibration (n = 901) est présentée en **Figure III-17** et **Figure III-18**. Il ressort de ces deux figures que les valeurs de la pente (**Figure III-17.b**), du nombre de mois moyens de pluie (Nwet) (**Figure III-18.a**) au niveau des stations de calibration sont légèrement supérieures à la plupart des valeurs prises en compte au niveau des bassins versants amont associés aux tronçons de rivière. Pour les valeurs d'altitudes (Alti) (**Figure III-17.c**), de l'indice topo (Topowi) (**Figure III-17.d**), de

l'indice (Tci) (Figure III-17.e), de l'indice d'aridité (Arid) (Figure III-18.b), du stockage des aquifères (Grwstor)(Figure III-18.e), de la profondeur à la nappe (Wtd)(Figure III-18.f) et de la perméabilité (Perm) (Figure III-18.g), les valeurs au niveau des stations de calibration sont en moyenne les mêmes qu'au niveau des bassins versants amont associés aux tronçons de rivières.

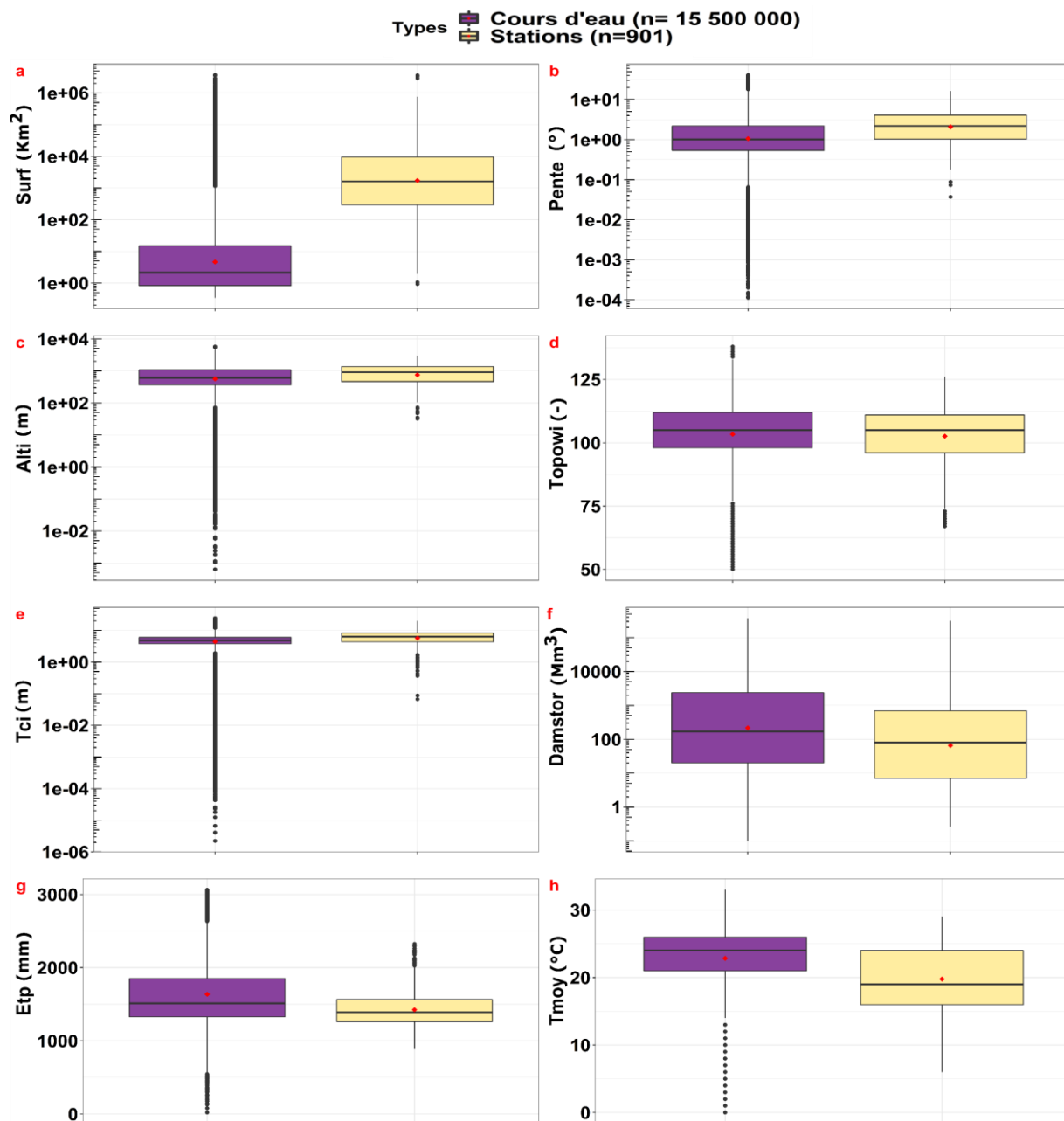
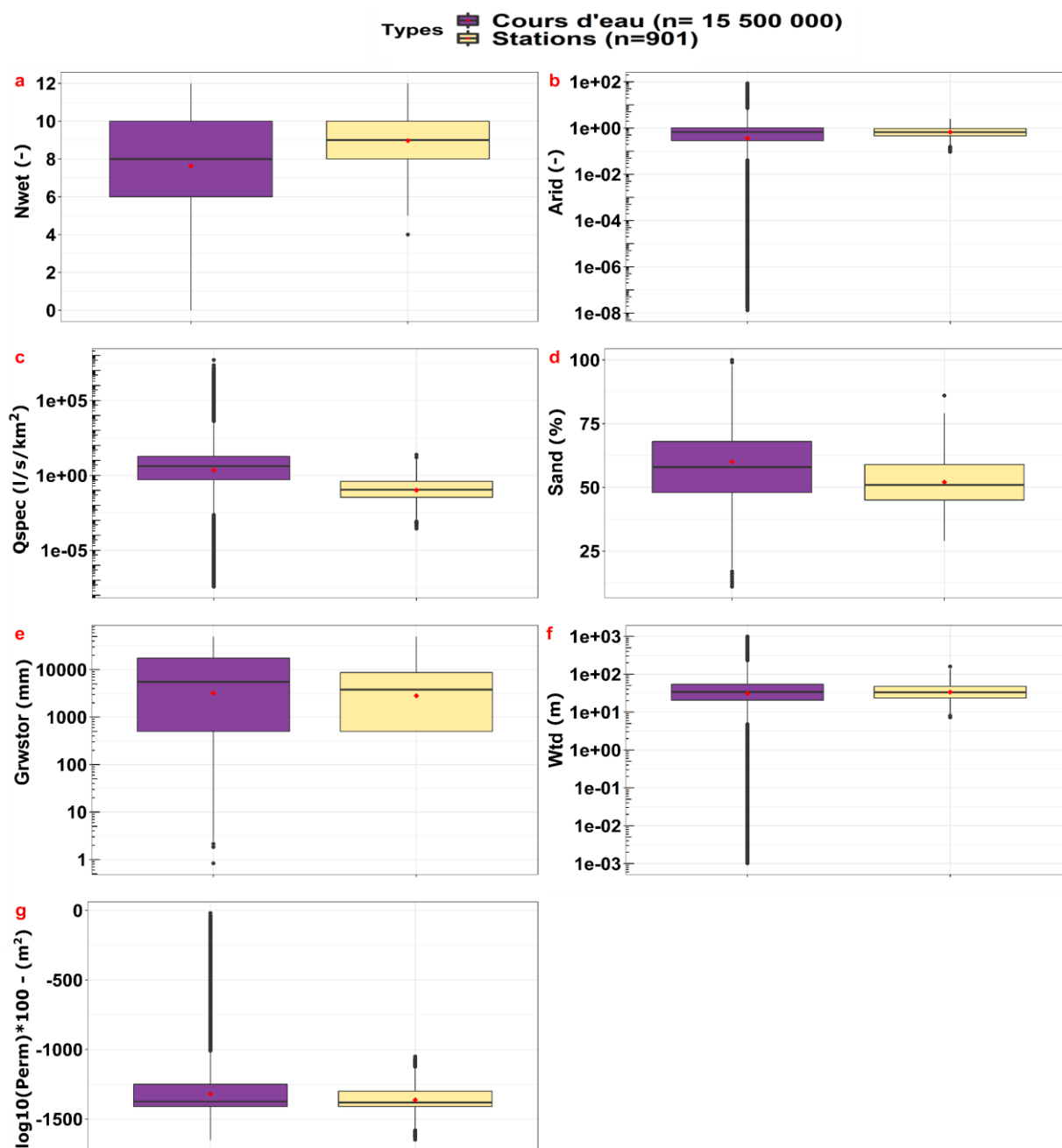


Figure III-17 : Distribution des variables hydro-environnementales pour 15 500 000 bassins versants amont associés aux tronçons de rivières (couleur violette) et 901 bassins versants amont associés aux stations de jaugeage utilisées en calibration des modèles. a) Surf, b) Pente, c) Alti, d) Topowi, e) Tci, f) Damstor, g) Etp, h) Tmoy. Les noms complets des variables sont définis dans le Tableau III-2.



**Figure III-18** : Distribution des variables hydro-environnementales pour 15 500 000 bassins versants amont associés aux tronçons de rivières (couleur violette) et 901 bassins versants amont associés aux stations de jaugeage utilisés en calibration des modèles. **a)** Nwet, **b)** Arid, **c)** Qspec, **d)** Sand, **e)** Grwstor, **f)** Wtd, **g)** Perm. Les noms complets des variables sont définis dans le **Tableau III-2**.

Enfin, pour les variables telles que la capacité de stockage amont des barrages (Damstor) (**Figure III-17.f**), l'évapotranspiration potentielle (Etp) (**Figure III-17.g**), la température moyenne (Tmoy) (**Figure III-17.h**), les débits spécifiques (Qspec) (**Figure III-18.c**) et le pourcentage de sable (Sand)(**Figure III-18.d**), les valeurs au niveau des stations de calibration sont en moyenne inférieures aux valeurs prises en compte au niveau des bassins versants amont associés aux tronçons de rivières. L'absence d'un nombre important de



stations de jaugeage sur des bassins versants de petite taille est très certainement une cause majeure de ces quelques différences dans la distribution des valeurs des variables au droit des stations et des bassins versants amont des cours d'eau.

Les résultats de prédiction de l'intermittence en Afrique par les modèles 3 et 4 sont respectivement présentés en (**Figure III-19, Figure III-20**) et (**Figure III-21, Figure III-22**). Les **Figure III-19.a** et **Figure III-21.a** montrent les prédictions pour tous les cours d'eau du réseau LCS. Dans ces figures, des vides « en blanc » apparaissent aux emplacements des lacs notamment dans la région des Grands Lacs en Afrique de l'Est. Dans les régions désertiques (Sahara, Kalahari), les vides « en blanc » représentent des zones où il n'y a pas de cours d'eau. La forte densité d'une classe d'intermittence donnée associée à une faible présence d'une autre classe aux mêmes endroits fait que certaines classes sont un peu masquées par endroits. Les **Figure 19.b-e** et **Figure 21.b-e** séparent les différents cours d'eau et stations en fonction de leur classe d'intermittence. Enfin, les **Figure 20.a,b** et **Figure 22.a,b** se focalisent sur les cours d'eau drainant au moins 100 km<sup>2</sup> et 6000 km<sup>2</sup>, pour les modèles 3 et 4 respectivement.

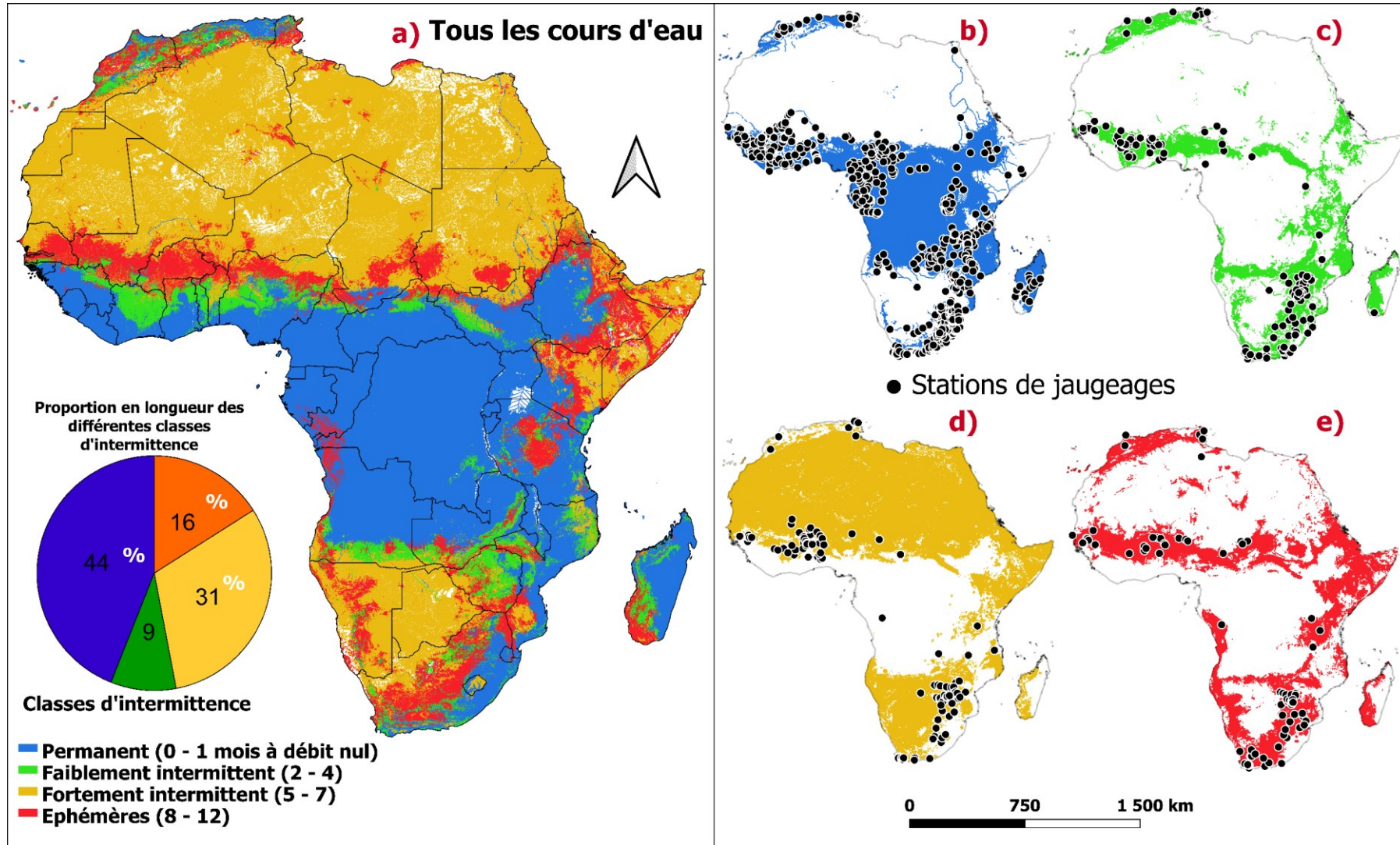


Figure III-19 : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le modèle 3. a) tous les cours d'eau du réseau LCS, b) les cours d'eau prédits permanents, c) les cours d'eau prédits faiblement intermittents, d) les cours d'eau prédits éphémères. Les stations de jaugeages représentées en noir sur les cartes b-e sont les stations classées dans la classe d'intermittence de chaque carte.

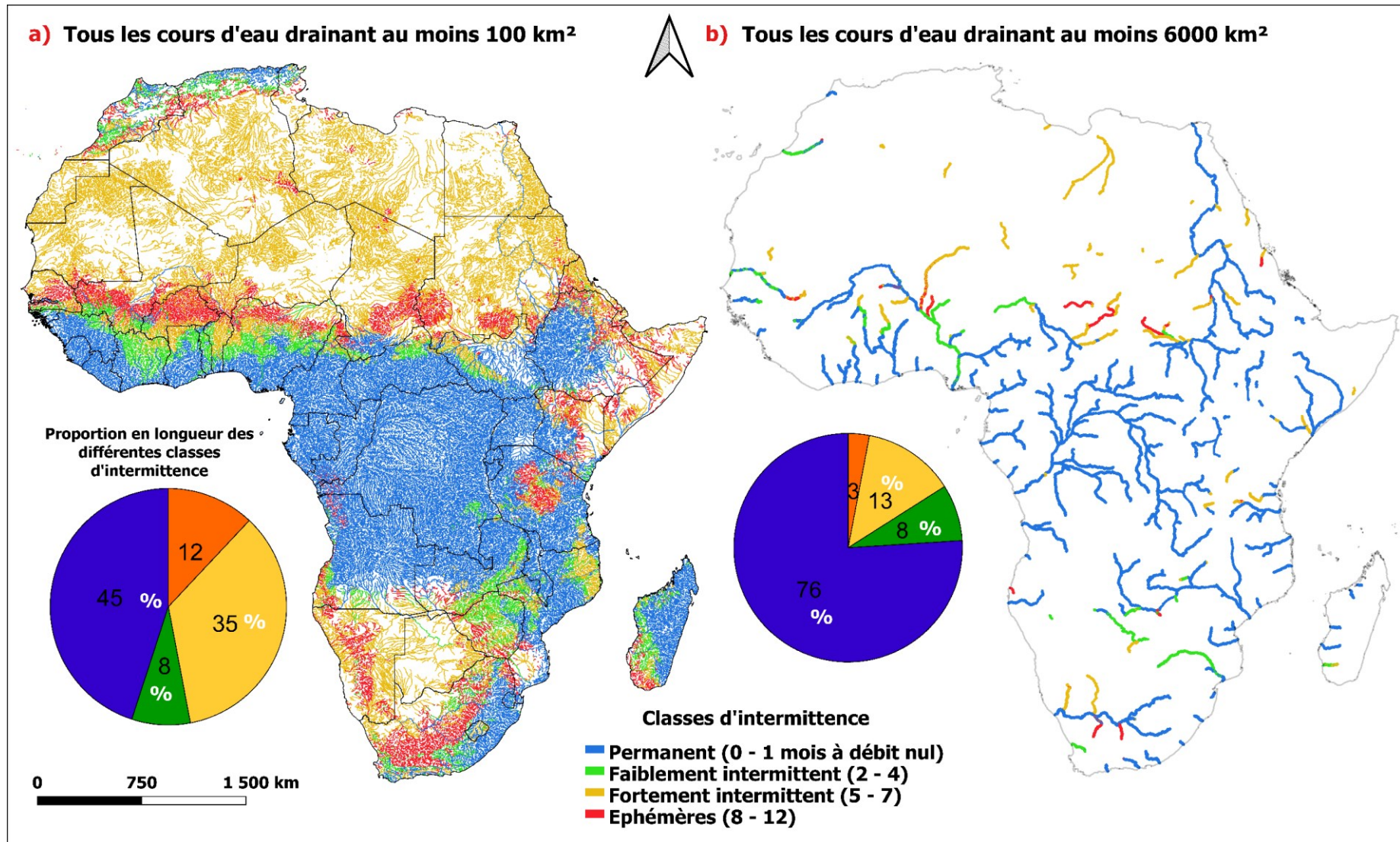


Figure III-20 : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le modèle 3. a) tous les cours d'eau drainant au moins 100 km<sup>2</sup>, b) tous les cours d'eau drainant au moins 6000 km<sup>2</sup>.

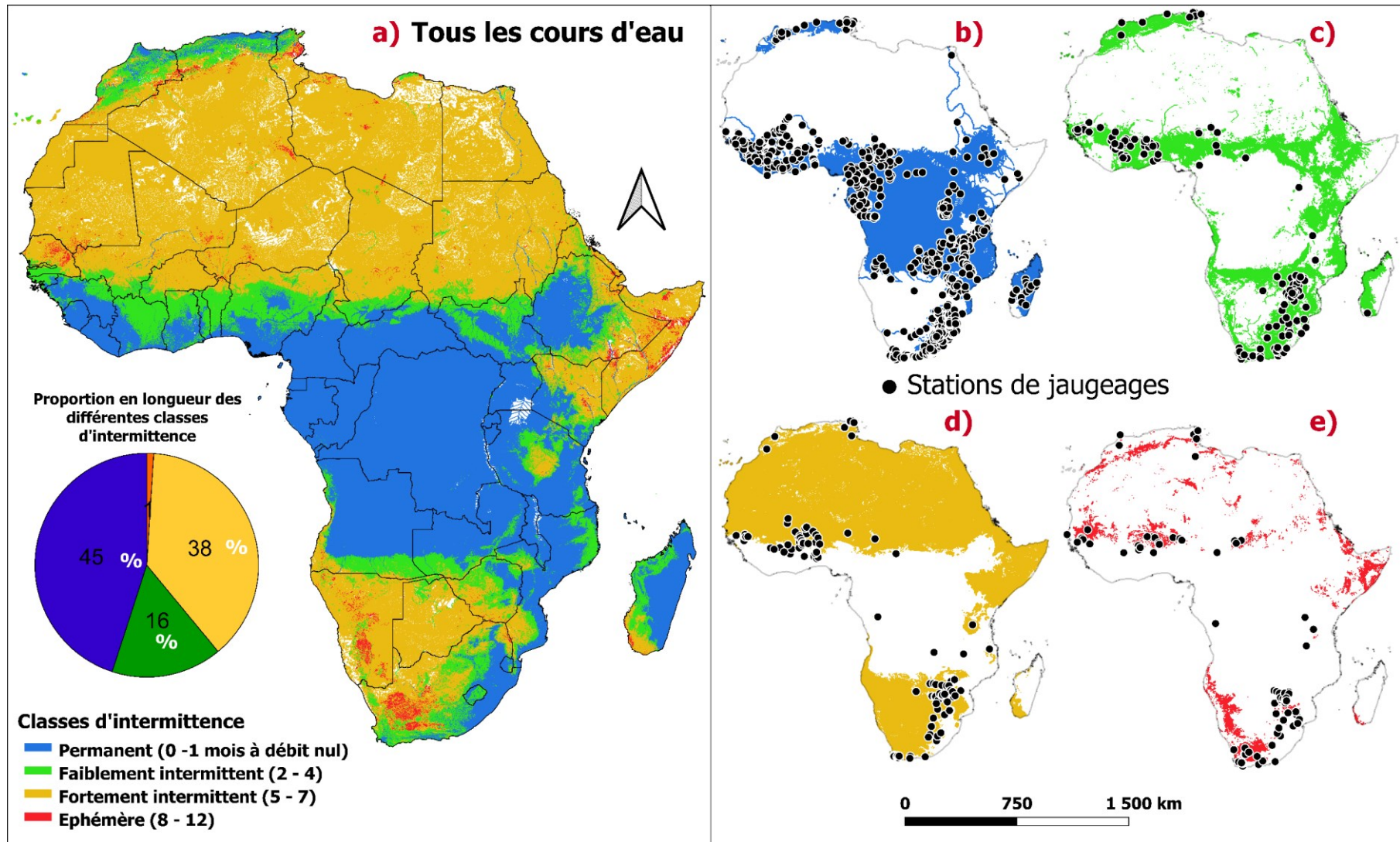


Figure III-21 : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le modèle 4. a) tous les cours d'eau du réseau LCS, b) les cours d'eau prédits permanents, c) les cours d'eau prédits faiblement intermittents, d) les cours d'eau prédits éphémères. Les stations de jaugeages représentées en noir sur les cartes b-e sont les stations classées dans la classe d'intermittence de chaque carte.

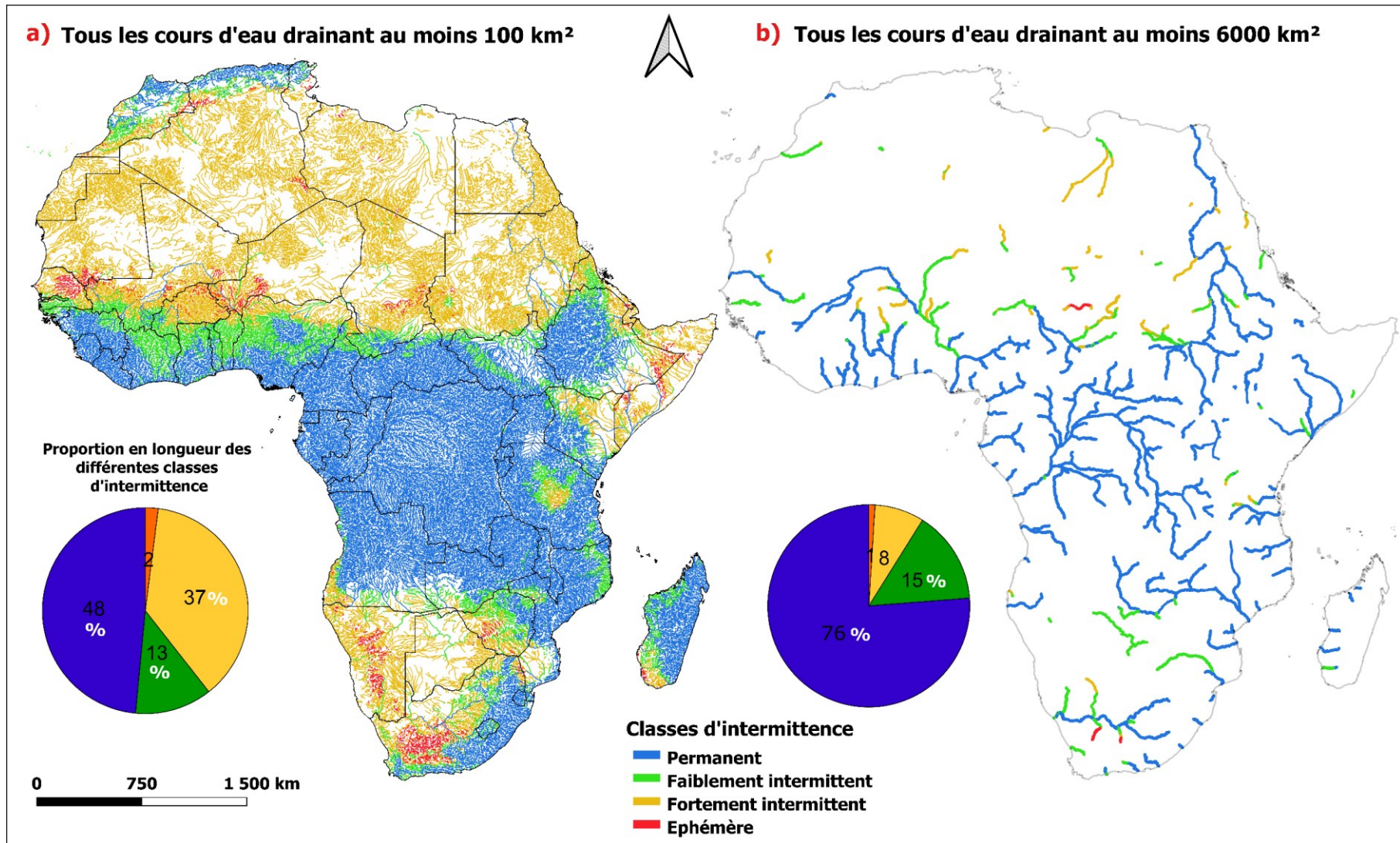


Figure III-22 : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le modèle 4. a) tous les cours d'eau drainant au moins 100 km<sup>2</sup>, b) tous les cours d'eau drainant au moins 6000 km<sup>2</sup>. Les diagrammes en camembert détaillent les pourcentages en longueur de chaque classe d'intermittence.

La prédiction de l'intermittence en Afrique pour tous les cours d'eau (**Figure III-19.a**, **Figure III-21.a**) à l'aide des modèles 3 et 4 donne respectivement les proportions en longueur suivante : 44% pour le modèle 3 (45% pour le modèle 4) des cours d'eau sont prédits permanents, et 56% (55%) sont intermittents tandis que dans les stations de jaugeages les cours d'eau permanents représentent 70,7 % des stations et le reste des stations intermittentes 29,3%. Cela montre que les stations de jaugeages sont majoritairement placées sur des cours d'eau permanents et n'échantillonnent pas une grande partie des cours d'eau qui sont intermittents (Krabbenhof *et al.*, 2022 ; Zimmer *et al.*, 2020). Les cours d'eau prédits sont donc à majorité intermittents et se répartissent en fonction des proportions en longueur des différentes classes d'intermittence pour le modèle 3 et 4 respectivement : 9% (16%) des cours d'eau qui sont considérés faiblement intermittents, 31% (38%) sont fortement intermittents et 16% (1%) sont prédits éphémères. Ces proportions par classe sont similaires à celles qu'on trouve sur les cours d'eau drainant plus de 100 km<sup>2</sup> (**Figure III-20.a** , **Figure III-22.a**). Les cours d'eau drainant plus de 6000 km<sup>2</sup> (**Figure III-20.b** , **Figure III-22.b**) sont à majorité permanents (plus de 70%) ce qui est lié au fait que plus les surfaces amont drainées sont très grandes plus les cours d'eau tendent à être permanents (Snelder *et al.*, 2013).

Les points communs entre les **Figure III-20.a** et **Figure III-22.a** sont que dans les régions d'Afrique où l'évapotranspiration potentielle annuelle est largement supérieure à la précipitation annuelle ( $P/ETP < 0,2$ ), notamment dans les régions arides (Sahel, les déserts du Sahara, Namib et du Kalahari), les cours d'eau sont majoritairement prédits fortement intermittents contrairement aux autres études (Jacobson et Jacobson, 2013 ; Lin *et al.*, 2021 ; Messenger *et al.*, 2021 ; Schneider *et al.*, 2017) où les cours d'eau sont à majorité éphémères dans les zones désertiques (**Figure I-3**). Cela pourrait être dû au fait que nous ne disposons pas de stations de jaugeage à ces endroits dans nos données (en calibration comme en validation) et donc l'extrapolation des modèles à ces zones moins bien représentées augmente l'incertitude de prédiction. Dans ce sens, l'étude de Krabbenhof *et al.* (2022) souligne également la tendance à la sous-représentation des climats extrêmes dans les stations de jaugeage à l'échelle globale. Ensuite, on peut penser que cette sous-estimation de l'intermittence dans les zones désertiques est liée au pourcentage de sable. Dans les zones désertiques, le pourcentage de sable est très important (**Figure A1-4.b**) et dans notre jeu de données de calibration, les stations qui ont un pourcentage moyen de sable qui est supérieur à 70% sont à majorité des stations classées fortement intermittentes. Cela aurait pu favoriser la prédiction de la classe fortement intermittente par le modèle à ces endroits. La sous-

estimation des cours d'eau éphémères en zone désertique est un problème évident de nos modèles 3 et 4.

La majorité (77%) du nombre des cours d'eau prédits intermittents (faiblement intermittent, fortement intermittent, éphémères) ont une température moyenne annuelle de l'air supérieure à 21°C (**Figure A1-1.c**). Les zones à pourcentage élevé de sable (> 60%) (**Figure A1-3.b**) sont généralement situées dans les zones arides et semi-arides et les cours d'eau à ces endroits sont majoritairement prédits intermittents (87%). Dans ces régions arides, il existe aussi des cours d'eau permanents, par le jeu d'autres facteurs de contrôle principaux dont une grande surface amont. Ainsi, des fleuves avec de nombreuses têtes de bassin situées dans les climats très humides peuvent maintenir des régimes d'écoulement permanents même si la majorité de leur longueur traverse des zones arides. Ils sont souvent qualifiés de fleuves allogènes (Bethemont, 1995). C'est le cas par exemple du fleuve Orange, du fleuve Nil, du fleuve Niger et du fleuve Sénégal qui prennent leur source en zone montagneuse humide et s'écoulent à travers des zones très arides.

En **Figure III-20.b** et **Figure III-22.b** les cinq plus grands fleuves d'Afrique par leur surface, à savoir le Congo, le Nil, le Niger, le Zambèze et l'Orange, sont majoritairement prédits permanents sur leur réseau hydrographique principal (surface amont > 6000 km<sup>2</sup>). La principale exception se trouve sur le Niger, entre Niamey et la confluence avec la rivière Bénoué au Nigéria, où il est classé faiblement intermittent par les deux modèles. Après avoir traversé les zones semi-arides du Mali et du Niger, le fleuve perd une grande partie de son débit, et plus de la moitié du débit du bas Niger vient de la rivière Bénoué qui prend sa source au Tchad et au Cameroun (Mahé *et al.*, 2011).

Dans les zones humides où l'évapotranspiration potentielle annuelle est inférieure à la précipitation annuelle ( $P/ETP > 1$ ) comme dans les zones équatoriales, la majorité des cours d'eau sont prédits permanents par les deux modèles.

Les éléments divergents entre les **Figure III-20.a** et **Figure III-22.a** se manifestent principalement dans les zones de transition climatique : entre les zones semi-arides et tropicales, et entre les zones arides au Sahara et le pourtour méditerranéen en Afrique du Nord. Dans ces zones de transition, le modèle 3 prédit majoritairement des cours d'eau éphémères, fortement intermittents et faiblement intermittents tandis que le modèle 4 prédit des cours d'eau à majorité faiblement intermittents et fortement intermittents. Le modèle 3 favorise donc les classes les plus sévères d'intermittence, alors que le modèle 4 prédit une

proportion très faible (1%) de cours d'eau éphémères. Cette caractéristique du modèle 4 se manifestait déjà dans la matrice de confusion (**Figure III-12**) en validation par un *recall* de la classe éphémère plus faible (0,19) que celui du modèle 3 (0,31).

Aux endroits où l'on a presque une même densité de stations éphémères (45) et fortement intermittentes (39) notamment en Afrique Australe (Afrique du Sud, Lesotho, Botswana, Zimbabwe), le modèle 4 prédit majoritairement des cours d'eau fortement intermittents (**Figure III-21.d,e**). Le modèle 4 reporterait une part importante des cours d'eau éphémères dans la classe voisine à savoir la classe fortement intermittente. Il est bon de rappeler que les données de calibration du modèle 3 ont subi un prétraitement visant à équilibrer les classes d'intermittence tandis que celles du modèle 4 (régression) sont calibrées sur des données continues (nombre moyen de mois secs). Bien que l'effet du déséquilibre dans les données se fasse généralement moins ressentir en régression, les prédictions du modèle 4 laissent voir ce déséquilibre avec les classes éphémères qui sont les moins prédites et qui représentent aussi les valeurs les moins courantes dans le jeu de données de calibration en régression. C'est ce qui expliquerait le fait que ce modèle prédit moins de classe éphémère et les reporterait dans la classe voisine notamment la classe fortement intermittente.

Face aux inconsistances du modèle 4 évoquées ci-dessus, dans la suite, nous avons retenu les prédictions faites par le modèle 3 en guise de comparaison avec les cartographies d'intermittence nationale de certains pays.

### III - 4 - 4 Comparaisons de la distribution spatiale de l'intermittence à l'échelle de l'Afrique

#### III- 4 - 4 - 1 Comparaison avec les cartographies à l'échelle nationale

M3-LCS et M3-LCS-100 km<sup>2</sup> désignent respectivement les prédictions d'intermittence du modèle 3 appliqué au réseau LCS (**Figure III-19.a**) et du modèle 3 appliqué au réseau LCS sur les cours d'eau drainant au minimum 100 km<sup>2</sup> (**Figure III-20.a**). Bien que le modèle M3-LCS-100 km<sup>2</sup> soit plus représentatif des prédictions faites à cause du biais d'échantillonnage des stations (très peu de stations de jaugeage localisées sur les cours d'eau drainant moins de 100 km<sup>2</sup>), la comparaison de ce modèle avec les réseaux hydrographiques de référence ainsi que d'autres études n'est pas cohérente en raison d'un manque trop important de cours d'eau. Les pourcentages des différentes classes d'intermittence étant quasi identiques entre M3-LCS



et M3-LCS-100 km<sup>2</sup>, nous nous sommes basés plutôt sur le modèle M3-LCS (**Figure III-19.a**) pour pouvoir comparer les prédictions du modèle 3. Les cartes d'intermittence de référence collectées auprès des instituts géographiques nationaux de quelques pays en Afrique et présentées au **chapitre I** ont été utilisées. Pour faciliter la comparaison, les classes d'intermittence prédites ont été regroupées en deux classes à savoir une classe permanente et une classe intermittente qui regroupent les classes faiblement intermittentes, fortement intermittentes et éphémères. Le **Tableau III-5** présente les critères de performances statistiques obtenues en comparant les différentes cartes d'intermittence (**Figure III-25**, **Figure III-29**). Dans ce **Tableau III-5**, à l'exception du pourcentage d'intermittence (% interm.) qui est le pourcentage en longueur des cours d'eau classés intermittents sur la longueur totale des cours d'eau dans un territoire ou pays donné, les autres paramètres statistiques sont calculés à partir du pourcentage d'intermittence dans les mailles de 5' (environ 8 km x 8 km). Pour chaque maille, le pourcentage d'intermittence est calculé en faisant la somme de la longueur des cours d'eau intermittents sur la longueur totale des cours d'eau dans la maille. Le pourcentage moyen d'intermittence (% moyen interm.) dans un territoire ou pays donné est calculé en faisant la moyenne des pourcentages d'intermittence dans les mailles.

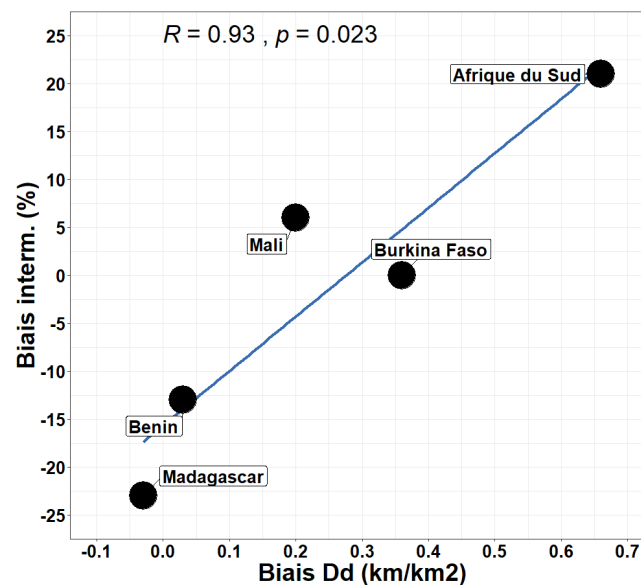
**Tableau III-5** : Paramètres statistiques de comparaison des prédictions d'intermittence du modèle 3 appliqué au réseau LCS (M3-LCS) par rapport aux réseaux de référence du Burkina Faso, Mali, Bénin, Madagascar et Afrique du Sud.

Pays (Échelle)	Réseau hydro	Paramètres statistiques							
		Dd <sub>moy</sub> (km/km <sup>2</sup> )	Biais Dd (km/km <sup>2</sup> )	% interm. (%)	Biais interm. (%)	% moyen interm. (%)	Écart-type (%)	Corr. (-)	RMSE (%)
Bénin (1 : 50 000)	IGN (référence)	1,03	0	94	0	93	16	1	0
	M3-LCS	1,06	0,03	81	-13	80	33	0,41	33
Burkina Faso (1 : 200 000)	IGB (référence)	0,84	0	98	0	98	7,4	1	0
	M3-LCS	1,2	0,36	98	0	98	7	0,1	10
Mali (1 : 200 000)	IGM (référence)	0,36	0	93	0	94	15	1	0
	M3-LCS	0,56	0,2	99	6	99	6	0,4	15
Madagascar (1 : 200 000)	FTM (référence)	1	0	55	0	55	22	1	0
	M3-LCS	0,97	-0,03	32	-23	36	43	0,39	45

Afrique du Sud (1 : 500 000)	DWS (référence)	0,17	0	48	0	54	47	1	0
	M3-LCS	0,83	0,66	69	21	73	38	0,48	47

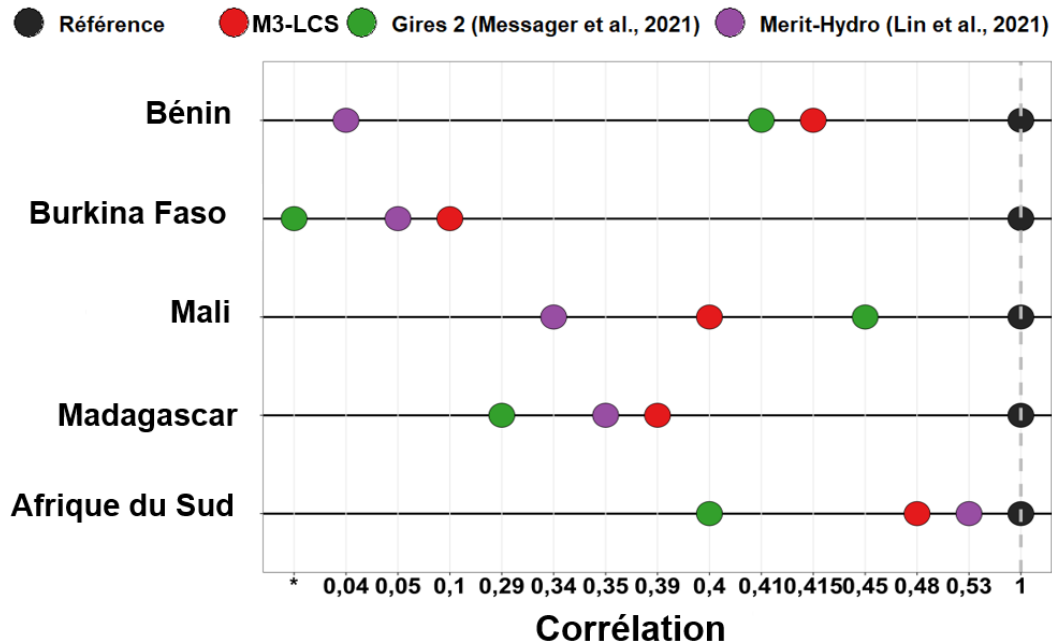
$Dd_{moy}$  : densité de drainage ( $km/km^2$ ) moyenne dans les mailles de 5' (environ 8 km x 8 km) – Biais Dd ( $km/km^2$ ) : Différence entre  $Dd_{moy}$  du réseau M3-LCS avec  $Dd_{moy}$  du réseau de référence - % interm. : pourcentage (%) en longueur des cours d'eau intermittents sur la longueur totale des cours d'eau dans le pays ou territoire considéré – Biais interm. (%) : Différence entre  $Dd_{moy}$  du réseau M3-LCS avec  $Dd_{moy}$  du réseau de référence - % moyen interm. : pourcentage (%) moyen dans les mailles de 5' des cours d'eau intermittents – Écart-type (%) du pourcentage d'intermittence dans les mailles de 5' – Corr. : Coefficient de corrélation de Pearson (-) du pourcentage d'intermittence entre les réseaux de référence et les réseaux prédits dans les mailles de 5'. RMSE (%) : Racine carrée de l'erreur quadratique moyenne du pourcentage d'intermittence entre les réseaux de référence et les réseaux prédits dans les mailles de 5'.

En général, dans tous les pays investigués, le pourcentage d'intermittence en longueur (% interm.) et le pourcentage moyen d'intermittence dans les mailles de 5' (% moyen interm.) sont presque les mêmes. Les densités de drainage moyennes du réseau LCS sont en général supérieures à celles des réseaux de référence collectés à cause de son échelle plus fine (1 : 50 000) à l'exception du Bénin où les échelles des cartes sont les mêmes et donc les densités de drainage moyennes sont similaires. Les pourcentages d'intermittence calculés dans les réseaux prédits sont du même ordre que ceux observés dans les réseaux de référence au Burkina Faso, au Mali, légèrement inférieurs au Bénin, à Madagascar et supérieurs en Afrique du Sud. Dans le **Tableau III-5**, les différences sur le pourcentage d'intermittence (Biais interm.) sont en partie liées aux différences de densité de drainage (Biais Dd) entre les réseaux de référence et le réseau LCS (**Figure III-23**) avec un coefficient de corrélation élevé ( $r = 0,93$ ).



**Figure III-23** : Relation linéaire du biais d'intermittence (%) et du biais de densité de drainage ( $km/km^2$ ) entre le réseau M3-LCS et les réseaux hydrographiques de référence du Bénin, Burkina Faso, Mali, Madagascar, Afrique du Sud.

En général, le modèle M-LCS est le plus corrélé aux réseaux de référence (3/5) dans les pays investigués par rapport aux études continentales précédentes (**Tableau I-5, Tableau III-5, Figure III-24**).



**Figure III-24** : Corrélation (en % d'intermittence par maille) des prédictions M3-LCS et des études précédentes (Lin et al., 2021 ; Messenger et al., 2021) par rapport aux réseaux de références de quelques pays investigués.

Au Burkina Faso, la corrélation des pourcentages d'intermittence entre le modèle M3-LCS et le réseau de référence est faible (0,10). Cela s'explique par le fait que le modèle 3 prédit beaucoup plus de cours d'eau permanents sur les petits cours d'eau à l'ouest du Burkina Faso qu'il n'y en a dans le réseau de référence (**Figure III-25.a-d**), notamment dans la zone sédimentaire qui comprend le Mouhoun supérieur. Cela est sans doute porté par l'important stockage des aquifères dans cette zone et par une perméabilité plus forte que dans le reste du pays. Sur les cours d'eau les plus grands, les principales différences sont sur la Comoé, la Sissili, le Nazinon, le Nakanbé et la Sirba qui sont classés permanents dans le réseau de référence et prédits faiblement intermittents pour la Comoé et fortement intermittents et éphémères pour le reste (**Figure III-25.a-b**). Comme déjà discutée dans l'analyse sur la prédiction spatiale de l'intermittence au Burkina Faso (section II-8 du **chapitre II**), la concordance entre les stations de jaugeage (49) et la classification du réseau de référence BNDT aux emplacements des stations de jaugeage n'est que de 57,1%, les observations de jaugeages ainsi que les prédictions confirment que ces cours d'eau sont faiblement intermittents pour la Comoé et fortement intermittents et éphémères pour le reste. Cela traduit une mauvaise classification du régime d'intermittence de ces cours d'eau dans la base BNDT.

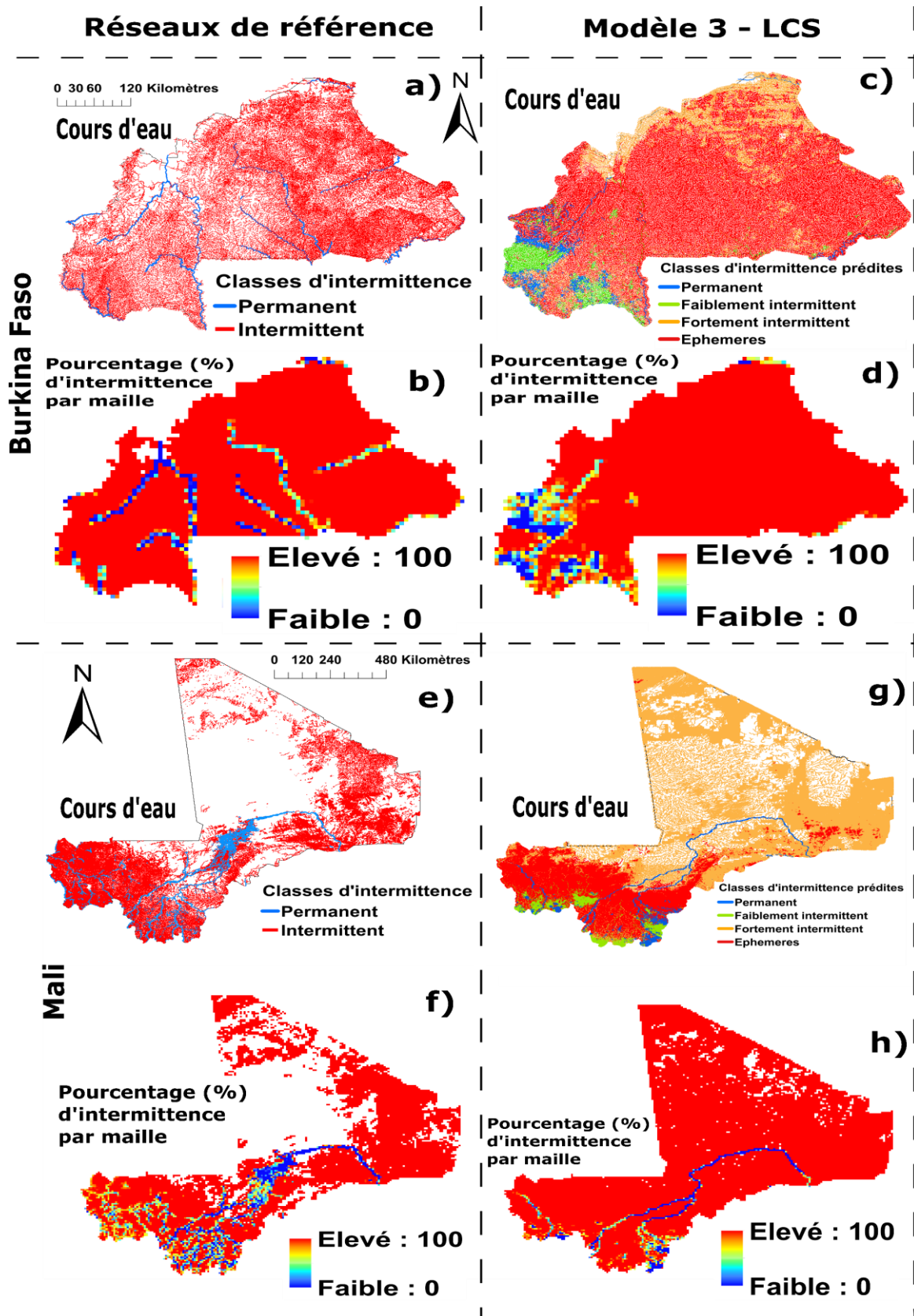


Figure III-25 : Comparaison de la distribution spatiale des cartes nationales de l'intermittence aux prédictions d'intermittence (M3-LCS) au Burkina Faso (a,b,c,d), Mali (e,f,g,h). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe.

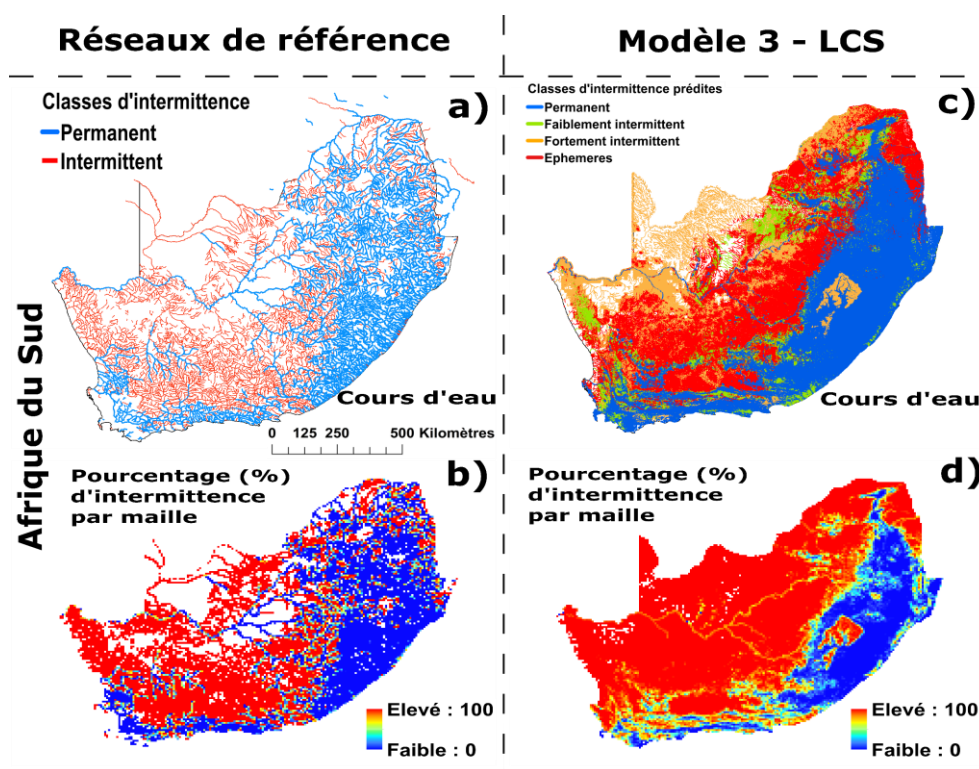
Au Mali, le coefficient de corrélation spatiale des pourcentages d'intermittence entre le réseau de référence (**Figure III-25.e**) et le réseau prédit (**Figure III-25.g**) est plus élevé qu'au Burkina Faso et assez satisfaisant (coefficient de corrélation de l'ordre de 0,4) et la distribution spatiale de l'intermittence prédite est assez similaire à ce qui s'observe sur le réseau de référence. Le fleuve Niger est prédit permanent tout au long de sa traversée au Mali tout comme la partie amont du fleuve Sénégal, tout cela en bon accord avec le réseau de référence. La principale différence avec la carte de référence est que certains affluents de ces fleuves sont prédits éphémères alors qu'ils sont classés permanents dans la base de référence. La comparaison de la classification de l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans le réseau hydrographique de référence du Mali aux emplacements des stations de jaugeage montre une bonne concordance avec une *accuracy* de 74%, un *recall* de 1 pour les stations permanentes (**Figure III-26**), ce qui veut dire que tous les tronçons de cours d'eau sur lesquels les stations permanentes sont situées ont été correctement classés dans la base de référence, mais par contre on a un *recall* de 0 pour la classe intermittente ce qui signifie qu'aucune station intermittente n'est correctement classée au tronçon de cours d'eau correspondant selon la base de référence et donc ces stations intermittentes (6/23) sont classées permanentes dans le tronçon de cours d'eau correspondant selon la base de référence.

		Classification du réseau de référence (IGM-Mali)		Total	Recall	Score F1
		Permanent	Intermittent			
Classification au droit des stations de jaugeages	Permanent	17	0	17	1	0,85
	Intermittent	6	0	6	0	0
	Total	23	0	23		
	Precision	0,74	0			

**Figure III-26** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence IGM-Mali aux emplacements des stations de jaugeages.

En Afrique du Sud, la distribution spatiale de l'intermittence observée (**Figure III-27.a**) est très similaire à celle prédite (**Figure III-27.b**). Les coefficients de corrélation spatiale sont satisfaisants et de l'ordre de 0,48 pour le réseau M3-LCS, mais avec une RMSE élevée, sans doute à cause de la densité de drainage plus élevée du réseau LCS, qui correspond à une échelle cartographique 1 :50 000. Les principaux fleuves tels que l'Orange sont correctement prédits permanents. Une des raisons de cette bonne prédiction est la présence d'un nombre important de stations de jaugeage dans le pays ainsi qu'un nombre d'années de disponibilité de données assez consistant (**Figure III-4**). La comparaison de la classification de

l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans le réseau hydrographique de référence de l'Afrique du Sud aux emplacements des stations de jaugeage montre une très bonne concordance avec une *accuracy* de 75% avec un *recall* de 0,93 pour les stations permanentes (**Figure III-28**), ce qui signifie que 93% des stations permanentes (202/217) sont correctement classées dans le tronçon correspondant dans la base de référence tandis que nous avons un *recall* de 0,22 pour les stations intermittentes ce qui signifie que seulement 22% des stations intermittentes (16/74) sont correctement classées dans le tronçon correspondant dans la base de référence.



**Figure III-27** : Comparaison de la distribution spatiale de la carte nationale de l'intermittence aux prédictions d'intermittence (M3-LCS) en Afrique du Sud (a,b,c,d). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères du modèle M3-LCS ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe.

		Classification du réseau de référence (DWS-Afrique du Sud)				Recall	Score F1
		Permanent	Intermittent	Total			
Classification au droit des stations de jaugeages	Permanent	202	15	217	0,93	0,85	
	Intermittent	58	16	74	0,22	0,30	
	Total	260	31	291			
Precision		0,78	0,52				

**Figure III-28** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence DWS-Afrique du Sud aux emplacements des stations de jaugeages.

Au Bénin, les corrélations spatiales du réseau de référence avec les prédictions sont assez satisfaisantes ( $r = 0,41$ ), la distribution spatiale de l'intermittence prédite suit généralement celle du réseau de référence (Figure III-29.a-d) avec un gradient Nord-Sud (beaucoup de cours d'eau permanents au Sud et beaucoup de cours d'eau éphémères au Nord) qui suit le gradient pluviométrique.

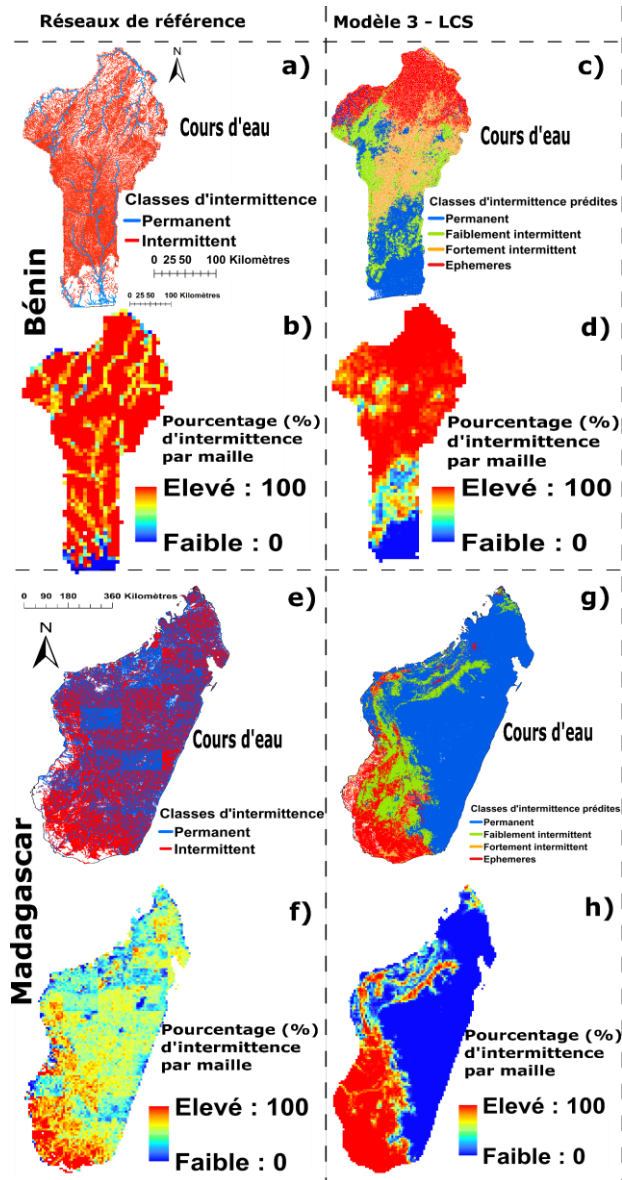


Figure III-29 : Comparaison de la distribution spatiale des prédictions d'intermittence (M3-LCS et M3-LCS-100 km<sup>2</sup>) aux cartes nationales de l'intermittence au Bénin (a,b,c) et Madagascar (d,e,f). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe.

La comparaison de la classification de l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans le réseau hydrographique de référence du Bénin aux emplacements des stations de jaugeage montre une assez bonne concordance avec une *accuracy* de 50% avec un *recall* de 0,18 pour les stations intermittentes

(Figure III-30), ce qui signifie qu'une large part des stations intermittentes (9/11) sont positionnées sur des tronçons classés permanents dans la base de référence, c'est notamment le cas d'un certain nombre de cours d'eau dans la partie Nord du pays.

		Classification du réseau de référence (IGN-Bénin)				
		Permanent	Intermittent	Total	Recall	Score F1
Classification au droit des stations de jaugeages	Permanent	8	1	9	0,89	0,62
	Intermittent	9	2	11	0,18	0,29
	Total	17	3	20		
	Precision	0,47	0,67			

Figure III-30 : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence IGN-Bénin aux emplacements des stations de jaugeages.

À Madagascar, on a une similitude de la distribution spatiale de l'intermittence avec un même partage de l'île entre cours d'eau éphémère au Sud-Ouest et plus permanent sur une grande partie Nord-Ouest (Figure III-29.e-h). Les principales différences semblent se situer dans la partie Est et Nord de l'île où le pourcentage d'intermittence est plus faible dans le modèle M3-LCS que dans le réseau de référence. Les coefficients de corrélations spatiales entre le réseau de référence avec les prédictions sont faibles ( $r = 0,39$ ). Sur la Figure III-29.e, on observe également un motif rectangulaire comme discuté au chapitre I et qui est sans doute lié aux dalles d'images satellites analysées. La comparaison de la classification de l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans le réseau hydrographique de référence de Madagascar aux emplacements des stations de jaugeage montre une très bonne concordance (Figure III-31), avec une *accuracy* de 94% avec un *recall* de 0 pour la seule station intermittente localisée au Sud de l'île ce qui signifie que la station est classée permanente dans le tronçon de cours d'eau correspondant dans la base de référence.

		Classification du réseau de référence (FTM-Madagascar)				
		Permanent	Intermittent	Total	Recall	Score F1
Classification au droit des stations de jaugeages	Permanent	16	0	16	1	0,97
	Intermittent	1	0	1	0	0
	Total	17	0	17		
	Precision	0,94	0			

Figure III-31 : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence FTM-Madagascar aux emplacements des stations de jaugeages.



La comparaison de la classification de l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans les réseaux hydrographiques de référence aux tronçons de rivière correspondant aux emplacements des stations de jaugeage montre qu'en général, l'intermittence est sous-estimée dans les bases hydrographiques de référence où les grands cours d'eau ont tendance à être systématiquement classés permanents. Ceci contribue certainement aux faibles coefficients de corrélation spatiale entre le réseau de référence avec la prédiction du modèle notamment au Burkina Faso. Parmi les autres études comparées au réseau de référence au Bénin, Burkina Faso, Mali, Madagascar, et Afrique du Sud (**Tableau I-5, Tableau III-5**), le modèle M3-LCS est toujours parmi les études qui montrent une bonne corrélation avec le réseau de référence. De façon générale, le modèle M3-LCS capture la distribution spatiale de l'intermittence dans les bases hydrographiques de référence investiguées. Les différences fortes en M3-LCS et les réseaux de référence investigués soulignent les zones où il faudrait des études hydrographiques plus fines.

#### III- 4 - 4 - 2 Comparaison avec les estimations d'études antérieures

Pour appréhender les similitudes et les différences avec les études globales récentes qui cartographient l'intermittence (Lin *et al.*, 2021 ; Messenger *et al.*, 2021), une comparaison est effectuée à l'échelle de l'Afrique. Dans Lin *et al.* (2021), une estimation de la proportion des cours d'eau intermittents (*fi*) par bassin versant à partir du réseau Merit hydro (Lin *et al.*, 2021) est proposée, tandis que dans Messenger *et al.* (2021), la base GIRES fournit pour chaque tronçon de cours d'eau du réseau hydroRIVERS (Linke *et al.*, 2019) la probabilité d'être intermittent. Les résultats de prédiction de ce chapitre ne sont pas directement comparables avec les estimations précédentes sans calcul supplémentaire, car notre classification est multiclasse alors que celles des études récentes est en réalité binaire. Pour comparer ces estimations à la prédiction faite dans ce chapitre (M3-LCS), nous nous sommes basés sur la classification GIRES Classe 2 à pas de temps mensuel qui est plus adapté à la comparaison avec nos prédictions mensuelles d'intermittence. Dans GIRES Classe 2 les cours d'eau intermittents sont définis comme des cours d'eau s'arrêtant de s'écouler en moyenne au moins un mois (au moins 30 jours) par an et inversement pour ceux définis permanents (< 30 jours en moyenne par an). Pour les prédictions effectuées dans ce chapitre, les classes faiblement intermittentes, fortement intermittentes et éphémères ont été regroupées en une seule classe intermittente et la classe permanente ( $\leq 1$  mois c'est-à-dire  $\leq 30$  jours sans

écoulement en moyenne) en une seule classe ce qui correspond à une définition presque équivalente à celle de GIRES Classe 2. Ainsi, pour les estimations de Messenger *et al.* (2021) et celles de ce chapitre, des pourcentages d'intermittence sont calculés par mailles de 5' (environ 8 km x 8 km) en faisant la somme des longueurs des cours d'eau classés intermittents sur la longueur totale des cours d'eau dans la maille. Les estimations de Lin *et al.* (2021) étant des proportions d'intermittence moyennes dans des bassins versants, nous les avons rastérisés à la résolution de 5' pour avoir des proportions d'intermittence dans des mailles de 5'. Ces calculs permettent une comparaison de ces estimations en termes de proportion d'intermittence, comme récapitulés en **Tableau III-6** à l'aide de plusieurs critères statistiques de comparaison.

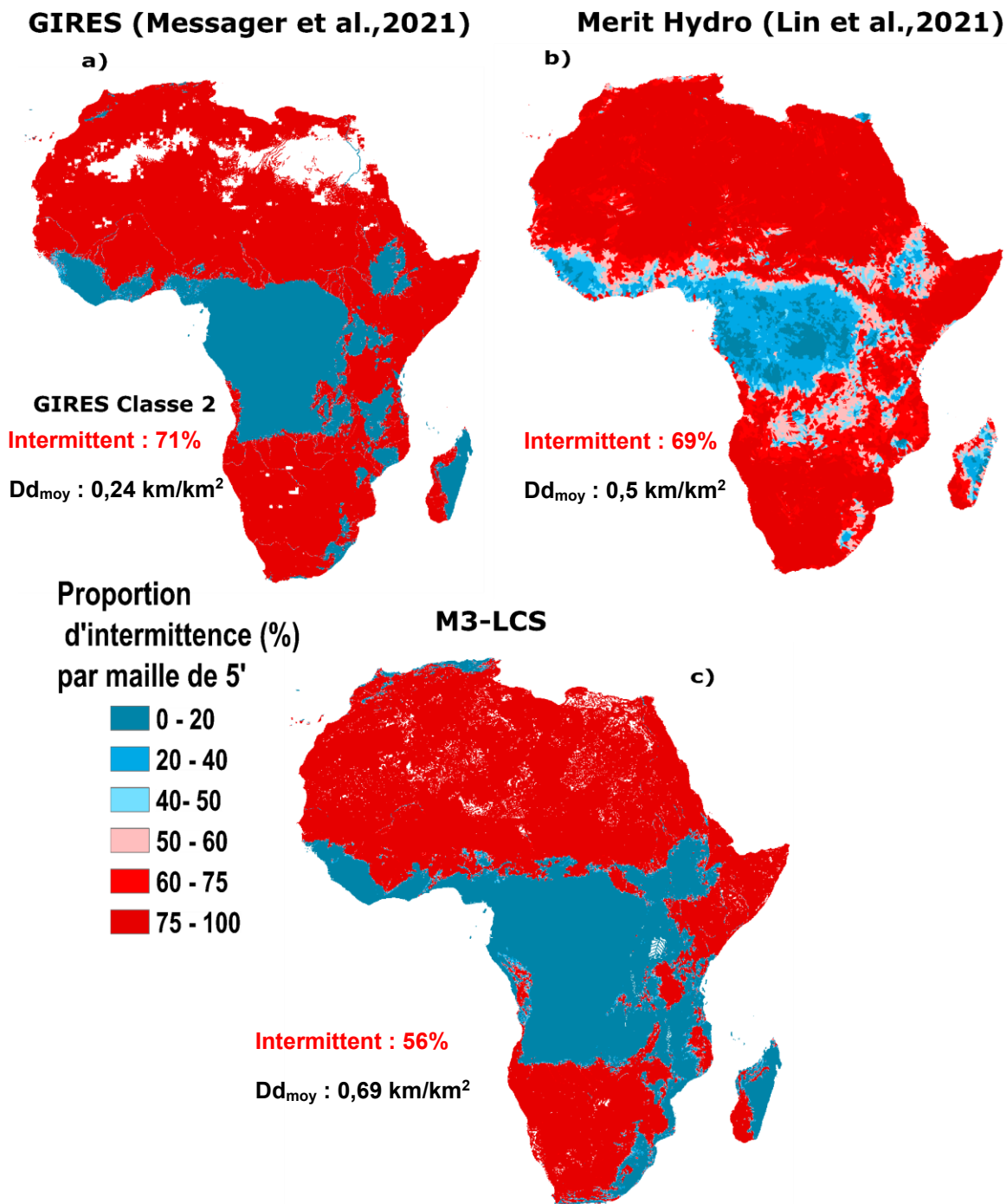
Nos prédictions M3-LCS sont très fortement corrélées aux estimations de GIRES-Classe 2 et de Lin *et al.* (2021) avec respectivement des coefficients  $r = 0,79$  et  $r = 0,76$  (**Tableau III-6**).

**Tableau III-6** : Paramètres statistiques de comparaison en Afrique des prédictions d'intermittence du modèle 3 appliqué au réseau LCS (M3-LCS) par rapport aux prédictions d'intermittence de l'étude de Messenger *et al.* (2021) sur le réseau hydroRIVERS (base GIRES) et de Lin *et al.* (2021).

Études / Base de données	Paramètres statistiques					
	Dd <sub>moy</sub> (km/km <sup>2</sup> )	% intermittent (%)	% moyen intermittent (%)	Écart-type (%)	Corr. (-)	RMSE (%)
(Messenger <i>et al.</i> , 2021) <b>GIRES-Classe2</b>	0,24	71	74	43	1	0
<b>M3-LCS</b>	0,69	56	67	45	0,79	30
(Lin <i>et al.</i> , 2021) <b>Merit-hydro</b>	0,5	69	73	25	1	0
<b>M3-LCS</b>	0,69	56	67	45	0,76	31

*Dd<sub>moy</sub>* : densité de drainage (km/km<sup>2</sup>) moyenne dans les mailles de 5' (environ 8 km x 8 km) - % interm. : pourcentage (%) en longueur des cours d'eau intermittents sur la longueur totale des cours d'eau dans le pays ou territoire considéré - % moyen interm. : pourcentage (%) moyen dans les mailles de 5' des cours d'eau intermittents – Écart-type (%) du pourcentage d'intermittence dans les mailles de 5' – Corr. : Coefficient de corrélation de Pearson (-) du pourcentage d'intermittence entre les réseaux de référence et les réseaux prédits dans les mailles de 5'. RMSE (%) : Racine carrée de l'erreur quadratique moyenne du pourcentage d'intermittence entre les réseaux de référence et les réseaux prédits dans les mailles de 5'.

La **Figure III-32** présente, à l'échelle de l'Afrique, la distribution spatiale des proportions d'intermittence des prédictions faites dans ce chapitre (M3-LCS) ainsi que celles des études antérieures. De façon générale, nous prédisons à l'échelle de l'Afrique moins de cours d'eau intermittents (56%) que les autres études, avec 69% pour Lin *et al.* (2021), et 71% pour GIRES-Classe 2.



**Figure III-32** : Comparaison de la distribution spatiale des proportions d'intermittence par maille de 5' a) GIRES-Classe 2 (Messenger et al., 2021), b) Merit hydro (Lin et al., 2021), c) M3-LCS. Dans chacune des figures, les zones en « blanc » correspondant à des zones où il n'y pas de cours d'eau.

Le fait de prédire moins de cours d'eau intermittents en Afrique pourrait s'expliquer partiellement par l'usage d'un réseau hydrographique à densité de drainage variable. Les densités de drainage moyennes du réseau LCS, Merit-hydro et hydroRIVERS en Afrique sont respectivement de 0,69, 0,5 et 0,24 km/km<sup>2</sup> (Tableau III-6, Tableau I-1). En effet, les fortes densités de drainage du réseau LCS (Schneider et al., 2017) sont concentrées dans les zones prédites comme étant majoritairement permanentes notamment dans les zones équatoriales

(**Figure III-32.c**), tandis que dans les zones désertiques où la majorité des cours d'eau sont intermittents, la densité de drainage donc la longueur totale des cours intermittents est plus faible (**Figure I-2.g**).

GIRES-Classe 2 est prédit sur le réseau hydrographique hydroRIVERS (Linke *et al.*, 2019) qui présente des densités de drainage plus faibles en général dans les zones désertiques comme dans les zones équatoriales plus humides (**Figure I-2.e**). Les études de Lin *et al.* (2021) se basent sur le réseau Merit-Hydro pour proposer un réseau hydrographique à densité de drainage variable spatialement en Afrique avec des densités de drainage moins élevées que LCS en zones équatoriales et de plus faibles densités de drainage dans les zones désertiques (**Figure I-2.f**). Ces éléments évoqués ci-dessus peuvent être une des raisons qui expliquent les différences en termes de proportion entre les résultats obtenus dans ce chapitre et ceux de Messenger *et al.* (2021) ; Lin *et al.* (2021). Enfin, une autre des raisons qui pourrait expliquer ces différences est le fait que notre modèle tend à surestimer les cours d'eau permanents, étant donné que les stations classées permanentes constituent la majorité (636/901) des stations de calibration.

La similarité entre ces études antérieures et celle du présent chapitre est très fortes (**Figure III-32**). Cependant, le modèle 3 prédit davantage de cours d'eau permanents que les autres études (**Figure III-32**) en Afrique du Nord (Tunisie et Maroc), Afrique de l'Est (sud du Soudan du Sud et Tanzanie), Afrique Australe (Mozambique et Afrique du Sud). Le présent chapitre fournit par ailleurs une classification plus détaillée en 4 classes d'intermittence (**Figure III-19.a**) qui sont regroupées en deux classes dans cette section pour faciliter la comparaison avec les études récentes qui fournissent en réalité une classification binaire de l'intermittence. Le modèle présente par ailleurs des erreurs notamment dans les zones désertiques où les cours d'eau éphémères sont sous-estimés.

## III - 5 Discussion

Dans ce chapitre, le modèle M3-LCS finalement sélectionné a montré une bonne efficacité dans la classification de l'intermittence avec une *accuracy* globale de 69%, un critère de Kappa de 0,37 et Kappa pondéré de 0,42 en validation. Les différentes classes d'intermittence des stations ont été classées avec des scores F1 de 0,85, 0,36, 0,27 et 0,34 respectivement pour les classes permanentes, faiblement intermittentes, fortement intermittentes et éphémères. Ce niveau de précision est comparable à certaines études portant sur la

classification de l'intermittence des cours d'eau. Price *et al.* (2021) ont estimé l'*accuracy* globale en validation de 61%. Cette valeur est plus faible que celle obtenue dans cette étude, mais avec un score F1 de 0,38, 0,22, 0,72 et 0,61 respectivement pour les classes 1 à 4 qui sont un peu élevés que ceux obtenus dans cette étude. Le Kappa obtenu en validation dans cette étude (0,37) est supérieur à celui obtenu par Snelder *et al.* (2013) dont la valeur était de 0,32. Le modèle 3 a plus de difficultés à correctement classer les stations intermittentes (faiblement intermittente, fortement intermittente et éphémère) notamment les stations classées fortement intermittentes qui n'ont été correctement identifiées que dans 29% des cas. Cela est certainement dû au faible nombre de stations intermittentes (29,3% contre 70,7% pour la classe permanente). Cette faiblesse du modèle M3-LCS renvoie au fait que les stations de jaugeage sont majoritairement placées sur des cours d'eau permanents (Krabbenhof *et al.*, 2022), si bien qu'une partie de l'information sur les cours d'eau intermittents n'est pas accessible (Ward *et al.*, 2020), alors qu'ils semblent constituer la majorité des cours d'eau, même par notre modèle qui les sous-estime par rapport aux autres études de la littérature (**Tableau III-6**).

Pour les deux modèles 3 et 4 utilisés pour analyser à l'échelle de l'Afrique l'importance des variables environnementales dans la prédiction de l'intermittence, il est ressorti que les variables liées au climat et à la topographie sont les plus prépondérantes. Pour le modèle 3, les trois variables environnementales les plus importantes par ordre sont : (1) l'indice d'aridité (P/Etp), (2) la surface amont (Surf) et (3) l'évapotranspiration potentielle annuelle (Etp) tandis que pour le modèle 4, les variables par ordre d'importance sont : (1) l'indice d'aridité (P/Etp), (2) la température moyenne annuelle de l'air (Tmoy) et (3) la surface amont (Surf). De l'échelle locale à l'échelle globale, plusieurs travaux antérieurs ont trouvé que différentes caractéristiques environnementales influencent l'intermittence des cours d'eau, comme synthétisés en **Tableau III-7**.

**Tableau III-7** : Facteurs de contrôles de l'intermittence identifiés dans quelques études à travers le monde. Les études sont classées par ordre croissant de l'échelle du bassin versant à l'échelle globale.

Études	Zones d'études	Facteurs de contrôle identifiés comme les plus importants de l'intermittence par ordre	Méthodes
(Sando et Blasch, 2015)	Bassin versant du Lower Clark Fork, États-Unis	Pourcentage d'occupation du manteau neigeux, moyenne annuelle de la température minimale mensuelle pluriannuelle, géologie	Random Forest

(González-Ferreras et Barquín, 2017)	Bassin versant Deva-Cares, Espagne	Surface des bassins versants, pourcentage d'occupation de la forêt, précipitation moyenne du mois d'août	Random Forest
(Kaplan <i>et al.</i> , 2020)	Bassin versant de l'Attert, Luxembourg	Surface des bassins versants, courbure du profil de pente, conductivité hydraulique à saturation	Régression logistique
(Yu <i>et al.</i> , 2018)	Bassins versants dans le sud-est du Queensland, Australie	Moyenne pluriannuelle de la température journalière maximale annuelle, surface du bassin versant, pente	Régression linéaire
(Beck <i>et al.</i> , 2017)	Bassin versant de Santa Ana - San Diego, États-Unis	Moyenne sur trois mois des précipitations mensuelles, moyenne sur trois mois de la température moyenne mensuelle, surface amont du bassin versant	<i>Generalized boosted modeling (gbm)</i>
(Jaeger <i>et al.</i> , 2019)	Bassins versants de la région du nord-ouest pacifique (code NHD 17), États-Unis	Précipitation annuelle, moyenne annuelle de la température minimale journalière pluriannuelle, pourcentage de forêt	Random Forest
(Snelder <i>et al.</i> , 2013)	France métropolitaine	Surface des bassins versants, forme du bassin versant, température minimale moyenne de l'air en janvier	Random Forest
(Perez-Saez <i>et al.</i> , 2017)	Burkina Faso	Débit moyen annuel, pente, indice d'aridité	<i>Gradient boosting tree (GBT)</i>
(Belemtougri <i>et al.</i> , 2021a)	Burkina Faso	Perméabilité, surface amont des bassins versants, précipitation moyenne annuelle	Analyse en composante principale (ACP)
(Sauquet <i>et al.</i> , 2021)	Australie, France, Angleterre, États-Unis	Indice d'aridité, température moyenne annuelle, précipitation moyenne annuelle	Analyse statistique d'homogénéité des classes
(Zipper <i>et al.</i> , 2021)	États-Unis	Indice d'aridité, surface amont des bassins versants, profondeur du substratum rocheux	Random Forest
(Hammond <i>et al.</i> , 2021)	États-Unis	Indice d'aridité, pourcentage des surfaces imperméables, température moyenne annuelle	Random Forest
(Messenger <i>et al.</i> , 2021)	Global	Indice d'aridité, teneur en eau du sol, débit moyen annuel (WaterGAP v2.2)	Random Forest
(Lin <i>et al.</i> , 2021)	Global	Altitude, débit moyen annuel, indice d'aridité	<i>Gradient boosting tree (GBT)</i>

À l'échelle régionale du Burkina Faso, l'analyse par *Random Forest* (section II-8 du chapitre II) confirme comme dans l'ACP, le rôle prépondérant de la perméabilité amont des bassins versants dans la discrimination des différentes classes d'intermittence. Parmi les variables identifiées comme contrôle de l'intermittence au Burkina Faso (Perméabilité, surface amont, précipitation moyenne annuelle, et ordre de Strahler) seule la perméabilité ne ressort pas

parmi les contrôles les plus importants de l'intermittence dans les modèles à larges échelles (modèle 3 et modèle 4), les autres variables (surface amont, ordre de Stahler qui sont très corrélés, **Figure III-6**) et (précipitation moyenne annuelle et indice d'aridité qui sont très corrélés, **Figure III-6**) ressortent en quelque sorte comme contrôle important de l'intermittence à l'échelle de l'Afrique (**Figure III-13**). Dans ce **Tableau III-7**, les variables les plus fréquentes sont la surface amont du bassin versant, la précipitation, l'indice d'aridité et la température qui sont aussi identifiés parmi les trois variables les plus importantes dans le modèle 3 et 4. Concernant les propriétés hydrogéologiques, qui pourraient logiquement contribuer à soutenir des cours d'eau permanents, mais aussi à expliquer de l'intermittence, elles ne ressortent pas dans les modèles à grande échelle y compris ceux de ce chapitre. Ceci peut s'expliquer par les incertitudes des données hydrogéologiques à grande échelle, mais aussi par une influence moindre en général que celle des variables climatiques. À grande échelle, l'indice d'aridité est considéré comme la variable la plus importante dans la discrimination des cours d'eau intermittents. En effet, plus le climat est aride, plus la probabilité de l'intermittence est élevée, comme également mise en évidence par la **Figure III-13** dans laquelle la majorité des stations intermittentes dans le jeu de données se retrouvent dans les zones où la précipitation moyenne annuelle est inférieure à l'évapotranspiration potentielle annuelle (aridité < 1). Par contre, une proportion non négligeable de cours d'eau permanents se trouvent dans des milieux arides et semi-arides, ce qui pourrait s'expliquer forcément par des processus autres que climatiques, faisant intervenir la surface du bassin versant amont (c'est le cas du Nil par exemple) et la géologie (cas de nombreuses stations permanentes situées sur le fleuve Mouhoun). Cependant l'analyse d'importance faite avec le modèle RF dans ce chapitre ne permet pas d'évaluer l'importance des variables à des endroits spécifiques, mais à l'échelle continentale. Dans les études à l'échelle locale (**Tableau III-7**), bien que l'influence du climat reste importante, l'indice d'aridité ne ressort pas partout comme contrôle majeur de l'intermittence, alors qu'il est presque toujours examiné à l'exception par exemple des zones de montagnes (Sando et Blasch, 2015). Pour évaluer l'interaction des variables environnementales sur l'intermittence des cours d'eau dans des zones plus réduites, il aurait fallu explorer plusieurs modèles de RF dans différentes zones climatiques délimitées peut-être par des intervalles d'indice d'aridité comme ceux proposés par l'UNEP<sup>34</sup>. Cette procédure a récemment été explorée aux États-Unis par Hammond *et al.* (2021) où, en complément du modèle RF continental (*conus*),

---

<sup>34</sup> [https://fr.wikipedia.org/wiki/Indices\\_d%27aridit%C3%A9](https://fr.wikipedia.org/wiki/Indices_d%27aridit%C3%A9)

l'influence de variables spécifiques sur l'intermittence des cours d'eau a été identifiée par écorégion. La limite de cette approche en Afrique est cependant la disponibilité spatiale des stations de jaugeage qui sont peu nombreuses et peu dispersées (**Figure III-4**).

Dans l'étude de Messenger *et al.* (2021), la teneur en eau du sol est la deuxième variable la plus importante après l'indice d'aridité. Dans cette étude, une corrélation très forte a été trouvée (0,91) entre l'indice d'aridité (Arid) et la teneur en eau du sol (Swc) (**Figure III-6**) d'où la non-prise en compte de cette variable dans le modèle RF pour limiter l'autocorrélation et donc la redondance d'informations. Lin *et al.* (2021) n'ont pas fourni un ordre d'importance des variables dans la prédiction de la proportion d'intermittence ( $f_i$ ). Les échanges effectués avec ces auteurs ont permis d'acquérir les informations renseignées dans le **Tableau III-7**.

Peu d'études ont essayé de fournir une comparaison quantitative de l'intermittence, la majeure partie des études se limitant à une comparaison qualitative de la distribution spatiale de l'intermittence. Dans ce chapitre, la comparaison à l'échelle nationale de la prédiction d'intermittence avec celle des bases nationales montre des résultats très satisfaisants pour certains pays comme l'Afrique du Sud et le Mali (**Figure III-25**) avec une bonne ressemblance spatiale entre réseaux prédits et réseaux de référence. Un réseau hydrographique de référence de l'Afrique du Sud avec les informations sur l'intermittence à une échelle plus fine aurait peut-être montré encore plus de similarité avec nos prédictions. La configuration spatiale des classes d'intermittence prédites montre aussi une bonne concordance avec les réseaux de référence au Bénin et à Madagascar avec un rendu passable par rapport à la base de référence du Burkina Faso à cause de la classification erronée dans la base de référence pour certains cours d'eau (**Figure III-25**, **Figure III-27**, **Figure III-29**). La comparaison de la classification de l'intermittence observée au droit des stations de jaugeage avec la classification de l'intermittence présente dans les réseaux hydrographiques de référence montre qu'en général ces derniers sous-estiment l'intermittence des cours d'eau. Les cartographies de l'intermittence dans certaines bases de données de référence sont plus récentes (de 1994 pour l'Afrique du Sud à 2018 pour le Bénin) alors que notre période d'étude s'étale de 1958-1991 ce qui pourrait être une source de quelques différences.

À l'aide du modèle M3-LCS, nous estimons qu'à l'échelle de l'Afrique, 56% de la longueur totale des cours d'eau présente un régime intermittent (c.-à-d. nombre moyen de mois à débit nul  $\overline{Ndry} > 1$  mois/an). Les proportions des classes d'intermittence estimées par le modèle 3 sur des cours d'eau drainant plus de 100 km<sup>2</sup> du réseau LCS sont pratiquement les mêmes en



tenant compte de tous les cours d'eau LCS. Cependant, il faut souligner que les estimations de l'intermittence sont très incertaines pour les cours d'eau drainant moins de 100 km<sup>2</sup>, car très peu des stations utilisées sont situées sur ces cours d'eau (4%). La proportion d'intermittence estimée (56%) est nettement inférieure à celle donnée par d'autres études récentes qui couvrent l'Afrique et comprises entre 69 et 71% (Lin *et al.*, 2021 ; Messenger *et al.*, 2021). Dans ce chapitre, 1125 stations de jaugeages avec au minimum 4 ans de données complètes sur la période 1958-1991 ont été réparties en 4 classes d'intermittence, la classe permanente étant largement prédominante (70,7%) dans l'échantillon avec une grande part des petits bassins versants jaugée qui sont aussi permanents (**Figure III-10.a**) et majoritairement localisés en Afrique du Sud, ce qui crée un biais d'échantillonnage dans le jeu de donnée, car les informations sur les petits cours d'eau intermittents dans le continent ne sont pas assez nombreuses dans l'échantillon ce qui aurait pu favoriser la prédiction d'un certain nombre de petits cours d'eau permanents par le modèle. Dans l'étude de Messenger *et al.* (2021), qui est globale, seulement 354 stations de jaugeages sont considérées en Afrique avec au moins 10 ans de données pour une proportion 41% pour les stations de jaugeage classées permanentes. En ayant très peu de stations en Afrique et majoritairement concentrées en Afrique australe, il est possible qu'un certain nombre de processus dans certaines zones géographiques comme l'Afrique du Nord et de l'Ouest soit occulté par cette dernière étude même si les stations prises en considération ont généralement une chronique qui est plus longue que celle de notre étude. Ces différences dans l'échantillon de départ en plus des éléments évoqués sur la densité de drainage des cours d'eau (section III-4-4-2) sont peut-être à l'origine des différences des proportions calculées sur les cartes d'intermittence (**Figure III-32.a-c**).

Malgré un traitement des données dans la phase de calibration du modèle 3 pour réduire l'influence du déséquilibre entre les classes par la technique SMOTE (Chawla *et al.*, 2002), les cours d'eau éphémères des déserts (Sahara, Kalahari) sont prédits fortement intermittents au lieu d'éphémères comme attendu (**Figure III-19, Figure III-20**), ce qui suggère que la technique SMOTE a un intérêt limité. Une simple méthode qui pourrait compléter cette technique et améliorer sensiblement les prédictions du modèle serait de placer dans ces zones désertiques quelques stations fictives définies comme éphémères.

Il est aussi possible que l'échelle des cartes de perméabilité (> 1 : 1 000 000) et de stockage en eau souterraine (30 km x 30 km) utilisées pour définir les variables hydrogéologiques prédictives soit trop grossière pour discriminer certaines variations hydrogéologiques à l'échelle de l'Afrique. Aux endroits où l'intermittence des cours d'eau est liée à la dynamique

des eaux souterraines, des données spatiales à fine échelle correspondant à la perméabilité des bassins et à d'autres facteurs hydrogéologiques pourraient améliorer la capacité à modéliser les cours d'eau intermittents avec précision, mais ces données sont rarement disponibles à fine échelle. Tant que ce manque de précision sur certaines catégories de données telles que les données hydrogéologique, géologique, lithologique, et pédologique n'aura pas été comblé, les modèles prédictifs auront toujours une précision limitée sur des processus spatialement hétérogènes comme l'intermittence.

Une autre solution pour améliorer les prédictions actuelles de l'intermittence des cours d'eau en Afrique est probablement l'amélioration de la qualité et de la quantité des données de débits à travers l'installation de nombreuses stations de jaugeage (en particulier sur les rivières intermittentes) qui ont connu une baisse depuis les années 1970 jusqu'à présent comme illustrée par la **Figure III-1**. Cependant, ces initiatives sont coûteuses et exigent un investissement considérable de ressources. Dans d'autres pays et continents, les études sur l'intermittence des cours d'eau ont bénéficié d'un nombre plus important de stations hydrométriques bien jaugées et assez bien réparties spatialement (Kennard *et al.*, 2010 ; Price *et al.*, 2021 ; Snelder *et al.*, 2013 ; Zipper *et al.*, 2021), ce qui n'est pas le cas en Afrique où les données des stations hydrométriques sont courtes et moins denses. Une autre piste serait de compléter les stations de jaugeages déjà exploitées dans ce chapitre par des stations provenant d'autres continents pour la calibration du modèle RF.

En basses eaux, les faibles débits sont parfois difficiles à mesurer et peuvent être mal évalués soit par de faux zéros ou des valeurs constantes (Zimmer *et al.*, 2020) ou dans certains cas des lacunes transformées en des valeurs nulles ou des données stockées sans décimales (Meerveld *et al.*, 2020). Dans cette étude malheureusement, dans le large jeu de données de débits collectées, nous n'avons pas fait d'analyse de stationnarité, les données ont été utilisées telles que téléchargées ou collectées auprès des gestionnaires de données.

## III - 6 Conclusion

Ce chapitre visait à comprendre les facteurs de contrôle de l'intermittence ainsi que la distribution spatiale des cours d'eau intermittents à l'échelle continentale de l'Afrique. À l'aide d'une méthode d'apprentissage automatique (*Random Forest*), basée sur un jeu de 1125 chroniques de débit avec au minimum 4 ans de données complètes sur la période 1958-1991,

il a été mis en évidence que les trois facteurs de contrôle les plus importants de l'intermittence à l'échelle de l'Afrique sont par ordre décroissant : l'indice d'aridité, la surface amont que drainent les cours d'eau, et l'évapotranspiration moyenne annuelle. L'extrapolation du modèle *Random Forest* entraîné sur ces données indique que 56% de la longueur totale des cours d'eau en Afrique sont intermittents (faiblement intermittent (8%), fortement intermittent (35%), et éphémères (12%)). La proportion d'intermittence estimée en Afrique dans ce chapitre (56%) est inférieure à celles des études antérieures 69 et 71% (Lin *et al.*, 2021 ; Messenger *et al.*, 2021). Les cours d'eau permanents sont majoritairement localisés dans les zones humides ( $P > Etp$ , c.-à-d. aridité  $> 1$ ) avec les cours d'eau intermittents majoritairement présents dans les zones de transition climatiques ( $P$  proche de  $Etp$ ) et les zones désertiques ( $P$  largement inférieure à  $Etp$ ). Dans les zones désertiques, le modèle a tendance à sous-estimer les cours d'eau éphémères au profit des cours d'eau fortement intermittents qui sont prédits à ces endroits. Cette erreur manifeste pourrait être corrigée en complétant le jeu des stations de jaugeages avec des stations provenant des zones arides situées dans d'autres continents ou en plaçant des stations fictives définies éphémères dans les zones désertiques. Les études menées dans ce chapitre couvrent la période (1958-1991), ce qui implique que d'éventuels évolutions des variables climatiques dont l'indice d'aridité à travers le temps comme attendu sur la période 2080-2099 avec une augmentation significative de l'aridité en Afrique Australe (Greve *et al.*, 2019, Figure 2) contribueront à un changement significatif dans la distribution spatiale de l'intermittence des cours d'eau et donc aussi de la proportion des différentes classes d'intermittence en Afrique. Le modèle *Random Forest* développé dans ce chapitre peut être utilisé en mettant à jour les variables climatiques par les projections climatiques futures ce qui permettra de cartographier l'évolution de l'intermittence du débit en Afrique sous l'influence du changement climatique anthropique. Ces informations sont importantes dans l'élaboration de stratégies de conservation de la biodiversité et sont d'un intérêt sur le plan biologique, géomorphologique et biogéochimique. L'information sur la distribution spatiale de l'intermittence des cours d'eau renseigne sur la disponibilité spatio-temporelle en eau de surface ce qui permet d'avoir une idée du potentiel de développement de l'irrigation dans une zone donnée et de faciliter la mobilisation de la ressource en eau.

## Chapitre IV - Développement d'une méthode pour extraire les cours d'eau à partir d'aires contributives variables en Afrique



Nous avons montré au **chapitre I** que les bases hydrographiques globales ne capturent pas bien la variabilité spatiale de la densité de drainage observée sur les réseaux hydrographiques de référence de quelques pays en Afrique. Le réseau LCS comme mis en évidence dans le **chapitre I** était parmi les réseaux hydrographiques globaux pris en compte, celui qui représentait au mieux la variabilité de la densité de drainage des réseaux hydrographiques de référence des quelques pays d'Afrique investigués. Au **chapitre II** et au **chapitre III**, nous nous sommes servis de ce réseau comme support pour la cartographie des cours d'eau intermittents. Dans ce dernier **chapitre IV**, nous avons exploré une nouvelle approche méthodologique devant aboutir à une meilleure représentation (que celles des bases de données globales actuelles) de la variabilité spatiale de la densité de drainage des cours d'eau à l'échelle de l'Afrique en intégrant les données observées (recueillies auprès des instituts géographiques nationaux) ainsi que des variables environnementales dans l'extraction des cours d'eau à l'aide de MNT. Les résultats de la première étape de cette méthodologie ont fait l'objet d'un article (Belemtougri *et al.*, 2021b), publié dans le cadre de la 4<sup>e</sup> conférence internationale sur « *l'hydrologie des grands fleuves africains* », Cotonou, Bénin, novembre 2021 :

**Belemtougri, A.P.**, Ducharne, A., Karambiri, H., 2021b. Improvement of river network representation in Africa: application of an approach based on digital elevation data and environmental characteristics. Proceedings of the International Association of Hydrological Sciences, 384, 19–23. <https://doi.org/10.5194/piahs-384-19-2021>

## IV - 1 Résumé

Dans de nombreuses bases de données continentales représentant les cours d'eau à l'échelle de l'Afrique, l'information sur les caractéristiques des cours d'eau (longueur, largeur, caractère intermittent ou permanent) est souvent incomplète et les petits cours d'eau sont largement sous-représentés. Heureusement, l'exploitation de l'information topographique des modèles numériques de terrain (MNT) fournit une occasion d'avoir sur de vastes domaines plus de détails sur les cours d'eau et en particulier les petits cours d'eau. Une approche commune pour extraire les cours d'eau à partir d'un MNT consiste à considérer comme cours d'eau, toutes les cellules du MNT qui drainent au minimum une certaine surface amont et couramment appelée surface contributive minimum ( $A_{\min}$ ). Ce paramètre ( $A_{\min}$ ) est généralement défini de façon uniforme sur de grands domaines, tenant compte uniquement de l'influence de la topographie sur l'initiation des cours d'eau et négligeant ainsi les conditions climatiques et paysagères très variables ce qui fait que les cours d'eau générés ont une répartition spatiale plutôt uniforme. Pour pallier à cette limitation, dans cette étude, nous nous sommes d'abord intéressés au Burkina Faso où une calibration du réseau hydrographique national de référence (BNDT) a été effectuée en utilisant un MNT de 3'' (environ 90m x 90m). Cette calibration nous a permis d'établir avec des performances très satisfaisantes (coefficient de détermination  $r^2 = 0,86$ ), une première relation donnant des surfaces contributives minimales ( $A_{\min}$ ) variables spatialement en fonction de la densité de drainage observée ( $Dd_{\text{obs}}$ ). La validation de cette première relation au Mali et à Madagascar a donné également des performances très satisfaisantes avec respectivement  $r^2 = 0,92$  et  $r^2 = 0,86$ . Une seconde relation entre les densités de drainage observées ( $Dd_{\text{obs}}$ ) et les variables environnementales (indice d'aridité, altitude, pente, stock d'eau souterrain, profondeur de la nappe) a été établie sur l'ensemble du jeu de donnée (Burkina Faso, Mali, Madagascar) par l'approche *Random Forest* et a permis d'avoir des performances satisfaisantes ( $r^2 = 0,81$ ) en calibration et en validation ( $r^2 = 0,78$ ). Cette deuxième relation combinée à la première relation a permis d'estimer des valeurs de  $A_{\min}$  variables spatialement en Afrique, ces valeurs de  $A_{\min}$  ont finalement servis à l'extraction des cours d'eau par MNT. Ce travail constitue une avancée importante pour mieux décrire les cours d'eau dans les nombreuses régions d'Afrique qui manquent de données hydrographiques précises. Ces informations peuvent avoir une finalité de prévision, de prévention à l'échelle régionale et continentale.

## IV - 2 Introduction

La représentation précise des cours d'eau sert de nos jours d'aide à la décision pour les gestionnaires des territoires pour un suivi et une exploitation efficace de la ressource (Heine *et al.*, 2004 ; Ouellet Dallaire *et al.*, 2019). Les modèles et de nombreuses études sur les processus hydrologiques sont sensibles à la représentation exacte et à la densité des cours d'eau (Brakebill *et al.*, 2011 ; Yan *et al.*, 2020). Cependant, dans de nombreuses bases de données continentales représentant les cours d'eau en Afrique, les informations sur les caractéristiques des cours d'eau sont souvent incomplètes, et les petits cours d'eau sont largement sous-représentés (Lehner *et al.*, 2008). Les bases de données hydrographiques nationales contiennent des informations hydrographiques fiables dans ce sens, qu'elles soient localisées ou nationales. Cependant, l'accessibilité à ces bases de données hydrographiques nationales reste difficile en Afrique en raison du coût requis et de la réticence de certains pays à partager ces informations.

Heureusement, l'exploitation de l'information topographique des modèles numériques de terrain (MNT) fournit une occasion d'avoir sur de vastes zones, plus de détails sur les cours d'eau (Liu *et al.*, 2005). À l'aide des Systèmes d'Information Géographiques (SIG), cette extraction passe généralement par la définition d'un nombre de pixels rivière seuil ( $N_p$ ) qui est une valeur à partir de laquelle on considère un pixel du MNT comme un cours d'eau. La surface amont que draine ce nombre de pixels ( $N_p$ ) correspond à la surface contributive minimale ( $A_{min}$ ) (O'Callaghan et Mark, 1984). Un choix de seuil  $A_{min}$  faible augmente la longueur des cours d'eau extraits et inversement (**Figure i-3**). Le choix de la valeur de  $A_{min}$ , va donc conditionner la longueur des cours d'eau extraits et influence la densité de drainage ( $D_d$ ) qui peut être définie comme le rapport entre la longueur totale des cours d'eau d'un bassin versant et sa surface (Horton, 1945). En général, pour l'extraction des cours d'eau,  $A_{min}$  est défini de manière plutôt arbitraire et constante sur de grands domaines par manque d'informations, ce qui fait que les cours d'eau générés ont une répartition spatiale plutôt uniforme et ne correspondent pas souvent à ceux observés sur le terrain (Papageorgaki et Nalbantis, 2017 ; Passalacqua *et al.*, 2010). C'est l'exemple de la base de données Hydro1k (U.S. Geological Survey, 2000) et AQUAMAPS (FAO, 2014), qui proposent une cartographie des cours d'eau extraite respectivement à partir de MNT à 30" (1 km x 1 km) et 15" (500 m x 500 m) pour des valeurs  $A_{min}$  constantes respectives de 1000 km<sup>2</sup> et 100 km<sup>2</sup>. Dans le milieu naturel, la densité de drainage varie spatialement en fonction de multiples

variables environnementales (climat, lithologie, pente, couvert végétal). La végétation et la précipitation jouent un rôle primordial dans le contrôle de la densité de drainage (Moglen *et al.*, 1998). Tucker *et al.* (2001) montrent que les fortes variations de la densité de drainage observée dans les bassins versants de l'Apennin au nord de l'Italie sont étroitement liées à la lithologie sous-jacente et à l'altitude. Ainsi, pour capturer la variabilité spatiale de la densité de drainage, il est nécessaire d'utiliser des seuils  $A_{\min}$  variables spatialement (Da Ros et Borga, 1997 ; Tucker *et al.*, 2001).

Récemment, Schneider *et al.* (2017) ont proposé une approche permettant de générer une série de valeurs variables d' $A_{\min}$  en fonction de variables environnementales. Ces valeurs d' $A_{\min}$  sont calibrées sur différents types de paysages (constitués de l'intersection spatiale de la lithologie, de la précipitation moyenne annuelle et de la pente) avec un MNT de 500 m de la base de données HydroSHEDS pour correspondre aux informations hydrographiques observées provenant des bases de données hydrographiques nationales de la France et de l'Australie. Les valeurs de  $A_{\min}$  calibrées ont été extrapolées à l'échelle globale, en tenant compte des caractéristiques du paysage local, de manière à extraire un réseau hydrographique global, appelé LCS. Malgré une validation satisfaisante au Brésil et aux États-Unis par rapport aux bases de données hydrographiques nationales, ce réseau LCS montre des difficultés à reproduire les très fortes densités de drainage observées (Schneider *et al.*, 2017). Le réseau LCS, par exemple, capture partiellement la variabilité spatiale de la densité de drainage du réseau hydrographique national du Burkina Faso (**Tableau I-4, Figure I-8**), ce qui est probablement dû à l'échelle grossière de la couche lithologique utilisée.

Sur la base de la méthodologie utilisée par Schneider *et al.* (2017), notre objectif est d'utiliser la précision des informations locales collectées auprès des instituts géographiques nationaux et des informations climatiques, physiographiques et de MNT largement disponibles pour établir une première relation entre la surface contributive minimale ( $A_{\min}$ ) et la densité de drainage (Dd), ensuite d'établir une seconde relation entre la densité de drainage et les variables environnementales (lithologie, climat, géologie et propriétés des sols). La combinaison de ces deux relations devrait permettre d'estimer des valeurs de  $A_{\min}$  variables spatialement en fonction des caractéristiques environnementales. Ces valeurs de  $A_{\min}$  permettront finalement d'extraire des cours d'eau en Afrique dont les densités de drainage correspondront au mieux à celles observées.

## IV - 3 Matériel et Méthodes

### IV - 3 - 1 Zones d'étude

Nous nous sommes intéressés principalement au Burkina Faso à Madagascar et au Mali (**Figure IV-1**) où les bases hydrographiques de référence ont la même échelle (1 : 200 000). Ces pays peuvent être présentés comme suit :

- Le Burkina Faso est un pays enclavé de 274 200 km<sup>2</sup> situé en Afrique de l'Ouest. Le pays couvre des zones climatiques assez variées avec un gradient pluviométrique annuel entre le Sud et le Nord (1200 à moins de 600 mm/an respectivement) (PANA, 2007). Les valeurs des altitudes sont principalement comprises entre 250 m et 350 m, avec une altitude maximale autour de 750 m rendant le relief relativement plat.
- Le Mali est un pays enclavé de 1 241 238 km<sup>2</sup>, il couvre également des zones climatiques assez variées avec un gradient pluviométrique annuel entre le Sud et le Nord (plus de 1000 à moins de 200 mm/an respectivement). Les valeurs des altitudes sont principalement comprises entre 250 m et 500 m, le point culminant est Hombori Tondo (1153 m), situé dans le centre du pays<sup>35</sup>.
- Madagascar est une île de 587 040 km<sup>2</sup> située dans l'océan Indien à 400 km à l'est des côtes africaines. La précipitation annuelle oscille entre 300 mm à 3500 mm avec des valeurs d'altitudes<sup>36</sup> comprises entre 0 et 2876 m.

---

<sup>35</sup> [https://fr.wikipedia.org/wiki/G%C3%A9ographie\\_du\\_Mali](https://fr.wikipedia.org/wiki/G%C3%A9ographie_du_Mali)

<sup>36</sup> [https://fr.wikipedia.org/wiki/G%C3%A9ographie\\_de\\_Madagascar](https://fr.wikipedia.org/wiki/G%C3%A9ographie_de_Madagascar)



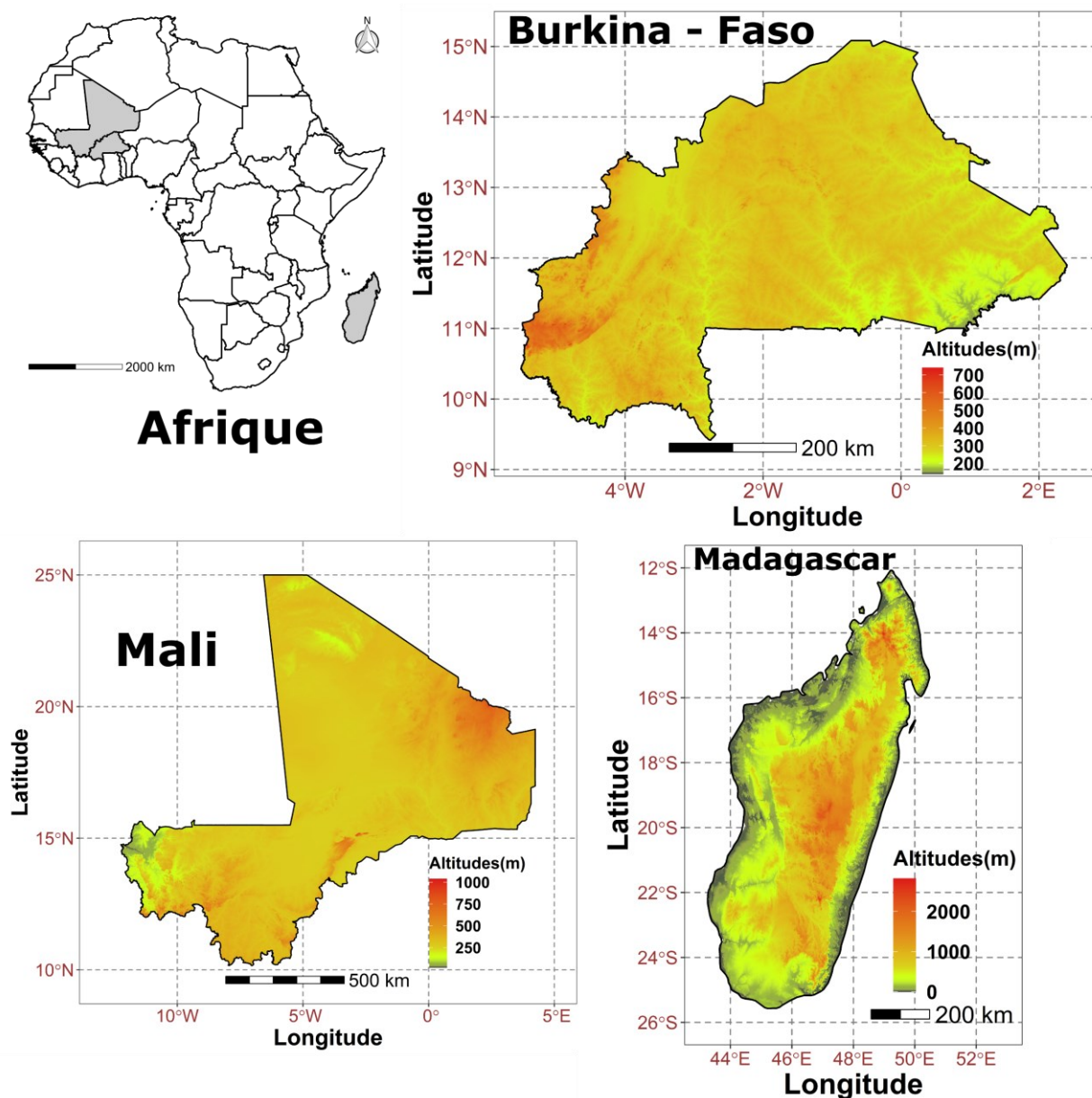


Figure IV-1 : Localisation en Afrique et description des principaux pays investigués (Burkina Faso, Mali, Madagascar).

### IV - 3 - 2 Description des données

Le MNT utilisé dans cette étude provient de la base de données hydroSHEDS (Lehner *et al.*, 2008), qui est la plus utilisée actuellement pour des applications hydrologiques. La résolution choisie est 3" (90m x 90m environ à l'équateur), c'est-à-dire la plus petite résolution proposée par la base. La base de données hydrographiques de référence prise en compte au Burkina Faso est la BNDT disponible à l'échelle 1 : 200 000 et distribuée par l'IGB (Institut Géographique du Burkina). Les bases de données hydrographiques de référence prises en compte au Mali et à Madagascar proviennent respectivement de l'IGM et FTM à l'échelle 1 :

200 000. Les bases de données hydrographiques de référence considérées en France et au Bénin sont respectivement la BD CARTHAGE et la BDI, disponibles à la même échelle de 1 : 50 000, et distribuées par l'IGN (Institut Géographique National) France et l'IGN Bénin. Dans tous ces réseaux hydrographiques de référence considérés, il a été possible de restreindre les cours d'eau à ceux qui sont naturels uniquement en excluant les cours d'eau artificiels.

Un certain nombre de variables environnementales a été sélectionné pour estimer la variabilité spatiale de la densité de drainage (Dd). Les variables investiguées sont les mêmes que celles utilisées au **chapitre III (Tableau III-2)** à l'exception de la surface des bassins versants, l'ordre de Strahler et le débit spécifique, étant donné que nous travaillons à l'échelle des mailles dans ce chapitre ces variables n'ont pas été prises en compte. Une résolution plus fine des données d'altitudes et de pentes (90 m) a été prise en compte dans ce chapitre pour rester en conformité avec la résolution du MNT (90 m x 90 m) utilisée pour extraire les cours d'eau.

### IV - 3 - 3 Méthodes

Toutes les cartes et le traitement spatial sont réalisés à l'aide du logiciel ArcGIS (10.6.1) et les graphiques à l'aide du logiciel R version 3.6.1. L'outil ArcGIS a été utilisé pour extraire 400 réseaux hydrographiques au Burkina Faso à l'aide du MNT d'où sont extraites progressivement les grilles de direction et d'accumulation. Pour couvrir une large gamme de densités de drainage, le nombre de pixels définis pour extraire ces réseaux ( $N_p$ ) varie entre 1-40000, ce qui correspond à des  $A_{min}$  entre 0,0081-324 km<sup>2</sup>. Le territoire du Burkina Faso a été divisé en des mailles de 5' (environ 64 km<sup>2</sup> à l'équateur) et tous les calculs sont effectués sur ce maillage. Pour chacun des réseaux extraits précédemment, la densité de drainage est calculée comme la somme totale des longueurs de cours d'eau (km) dans une maille de 5' divisée par la surface de la maille (km<sup>2</sup>). La même procédure est utilisée pour calculer la densité de drainage observée ( $Dd_{obs}$ ) en utilisant les informations hydrographiques contenues dans la base de données hydrographiques de référence du Burkina Faso (**Figure IV-2.a**), du Mali, de Madagascar, du Bénin et de la France.

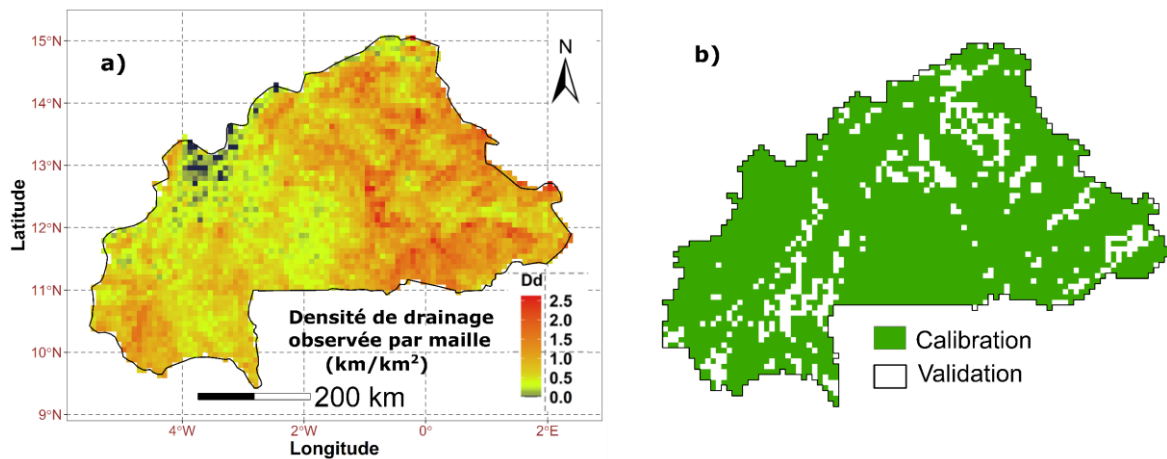


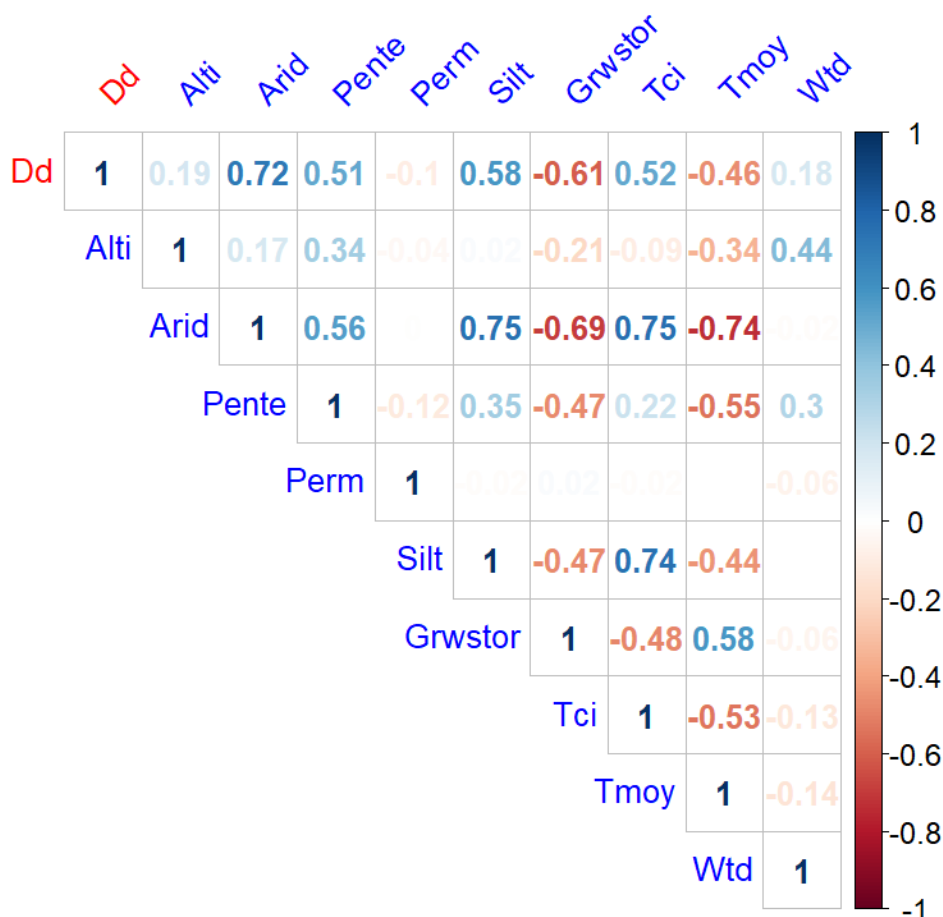
Figure IV-2 : Processus de calibration des relations au Burkina Faso. (a) Représentation de la densité de drainage observée dans les mailles de 5'. (b) Zones de calibration et de validation.

Au Burkina Faso, pour chaque maille, la valeur d' $A_{\min}$  sélectionnée est définie par le réseau extrait (parmi les 400 possibilités) qui approche le mieux la densité de drainage observée ( $Dd_{\text{obs}}$ ) de la couche BNDT. Ainsi, l' $A_{\min}$  sélectionné est celui qui minimise dans la maille, la différence de densité de drainage entre les réseaux extraits et le réseau observé de la BNDT. Nous avons sélectionné aléatoirement plusieurs mailles qui ont été exclues du processus de calibration et qui représentent 14 % du territoire (Figure IV-2.b). C'est sur ces mailles que la validation de la relation établie entre  $A_{\min}$  calibré et  $Dd_{\text{obs}}$  a été effectuée au Burkina Faso. Cette première relation est également validée au Mali et à Madagascar qui ont la même échelle de réseau hydrographique qu'au Burkina Faso (échelle 1 : 200 000). Le même processus de calibration a été appliqué au Bénin et en France (1 : 50 000) pour évaluer l'effet de l'échelle de la base de données hydrographiques dans la relation entre  $A_{\min}$  et  $Dd_{\text{obs}}$ .

Ensuite, une deuxième relation reliant les densités de drainage observées et les variables environnementales a été établie sur 86% des mailles de l'ensemble du jeu de donnée constitué du Burkina Faso de Madagascar et du Mali car la relation entre la densité de drainage et les variables environnementales varie fortement dans l'espace (Vogt *et al.*, 2003). Pour tenir compte des relations complexes et non linéaires entre la densité de drainage observée et les variables environnementales, le modèle *Random Forest* a été implémenté comme dans le chapitre III précédent. Les méthodes d'apprentissage automatique ont également été utilisées par Lin *et al.* (2021) pour estimer la densité de drainage aux États-Unis. Cette relation établie entre  $Dd_{\text{obs}}$  et les variables environnementales permet de prédire spatialement la densité de drainage en fonction des variables environnementales. La relation établie est ensuite validée sur 14% des mailles restantes de l'ensemble du jeu de donnée. Une fois validée avec des

performances satisfaisantes, cette relation devra par la suite être combinée à la première relation entre  $A_{\min}$  et  $Dd_{\text{obs}}$  pour déduire spatialement des valeurs de  $A_{\min}$  qui sont nécessaires pour l'extraction des cours d'eau.

Dans la même démarche que ce qui a été fait au **chapitre III**, nous avons généré une matrice de corrélation basée sur le rang de Spearman dans le but d'éliminer les variables environnementales redondantes. Une corrélation  $|r| > 0,75$  est considérée comme une corrélation très forte entre les variables. L'information sur la perméabilité est issue de Huscroft *et al.* (2018) disponible à large échelle au lieu de la perméabilité estimée à partir de la carte géologique du BUMIGEB reclassifiée, car nous n'avons pas d'information à ce niveau de détail hors du Burkina Faso. La matrice de corrélation résultante de la sélection des variables non redondantes est présentée en **Figure IV-3**. La **Figure IV-4** représente la distribution spatiale des variables environnementales retenues à l'échelle des pays investigués.



**Figure IV-3** : Matrice de corrélation de Spearman des variables (9) dont la corrélation est inférieure ou égale à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ). Les noms complets des variables sont définis dans le **Tableau III-2**.

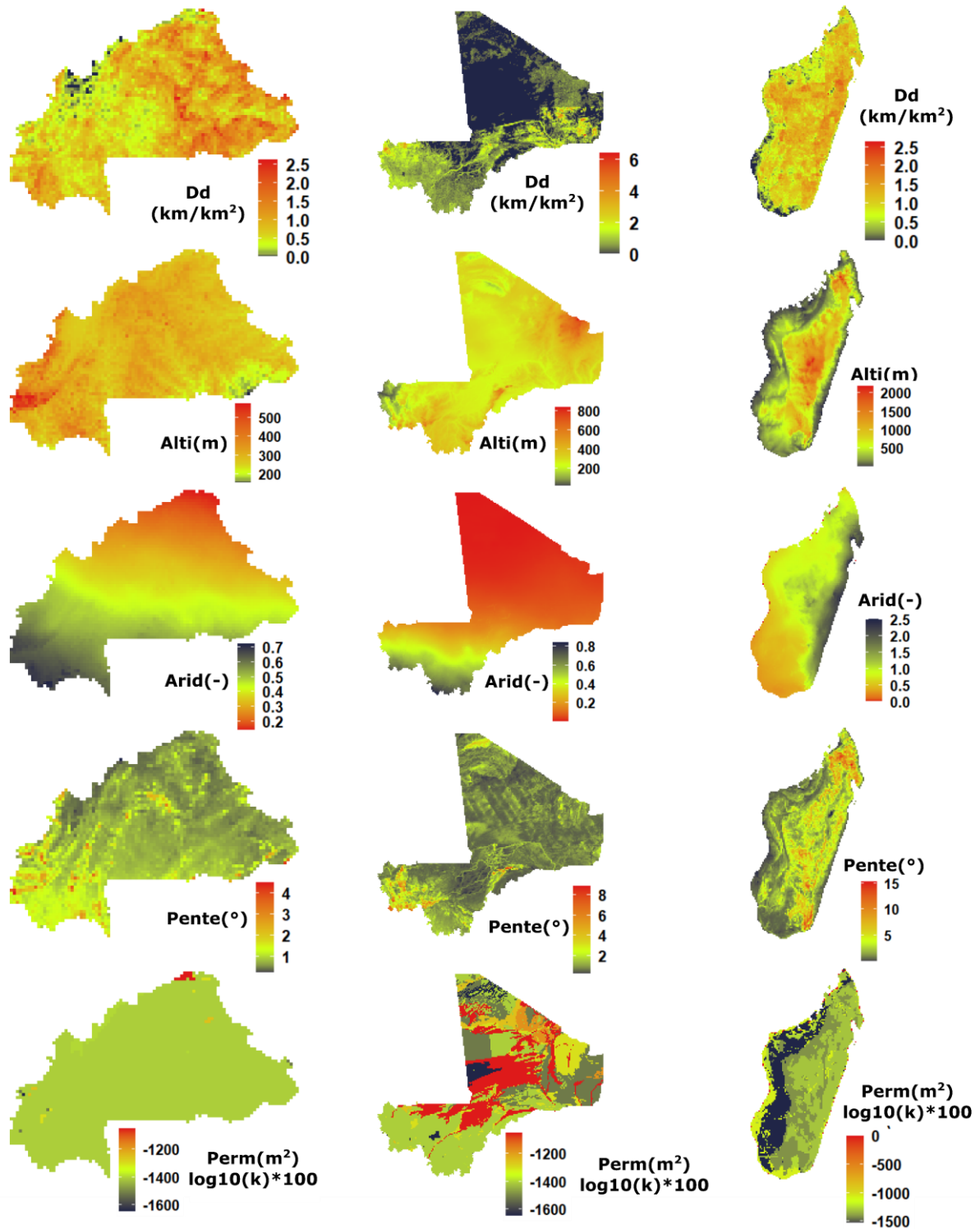


Figure IV-4 : Distribution spatiale (au Burkina Faso, Mali et Madagascar) des variables environnementales non redondantes (Figure IV-3) utilisées pour estimer la densité de drainage. Les noms complets des variables sont définis dans le Tableau III-2.

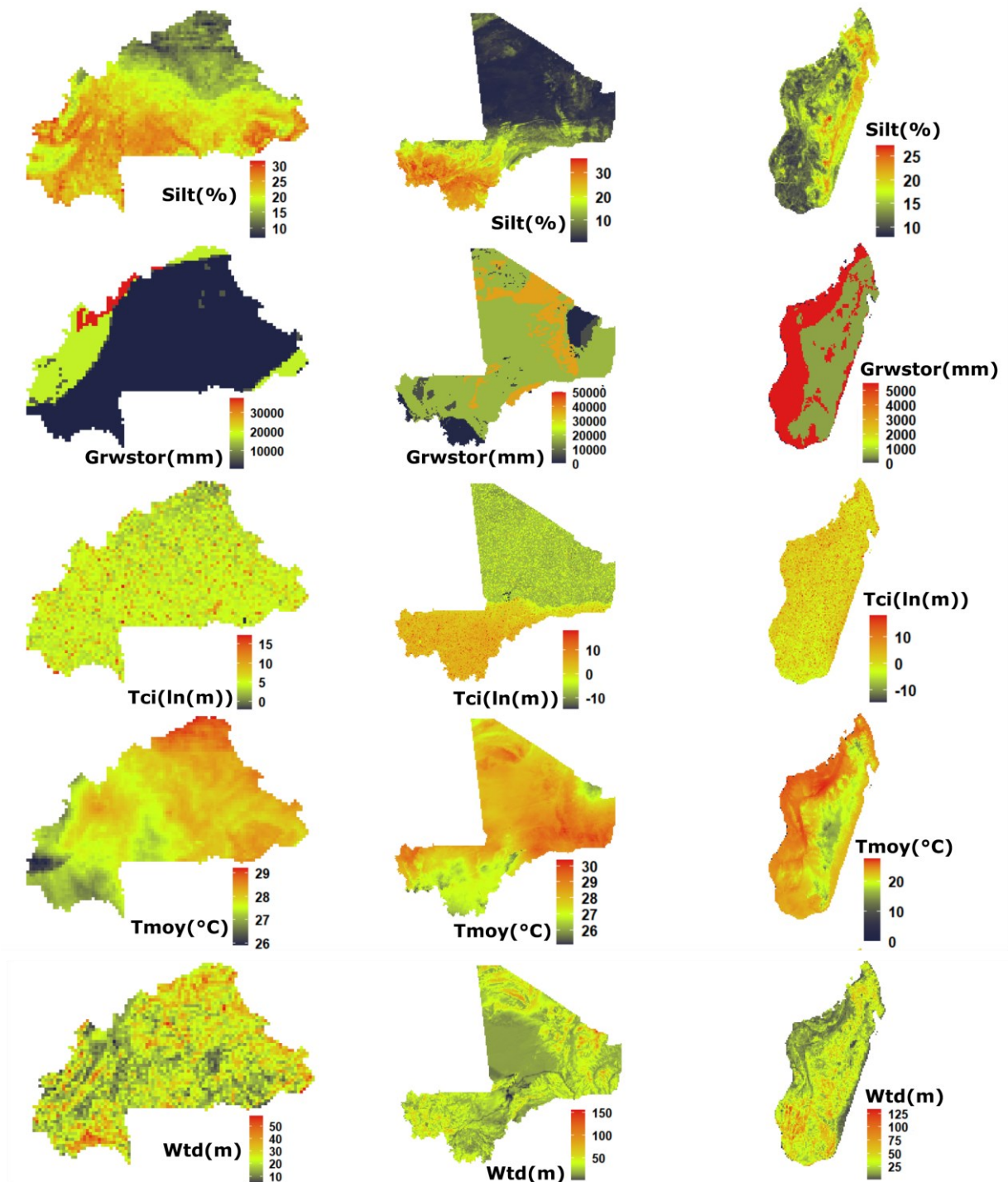


Figure IV-4 : (Suite)

## IV - 4 Résultats et Discussion

Pour la première relation établie entre  $A_{\min}$  et  $D_{\text{obs}}$  au Burkina Faso, les valeurs calibrées d' $A_{\min}$  obtenue varient entre 0,084 et 318 km<sup>2</sup> (Figure IV-5). La valeur minimale obtenue (0,084 km<sup>2</sup>) est beaucoup plus faible que celles utilisées dans les bases de données

continentales (Hydro1k et Aquamaps à des valeurs constantes d' $A_{min}$  de 1000 km<sup>2</sup> et 100 km<sup>2</sup>, respectivement, LCS avec des valeurs variables qui se situent entre 0,3-193 km<sup>2</sup>). Ceci peut refléter la prise en compte de densités de drainage plus élevées au Burkina Faso par rapport aux valeurs obtenues dans les bases de données mentionnées ci-dessus, étant donné que plus  $A_{min}$  devient faible, plus  $Dd_{obs}$  sera élevé.

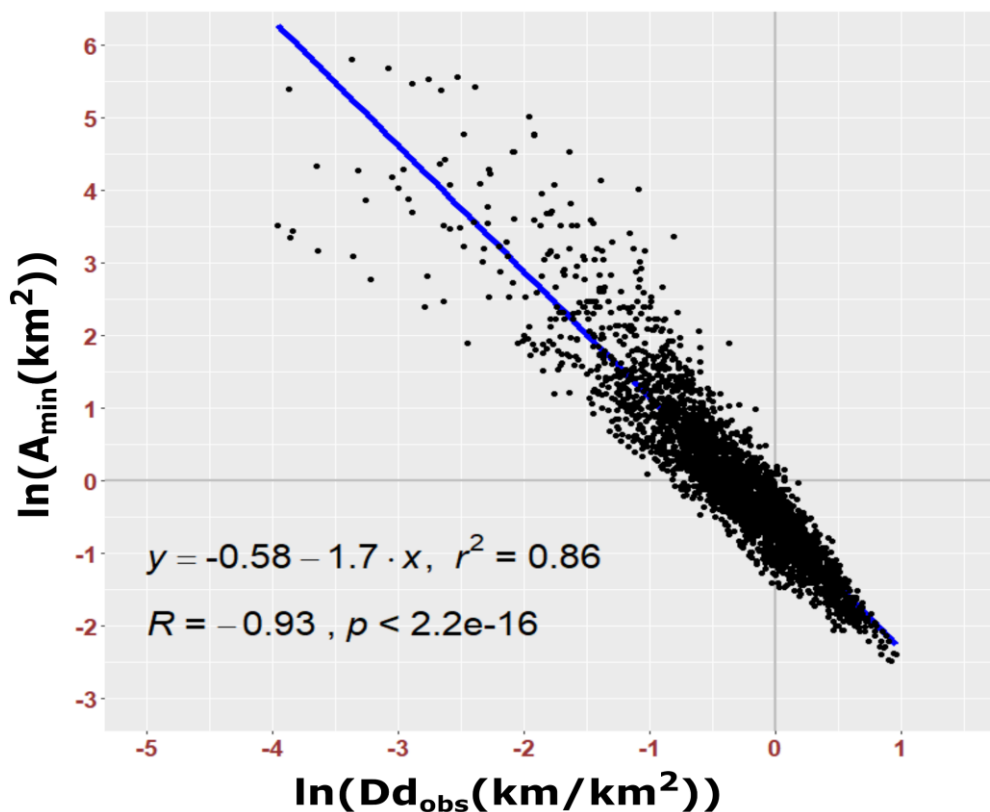


Figure IV-5 : Résultats de la première relation au Burkina Faso, relation entre les valeurs calibrées de  $A_{min}$  et les valeurs de  $Dd_{obs}$  dans les mailles de 5', avec les axes en logarithme.

Les résultats de cette première étape de la mise en œuvre du modèle nous ont permis d'établir une relation (Figure IV-5) avec des performances très satisfaisantes (Coefficient de détermination  $r^2 = 0,86$ , coefficient de corrélation  $R = -0,93$ , p-value  $< 2,2 \times 10^{-16}$ ) :

$$A_{min} = 0,56 * Dd_{obs}^{-1,7} \quad (IV-1)$$

avec  $A_{min}$  (km<sup>2</sup>) et  $Dd_{obs}$  (km/km<sup>2</sup>).

L'équation (IV-1) montre une diminution de la surface minimale contributive ( $A_{min}$ ) en fonction de l'augmentation de la densité de drainage, sous la forme d'une relation puissance comme mentionnée dans certaines études (Colombo *et al.*, 2007 ; Moglen *et al.*, 1998). Elle devrait permettre à partir de la densité de drainage observée de trouver l' $A_{min}$  correspondant, un paramètre nécessaire pour l'extraction des cours d'eau.

La validation de l'équation (IV-1) sur les 14% restants du territoire (Figure IV-2.b) et sur l'ensemble du Mali et Madagascar a permis d'extraire des cours d'eau dont les densités de drainage calculées ( $Dd_{sim}$ ) sont presque similaires à celles observées ( $Dd_{obs}$ ), avec un coefficient de détermination  $r^2$  de 0,86, 0,92 et 0,86 respectivement pour le Burkina Faso, le Mali et Madagascar (Tableau IV-1).

Tableau IV-1 : Performance de la validation de la première étape du modèle au Burkina Faso (14% du territoire), au Mali et à Madagascar.

Paramètres statistiques	Burkina Faso		Mali		Madagascar	
	$Dd_{obs}$	$Dd_{sim}$	$Dd_{obs}$	$Dd_{sim}$	$Dd_{obs}$	$Dd_{sim}$
Moyenne (km/km <sup>2</sup> )	1,01	0,99	0,36	0,45	1	1
Écart-type (km/km <sup>2</sup> )	0,42	0,41	0,52	0,49	0,4	0,35
Coefficient de corrélation (-)	1	0,93	1	0,96	1	0,93
Coefficient de détermination (-)	1	0,86	1	0,92	1	0,86
RMSE (km/km <sup>2</sup> ) (Erreur quadratique moyenne)	0	0,15	0	0,18	0	0,16

Il y a une similitude dans les équations entre  $A_{min}$  en fonction de  $Dd_{obs}$  à différentes échelles et dans différents pays avec des performances  $r^2$  entre 0,82 et 0,88 (Tableau IV-2).

Tableau IV-2 : Relation entre  $A_{min}$  et  $Dd_{obs}$  dans différents pays et à différentes échelles.

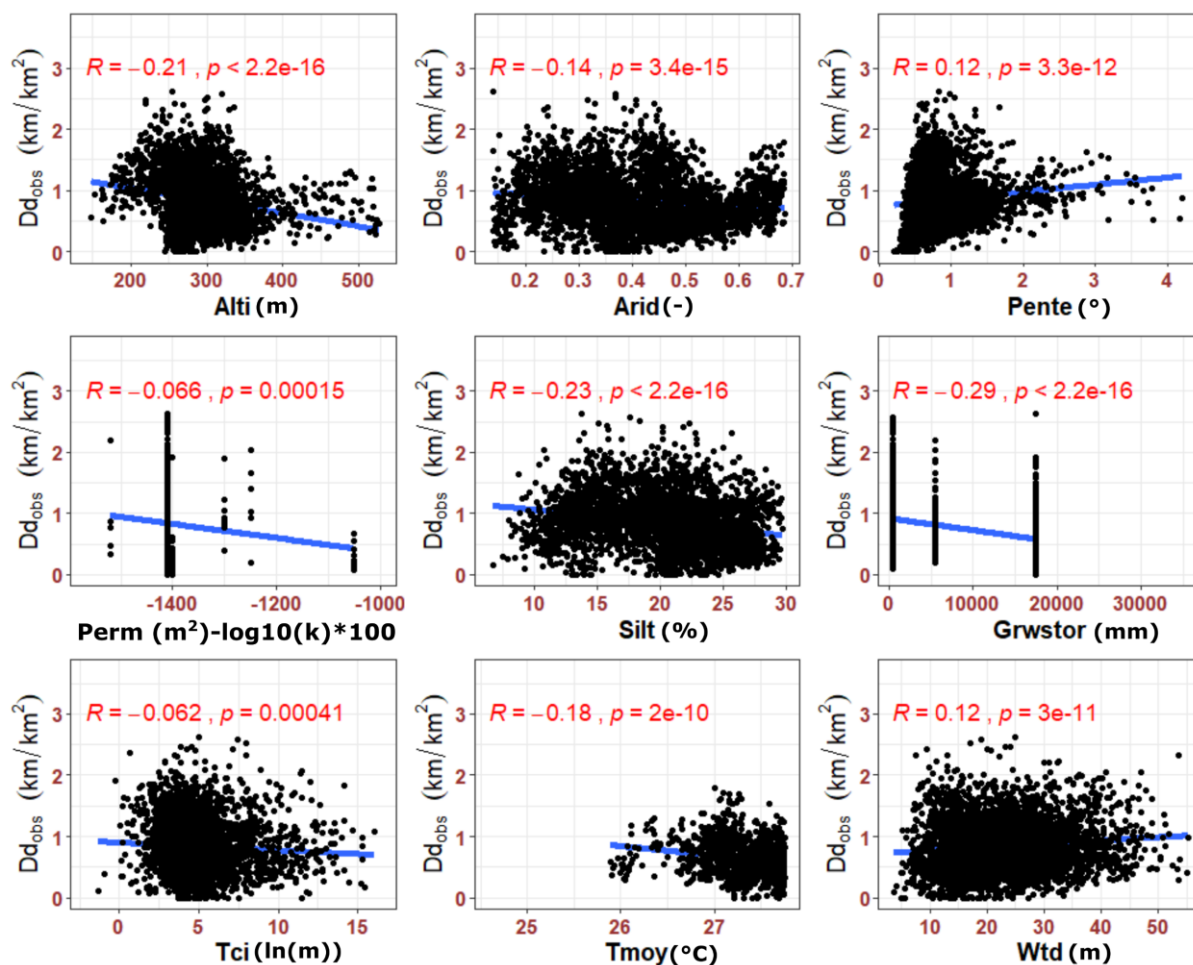
Pays	Équations	$r^2$	R	Échelle
Bénin	$A_{min} = 0,41 * Dd^{-1,8}$	0,88	-0,94	1:50 000
France	$A_{min} = 0,53 * Dd^{-1,7}$	0,82	-0,91	1:50 000
Burkina Faso	$A_{min} = 0,56 * Dd^{-1,7}$	0,86	-0,93	1:200 000

Il est probable que cette relation ne dépende pas de l'échelle de la base de données hydrographiques. La calibration d'autres bases de données hydrographiques à différentes échelles et dans d'autres pays permettrait de confirmer cette hypothèse. La première relation établie au Burkina Faso et validée au Mali et à Madagascar avec de très bonnes performances



permet déjà à partir de densité de drainage observée partout en Afrique d'extraire des cours d'eau à l'échelle de 1 : 200 000.

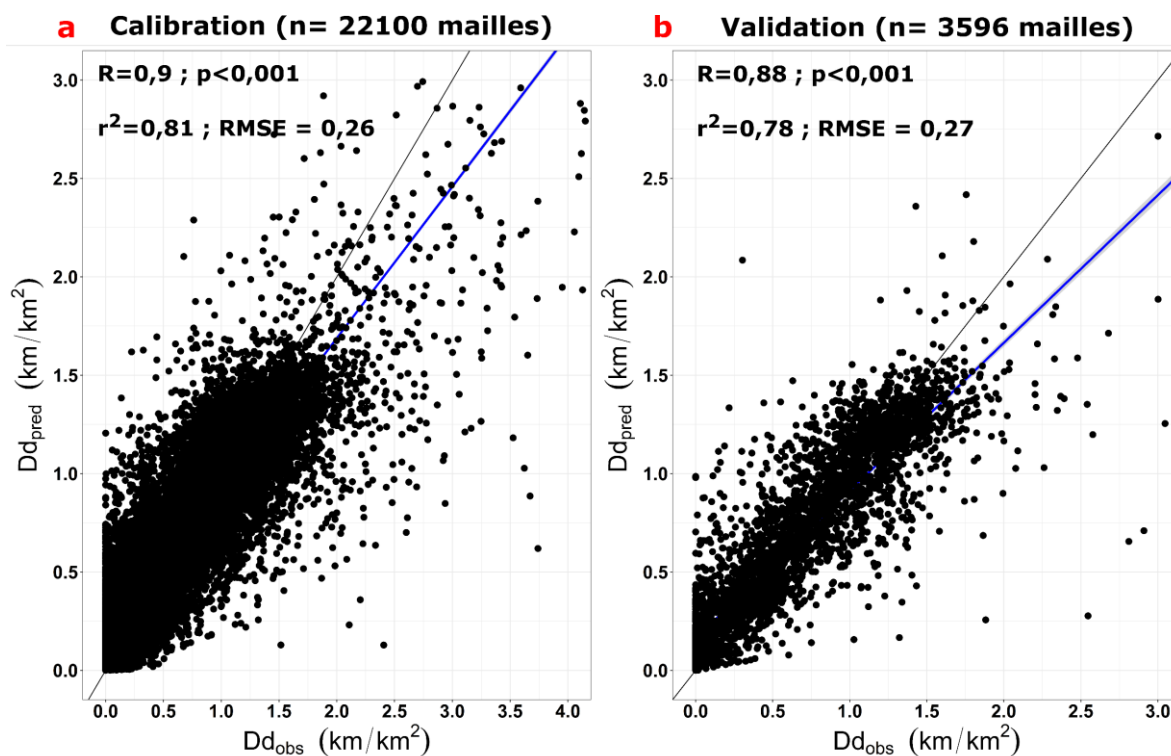
La **Figure IV-6** représente au Burkina Faso, les relations linéaires entre les variables environnementales investiguées et la densité de drainage observée dans les mailles de 5'. En général, la **Figure IV-6** nous montre des relations non linéaires entre la densité de drainage observée et les variables environnementales avec des coefficients de corrélation faible. Dans ce sens, l'étude de Luo *et al.* (2016) mentionne que les variables environnementales sont censées interagir de manière non linéaire pour influencer la valeur de la densité de drainage.



**Figure IV-6** : Régression linéaire entre la densité de drainage observée ( $Dd_{obs}$ ) et les variables environnementales. Pour la signification des codes des variables environnementales, voir le Tableau III-2.

Pour tenir compte des relations non linéaires entre la densité de drainage et les variables environnementales, l'approche *Random Forest* en régression a été utilisée pour relier les densités de drainage observées aux valeurs moyennes des variables environnementales dans les mailles de 5'. Les relations entre la densité de drainage et les variables environnementales varient en fonction du domaine géographique ce qui justifie la prise en compte de l'ensemble

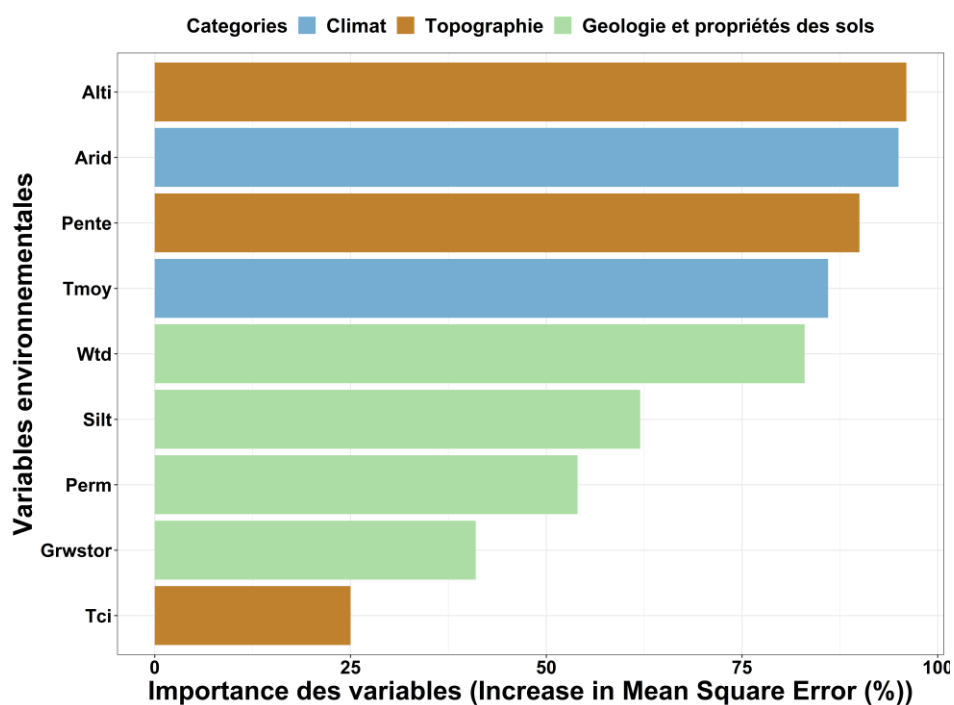
du jeu de donnée constitué du Burkina Faso du Mali et de Madagascar dans la mise en place de cette deuxième relation. Le nombre d'arbres à entraîner (*ntree*) a été fixé 800 comme au chapitre précédent et la valeur calibrée pour le paramètre *mtry* est de 6. Le modèle a montré de bonnes performances  $r^2 = 0,78$  et une corrélation de 0,88 en validation. La **Figure IV-7** résume les critères de performances obtenus en calibration et validation du modèle.



**Figure IV-7 :** Critères de performance en **a)** calibration et **b)** validation du modèle sur l'ensemble du jeu de données (Burkina Faso, Mali et Madagascar) pour l'estimation de la densité de drainage.

Les variables les plus importantes dans le modèle sont l'altitude, l'indice d'aridité et la pente (**Figure IV-8**). Le **Tableau IV-3** synthétise des travaux antérieurs un peu partout dans le monde qui ont évalué les facteurs de contrôle de la densité de drainage. Dans ce tableau, parmi les facteurs de contrôles, l'altitude, la pente et les caractéristiques hydrogéologiques (lithologie, perméabilité) ressortent principalement comme contrôle de la densité de drainage.

Bien que les cartes plus précises de lithologie ou de perméabilité estimées au Burkina Faso ne soient pas prises en compte dans la calibration du modèle à cause de problèmes d'extrapolation, elles ont déjà montré une bonne corrélation avec la densité de densité de drainage, les lithologies les plus perméables montrant des densités de drainage plus faibles et inversement (**Figure II-A-5**). La perméabilité de Huscroft *et al.* (2018) prise en compte dans cette étude et dont l'échelle est grossière pourrait expliquer la moindre influence de cette variable dans le contrôle de la densité de drainage.

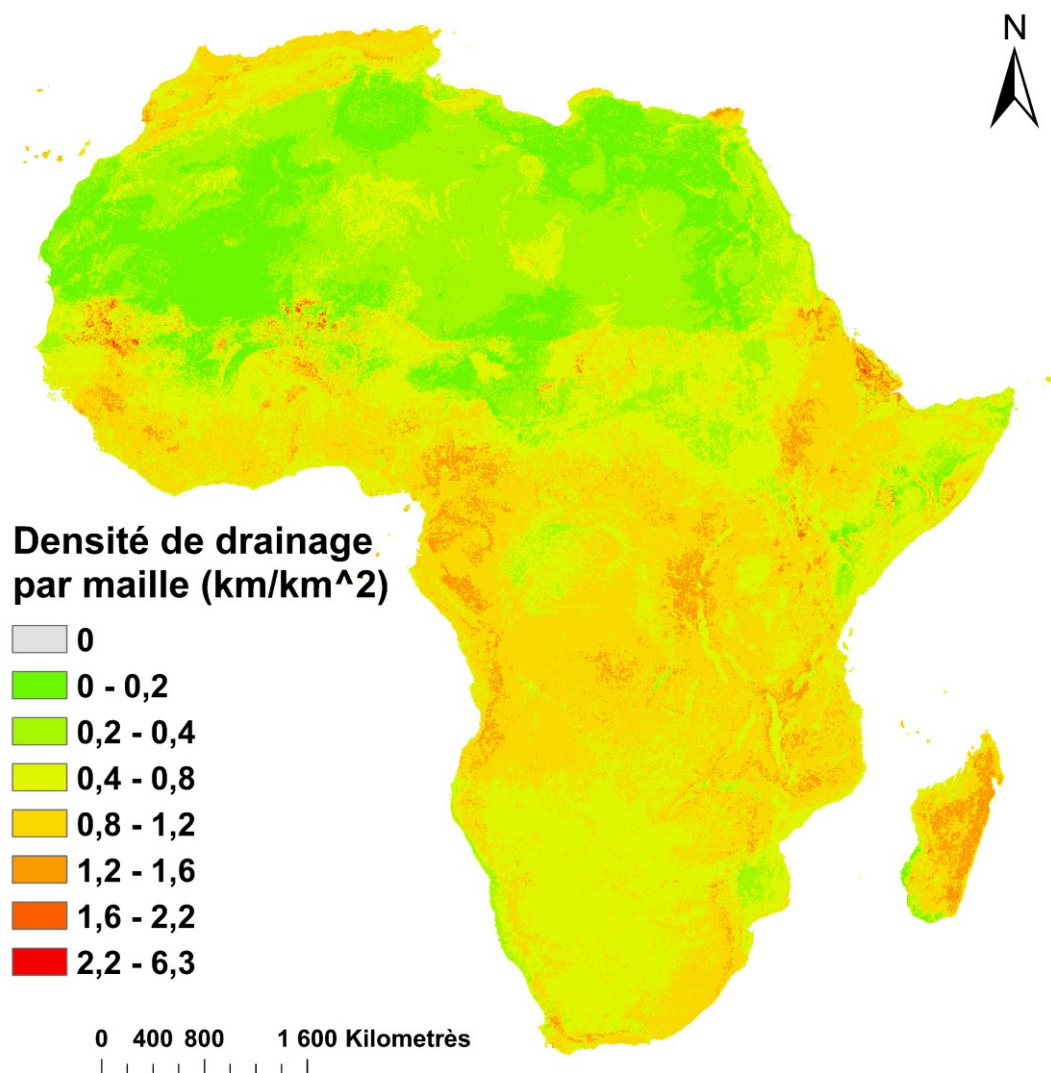


**Figure IV-8 :** Importance des variables dans le modèle Random Forest appliqué en régression pour estimer la densité de drainage. L'importance est appréciée par l'augmentation de l'erreur quadratique moyenne (MSE) si la variable est supprimée du modèle. Les noms complets des variables sont définis dans le **Tableau III-2**.

**Tableau IV-3 :** Facteurs de contrôles de la densité de drainage identifiés dans quelques études à travers le monde. Les études sont classées par date de publication.

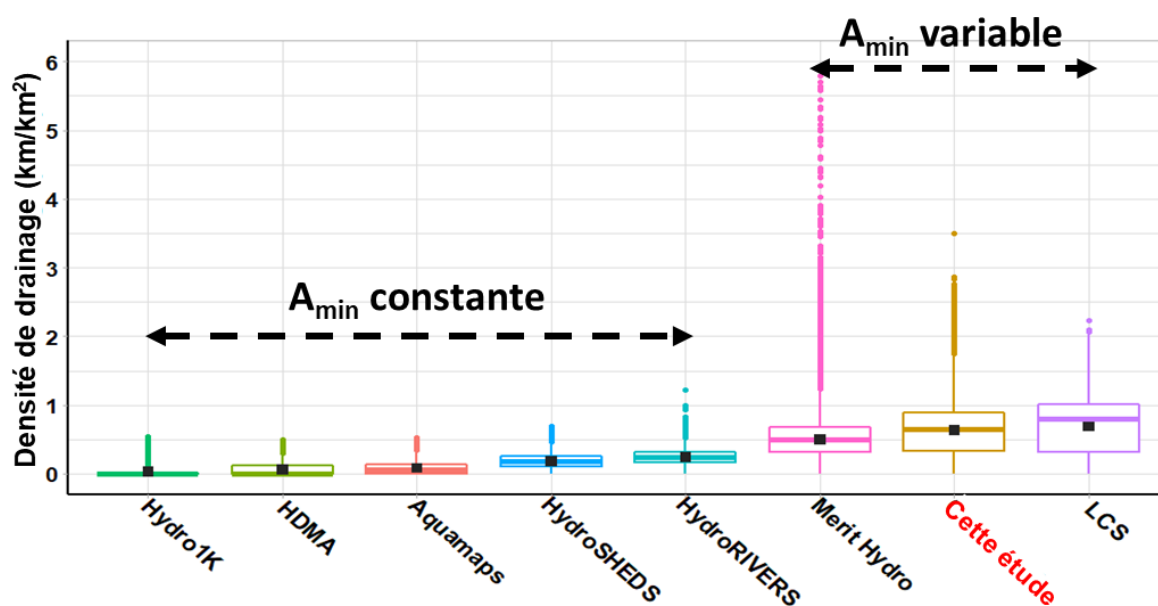
Études	Zones d'études	Facteurs de contrôle identifiés comme les plus importants de la densité de drainage
(Madduma Bandara, 1974)	Sri Lanka	Thornthwaite précipitation index et végétation
(Day, 1980)	Armidable, New england (Australie)	Lithologie
(Oguchi, 1997)	Matsumoto (Japon)	Altitudes et lithologie
(Tucker <i>et al.</i> , 2001)	Nord Apenny (Italy)	Lithologie et altitude
(Nag et Chakraborty, 2003)	West Benghal (Inde)	Lithologie, végétation et altitude
(Strobl et Forte, 2007)	Pennsylvanie (USA)	Perméabilité, texture du sol
	Sud de l'Italie- Région de Puglia (péninsule salentine)	Altitudes et pente
(Luo <i>et al.</i> , 2007)	Cascade range, Oregon (USA)	Âge géologique des formations
(Sangireddy <i>et al.</i> , 2016)	États-Unis	Précipitation, couverture végétale, altitude
(Schneider <i>et al.</i> , 2017)	Globale	Lithologie, précipitation et pente
(Lin <i>et al.</i> , 2021)	Globale	Altitude, profondeur de la nappe, débit moyen annuel

Après avoir mis en place et validé la deuxième relation (densité de drainage en fonction de variables environnementales) sur une partie de l'ensemble du jeu de données (Burkina Faso, Mali et Madagascar) avec des performances satisfaisantes (**Figure IV-7**), le modèle calibré est appliqué en extrapolation à partir des valeurs de variables environnementales (9) réparties spatialement à l'échelle de l'Afrique afin de prédire la densité de drainage. La première relation (équation **(IV-1)**) entre les aires contributives minimales ( $A_{\min}$ ) et la densité de drainage ( $D_{\text{obs}}$ ) qui montre une certaine universalité est ensuite mise à profit pour déduire spatialement les valeurs de  $A_{\min}$  à l'échelle de l'Afrique. Ces  $A_{\min}$  ont finalement servi à extraire des cours d'eau dont les densités de drainage correspondantes sont synthétisées dans les mailles de 5' (**Figure IV-9**).



**Figure IV-9 :** Densités de drainage correspondantes aux cours d'eau extraits à partir d'aires contributives variables en Afrique.

Sur la **Figure IV-9**, on constate une variabilité spatiale de la densité de drainage avec les faibles densités de drainage principalement localisées dans les zones désertiques notamment le Sahara et une grande partie des fortes densités de drainage qui se localise dans la zone équatoriale de l'Afrique où l'indice d'aridité (P/ETP) est fort donc les précipitations sont plus abondantes. Comparativement aux études globales qui représentent les cours d'eau en Afrique (**Figure I-2**), la **Figure IV-10** nous permet de comparer la distribution de leur densité drainage. On voit que pour les cours d'eau qui sont extraits avec des  $A_{min}$  constante on note une faible variabilité de la densité de drainage tandis que ceux qui sont extraits avec  $A_{min}$  variables présentent une plus grande variabilité de la densité de drainage. Les densités de drainage des cours d'eau extraits dans cette étude sont en moyenne supérieure à ceux du réseau Merit Hydro et inférieures à ceux du réseau LCS. Par ailleurs, ces réseaux LCS et Merit Hydro ne représentent pas très bien les cours d'eau au Burkina Faso, Mali et Madagascar par rapport aux réseaux hydrographiques de ces pays (**Tableau I-4**). La collecte de nouvelles données hydrographiques de référence à l'échelle de 1 : 200 000 dans quelques pays en Afrique devrait permettre une validation supplémentaire de la carte de densité de drainage calculée dans cette étude.



**Figure IV-10** : Distribution des densités de drainage ( $\text{km}/\text{km}^2$ ) des principales bases de données qui représentent les cours d'eau en Afrique. Pour une description détaillée de ces bases voir **Tableau I-1** et **Figure I-1**.

## IV - 5 Conclusion

Cette étude propose une méthode simple pour obtenir des surfaces contributives minimales ( $A_{\min}$ ) en fonction de variables environnementales (perméabilité, précipitation, profondeur de la nappe, altitude, pente). La première étape de la mise en place du modèle a conduit à une relation très performante entre  $A_{\min}$  et  $Dd_{\text{obs}}$  au Burkina Faso. Cette première relation établie également dans différents pays (Bénin, France) avec des échelles de réseaux hydrographiques différents a montré une certaine invariance d'échelle. La validation de cette relation sur une partie du Burkina Faso, et dans d'autres pays (Mali, Madagascar) ayant des caractéristiques environnementales différentes du Burkina Faso a montré des performances très satisfaisantes. Cette première relation permet d'estimer localement le meilleur  $A_{\min}$  à partir des densités de drainage observées sans avoir à extraire 400 réseaux hydrographiques différents par exemple. Cependant, l'objectif final était de caractériser spatialement l' $A_{\min}$  dans les zones dépourvues d'informations hydrographiques, où les densités de drainage observées ne sont donc pas disponibles. À cette fin, nous avons établi une deuxième relation avec des performances satisfaisantes, reliant les densités de drainage observées et les variables environnementales sur l'ensemble des pays investigués (Burkina Faso, Mali et Madagascar). La validation de cette deuxième relation sur une partie du jeu de données (14%) a montré également de bonnes performances. La combinaison de ces deux relations nous a permis d'extraire des cours d'eau à l'échelle de l'Afrique dont les densités de drainage correspondent à ceux observées dans de nombreux pays en Afrique. Nous prévoyons que ce nouveau réseau hydrographique pourra être utilisé en entrée des modèles globaux et régionaux pour une meilleure prise en compte de l'information hydrographique.

# CONCLUSIONS GENERALES ET PERSPECTIVES

## Principales contributions

La caractérisation de l'intermittence du débit des cours d'eau à large échelle et surtout dans des zones à données limitées comme l'Afrique est un défi et c'est sans doute la raison pour laquelle peu d'études se sont appesanties sur la caractérisation de l'intermittence. L'objectif général de cette thèse était d'apporter une meilleure compréhension sur les facteurs qui contrôlent l'intermittence des cours d'eau à différentes échelles et de proposer une approche méthodologique de caractérisation des cours d'eau, en termes d'intermittence du débit, sur de grands domaines jusqu'à l'échelle du continent, afin de pallier les lacunes d'observations.

Le premier objectif spécifique de cette thèse (**chapitre II**) s'est focalisé au Burkina Faso afin d'identifier le rôle des variables environnementales dans la variation géographique de l'intermittence du débit des cours d'eau. Les variables clés identifiées ont servi à prédire spatialement l'intermittence à l'échelle du Burkina Faso par l'approche *Random Forest*. La période investiguée s'étend de 1955 à 1985 et permet de maximiser la disponibilité des stations de jaugeages. Il ressort de cette étude que :

- Bien que la précipitation moyenne annuelle et l'ordre de Strahler soient importants dans la discrimination des classes d'intermittence, la perméabilité du sol et la surface amont des bassins versants expliquent principalement l'intermittence du débit des cours d'eau au Burkina Faso.
- La répartition géographique des quatre classes d'intermittence au Burkina Faso présente une cohérence spatiale. Les stations les plus intermittentes (fortement intermittentes et éphémères) sont majoritairement localisées dans les bassins versants du Nakanbe et du Niger tandis que les plus permanentes (permanent et faiblement intermittent) sont localisées en partie dans les bassins versants du Mouhoun et de la Comoé. Cette répartition des stations met en évidence le contraste des conditions hydrogéologiques entre les bassins versants du Mouhoun et de la Comoé d'une part, et celles du Nakanbe et du Niger.
- La configuration spatiale de l'intermittence entre la sous période normale (1955-1969) et la sous-période sèche (1970-1985) a montré que les classes d'intermittence sont peu sensibles aux variations importantes des précipitations annuelles moyennes. Cette faible sensibilité des cours d'eau à la précipitation moyenne annuelle est liée à la très forte saisonnalité de la précipitation dans le pays : par défaut, le débit des cours d'eau est également saisonnier, quelles que soient les précipitations moyennes annuelles, à moins que des processus particuliers du bassin versant, ici l'écoulement des eaux souterraines, ne le rendent pérenne. Ceci implique que les évolutions futures de la durée de la saison des pluies et des processus de recharge des eaux souterraines

peuvent être plus importantes pour l'évolution de l'intermittence au Burkina Faso que celles de la précipitation moyenne annuelle.

- 83% de la longueur des cours d'eau au Burkina Faso sont prédits intermittents (faiblement intermittent, fortement intermittent et éphémère) contre 98% dans la base hydrographique de référence. Cette sous-estimation de l'intermittence pourrait en partie s'expliquer par le fait que la calibration du modèle à partir de stations situées sur de grands cours d'eau (ordre de Strahler de 5 à 9) donne des résultats moins bons en extrapolation sur les plus petits cours d'eau (ordre de Strahler 1 à 4) qui constituent la majorité des cours d'eau dans un réseau hydrographique.
- Bien que les classifications de l'intermittence présente dans les réseaux hydrographiques de référence soient une base utile pour la comparaison avec les résultats de prédiction de l'intermittence, l'intermittence observée au droit des stations de jaugeage divergent très souvent avec la classification présente dans les réseaux hydrographiques de référence, le même constat est fait aux États-Unis par Hafen *et al.* (2020).

Dans le second objectif spécifique (**chapitre III**) on s'est intéressé à l'échelle de l'Afrique aux contrôles de l'intermittence sur la période 1958-1991, une période qui maximise la disponibilité des stations de jaugeage. Les contrôles identifiés ont été reliés aux différentes classes d'intermittence observée aux droits des stations de jaugeage dans le but de prédire sur le réseau LCS (Schneider *et al.*, 2017), la distribution spatiale des cours d'eau intermittents. Cette étude a permis d'établir les constats suivants :

- L'indice d'aridité (P/ETP), la surface amont des bassins versants et l'évapotranspiration potentielle moyenne annuelle sont les facteurs de contrôle les plus importants de l'intermittence à l'échelle continentale de l'Afrique. Ces variables de contrôle ressortent également comme les plus importantes dans quelques études continentales et globales axées sur l'intermittence (Hammond *et al.*, 2021 ; Lin *et al.*, 2021 ; Messenger *et al.*, 2021 ; Sauquet *et al.*, 2021 ; Zipper *et al.*, 2021). Les variables climatiques sont donc les déterminants les plus importants de l'intermittence du débit des cours d'eau à l'échelle de l'Afrique.
- Une très forte similarité spatiale (coefficient de corrélation  $r = 0,79$  ;  $r = 0,76$ ) entre la carte d'intermittence de cette étude et respectivement celle des études précédentes de Lin *et al.* (2021) et Messenger *et al.* (2021) a été observée. Cette étude estime que 56% de la longueur des cours d'eau à l'échelle de l'Afrique sont intermittents tandis que 44% sont prédits permanents. Ce résultat sous-estime l'intermittence par rapport aux études précédentes de Lin *et al.* (2021) et Messenger *et al.* (2021) avec respectivement 69% et 71% des cours d'eau prédits intermittents à l'échelle de l'Afrique. En conclusion, les cours d'eau intermittents sont répandus et constituent la majorité des cours d'eau en Afrique selon les résultats du **chapitre III** de cette thèse et les récentes études (Lin *et al.*, 2021 ; Messenger *et al.*, 2021).



- Le modèle M3-LCS développé dans cette étude a tendance à surestimer les cours d'eau permanents et cela s'explique en partie par un pourcentage très élevé de stations permanentes (70,7%) dans l'échantillon. Dans les zones désertiques (Kalahari, Sahara), les prédictions du modèle sous-estiment largement l'intermittence des cours d'eau en classant tous les cours d'eau dans cette zone comme étant fortement intermittentes au lieu de la classe éphémère. Cette erreur révèle le manque de stations de jaugeage de ces zones dans l'échantillon de calibration.
- Les raisons qui expliquent aussi la différence du pourcentage d'intermittence entre les études à l'échelle de l'Afrique sont liées à la densité de drainage du réseau hydrographique considéré ensuite à la définition de l'intermittence qui est souvent différente entre les études.
- De façon générale, le modèle M3-LCS capture la distribution spatiale de l'intermittence dans les bases hydrographiques de référence investiguées (Bénin, Mali, Madagascar, et Afrique du Sud) à l'exception du Burkina Faso où cette ressemblance spatiale est moins forte en raison de quelques cours d'eau mal classés dans le réseau de référence.
- En général, l'intermittence est sous-estimée dans les bases hydrographiques de référence où les grands cours d'eau ont tendance à être systématiquement classés permanents.

La collecte d'un large jeu de données hydrographiques à l'échelle continentale de l'Afrique et à l'échelle de plusieurs pays (Bénin, Burkina Faso, Mali, Madagascar et Afrique du Sud) a permis de mettre en évidence que les bases de données hydrographiques continentales ne reproduisent pas correctement la variabilité spatiale de la densité de drainage observée sur les réseaux hydrographiques de référence des pays investigués (**chapitre I**). Cela est dû en partie à une extraction des cours d'eau par MNT avec une aire contributive constante spatialement ce qui tient compte uniquement de l'influence de la topographie dans l'initiation des cours d'eau négligeant par l'occasion l'influence d'autres variables telles que la précipitation et la géologie.

Le troisième objectif spécifique nous a amené à développer (**chapitre IV**) en deux étapes une méthode simple pour extraire des cours d'eau qui représentent au mieux la variabilité spatiale de la densité de drainage observée en Afrique. Des aires contributives variables spatialement ont pu être calibrées en fonction de variables environnementales. Cette étude montre que les variables topographiques et climatiques sont les contrôles principaux de la densité de drainage dans les pays investigués (Burkina Faso, Mali et Madagascar). Les densités de drainage correspondants aux cours d'eau extraits à partir de la méthodologie développée dans ce chapitre montrent une variabilité spatiale en Afrique avec les zones désertiques comportant les faibles densités de drainage. Les analyses menées dans ce chapitre ont permis d'améliorer

la représentation des cours d'eau à l'échelle de l'Afrique en tenant compte de leur densité de drainage variable dans l'espace.

### **Perspectives et recommandations**

Les méthodes proposées et les modèles développés dans cette thèse fournissent des informations précieuses aussi bien sur la dynamique spatiale de l'intermittence des cours d'eau que sur la variabilité spatiale de la densité de drainage. Cependant, plusieurs défis demeurent et fournissent en même temps, des pistes pour de futures améliorations et orientations de recherche à plusieurs niveaux :

- L'étude menée au Burkina Faso au **chapitre II** s'est appuyée sur des stations de jaugeages situées sur de grands cours d'eau et qui, depuis les années 1970, n'ont fait que diminuer. Les stations prises en compte dans cette étude ne sont pas représentatives des mécanismes qui se produisent sur les très petits cours d'eau.
- Les données d'occupation du sol n'ont pas pu être considérées comme des variables explicatives potentielles de l'intermittence au Burkina Faso en raison de nombreux problèmes concernant la qualité et la couverture spatiale et temporelle des ensembles de données que nous avons pu obtenir pour la période d'étude qui s'étend de 1955 à 1985.
- Dans l'échantillon de stations de jaugeage considérées à l'échelle de l'Afrique, un déséquilibre existe entre les classes d'intermittence avec notamment les stations permanentes qui sont prépondérantes dans l'échantillon en plus de nombreuses stations permanentes situées sur des petits cours d'eau qui se retrouvent en Afrique du Sud. Malgré l'implémentation dans cette étude d'une technique statistique de gestion des données déséquilibrées, ce déséquilibre peut avoir favorisé la prédiction de cours d'eau permanents à certains endroits.
- Pour améliorer les performances de prédiction du modèle en Afrique notamment dans les zones désertiques où il n'y a pas de stations de jaugeage, il serait envisageable de placer dans ces zones, des stations fictives définies « éphémères » ou d'exploiter des stations hydrométriques disponibles dans d'autres continents bien jaugés.
- Pour limiter l'effet du déséquilibre des classes dans l'échantillon, il serait possible d'écarter les nombreuses stations permanentes situées sur des petits cours d'eau et majoritairement localisées en Afrique du Sud dans le but d'éviter un biais d'échantillonnage.
- En matière d'outils à explorer, les réseaux de neurones (ANN) qui sont également basées sur une méthode d'intelligence artificielle pourraient être moins sensibles au cas des données déséquilibrées et présenter des performances de prédiction plus satisfaisantes que Random Forest (Beaufort *et al.*, 2019).

- Le changement climatique devrait entraîner une augmentation globale de la température de l'air. L'une des principales conséquences de cette augmentation est le risque d'une augmentation de l'aridité. L'indice d'aridité est le principal facteur de contrôle de l'intermittence identifié dans cette thèse à l'échelle de l'Afrique. Les projections climatiques futures pourront être prises en compte dans le modèle *Random Forest* développé dans cette thèse afin de prédire l'évolution des classes d'intermittence. Les résultats issus de cette perspective devraient permettre d'identifier des zones en Afrique où la durée de l'écoulement des cours d'eau devrait diminuer ou augmenter.
- La caractérisation de l'intermittence dans cette thèse s'est plutôt basée sur les périodes historiques les années 1950 à 1990, car elle permettait d'avoir un maximum de stations disponibles. Bien que le nombre de stations de jaugeage ait diminué depuis les années 1970, il serait intéressant d'étudier sur les stations ayant de longues chroniques, la tendance d'évolution de l'intermittence entre les périodes historiques et celles récentes comme étudiée aux États-Unis par Zipper *et al.* (2021).
- La collecte d'un nouveau jeu de données hydrographiques de référence à l'échelle de 1 : 200 000 dans quelques pays d'Afrique devrait permettre une validation supplémentaire des densités de drainage calculées au **chapitre IV**.
- Pour mieux considérer les petits cours d'eau à l'amont du réseau hydrographique, l'implantation d'enregistreurs de données avec capteurs par les services hydrologiques des différents pays en Afrique devrait permettre de mesurer l'absence ou la présence d'eau en continu et ils sont moins coûteux que les stations de jaugeage traditionnelles. Cela devrait contribuer à améliorer notre compréhension sur les mécanismes de contrôle de l'intermittence ainsi que notre capacité de prédiction de ces cours d'eau. Cependant, la tendance actuelle dans de nombreux pays africains comme le Burkina Faso est à l'optimisation du réseau hydrométrique, basée sur les moyens financiers disponibles, pour ne garder qu'un nombre limité de stations de jaugeage sur des cours d'eau d'intérêt.
- La mise à jour et à disposition de carte exploitable sur l'occupation du sol en Afrique de l'Ouest pour les années 1970 permettrait de mieux considérer l'influence de l'occupation du sol dans le contrôle de l'intermittence des cours d'eau au Burkina Faso. Dans ce sens, des initiatives sont en cours au sein de l'institut de recherche pour le développement (IRD) pour numériser les anciennes cartes d'occupations du sol dans de nombreux pays en Afrique de l'Ouest.
- Des données spatiales à fine échelle correspondant à la perméabilité des bassins et à d'autres facteurs hydrogéologiques pourraient améliorer notre capacité à modéliser les cours d'eau intermittents avec précision compte tenu de l'échelle grossière des cartes continentales utilisées. Des efforts sont faits dans ce sens avec les nouvelles cartes

hydrogéologiques<sup>37</sup> de l'Afrique de l'Ouest que propose le BRGM (Bureau de recherches géologiques et minières de France) et ses partenaires, mais l'échelle des cartes reste grossière 1 : 5 000 000 et ces cartes ne sont pas encore disponible sous format numérique.

Les cours d'eau intermittents constituent un refuge pour de nombreuses espèces aquatiques et la caractérisation spatiale de l'intermittence permet d'avoir des informations précieuses qui facilitent la mobilisation et la gestion de la ressource en eau de surface ainsi que la conservation de la biodiversité (Datry *et al.*, 2017 ; Leigh *et al.*, 2016). À cette fin, les travaux menés dans cette thèse ont fait progresser notre compréhension scientifique sur les facteurs de contrôle de l'intermittence de l'échelle régionale à l'échelle continentale de l'Afrique ainsi que sur la dynamique spatiale de ces systèmes. Les facteurs qui contrôlent l'intermittence des cours d'eau dépendent de l'échelle et de la région considérée. Des approches innovantes ont été développées pour caractériser l'intermittence des cours d'eau au Burkina Faso et en Afrique respectivement dans le **chapitre II** et le **chapitre III** et nous avons exploré une nouvelle approche pour extraire des cours d'eau plus réalistes en Afrique en **chapitre IV**. La recherche menée dans cette thèse contribuera certainement à améliorer la gestion des ressources en eau en Afrique.

---

<sup>37</sup> <https://www.brgm.fr/fr/actualite/actualite/nouvelle-carte-hydrogeologique-afrique-ouest-cedeao>

## REFERENCES BIBLIOGRAPHIQUES

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. et Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191. <https://doi.org/10.1038/sdata.2017.191>
- Acuña, V., Datry, T., Marshall, J., Barcelo, D., Dahm, C. N., Ginebreda, A., McGregor, G., Sabater, S., Tockner, K. et Palmer, M. A. (2014). Why Should We Care About Temporary Waterways? *Science*, 343(6175), 1080-1081. <https://doi.org/10.1126/science.1246666>
- Acuña, V., Hunter, M. et Ruhí, A. (2017). Managing temporary streams and rivers as unique rather than second-class ecosystems. *Biological Conservation*, 211, 12-19.
- Addinsoft, S. X. V. (2015). Data Analysis and Statistics Software for Microsoft Excel. *Addinsoft: Paris, France*.
- Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. et Clark, M. P. (2018). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, 54(11), 8792-8812. <https://doi.org/10.1029/2018WR022606>
- Alcázar, J. et Palau, A. (2010). Establishing environmental flow regimes in a Mediterranean watershed based on a regional classification. *Journal of Hydrology*, 388(1-2), 41-51. <https://doi.org/10.1016/j.jhydrol.2010.04.026>
- Ali, M., Son, D.-H., Kang, S.-H. et Nam, S.-R. (2017). An Accurate CT Saturation Classification Using a Deep Learning Approach Based on Unsupervised Feature Extraction and Supervised Fine-Tuning Strategy. *Energies*, 10(11). <https://doi.org/10.3390/en10111830>
- Allen, R. G., Pereira, L. S., Raes, D. et Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9), D05109.
- Al-Muqdad, S. W. et Merkel, B. J. (2011). Automated watershed evaluation of flat terrain. *Journal of Water Resource and Protection*, 3(12), 892. <https://doi.org/10.4236/jwarp.2011.312099>
- Álvarez-Cabria, M., Barquín, J. et Peñas, F. J. (2016). Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors. *Science of The Total Environment*, 545-546, 152-162. <https://doi.org/10.1016/j.scitotenv.2015.12.109>
- Amaratunga, D., Cabrera, J. et Lee, Y.-S. (2008). Enriched random forests. *Bioinformatics*, 24(18), 2010-2014. <https://doi.org/10.1093/bioinformatics/btn356>
- Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M., Shen, L. et Domisch, S. (2022). Hydrography90m: A new high-resolution global hydrographic dataset. *Earth System Science Data*, 2022, 1-43. <https://doi.org/10.5194/essd-2022-9>

- Apley, D. W. et Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4), 1059-1086. <https://doi.org/10.1111/rssb.12377>
- Askew, A. J. (1970). Variation in lag time for natural catchments. *Journal of the Hydraulics Division*, 96(2), 317-330.
- Assani, A. A., Tardif, S. et Lajoie, F. (2006). Statistical analysis of factors affecting the spatial variability of annual minimum flow characteristics in a cold temperate continental region (southern Québec, Canada). *Journal of Hydrology*, 328(3-4), 753-763. <https://doi.org/10.1016/j.jhydrol.2006.01.015>
- Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H. et Schipper, A. M. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Scientific Data*, 5(1). <https://doi.org/10.1038/sdata.2018.52>
- Barbé, L. L. et Lebel, T. (1997). Rainfall climatology of the HAPEX-Sahel region during the years 1950–1990. *Journal of Hydrology*, 188-189, 43-73. [https://doi.org/10.1016/S0022-1694\(96\)03154-X](https://doi.org/10.1016/S0022-1694(96)03154-X)
- Beaufort, A., Carreau, J. et Sauquet, E. (2019). A classification approach to reconstruct local daily drying dynamics at headwater streams. *Hydrological Processes*, 33(13), 1896-1912. <https://doi.org/10.1002/hyp.13445>
- Beaufort, A., Lamouroux, N., Pella, H., Datry, T. et Sauquet, E. (2018). Extrapolating regional probability of drying of headwater streams using discrete observations and gauging networks. *Hydrology and Earth System Sciences*, 22(5), 3033-3051. <https://doi.org/10.5194/hess-22-3033-2018>
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A. et Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1), 180214. <https://doi.org/10.1038/sdata.2018.214>
- Beck, M., Mazor, R. D., Stein, E. D., Maas, R., De Mello, D. et Bram, D. (2017). Mapping of non-perennial and ephemeral streams in the Santa Ana Region. *California Coastal Water Research Project. Retrieved August, 7, 2020.*
- Bejarano, M. D., Marchamalo, M., Jalón, D. G. de et Tánago, M. G. del. (2010). Flow regime patterns and their controlling factors in the Ebro basin (Spain). *Journal of Hydrology*, 385(1-4), 323-335. <https://doi.org/10.1016/j.jhydrol.2010.03.001>
- Belemtougri, A. P., Ducharne, A. et Karambiri, H. (2021a). Improvement of river network representation in Africa: application of an approach based on digital elevation data and environmental characteristics. *Proceedings of the International Association of Hydrological Sciences*, 384, 19-23. <https://doi.org/10.5194/piahs-384-19-2021>
- Belemtougri, A. P., Ducharne, A., Tazen, F., Oudin, L. et Karambiri, H. (2021b). Understanding key factors controlling the duration of river flow intermittency: Case of Burkina Faso in West Africa. *Journal of Hydrology: Regional Studies*, 37, 100908. <https://doi.org/10.1016/j.ejrh.2021.100908>

- Belgiu, M. et Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Benejam, L., Angermeier, P. L., Munne, A. et Garcia-Berthou, E. (2010). Assessing effects of water abstraction on fish assemblages in Mediterranean streams. *Freshwater Biology*, 55(3), 628-642. <https://doi.org/10.1111/j.1365-2427.2009.02299.x>
- Berhanu, B., Seleshi, Y., Demisse, S. et Melesse, A. (2015). Flow Regime Classification and Hydrological Characterization: A Case Study of Ethiopian Rivers. *Water*, 7(12), 3149-3165. <https://doi.org/10.3390/w7063149>
- Bertoldi, W., Piégay, H., Buffin-Bélanger, T., Graham, D. et Rice, S. (2012). Applications of close-range imagery in river research. *Fluvial remote sensing for science and management*, 341-366. <https://doi.org/10.1002/9781119940791.ch15>
- Bethemont, J. (1995). Les fleuves allogènes : espoirs et réalités. *Géocarrefour : Revue de géographie de Lyon*, 70(3), 275-278.
- Bhamjee, R., Lindsay, J. B. et Cockburn, J. (2016). Monitoring ephemeral headwater streams: a paired-sensor approach. *Hydrological Processes*, 30(6), 888-898. <https://doi.org/10.1002/hyp.10677>
- Biasutti, M. (2019). Rainfall trends in the African Sahel: Characteristics, processes, and causes. *Wiley Interdisciplinary Reviews: Climate Change*, 10(4), e591. <https://doi.org/10.1002/wcc.591>
- Biau, G. et Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197-227. <https://doi.org/10.1007/s11749-016-0481-7>
- Bird, S., Hogan, D. et Schwab, J. (2010). Photogrammetric monitoring of small streams under a riparian forest canopy. *Earth Surface Processes and Landforms*, 35(8), 952-970. <https://doi.org/10.1002/esp.2001>
- Bischel, H. N., Lawrence, J. E., Halaburka, B. J., Plumlee, M. H., Bawazir, A. S., King, J. P., McCray, J. E., Resh, V. H. et Luthy, R. G. (2013). Renewing urban streams with recycled water for streamflow augmentation: hydrologic, water quality, and ecosystem services management. *Environmental Engineering Science*, 30(8), 455-479. <https://doi.org/10.1089/ees.2012.0201>
- Borg Galea, A., Sadler, J. P., Hannah, D. M., Datry, T. et Dugdale, S. J. (2019). Mediterranean intermittent rivers and ephemeral streams: Challenges in monitoring complexity. *Ecohydrology*, 12(8). <https://doi.org/10.1002/eco.2149>
- Börker, J., Hartmann, J., Amann, T. et Romero-Mujalli, G. (2018). Terrestrial sediments of the earth: development of a global unconsolidated sediments map database (GUM). *Geochemistry, Geophysics, Geosystems*, 19(4), 997-1024. <https://doi.org/10.1002/2017GC007273>
- Boschetto, R. G., Mohamed, R. M. et Arrigotti, J. (2010). Vulnerability to desertification in a sub-Saharan region: a first local assessment in five villages of southern region of

- Malawi. *Italian Journal of Agronomy*, 5(s3), 91-102. <https://doi.org/10.4081/ija.2010.s3.91>
- Boulesteix, A., Janitza, S., Kruppa, J. et König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), 493-507. <https://doi.org/10.1002/widm.1072>
- Boulton, A. J., Rolls, R. J., Jaeger, K. L. et Datry, T. (2017). Hydrological Connectivity in Intermittent Rivers and Ephemeral Streams. Dans *Intermittent Rivers and Ephemeral Streams* (p. 79-108). Elsevier. <https://doi.org/10.1016/B978-0-12-803835-2.00004-8>
- Bourke, S. A., Degens, B., Searle, J., de Castro Tayer, T. et Rothery, J. (2021). Geological permeability controls streamflow generation in a remote, ungauged, semi-arid drainage system. *Journal of Hydrology: Regional Studies*, 38, 100956. <https://doi.org/10.1016/j.ejrh.2021.100956>
- Boyer, J., Dieulin, C., Rouche, N., Cres, A., Servat, E. et Paturel, J. (2006). SIEMEM: an environmental information system for water resources. *Climate Variability and Change--hydrological Impacts*, (308), 19.
- Brakebill, J. W., Wolock, D. M. et Terziotti, S. (2011). Digital Hydrologic Networks Supporting Applications Related to Spatially Referenced Regression Modeling 1. *JAWRA Journal of the American Water Resources Association*, 47(5), 916-932. <https://doi.org/10.1111/j.1752-1688.2011.00578.x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A. et Stone, C. J. (1984). Classification and regression trees. Belmont, CA: Wadsworth. *International Group*, 432, 151-166.
- Brooks, R. T. et Colburn, E. A. (2011). Extent and Channel Morphology of Unmapped Headwater Stream Segments of the Quabbin Watershed, Massachusetts1: Extent and Channel Morphology of Unmapped Headwater Stream Segments of the Quabbin Watershed, Massachusetts. *JAWRA Journal of the American Water Resources Association*, 47(1), 158-168. <https://doi.org/10.1111/j.1752-1688.2010.00499.x>
- Brousseau, D., Assani, A. A., Kalenga, F. et Mbenza, M. (2010). Classification des régimes hydrologiques saisonniers de l'Afrique: un outil potentiel pour le suivi des changements environnementaux. *Sécheresse*, 21(2), 147-153. <https://doi.org/10.1684/sec.2010.0238>
- Budic, L., Didenko, G. et Dormann, C. F. (2016). Squares of different sizes: effect of geographical projection on model parameter estimates in species distribution modeling. *Ecology and Evolution*, 6(1), 202-211. <https://doi.org/10.1002/ece3.1838>
- Budyko, M. I., Miller, D. H. et Miller, D. H. (1974). *Climate and life* (vol. 508). Academic press New York.



- Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202(1), 212-230. [https://doi.org/10.1016/S0022-1694\(97\)00068-1](https://doi.org/10.1016/S0022-1694(97)00068-1)
- Busch, M. H., Costigan, K. H., Fritz, K. M., Datry, T., Krabbenhoft, C. A., Hammond, J. C., Zimmer, M., Olden, J. D., Burrows, R. M., Dodds, W. K., Boersma, K. S., Shanafield, M., Kampf, S. K., Mims, M. C., Bogan, M. T., Ward, A. S., Perez Rocha, M., Godsey, S., Allen, G. H., ... Allen, D. C. (2020). What's in a Name? Patterns, Trends, and Suggestions for Defining Non-Perennial Rivers and Streams. *Water*, 12(7), 1980. <https://doi.org/10.3390/w12071980>
- Buttle, J. M., Boon, S., Peters, D. L., Spence, C., (Ilja) van Meerveld, H. J. et Whitfield, P. H. (2012). An Overview of Temporary Stream Hydrology in Canada. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 37(4), 279-310. <https://doi.org/10.4296/cwrj2011-903>
- Calle, M. L. et Urrea, V. (2011). Letter to the Editor: Stability of Random Forest importance measures. *Briefings in Bioinformatics*, 12(1), 86-89. <https://doi.org/10.1093/bib/bbq011>
- Capitaine, L. (2020). *Forêts aléatoires pour données longitudinales de grande dimension*. [Thèse de doctorat, Université de Bordeaux]. <https://tel.archives-ouvertes.fr/tel-03151753>
- Carrier, C., Wirth, S. B., Cochand, F., Hunkeler, D. et Brunner, P. (2018). Geology controls streamflow dynamics. *Journal of Hydrology*, 566, 756-769. <https://doi.org/10.1016/j.jhydrol.2018.08.069>
- Casse, C., Gosset, M., Vischel, T., Quantin, G. et Tanimoun, B. A. (2016). Model-based study of the role of rainfall and land use–land cover in the changes in the occurrence and intensity of Niger red floods in Niamey between 1953 and 2012. *Hydrology and Earth System Sciences*, 20(7), 2841-2859. <https://doi.org/10.5194/hess-20-2841-2016>
- Castaing, C., Le Metour, J., Billa, M., Donzeau, M., Chevremont, P., Egal, E., Zida, B., Ouédraogo, I., Koté, S., Kaboré, B. E., Ouedraogo, C., Thiéblemont, D., Guerrot, C., Cocherie, A., Tegye, M., Milési, J. P. et Itard, Y. (2003). *Carte géologique et minière du Burkina Faso à 1/1 000 000*.
- Castelli, G., Bresci, E., Castelli, F., Hagos, E. Y. et Mehari, A. (2018). A participatory design approach for modernization of spate irrigation systems. *Agricultural Water Management*, 210, 286-295. <https://doi.org/10.1016/j.agwat.2018.08.030>
- Cecchi, P., Meunier-Nikiema, A., Moiroux, N. et Sanou, B. (2009). *Towards an atlas of lakes and reservoirs in Burkina Faso*. International Water Management Institute (IWMI).
- Cecchi, P., Meunier-Nikiema, A., Moiroux, N., Sanou, B. et Bougaire, F. (2007). Why an atlas of lakes and reservoirs in Burkina Faso. *Small Reservoirs Project*.
- Chawla, N. V., Bowyer, K. W., Hall, L. O. et Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357. <https://doi.org/10.48550/arXiv.1106.1813>

- Chiu, M.-C., Leigh, C., Mazor, R., Cid, N. et Resh, V. (2017). Anthropogenic Threats to Intermittent Rivers and Ephemeral Streams. Dans *Intermittent Rivers and Ephemeral Streams* (p. 433-454). Elsevier. <https://doi.org/10.1016/B978-0-12-803835-2.00017-6>
- Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4), 213. <https://doi.org/10.1037/h0026256>
- Colombo, R., Vogt, J. V., Soille, P., Paracchini, M. L. et de Jager, A. (2007). Deriving river networks and catchments at the European scale from medium resolution digital elevation data. *CATENA*, 70(3), 296-305. <https://doi.org/10.1016/j.catena.2006.10.001>
- Cooney, P. B. et Kwak, T. J. (2013). Spatial Extent and Dynamics of Dam Impacts on Tropical Island Freshwater Fish Assemblages. *BioScience*, 63(3), 176-190. <https://doi.org/10.1525/bio.2013.63.3.6>
- Costigan, K. H., Daniels, M. D. et Dodds, W. K. (2015). Fundamental spatial and temporal disconnections in the hydrology of an intermittent prairie headwater network. *Journal of Hydrology*, 522, 305-316. <https://doi.org/10.1016/j.jhydrol.2014.12.031>
- Costigan, K. H., Jaeger, K. L., Goss, C. W., Fritz, K. M. et Goebel, P. C. (2016). Understanding controls on flow permanence in intermittent rivers to aid ecological research: integrating meteorology, geology and land cover: Integrating Science to Understand Flow Intermittence. *Ecohydrology*, 9(7), 1141-1153. <https://doi.org/10.1002/eco.1712>
- Costigan, K. H., Kennard, M. J., Leigh, C., Sauquet, E., Datry, T. et Boulton, A. J. (2017). Flow Regimes in Intermittent Rivers and Ephemeral Streams. Dans *Intermittent Rivers and Ephemeral Streams* (p. 51-78). Elsevier. <https://doi.org/10.1016/B978-0-12-803835-2.00003-6>
- Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A. et Arheimer, B. (2020). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, 65(5), 699-711. <https://doi.org/10.1080/02626667.2019.1659509>
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. et Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. <https://doi.org/10.1890/07-0539.1>
- Da Ros, D. et Borga, M. (1997). Use of digital elevation model data for the derivation of the geomorphological instantaneous unit hydrograph. *Hydrological processes*, 11(1), 13-33. [https://doi.org/10.1002/\(SICI\)1099-1085\(199701\)11:1<13::AID-HYP400>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-1085(199701)11:1<13::AID-HYP400>3.0.CO;2-M)
- Dakoure, D. (2003). *Etude hydrogéologique et géochimique de la bordure sud-est du bassin sédimentaire de Taoudeni (Burkina Faso-Mali): essai de modélisation* [Thèse de doctorat, Université Pierre et Marie Curie]. <https://tel.archives-ouvertes.fr/tel-00608860>
- D'Ambrosio, E., De Girolamo, A. M., Barca, E., Ielpo, P. et Rulli, M. C. (2017). Characterising the hydrological regime of an ungauged temporary river system: a case

- study. *Environmental Science and Pollution Research*, 24(16), 13950-13966. <https://doi.org/10.1007/s11356-016-7169-0>
- Danielson, J. J. et Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. US Geological Survey.
- Datry, T., Arscott, D. B. et Sabater, S. (2011). Recent perspectives on temporary river ecology. *Aquatic Sciences*, 73(4), 453-457. <https://doi.org/10.1007/s00027-011-0236-1>
- Datry, T., Bonada, N. et Boulton, A. J. (2017). General Introduction. Dans *Intermittent Rivers and Ephemeral Streams* (p. 1-20). Elsevier. <https://doi.org/10.1016/B978-0-12-803835-2.00001-2>
- Datry, T., Boulton, A. J., Bonada, N., Fritz, K., Leigh, C., Sauquet, E., Tockner, K., Hugueny, B. et Dahm, C. N. (2018). Flow intermittence and ecosystem services in rivers of the Anthropocene. *Journal of Applied Ecology*, 55(1), 353-364. <https://doi.org/10.1111/1365-2664.12941>
- Datry, T., Corti, R. et Philippe, M. (2012). Spatial and temporal aquatic-terrestrial transitions in the temporary Albarine River, France: responses of invertebrates to experimental rewetting: Invertebrates in dry river sediments. *Freshwater Biology*, 57(4), 716-727. <https://doi.org/10.1111/j.1365-2427.2012.02737.x>
- Datry, T., Larned, S. T. et Tockner, K. (2014). Intermittent Rivers: A Challenge for Freshwater Ecology. *BioScience*, 64(3), 229-235. <https://doi.org/10.1093/biosci/bit027>
- Day, D. G. (1980). Lithologic controls of drainage density: A study of six small rural catchments in New England, N.S.W. *CATENA*, 7(4), 339-351. [https://doi.org/10.1016/0341-8162\(80\)90018-1](https://doi.org/10.1016/0341-8162(80)90018-1)
- De Girolamo, A. M., Bouraoui, F., Buffagni, A., Pappagallo, G. et Lo Porto, A. (2017). Hydrology under climate change in a temporary river system: Potential impact on water balance and flow regime. *River Research and Applications*, 33(7), 1219-1232. <https://doi.org/10.1002/rra.3165>
- De la Torre, J., Puig, D. et Valls, A. (2018). Weighted kappa loss function for multi-class classification of ordinal data in deep learning. *Machine Learning and Applications in Artificial Intelligence*, 105, 144-154. <https://doi.org/10.1016/j.patrec.2017.05.018>
- De Marsily, G. (2009). *L'eau, un trésor en partage*. Dunod Paris.
- De Wit, M. (2006). Changes in Surface Water Supply Across Africa with Predicted Climate Change. *Science*, 311(5769), 1917-1921. <https://doi.org/10.1126/science.1119929>
- Dettinger, M. D. et Diaz, H. F. (2000). Global characteristics of stream flow seasonality and variability. *Journal of hydrometeorology*, 1(4), 289-310. [https://doi.org/10.1175/1525-7541\(2000\)001<0289:GCOSFS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2)
- Dey, P. et Mishra, A. (2017). Separating the impacts of climate change and human activities on streamflow: A review of methodologies and critical assumptions. *Journal of Hydrology*, 548, 278-290. <https://doi.org/10.1016/j.jhydrol.2017.03.014>

- Dieulin, C., Mahé, G., Paturel, J.-E., Ejjiyar, S., Trambly, Y., Rouché, N. et EL Mansouri, B. (2019). A New 60-year 1940/1999 Monthly-Gridded Rainfall Data Set for Africa. *Water*, 11(2), 387. <https://doi.org/10.3390/w11020387>
- Dingman, S. L. (2015). *Physical hydrology*. Waveland press.
- Dixon, H., Sandström, S., Cudennec, C., Lins, H. F., Abrate, T., Béro, D., Chernov, I., Ravalitera, N., Sighomnou, D. et Teichert, F. (2020). Intergovernmental cooperation for hydrometry – what, why and how? *Hydrological Sciences Journal*, 1-15. <https://doi.org/10.1080/02626667.2020.1764569>
- Do, H. X., Gudmundsson, L., Leonard, M. et Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, 10(2), 765-785. <https://doi.org/10.5194/essd-10-765-2018>
- Dobos, E., Daroussin, J. et Montanarella, L. (2005). An SRTM-based procedure to delineate SOTER Terrain Units on 1: 1 and 1: 5 million scales, EUR 21571 EN. *Office for Official Publications of the European Communities, Luxembourg*, 55.
- Dodds, W. K. et Oakes, R. M. (2008). Headwater Influences on Downstream Water Quality. *Environmental Management*, 41(3), 367-377. <https://doi.org/10.1007/s00267-007-9033-y>
- Döll, P. et Fiedler, K. (2008). Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences*, 12(3), 863-885. <https://doi.org/10.5194/hess-12-863-2008>
- Döll, P., Kaspar, F. et Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, 270(1), 105-134. [https://doi.org/10.1016/S0022-1694\(02\)00283-4](https://doi.org/10.1016/S0022-1694(02)00283-4)
- Döll, P. et Schmied, H. M. (2012). How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environmental Research Letters*, 7(1), 014037. <https://doi.org/10.1088/1748-9326/7/1/014037>
- Downing, J. A., Cole, J. J., Duarte, C. M., Middelburg, J. J., Melack, J. M., Prairie, Y. T., Kortelainen, P., Striegl, R. G., McDowell, W. H. et Tranvik, L. J. (2012). Global abundance and size distribution of streams and rivers. *Inland Waters*, 2(4), 229-236. <https://doi.org/10.5268/IW-2.4.502>
- ESRI. (1993). Digital chart of the world, 1:1M (Redlands, Calif).
- Fadul, E., de Fraiture, C. et Masih, I. (2021). Flexibility as a strategy to cope with uncertain water supply in spate irrigation. *Irrigation and Drainage*. <https://doi.org/10.1002/ird.2608>
- Fan, Y., Li, H. et Miguez-Macho, G. (2013). Global Patterns of Groundwater Table Depth. *Science*, 339(6122), 940-943. <https://doi.org/10.1126/science.1229881>
- Fang, X., Cleveland, T., Garcia, C., Thompson, D. et Malla, R. (2005). Literature review on timing parameters for hydrographs. *Department of Civil Engineering, Lamar University, Beaumont*, 83.

- FAO. (2014). AQUAMAPS Global spatial database on water and agriculture, Food and Agriculture Organization of the United Nations. [www.fao.org/nr/water/aquamaps/](http://www.fao.org/nr/water/aquamaps/)
- Farr, T. G. et Kobrick, M. (2000). Shuttle Radar Topography Mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, 81(48), 583-585. <https://doi.org/10.1029/EO081i048p00583>
- Fekete, B. M. et Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *Proceedings of the International Association of Hydrological Sciences*, 309, 129-136.
- Fick, S. E. et Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12), 4302-4315. <https://doi.org/10.1002/joc.5086>
- Fovet, O., Belemtougri, A., Boithias, L., Braud, I., Charlier, J., Cottet, M., Daudin, K., Dramais, G., Ducharne, A., Folton, N., Grippa, M., Hector, B., Kuppel, S., Le Coz, J., Legal, L., Martin, P., Moatar, F., Molénat, J., Probst, A., ... Datry, T. (2021). Intermittent rivers and ephemeral streams: Perspectives for critical zone science and research on socio-ecosystems. *Wiley Interdisciplinary Reviews: Water*, 8(4), e1523. <https://doi.org/10.1002/wat2.1523>
- Fowe, T., Karambiri, H., Paturel, J.-E., Poussin, J.-C. et Cecchi, P. (2015). Water balance of small reservoirs in the Volta basin: A case study of Boura reservoir in Burkina Faso. *Agricultural Water Management*, 152, 99-109. <https://doi.org/10.1016/j.agwat.2015.01.006>
- Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington, P. J., Leibowitz, S. G., Comeleo, R. L., Ebersole, J. L. et Nadeau, T.-L. (2013). Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *JAWRA Journal of the American Water Resources Association*, 49(4), 867-882. <https://doi.org/10.1111/jawr.12040>
- Fritz, K. M., Nadeau, T.-L., Kelso, J. E., Beck, W. S., Mazor, R. D., Harrington, R. A. et Topping, B. J. (2020). Classifying streamflow duration: the scientific basis and an operational framework for method development. *Water*, 12(9), 2545. <https://doi.org/10.3390/w12092545>
- Gaetani, M., Janicot, S., Vrac, M., Famien, A. M. et Sultan, B. (2020). Robust assessment of the time of emergence of precipitation change in West Africa. *Scientific Reports*, 10(1), 7670. <https://doi.org/10.1038/s41598-020-63782-2>
- Gallart, F., Prat, N., García-Roger, E. M., Latron, J., Rieradevall, M., Llorens, P., Barberá, G. G., Brito, D., De Girolamo, A. M., Lo Porto, A., Buffagni, A., Erba, S., Neves, R., Nikolaidis, N. P., Perrin, J. L., Querner, E. P., Quiñero, J. M., Tournoud, M. G., Tzoraki, O., ... Froebrich, J. (2012). A novel approach to analysing the regimes of temporary streams in relation to their controls on the composition and structure of aquatic biota. *Hydrology and Earth System Sciences*, 16(9), 3165-3182. <https://doi.org/10.5194/hess-16-3165-2012>

- Gallo, E. L., Meixner, T., Lohse, K. A. et Nicholas, H. (2020). Estimating Surface Water Presence and Infiltration in Ephemeral to Intermittent Streams in the Southwestern US. *Frontiers in Water*, 2, 572950. <https://doi.org/10.3389/frwa.2020.572950>
- Gardiner, V. (1995). Channel networks: progress in the study of spatial and temporal variations of drainage density. *Changing river channels*, 65-85.
- Gebrehiwot, S. G., Ilstedt, U., Gärdenas, A. I. et Bishop, K. (2011). Hydrological characterization of watersheds in the Blue Nile Basin, Ethiopia. *Hydrology and Earth System Sciences*, 15(1), 11-20. <https://doi.org/10.5194/hess-15-11-2011>
- Gleeson, T., Moosdorf, N., Hartmann, J. et van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, 41(11), 3891-3898. <https://doi.org/10.1002/2014GL059856>
- Godsey, S. E. et Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: hydrologically driven variations in active drainage density, flowing channels and stream order: SCIENTIFIC BRIEFING. *Hydrological Processes*, 28(23), 5791-5803. <https://doi.org/10.1002/hyp.10310>
- González-Ferreras, A. M. et Barquín, J. (2017). Mapping the temporary and perennial character of whole river networks. *Water Resources Research*, 53(8), 6709-6724. <https://doi.org/10.1002/2017WR020390>
- Goodrich, D. C., Kepner, W. G., Levick, L. R. et Wigington, P. J. (2018). Southwestern Intermittent and Ephemeral Stream Connectivity. *JAWRA Journal of the American Water Resources Association*, 54(2), 400-422. <https://doi.org/10.1111/1752-1688.12636>
- Grandini, M., Bagli, E. et Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*. <https://doi.org/10.48550/arXiv.2008.05756>
- GRDC. (2019). *Global Runoff Data Center*. [https://www.bafg.de/GRDC/EN/02\\_srvcs/22\\_gslrs/gislayers\\_node.html](https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/gislayers_node.html)
- Greve, P., Roderick, M., Ukkola, A. et Wada, Y. (2019). The aridity index under global warming. *Environmental Research Letters*, 14(12), 124006. <https://doi.org/10.1088/1748-9326/ab5046>
- Gupta, S., Hengl, T., Lehmann, P., Bonetti, S. et Or, D. (2021). SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications. *Earth System Science Data*, 13(4), 1593-1612. <https://doi.org/10.5194/essd-13-1593-2021>
- Hafen, K., Blasch, K., Rea, A. H., Sando, R. et Paul Gessler. (2020). The influence of climate variability on the accuracy of NHD perennial and non-perennial stream classifications. *Journal of the American Water Resources Association*, 56(5), 903-916. <https://doi.org/10.1111/1752-1688.12871>

- Haile Ghebremariam, B. et van Steenbergen, F. (2006). Agricultural Water Management in Ephemeral Rivers: Community Management in Spate Irrigation in Eritrea. *African Water Journal*, 1(1), 54-71.
- Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., Zipper, S. C., Burrows, R. M., Kampf, S. K. et Dodds, W. (2021). Spatial patterns and drivers of non-perennial flow regimes in the contiguous United States. *Geophysical Research Letters*, 48(2), e90794. <https://doi.org/10.1029/2020GL090794>
- Hannah, D. M., Demuth, S., van Lanen, H. A., Looser, U., Prudhomme, C., Rees, G., Stahl, K. et Tallaksen, L. M. (2011). Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, 25(7), 1191-1200. <https://doi.org/10.1002/hyp.7794>
- Harris, I., Jones, P. D., Osborn, T. J. et Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *International journal of climatology*, 34(3), 623-642. <https://doi.org/10.1002/joc.3711>
- Harris, I., Osborn, T. J., Jones, P. et Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data*, 7(1), 1-18. <https://doi.org/10.1038/s41597-020-0453-3>
- Hartmann, J. et Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface: TECHNICAL BRIEF. *Geochemistry, Geophysics, Geosystems*, 13(12). <https://doi.org/10.1029/2012GC004370>
- Hastie, T., Tibshirani, R., Friedman, J. H. et Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (vol. 2). Springer.
- Hedman, E. R. et Osterkamp, W. R. (1982). *Streamflow characteristics related to channel geometry of streams in western United States* [Water-Supply Paper 2193]. <https://doi.org/10.3133/wsp2193>
- Heine, R. A., Lant, C. L. et Sengupta, R. R. (2004). Development and Comparison of Approaches for Automated Mapping of Stream Channel Networks. *Annals of the Association of American Geographers*, 94(3), 477-490. <https://doi.org/10.1111/j.1467-8306.2004.00409.x>
- Hengl, T. (2018). Global DEM derivatives at 250 m, 1 km and 2 km based on the MERIT DEM. *Zenodo*, (1.0). <https://doi.org/10.5281/zenodo.1447210>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangquan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. et Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Hoekstra, A. Y. et Mekonnen, M. M. (2012). The water footprint of humanity. *Proceedings of the National Academy of Sciences*, 109(9), 3232-3237. <https://doi.org/10.1073/pnas.1109936109>

- Hoens, T. R. et Chawla, N. V. (2013). Imbalanced datasets: from sampling to classifiers. *Imbalanced learning: Foundations, algorithms, and applications*, 43-59. <https://doi.org/10.1002/9781118646106.ch3>
- Hooshyar, M., Singh, A., Wang, D. et Foufoula-Georgiou, E. (2019). Climatic Controls on Landscape Dissection and Network Structure in the Absence of Vegetation. *Geophysical Research Letters*, 46(6), 3216-3224. <https://doi.org/10.1029/2019GL082043>
- Hornby, D. (2020). RivEX (Version 10.35) [Software]. <http://www.rivex.co.uk>
- Horton, R. E. (1932). Drainage-basin characteristics. *Eos, transactions american geophysical union*, 13(1), 350-361. <https://doi.org/10.1029/TR013i001p00350>
- Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological society of America bulletin*, 56(3), 275-370. <https://doi.org/10.1177/030913339501900406>
- Hughes, D. A. (2005). Hydrological issues associated with the determination of environmental water requirements of ephemeral rivers. *River Research and Applications*, 21(8), 899-908. <https://doi.org/10.1002/rra.857>
- Huneau, F., Dakoure, D., Celle-Jeanton, H., Vitvar, T., Ito, M., Traore, S., Compaore, N. F., Jirakova, H. et Le Coustumer, P. (2011). Flow pattern and residence time of groundwater within the south-eastern Taoudeni sedimentary basin (Burkina Faso, Mali). *Journal of Hydrology*, 409(1-2), 423-439. <https://doi.org/10.1016/j.jhydrol.2011.08.043>
- Huscroft, J., Gleeson, T., Hartmann, J. et Börker, J. (2018). Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, 45(4), 1897-1904. <https://doi.org/10.1002/2017GL075860>
- Ibrahim, B. (2012). *Caractérisation des saisons de pluies au Burkina Faso dans un contexte de changement climatique et évaluation des impacts hydrologiques sur le bassin du Nakanbé* [Thèse de doctorat, Université Pierre et Marie Curie - Paris VI]. <https://hal.inria.fr/tel-00827764/>
- IGB, I. G. du B. F. (2012). BNDT (Base Nationale des Données Topographiques). [http://www.igb.bf/?page\\_id=85](http://www.igb.bf/?page_id=85)
- Jacobson, P. J. et Jacobson, K. M. (2013). Hydrologic controls of physical and ecological processes in Namib Desert ephemeral rivers: Implications for conservation and management. *Journal of Arid Environments*, 93, 80-93. <https://doi.org/10.1016/j.jaridenv.2012.01.010>
- Jaeger, K. L., Olden, J. D. et Pelland, N. A. (2014). Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. *Proceedings of the National Academy of Sciences*, 111(38), 13894-13899. <https://doi.org/10.1073/pnas.1320890111>



- Jaeger, K. L., Sando, R., McShane, R. R., Dunham, J. B., Hockman-Wert, D. P., Kaiser, K. E., Hafen, K., Risley, J. C. et Blasch, K. W. (2019). Probability of Streamflow Permanence Model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. *Journal of Hydrology X*, 2, 100005. <https://doi.org/10.1016/j.hydroa.2018.100005>
- Jalloh, A., Nelson, G. C., Thomas, T. S., Zougmore, R. B. et Roy-Macauley, H. (2013). *Chapter 4 in West African agriculture and climate change: a comprehensive analysis*. IFPRI Research Monograph. Washington, D.C. Intl Food Policy Res Inst. <http://dx.doi.org/10.2499/9780896292048>
- James, G., Witten, D., Hastie, T. et Tibshirani, R. (2013). *An Introduction to Statistical Learning* (vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jaud, M., Rouveure, R., Faure, P. et Monod, M. (2013). Radar hyperfréquence embarqué appliqué à la cartographie du corridor fluvial (p. 12-p). SAGEO Conférence internationale de Géomatique et d'analyse spatiale. <https://hal.archives-ouvertes.fr/hal-00948369>
- Jenson, S. K. et Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information-system analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11), 1593-1600. <http://pubs.er.usgs.gov/publication/70142175>
- Jolliffe, I. T. et Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141-151. <https://doi.org/10.1177/001316446002000116>
- Kaplan, N. H., Blume, T. et Weiler, M. (2020). Predicting probabilities of streamflow intermittency across a temperate mesoscale catchment. *Hydrology and Earth System Sciences*, 24(11), 5453-5472. <https://doi.org/10.5194/hess-24-5453-2020>
- Kaplan, N. H., Sohr, E., Blume, T. et Weiler, M. (2019). Monitoring ephemeral, intermittent and perennial streamflow: a dataset from 182 sites in the Atttert catchment, Luxembourg. *Earth System Science Data*, 11(3), 1363-1374. <https://doi.org/10.5194/essd-11-1363-2019>
- Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L. et Marsh, N. (2010). Classification of natural flow regimes in Australia to support environmental flow management: Classification of natural flow regimes in Australia. *Freshwater Biology*, 55(1), 171-193. <https://doi.org/10.1111/j.1365-2427.2009.02307.x>
- Khorshidi, H. A. et Aickelin, U. (2020). A Synthetic Over-sampling method with Minority and Majority classes for imbalance problems. *arXiv preprint arXiv:2011.04170*. <https://doi.org/10.48550/arXiv.2011.04170>

- Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C. et Endo, H. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5-48. <https://doi.org/10.2151/jmsj.2015-001>
- Koïta, M., Sandwidi, W. J. P. et Dara, A. E. (2017). Recharge Estimation of Hard Rock Aquifers under Sahelian Climate Conditions Using Water Table Fluctuation: Case Study of Tougou Catchment, Burkina Faso. *Journal of Water Resource and Protection*, 9(12), 1428-1448. <https://doi.org/10.4236/jwarp.2017.912092>
- Kostianoy, A. G., Soloviev, D. M. et Pešić, V. (2020). Application of Google Earth in Mapping Intermittent Rivers of Montenegro. Dans V. Pešić, M. Paunović et A. G. Kostianoy (dir.), *The Rivers of Montenegro* (vol. 93, p. 253-263). Springer International Publishing. [https://doi.org/10.1007/698\\_2020\\_488](https://doi.org/10.1007/698_2020_488)
- Kouanda, B., Coulibaly, P., Niang, D., Fowe, T., Karambiri, H. et Paturel, J. E. (2018). Analysis of the Performance of Base Flow Separation Methods Using Chemistry and Statistics in Sudano-Sahelian Watershed, Burkina Faso. *Hydrology: Current Research*, 09(02). <https://doi.org/10.4172/2157-7587.1000300>
- Koussoube, Y. (2010). *Hydrogéologie des séries sédimentaires de la dépression piézométrique du Gondo (bassin du Sourou): Burkina Faso/Mali* [Thèse de doctorat, Université Pierre et Marie Curie-Paris VI]. <https://tel.archives-ouvertes.fr/tel-00815287>
- Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., DelVecchia, A. G., Fritz, K. M., Shanafield, M., Burgin, A. J., Zimmer, M. A., Detry, T., Dodds, W. K., Jones, C. N., Mims, M. C., Franklin, C., Hammond, J. C., Zipper, S., Ward, A. S., ... Olden, J. D. (2022). Assessing placement bias of the global river gauge network. *Nature Sustainability*. <https://doi.org/10.1038/s41893-022-00873-0>
- Kranjac-Berisavljevic, G., Abdul-Ghanyu, S., Gandaa, B. Z. et Abagale, F. K. (2014). Dry Spells Occurrence in Tamale, Northern Ghana-Review of Available Information. *Journal of Disaster Research*, 9(4), 468-474. <https://doi.org/10.20965/jdr.2014.p0468>
- Kuhn, M. et Johnson, K. (2013). *Applied predictive modeling* (vol. 26). Springer.
- Kuhn, M. et Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B. et Team, R. C. (2020). Package ‘caret’. *The R Journal*, 223.
- Kumar, N., Nayak, S. B., Adiga, M. et Aithal, A. P. (2018). Determination of Spearman Correlation Coefficient. *Dermatology research and practice*, 2018, 1-6.
- Kursa, M. B. et Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1-13. <https://doi.org/10.18637/jss.v036.i11>
- Kustu, M. D., Fan, Y. et Robock, A. (2010). Large-scale water cycle perturbation due to irrigation pumping in the US High Plains: A synthesis of observed streamflow

- changes. *Journal of Hydrology*, 390(3-4), 222-244. <https://doi.org/10.1016/j.jhydrol.2010.06.045>
- Landis, J. R. et Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174. <https://doi.org/10.2307/2529310>
- Lane, B. A., Dahlke, H. E., Pasternack, G. B. et Sandoval-Solis, S. (2017). Revealing the Diversity of Natural Hydrologic Regimes in California with Relevance for Environmental Flows Applications. *JAWRA Journal of the American Water Resources Association*, 53(2), 411-430. <https://doi.org/10.1111/1752-1688.12504>
- Lange, H. et Sippel, S. (2020). Machine Learning Applications in Hydrology. Dans D. F. Levia, D. E. Carlyle-Moses, S. Iida, B. Michalzik, K. Nanko et A. Tischer (dir.), *Forest-Water Interactions* (p. 233-257). Springer International Publishing. [https://doi.org/10.1007/978-3-030-26086-6\\_10](https://doi.org/10.1007/978-3-030-26086-6_10)
- Larned, S. T., Datry, T., Arscott, D. B. et Tockner, K. (2010). Emerging concepts in temporary-river ecology. *Freshwater Biology*, 55(4), 717-738. <https://doi.org/10.1111/j.1365-2427.2009.02322.x>
- Lebel, T. et Ali, A. (2009). Recent trends in the Central and Western Sahel rainfall regime (1990–2007). *Journal of Hydrology*, 375(1-2), 52-64. <https://doi.org/10.1016/j.jhydrol.2008.11.030>
- Lehner, B. (2012). *Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network*. Global Runoff Data Centre in the Federal Institute of Hydrology (BFG). [https://www.bafg.de/GRDC/EN/02\\_srvcs/22\\_gslrs/222\\_WSB/methodology\\_Lehner.html](https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/222_WSB/methodology_Lehner.html)
- Lehner, B. et Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171-2186. <https://doi.org/10.1002/hyp.9740>
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K. et Magome, J. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9), 494-502.
- Lehner, B., Verdin, K. et Jarvis, A. (2006). HydroSHEDS technical documentation, version 1.0. *World Wildlife Fund US, Washington, DC*, 1-27.
- Lehner, B., Verdin, K. et Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, 89(10), 93. <https://doi.org/10.1029/2008EO100001>
- Leigh, C., Boulton, A. J., Courtwright, J. L., Fritz, K., May, C. L., Walker, R. H. et Datry, T. (2016). Ecological research and management of intermittent rivers: an historical review and future directions. *Freshwater Biology*, 61(8), 1181-1199. <https://doi.org/10.1111/fwb.12646>

- Leigh, C. et Datry, T. (2017). Drying as a primary hydrological determinant of biodiversity in river systems: a broad-scale analysis. *Ecography*, 40(4), 487-499. <https://doi.org/10.1111/ecog.02230>
- Liaw, A. et Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lin, P., Pan, M., Allen, G. H., de Frasson, R. P., Zeng, Z., Yamazaki, D. et Wood, E. F. (2020). Global estimates of reach-level bankfull river width leveraging big data geospatial analysis. *Geophysical Research Letters*, 47(7), e2019GL086405. <https://doi.org/10.1029/2019GL086405>
- Lin, P., Pan, M., Wood, E. F., Yamazaki, D. et Allen, G. H. (2021). A new vector-based global river network dataset accounting for variable drainage density. *Scientific Data*, 8(1), 28. <https://doi.org/10.1038/s41597-021-00819-9>
- Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F. et Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, 6(1), 283. <https://doi.org/10.1038/s41597-019-0300-6>
- Liu, X., Peterson, J. et Zhang, Z. (2005). High-resolution DEM generated from LiDAR data for water resource management (p. 1402-1408). Proceedings of the International Congress on Modelling and Simulation (MODSIM05), Modelling and Simulation Society of Australia and New Zealand Inc.
- Lovill, S. M., Hahm, W. J. et Dietrich, W. E. (2018). Drainage from the Critical Zone: Lithologic Controls on the Persistence and Spatial Extent of Wetted Channels during the Summer Dry Season. *Water Resources Research*, 54(8), 5702-5726. <https://doi.org/10.1029/2017WR021903>
- Luo, W., Jasiewicz, J., Stepinski, T., Wang, J., Xu, C. et Cang, X. (2016). Spatial association between dissection density and environmental factors over the entire conterminous United States. *Geophysical Research Letters*, 43(2), 692-700. <https://doi.org/10.1002/2015GL066941>
- Luo, W., Stepinski, T. et Qi, Y. (2007). Drainage Density and Controlling Factors in Cascade Range, Oregon, USA. *proceedings of Geoinformatics*, 11.
- Luoto, M. (2007). New Insights into Factors Controlling Drainage Density in Subarctic Landscapes. *Arctic, Antarctic, and Alpine Research*, 39(1), 117-126. [https://doi.org/10.1657/1523-0430\(2007\)39\[117:NIIFCD\]2.0.CO;2](https://doi.org/10.1657/1523-0430(2007)39[117:NIIFCD]2.0.CO;2)
- MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó. et Taylor, R. G. (2012). Quantitative maps of groundwater resources in Africa. *Environmental Research Letters*, 7(2), 024009. <https://doi.org/10.1088/1748-9326/7/2/024009>
- MacDonald, A. M., Lark, R. M., Taylor, R. G., Abiye, T., Fallas, H. C., Favreau, G., Goni, I. B., Kebede, S., Scanlon, B., Sorensen, J. P. R., Tijani, M., Upton, K. A. et West, C. (2021). Mapping groundwater recharge in Africa from ground observations and implications for water security. *Environmental Research Letters*, 16(3), 034012. <https://doi.org/10.1088/1748-9326/abd661>

- Madduma Bandara, C. M. (1974). Drainage density and effective precipitation. *Journal of Hydrology*, 21(2), 187-190. [https://doi.org/10.1016/0022-1694\(74\)90036-5](https://doi.org/10.1016/0022-1694(74)90036-5)
- Mahé, G. (2006). *Variabilité pluie-débit en Afrique de l'Ouest et Centrale au 20ème siècle : changements hydro-climatiques, occupation du sol et modélisation hydrologique* [Mémoire Habilitation à Diriger des Recherches, Université de Montpellier II].
- Mahé, G. (2009). Surface/groundwater interactions in the Bani and Nakambe rivers, tributaries of the Niger and Volta basins, West Africa. *Hydrological Sciences Journal*, 54(4), 704-712. <https://doi.org/10.1623/hysj.54.4.704>
- Mahé, G., Lienou, G., Bamba, F., Paturol, J.-E., Adeaga, O., Descroix, L., Mariko, A., Olivry, J.-C., Sangaré, S. et Ogilvie, A. (2011). Le fleuve Niger et le changement climatique au cours des 100 dernières années (vol. 344, p. 131-7). Hydroclimatology variability and change, Proceedings of symposium held during IUGG.
- Maidment, D. R. (1993). *Handbook of hydrology* (vol. 9780070). McGraw-Hill New York.
- Makarewicz, A. N. (2020). *Quantifying Intermittent Streamflow with Low-cost Sensors and Fully Integrated Hydrological Modelling of Upstream Runoff* [Thèse de doctorat, Université de Adélaïde]. <https://hdl.handle.net/2440/127014>
- Mardhel, V., Pinson, S. et Allier, D. (2020). Description of an indirect method (IDPR) to determine spatial distribution of infiltration and runoff and its hydrogeological applications to the French territory. *Journal of Hydrology X*, 100065. <https://doi.org/10.1016/j.hydroa.2020.100065>
- Mattivi, P., Franci, F., Lambertini, A. et Bitelli, G. (2019). TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, 4(1), 6. <https://doi.org/10.1186/s40965-019-0066-y>
- McManamay, R. A. et DeRolph, C. R. (2019). A stream classification system for the conterminous United States. *Scientific Data*, 6(1), 190017. <https://doi.org/10.1038/sdata.2019.17>
- Meerveld, H. J. I., Sauquet, E., Gallart, F., Sefton, C., Seibert, J. et Bishop, K. (2020). Aqua temporaria incognita. *Hydrological Processes*, 34(26), 5704-5711. <https://doi.org/10.1002/hyp.13979>
- Merot, P., Squvidant, H., Arousseau, P., Hefting, M., Burt, T., Maitre, V., Kruk, M., Butturini, A., Thenail, C. et Viaud, V. (2003). Testing a climato-topographic index for predicting wetlands distribution along an European climate gradient. *Ecological modelling*, 163(1-2), 51-71. [https://doi.org/10.1016/S0304-3800\(02\)00387-3](https://doi.org/10.1016/S0304-3800(02)00387-3)
- Merritt, A. M., Lane, B. et Hawkins, C. P. (2021). Classification and Prediction of Natural Streamflow Regimes in Arid Regions of the USA. *Water*, 13(3), 380. <https://doi.org/10.3390/w13030380>
- Messenger, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C. et Datry, T. (2021). Global prevalence of non-perennial rivers and streams. *Nature*, 594(7863), 391-397. <https://doi.org/10.1038/s41586-021-03565-5>

- Messenger, M. L., Lehner, B., Grill, G., Nedeva, I. et Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1), 13603. <https://doi.org/10.1038/ncomms13603>
- Meybeck, M. et Ragu, A. (1997). *River discharges to the oceans: an assessment of suspended solids, major ions and nutrients* (vol. 245). UNEP.
- Moglen, G. E., Eltahir, E. A. B. et Bras, R. L. (1998). On the sensitivity of drainage density to climate change. *Water Resources Research*, 34(4), 855-862. <https://doi.org/10.1029/97WR02709>
- Moliere, D. R., Lowry, J. B. C. et Humphrey, C. L. (2009). Classifying the flow regime of data-limited streams in the wet-dry tropical region of Australia. *Journal of Hydrology*, 367(1-2), 1-13. <https://doi.org/10.1016/j.jhydrol.2008.12.015>
- Molinaro, A. M., Simon, R. et Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21(15), 3301-3307. <https://doi.org/10.1093/bioinformatics/bti499>
- Molnar, C., Casalicchio, G. et Bischl, B. (2018). iml: An R package for interpretable machine learning. *Journal of Open-Source Software*, 3(26), 786. <https://doi.org/10.21105/joss.00786>
- Morin, E., Grodek, T., Dahan, O., Benito, G., Kulls, C., Jacoby, Y., Langenhove, G. V., Seely, M. et Enzel, Y. (2009). Flood routing and alluvial aquifer recharge along the ephemeral arid Kuiseb River, Namibia. *Journal of Hydrology*, 368(1-4), 262-275. <https://doi.org/10.1016/j.jhydrol.2009.02.015>
- Nadeau, T.-L. et Rains, M. C. (2007). Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy1: Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy. *JAWRA Journal of the American Water Resources Association*, 43(1), 118-133. <https://doi.org/10.1111/j.1752-1688.2007.00010.x>
- Nag, S. K. et Chakraborty, S. (2003). Influence of rock types and structures in the development of drainage network in hard rock area. *Journal of the Indian Society of Remote Sensing*, 31(1), 25-35. <https://doi.org/10.1007/BF03030749>
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C. et Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3), e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- Nka Nnomo, B. (2016). *Contribution à l'Actualisation des Normes Hydrologiques en relation avec les Changements Climatiques et Environnementaux en Afrique de l'Ouest* [Thèse de doctorat, Université Pierre et Marie Curie-Paris VI; Institut international d'ingénierie de l'eau et de l'environnement]. <https://tel.archives-ouvertes.fr/tel-01398353>
- O'Callaghan, J. F. et Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, 28(3), 323-344. [https://doi.org/10.1016/S0734-189X\(84\)80011-0](https://doi.org/10.1016/S0734-189X(84)80011-0)

- Oguchi, T. (1997). Drainage Density and Relative Relief in Humid Steep Mountains with Frequent Slope Failure. *Earth Surface Processes and Landforms*, 22(2), 107-120. [https://doi.org/10.1002/\(SICI\)1096-9837\(199702\)22:2<107::AID-ESP680>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1096-9837(199702)22:2<107::AID-ESP680>3.0.CO;2-U)
- Olden, J. D., Messenger, M. L., Tharme, R. E., Kashaigili, J. J., Munkyaala, D., Zielinski, L. et Warner, A. (2021). Hydrologic classification of Tanzanian rivers to support national water resource policy. *Ecohydrology*, 14(4), e2282. <https://doi.org/10.1002/eco.2282>
- Ouédraogo, I. (1994). *Géologie et hydrogéologie des formations sédimentaires de la boucle du Mouhoun (Burkina Faso)* [Thèse de doctorat, Université Cheikh Anta Diop].
- Ouedraogo, M., Paturel, J., Mahe, G. et Servat, E. (2002). Conséquences de la sécheresse observée depuis le début des années 1970 en Afrique de l'Ouest et centrale: normes météorologiques et hydrologiques. *IAHS-AISH publication*, 149-155.
- Ouellet Dallaire, C., Lehner, B., Sayre, R. et Thieme, M. (2019). A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environmental Research Letters*, 14(2), 024003. <https://doi.org/10.1088/1748-9326/aad8e9>
- Oueslati, O., De Girolamo, A. M., Abouabdillah, A., Kjeldsen, T. R. et Lo Porto, A. (2015). Classifying the flow regimes of Mediterranean streams using multivariate analysis. *Hydrological Processes*, 29(22), 4666-4682. <https://doi.org/10.1002/hyp.10530>
- PANA, B. (2007). Programme National d'adaptation de la variabilité et aux changements climatiques. *Ministère de l'environnement et du cadre du vie, Burkina Faso*.
- Papa, F., Crétaux, J.-F., Grippa, M., Robert, E., Trigg, M., Tshimanga, R. M., Kitambo, B., Paris, A., Carr, A., Fleischmann, A. S., de Fleury, M., Gbetkom, P. G., Calmettes, B. et Calmant, S. (2022). Water Resources in Africa under Global Change: Monitoring Surface Waters from Space. *Surveys in Geophysics*. <https://doi.org/10.1007/s10712-022-09700-9>
- Papageorgaki, I. et Nalbantis, I. (2017). Definition of critical support area revisited. *European Water*, 57, 273-278.
- Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G. et Dietrich, W. E. (2010). A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. *Journal of Geophysical Research*, 115(F1). <https://doi.org/10.1029/2009JF001254>
- Paturel, J., Boubacar, I., L'aour-Cres, A. et Mahé, G. (2010). Research note: Monthly rainfall grids in West and Central Africa. *Revue des sciences de l'eau*, 23(4), 325-333. <https://doi.org/10.7202/045095ar>
- Pavelic, P., Giordano, M., Keraita, B. N., Ramesh, V. et Rao, T. (2012). *Groundwater availability and use in Sub-Saharan Africa: a review of 15 countries*. International Water Management Institute (IWMI).

- Peel, M. C., Finlayson, B. L. et McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5), 1633-1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Peirce, S. E. et Lindsay, J. B. (2015). Characterizing ephemeral streams in a southern Ontario watershed using electrical resistance sensors. *Hydrological Processes*, 29(1), 103-111. <https://doi.org/10.1002/hyp.10136>
- Perez-Saez, J., Mande, T., Larsen, J., Ceperley, N. et Rinaldo, A. (2017). Classification and prediction of river network ephemerality and its relevance for waterborne disease epidemiology. *Advances in Water Resources*, 110, 263-278. <https://doi.org/10.1016/j.advwatres.2017.10.003>
- Petheram, C., McMahon, T. A. et Peel, M. C. (2008). Flow characteristics of rivers in northern Australia: implications for development. *Journal of Hydrology*, 357(1-2), 93-111. <https://doi.org/10.1016/j.jhydrol.2008.05.008>
- Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K. et McDonnell, J. J. (2017). Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. *Hydrological Processes*, 31(10), 1828-1845. <https://doi.org/10.1002/hyp.11134>
- Pinzon, J. E. et Tucker, C. J. (2014). A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote sensing*, 6(8), 6929-6960. <https://doi.org/10.3390/rs6086929>
- Piquemal, D. (1991). Inventaire et bilan des retenues d'eau au Burkina Faso. *Les cahiers d'outre-mer*, 44(175), 259-280.
- Poff, N. L. et Ward, J. V. (1989). Implications of Streamflow Variability and Predictability for Lotic Community Structure: A Regional Analysis of Streamflow Patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(10), 1805-1818. <https://doi.org/10.1139/f89-228>
- Prancevic, J. P. et Kirchner, J. W. (2019). Topographic Controls on the Extension and Retraction of Flowing Streams. *Geophysical Research Letters*, 46(4), 2084-2092. <https://doi.org/10.1029/2018GL081799>
- Price, A. N., Jones, C. N., Hammond, J. C., Zimmer, M. A. et Zipper, S. C. (2021). The Drying Regimes of Non-Perennial Rivers and Streams. *Geophysical Research Letters*, 48(14). <https://doi.org/10.1029/2021GL093298>
- Probst, P. et Boulesteix, A.-L. (2017). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research*, 18(1), 6673-6690. <https://doi.org/10.48550/arXiv.1705.05654>
- Probst, P., Wright, M. N. et Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3). <https://doi.org/10.1002/widm.1301>
- Pumo, D., Caracciolo, D., Viola, F. et Noto, L. V. (2016). Climate change effects on the hydrological regime of small non-perennial river basins. *Science of The Total Environment*, 542, 76-92. <https://doi.org/10.1016/j.scitotenv.2015.10.109>



- R Core Team. (2021). *R: A language and environment for statistical computing (R Version 4.0.3, R Foundation for Statistical Computing, Vienna, Austria, 2020)*.
- Rabba, Z. A., Fatoyinbo, B. et Stretch, D. D. (2018). Applications of the PyTOPKAPI model to ungauged catchments. *Water SA*, 44(2), 162-175. <https://doi.org/10.4314/wsa.v44i2.03>
- Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P. et Guth, P. (2013). Global carbon dioxide emissions from inland waters. *Nature*, 503(7476), 355-359. <https://doi.org/10.1038/nature12760>
- Reid, L. M. et Ziemer, R. R. (1994). Evaluating the biological significance of intermittent streams. Issues in watershed analysis. Discussions at interdisciplinary and interagency workshops held at the Humboldt Interagency Watershed Analysis Center in McKinleyville, California. 13 p.
- Reynolds, L. V., Shafroth, P. B. et Poff, N. L. (2015). Modeled intermittency risk for small streams in the Upper Colorado River Basin under climate change. *Journal of Hydrology*, 523, 768-780. <https://doi.org/10.1016/j.jhydrol.2015.02.025>
- Richter, B. D., Baumgartner, J. V., Powell, J. et Braun, D. P. (1996). A Method for Assessing Hydrologic Alteration within Ecosystems. *Conservation Biology*, 10(4), 1163-1174. <https://doi.org/10.1046/j.1523-1739.1996.10041163.x>
- Rivers-Moore, N., Schulze, R. et Davis, N. (2020). A regional flow type classification for South African perennial rivers. *Water SA*, 46(3), 383-392. <https://doi.org/10.17159/wsa/2020.v46.i3.8648>
- Rusagara, R., Koïta, M., Plagnes, V. et Jost, A. (2022). Groundwater recharge pathways to a weathered-rock aquifer system in a dryland catchment in Burkina Faso. *Hydrogeology Journal*, 30(5), 1489-1512. <https://doi.org/10.1007/s10040-022-02506-4>
- Saadi, M., Oudin, L. et Ribstein, P. (2019). Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters. *Water*, 11(8), 1540. <https://doi.org/10.3390/w11081540>
- Sánchez-Montoya, M. M., von Schiller, D., Barberá, G. G., Díaz, A. M., Arce, M. I., del Campo, R. et Tockner, K. (2018). Understanding the effects of predictability, duration, and spatial pattern of drying on benthic invertebrate assemblages in two contrasting intermittent streams. *PLOS ONE*, 13(3), e0193933. <https://doi.org/10.1371/journal.pone.0193933>
- Sando, R. et Blasch, K. W. (2015). Predicting alpine headwater stream intermittency: a case study in the northern Rocky Mountains. *Ecohydrology & Hydrobiology*, 15(2), 68-80. <https://doi.org/10.1016/j.ecohyd.2015.04.002>
- Sangireddy, H., Carothers, R. A., Stark, C. P. et Passalacqua, P. (2016). Controls of climate, topography, vegetation, and lithology on drainage density extracted from high resolution topography data. *Journal of Hydrology*, 537, 271-282. <https://doi.org/10.1016/j.jhydrol.2016.02.051>

- Sauquet, E., Shanafield, M., Hammond, J. C., Sefton, C., Leigh, C. et Datry, T. (2021). Classification and trends in intermittent river flow regimes in Australia, northwestern Europe and USA: A global perspective. *Journal of Hydrology*, 597, 126170. <https://doi.org/10.1016/j.jhydrol.2021.126170>
- Sauret, E. (2013). *Etude des potentialités hydrogéologiques d'une plaine alluviale en relation avec les eaux souterraines et de surface dans un contexte d'agriculture irriguée (Burkina Faso)* [Thèse de doctorat, Université de Liège]. <https://hdl.handle.net/2268/152969>
- Schacht, B. S. (2019). *Identification and estimation of ecologically relevant flow indices for non-perennial rivers, South Africa* [Thèse de doctorat, University of the Western Cape]. <http://hdl.handle.net/11394/7115>
- Scherer, S., Rehder, J., Achar, S., Cover, H., Chambers, A., Nuske, S. et Singh, S. (2012). River mapping from a flying robot: state estimation, river detection, and obstacle mapping. *Autonomous Robots*, 33(1), 189-214. <https://doi.org/10.1007/s10514-012-9293-0>
- Schneider, A., Jost, A., Coulon, C., Silvestre, M., Théry, S. et Ducharne, A. (2017). Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density: Global-Scale River Network Extraction. *Geophysical Research Letters*, 44(6), 2773-2781. <https://doi.org/10.1002/2016GL071844>
- Schwarz, D. F., König, I. R. et Ziegler, A. (2010). On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. *Bioinformatics*, 26(14), 1752-1758. <https://doi.org/10.1093/bioinformatics/btq257>
- Shanafield, M. et Cook, P. G. (2014). Transmission losses, infiltration and groundwater recharge through ephemeral and intermittent streambeds: A review of applied methods. *Journal of Hydrology*, 511, 518-529. <https://doi.org/10.1016/j.jhydrol.2014.01.068>
- Shangguan, W., Hengl, T., de Jesus, J. M., Yuan, H. et Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, 9(1), 65-88. <https://doi.org/10.1002/2016MS000686>
- Skoulikidis, N. T., Sabater, S., Datry, T., Morais, M. M., Buffagni, A., Dörflinger, G., Zogaris, S., del Mar Sánchez-Montoya, M., Bonada, N., Kalogianni, E., Rosado, J., Vardakas, L., De Girolamo, A. M. et Tockner, K. (2017). Non-perennial Mediterranean rivers in Europe: Status, pressures, and challenges for research and management. *Science of The Total Environment*, 577, 1-18. <https://doi.org/10.1016/j.scitotenv.2016.10.147>
- Skoulikidis, N. Th. (2009). The environmental state of rivers in the Balkans—A review within the DPSIR framework. *Science of The Total Environment*, 407(8), 2501-2516. <https://doi.org/10.1016/j.scitotenv.2009.01.026>
- Smakhtin, V. U. (2001). Low flow hydrology: a review. *Journal of Hydrology*, 240(3-4), 147-186. [https://doi.org/10.1016/S0022-1694\(00\)00340-1](https://doi.org/10.1016/S0022-1694(00)00340-1)

- Snelder, T. H., Datry, T., Lamouroux, N., Larned, S. T., Sauquet, E., Pella, H. et Catalogne, C. (2013). Regionalization of patterns of flow intermittence from gauging station records. *Hydrology and Earth System Sciences*, 17(7), 2685-2699. <https://doi.org/10.5194/hess-17-2685-2013>
- Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, 38(6), 913-920. <https://doi.org/10.1029/TR038i006p00913>
- Strobl, C., Boulesteix, A.-L., Zeileis, A. et Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
- Strobl, R. O. et Forte, F. (2007). Artificial neural network exploration of the influential factors in drainage network derivation. *Hydrological Processes*, 21(22), 2965-2978. <https://doi.org/10.1002/hyp.6506>
- Stubington, R., Paillex, A., England, J., Barthès, A., Bouchez, A., Rimet, F., Sánchez-Montoya, M. M., Westwood, C. G. et Datry, T. (2019). A comparison of biotic groups as dry-phase indicators of ecological quality in intermittent rivers and ephemeral streams. *Ecological Indicators*, 97, 165-174. <https://doi.org/10.1016/j.ecolind.2018.09.061>
- Tafasca, S. (2020). *Evaluation de l'impact des propriétés du sol sur l'hydrologie simulée dans le modèle ORCHIDEE* [Thèse de doctorat, Sorbonne Université]. <https://tel.archives-ouvertes.fr/tel-03220775>
- Therneau, T. M. et Atkinson, E. J. (1997). *An introduction to recursive partitioning using the RPART routines*, 61, 452.
- Tirogo, J., Jost, A., Biaou, A., Valdes-Lao, D., Koussoubé, Y. et Ribstein, P. (2016). Climate Variability and Groundwater Response: A Case Study in Burkina Faso (West Africa). *Water*, 8(5), 171. <https://doi.org/10.3390/w8050171>
- Tootchifatidehi, A. (2019). *Development of a global wetland map and application to describe hillslope hydrology in the ORCHIDEE land surface model* [Thèse de doctorat, Sorbonne Université]. <https://tel.archives-ouvertes.fr/tel-03020179>
- Tooth, S. (2000). Process, form and change in dryland rivers: a review of recent research. *Earth-Science Reviews*, 51(1-4), 67-107. [https://doi.org/10.1016/S0012-8252\(00\)00014-3](https://doi.org/10.1016/S0012-8252(00)00014-3)
- Trabucco, A. et Zomer, R. (2010). High-Resolution Global Soil-Water Balance explicit for Climate–Standard Vegetation and Soil Conditions. *CGIAR Consortium for Spatial Information*. <https://doi.org/10.6084/m9.figshare.7707605.v3>
- Trabucco, A. et Zomer, R. (2019). Global high-resolution soil-water balance. *Figshare. Fileset*, 10, m9. <https://doi.org/10.6084/m9.figshare.7707605.v3>
- Tramblay, Y., Rouché, N., Paturel, J.-E., Mahé, G., Boyer, J.-F., Amoussou, E., Bodian, A., Dacosta, H., Dakhlaoui, H., Dezetter, A., Hughes, D., Hanich, L., Peugeot, C., Tshimanga, R. et Lachassagne, P. (2021a). ADHI: the African Database of

- Hydrometric Indices (1950–2018). *Earth System Science Data*, 13(4), 1547-1560. <https://doi.org/10.5194/essd-13-1547-2021>
- Tramblay, Y., Rutkowska, A., Sauquet, E., Sefton, C., Laaha, G., Osuch, M., Albuquerque, T., Alves, M. H., Banasik, K., Beaufort, A., Brocca, L., Camici, S., Csabai, Z., Dakhlaoui, H., DeGirolamo, A. M., Dörflinger, G., Gallart, F., Gauster, T., Hanich, L., ... Datry, T. (2021b). Trends in flow intermittence for European rivers. *Hydrological Sciences Journal*, 66(1), 37-49. <https://doi.org/10.1080/02626667.2020.1849708>
- Trancoso, R., Phinn, S., McVicar, T. R., Larsen, J. R. et McAlpine, C. A. (2017). Regional variation in streamflow drivers across a continental climatic gradient. *Ecohydrology*, 10(3), e1816. <https://doi.org/10.1002/eco.1816>
- Tucker, G. E., Catani, F., Rinaldo, A. et Bras, R. L. (2001). Statistical analysis of drainage density from digital terrain data. *Geomorphology*, 36(3-4), 187-202. [https://doi.org/10.1016/S0169-555X\(00\)00056-8](https://doi.org/10.1016/S0169-555X(00)00056-8)
- Turner, D. S. et Richter, H. E. (2011). Wet/Dry Mapping: Using Citizen Scientists to Monitor the Extent of Perennial Surface Flow in Dryland Regions. *Environmental Management*, 47(3), 497-505. <https://doi.org/10.1007/s00267-010-9607-y>
- Tyralis, H., Papacharalampous, G. et Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, 11(5), 910. <https://doi.org/10.3390/w11050910>
- U.S. Geological Survey. (2000). HYDRO1k Elevation Derivative Database (Cent. for Earth Resour. Obs. and Sci., Sioux Falls, S. D.). <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-hydro1k>
- Uys, M. C. et O’Keeffe, J. H. (1997). Simple Words and Fuzzy Zones: Early Directions for Temporary River Research in South Africa. *Environmental Management*, 21(4), 517-531. <https://doi.org/10.1007/s002679900047>
- Vadher, A. N., Millett, J., Stubbington, R. et Wood, P. J. (2018). Drying duration and stream characteristics influence macroinvertebrate survivorship within the sediments of a temporary channel and exposed gravel bars of a connected perennial stream. *Hydrobiologia*, 814(1), 121-132. <https://doi.org/10.1007/s10750-018-3544-9>
- van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. et van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(1), 142. <https://doi.org/10.1186/1471-2164-7-142>
- Vardakas, L., Kalogianni, E., Economou, A. N., Koutsikos, N. et Skoulikidis, N. T. (2017). Mass mortalities and population recovery of an endemic fish assemblage in an intermittent river reach during drying and rewetting. *Fundamental and Applied Limnology*, 190(4), 331-347. <https://doi.org/10.1127/fal/2017/1056>
- Verdin Kristine. (2017). Hydrologic Derivatives for Modeling and Applications (HDMA) database: U.S. Geological Survey data release (USGS). <https://www.sciencebase.gov/catalog/item/591f1d4afe4b0ac16dbdde044>

- Vicente-Serrano, S. M., Beguería, S. et López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7), 1696-1718. <https://doi.org/10.1175/2009JCLI2909.1>
- Vidal-Abarca, M., Gómez, R., Sánchez-Montoya, M., Arce, M., Nicolás, N. et Suárez, M. (2020). Defining Dry Rivers as the Most Extreme Type of Non-Perennial Fluvial Ecosystems. *Sustainability*, 12(17), 7202. <https://doi.org/10.3390/su12177202>
- Vogt, J. V., Colombo, R. et Bertolo, F. (2003). Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics. *Geomorphology*, 53(3-4), 281-298. [https://doi.org/10.1016/S0169-555X\(02\)00319-7](https://doi.org/10.1016/S0169-555X(02)00319-7)
- Wainer, J. (2016). Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*. <https://doi.org/10.48550/arXiv.1606.00930>
- Wang-Erlandsson, L., Bastiaanssen, W. G., Gao, H., Jägermeyr, J., Senay, G. B., Van Dijk, A. I., Guerschman, J. P., Keys, P. W., Gordon, L. J. et Savenije, H. H. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrology and Earth System Sciences*, 20(4), 1459-1481. <https://doi.org/10.5194/hess-20-1459-2016>
- Ward, A. S., Wondzell, S. M., Schmadel, N. M. et Herzog, S. P. (2020). Climate Change Causes River Network Contraction and Disconnection in the H.J. Andrews Experimental Forest, Oregon, USA. *Frontiers in Water*, 2, 7. <https://doi.org/10.3389/frwa.2020.00007>
- Warix, S. R., Godsey, S. E., Lohse, K. A. et Hale, R. L. (2021). Influence of groundwater and topography on stream drying in semi-arid headwater streams. *Hydrological Processes*, 35(5), e14185. <https://doi.org/10.1002/hyp.14185>
- Wekesa, S. S., Stigter, T. Y., Olang, L. O., Oloo, F., Fouchy, K. et McClain, M. E. (2020). Water Flow Behavior and Storage Potential of the Semi-Arid Ephemeral River System in the Mara Basin of Kenya. *Frontiers in Environmental Science*, 8, 95. <https://doi.org/10.3389/fenvs.2020.00095>
- Wilding, N. A., White, J. C., Chadd, R. P., House, A. et Wood, P. J. (2018). The influence of flow permanence and drying pattern on macroinvertebrate biomonitoring tools used in the assessment of riverine ecosystems. *Ecological Indicators*, 85, 548-555. <https://doi.org/10.1016/j.ecolind.2017.10.059>
- Williams, D. D. (2006). *The biology of temporary waters*. Oxford University Press.
- Willmott, C. J. et Robeson, S. M. (1995). Climatologically aided interpolation (CAI) of terrestrial air temperature. *International Journal of Climatology*, 15(2), 221-229. <https://doi.org/10.1002/joc.3370150207>
- Willmott, C. J., Rowe, C. M. et Mintz, Y. (1985). Climatology of the terrestrial seasonal water cycle. *Journal of Climatology*, 5(6), 589-606. <https://doi.org/10.1002/joc.3370050602>

- Wirth, S., Carlier, C., Cochand, F., Hunkeler, D. et Brunner, P. (2020). Lithological and Tectonic Control on Groundwater Contribution to Stream Discharge During Low-Flow Conditions. *Water*, 12(3), 821. <https://doi.org/10.3390/w12030821>
- Wohl, E. (2017). The significance of small streams. *Frontiers of Earth Science*, 11(3), 447-456. <https://doi.org/10.1007/s11707-017-0647-y>
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S. et Bates, P. D. (2017). A high-accuracy map of global terrain elevations: Accurate Global Terrain Elevation map. *Geophysical Research Letters*, 44(11), 5844-5853. <https://doi.org/10.1002/2017GL072874>
- Yameogo, D. (1988). *Hydrogéologie des formations fissurées de la partie sud du plateau Mossi entre le Nazino et le Nakambe (région de Kombissiri-Manga), Burkina-Faso* [Thèse de doctorat, Université de Grenoble]. <https://tel.archives-ouvertes.fr/tel-00756428>
- Yan, Y., Lidberg, W., Tenenbaum, D. E. et Pilesjö, P. (2020). The accuracy of drainage network delineation as a function of environmental factors: A case study in Central and Northern Sweden. *Hydrological Processes*, 34(26), 5489-5504. <https://doi.org/10.1002/hyp.13963>
- Yang, J., Griffiths, J. et Zammit, C. (2019). National classification of surface-groundwater interaction using random forest machine learning technique. *River Research and Applications*, 35(7), 932-943. <https://doi.org/10.1002/rra.3449>
- Yonaba, R., Biao, A. C., Koïta, M., Tazen, F., Mounirou, L. A., Zouré, C. O., Queloz, P., Karambiri, H. et Yacouba, H. (2021a). A dynamic land use/land cover input helps in picturing the Sahelian paradox: Assessing variability and attribution of changes in surface runoff in a Sahelian watershed. *Science of The Total Environment*, 757, 143792. <https://doi.org/10.1016/j.scitotenv.2020.143792>
- Yonaba, R., Koïta, M., Mounirou, L., Tazen, F., Queloz, P., Biao, A., Niang, D., Zouré, C., Karambiri, H. et Yacouba, H. (2021b). Spatial and transient modelling of land use/land cover (LULC) dynamics in a Sahelian landscape under semi-arid climate in northern Burkina Faso. *Land Use Policy*, 103, 105305. <https://doi.org/10.1016/j.landusepol.2021.105305>
- Yu, S., Bond, N. R., Bunn, S. E. et Kennard, M. J. (2019). Development and Application of Predictive Models of Surface Water Extent to Identify Aquatic Refuges in Eastern Australian Temporary Stream Networks. *Water Resources Research*, 55(11), 9639-9655. <https://doi.org/10.1029/2019WR025216>
- Yu, S., Bond, N. R., Bunn, S. E., Xu, Z. et Kennard, M. J. (2018). Quantifying spatial and temporal patterns of flow intermittency using spatially contiguous runoff data. *Journal of Hydrology*, 559, 861-872. <https://doi.org/10.1016/j.jhydrol.2018.03.009>
- Zeng, M., Zou, B., Wei, F., Liu, X. et Wang, L. (2016). Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data (p. 225-228). 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS), IEEE. <https://doi.org/10.1109/ICOACS.2016.7563084>

- Ziegler, A. et König, I. R. (2014). Mining data with random forests: current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1), 55-63. <https://doi.org/10.1002/widm.1114>
- Zimmer, M. A., Kaiser, K. E., Blaszcak, J. R., Zipper, S. C., Hammond, J. C., Fritz, K. M., Costigan, K. H., Hosen, J., Godsey, S. E., Allen, G. H., Kampf, S., Burrows, R. M., Krabbenhoft, C. A., Dodds, W., Hale, R., Olden, J. D., Shanafield, M., DelVecchia, A. G., Ward, A. S., ... Allen, D. C. (2020). Zero or not? Causes and consequences of zero-flow stream gage readings. *WIREs Water*, 7(3). <https://doi.org/10.1002/wat2.1436>
- Zipper, S. C., Hammond, J. C., Shanafield, M., Zimmer, M., Datry, T., Jones, C. N., Kaiser, K. E., Godsey, S. E., Burrows, R. M. et Blaszcak, J. R. (2021). Pervasive changes in stream intermittency across the United States. *Environmental Research Letters*, 16(8), 084033. <https://doi.org/10.1088/1748-9326/ac14ec>
- Zouré, C., Queloz, P., Koïta, M., Niang, D., Fowé, T., Yonaba, R., Consuegra, D., Yacouba, H. et Karambiri, H. (2019). Modelling the water balance on farming practices at plot scale: Case study of Tougou watershed in Northern Burkina Faso. *CATENA*, 173, 59-70. <https://doi.org/10.1016/j.catena.2018.10.002>

## LISTE DES FIGURES

- Figure i-1:** Le Massili à Gonsé - Burkina Faso : Différentes périodes d'un cours d'eau intermittent, **a)** période sèche (13/04//2019) ; **b)** période humide (04/08/2019) – Coord. Géog. (-1,39 ; 12,48°). Crédit : Axel Patindé Belemtougri. .... 3
- Figure i-2:** Échelle spatio-temporelle (axe des abscisses) des contrôles météorologiques (**A**), géologiques (**B**) et de l'occupation des sols (**C**) sur le régime d'écoulement, avec les facteurs favorisant l'intermittence (flèches noires supérieures) ou l'écoulement pérenne (flèches grises inférieures) le long des axes des ordonnées, l'axe des y. Cette figure est extraite de (Costigan et al., 2016). .... 5
- Figure i-3:** Cours d'eau du bassin versant du Nakanbé à Wayen (Burkina Faso) extraits à partir d'un MNT (500m x 500m) issu de la base HydroSHEDS. **a)**  $A_{min} > 2,5 \text{ km}^2$  ; **b)**  $A_{min} > 25 \text{ km}^2$ . .... 8
- Figure I-1:** Distribution des densités de drainage ( $\text{km}/\text{km}^2$ ) calculées dans les mailles de 5' (environ 8km x 8km) par base de données collectées. Dans chaque boîte à moustache, le trait horizontal inférieur est le premier quartile qui représente la valeur correspondant à 25 % de la distribution des données tandis que les valeurs moyennes sont matérialisées par un carré au centre de chaque boîte à moustache, le second trait horizontal est la médiane qui est le deuxième quartile et qui représente la valeur correspondant à 50 % de la distribution des données. Le trait horizontal supérieur est le troisième quartile et représente la valeur correspondant à 75 % de la distribution des données, enfin les valeurs extrêmes sont matérialisées par des points. .... 17
- Figure I-2:** Représentation en Afrique, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) des bases de données. Les cartes sont affichées dans la projection WGS84, mais les différents calculs sont faits dans la projection Africa Albers Equal Area conic (code ESRI : 102022). .... 18
- Figure I-3 :** Distribution spatiale de l'intermittence estimée en Afrique par les bases de données et études **a)** AQUAMAPS (FAO, 2014), **b)** LCS (Schneider et al., 2017), **c)** GIRES (Messenger et al., 2021) estimation mensuelle avec les cours d'eau intermittents définis comme des cours d'eau s'arrêtant de s'écouler en moyenne un mois par an et inversement pour ceux définis permanents, **d)** GIRES (Messenger et al., 2021) estimation journalière avec les cours d'eau intermittents définis comme des cours d'eau s'arrêtant de s'écouler en moyenne un jour par an et inversement pour ceux définis permanents, **e)** Merit-hydro (Lin et al., 2021) proportion des cours d'eau intermittents (fi) par bassin versant. .... 23
- Figure I-4 :** Carte de la répartition des types de climats (période 1980-2016) selon la classification de Köppen-Geiger (Beck et al., 2018 ; Peel et al., 2007). Les contours des pays initialement ciblés pour la collecte des données hydrographiques sont matérialisés en noir... 25
- Figure I-5 :** Représentation spatiale de l'intermittence dans les bases hydrographiques de référence du **a)** Burkina Faso, **b)** Mali, **c)** Bénin, **d)** Madagascar, **e)** Afrique du Sud. .... 28
- Figure I-6 :** Variabilités de la densité de drainage observées ( $D_{d_{obs}}$ ) par pays où les données hydrographiques de références ont été collectées. La largeur des boîtes à moustaches varie en fonction du nombre d'observations. Les densités de drainage ont été calculées dans des mailles de 5' (environ 8km x 8km) par base de données collectées. .... 30



- Figure I-7** : Représentations au Bénin, des cours d'eau (à gauche) et densités de drainage (à droite) correspondantes de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Bénin sont faits dans la projection WGS84 UTM zone 31 N (code EPSG : 32631). ..... 36
- Figure I-8** : Représentations au Burkina Faso, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Burkina Faso sont faits dans la projection WGS84 UTM zone 30 N (code EPSG : 32630)..... 37
- Figure I-9** : Représentations au Mali, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, les différents calculs au Mali sont faits dans la projection WGS84 UTM zone 30 N (code EPSG : 32630). ..... 38
- Figure I-10** : Représentations à Madagascar, des cours d'eau (à gauche) et densités de drainage correspondantes (à droite) de quelques bases de données. Les cartes sont affichées dans la projection WGS84, mais les différents calculs sont faits dans la projection Tananarive / Laborde Grid (code EPSG : 29701). ..... 39
- Figure I-11** : Comparaison de la distribution spatiale des cartes nationales de l'intermittence du Bénin, Burkina Faso, Mali, Madagascar (Référence) avec celles des études (FAO, 2014 ; Lin et al., 2021 ; Messenger et al., 2021 ; Perez-Saez et al., 2017 ; Schneider et al., 2017). Dans l'étude de Perez-Saez et al. (2017), les cours d'eau classés « permanent, permanent low variability et permanent high variability » et ceux « mid ephemeral, strongly ephemeral » ont été regroupés respectivement dans cette représentation en « cours d'eau permanent » et « cours d'eau intermittent ». ..... 40
- Figure I-12**. Noms et localisations des stations hydrométriques au Burkina Faso (juin 2003). Source (extrait de l'annuaire hydrologique national 2011 du Burkina Faso, DGRE). ..... 43
- Figure I-13** : Aperçus sur les tronçons de rivières, chaque tronçon est représenté par une ligne d'une couleur différente (a), les points exutoires des tronçons de rivières (b). Les figures (a), (b) sont extraites de Linke et al. (2019). ..... 55
- Figure I-14** : Schéma d'un arbre de décision. .... 58
- Figure I-15** : (a) Découpage des données d'entrée en différents sous-espaces (type de fleur en fonction des variables Sepal.Length et Sepal.width), (b) arbre de décision résultant du découpage. .... 59
- Figure I-16** : (a) Exemple d'un arbre de régression entraîné pour prédire l'indice de débit de base. (b) Comparaison des observations de l'indice de débit de base avec les prédictions de l'arbre de régression présenté en (a) et d'une forêt aléatoire. La figure est extraite de Addor et al. (2018). ..... 60
- Figure I-17** : Fonctionnement de l'algorithme de Random Forest. La figure est adaptée de Boulesteix et al. (2012). ..... 61
- Figure I-18** : Répartition des données en ensembles d'apprentissages et de tests. .... 63
- Figure I-19** : Illustration d'une validation croisée à k blocs avec  $k = 4$ . Cette figure est adaptée de Ali et al. (2017). ..... 64
- Figure I-20** : Exemple de la représentation de l'importance des variables (% incMSE) dans un modèle random Forest en régression, calibré pour prédire le nombre de jours moyens sans débit. Cette figure est extraite de Zipper et al. (2021). ..... 67

**Figure I-21** : Exemple de la représentation de l'importance des variables (Mean Decrease Gini) dans un modèle Random Forest calibré pour prédire l'intermittence des cours d'eau (González-Ferreras et Barquín, 2017). ..... 67

**Figure II-1** : Hydroclimatic characteristics of the study area. Burkina Faso is shown in dark grey. **a)** The main catchments' boundaries are derived from the HydroSHEDS database (Lehner et al., 2008). The main rivers and water reservoirs (light blue lines) are derived from the Digital Chart of the World (DCW) (ESRI, 1993). Isohyets (red dashed lines) represent average annual rainfall (mm/yr) calculated from Climatic Research Unit (Harris et al., 2020) data at 0.5° spatial resolution over the period 1950-2018. The black and green dots give respectively the location of 49 streamflow gauges and 3 rain gauge stations provided by the hydrological and meteorological services; some gauges are superimposed in the representation due to their spatial proximity. **b)** Focus on the different watersheds and sub-watersheds of Volta (Mouhoun and Nakanbe) in Burkina Faso. .... 79

**Figure II-2:** Normal and drought period from 1954 to 2017. **a)** The Standardized Precipitation Index (SPI) was calculated considering three rain gauge stations located on the country's different climatic zones. The Dori rain gauge station (World Meteorological Organization Code: 1200010000) is located in the Sahelian zone; the Ouagadougou rain gauge station (WMO Code: 1200000100) is located in the Sudano-Sahelian zone; and the Bobo-Dioulasso rain gauge station (WMO Code: 1200004000) in the Sudanian zone. The SPI is calculated using the average annual precipitation of the three selected stations. The National Meteorology Agency (ANAM-BF) provided the meteorological data (Météo\_Burkina) in Burkina Faso. Other databases are considered, and precipitation values are extracted at the locations of the synoptic stations. CCKP data are taken from the Climate Change Knowledge Platform, <http://sdwebx.worldbank.org/climateportal/>; Climatic Research Unit (CRU) (Harris et al., 2014); TerraClimate (Abatzoglou et al., 2018). **b)** Distribution of the number of gauging stations available per year. .... 81

**Figure II-3** : Boxplot of the number of dry months per year during the normal and drought periods (1955-1969 and 1970-1985, respectively) at the 20 common gauging stations. The outliers of the boxplots are displayed in black dots and the gauges are ordered by watershed. .... 85

**Figure II-4** : Geographical variation of flow intermittency classes based on the mean number of dry months per year: **a)** over 1955-1985; **b)** over 1955-1969 (normal sub-period); **c)** over 1970-1985 (drought sub-period). In b and c, gauging stations that have a different intermittency class between both sub-periods are evidenced by crosses inside the circles. River line thickness is proportional to Strahler's order..... 88

**Figure II-5** : **a)** Proportion of gauging stations within large basins and flow intermittency classes over 1955-1985. **b)** Distribution of the number of dry months for each individual year of each station belonging to an intermittency class over 1955-1985. .... 89

**Figure II-6** : Percentage of dry months per class **(a)**, the long-term average of monthly discharge per class **(b)**, average precipitation regime pattern **(c)**, average potential evapotranspiration regime pattern **(d)**. The width of the graphic lines is proportional to the mean catchment area per class. The considered period is 1955-1985. .... 90

**Figure II-7** : Boxplot of estimated lag time value according to the intermittency classes over 1955-1985. The lag time (in month) is estimated for each catchment through a cross-correlation between the average monthly precipitation (P) and the average monthly discharge (Q). .... 91

- Figure II-8** : Biplot of the first two axes resulting from the PCA (n = 49 catchments, over 1955-1985) for significant environmental variables and the flow index (*Ndry*) introduced as a supplementary variable. Q1 to Q4 designates quadrants 1 to 4. The catchments are classified as perennial (blue dots), weakly intermittent classes (green dots), strongly intermittent classes (orange dots), and ephemeral classes (red dots). Legend: P - average precipitation (mm/yr), Area - catchment area (km<sup>2</sup>), K – permeability (m<sup>2</sup>), Strahler – Strahler order (-). Some catchments are superimposed in the representation due to their spatial proximity..... 93
- Figure II-9** : Comparisons of intermittency classes in Burkina Faso: **a)** The present study over 1955-1985, **b)** Adapted from (Perez-Saez et al., 2017). ..... 95
- Figure II-10** : The observed number of dry months per gauging stations over 1955-1985. The middle line represents the median in each boxplot, whereas the black dot represents the average value. The boxplot thickness is proportional to the number of observations. The gauged stations are ordered according to the increasing latitude. The trend curve (black line), with a 95% confidence interval (grey zone), is obtained by local weighted least-squares smoothing of the raw counts of dry months. According to the climatic zones, the boxplots' color is defined from the Sudanian zone in blue to the Sahelian zone in light green. The color of gauging station names depends on their intermittency class: perennial (blue), weakly intermittent (green), highly intermittent (orange), ephemeral (red). ..... 97
- Figure II-A-1** : **a)** Cumulative distribution of gauging stations in relation to the length of records. The majority of gauges (>50 %) have at least ten years of data. **b)** Comparing delineated catchment areas with catchment area information in the GRDC and DGRE databases (43/49 catchments). The consistency is satisfactory ( $r^2 = 0.99$ )..... 102
- Figure II-A-2** : Distribution of the 49 gauging stations studied against drainage area. The majority of catchments are large. .... 102
- Figure II-A-3** : Map of lithological formations of Burkina Faso and neighbouring countries. **a)** Lithological map of Hartmann and Moosdorf (2012). **b)** National lithological map. **c)** National lithological map reclassified according to Hartmann and Moosdorf (2012). In **c)** Outside the borders of Burkina Faso, lithology is represented by Hartmann and Moosdorf (2012) map at 1/3 750 000 scales. In Burkina Faso, lithological formations are represented by a national lithological map at the scale of 1/1 000 000 reclassified according to (Hartmann et Moosdorf, 2012) nomenclature to have nomenclature compatibility between the two databases. This allowed considering the catchment areas with parts outside of Burkina Faso. .... 103
- Figure II-A-4** : Percentage of area (%) occupied by the different types of lithologies per catchment. Legend: mt (metamorphic rocks), pb (basic plutonic rocks), pi (intermediate plutonic rocks), sc (carbonate sedimentary rocks), sm (mixed sedimentary rocks), ss (siliciclastic sedimentary rocks), su (unconsolidated sediments), va (acid volcanic rocks), vb (basic volcanic rocks). ..... 104
- Figure II-A-5** : Mean drainage density by type of lithology in Burkina-Faso (see lithology code description in **Figure II-A-3.c)**. ..... 106
- Figure II-A-6** : Estimated permeability map. For each catchment, the average permeability is obtained by weighting the permeabilities obtained for each lithology type (**Tableau II-A-1**) by the area occupied by each lithology type in the catchment. .... 107
- Figure II-A-7** : Hydrogeology maps showing the combination of aquifer type and aquifer productivity. The data are downloaded from Africa Groundwater Atlas at <https://www2.bgs.ac.uk/africagroundwateratlas/downloadGIS.html>..... 107

- Figure II-A-8** : Geographical variation of flow intermittency classes based on the mean number of dry months per year over 1955-1985. In this figure, the dams in operation over the period 1947-1985 (n=49) are represented by crosses (purple color); their size varies according to capacity (Mm<sup>3</sup>). The data are from the GRanD database (Lehner et al., 2011). ..... 108
- Figure II-A-9** : Pearson correlation matrix of the 11-initial set of variables. Legend: P (Precipitation), PET (Potential evapotranspiration), AET (Actual evapotranspiration), Aridity (Aridity index), Tmoy (Mean air temperature), Elevation, Topo\_wi (Topography wetness index), Area, Slope, Strahler (Strahler order), K (Permeability). The empty tiles refer to non-significant correlation values (p-value > 5%), whereas the displayed values are correlation coefficients found to be significant at the 5% level (p-value < 0.05). ..... 110
- Figure II-A-10** : Spatial representation of potential environmental variables of intermittence (see variables code description in ). Note that the Strahler order and the permeability (K) are respectively represented in **Figure II-4** in the main text and in **Figure II-A-8**. ..... 111
- Figure II-A-11** : Significant environmental variables contribution to principal component 1 (PC1) and principal component 2 (PC2). ..... 112
- Figure II-A-12** : Boxplots of normalized data. The boxplots show the distribution values of a) Area, b) Precipitation, c) Strahler order, and d) Permeability according to the different flow intermittent class. .... 113
- Figure II-A-13** : Boxplot of the number of dry months per year during the periods (1955-1985 and 1986-2017, respectively) at the 8 common gauging stations. The outliers of the boxplots are displayed in black dots, and the gauges are ordered by watershed. .... 113
- Figure II-A-14** : Matrice de confusion en calibration et validation pour les différents modèles testés au Burkina Faso. Notons que les cellules de la diagonale indiquent les cas où les classes sont correctement prédites. En régression les valeurs prédites sont numériques (nombre moyen de mois sec) et ont été converties en classe d'intermittence selon le Tableau II-1 pour générer les matrices de confusion correspondantes. .... 116
- Figure II-A-15** : Ordre d'importance des variables dans le modèle Random Forest en régression. K (Perméabilité moyenne amont du bassin), P (Précipitation moyenne annuelle amont du bassin), Area (Surface amont du bassin), Strahler (Ordre de Strahler). ..... 118
- Figure II-A-16** : Répartition spatiale des classes d'intermittences prédites au Burkina Faso sur le réseau LCS (ordre de Strahler 1-9) par le modèle RF en régression. .... 118
- Figure II-A-17** : Répartition spatiale des classes d'intermittences prédites au Burkina Faso sur le réseau LCS (ordre de Strahler 5-9) par le modèle RF en régression. .... 119
- Figure II-A-18** : Comparaison de la distribution spatiale de l'intermittence prédite par M2-LCS-BF et les études (Lin et al., 2021 ; Messenger et al., 2021 ; Perez-Saez et al., 2017) à la carte nationale de l'intermittence (IGB-BNDT) du Burkina Faso. .... 121
- Figure II-A-19** : Comparaison de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence BNDT du Burkina Faso aux emplacements des stations de jaugeages. **a**) matrice de confusion, **b**) distribution spatiale des classes d'intermittence, la première moitié du cercle correspond à la classification calculée aux stations et la seconde à la classification observée au niveau du réseau BNDT. .... 122
- Figure III-1** : Distribution du nombre de stations disponibles par an entre 1871 et 2016. La période d'étude finalement sélectionnée 1958-1991 est matérialisée en bleu. .... 130

- Figure III-2** : Comparaisons des surfaces des bassins versants récupérés d'études ou délimités par MNT avec celles reportées dans les métadonnées d'origine avec une différence de surface de moins de 15 %. Dans ce graphique, chaque colonne représente une étude ou donnée délimitée par MNT. En dernière colonne on peut lire par exemple que les surfaces de 633 bassins récupérés de l'étude de Do et al. (2018) ont pu être comparées avec les informations renseignées dans les métadonnées. .... 132
- Figure III-3** : Répartition spatiale des bassins versants étudiés (1125) à l'échelle de l'Afrique sur la période 1958-1991 **(a)**, pourcentage cumulé des stations en fonction de la surface amont des bassins versants **(b)**, distribution des stations étudiées en fonction de différentes classes de surface **(c)**. .... 133
- Figure III-4** : Répartition des stations de jaugeage par base de données sur la période 1958-1991 **(a)**, et par durée de chronique **(b)**. .... 134
- Figure III-5** : Pourcentage cumulé des stations en fonction de la durée de la chronique sur la période 1958-1991. .... 135
- Figure III-6** : Matrice de corrélation de Spearman des variables (11) dont la corrélation est supérieure à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ). .... 141
- Figure III-7** : Matrice de corrélation de Spearman des variables (15) dont la corrélation est inférieure ou égale à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ). .... 142
- Figure III-8** : Aperçu global de l'approche méthodologique développée dans cette étude. . 143
- Figure III-9** : Variation géographique des classes d'intermittence du débit (sur la base du nombre moyen de mois secs par an) pour 1125 stations avec des données mensuelles  $\geq 4$  ans. Les histogrammes en marge de la figure sur l'axe x et l'axe y représentent respectivement la distribution des stations selon la longitude et la latitude. .... 146
- Figure III-10** : Répartition de la surface amont des stations de jaugeage en fonction de leur classe d'intermittence **(a)**, évolution de la proportion d'intermittence sur la période 1958-1991 **(b)**. .... 147
- Figure III-11** : Mois moyen d'occurrence de l'intermittence **(a)** ainsi que la variabilité du mois d'occurrence de l'intermittence **(b)** à l'échelle de l'Afrique pour 532 stations comportant au moins un mois sec sur la période 1958-1991. .... 148
- Figure III-12** : Matrice de confusion en validation pour les différents modèles testés. Notons que les cellules de la diagonale indiquent les cas où les classes sont correctement prédites. 151
- Figure III-13** : Ordre d'importance des variables dans le modèle **a)** modèle 3, l'importance est appréciée par la décroissance moyenne de l'indice de Gini dans le modèle, **b)** modèle 4, l'importance est appréciée par l'augmentation de l'erreur quadrique moyenne (MSE) si la variable est supprimée du modèle. Les noms complets des variables sont définis dans le **Tableau III-2**. .... 152
- Figure III-14** : Graphiques de dépendance partielle pour les quatre variables les plus influentes du **modèle 3** en fonction des différentes classes d'intermittence, les graphiques sont classés par ordre d'importance des variables dans le modèle **a)** Arid, **b)** Surf, **c)** Etp, **d)** Sand, voir la description complète du code des variables au **Tableau III-2**. Le "tapis" en bas (couleur grise) montre la distribution des valeurs prises par les stations (901) pour chaque variable représentée. .... 153

- Figure III-15** : Graphiques de dépendance partielle pour les quatre variables les plus influentes du modèle 4, les graphiques sont classés par ordre d'importance des variables dans le **modèle 4 a)** Arid, **b)** Surf, **c)** Etp, **d)** Sand, voir la description complète du code des variables au **Tableau III-2**. Le "tapis" en bas (couleur grise) montre la distribution des valeurs prises par les stations (901) pour chaque variable représentée. .... 155
- Figure III-16** : Relation linéaire simple entre la moyenne du nombre de mois secs et l'indice d'aridité (Arid) pour les 901 stations de calibration. .... 156
- Figure III-17** : Distribution des variables hydro-environnementales pour 15 500 000 bassins versants amont associés aux tronçons de rivières (couleur violette) et 901 bassins versants amont associés aux stations de jaugeage utilisées en calibration des modèles. **a)** Surf, **b)** Pente, **c)** Alti, **d)** Topowi, **e)** Tci, **f)** Damstor, **g)** Etp, **h)** Tmoy. Les noms complets des variables sont définis dans le **Tableau III-2**. .... 158
- Figure III-18** : Distribution des variables hydro-environnementales pour 15 500 000 bassins versants amont associés aux tronçons de rivières (couleur violette) et 901 bassins versants amont associés aux stations de jaugeage utilisés en calibration des modèles. **a)** Nwet, **b)** Arid, **c)** Qspec, **d)** Sand, **e)** Grwstor, **f)** Wtd, **g)** Perm. Les noms complets des variables sont définis dans le **Tableau III-2**. .... 159
- Figure III-19** : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le **modèle 3**. **a)** tous les cours d'eau du réseau LCS, **b)** les cours d'eau prédits permanents, **c)** les cours d'eau prédits faiblement intermittents, **d)** les cours d'eau prédits éphémères. Les stations de jaugeages représentées en noir sur les cartes b-e sont les stations classées dans la classe d'intermittence de chaque carte. .... 161
- Figure III-20** : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le **modèle 3**. **a)** tous les cours d'eau drainant au moins 100 km<sup>2</sup>, **b)** tous les cours d'eau drainant au moins 6000 km<sup>2</sup>. .... 162
- Figure III-21** : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le **modèle 4**. **a)** tous les cours d'eau du réseau LCS, **b)** les cours d'eau prédits permanents, **c)** les cours d'eau prédits faiblement intermittents, **d)** les cours d'eau prédits éphémères. Les stations de jaugeages représentées en noir sur les cartes b-e sont les stations classées dans la classe d'intermittence de chaque carte. .... 163
- Figure III-22** : Distribution spatiale des classes d'intermittence prédites à l'échelle de l'Afrique par le **modèle 4**. **a)** tous les cours d'eau drainant au moins 100 km<sup>2</sup>, **b)** tous les cours d'eau drainant au moins 6000 km<sup>2</sup>. Les diagrammes en camembert détaillent les pourcentages en longueur de chaque classe d'intermittence. .... 164
- Figure III-23** : Relation linéaire du biais d'intermittence (%) et du biais de densité de drainage (km/km<sup>2</sup>) entre le réseau M3-LCS et les réseaux hydrographiques de référence du Bénin, Burkina Faso, Mali, Madagascar, Afrique du Sud. .... 169
- Figure III-24** : Corrélation (en % d'intermittence par maille) des prédictions M3-LCS et des études précédentes (Lin et al., 2021 ; Messenger et al., 2021) par rapport aux réseaux de références de quelques pays investigués. .... 170
- Figure III-25** : Comparaison de la distribution spatiale des cartes nationales de l'intermittence aux prédictions d'intermittence (M3-LCS) au Burkina Faso (**a,b,c,d**), Mali (**e,f,g,h**). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe. .... 171

- Figure III-26** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence IGM-Mali aux emplacements des stations de jaugeages. .... 172
- Figure III-27** : Comparaison de la distribution spatiale de la carte nationale de l'intermittence aux prédictions d'intermittence (M3-LCS) en Afrique du Sud (**a,b,c,d**). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères du modèle M3-LCS ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe. .... 173
- Figure III-28** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence DWS-Afrique du Sud aux emplacements des stations de jaugeages. .... 173
- Figure III-29** : Comparaison de la distribution spatiale des prédictions d'intermittence (M3-LCS et M3-LCS-100 km<sup>2</sup>) aux cartes nationales de l'intermittence au Bénin (**a,b,c**) et Madagascar (**d,e,f**). Pour faciliter les comparaisons, les classes faiblement intermittentes, fortement intermittentes et éphémères ont été regroupées en une seule classe intermittente et la classe permanente en une seule classe. .... 174
- Figure III-30** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence IGN-Bénin aux emplacements des stations de jaugeages. .... 175
- Figure III-31** : Matrice de confusion de la classification d'intermittence calculée au droit des stations de jaugeages à la classification d'intermittence du réseau de référence FTM-Madagascar aux emplacements des stations de jaugeages. .... 175
- Figure III-32** : Comparaison de la distribution spatiale des proportions d'intermittence par maille de 5' **a**) GIRES-Classe 2 (Messenger et al., 2021), **b**) Merit hydro (Lin et al., 2021), **c**) M3-LCS. Dans chacune des figures, les zones en « blanc » correspondant à des zones où il n'y pas de cours d'eau. .... 178
- Figure IV-1** : Localisation en Afrique et description des principaux pays investigués (Burkina Faso, Mali, Madagascar). .... 192
- Figure IV-2** : Processus de calibration des relations au Burkina Faso. (**a**) Représentation de la densité de drainage observée dans les mailles de 5'. (**b**) Zones de calibration et de validation. .... 194
- Figure IV-3** : Matrice de corrélation de Spearman des variables (9) dont la corrélation est inférieure ou égale à 0,75. Les corrélations dans les cases entièrement vides correspondent à des valeurs de corrélation non significatives (valeur  $p > 5\%$ ). Les noms complets des variables sont définis dans le **Tableau III-2**. .... 195
- Figure IV-4** : Distribution spatiale (au Burkina Faso, Mali et Madagascar) des variables environnementales non redondantes (**Figure IV-3**) utilisées pour estimer la densité de drainage. Les noms complets des variables sont définis dans le **Tableau III-2**. .... 196
- Figure IV-5** : Résultats de la première relation au Burkina Faso, relation entre les valeurs calibrées de  $A_{\min}$  et les valeurs de  $Dd_{\text{obs}}$  dans les mailles de 5', avec les axes en logarithme. 198
- Figure IV-6** : Régression linéaire entre la densité de drainage observée ( $Dd_{\text{obs}}$ ) et les variables environnementales. Pour la signification des codes des variables environnementales, voir le **Tableau III-2**. .... 200

**Figure IV-7** : Critères de performance en **a)** calibration et **b)** validation du modèle sur l'ensemble du jeu de données (Burkina Faso, Mali et Madagascar) pour l'estimation de la densité de drainage. .... 201

**Figure IV-8** : Importance des variables dans le modèle Random Forest appliqué en régression pour estimer la densité de drainage. L'importance est appréciée par l'augmentation de l'erreur quadrique moyenne (MSE) si la variable est supprimée du modèle. Les noms complets des variables sont définis dans le **Tableau III-2**. .... 202

**Figure IV-9** : Densités de drainage correspondantes aux cours d'eau extraits à partir d'aires contributives variables en Afrique. .... 203

**Figure IV-10** : Distribution des densités de drainage ( $\text{km}^3/\text{km}^2$ ) des principales bases de données qui représentent les cours d'eau en Afrique. Pour une description détaillée de ces bases voir Tableau I-1 et Figure I-1. .... 204



## LISTE DES TABLEAUX

<b>Tableau I-1:</b> Description des bases de données hydrographiques principales couvrant l’Afrique, les densités de drainages (Dd) ainsi que l’écart-type (Sd) sont calculés dans des mailles de 5’ à l’échelle de l’Afrique. Les bases de données sont classées par ordre chronologique croissantes. ....	15
<b>Tableau I-2 :</b> Classification du régime d’intermittence basée sur l’ordre de Strahler et l’indice d’aridité (FAO, 2014). Le terme P pour « Permanent » et le terme I pour « Intermittent ». On peut lire par exemple que tous les cours d’eau d’ordre de Strahler 1 ayant un indice d’aridité inférieur à 0,35 sont classés intermittents et inversement pour les cours d’eau permanents. ..	19
<b>Tableau I-3 :</b> Définition et quantification de l’intermittence dans quelques études qui proposent une représentation du régime d’intermittence des cours d’eau en Afrique. ....	24
<b>Tableau I-4 :</b> Performances des bases de données par rapport au réseau hydrographique de référence du Bénin, Burkina Faso, Mali, Madagascar. Les paramètres s’appliquent pour les différents réseaux sur les densités de drainage (km/km <sup>2</sup> ) synthétisées dans les mailles de 5’ (environ 8 km x 8 km). Les valeurs de corrélation et Rmse affichées sont celles entre le réseau de référence et les réseaux comparés. ....	34
<b>Tableau I-5 :</b> Paramètres statistiques de comparaison de l’intermittence estimée en Afrique par des études par rapport au réseau de référence du Bénin, Burkina Faso, Mali, Madagascar et Afrique du Sud. Les paramètres s’appliquent pour les différents réseaux sur le pourcentage d’intermittence (%) synthétisé dans les mailles de 5’ (environ 8 km x 8 km). ....	35
<b>Tableau I-6 :</b> Comparaison des paramètres statistiques de validation pour TerraClimate et CRU Ts4.0 pour la température moyenne annuelle (Tas), les précipitations cumulées (Pr), l’évapotranspiration de référence (ET0) et le débit cumulé (Q). La valeur médiane des statistiques de validation pour le coefficient de corrélation de Pearson ( $\rho$ ) et l’erreur absolue moyenne (MAE) est indiquée. L’erreur absolue moyenne relative est indiquée pour Pr, ET0 et Q, l’erreur absolue moyenne étant indiquée entre parenthèses. Cette illustration est extraite de (Abatzoglou et al., 2018). ....	46
<b>Tableau I-7 :</b> Exemple de matrice de confusion pour une classification multiclasse. Le calcul des faux positifs (FP), faux négatif (FN), precision, recall de chaque classe ainsi que l’accuracy globale sont détaillés. ....	69
<b>Tableau II-1 :</b> Flow intermittency classes and their description. ....	83
<b>Tableau II-2 :</b> Set of environmental variables considered in this study as potential flow intermittency controls. In bold, the independent variables selected for further analyses. ....	86
<b>Tableau II-A-1 :</b> Permeability estimated values for all lithological units in the study area. .	105
<b>Tableau II-A-2 :</b> Classification in terms of intermittency of gauges based on monthly versus daily data. ....	109
<b>Tableau II-A-3 :</b> Eigenvalues for the four principal components. Only the first two components have an eigenvalue greater than 1. They are retained for interpretation. ....	112
<b>Tableau II-A-4 :</b> Correlations between variables and factors. <i>Ndry</i> added to PCA as a supplementary variable shows a high negative correlation with PC2. ....	112
<b>Tableau II-A-5 :</b> Critères de performances obtenus en calibration et validation pour les différents modèles de RF explorés pour prédire l’intermittence au Burkina Faso. ....	116

**Tableau II-A-6** : Matrice de coût utilisée comme pondération pour le calcul du critère de Kappa. .... 117

**Tableau II-A-7** : Performances du modèle M2-LCS par rapport au réseau hydrographique de référence du Burkina Faso. Les paramètres s'appliquent pour les différents réseaux sur les densités de drainage ( $\text{km}/\text{km}^2$ ) synthétisées dans les mailles de 5' (environ 8 km x 8 km). . 120

**Tableau III-1** : Les classes d'intermittence des cours d'eau et leur description. .... 136

**Tableau III-2** : Liste des variables hydro-environnementales (23) examinées comme contrôle de l'intermittence. En gras, les variables indépendantes sélectionnées pour les analyses ultérieures. .... 138

**Tableau III-3** : Récapitulatif des différents modèles Random Forest explorés ..... 144

**Tableau III-4** : Critères de performance obtenus en calibration et validation pour les différents modèles de RF explorés. Les critères de performances en calibration du modèle 3 ne sont pas comparables aux modèles 1 et 2 parce que ce n'est pas le même échantillon. ... 150

**Tableau III-5** : Paramètres statistiques de comparaison des prédictions d'intermittence du modèle 3 appliqué au réseau LCS (M3-LCS) par rapport aux réseaux de référence du Burkina Faso, Mali, Bénin, Madagascar et Afrique du Sud. .... 168

**Tableau III-6** : Paramètres statistiques de comparaison en Afrique des prédictions d'intermittence du modèle 3 appliqué au réseau LCS (M3-LCS) par rapport aux prédictions d'intermittence de l'étude de Messenger et al. (2021) sur le réseau hydroRIVERS (base GIRES) et de Lin et al. (2021). .... 177

**Tableau III-7** : Facteurs de contrôles de l'intermittence identifiés dans quelques études à travers le monde. Les études sont classées par ordre croissant de l'échelle du bassin versant à l'échelle globale. .... 180

**Tableau IV-1** : Performance de la validation de la première étape du modèle au Burkina Faso (14% du territoire), au Mali et à Madagascar..... 199

**Tableau IV-2** : Relation entre  $A_{\min}$  et  $Dd_{\text{obs}}$  dans différents pays et à différentes échelles. . 199

**Tableau IV-3** : Facteurs de contrôles de la densité de drainage identifiés dans quelques études à travers le monde. Les études sont classées par date de publication. .... 202

## **ANNEXES**

## Annexe - 1 - Représentation spatiale des variables hydro-environnementales investiguées

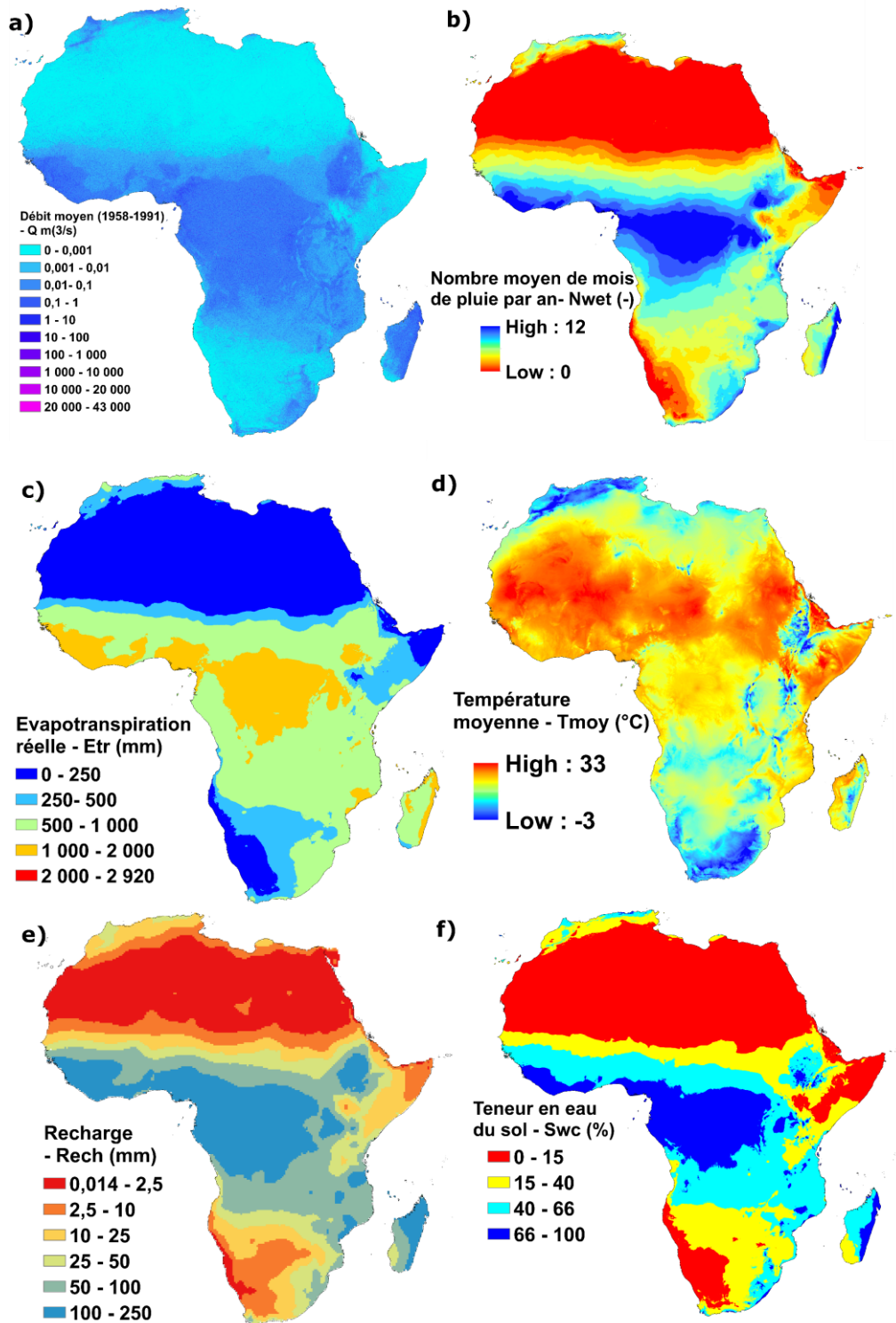


Figure A1-1 : Représentation spatiale des variables hydro-climatiques investiguées comme facteurs potentiels de l'intermittence en Afrique sur la période 1958-1991.

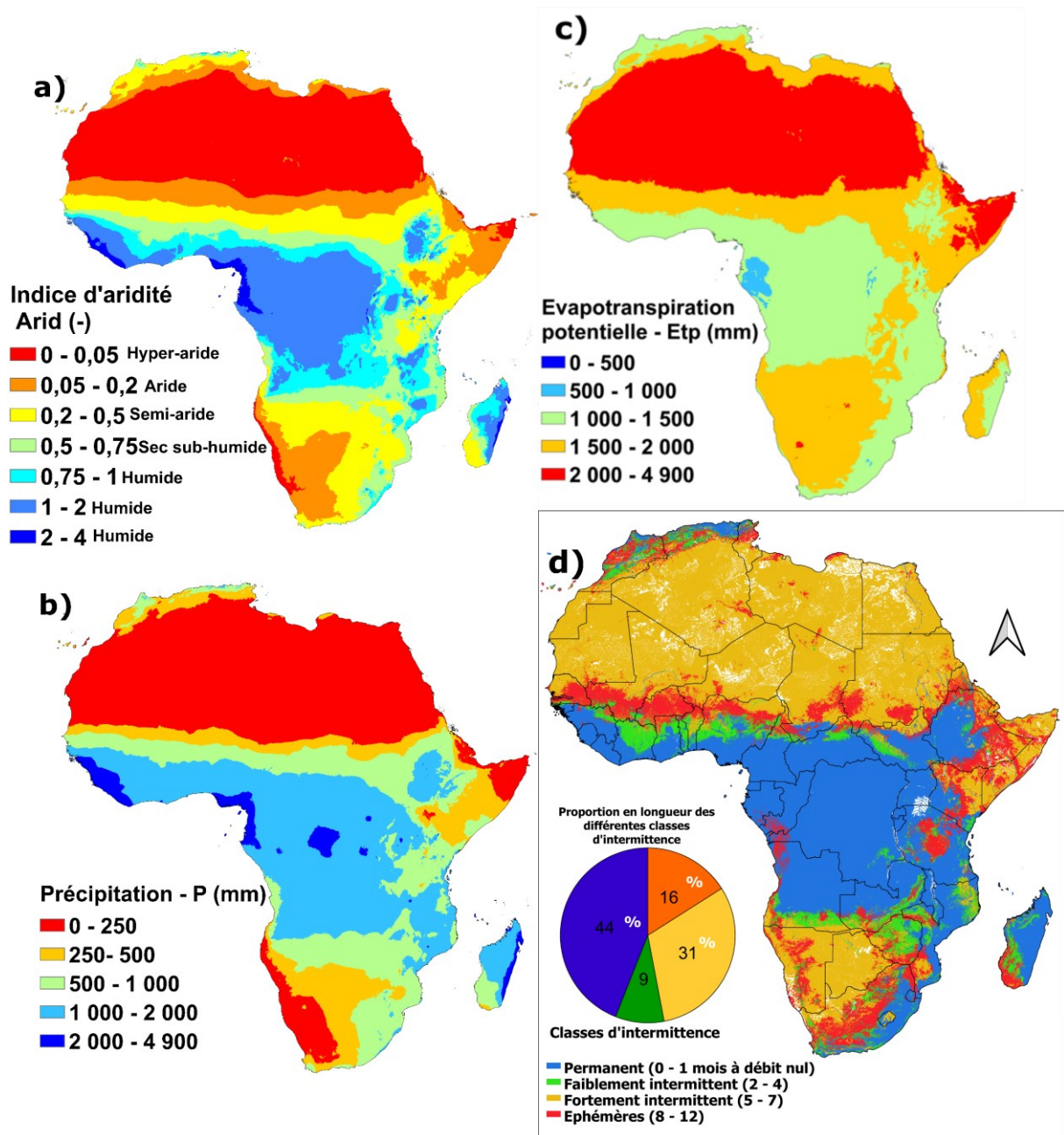


Figure A1-2 : Représentation spatiale des variables climatiques investiguées comme facteurs potentiels de l'intermittence en Afrique sur la période 1958-1991 (a-c) et des classes d'intermittence prédites par le modèle 3 (d). Les classes d'aridité sont définies en fonction de la classification de l'UNEP (Boschetto et al., 2010).

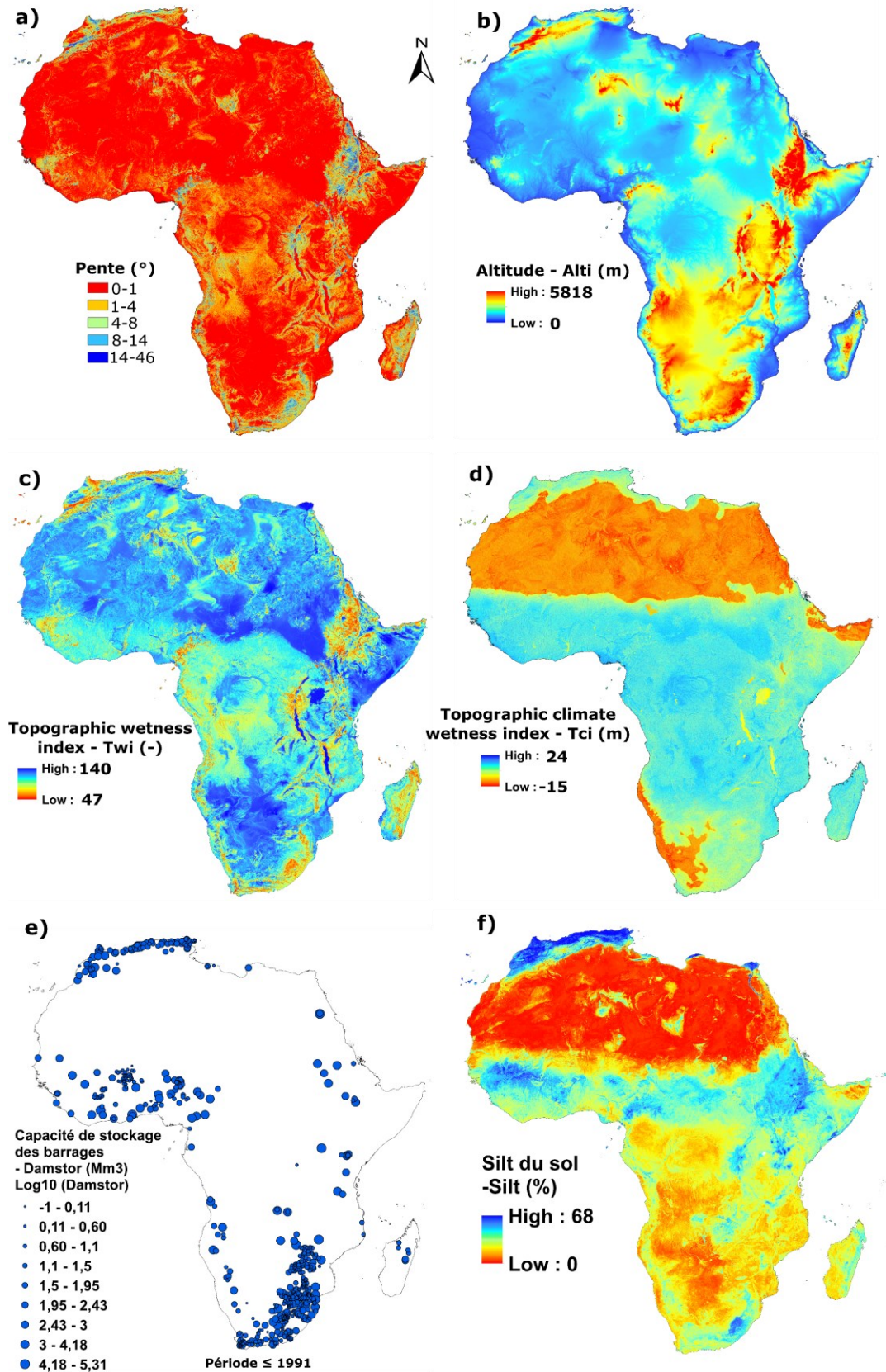


Figure A1-3 : Représentation spatiale des variables liées à la topographie, capacité de stockage des barrages, géologie et propriétés des sols investigués comme facteurs potentiels de l'intermittence en Afrique.

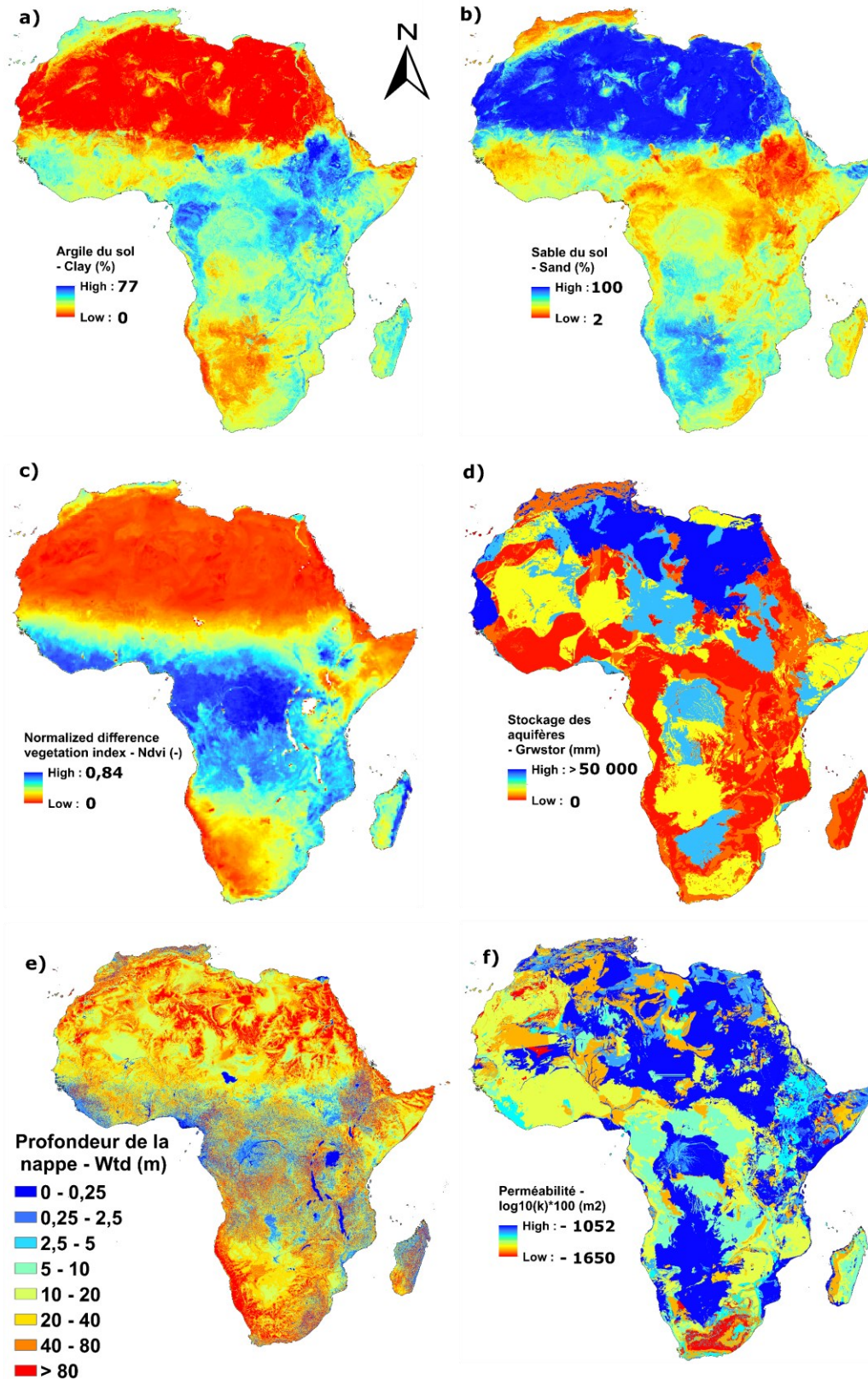


Figure A1-4 : Représentation spatiale des variables liées à la géologie et propriétés des sols investiguées comme facteurs potentiels de l'intermittence en Afrique.

## Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso

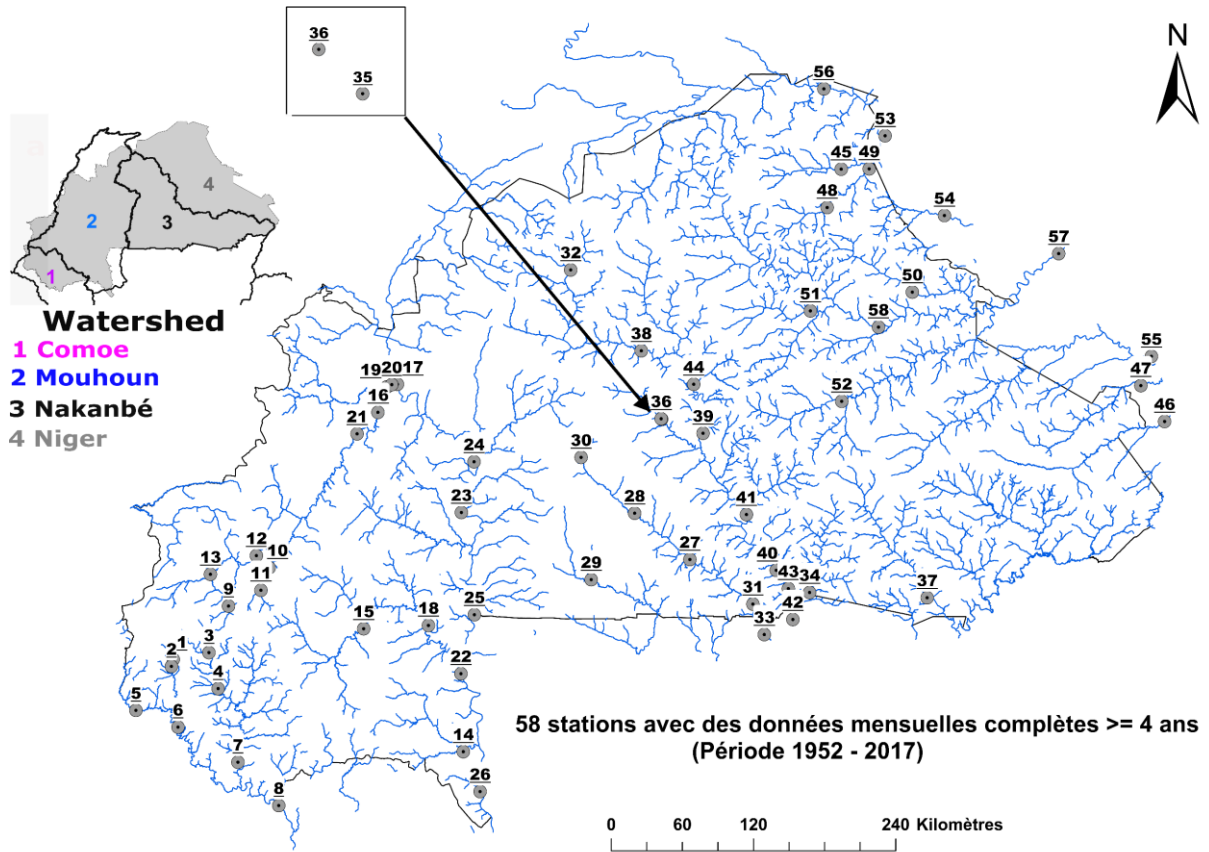
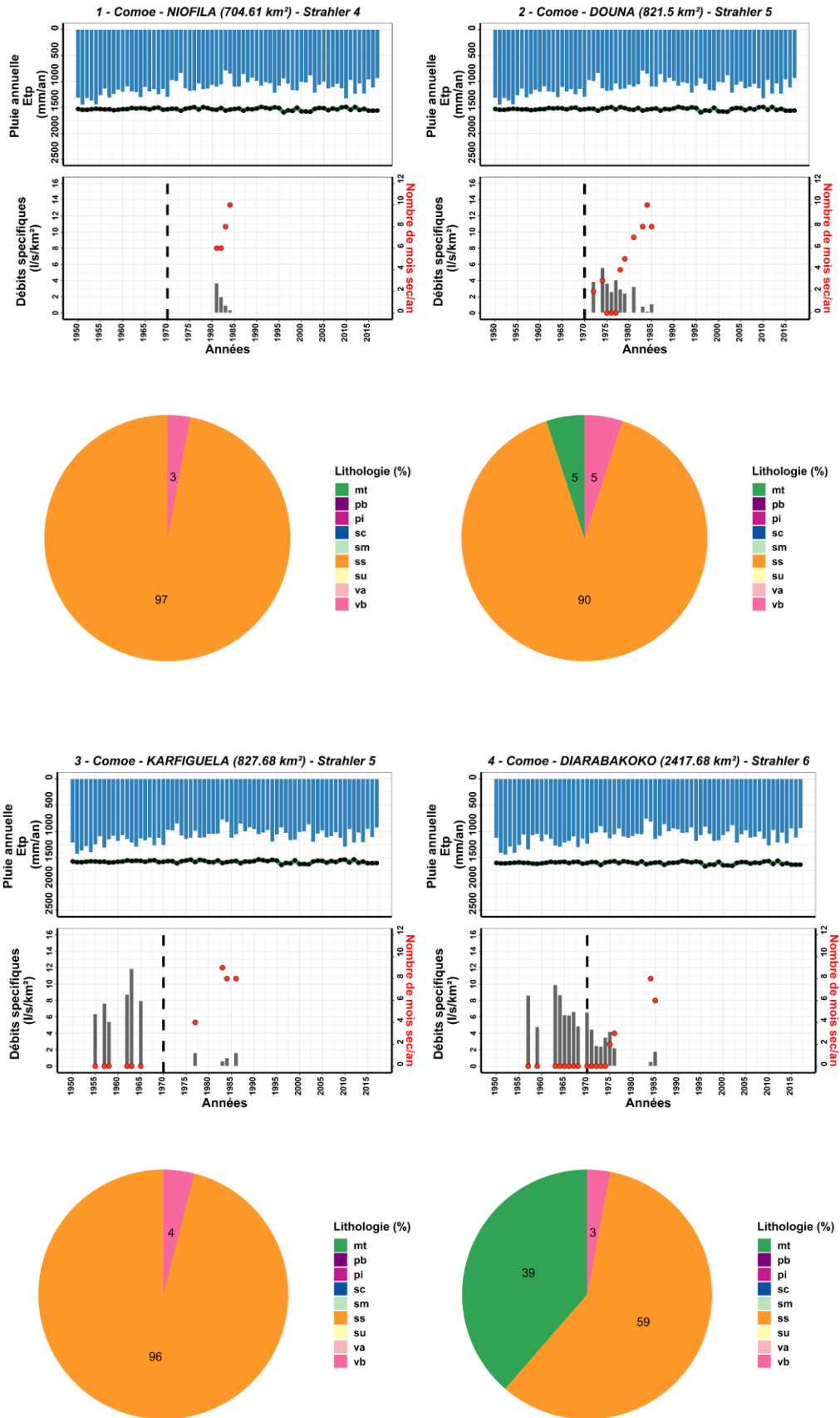


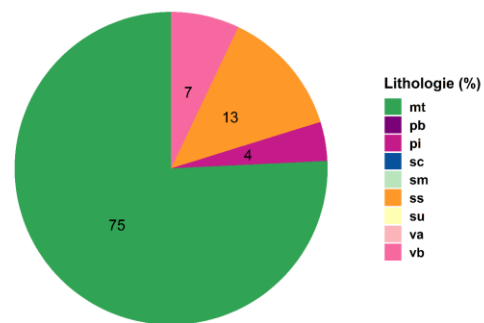
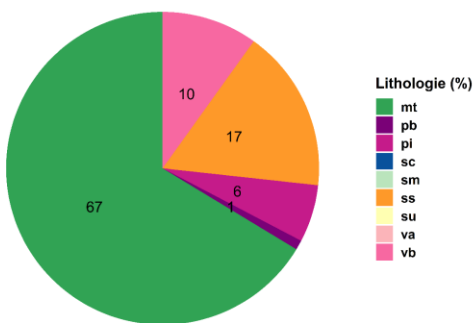
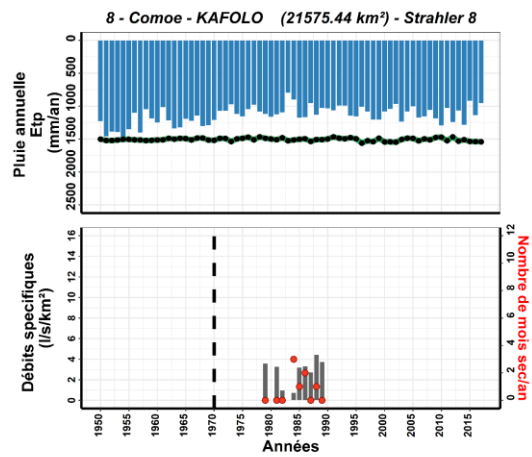
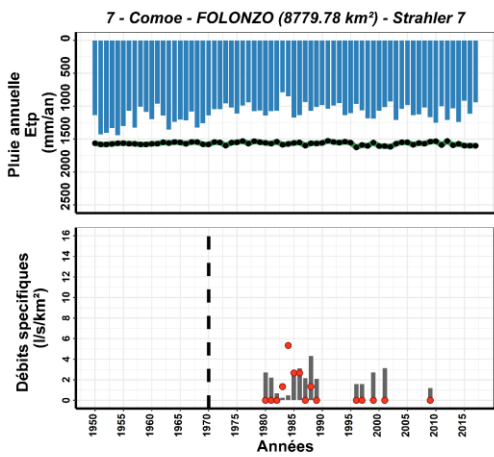
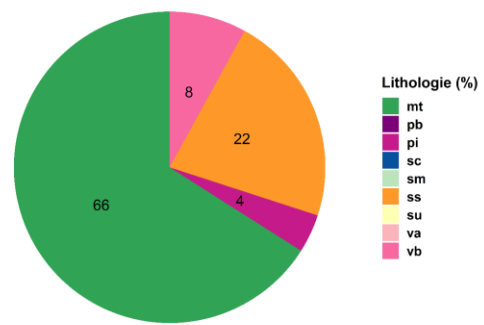
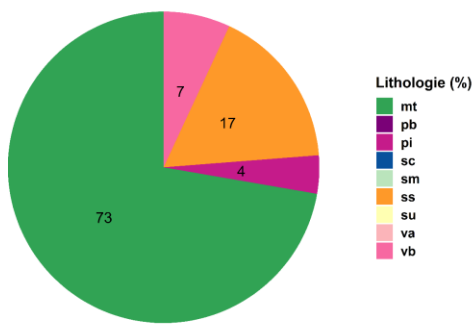
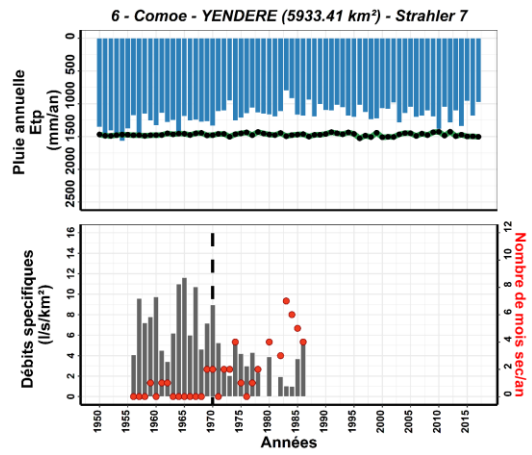
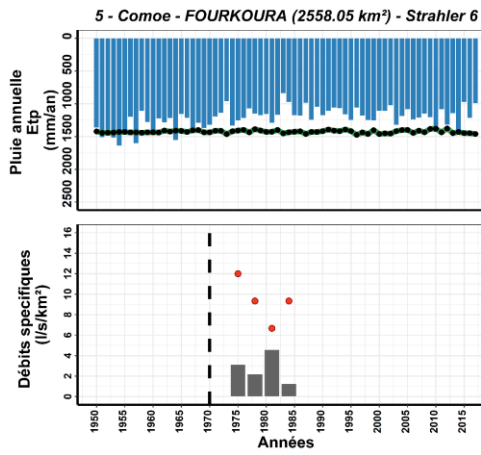
Figure A2-1 : Nomenclature des principaux cours d'eau jaugés investigués au Burkina Faso.



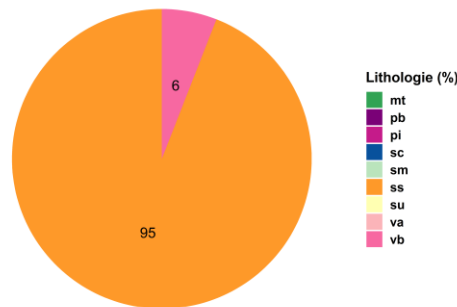
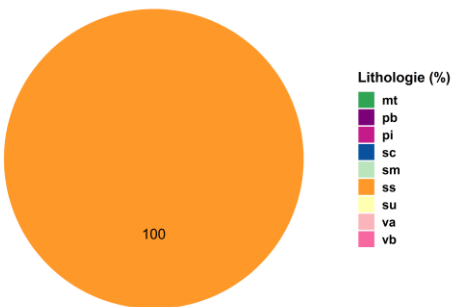
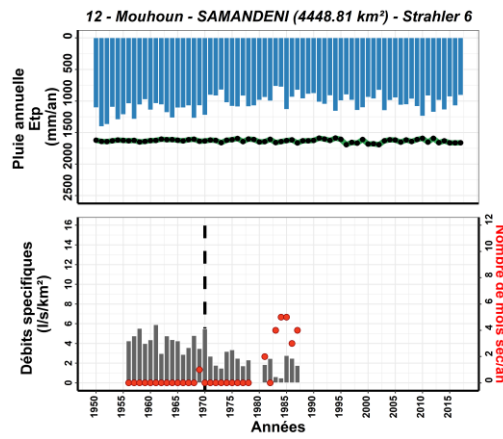
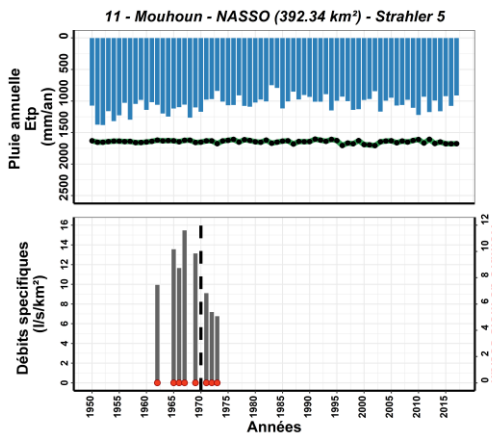
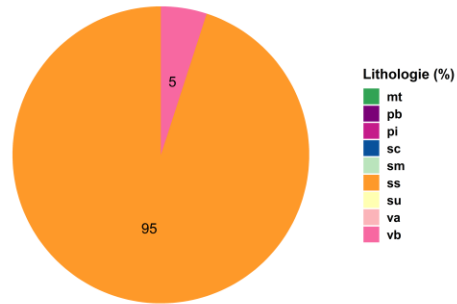
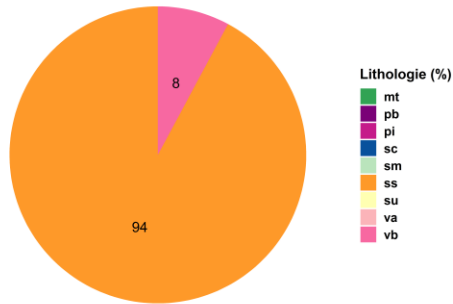
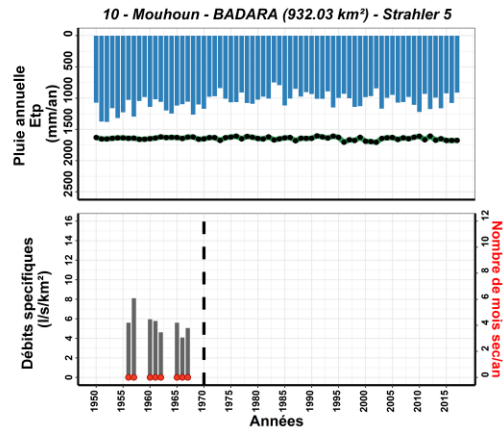
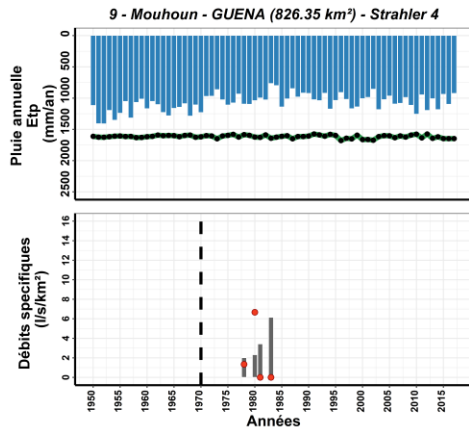
## Bassin de la Comoé



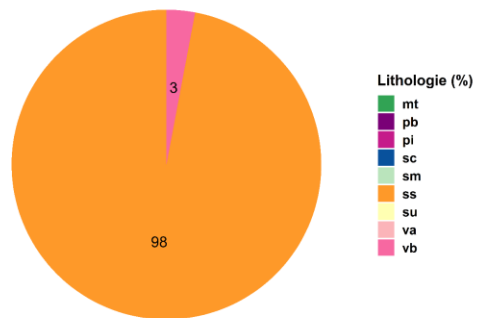
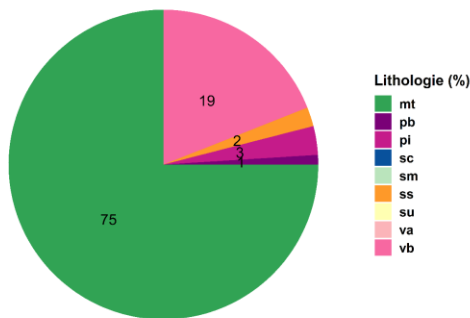
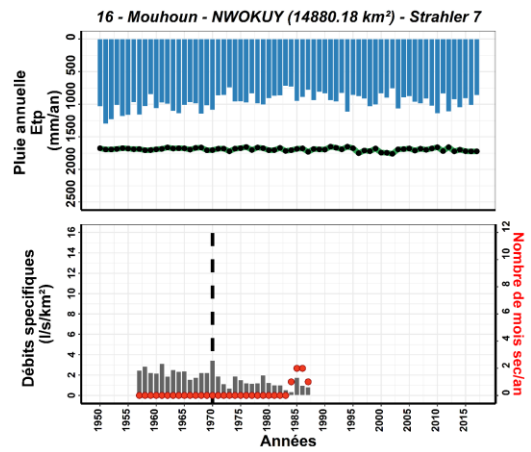
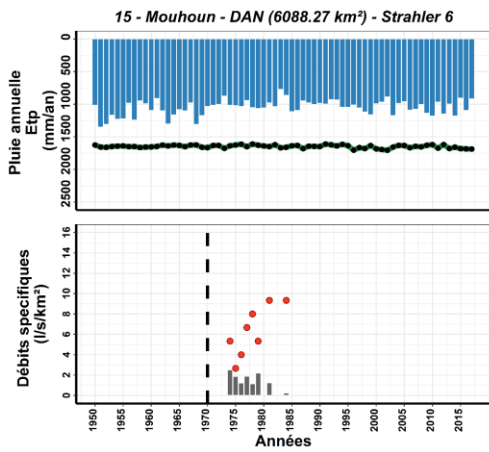
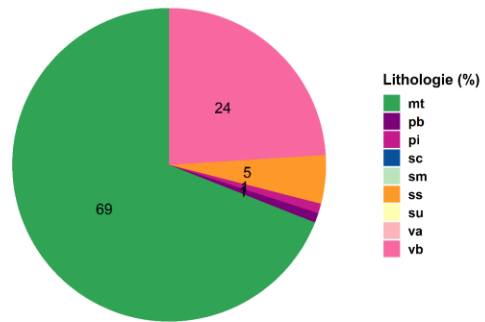
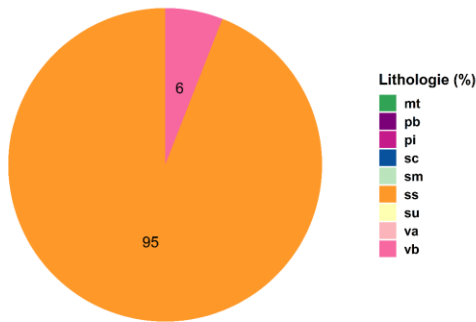
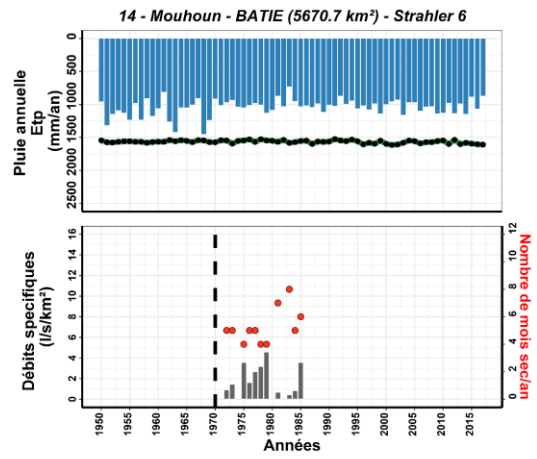
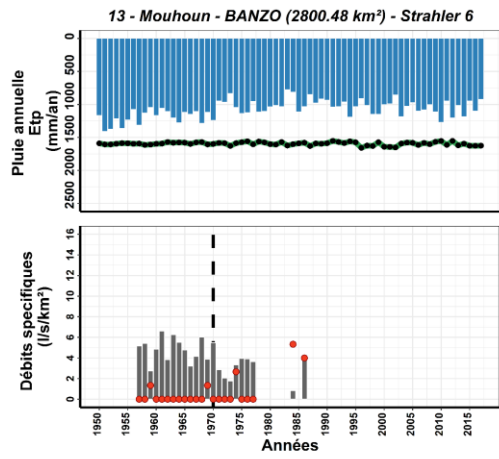
Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso



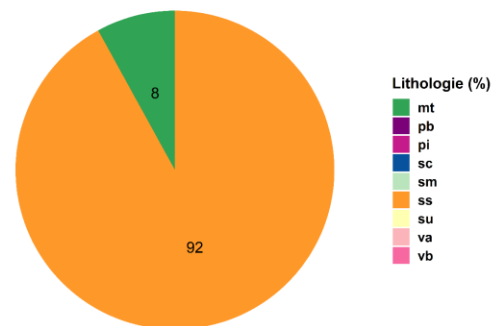
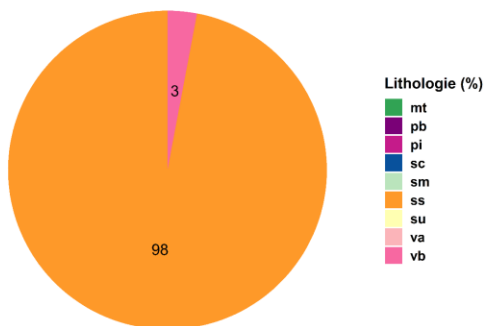
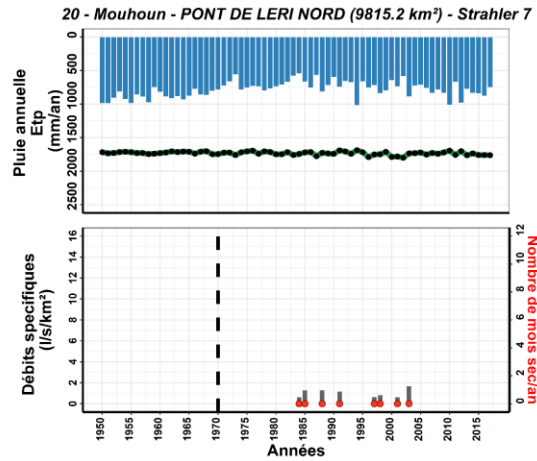
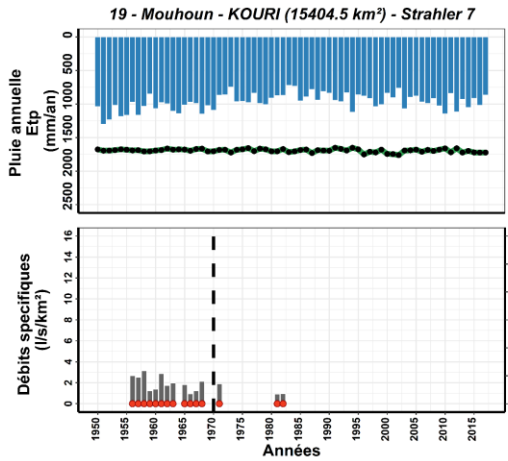
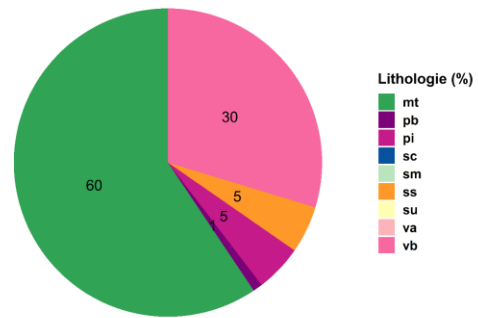
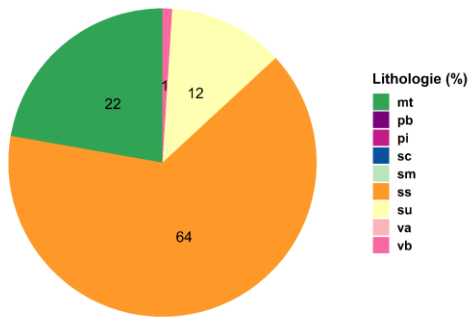
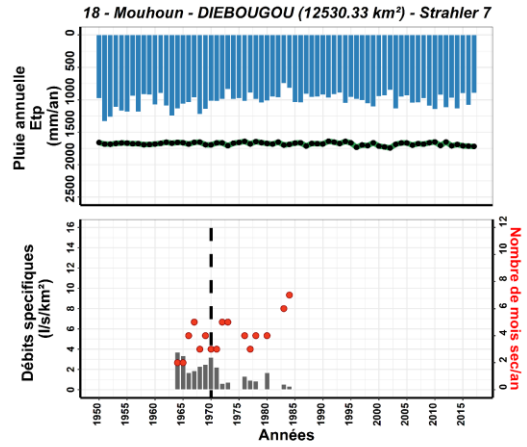
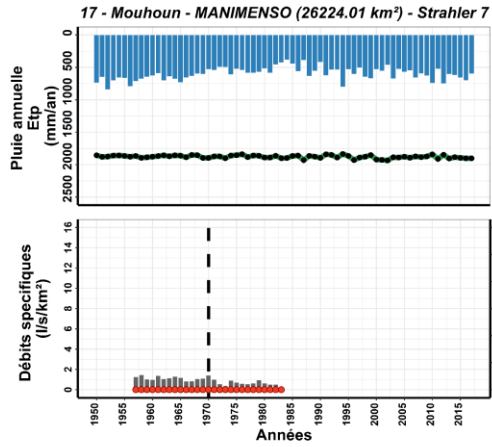
## Bassin du Mouhoun

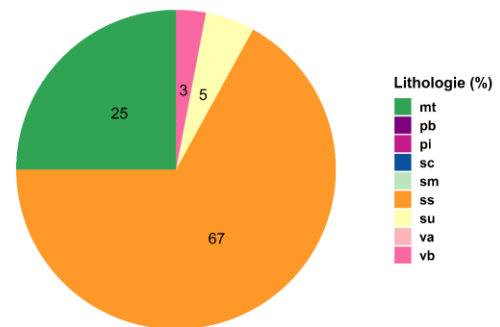
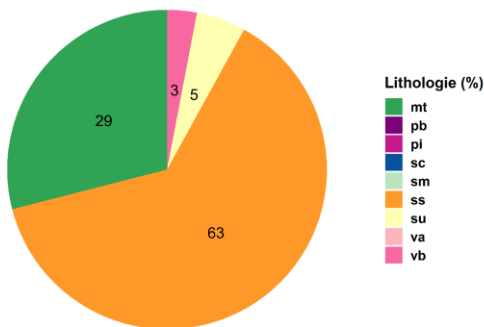
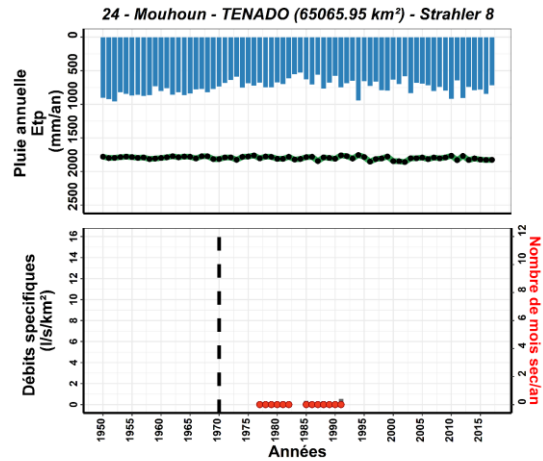
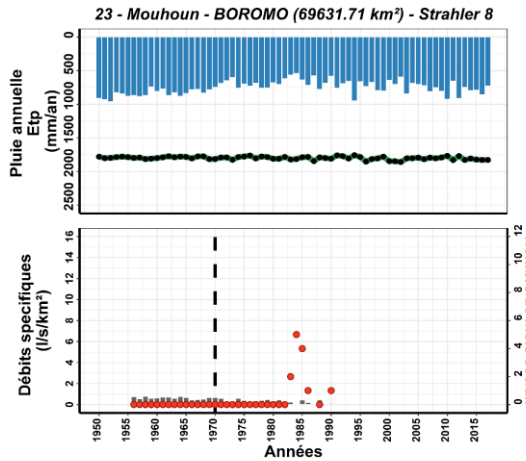
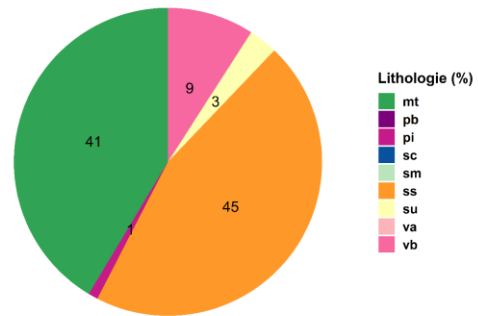
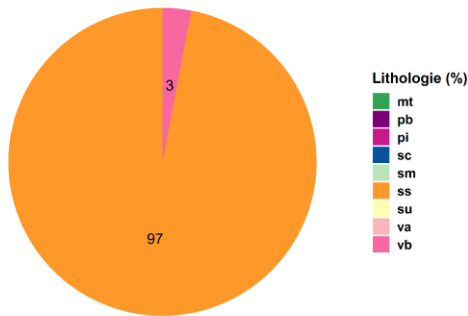
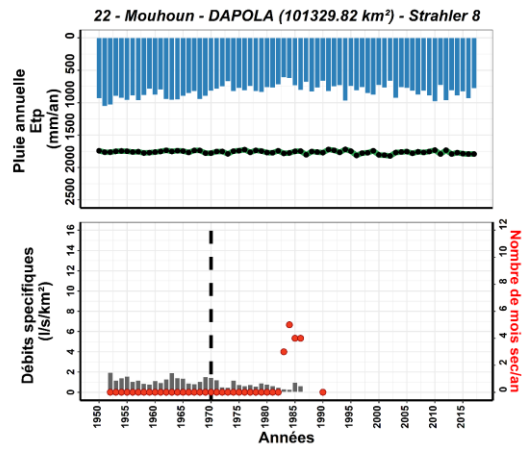
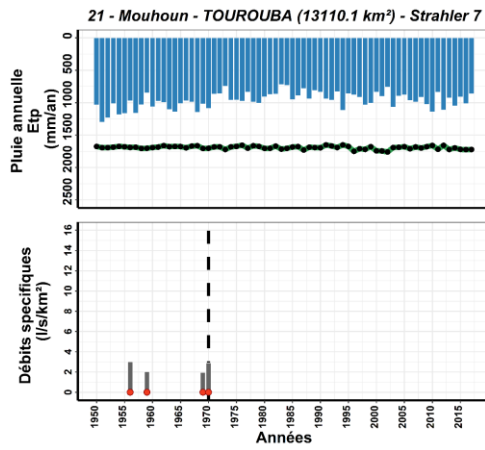


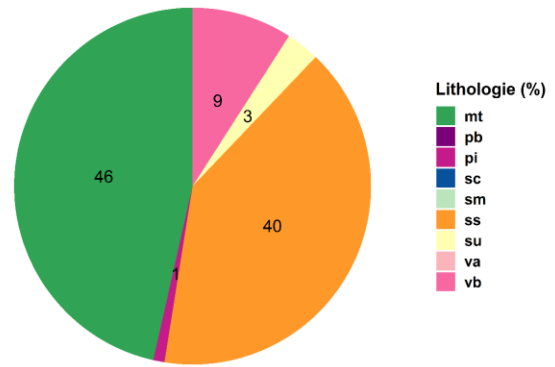
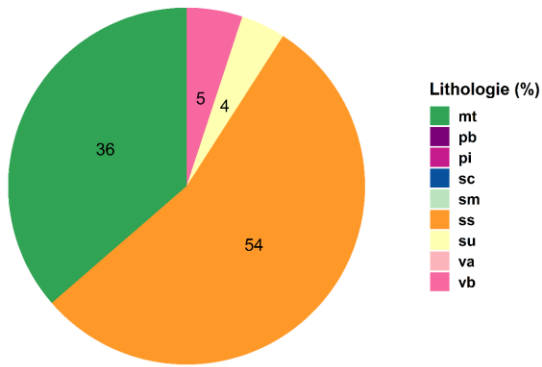
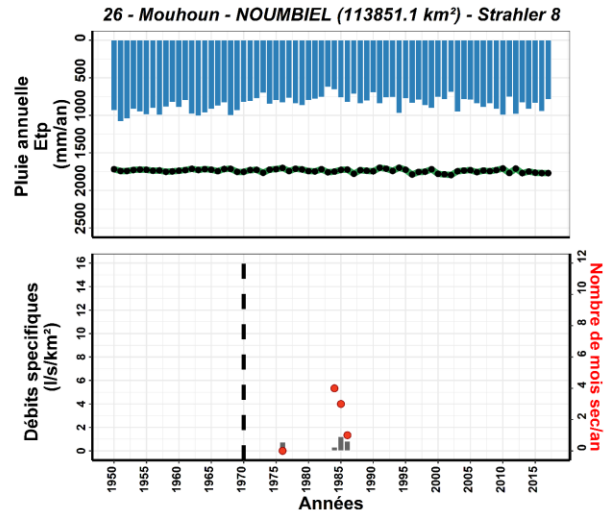
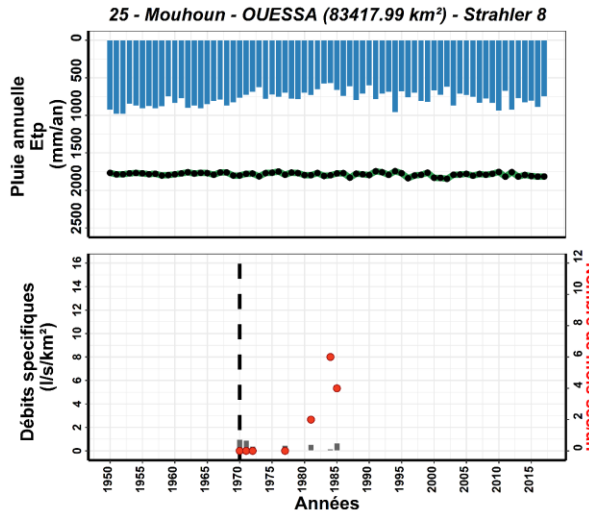
Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso



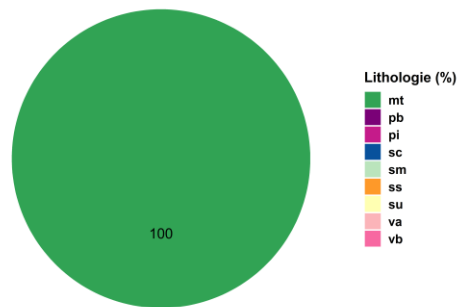
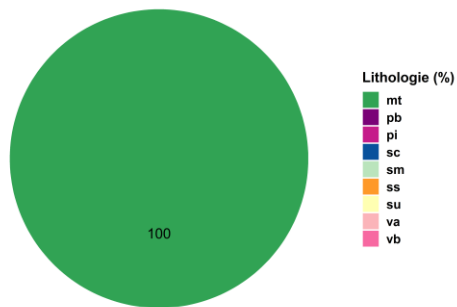
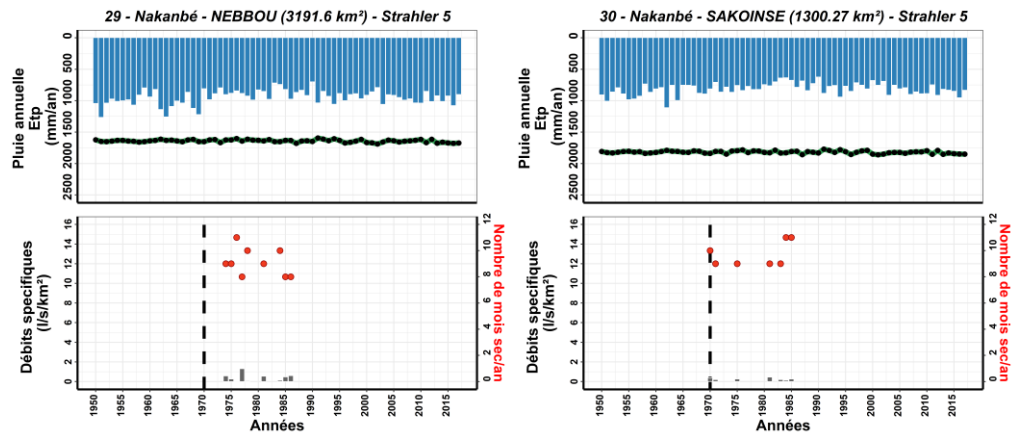
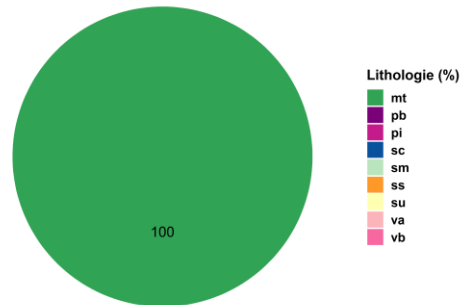
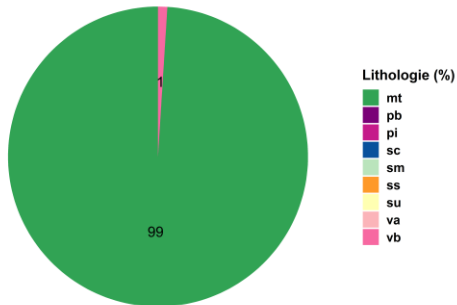
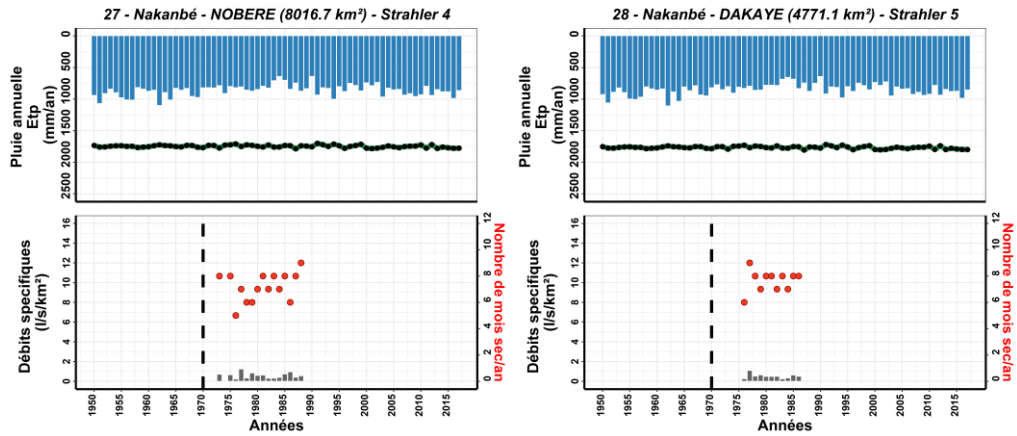
Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso





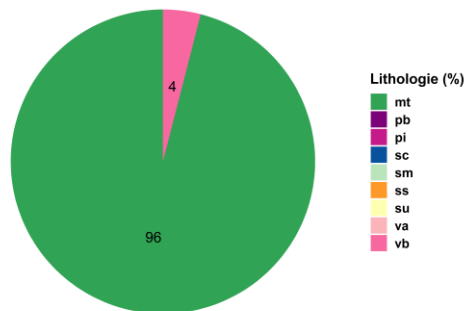
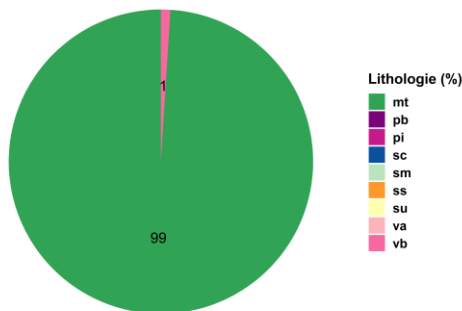
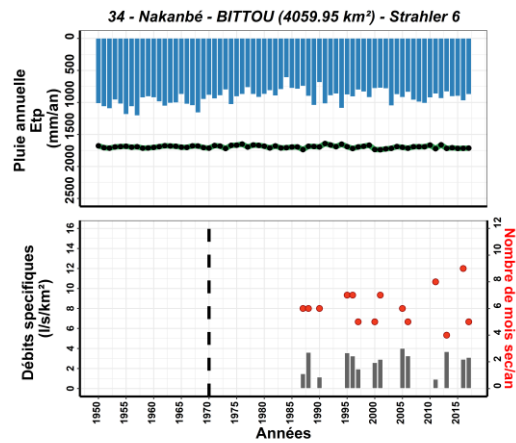
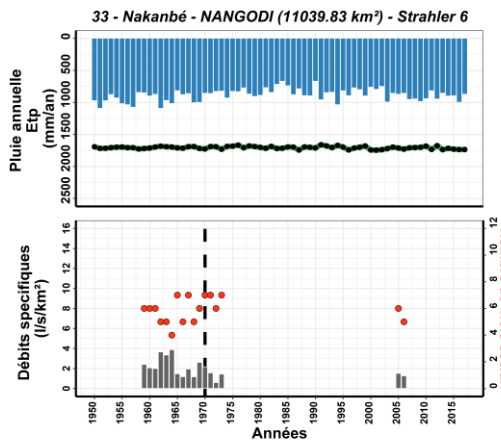
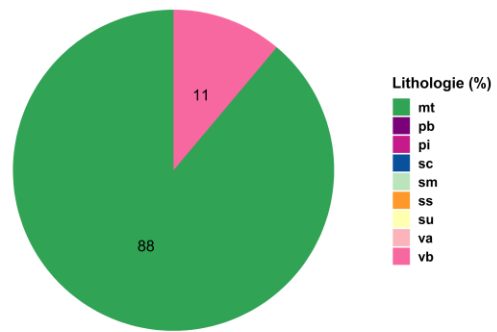
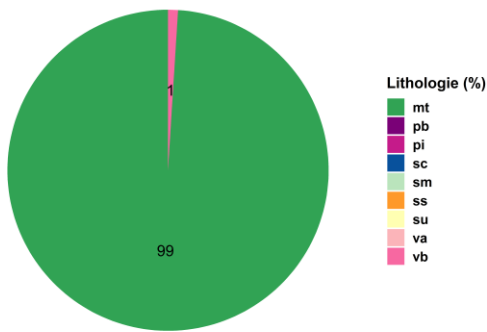
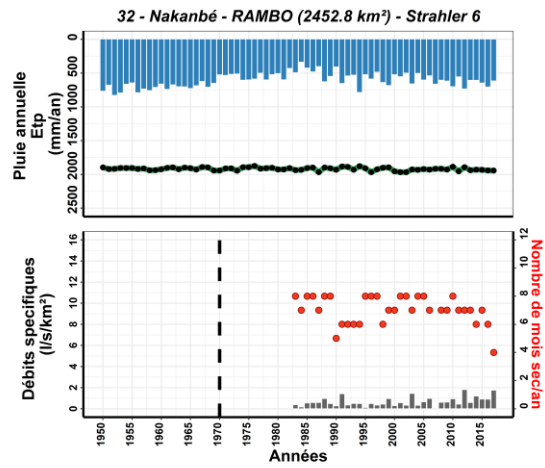
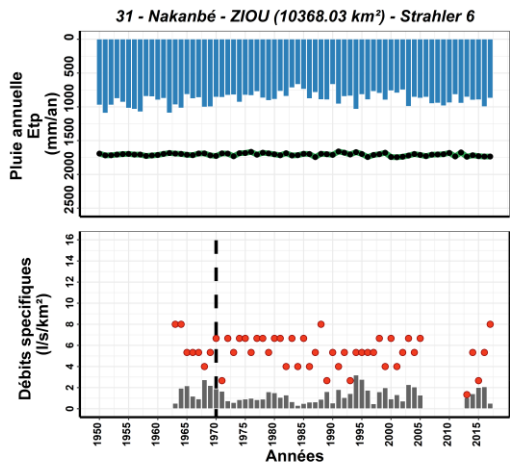


## Bassin du Nakanbé

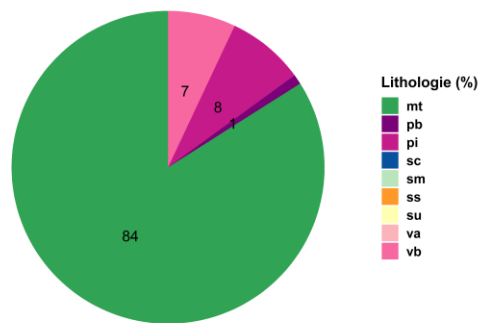
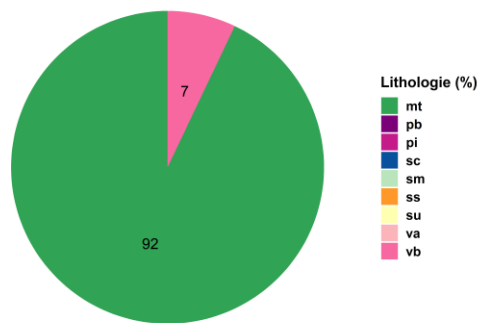
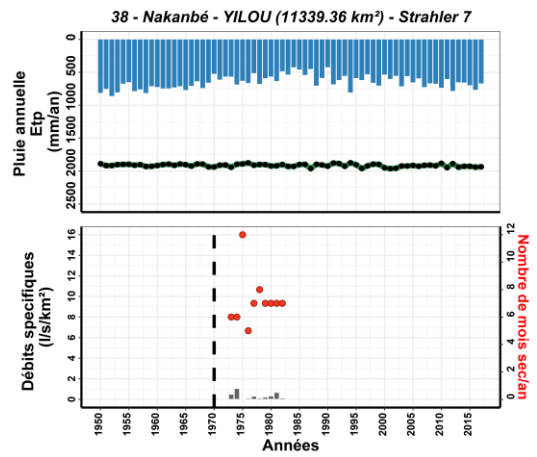
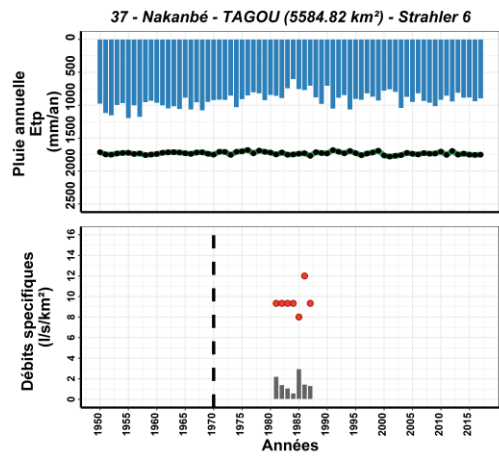
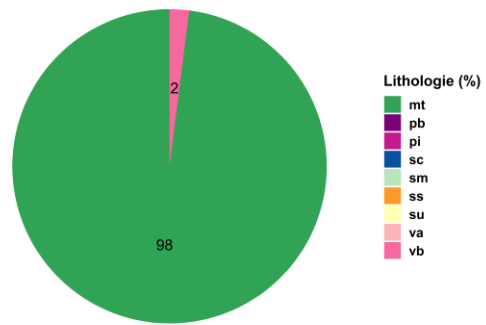
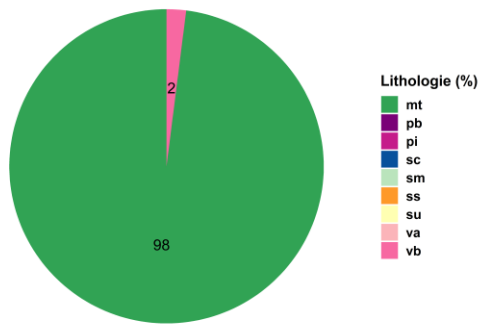
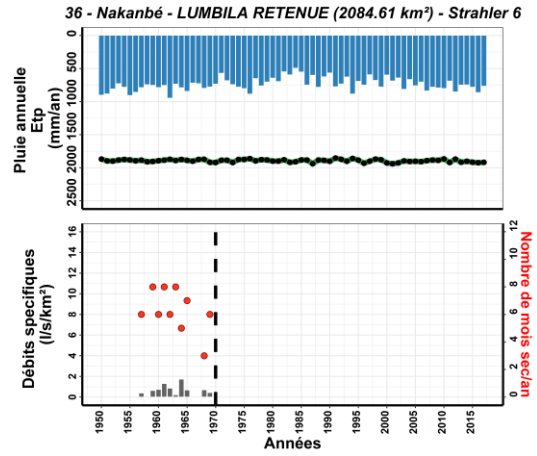
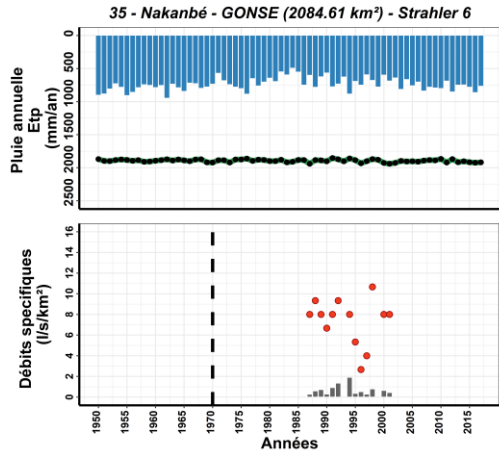




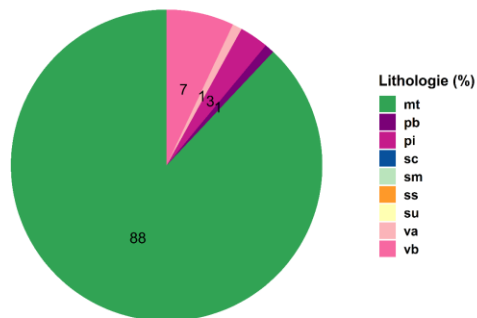
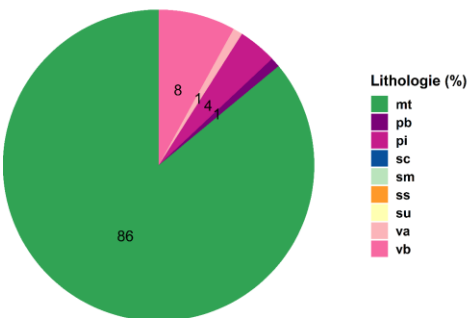
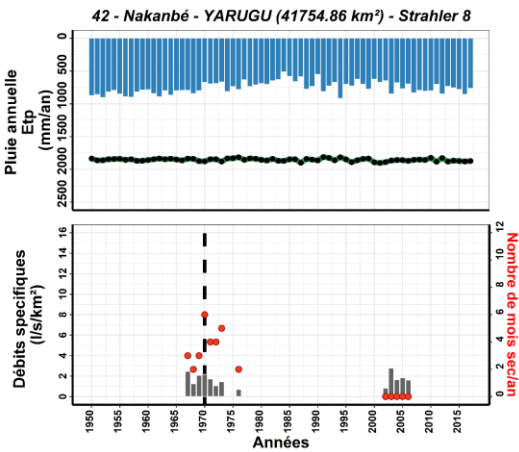
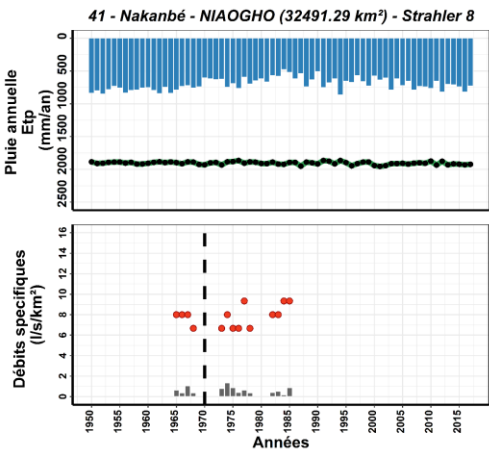
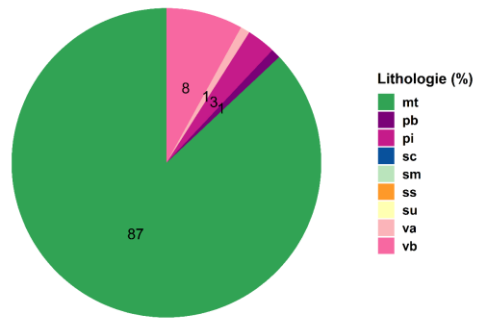
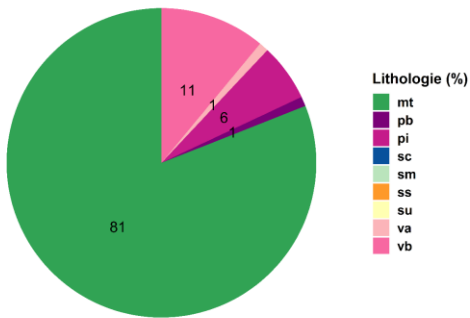
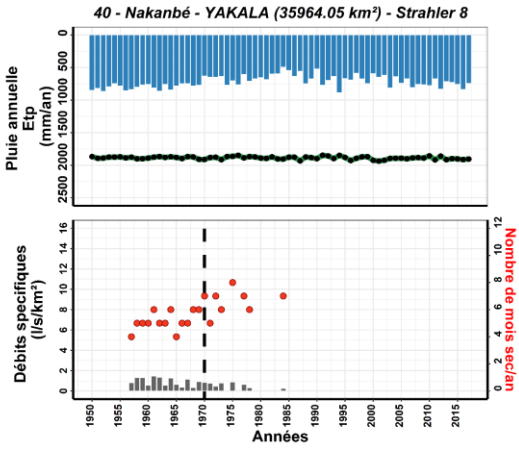
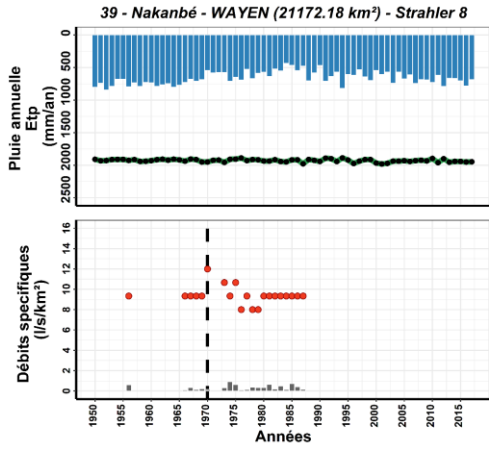
Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso

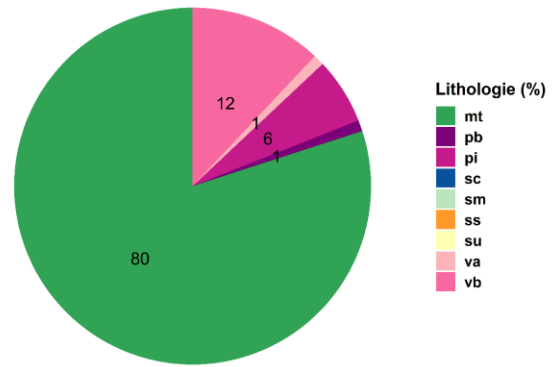
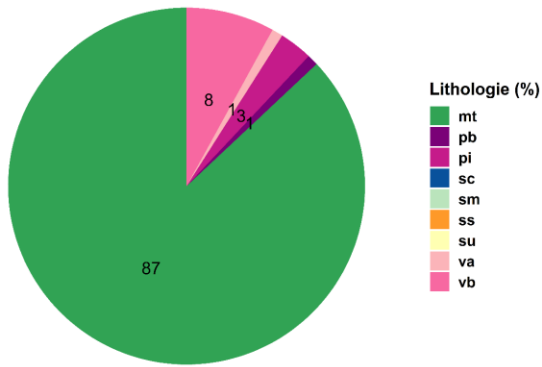
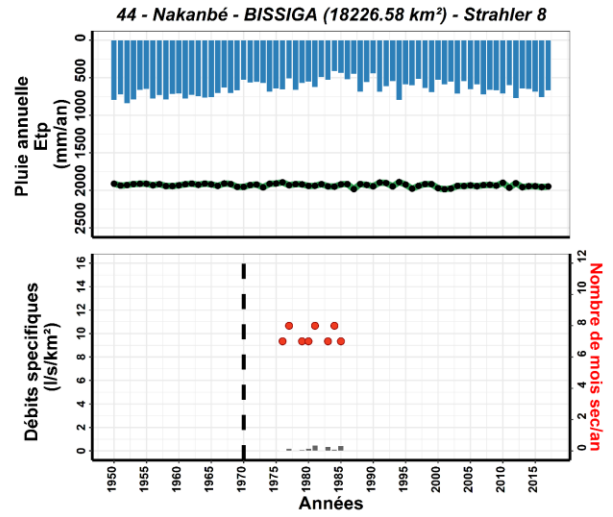
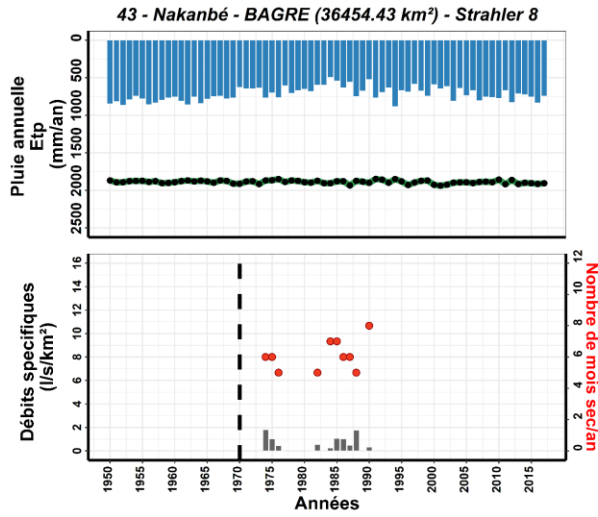


Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso



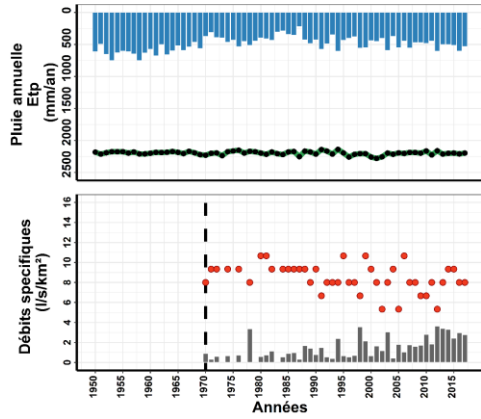
Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso



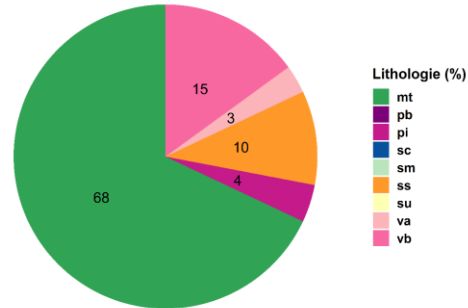
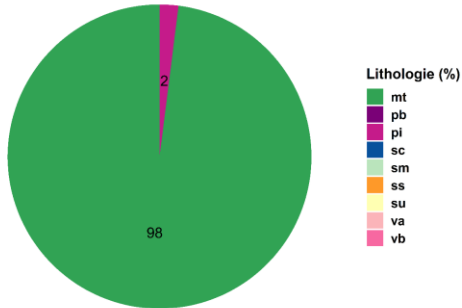
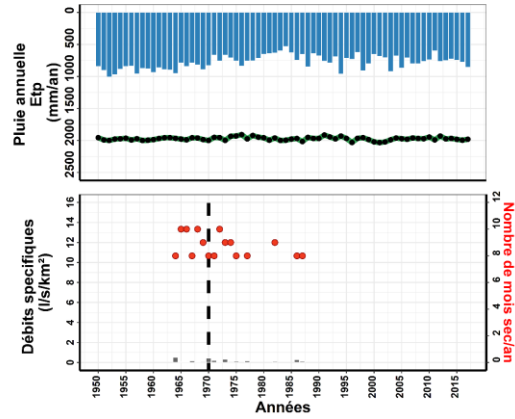


## Bassin du Niger

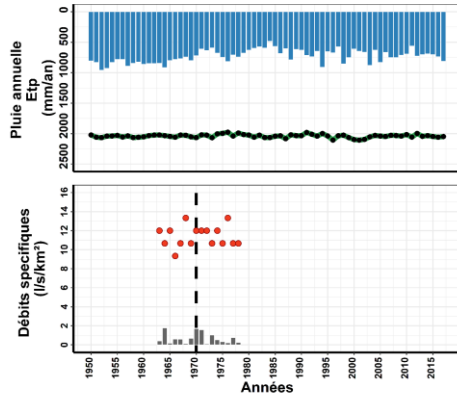
45 - Niger - KORIZZIENA (2569.07 km<sup>2</sup>) - Strahler 6



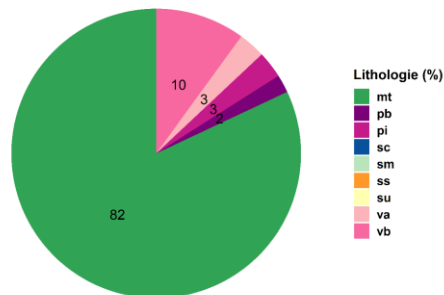
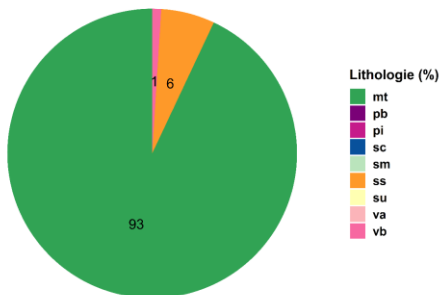
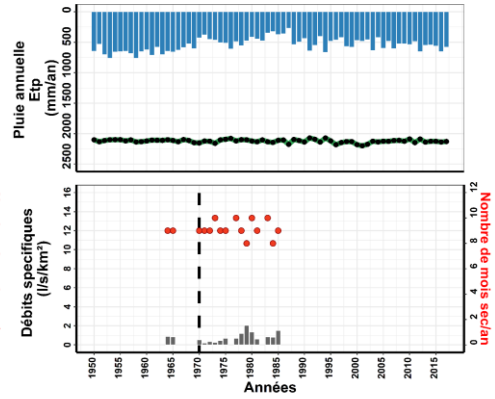
46 - Niger - CAMP W (5597.12 km<sup>2</sup>) - Strahler 6



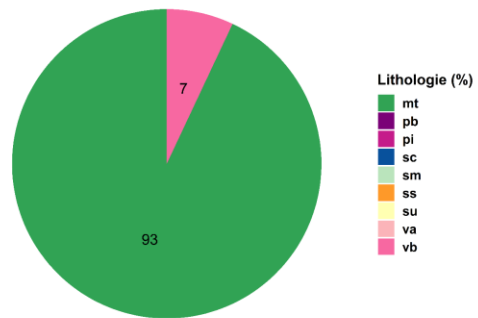
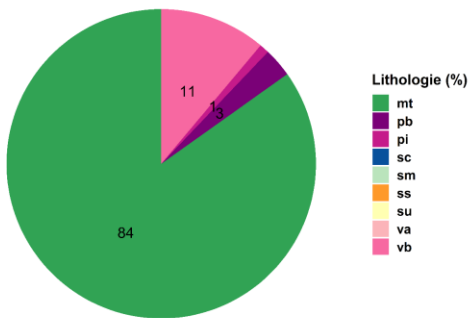
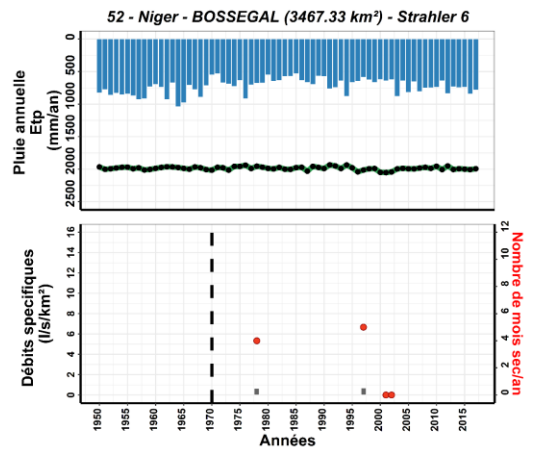
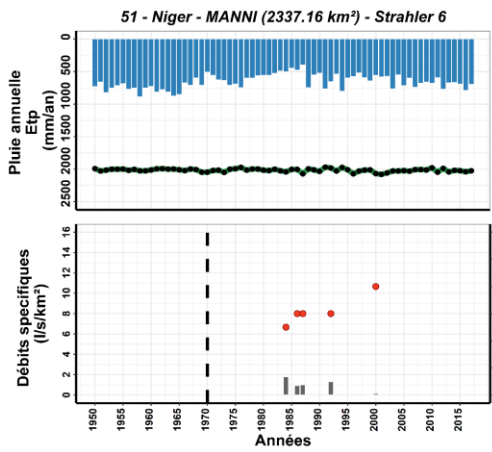
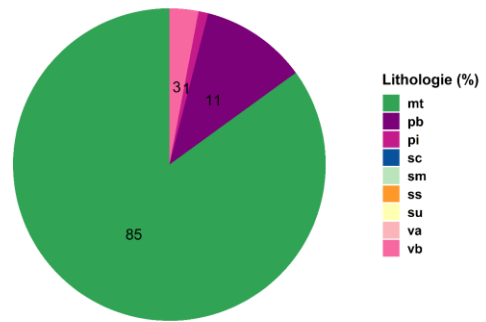
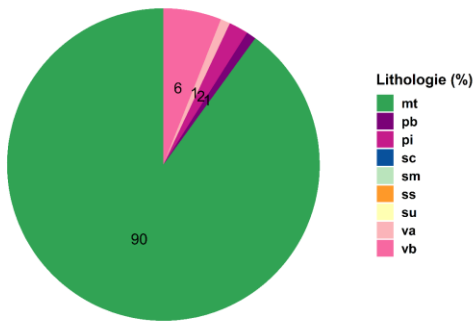
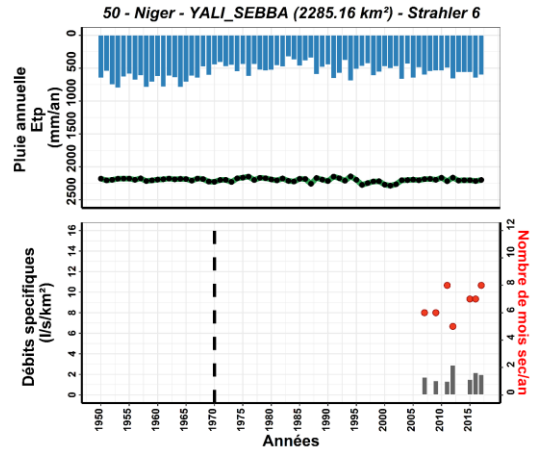
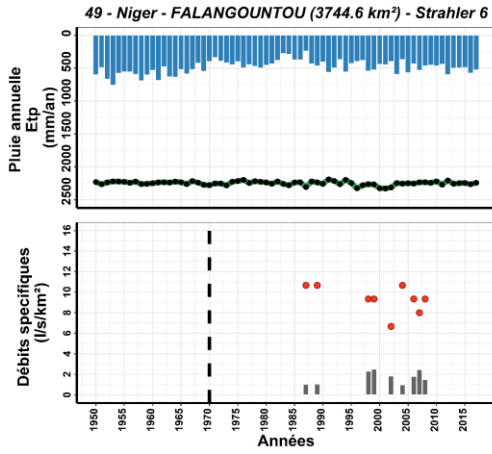
47 - Niger - TAMOU (4099.18 km<sup>2</sup>) - Strahler 6



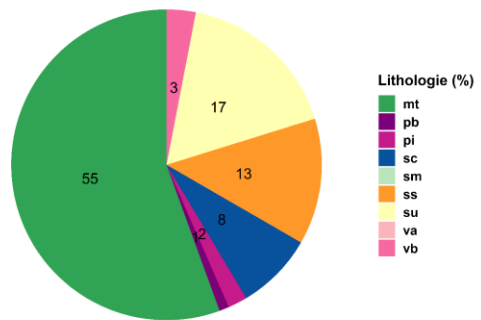
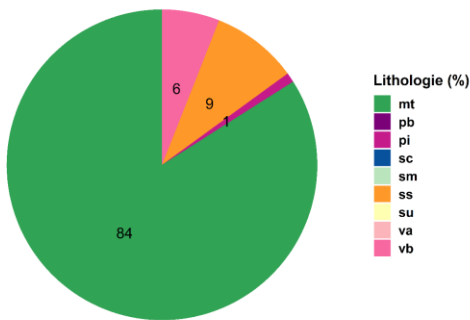
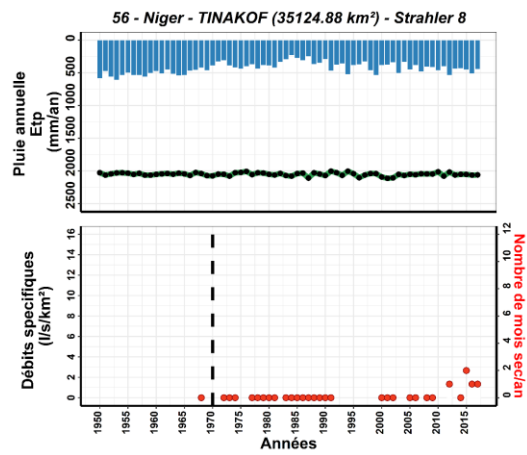
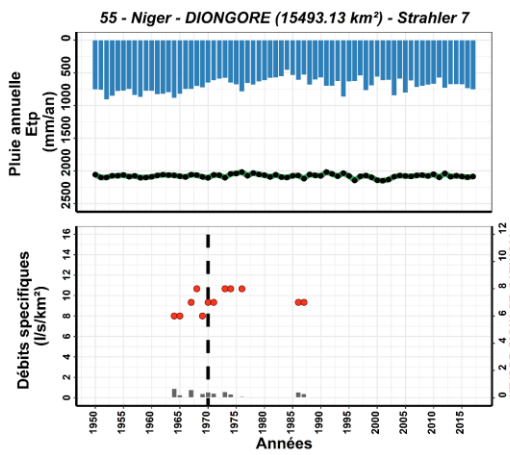
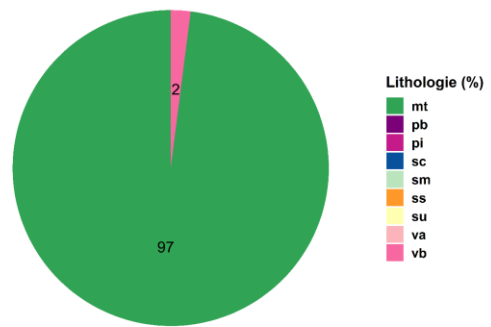
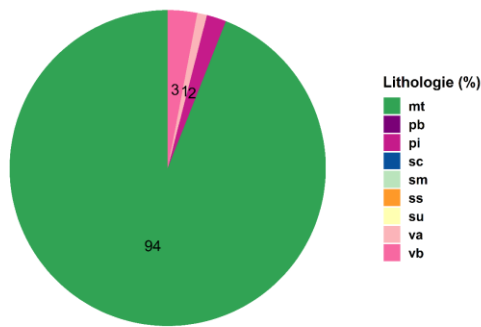
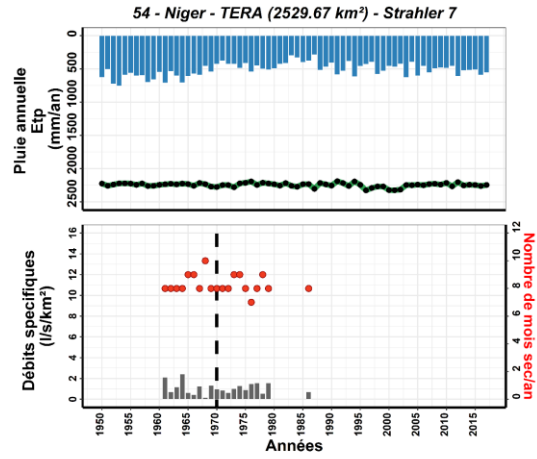
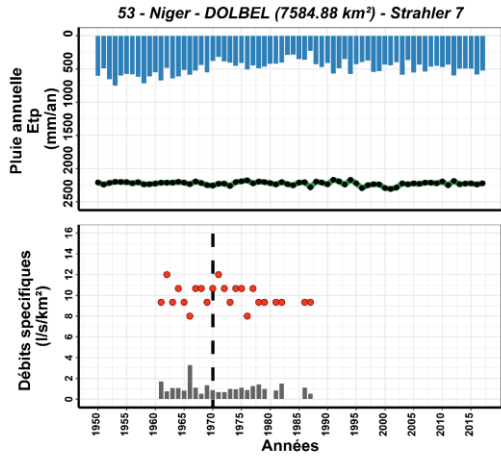
48 - Niger - YAKOUTA (1675.24 km<sup>2</sup>) - Strahler 6

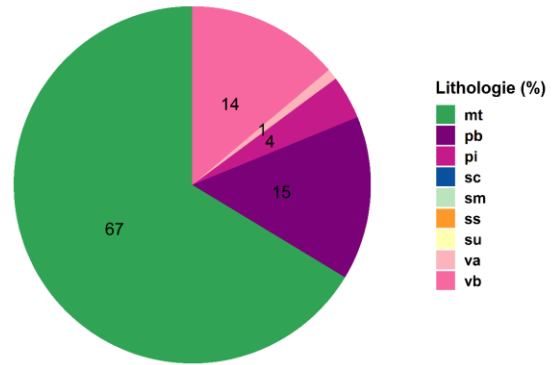
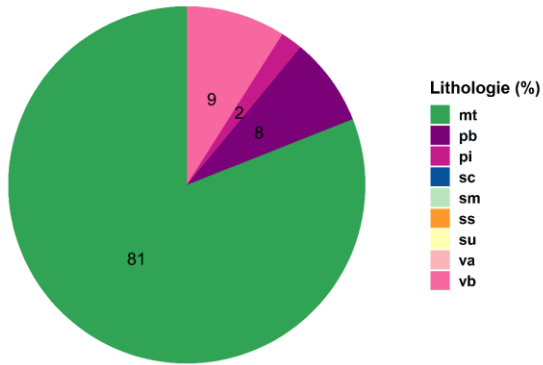
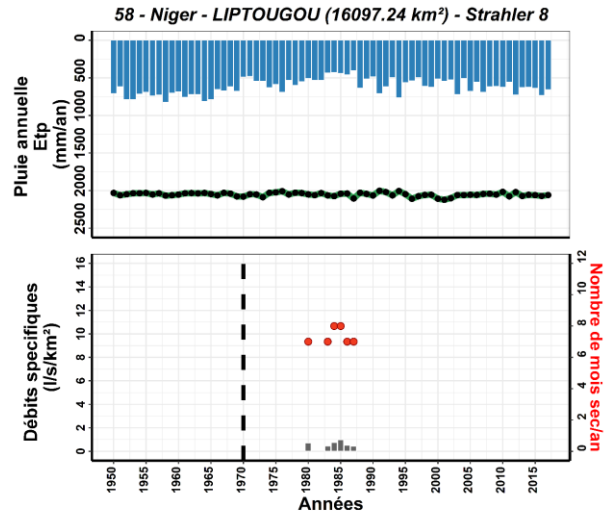
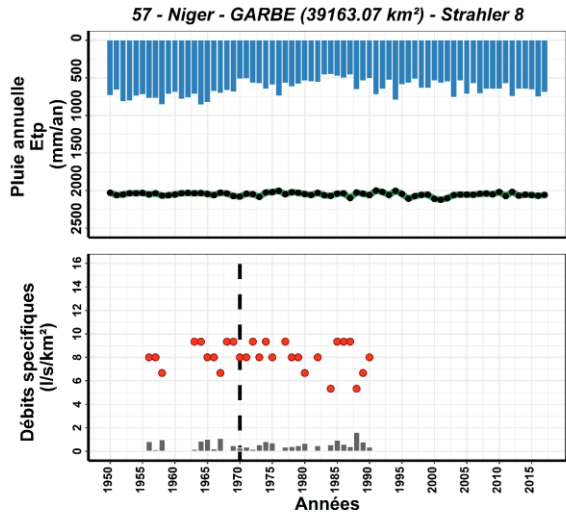


Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso



Annexe - 2 - Fiches descriptives des stations investiguées au Burkina Faso







## Annexe - 3 - Éléments supplémentaires sur la prédiction de l'intermittence des cours d'eau en Afrique

### Annexe - 3 - 1 Variabilité spatiale des caractéristiques hydro-environnementales par bassin versant en Afrique

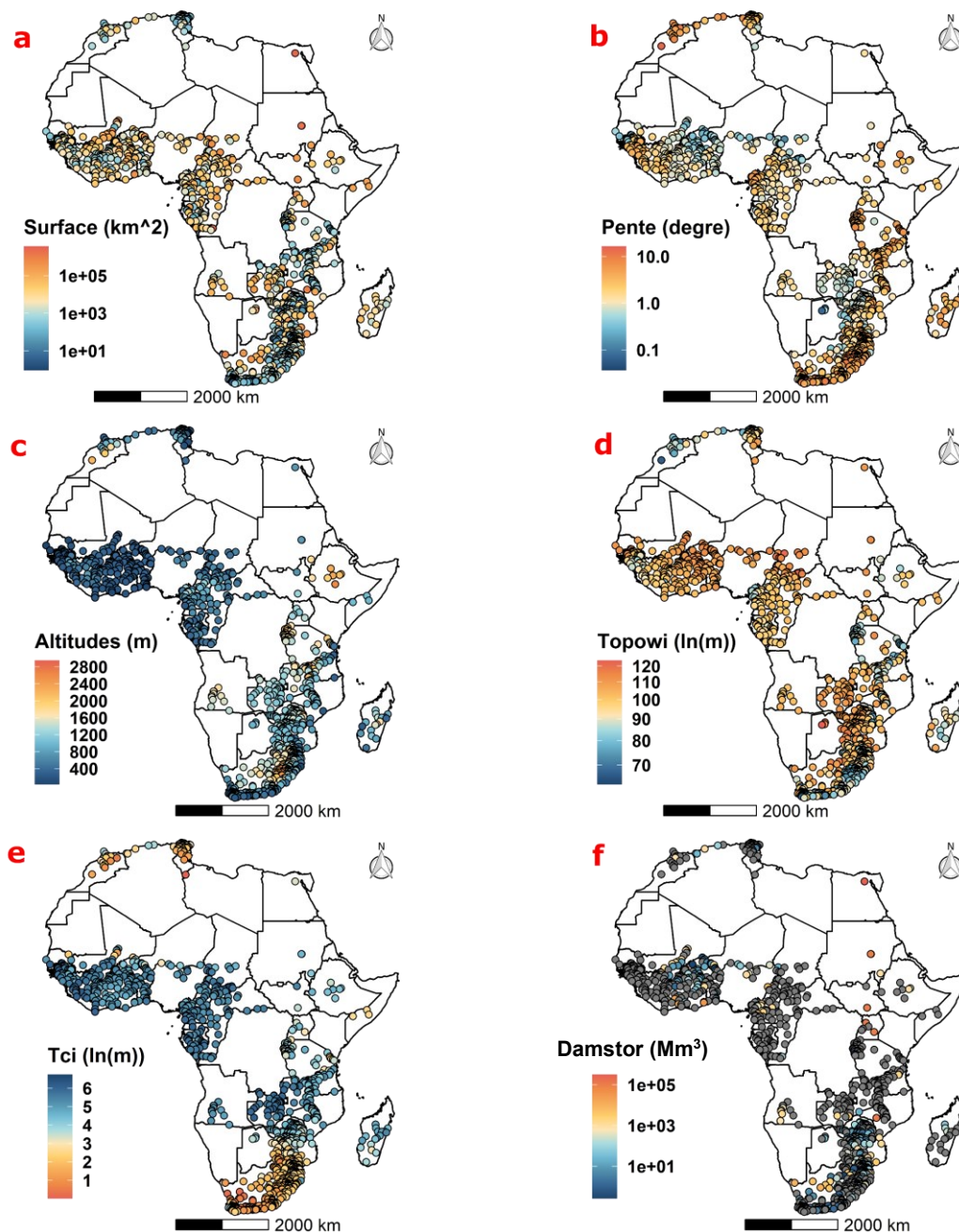


Figure A3-1 : Caractéristiques topographiques et influence anthropique des 1125 bassins versants en Afrique. Les points représentent les exutoires des bassins versants.

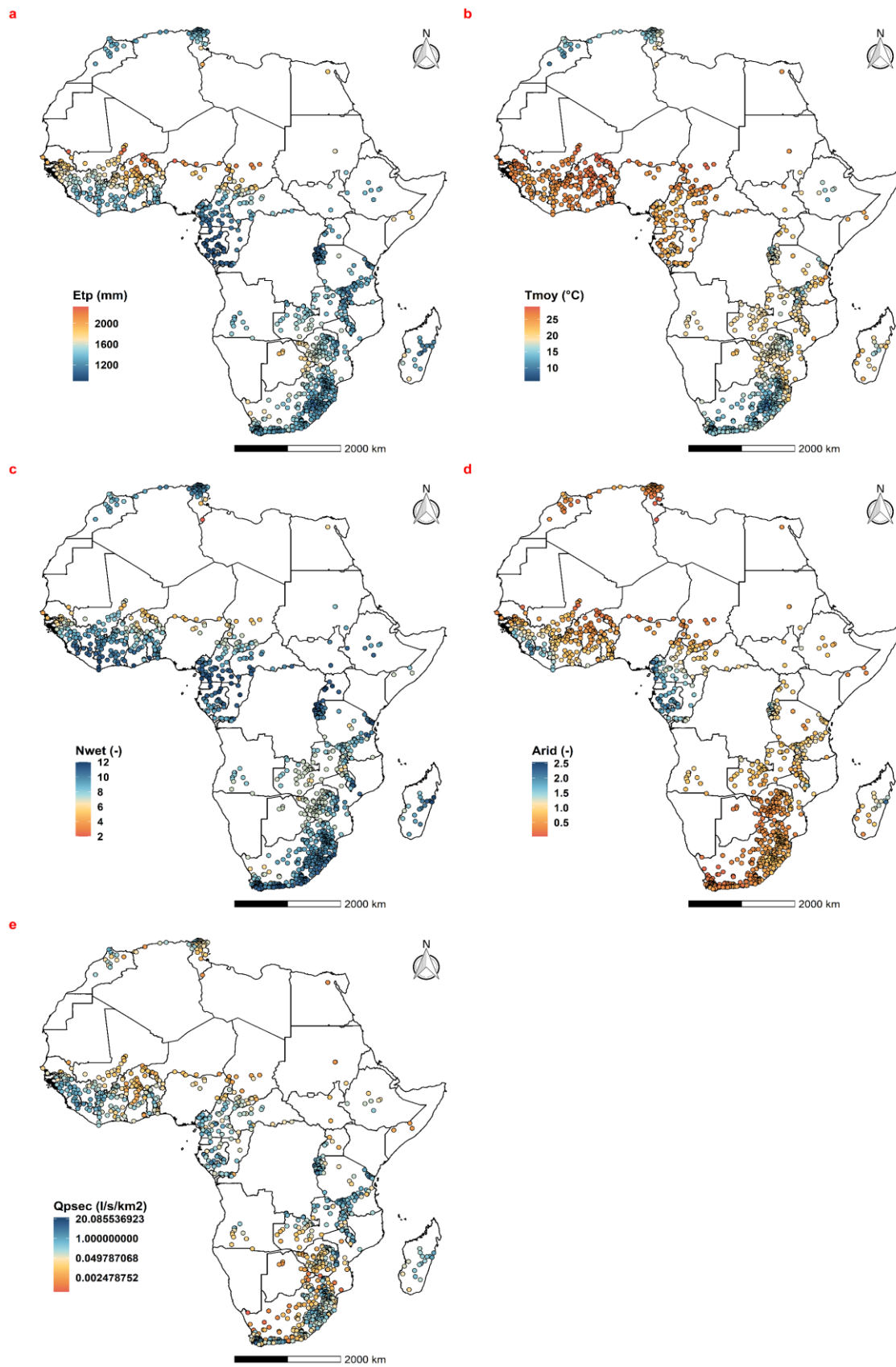


Figure A3-2 : Caractéristiques hydro-climatiques des 1125 bassins versants en Afrique. Les points représentent les exutoires des bassins versants.

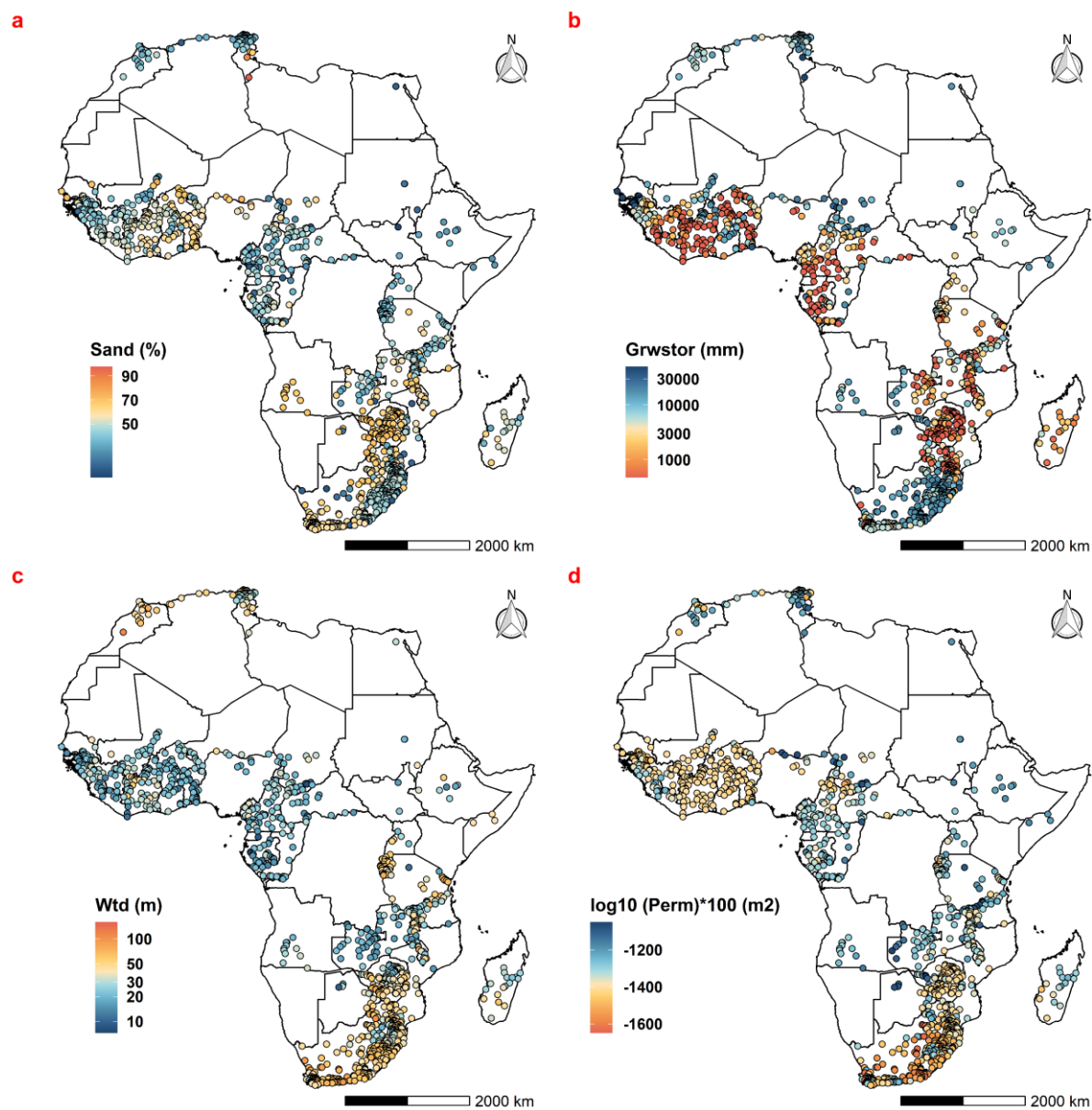
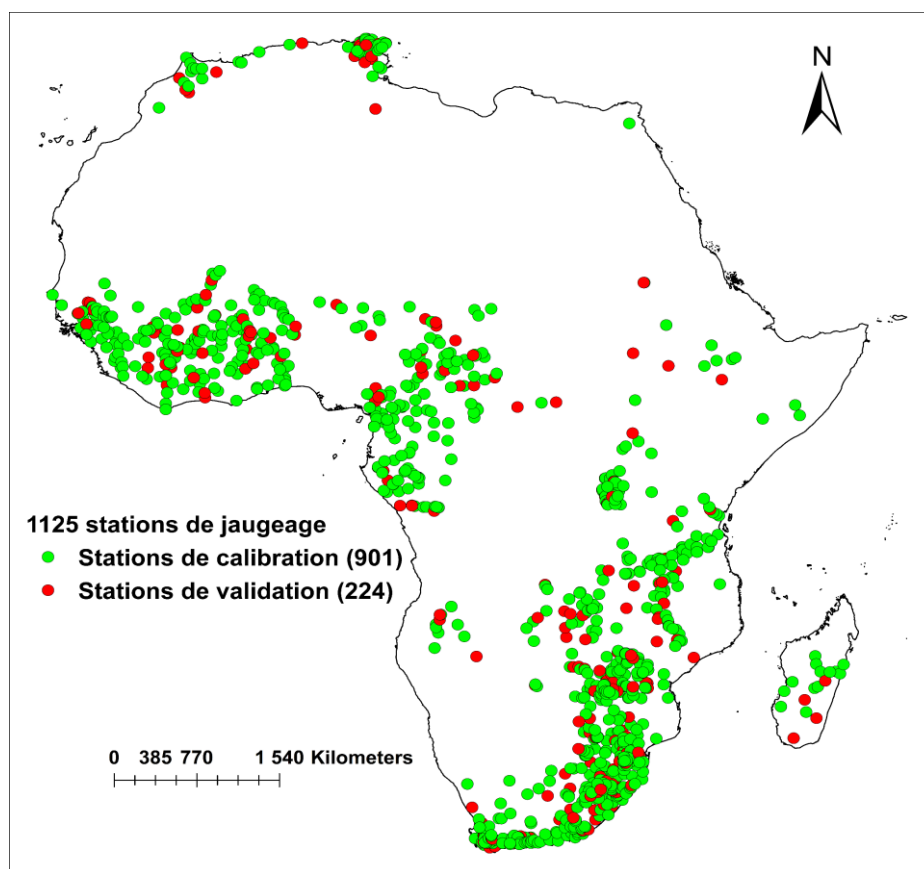


Figure A3-3 : Caractéristiques hydrogéologiques et propriétés des sols des 1125 bassins versants en Afrique. Les points représentent les exutoires des bassins versants.

## Annexe - 3 - 2 Détails sur l'implémentation de Random Forest

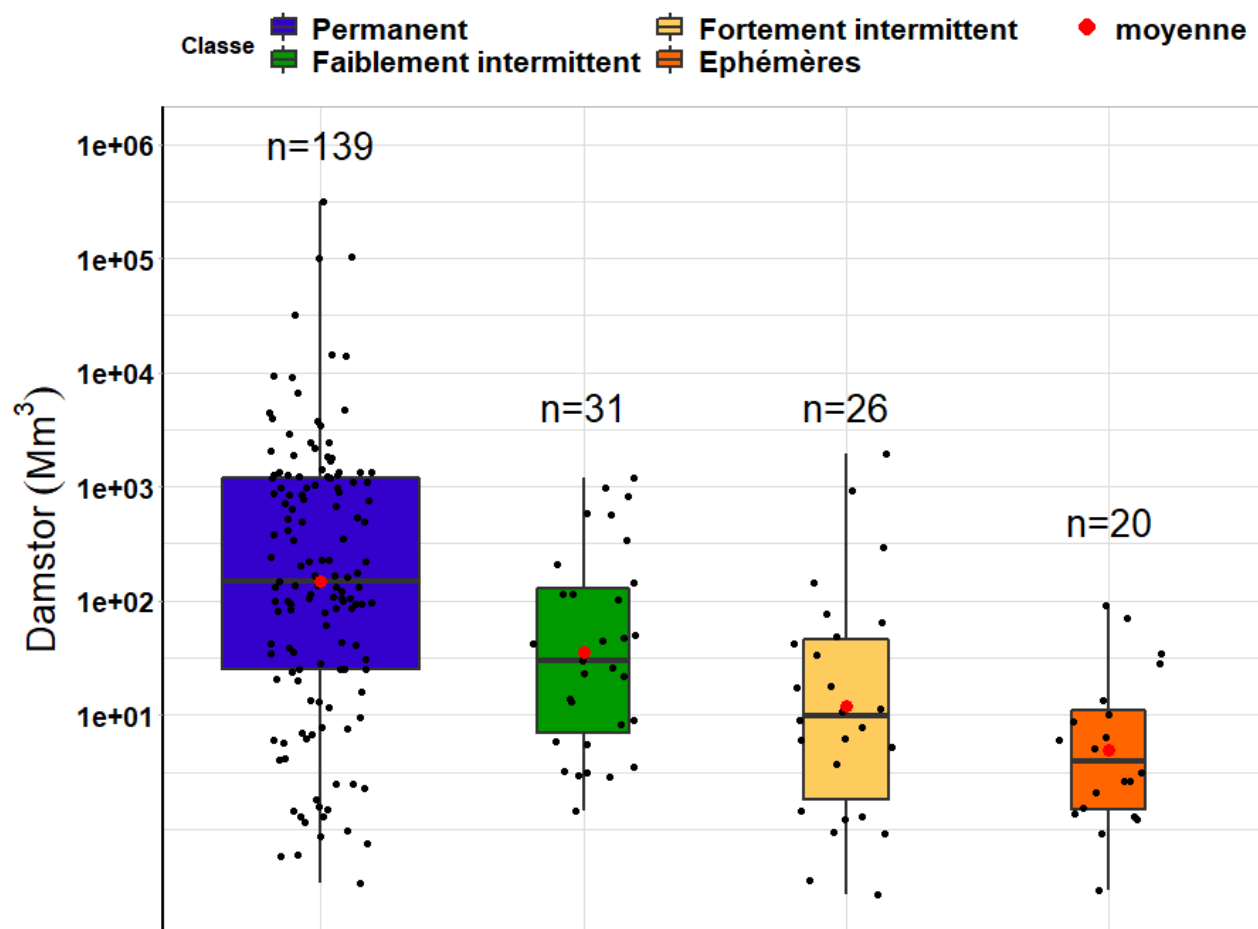


**Figure A3-4 :** Localisation des stations de calibration (901) et validation (224). Certaines stations sont superposées dans la représentation en raison de leur proximité spatiale.

Pour le calcul du Kappa pondéré au **chapitre III**, le coût de l'éloignement d'une classe prédite par rapport à sa classe réelle est défini par la différence entre les valeurs moyennes du nombre de mois secs des classes concernées à l'exception de la classe permanente où des valeurs de coût plus faible ont été définies. Ce choix de coût plus faible pour la classe permanente se justifie par le fait que dans l'ensemble les modèles prédisent déjà bien la classe permanente.

**Tableau A3-1:** Matrice de coût utilisé comme pondération pour le calcul du critère de Kappa. Par exemple, la différence entre les valeurs moyennes du nombre de mois secs de la classe faiblement intermittente et celle permanente est de 2,5 mois tandis qu'entre celle faiblement intermittente et fortement intermittente, elle est de 3 mois.

		Prédit			
		Perm	FaI	FoI	Eph
Observé	Perm	0	1	2	3
	FaI	2,5	0	3	7
	FoI	5,5	3	0	4
	Eph	9,5	7	4	0



**Figure A3-5 :** Capacité de stockage total des barrages à l'amont pour les 24% de stations en calibration (soit 216 stations) qui ont au moins un barrage à l'amont en fonction des différentes classes d'intermittence. La largeur des boxplot est proportionnelle au nombre d'observations par classe d'intermittence.