UNIVERSITÉ DU BURUNDI
GRENIER DU SAVOIR DU BURUNDI
Dépôt institutionnel officiel
https://repository.ub.edu.bi

Modélisation de la montée des eaux du lac Tanganyika

Niyukuri, Méthode; Sous la direction de : Dr David Niyukuri

2024-04
UB, FS
https://repository.ub.edu.bi/handle/123456789/2032

Document téléchargé depuis le **Grenier du Savoir du Burundi** — https://repository.ub.edu.bi

UNIVERSITE DU BURUNDI
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES
CENTRE DE RECHERCHE EN MATHEMATIQUES ET PHYSIQUE

Modélisation de la montée des eaux du lac Tanganyika

Par :

Méthode NIYUKURI

Mémoire
présenté et défendu publiquement en vue de l'obtention du Diplôme de Master en Mathématiques fondamentales et appliquées.

Sous la direction de :
Dr David NIYUKURI

Bujumbura, Avril 2024

Composition du Jury

Prof Servat NYANDWI (*Président du Jury*)

Dr Menus NKURUNZIZA (*Secrétaire du Jury*)

Dr David NIYUKURI (*Directeur de Mémoire*)

Dédicace

A **Dieu** tout puissant
A mon épouse **Floride Ndayishemeze** ;
à mes enfants :
Igiraneza Achlei solène-lava
Mahoro Worthy Meflo-Alliance
Niyukuri Samy Santinelle
Niyukuri Sara Sapience
Niyukuri Reine Esther
Niyukuri Dany Douceur
Niyukuri Amie Souriante
A mes **parents** et à mes **beaux-parents** ;
à mes chers **frères et soeurs** et à mes chers **beaux-frères et belles-soeurs** ;
à tous ceux qui sont fiers de ce pas franchi.
Je dédie ce mémoire !

Remerciements

Ce travail n'est pas seulement le fruit de mes seuls efforts. Plusieurs personnes ont contribué à sa réalisation. J'ai l'insigne honneur d'adresser mes profonds remerciements au **Dr David NIYUKURI**, Professeur de l'université du Burundi, qui, malgré ses occupations a accepté de diriger ce mémoire. Ses précieux conseils, sa disponibilité, sa rigueur scientifique et ses suggestions m'ont été d'une grande utilité. Qu'il trouve ici l'expression de ma profonde et respectueuse gratitude.

Je suis très reconnaissant à l'égard de **Mr. Nathanaël HATEGEKIMANA**, pour son accueil chaleureux, le temps passé ensemble et le partage de ses expertises au quotidien. Malgrè ses occupations, il a accepté de m'aider et à fournir quelques données nécessaires à mon sujet pour effectuer mon travail. Grâce aussi à sa confiance, j'ai pu accomplir mes missions.

Je remercie également tous les membres du jury pour le temps qu'ils ont consacré pour lire et évaluer ce travail. Leurs précieux commentaires pour l'amélioration de ce travail me seront d'une grande importance.

Mes sentiments de reconnaissance sont également dirigés à l'endroit de tous ceux qui ont contribué à ma formation de ce cycle de master, plus particulièrement à tous les professeurs de la Faculté des Sciences, surtout ceux du département de mathématiques et de physique. Qu'ils trouvent ici ma profonde gratitude pour m'avoir doté d'une formation tant humaine que scientifique.

J'exprime ma vive reconnaissance envers ma famille pour leur support tant matériel que moral et leur encouragement depuis le début de cette formation à nos jours, qu'ils trouvent ici le fruit de leurs efforts. Mes sentiments de gratitude sont également adressés à mes parents pour m'avoir montré le chemin de l'école et pour m'avoir soutenu moralement et matériellement durant notre scolarité.

Résumé

Les inondations sont l'une des catastrophes naturelles qui causent des dommages irréversibles et immenses à la vie humaine, à l'agriculture, aux infrastructures et aux aspects socio-économiques. La montée du niveau du lac Tanganyika est un phénomène qui doit être de plus en plus étudié puisqu'elle possède des répercussions directes sur la population. Cette catastrophe s'est manifestée par l'influence des facteurs tels que les précipitations, la densité de la population, la température, l'humidité et l'urbanisation. Pour estimer les niveaux et les vitesses de l'eau de ce lac sur une grande surface, nous avons combiné les données relatives aux précipitations, à la croissance démographique, à la température, à l'humidité et à l'urbanisation grandissante. Une prévision efficace repose sur l'utilisation d'un modèle de régression linéaire multiple (RLM) accompagné d'un modèle de Random Forest (RF) pour prédire avec précision la montée des eaux du lac Tanganyika. Les résultats obtenus par le modèle de prédiction basé sur la régression linéaire multiple ont montré que la montée des eaux du lac Tanganyika a beaucoup progressé depuis 2021 à 2022, mais l'an 2023, le niveau du lac Tanganyika a légèrement diminué avec un grand risque de remonter au cours de l'année 2024. Donc, ce document vise à réduire les risques extrêmes de cette catastrophe naturelle et contribue également aux suggestions politiques en fournissant une prédiction des inondations à l'aide de ce modèle d'apprentissage automatique (régression linéaire multiple et Random Forest).

Les mots clés : Montée des eaux, Régression linéaire multiple, Prévision, Urbanisation, Démographie, Précipitations, Température, Humidité, Random Forest.

Abstract

Flooding is one of the most catastrophic natural disasters, causing irreversible and immense damage to human life, agriculture, infrastructure, and socio-economic aspects. The rising level of Lake Tanganyika is a phenomenon that needs to be studied more and more, as it has direct repercussions on the population. This catastrophe has manifested itself through the influence of factors such as rainfall, population density, temperature, humidity, and urbanization. To estimate lake levels and velocities over a large area, we combined data on precipitation, population growth, temperature, humidity, and increasing urbanization. Effective forecasting relies on the use of a multiple linear regression (MLR) model accompanied by a random forest (RF) model to accurately predict the rising water levels of Lake Tanganyika. The results obtained by the prediction model based on multiple linear regression showed that Lake Tanganyika's water level rose significantly from 2021 to 2022, but in 2023, Lake Tanganyika's water level fell slightly, with a high risk of rising again in 2024. So, this paper aims to reduce the extreme risks of this natural disaster and also contributes to policy suggestions by providing a flood prediction using this machine learning model (multiple linear regression and Random Forest).

Keywords : Rising water levels, Multiple linear regression, Forecast, Urbanization, Demography, Precipitation, Temperature, Humidity, Random Forest.

Table des matières

Composition du Jury i

Dédicace ii

Remerciements iii

Résumé iv

Abstract v

Table des Matières vi

Liste des Tableaux viii

Liste des Figures ix

Liste des sigles et abréviations x

Avant-propos xi

1 Introduction 1

1 Problématique . . . 2

2 Objectif de recherche . . . 3

2.1 Objectif global . . . 3

2.2 Objectifs spécifiques . . . 3

2 Revue de la littérature 4

1 Influence des pluies torrentielles, de la température et de l'humidité sur la montée des eaux . . . 4

2 Influence de l'urbanisation sur la montée des eaux . . . 5

3 La croissance démographique influence la montée des eaux . . . 6

3 Matériel et Méthodes 7

1 Matériel . . . 7

1.1 Données climatiques . . . 7

1.2 Données géographiques 8
1.3 Localisation 8
2 Méthodes 10
2.1 Analyse de la statistique descriptive des données 10
2.2 Modèle de Régression linéaire multiple (RLM) 12
2.3 Modèle de Random Forest 12
2.4 Formulation du modèle 13

4 Résultats et leurs Interprétations **15**
1 Analyse de l'évolution et de la distribution du Niveau du Lac Tanganyika de 2010 à 2022 15
2 Analyse uni-variée de la distribution des variables explicatives 17
3 Analyse bivariée entre les variables indépendantes et le niveau du lac Tanganyika 18
4 Estimation du modèle linéaire simple pour chaque variable explicative . . . 19
5 Estimation du modèle linéaire multiple 20
6 Tests de différentes hypothèses 21
6.1 Test de significativité global 21
6.2 Test d'autocorrelation 22
6.3 Test d'hétéroscedasticité 24
6.4 Test de normalité des résidus 25
6.5 Test de spécification du modèle (Test de Ramsey) 25
7 Election des facteurs explicatifs 26
8 Prévision avec le modèle de régression linéaire multiple 27
9 Estimation du modèle Random Forest 28

5 Discussion et Conclusion **31**
1 Discussion 31
2 Conclusion et recommandations 32

ANNEXE **34**

Bibliographie **39**

Liste des tableaux

3.1 Analyse de la statistique descriptive des données 11

4.1 Estimation du modèle linéaire simple 19
4.2 Estimation du modèle linéaire multiple 20
4.3 Significativité individuelle des paramètres 21
4.4 Différentes statistiques de régression 22
4.5 Différentes statistiques **de validité** d'une régression 23
4.6 Breusch-Godfrey Serial Correlation LM Test(Multiplicateur de Lagrange) . 23
4.7 Correction finale d'autocorrelation 24
4.8 Test d'hétéroscédasticité (ARCH) 24
4.9 Test de spécification du modèle 25
4.10 Model Selection 27
4.11 Estimation du modèle final 27
4.12 Resampling Indicators 28

Table des figures

3.1 Bujumbura Mairie et Zone Gatumba 9

4.1 Niveaux du lac Tanganyika selon la période de 2010 à 2022 15
4.2 Analyse uni-variée de la distribution des variables explicatives 17
4.3 Analyse bivariée de la distribution des variables explicatives 18
4.4 Prédiction de la montée du lac Tanganyika 28
4.5 Chronogramme des valeurs prédites par Random Forest 29
4.6 Importance de chaque variable 30

5.1 Allure de l'humidité dans la Mairie de Bujumbura 34
5.2 Allure de la température dans la Mairie de Bujumbura 35
5.3 Allure du rainfall dans la Mairie de Bujumbura 35
5.4 Allure de la densitité dans la Mairie de Bujumbura 36
5.5 Allure de l'humidité dans la Province de Bujumbura 36
5.6 Allure de la température dans la Province de Bujumbura 37
5.7 Allure du rainfall dans la Province de Bujumbura 37
5.8 Allure de la densitité dans la Province de Bujumbura 38
5.9 Allure de l'urbanisation 38

Liste des abréviations

- ARCH : AutoRegressive Conditional Heteroskedasticity
- LM : Multiplicateur de Lagrange
- H0 : Hypothèse nulle
- H1 : Hypothèse alternative
- OIM : l'Organisation Internationale pour les Migrations
- BRARUDI : Brasserie et Limonaderie du Burundi
- COGERCO : Compagnie de gérance du coton
- IGEBU : Institut Géographique du Burundi
- RAFINA : Raffinerie Nationale
- SEP : Société de Gérance du Pétrole
- N_TANGA : Niveau du lac Tanganyika
- H_MAIRIE : Humidité en Mairie
- T_MAIRIE : Temperature en Mairie
- R_MAIRIE : précipitations en Mairie
- D_MAIRIE : Densité de la population en Mairie
- H_BUJA : Humidité à Bujumbura
- T_BUJA : Temperature à Bujumbura
- R_BUJA : Précipitations à Bujumbura
- D_BUJA : Densité de la population à Bujumbura
- URBA : Urbanization à Mairie
- RMSE : (Root Mean Squared Error)Racine carrée de l'erreur quadratique moyenne.
- INSBU : Institut national de la statistique du Burundi (INSBU), anciennement connu sous le nom de ISTEEBU,
- ISTEEBU : Institut de Statistiques et d'Etudes Economiques du Burundi
- SIG : Système d'information géographique
- MAE : Erreur moyenne absolue

Avant-propos

Le lac Tanganyika est l'un des plus grands lacs d'Afrique et représente une ressource vitale pour les populations riveraines. Cependant, ces dernières années, on a observé une montée inquiétante du niveau des eaux du lac, causant d'importants dommages et déplacements de populations. Comprendre les causes et les dynamiques de cette montée des eaux est donc devenu un enjeu majeur pour la région.

C'est dans ce contexte que la présente étude a été menée, avec pour objectif de développer un modèle prédictif de l'évolution du niveau du lac Tanganyika. En se basant sur une analyse approfondie des données historiques, climatiques et hydrologiques, ce modèle vise à identifier les facteurs-clés responsables de la hausse du niveau d'eau et à proposer des scénarios d'évolution à court et moyen terme.

Les résultats de cette modélisation fourniront aux décideurs politiques et aux communautés locales des outils essentiels pour anticiper les impacts futurs, mettre en place des mesures d'adaptation et de gestion durable du lac Tanganyika. Il s'agit d'une contribution importante à la compréhension et à la gestion d'un enjeu environnemental majeur pour l'Afrique centrale.

Chapitre 1

Introduction

Le changement climatique est un facteur important dans l'exposition de la population aux crues fluviales que le développement socio-économique, car la fréquence et l'intensité des crues fluviales devraient augmenter en raison du changement climatique dans de certaines régions [1]. Ce phénomène élargirait les zones sujettes aux inondations et rendrait les inondations plus susceptibles de se produire plus souvent dans ces zones [2]. L'adaptation des populations riveraines des cours d'eau face au risque d'inondations est en perpétuelle évolution. Ce risque doit de nos jours être de plus en plus contrôlé et quantifié. Le but de la démarche présentée dans ce travail est de minimiser les impacts néfastes de la montée des eaux dans notre zone d'étude où la pression urbanistique est de plus en plus forte. Le moyen privilégié de lutter contre les inondations est actuellement la mesure de protection [3]. Le changement climatique met les populations à risque dans les pays développés comme dans les pays en développement [4]. La montée des eaux causée par les facteurs tels que : les précipitations, l'urbanisation, la croissance démographique, la température et l'humidité bouleverse à ce point l'organisation politique, économique et sociale des cités que les communautés humaines ont très tôt cherché à en reconnaître les signes avant-coureurs et à établir des systèmes de prévision efficients [5].
Ces inondations entraînent l'abandon des habitations touchées et un relogement des sinistrés dans des conditions parfois précaires.
Il apparaît cependant que les pluies responsables des inondations de ces dernières années ne revêtent pas un caractère exceptionnel. Par contre, la vulnérabilité de la ville de Bujumbura Mairie face au risque d'inondations semble avoir fortement augmenté suite à un aménagement urbain déficient [6]. Ce travail se donne pour objectif d'investiguer les cas vécus dans la région, de ressortir le niveau d'exposition aux impacts directs et de produire des outils d'éclairage en la matière [7].
La modélisation de cette catastrophe peut également être utilisée pour aider les gouvernements et les organisations à prendre des décisions éclairées sur la préparation et la réponse, ainsi que pour informer le public sur les risques liés à ce type de catastrophe [8].

1 Problématique

Cette recherche s'inspire des cas vécus dans la région et propose une analyse axée sur la regression linéaire multiple et le Random Forest dans le but de proposer des recommandations [7]. La montée des eaux se produit souvent lorsqu'il y a de fortes pluies qui durent quelques heures, provoquant une montée de niveau et un débordement de l'eau vers le quartier résidentiel proche de la rivière [9]. L'étude de Bergonzini et ses coéquipiers montre qu'entre 1932 et 1995, le niveau du lac a chuté d'un mètre, soit en moyenne de 15 mm par an et que la fluctuation du niveau lacustre est caractérisée par la variabilité hydroclimatique qui a été simultanément constatée sur le lac Victoria et les autres lacs de la sous-région [10].

La montée des eaux qui s'observe par le débordement des lits des rivières qui traversent la ville de Bujumbura pendant une période relativement longue (cas du débordement de la Ntahangwa en avril 1986) où on a enregistré 200 mm de pluie en seulement 36 heures causa de multiples dégâts (chausse d'Uvira recouvert d'1m d'eau et la route vers l'aéroport sont interrompues, la hauteur de 2,20 m à la SEP, Fina actuel Engen emporté, etc.) ; cas de la Ntahangwa en janvier 1997 (hauteur de 2,1 m à la SEP, 1 m à la RAFINA, digue mise en place en 1990 submergée et endommagée) ; cas de la Mutimbuzi en 2018 par exemple sous les eaux. Cette montée s'observe souvent sur les cours d'eau et est responsable de multiples dégâts sur l'usine Brarudi, COGERCO, RAFINA, etc. chaque année [10].

Gatumba, près de Bujumbura, est l'une des zones les plus touchées. Des maisons ont été détruites, et des écoles, des centres de santé et des cultures ont été inondés. Selon l'Organisation internationale pour les migrations (OIM), environ 40 000 personnes ont été affectées par les inondations dans cette zone et, parmi elles, 13 000 ont dû quitter leurs maisons. Elles se sont installées dans des sites de déplacés ou sont logées dans des familles d'accueil. D'autres personnes sont restées dans leurs maisons inondées par manque d'alternative ou par peur d'abandonner leur quartier et de s'éloigner de leurs moyens de subsistance [11]. Les catastrophes occasionnent des crises majeures dont les conséquences humanitaires et environnementales sont souvent désastreuses, nécessitant une approche de prédiction stricte et efficace [12]. Selon les données de l'Institut géographique du Burundi (IGEBU), la montée du lac est un phénomène cyclique qui se produit tous les 50 à 60 ans. Les inondations actuelles sont largement attribuées aux changements climatiques, à l'urbanisation grandissante et à une croissance démographique très remarquable [13]. Les mécanismes d'éclaircissement de tels défis seront notre objectif.

Cela permet de prendre des mesures préventives et d'élaborer des stratégies de gestion des risques pour faire face aux défis posés par la montée des eaux dans la ville [14].

2 Objectif de recherche

2.1 Objectif global

Notre recherche a pour but de construire une plateforme de prédiction de la montée des eaux du lac Tanganyika dans la ville de Bujumbura à base des facteurs qui l'influencent. Le contrôle de ces facteurs peut diminuer l'ampleur de cette catastrophe.

2.2 Objectifs spécifiques

Pour atteindre l'objectif global ci-dessus, cette étude portera spécifiquement sur :

1 Identifier les facteurs influençant la montée des eaux du lac Tanganyika.

2 Prédire la montée des eaux du lac Tanganyika.

Chapitre 2

Revue de la littérature

Le phénomène d'inondations est au premier rang des risques majeurs sur le territoire burundais [15]. En effet, la ville de Bujumbura et ses environs est depuis des années submergée par une forte pression démographique qui est l'un des facteurs qui accentue la probabilité du risque dans la mesure où cela est suivi par la construction des lotissements sans études préalables ou sans le moindre souci de respect des normes et des règles urbanistiques recommandées. On observe des inondations par débordement direct des lits des rivières (rivière GASENYI en 2014) ou des accumulations des eaux ruisselées [16].

1 Influence des pluies torrentielles, de la température et de l'humidité sur la montée des eaux

Les précipitations sont une des composantes principales du cycle hydrologique. Elles représentent toutes les eaux météoriques qui tombent sur la surface de la terre sous forme liquide ou solide en provenance directe ou indirecte de la condensation de la vapeur d'eau atmosphérique. Elles sont mesurées en millimètres, le plus souvent grâce à des pluviomètres.
Il existe plusieurs formes différentes de précipitations : la pluie, la pluie verglaçante, le grésil et la neige, qui dépendent de l'épaisseur des nuages, de leur taux d'humidité, de la température de l'air ambiant et de celui du sol, mais aussi de la pression atmosphérique [17]. Le Sahel de Sousse en Tunisie a souffert des inondations majeures en 1969, 1973, 1995 et 2016. L'automne et l'hiver 2016 ont été marqués par des épisodes successifs de pluies intenses. Sept épisodes ont été recensés entre le 23 septembre et le 27 décembre 2016 (23 et 29 septembre ; 29 octobre ; 17 novembre ; 8, 16 et 27 décembre). Les 23 et 29 septembre 2016, les précipitations enregistrées à la station de Sousse ont totalisé 205 mm, représentant environ 0,62% du cumul annuel pour cette station. L'enchaînement rapide

des perturbations pluvieuses a entraîné des débordements des inondations et un ruissellement rapide dans les zones urbaines et périurbaines. L'événement pluvial a été à l'origine des processus hydrogéomorphologiques dommageables et des dégâts non négligeables [18]. Le Burundi est l'un des 20 premiers pays les plus vulnérables aux effets du changement climatique au monde. La majorité des besoins humanitaires dans le pays sont causés par des catastrophes naturelles récurrentes liées au changement climatique. Depuis 2018, 445 catastrophes naturelles ont affecté près de 270 000 personnes au Burundi. Parmi elles, 100 000 ont été déplacées [19].
L'augmentation de la température des océans perturbe les précipitations dans de nombreux pays du monde et le Burundi n'est pas épargné. Les saisons des pluies au Burundi sont de plus en plus éprouvantes pour les populations vulnérables. Les pluies torrentielles peuvent avoir un impact significatif sur la montée des eaux dans la ville de Bujumbura, qui est située au bord du lac Tanganyika. Lorsque de fortes précipitations se produisent, plusieurs facteurs peuvent contribuer à l'augmentation des niveaux d'eau [20]. Ces pluies torrentielles entraînent souvent un ruissellement important, car le sol peut devenir rapidement saturé et incapable d'absorber toute l'eau. Le ruissellement excessif peut entraîner des inondations urbaines, car les systèmes de drainage peuvent être dépassés, ce qui entraîne une accumulation d'eau dans les rues et les zones basses de la ville [19].
Notre étude vise à analyser cet événement exceptionnel, afin de tester l'hypothèse que les inondations survenues dans la zone d'étude sont liées à l'extension de l'intensité des pluies dues aux changements climatiques.

2 Influence de l'urbanisation sur la montée des eaux

L'urbanisation peut avoir un impact significatif sur la montée des eaux dans la ville de Bujumbura. Lorsque des zones rurales sont urbanisées, les sols naturels sont souvent recouverts de béton, d'asphalte et d'autres surfaces imperméables, telles que les routes, les trottoirs et les bâtiments. Cela limite la capacité du sol à absorber l'eau de pluie, ce qui entraîne un ruissellement accru. En conséquence, les précipitations sont évacuées plus rapidement vers les systèmes de drainage, ce qui peut augmenter la montée des eaux et le risque d'inondations [21].

Les impacts de l'urbanisation sur le cycle de l'eau sont nombreux. L'urbanisation peut également entraîner une expansion de la population et des infrastructures dans des zones à risque d'inondations, telles que les plaines inondables ou les zones côtières. Cela expose davantage de personnes et de biens aux dangers liés aux inondations, augmentant ainsi les conséquences potentielles de la montée des eaux [16].

L'une des conséquences les plus visibles de l'urbanisation est l'imperméabilisation des sols qui limite très fortement les possibilités d'infiltration de l'eau. En France, par exemple, la

surface imperméabilisée a décuplé entre 1955 et 1965 [22].

Ce phénomène entraîne en premier lieu une augmentation des volumes d'eau ruisselée. Cet élément est souvent mis en avant pour expliquer les inondations urbaines. En effet, le taux d'infiltration de la plupart des sols saturés, en l'absence de couvert forestier dense et bien établi (ce qui n'est pas le cas par exemple des terrains de montagne récemment reboisés), ou à l'exception de terrains très sableux, est très inférieur aux intensités que l'on peut rencontrer lors des événements pluvieux exceptionnels. Ainsi, dans ce type de situation, les terrains non revêtus donnent souvent lieu à des volumes ruisselés [22]. Cela peut inclure des mesures telles que la planification des zones de développement pour éviter les zones à risque, la création de zones tampons naturelles, la conception de systèmes de drainage adéquats et la promotion de pratiques de construction durables qui favorisent l'infiltration des eaux de pluie [23].

3 La croissance démographique influence la montée des eaux

Avec une croissance démographique rapide, la demande en infrastructures urbaines telles que les routes, les bâtiments, les systèmes de drainage et les réseaux d'approvisionnement en eau augmente. Lorsque ces infrastructures ne sont pas développées ou entretenues de manière adéquate pour répondre à la demande croissante, cela peut entraîner une augmentation du ruissellement des eaux de pluie et augmenter le risque d'inondations [24].

Lorsque la population de Bujumbura augmente, il peut y avoir une expansion de la ville vers des zones qui étaient auparavant non urbanisées ou moins développées. Cela peut inclure des zones telles que les plaines inondables ou les zones côtières qui sont naturellement sujettes aux inondations. L'occupation de ces zones à risque par une population croissante accroît l'exposition aux inondations et la vulnérabilité des habitants [25].

Une croissance démographique rapide peut aussi exercer une pression accrue sur les ressources naturelles, y compris les ressources en eau [26]. Cela peut entraîner une surexploitation des ressources en eau, une diminution des zones de captage naturel et une perturbation de l'équilibre hydrologique [27]. Ces facteurs peuvent contribuer à une augmentation de la montée des eaux lors de précipitations intenses.

Chapitre 3

Matériel et Méthodes

1 Matériel

1.1 Données climatiques

L'ensemble des données climatiques comprenne les variables comme la température, les précipitations et l' humidité de la province de Bujumbura et de la Mairie de Bujumbura. Les données utilisées couvrent une période de 13 ans, de 2010 à 2022, et ces données ont été enregistrées mensuellement. Les données seront recueillies auprès de la météorologie du Burundi.
Voici une représentation descriptive des noms de variables climatiques :

- H_MAIRIE : Humidité en Mairie
- T_MAIRIE : Température en Mairie
- R_MAIRIE : Précipitations en Mairie
- H_BUJA : Humidité à Bujumbura
- T_BUJA : Température à Bujumbura
- R_BUJA : Précipitations à Bujumbura

1.2 Données géographiques

1.2.1 Démographie de la zone

L'étude a utilisé les variables environnementales et démographiques qui sont la densité de la population en Mairie de Bujumbura, la densité de la population dans la province de Bujumbura, le niveau du lac Tanganyika et l'urbanisation dans la Mairie de Bujumbura. Les données utilisées couvrent une période de 13 ans, de 2010 à 2022, et ces données ont été enregistrées mensuellement. Ces données démographiques ont été recueillies auprès de l'Institut national de la statistique du Burundi (INSBU), anciennement connu sous le nom de l'ISTEEBU. Ces variables sont pour les données climatiques :

- N_TANGA : Niveau du lac Tanganyika
- D_MAIRIE : Densité de la population en Mairie
- D_BUJA : Densité de la population de Bujumbura
- URBA : Urbanisation en Mairie

1.2.2 Description de la zone d'Etude

Dans la ville de Bujumbura, plusieurs zones sont considérées comme étant à risque de montée des eaux et d'inondations. La ville est située près du lac Tanganyika et de la rivière Ruzizi, ce qui rend certaines parties de la zone sujettes aux inondations causées par les crues de ces cours d'eau. Les plaines adjacentes aux rives du lac Tanganyika, ainsi que les zones environnantes le long de la rivière Ruzizi, sont particulièrement vulnérables aux inondations [7].

Certains quartiers de Bujumbura sont situés dans des zones à basse altitude et sont souvent caractérisés par une infrastructure de drainage inadéquate et une capacité limitée à faire face aux excès d'eau.

Les zones côtières le long du lac Tanganyika sont très exposées aux inondations. Les quartiers sont densément peuplés et l'urbanisation est plus importante, en raison de l'imperméabilisation des sols, du manque d'espaces verts et de l'insuffisance des infrastructures de drainage. Les accumulations d'eau peuvent se produire plus facilement dans la zone en cas de fortes précipitations [28].

1.3 Localisation

La Mairie de Bujumbura est située au Burundi, en Afrique de l'Est, avec une latitude 3.3761° S et une longitude 29.3599° E. Elle est localisée au bord du lac Tanganyika, l'un

des plus grands lacs d'Afrique. La zone d'étude comprend la ville de Bujumbura qui est la capitale économique du pays et ses environs (zone Gatumba). Plusieurs rivières et cours d'eau traversent la zone d'étude.

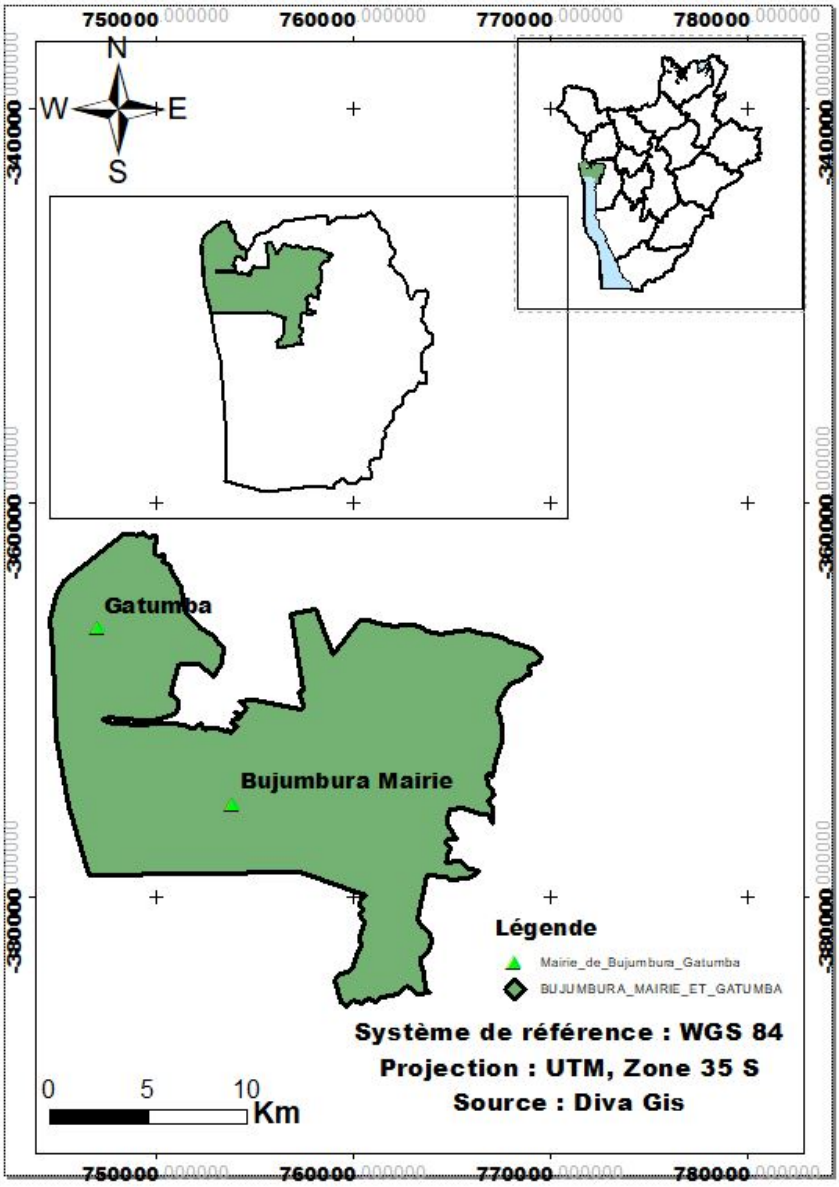


FIGURE 3.1 – Bujumbura Mairie et Zone Gatumba

2 Méthodes

2.1 Analyse de la statistique descriptive des données

Ces statistiques descriptives permettent de tester l'hypothèse de normalité de la distribution. La statistique de Jarque Bera teste l'hypothèse nulle (la série est normalement distribuée) et l'hypothèse alternative (la distribution de la série est anormale) [29]. La distribution sera normale pour les grandes valeurs de la probabilité, c'est-à-dire une probabilité supérieure à 0,05 [30]. La description statistique des différentes variables environnementales et démographiques est présentée dans le Tableau 3.1. Il y a 156 observations pour chaque variable. L'écart-type indique le degré de variabilité des données.

Tableau 3.1 – Analyse de la statistique descriptive des données

	N_TANGA	H_MAIRIE	T_MAIRIE	R_MAIRIE	D_MAIRIE	H_BUJA	T_BUJA	R_BUJA	D_BUJA	URBA
Mean	4.82	71.39	24.91	2.27	7425.39	73.42	20.79	3.38	560.39	1.54
Median	4.59	73.13	24.88	1.94	7440.03	75.74	20.85	3.45	561.50	2.00
Maximum	6.55	92.65	29.59	7.97	8516.55	85.58	22.43	11.63	642.74	2.00
Minimum	3.73	57.82	23.41	0.00	6276.49	58.03	19.06	0.00	473.69	1.00
Std. Dev.	0.68	5.55	0.70	1.84	698.74	5.72	0.62	2.38	52.73	0.50
Skewness	0.90	0.11	1.79	0.74	−0.06	−0.68	−0.36	0.31	−0.06	−0.15
Kurtosis	2.80	4.47	14.50	3.05	1.80	2.64	3.50	2.62	1.80	1.02
Jarque-Bera	21.19	14.32	943.04	14.37	9.51	12.97	4.94	3.37	9.51	26.00
Observations	156.00	156.00	156.00	156.00	156.00	156.00	156.00	156.00	156.00	156.00

2.2 Modèle de Régression linéaire multiple (RLM)

La régression linéaire multivariée (RLM) a été utilisée comme base de référence pour l'évaluation de la montée des eaux du lac Tanganyika et aussi pour l'évaluation de la concurrence. Les modèles de régression linéaire sont des modèles simples qui ont des paramètres linéaires et non linéaires pour les prédictions [31]. Pour les échantillons de petite taille, les modèles paramétriques de régression multilinéaire (RLM) sont capables d'établir les relations entre les variables prédictives et la variable dépendante à l'aide de l'ajustement des moindres carrés [32]. Dans cette étude, le niveau d'eau dépend des facteurs climatiques, de la densité de population et de l'urbanisation. Le modèle général de RLM s'exprime comme suit dans l'équation (3.1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \tag{3.1}$$

où :

- y : Variable dépandante
- β_0 : Constance de la régression
- $\beta_1, \beta_2, \beta_3, \ldots, \beta_n$: coefficients de régression multiple
- $x_1, x_2, x_3, \ldots, x_n$: Variables indépendentes / Variables prédictives
- e = Erreur

Après avoir explicité les hypothèses nécessaires et les termes du modèle, les notions d'estimation des paramètres du modèle , de prévision par intervalle de confiance, la signification des tests d'hypothèse sont discutées de même que les outils de diagnostics.

Le modèle linéaire multiple, une fois ces étapes franchies, nous a permis d'avoir un large choix pour la distribution de la variable à expliquer [33]. Nous avons également souligné l'importance de l'utilisation du logiciel R pour mener ces analyses de manière efficace. R offre une grande variété de packages et de fonctions dédiées à la modélisation statistique, ce qui facilite la mise en œuvre du modèle linéaire multiple ainsi que des analyses complémentaires pour en explorer toutes les facettes. De plus nous avons utilisé le logiciel de Cartographie appelé Ggis pour la localisation de la zone d'étude ainsi que le logiciel Eviews pour montrer l'allure de chaque variable intervenant dans notre étude.

2.3 Modèle de Random Forest

Afin d'améliorer la qualité de prédiction de la montée du niveau du lac Tanganyika, nous allons estimer un modèle Random Forest qui est parmi les méthodes d'ensemble.

En effet, le modèle Random Forest est un ensemble des arbres de décision estimés sur les différents sous-ensembles des données et des variables sélectionnées aléatoirement [34]. Enfin, pour prédire, le modèle Random Forest calcule la moyenne des valeurs prédites par ces différents arbres de décision. Étant donné qu'on ne dispose pas beaucoup de données pour les séparer en données d'entraînement et en données de test, nous avons estimé le modèle avec la technique de validation croisée K-Fold. Où K est égal à 10. C'est-à-dire que l'ensemble de nos données est divisé en aléatoire en 10 blocs et plusieurs modèles sont entraînés sur les différents 9 blocs et sont évalués sur le bloc qui n'a pas participé à l'entraînement [35]. Les indicateurs de performance comme la racine carrée de l'erreur quadratique moyenne (RMSE : Root Mean Squared Error) seront obtenus en calculant leur moyenne.

2.4 Formulation du modèle

Supposons que y soit une variable dépandante et que x_i $(i = 1, 2, \cdots, k)$ soient des variables indépendantes. Prenons que la valeur de la variable y soit la fonction linéaire de la variable indépendante x_i , et modifions la précision de l'erreur selon la variable d'erreur aléatoire indépendante e_i. L'hypothèse du modèle linéaire multiple est :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ \measuredangle y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n \end{aligned} \tag{3.2}$$

Sous forme matricielle, on a :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3.3}$$

où :

- y_i : Niveau d'eau du lac Tanganyika (variable dépendante observée)
- n : taille de l'échantillon avec i = 1, ..., 156
- x_i =H_MAIRIE, T_MAIRIE, R_MAIRIE, D_MAIRIE, H_BUJA, T_BUJA, R_BUJA, D_BUJA, URBA (variables prédictives explicatives)
- ϵ_i : résidu ou erreur pour l'individu
- β_i : coefficients de régression multiples

Dans l'équation 3.3, le niveau d'eau est la variable dépendante (y_i), déterminée par un ensemble de variables prédictives suivant l'équation (3.4) :

$$\begin{aligned} N_TANGA_t = \beta_0 &+ (\beta_1 \cdot H_MAIRIE_t) + (\beta_2 \cdot T_MAIRIE_t) + (\beta_3 \cdot R_MAIRIE_t) \\ &+ (\beta_4 \cdot D_MAIRIE_t) + (\beta_5 \cdot H_BUJA_t) + (\beta_6 \cdot T_BUJA_t) \\ &+ (\beta_7 \cdot R_BUJA_t) + (\beta_8 \cdot D_BUJA_t) + (\beta_9 \cdot URBA_t) + e_t \end{aligned} \tag{3.4}$$

Dans le contexte du modèle linéaire multidimensionnel, une variable quantitative y_i était expliquée par plusieurs variables quantitatives x_i, où i indiquait les observations et j ($j = 1, \ldots, p$) représentait les différentes variables explicatives. Le but de ce modèle était de relier la variable y à un ensemble de variables explicatives $x_1, \ldots, x_p$. Par exemple, dans notre cas, la variable y pouvait représenter la montée des eaux du lac Tanganyika, tandis que $x_1, \ldots, x_p$ pouvaient être des variables internes explicatives telles que les précipitations, l'urbanisation, la densité de la population, la température et l'humidité.

Dans cette situation, la variable quantitative y que l'on cherchait à expliquer était également appelée variable "endogène", "dépendante" ou "réponse", tandis que les variables $x_1, \ldots, x_p$ étaient désignées comme les variables "exogènes", "indépendantes" ou "régressives", car elles agissaient comme des facteurs explicatifs conditionnellement à leur observation.

L'ajout de "t" dans l'équation du modèle linéaire multidimensionnel était justifié par l'idée de modéliser la variable à expliquer y en fonction des variables explicatives x_1, ..., x_p et du temps t". L'ajout de la dimension temporelle permettait de prendre en compte l'évolution ou la variation de la variable y et des variables explicatives au fil du temps, ce qui pouvait être crucial pour capturer des tendances, des saisons ou des changements qui pouvaient influencer les relations entre les différentes variables. Ainsi, en incluant la dimension temporelle dans le modèle, on cherchait à améliorer la capacité du modèle à expliquer et prédire la variable y en intégrant le facteur temps dans l'analyse.

Chapitre 4

Résultats et leurs Interprétations

Ce chapitre présente les résultats statistiques obtenus de l'analyse des données. La statistique a fourni les résultats suivants, et nous allons maintenant les interpréter pour mieux comprendre les relations entre les variables et les implications pour notre étude.

1 Analyse de l'évolution et de la distribution du Niveau du Lac Tanganyika de 2010 à 2022

Les deux graphiques suivants nous aident à comprendre respectivement le comportement du niveau du Lac Tanganyika au cours du temps et la distribution statistique de ce niveau.

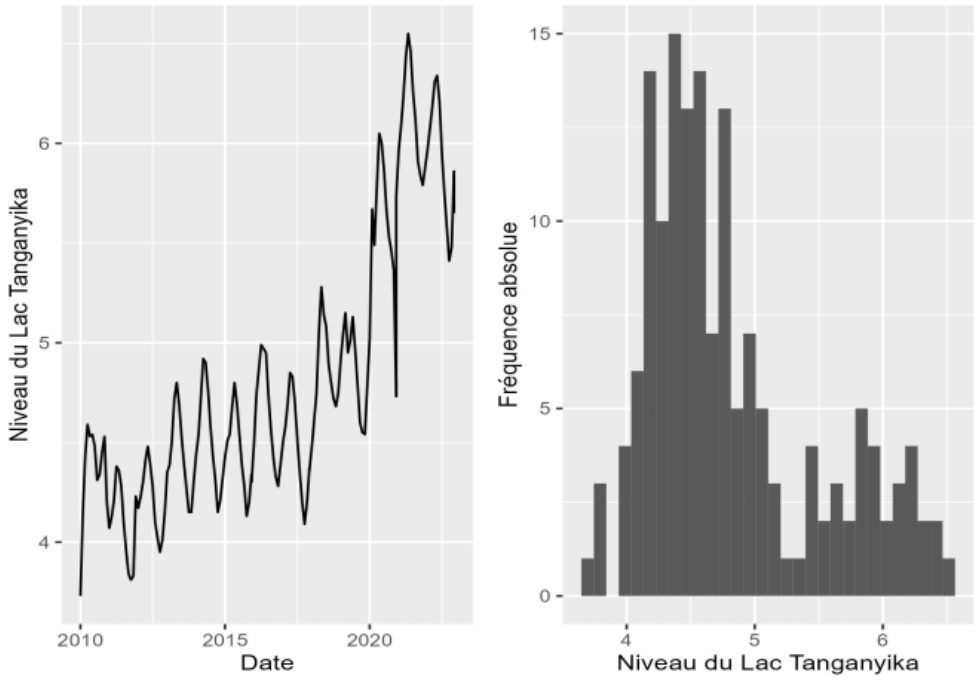


FIGURE 4.1 – Niveaux du lac Tanganyika selon la période de 2010 à 2022

Pour le graphique situé à gauche, le chronogramme indique que le niveau du lac Tanganyika admet une tendance relativement linéaire et un caractère périodique. Néanmoins, cette tendance semble avoir deux allures différentes : Le taux de croissance de la période d'avant 2020 semble brusquement changer à partir de 2020. Cela pourrait indiquer qu'il y a eu un changement au niveau de la distribution statistique qui génère la montée du niveau du Lac Tanganyika, l'histogramme à côté du chronogramme montre qu'effectivement la distribution statistique a changé depuis 2020. En effet, les niveaux du Lac qui sont supérieurs à 5.02 mètres (Niveau du Lac en Janvier 2020) semblent avoir une autre distribution statistique différente de celle d'avant 2020.

2 Analyse uni-variée de la distribution des variables explicatives

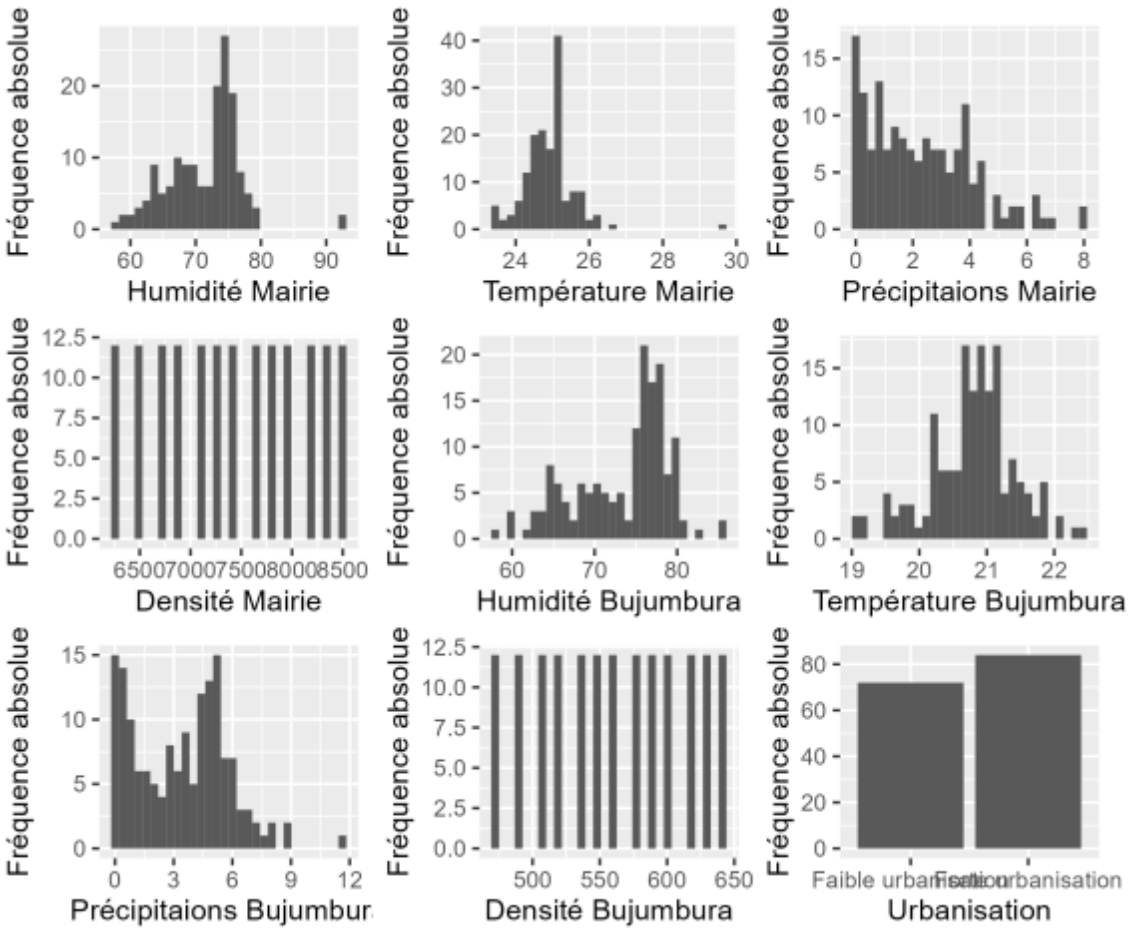


FIGURE 4.2 – Analyse uni-variée de la distribution des variables explicatives

La distribution de ces variables ne suit pas une loi normale, mais les valeurs se situant entre 73 et 75 pour l'humidité_mairie sont les plus observées, aussi la température, mais aussi presque toutes les distributions sont assymétriques. Pour les précipitations, on constate que certaines valeurs n'ont pas été bien observées, la distribution est presque uniforme.
Après avoir effectué une analyse uni-variée de la distribution des variables explicatives, nous avons examiné les liens dans une analyse bivariée entre les variables indépendantes et le niveau du Lac Tanganyika.

3 Analyse bivariée entre les variables indépendantes et le niveau du lac Tanganyika

Les nuages de points ci-dessous visualisent les relations qui existent entre les variables explicatives (Humidité, Température, Précipitations, Densité, pour la Mairie et la province de Bujumbura) et la montée des eaux du Lac Tanganyika. Tous ces graphiques comme

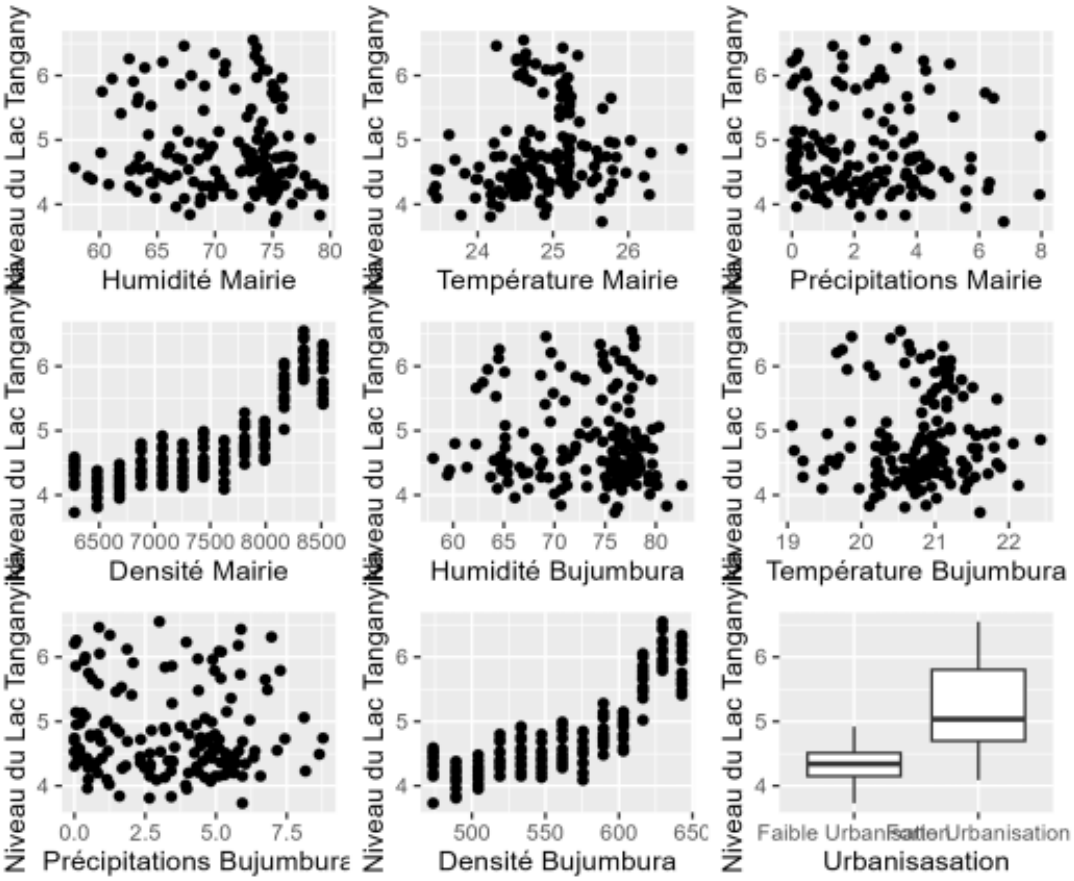


FIGURE 4.3 – Analyse bivariée de la distribution des variables explicatives

d'autres outils de la statistique descriptive ne permettent pas de conclure objectivement sur les relations existant entre les variables explicatives et la variable à expliquer. Cependant, ils nous permettent subjectivement d'avoir une idée sur ces relations. En effet, à partir de ces graphiques la densité Maire et la densité Bujumbura semblent évoluer dans le même sens que le niveau du Lac Tanganyika : L'augmentation du niveau du Lac Tanganyika est corrélée à l'augmentation de la densité de la population. Ainsi, le dernier graphique montre que plus il y a une forte urbanisation plus il y a la montée du Lac Tanganyika. Pour d'autres variables les relations linéaires ne semblent pas être évidentes.

4 Estimation du modèle linéaire simple pour chaque variable explicative

Les résultats des régressions linéaires simples effectuées pour chaque facteur potentiellement explicatif de la montée du niveau du lac Tanganyika sont présentés dans le tableau ci-dessus.

Tableau 4.1 – Estimation du modèle linéaire simple

Characteristic	N	Beta	95% CI[1]	p-value
H_MAIRIE	153	-0.0186	-0.0402, 0.0029	0.090
T_MAIRIE	153	0.1625	-0.0187, 0.3436	0.078
R_MAIRIE	153	-0.0240	-0.0828, 0.0348	0.4
D_MAIRIE	153	0.0008	0.0007, 0.0009	<0.001
H_BUJA	153	-0.0069	-0.0264, 0.0127	0.5
T_BUJA	153	0.0188	-0.1583, 0.1959	0.8
R_BUJA	153	-0.0171	-0.0627, 0.0286	0.5
D_BUJA	153	0.0105	0.0094, 0.0117	<0.001
URBA	153			
Faible urbanisation		—	—	
Forte urbanisation		0.8787	0.7112, 1.0462	<0.001

[1] $CI = ConfidenceInterval$

Les coefficients Beta indiquent la direction et l'intensité de la relation entre chaque variable explicative et le niveau du lac. Les intervalles de confiance à 95% (CI) montrent l'incertitude autour de ces estimations, et les valeurs de p ou p-value indiquent la signification statistique de chaque coefficient.

Les variables telles que l'Humidité en Mairie, les Précipitations en Mairie, l'Humidité à Bujumbura, les Précipitations à Bujumbura et l'Urbanisation présentent une relation négative avec la montée du niveau du lac Tanganyika. Cependant, ces relations ne sont pas statistiquement significatives au seuil de 5%. En revanche, les facteurs tels que la Température en Mairie, la Densité en Mairie, la Température à Bujumbura et la Densité à Bujumbura montrent une augmentation du niveau du lac Tanganyika avec l'augmentation de leurs valeurs. Seules la densité de population en Mairie et la densité de population à Bujumbura ont une influence statistiquement significative sur la montée du lac Tanganyika, avec des valeurs de p inférieures à 5%.

Ainsi, ces résultats suggèrent que la densité de la population dans les zones de Mairie et de Bujumbura pourraient jouer un rôle important dans l'explication de la variation du niveau du Lac Tanganyika, tandis que d'autres facteurs tels que l'humidité, les précipitations, la température et l'urbanisation pourraient également être des facteurs à considérer, bien que leur impact ne soit pas statistiquement significatif dans cette étude.

5 Estimation du modèle linéaire multiple

Les résultats du tableau montrent les estimations des coefficients du modèle linéaire complet pour l'élévation du niveau du lac Tanganyika en fonction des différentes variables explicatives.

Tableau 4.2 – Estimation du modèle linéaire multiple

Characteristic	Beta	95% CI[1]	p-value
(Intercept)	-2.2	-5.3, 0.87	0.2
H_MAIRIE	-0.01	-0.05, 0.02	0.4
T_MAIRIE	0.15	0.01, 0.28	0.030*
R_MAIRIE	-0.01	-0.07, 0.04	0.6
D_MAIRIE	-2.5	-5.0, -0.10	0.041*
H_BUJA	0.02	-0.01, 0.05	0.2
T_BUJA	-0.23	-0.39, -0.06	0.007*
R_BUJA	0.00	-0.05, 0.06	0.9
D_BUJA	34	1.4, 66	0.041*
URBA			
Faible urbanisation	-	-	
Forte urbanisation	-0.35	-0.60, -0.10	0.006

[1] $CI = ConfidenceInterval$

Le tableau ci-dessus représentent les résultats du modèle linéaire multiple définie par l'équation suivante :

$$\begin{aligned} N_TANGA_t = \beta_0 &+ (\beta_1 \cdot H_MAIRIE_t) + (\beta_2 \cdot T_MAIRIE_t) + (\beta_3 \cdot R_MAIRIE_t) \\ &+ (\beta_4 \cdot D_MAIRIE_t) + (\beta_5 \cdot H_BUJA_t) + (\beta_6 \cdot T_BUJA_t) + (\beta_7 \cdot R_BUJA_t) \\ &+ (\beta_8 \cdot D_BUJA_t) + (\beta_9 \cdot URBA_t) + e_t \end{aligned} \tag{4.1}$$

L'humidité en Mairie (H_MAIRIE) n'a pas d'un impact significatif sur la montée du niveau du lac, avec un coefficient estimé de -0.01 et une p-value de 0,4.

La température en Mairie (T_MAIRIE) a un effet positif significatif sur l'élévation du niveau du lac, avec un coefficient de 0,15 et une p-value de 0,030.

Les précipitations en Mairie (R_MAIRIE) n'influencent pas de manière significative le niveau du lac, avec un coefficient estimé de -0,01 et une p-value de 0,6.

La densité en Mairie (D_MAIRIE) a un effet négatif significatif sur l'élévation du niveau du lac, avec un coefficient de -2,5 et une p-value de 0,041.

Les autres variables comme l'humidité à Bujumbura, la température à Bujumbura, les précipitations à Bujumbura et le niveau d'urbanisation ont des effets et des significations variables sur le niveau du lac, comme indiqué dans le tableau.

Le modèle global indique que la température à Bujumbura, la densité en Mairie et à Bujumbura, ainsi que le niveau d'urbanisation, ont des effets significatifs sur la montée du niveau du lac Tanganyika. Ces résultats suggèrent que ces facteurs peuvent jouer un rôle important dans l'explication des variations observées.

6 Tests de différentes hypothèses

6.1 Test de significativité global

La statistique t-Student permet de tester la significativité individuelle des paramètres, c'est-à dire la contribution de chaque variable prise individuellement et la statistique F de Fischer permet de tester la significativité globale des paramètres [36]. Le tableau 4.3 montre la contribution de chaque variable prise individuellement suivant le modèle.

Tableau 4.3 – Significativité individuelle des paramètres

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	475.6926	46.00873	10.33918	0.0000
H_MAIRIE	-0.006761	0.011919	-0.567214	0.5714
T_MAIRIE	0.093535	0.051409	1.819412	0.0709
D_MAIRIE	-60.48535	5.824403	-10.38482	0.0000
R_MAIRIE	-0.060940	0.021334	-2.856465	0.0049
H_BUJA	0.021957	0.011889	1.846817	0.0668
T_BUJA	-0.156138	0.061472	-2.539962	0.0121
R_BUJA	0.008467	0.020419	0.414654	0.6790
D_BUJA	0.121835	0.010456	11.65264	0.0000
URBA	-0.285832	0.097029	-2.945841	0.0037

Le tableau ci-dessus présente les coefficients, les erreurs standard, les statistiques t et les valeurs de probabilité pour chaque variable du modèle. Une valeur de probabilité (Prob.) inférieure à 0,05 est généralement considérée comme statistiquement significative.

La probabilité « Prob(F-statistic) » égale à 0,000000 suggère qu'il y a au moins un coefficient dans le modèle qui est non nul, ce qui signifie qu'au moins une variable explique de manière significative la variabilité des données observées. Cela indique également que le modèle global est statistiquement significatif et que les variables incluses ont contribué ensemble de manière significative.

Le tableau présenté donne différentes statistiques de régression qui sont importantes pour évaluer la qualité du modèle utilisé pour analyser les données. En ce qui concerne le R-

Tableau 4.4 – Différentes statistiques de régression

R-squared	0.835674
Mean dependent var	4.815841
Adjusted R-squared	0.825544
S.D. dependent var	0.675011
S.E. of regression	0.281938
Akaike info criterion	0.367698
Sum squared resid	11.60542
Schwarz criterion	0.563202
Log likelihood	-18.68045
Hannan-Quinn criter.	0.447103
F-statistic	82.49725
Prob(F-statistic)	0.000000

squared, un coefficient élevé comme 0,836 indique que le modèle est assez efficace pour expliquer une grande partie de la variation de la variable dépendante à l'aide des variables indépendantes incluses. Cela suggère que le modèle est globalement pertinent et que les variables indépendantes contribuent de manière significative.

6.2 Test d'autocorrelation

La multicollinéarité est un phénomène dans lequel deux ou plusieurs variables indépendantes dans un modèle de régression sont fortement corrélées entre elles, ce qui peut poser des problèmes lors de l'estimation des coefficients du modèle. La multicollinéarité peut rendre les coefficients instables, réduire la précision des estimations, et ainsi rendre difficile l'interprétation des effets des variables sur la variable dépendante.

La statistique de Durbin-Watson, d'autre part, est utilisée pour détecter l'autocorrélation des erreurs, c'est-à-dire la corrélation sérielle des résidus du modèle. Une valeur de Durbin-Watson proche de 2 indique une absence d'autocorrélation des erreurs [37].

Dans le cas présent, nous avons mentionné que la valeur de la statistique de Durbin-Watson est de 0,453371 selon le tableau , ce qui suggère une forte autocorrélation positive des erreurs dans le modèle de régression. Cette information met en lumière un problème d'autocorrélation des erreurs, mais n'aborde pas directement le problème de la multicollinéarité.

Il est possible d'avoir simultanément des problèmes d'autocorrélation des erreurs et de multicollinéarité dans un modèle de régression. Il est important de diagnostiquer et de traiter ces problèmes séparément pour améliorer la fiabilité des résultats du modèle.

Tableau 4.5 – Différentes statistiques **de validité** d'une régression

R-squared	0.835674
Mean dependent var	4.815841
Adjusted R-squared	0.825544
S.D. dependent var	0.675011
S.E. of regression	0.281938
Akaike info criterion	0.367698
Sum squared resid	11.60542
Schwarz criterion	0.563202
Log likelihood	-18.68045
Hannan-Quinn criter.	0.447103
F-statistic	82.49725
Durbin-Watson stat	0.453371
Prob(F-statistic)	0.000000

Cette valeur de Durbin Watson est faible et fait penser à un problème d'autocorrélation positive. Après avoir consulté la table de Durbin Watson, les valeurs de d1 et d2 sont respectivement 1,57 et 1,78 [38].

Le tableau 4.6 presente les resultats du test de Breusch-Godfrey.

Tableau 4.6 – Breusch-Godfrey Serial Correlation LM Test(Multiplicateur de Lagrange)

Test Statistic	Value	Probability
F-statistic	369.8037	0.0000
Obs*R-squared	112.0609	0.0000

Le test de Breusch-Godfrey permet de tester l'hypothèse à l'aide de la statistique F de Fisher. La correction se matérialise par l'accroissement de la valeur de la statistique de Durbin Watson comme nous pouvons le remarquer ci-après :

Tableau 4.7 – Correction finale d'autocorrelation

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	412.8001	125.4968	3.289328	0.0013
H_MAIRIE	-0.002025	0.005536	-0.365863	0.7150
T_MAIRIE	0.027014	0.022478	1.201821	0.2314
D_MAIRIE	-52.27817	15.87250	-3.293632	0.0012
R_MAIRIE	-0.005920	0.009094	-0.650988	0.5161
H_BUJA	0.008005	0.005899	1.356951	0.1769
T_BUJA	-0.055045	0.031475	-1.748862	0.0824
R_BUJA	-0.002121	0.009442	-0.224626	0.8226
D_BUJA	0.102977	0.028388	3.627472	0.0004
URBA	0.053645	0.155083	0.345908	0.7299
AR(1)	0.869218	0.045447	19.12592	0.0000

Dans le tableau 4.7, la statistique t de Student permet de tester la significativité individuelle des paramètres. On conclut que **D_Mairie** et **D_Buja** sont significatives dans la montée du lac Tanganyika et leur contribution est respectivement négative et positive (leurs probabilités sont inférieures à 5%).

6.3 Test d'hétéroscedasticité

Le test est mené à l'aide de la statistique de Fisher (F-Statistics) suivant :

- H0 : Les erreurs sont homoscédastiques.
- H1 : Les erreurs sont hétéroscédastiques.

Toutefois, il importe d'utiliser aussi le test ARCH qui est considéré comme étant le plus déterminant comme le montre les résultats obtenus dans le tableau ci-après tenant aussi compte de R-carré (nombre d'observations multiplié par le R^2) :

Tableau 4.8 – Test d'hétéroscédasticité (ARCH)

Test Statistic	Value	Probability
F-statistic	0.061430	0.8046
Obs*R-squared	0.062213	0.8030

Les résultats relatifs au test d'hétéroscédasticité ARCH se trouvant au début du tableau 4.8 montrent la probabilité associée à la statistique F de Fisher pour ce test (Prob. F(1,152)= 0,8046). L'hypothèse H0 est acceptée : il n'y a pas de problème d'hétéroscédasticité.
Il est également mentionné que le test ARCH est considéré comme étant le plus déterminant en matière de détection de l'hétéroscédasticité. Dans votre cas, les valeurs du test

statistique F et de l'observation multipliées par le R-carré montrent des valeurs de probabilité élevées, respectivement 0,8046 et 0,8030, ce qui renforce la conclusion selon laquelle il n'y a pas de problème d'hétéroscédasticité dans le modèle.

6.4 Test de normalité des résidus

H_0 :les résidus suivent une loi normale
H_1 :les résidus ne suivent pas une loi normale

- Shapiro-Wilk normality test data : mod_comp residuals $W = 0.98702$, p-value = 0.1554
- Jarque Bera Test data : mod_comp residuals X-squared = 3.2655, df = 2, p-value = 0.1954

6.5 Test de spécification du modèle (Test de Ramsey)

Le résultat du test de Ramsey a donné la probabilité associée à la statistique F de Fisher = 0,7081, ce qui stipule que le modèle est bien spécifié [40], c'est à dire que les deux hypothèses suivantes ont été vérifié :

- H0 : Le modèle est bien spécifié
- H1 : Le modèle est mal spécifié

Tableau 4.9 – Test de spécification du modèle

	Value	df	Probability
t-statistic	0.375099	145	0.7081
F-statistic	0.140699	(1, 145)	0.7081
Likelihood ratio	0.151299	1	0.6973

Le test de spécification du modèle, tel que rapporté dans le tableau fourni, indique que la probabilité associée à la statistique F de Fisher est de 0,7081. Cette valeur de probabilité élevée suggère que l'hypothèse nulle selon laquelle le modèle est bien spécifié (H0) est acceptée. En d'autres termes, les résultats du test indiquent que le modèle est correctement spécifié, ce qui signifie que les variables indépendantes incluses dans le modèle contribuent de manière significative à expliquer la variable dépendante. Plus précisément, les différentes statistiques présentées dans le tableau : le t-statistique, la statistique F et le ratio de vraisemblance montrent des valeurs de probabilité similaires autour de 0,7081, renforçant ainsi la conclusion selon laquelle le modèle est bien spécifié. En utilisant le test

de Ramsey, qui examine la spécification du modèle, on peut confirmer que les hypothèses de spécification du modèle sont vérifiées et que le modèle est adéquat pour expliquer les relations entre les variables étudiées.

7 Election des facteurs explicatifs

Le modèle complet qu'on vient d'estimer est composé de plusieurs facteurs, comme les résultats obtenus le révèlent. Certains facteurs d'entre eux ont été jugés non explicatifs de la montée des eaux ; donc, il est bon d'éliminer les facteurs improbables (non significatifs) [41].
D'autres facteurs, malgré qu'ils soient explicatifs à la montée du niveau du lac, peuvent être redondants, c'est-à-dire que les deux facteurs ou plus peuvent rapporter la même information à l'explication de la montée des eaux. En outre, plus on augmente le nombre de facteurs dans un modèle linéaire, plus le modèle colle aux données de l'échantillon, mais perd la qualité de mieux expliquer le phénomène pour toute la population dont on n'a pas vu les données.
Afin d'éviter tous ces problèmes, on fait recours à la sélection des variables dans le but que seuls les facteurs indépendants entre eux et pertinents pour l'explication de la montée des eaux restent dans le modèle. Bien qu'il existe plusieurs méthodes pour la sélection des variables explicatives du phénomène, le choix de la technique basée sur la minimisation du critère BIC (Bayesian Information Criterion) est privilégié du fait que cette dernière pénalise l'augmentation du nombre de variables. Son expression est la suivante :

$$BIC = -2\ln(L) + \ln(n) \times (J + 1) \tag{4.2}$$

Où ln est le logarithme népérien, L est la vraisemblance du modèle, n le nombre total d'observations et J le nombre total de paramètres du modèle à estimer. La méthode de recherche des variables qui minimisent ce critère utilisée dans cette étude est celle de BACKWARD. Cette méthode consiste à l'élimination progressive des variables jusqu'à ce que ce critère devienne minimal.

Le tableau suivant nous décrit comment le processus s'est déroulé.

Le processus de sélection de modèle était observé, où plusieurs modèles ont été comparés en termes du critère d'information bayésien (BIC). Le modèle complet, qui inclut toutes les variables, avait un BIC de -303.7. Ensuite,les différents modèles ont été créés en éliminant progressivement des variables. Le premier modèle, qui a éliminé la variable R_BUJA, avait un BIC de -305.68, suivi du deuxième modèle avec l'élimination de la variable R_MAIRIE et un BIC de -307.34. Ce processus s'est répété jusqu'au quatrième

Tableau 4.10 – Model Selection

Modèles	Variables éliminées	BIC
Modèle complet	-	-303.7
1er modèle	R_BUJA	-305.68
2ème modèle	R_MAIRIE	-307.34
3ème modèle	H_MAIRIE	-308.41
4ème modèle	H_BUJA	-309.67

modèle, avec l'élimination successive des variables H_MAIRIE et H_BUJA, et un BIC en diminution de -309.67. Ces résultats suggèrent que le modèle final choisi avait un BIC optimal parmi ceux testés, et les variables T_MAIRIE, D_MAIRIE, T_BUJA, D_BUJA et URBA ont été retenues pour former le modèle final.

Tableau 4.11 – Estimation du modèle final

Characteristic	Beta	95% CI[1]	p-value
(Intercept)	-2.2	-4.8, 0.50	0.11
T_MAIRIE	0.17	0.04, 0.29	0.010
D_MAIRIE	-2.7	-5.1, -0.36	0.024
T_BUJA	-0.24	-0.38, -0.10	0.001
D_BUJA	36	4.8, 67	0.024
URBA			
Faible urbanisation	—	—	
Forte urbanisation	-0.36	-0.61, -0.12	0.004

[1] $CI = ConfidenceInterval$

Les résultats de l'estimation du modèle final sélectionné sont présentés dans le tableau 4.11. Les coefficients Beta associés à chaque variable, accompagnés de leurs intervalles de confiance à 95%, ainsi que les p-valeurs sont donnés. On constate que les variables T_MAIRIE, D_MAIRIE, T_BUJA, D_BUJA et URBA sont incluses dans le modèle final, et leur significativité est confirmée par des p-valeurs inférieures à 0.05. Cela signifie que ces variables ont un effet significatif sur la variable dépendante N_TANGA, et sont donc pertinentes pour expliquer les variations de cette dernière dans le modèle.

8 Prévision avec le modèle de régression linéaire multiple

La validation du modèle nous a permis de prédire la montée du lac Tanganyika jusqu'à l'an 2023 et les prévisions du niveau du lac Tanganyika sont reprensentées dans le tableau ci-après :

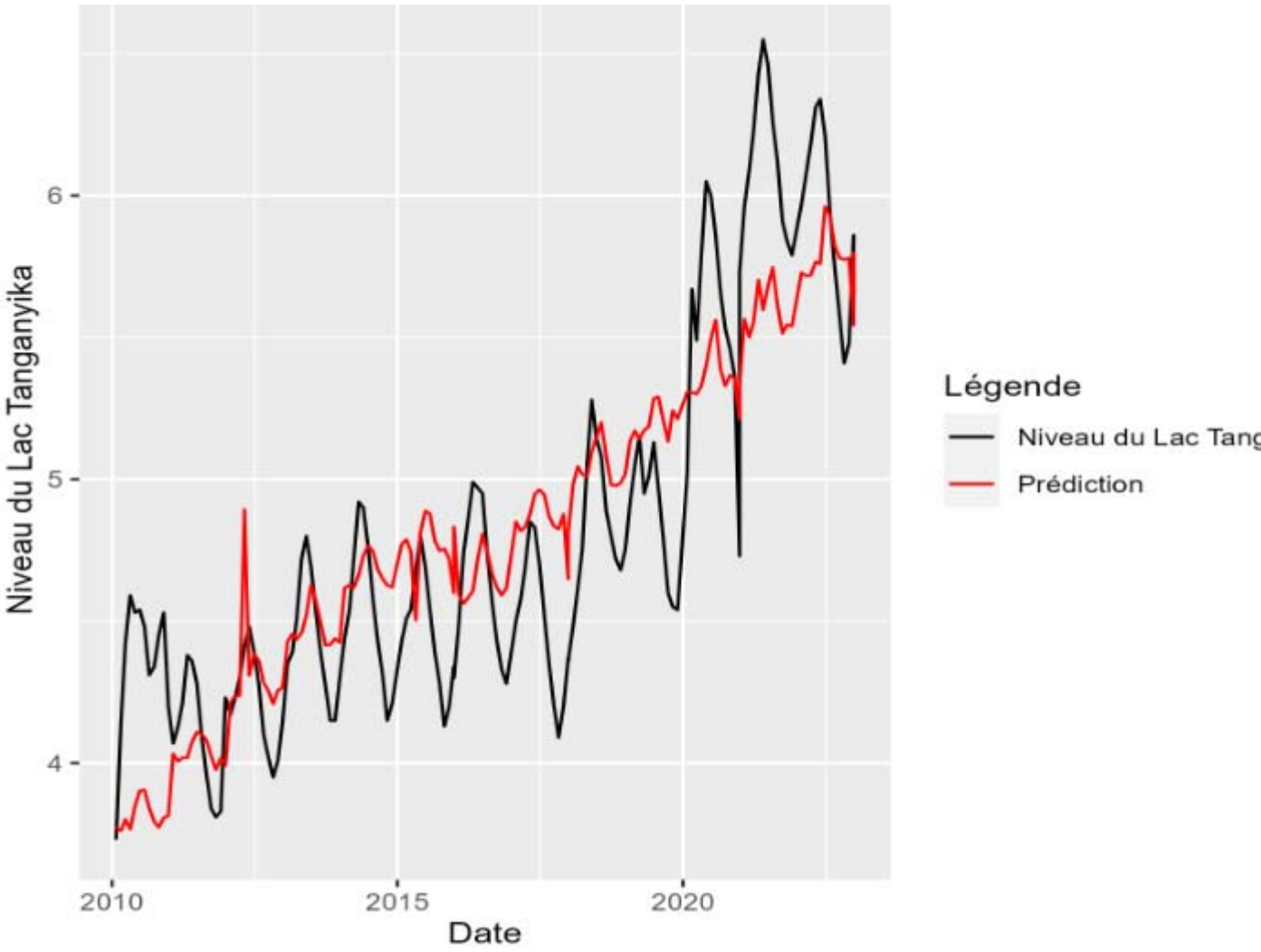


FIGURE 4.4 – Prédiction de la montée du lac Tanganyika

9 Estimation du modèle Random Forest

L'arbre de régression est donc l'arbre qui cherche à expliquer une variable quantitative. Alors qu'un modèle de régression va chercher une formule globale et commune pour expliquer la variable dépendante, l'arbre de régression va plutôt créer des regroupements constitués des classes les plus homogènes. Nous avons trouvé que 500 est le nombre d'arbres décisionnels qui minimisent le RMSE. Les résultats obtenus sont présentés ci-dessous.

TABLEAU 4.12 – Resampling Indicators

Resampling	Indicators	Values
Cross-Validation	RMSE	0.241
	R squared	0.8822823
	MAE	0.196756

Avec ces résultats obtenus, nous constatons que Random Forest améliore la qualité de prédiction relativement au modèle linéaire. En effet, l'indicateur de performance RMSE de Random Forest est maintenant égal à 0,241 contre 0,3566 du modèle linéaire ainsi que le R-Carré est maintenant à 88,22% contre 83,5% du modèle linéaire.
En confrontant graphiquement les valeurs observées du niveau du lac Tanganyika et ses valeurs prédites par le modèle Random Forest comme nous l'avons fait pour la régression linéaire, nous constatons que le chronogramme des valeurs prédites par Random Forest semble être bien ajusté au chronogramme des valeurs réelles du niveau du lac Tanganyika.

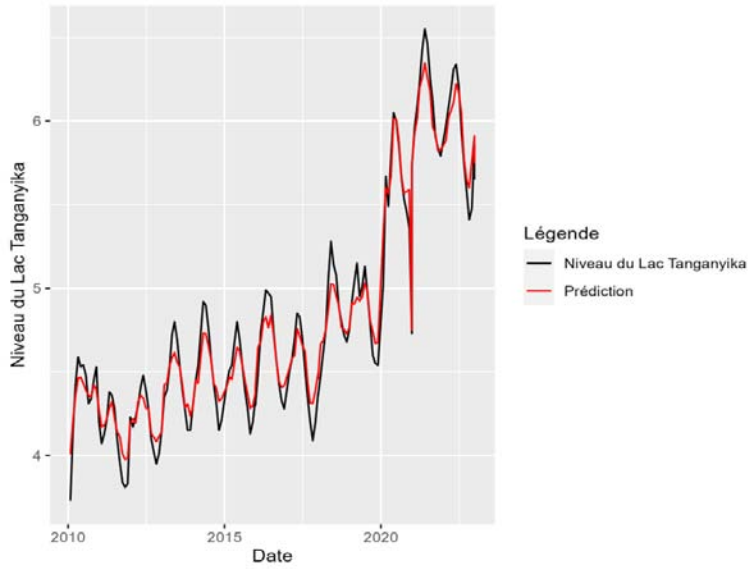


FIGURE 4.5 – Chronogramme des valeurs prédites par Random Forest

Importance des Variables

Au-delà de l'amélioration de prédiction fourni par Random Forest, le modèle nous permet aussi de calculer la contribution de chaque variable dans la prédiction du niveau du lac Tanganyika. La figure ci-dessous nous montre l'importance de chaque variable.

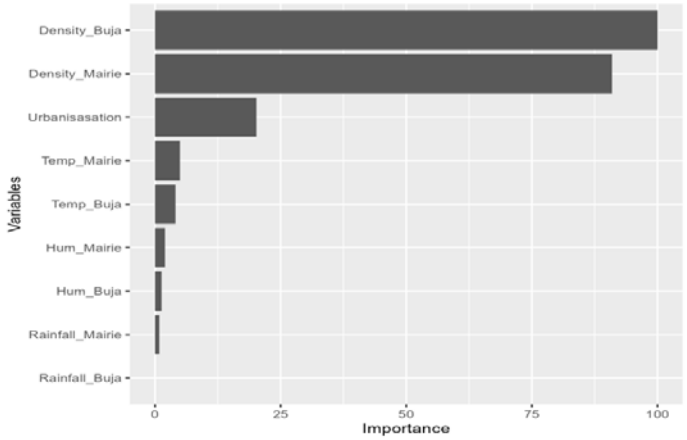


FIGURE 4.6 – Importance de chaque variable

A partir de ces résultats, on constate que les variables comme D_Maire (90.95) et D_Buja (100.00) ont une grande importance dans la prédiction du niveau du lac Tanganyika comme le montre la présentation (Figure 6).

Chapitre 5

Discussion et Conclusion

1 Discussion

Les résultats de notre analyse descriptive ont permis de tester l'hypothèse de normalité de la distribution des données. Les tests de Jarque Bera et Shapiro-Wilk étaient utilisés pour évaluer et confronter cette hypothèse, où H0 stipule une distribution normale et H1 une distribution anormale [29]. Si la probabilité associée à ces statistiques était inférieure à 5% (0,05), l'hypothèse nulle est rejetée [42].
En utilisant la statistique t de Student, nous avons pu évaluer la significativité individuelle de chaque paramètre, tandis que la statistique de Fischer (F) permet d'évaluer la significativité globale des paramètres [43]. Dans notre étude, le test de l'hypothèse nulle (tous les coefficients nuls) pour vérifier la significativité globale a été rejeté, avec une probabilité Prob(F-statistic) inférieure à 0.001, indiquant qu'au moins un coefficient est non nul [44].
L'analyse de la contribution de chaque variable dans le modèle final a révélé que la température en mairie (T_MAIRIE), la densité en mairie (D_MAIRIE), la température à Bujumbura (T_BUJA), la densité à Bujumbura (D_BUJA) et l'urbanisation (URBA) ont joué un rôle dans l'augmentation du niveau du lac Tanganyika, leurs coefficients étant statistiquement significatifs à un seuil de 5% (probabilités inférieures à 5%). Ces résultats ont mis en lumière un possible problème d'autocorrélation, étant donné que la valeur estimée était significativement éloignée de 2, avec une valeur de 0.453371 selon Durbin Watson [45].
Dans l'interprétation des résultats, il est crucial de tenir compte des éventuels problèmes tels que l'autocorrélation des erreurs ou l'hétéroscédasticité qui pourraient violer les hypothèses classiques [46]. Le coefficient de détermination R^2 nous indique que les variables exogènes expliquent à 83,5% des variations de la variable endogène N_TANGA, reflétant l'importance de ces variables dans le modèle [47].

En analysant les signes des coefficients et leur significativité, nous avons observé des re-

lations positives et négatives entre les variables explicatives et la variable à expliquer. T_MAIRIE, H_BUJA, R_BUJA, D_BUJA ont une influence positive et H_MAIRIE, D_MAIRIE, R_MAIRIE, T_BUJA et URBA ont montré une influence négative sur la variable dépendante N_TANGA.
Les tests de significativité, à la fois globaux et individuels, ont confirmé la qualité du modèle. La probabilité associée à F-Stat était de 0.00000, indiquant un modèle globalement fiable. En revanche, au niveau individuel, les variables H_MAIRIE, R_MAIRIE et R_BUJA ont montré des coefficients statistiquement nuls, tandis que T_MAIRIE, D_MAIRIE, H_BUJA, T_BUJA, D_BUJA et URBA ont influencé de manière significative le niveau du lac Tanganyika.
Les résultats de l'estimation du modèle Random Forest montrent une nette amélioration de la qualité de prédiction par rapport au modèle de régression linéaire précédemment étudié. En effet, l'indicateur de performance RMSE du modèle Random Forest est désormais de 0,241, contre 0,3566 pour le modèle linéaire. De même, le coefficient de détermination R-carré atteint 88,22%, contre 83,5% pour le modèle linéaire. Cette amélioration substantielle des performances prédictives s'explique par la capacité du Random Forest à mieux capturer les relations complexes entre les variables explicatives et la variable d'intérêt.
L'analyse graphique du chronogramme des valeurs prédites par le modèle Random Forest confirme ce constat. On observe en effet un ajustement très satisfaisant des valeurs prédites aux valeurs observées du niveau du lac Tanganyika. Cela démontre la pertinence de l'utilisation d'un modèle d'apprentissage automatique tel que le Random Forest pour modéliser un phénomène aussi complexe que la dynamique du niveau d'un lac.
Ces résultats soulignent l'intérêt d'explorer des approches alternatives à la régression linéaire classique lorsqu'on est confronté à des problématiques de prédiction de variables quantitatives. Le Random Forest, de par sa capacité à s'adapter à des relations non linéaires et à prendre en compte des interactions complexes entre variables, s'avère particulièrement adapté dans ce type de contexte.

2 Conclusion et recommandations

Cette étude sur la prévision du niveau du lac Tanganyika à Bujumbura a permis de mettre en évidence plusieurs éléments clés.

Tout d'abord, les résultats du modèle Random Forest ont montré une nette amélioration de la qualité de prédiction par rapport au modèle de régression linéaire initial. Avec un RMSE de 0,241 et un R-carré de 88,22%, le Random Forest s'est avéré particulièrement pertinent pour modéliser la dynamique complexe du niveau du lac. L'analyse graphique a confirmé ce constat, avec un ajustement très satisfaisant des valeurs prédites aux valeurs

observées.

Ces résultats ont souligné l'importance d'explorer des approches alternatives à la régression linéaire classique lorsqu'on était confronté à des problématiques de prévision de variables quantitatives. Les modèles d'apprentissage automatique tels que le Random Forest avaient en effet la capacité de mieux capturer les relations non linéaires et les interactions complexes entre variables.

Sur le plan opérationnel, les prévisions du niveau du lac pour l'année 2023, bien que montrant une tendance à la baisse, indiquent que le niveau reste toujours dans une dynamique de hausse. Il est donc essentiel de prendre en compte ces risques dans la planification urbaine et la gestion des inondations à Bujumbura.

Parmi les recommandations clés, on peut citer :

- L'amélioration des systèmes de drainage urbain
- La construction d'infrastructures résilientes aux inondations
- L'établissement de zones tampon naturelles le long des rives
- La sensibilisation des populations aux risques d'inondation
- La promotion de pratiques de construction adaptées dans les zones à risque
- Une gestion démographique raisonnée pour limiter l'exposition aux risques

L'étude de la montée des eaux à Bujumbura revêt d'une importance cruciale pour protéger les populations, les biens et les infrastructures, tout en planifiant un développement urbain durable et en s'adaptant aux effets du changement climatique. Les résultats et les recommandations de cette étude constituent des éléments clés pour guider les décideurs dans la mise en place de stratégies de gestion des risques d'inondation efficace.

ANNEXE

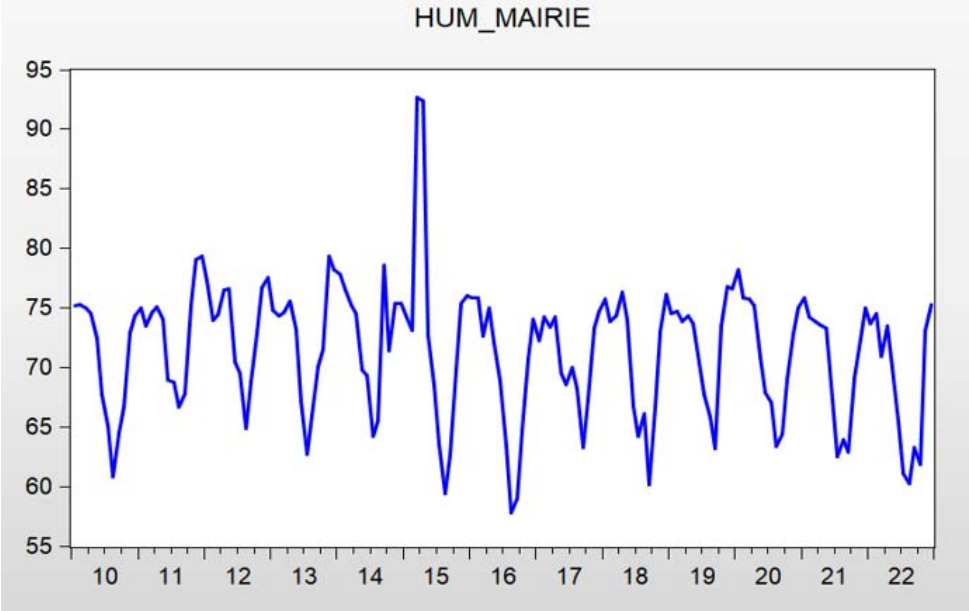


FIGURE 5.1 – Allure de l'humidité dans la Mairie de Bujumbura

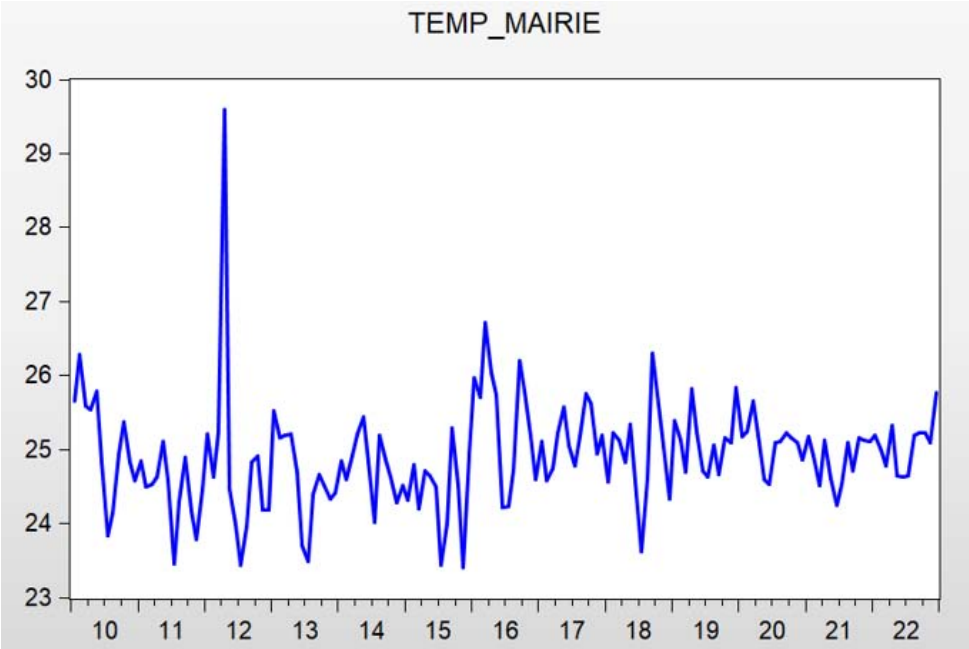


FIGURE 5.2 – Allure de la température dans la Mairie de Bujumbura

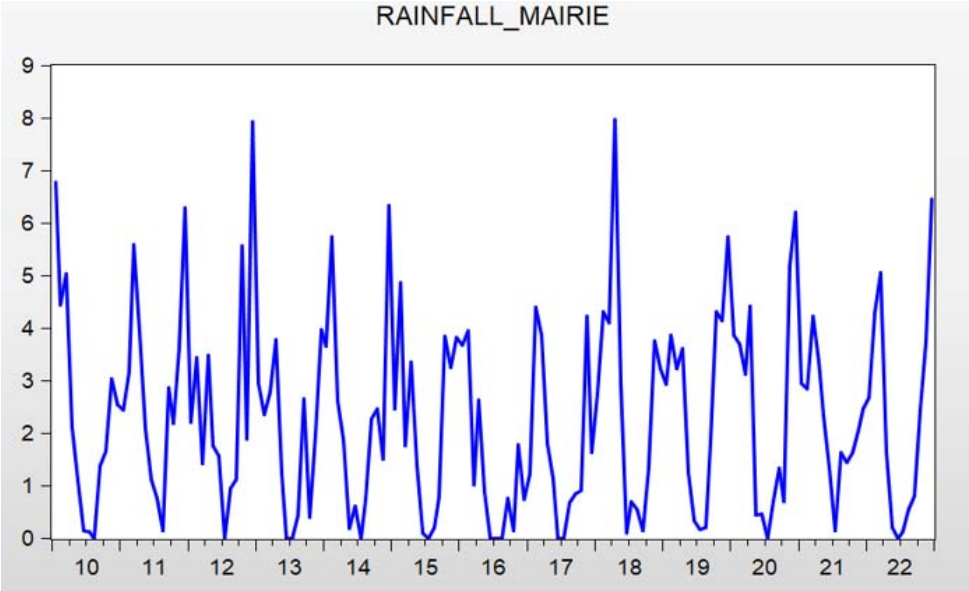


FIGURE 5.3 – Allure du rainfall dans la Mairie de Bujumbura

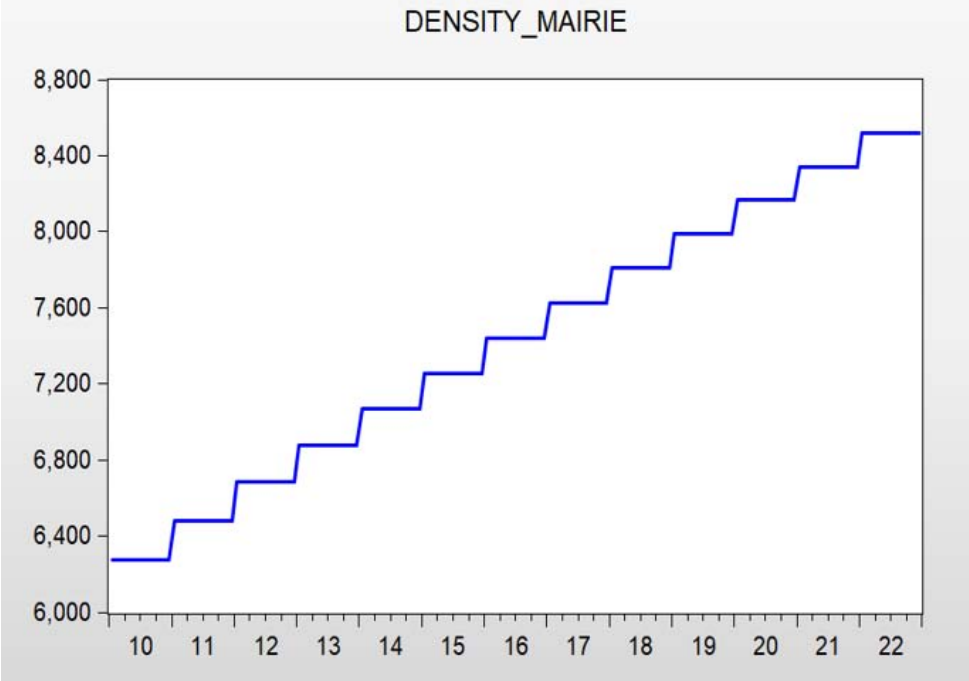


FIGURE 5.4 – Allure de la densitité dans la Mairie de Bujumbura

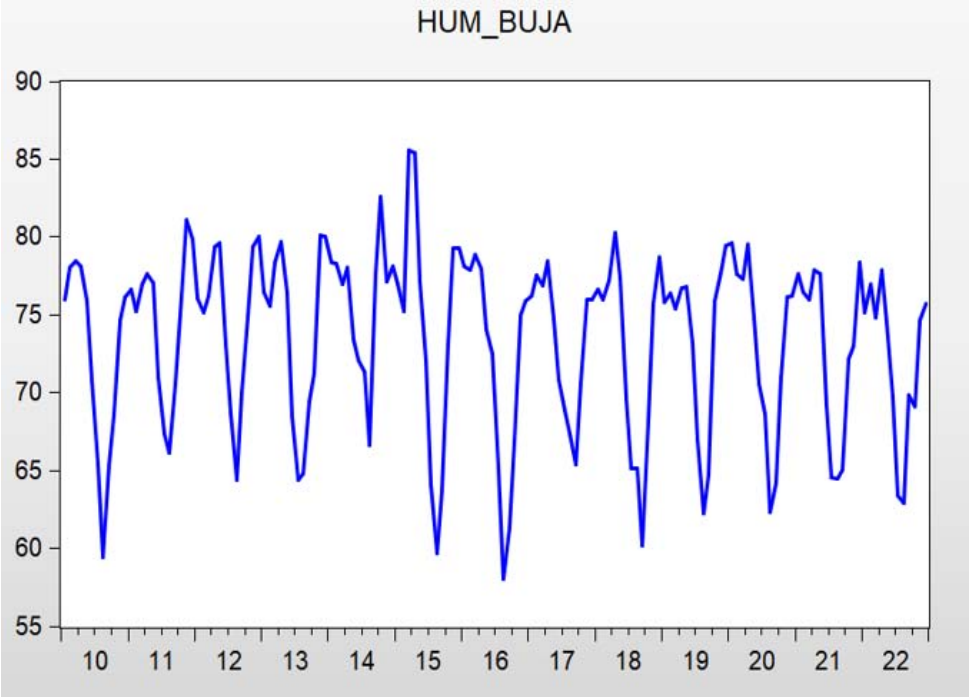


FIGURE 5.5 – Allure de l'humidité dans la Province de Bujumbura

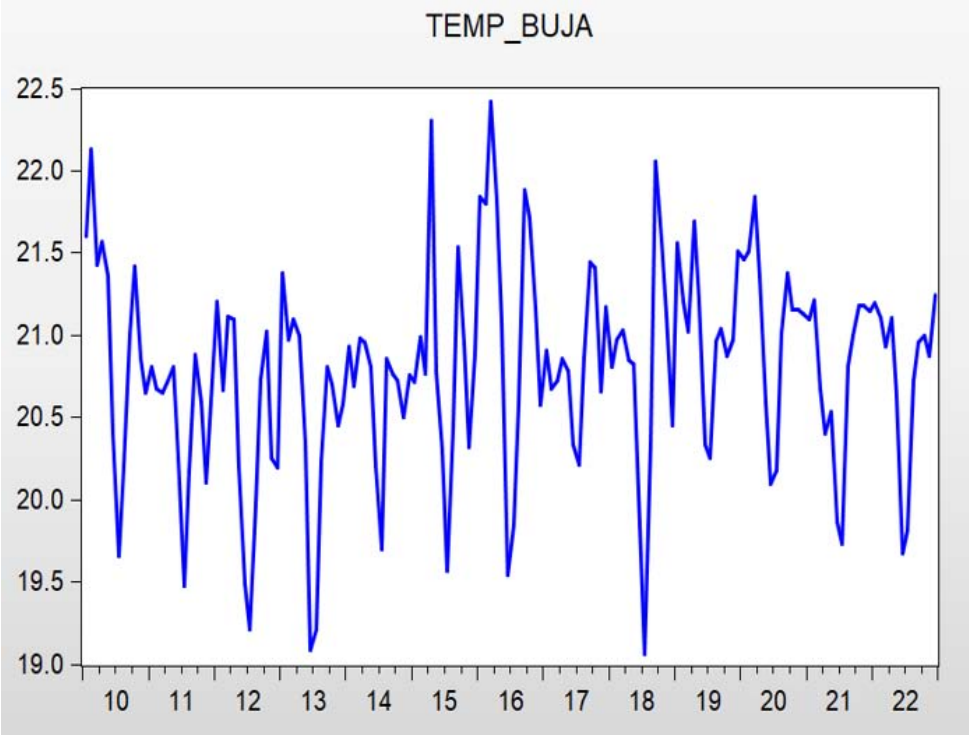


FIGURE 5.6 – Allure de la température dans la Province de Bujumbura

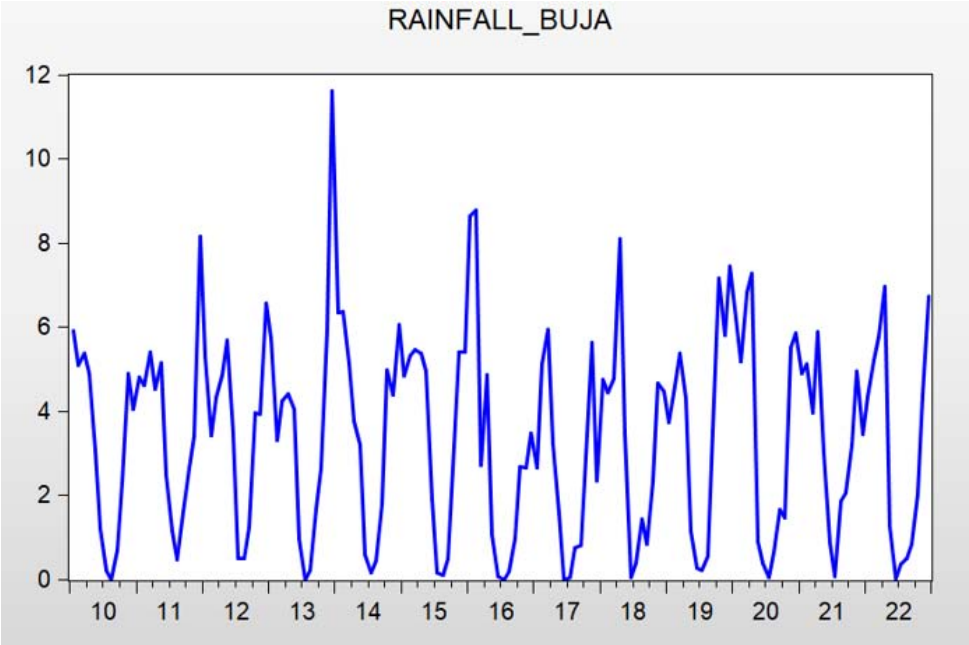


FIGURE 5.7 – Allure du rainfall dans la Province de Bujumbura

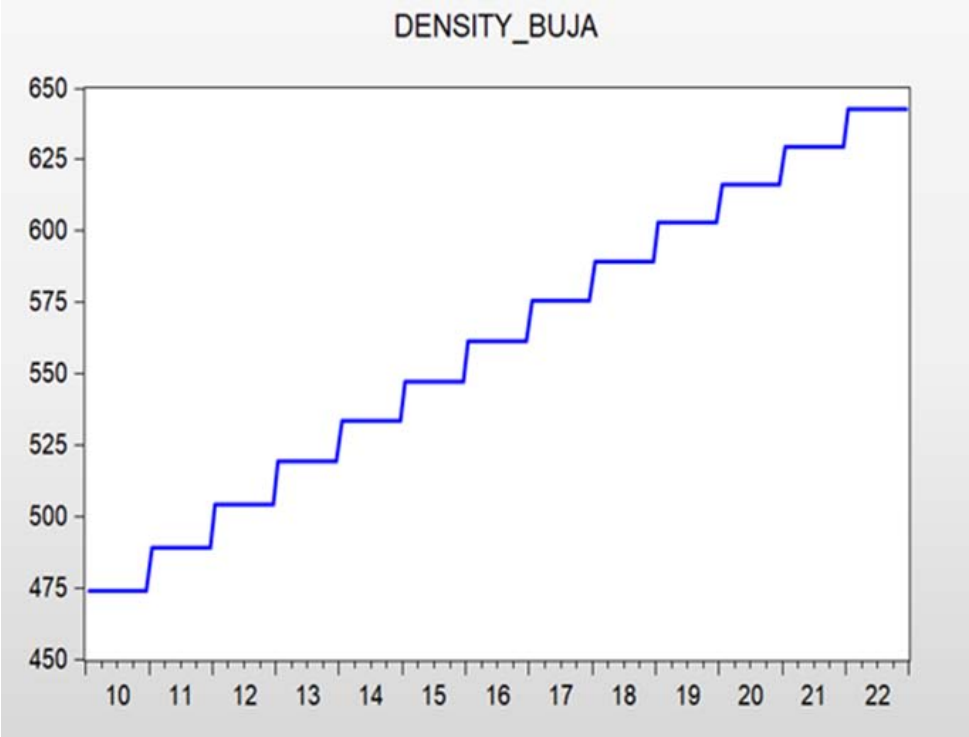


FIGURE 5.8 – Allure de la densitité dans la Province de Bujumbura

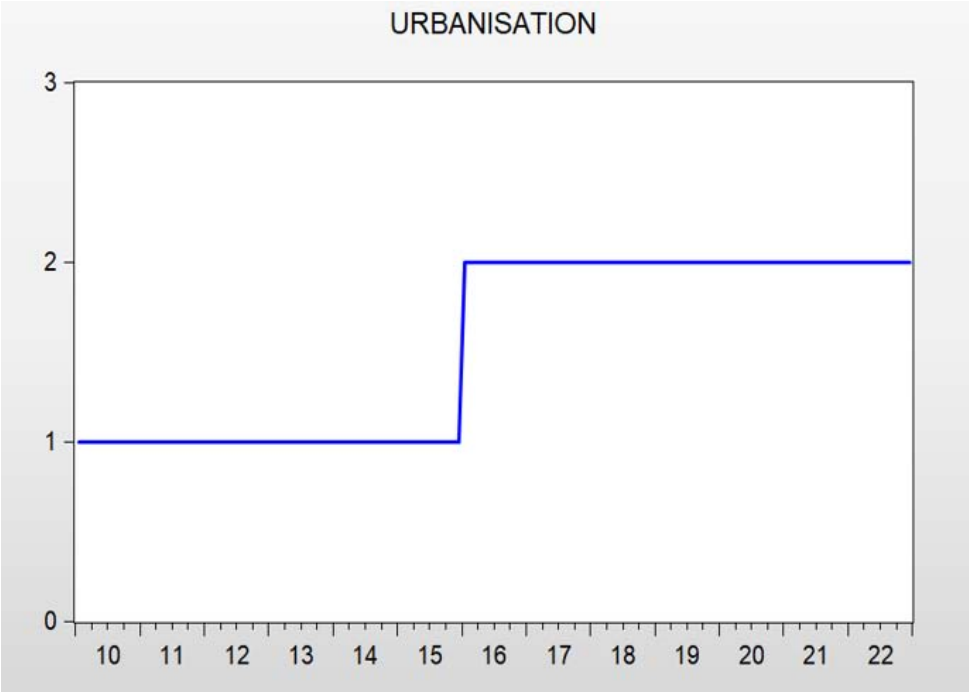


FIGURE 5.9 – Allure de l'urbanisation

Bibliographie

[1] Groupe interministériel Impacts du changement climatique, adaptation et coûts associés en France, Direction des Etudes Economiques et de l'Evaluation Environnementale, Observatoire National sur les Effets du Réchauffement Climatique.

[2] Boudon, R. (Ed.) (2016). Le vocabulaire des sciences sociales : concepts et indices (Vol. 1). Walter de Gruyter GmbH, Co KG.

[3] Dewals, B. J., Detrembleur, S., Archambeau, P., Erpicum, S., Ernst, J., & Pirotton, M. (2011). Caractérisation micro-echelle du risque d'inondation : modélisation hydraulique détaillée et quantification des impacts socio-économiques. La Houille Blanche, (2), 28-34.

[4] Luo, T., Maddocks, A., Iceland, C., Ward, P., & Winsemius, H.

[5] `htpps://books.openedition.org/efa/1571?lang=fr`, consulté le 30/8/2023

[6] Sene, S., Ozer, P. (2002). Evolution pluviométrique et relation inondations–événements pluvieux au Sénégal. Bulletin de la Société géographique de Liège, 42.

[7] Nibigira, L. (2019). Etude des risques naturels liés aux interactions entre les mouvements de masse et le réseau hydrographique dans la région des lacs Kivu et Tanganyika.

[8] Hypolithe NAHABAKOMEYE. Evaluation et gestion des risques catastrophiques naturelles en zone urbaine et peri-urbaine de Bujumbura(proposal).

[9] A M Wahab1 and A N Muhamad Ludin2
1 Student, Department of Urban and Town Planning, Faculty of Built Environment, Universiti Teknologi Malaysia
2 Professor, Centre for Innovative Planning & Development, Faculty of Built Environment, Universiti Teknologi Malaysia
Flood vulnerability assessment using artificial neural networks in Muar Region, Johor Malaysia

[10] J Claude IYANKUZE. Enjeux, risque d'inondation et vulnérabilité de la zone peri-urbaine de la ville de Bujumbura (Burundi) et impacts associés : exemple de la rive droite du lac Tanganyika. Mémoire de Master, UB, 2023.

[11] Lopez, Albert. Les paysans pêcheurs des basses terres du Burundi : contribution à l'étude de la complémentarité de deux activités à partir de deux exemples types Kitaga et Rutanga, '(1978)

[12] Kung, Hsu-Yang & al. Designing intelligent disaster prediction models and systems for debris-flow disasters in Taiwan. *Expert Systems with Applications*. volume (39), number (5), pages (5838–5856), 2012, Elsevier.

[13] Van Gameren, V., Weikmans, R., Zaccai, E. (2014). L'adaptation au changement climatique. Paris : La découverte.

[14] Saidi, T. B. J. (2016). Enjeux environnementaux de l'urbanisation sur le littoral de Mohammedia : quels impacts des changements climatiques sur les zones urbaines et les sites industriels à haut risque. Paysages géographiques N 2, 145.

[15] STRATEGIE NATIONALE DE PREVENTION DES RISQUES ET DE GESTION DES CATASTROPHES ET PLAN D'ACTION NATIONAL 2012-2015

[16] Burundi : ProfiL urbain de Bujumbura Programme des Nations Unies Pour les Établissements Humains P.O Box 30030 - 00100, Nairobi, Kenya Tel : +254-20-7623120 Fax : +254-20-7623426/7 (Central Office) `infohabitat@unhabitat.orgwww.unhabitat.org/publications`

[17] de Soutenance, J. Analyse du risque des inondations du bassin versant de l'oued Moudjar par le couplage du modèle hydraulique HEC-RAS et le SIG.

[18] Chouari, W. (2020). Pluviométries exceptionnelles et occupation des sols mal maîtrisée : l'exemple d'inondations des 23 et 29 septembre 2016 dans le Sahel de Sousse (Tunisie centre-orientale). La Houille Blanche, 106(1), 50-59.

[19] La montée des eaux du lac Tanganyika provoque des besoins humanitaires urgents au Burundi.OIM 2021/Triffin Ntore Une école abandonnée en raison de la montée des eaux du lac Tanganyika.19 juillet 2021 Climat et environnement (/fr/news/topic/climate-change)

[20] Marion TANGUY, CARTOGRAPHIE DU RISQUE D'INONDATION EN MILIEU URBAIN ADAPTÉE À LA GESTION DE CRISE ANALYSE PRÉLIMINAIRE, Version finale : Lundi 26 novembre 2012

[21] L'état des Villes Africaines 2010 Gouvernance, Inégalité et Marchés Fonciers Urbains Nairobi, Novembre 2010 Johanna Drach, Lisa Junghans et Katharina Rochel ont apporté leur contribution à ce rapport. Thierry Naudin a assuré la révision de la version anglaise, ainsi que la traduction française avec Martin Ebele-Tobbo.

[22] Chocat, B. (1997). Le rôle possible de l'urbanisation dans l'aggravation du risque d'inondation : l'exemple de l'Yseron à Lyon/The potential role of urbanization in increasing the risk of flooding : the example of the Yzeron in Lyon. Géocarrefour, 72(4), 273-280.

[23] Boubazine, L. (2023). Evaluationet cartographiede l'aléaglissementde terrainet risque associédans la région d'èl milia (jijel ne algérien) (Doctoral dissertation).

[24] EXPANSION DÉMOGRAPHIQUE ET DÉVELOPPEMENT AU BURUNDI : L'IMPOSSIBLE ADÉQUATION ? Gervais BARAMPANZE, François NKURUNZIZA

[25] Bani, S. S., Yonkeu, S. (2016). Risques d'inondation dans la ville de Ouagadougou : Cartographie des zones à risques et mesures de prévention. Journal Ouest-Africain des Sciences de Gestion, 1(1), 1-18.

[26] Jellali, M. (1997). Développement des ressources en eau au Maroc.

[27] Beguería, S., Caballero, Y., Serrano, E., Haver, M., Zabaleta, A., Jiménez, L. J. L. (2018). Cycle hydrologique et ressources en eau. Le changement climatique dans les Pyrénées : impacts, vulnérabilités et adaptation. Bases de connaissances pour la future stratégie d'adaptation au changement climatique dans les Pyrénées, 66-77.

[28] Lelièvre, M. A., Buffin-Bélanger, T., Morneau, F. (2008, May). L'approche hydrogéomorphologique pour la cartographie des zones à risque d'inondation dans les vallées de petites et moyennes tailles : un exemple commenté pour la vallée de la Rivière-au-Renard. In Comptes rendus de la 4 ème Conférence canadienne sur les géorisques : des causes à la gestion.

[29] ELHAMI, M., HEFNAOUI, A. (2018). L'Efficience du Marché dans les Marchés Émergents et Frontières de la Zone MENA. Finance, Finance Internationale, (10).

[30] Morice, E. (1972). Tests de normalité d'une distribution observée. Revue de statistique appliquée, 20(2), 5-35.

[31] Bélanger, M., El-Jabi, N., Caissie, D., Ashkar, F., Ribi, J. M. (2005). Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple. Revue des Sciences de l'Eau, 18(3), 403-421.

[32] Montreer, P. (2019). Identification des molécules responsables de l'odeur d'un mélange gazeux–Vers une automatisation de la démarche (Doctoral dissertation, IMT-MINES ALES-IMT-Mines Alès Ecole Mines-Télécom).

[33] Martinez, M. J. (2006). Modèles linéaires généralisés à effets aléatoires : contributions au choix de modèle et au modèle de mélange (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).

[34] Brostaux, Y. (2005). Etude du classement par forêts aléatoires d'échantillons perturbés à forte structure d'interaction.

[35] Cornec, M., Bertail, P. (2009). Validation Croisée et Modèles Statistiques Appliquées. Nanterre, Paris (France).

[36] Diaw, O. (2020). Impact des critères extra-financiers (ESG) sur l'investissement socialement responsable.

[37] Robillart, M. (2015). Etude de stratégies de gestion en temps réel pour des bâtiments énergétiquement performants (Doctoral dissertation, Ecole Nationale Supérieure des Mines de Paris).

[38] LAM, M. L. A. D ipl ô mede S tatisticien M ention A ctuariat.

[39] Caulet, R., Peguin-Feissolle, A. (2000). Un test d'hétéroscédasticité conditionnelle inspiré de la modélisation en termes de réseaux neuronaux artificiels. Annales d'Economie et de Statistique, 177-197.

[40] Crocco, G., Herzig, A. (2002). Les opérations de changement basées sur le test de Ramsey.

[41] Ricco, R. (2017). Pratique de la Régression Logistique Régression Logistique Binaire et Polytomique Université Lumière Lyon 2.

[42] SAIDI, O., MALLOUCHE, D., MANSOUR, N. B., HAJEM, S., HSAIRI, M. TESTS STATISTIQUES : PRINCIPE ET APPLICATIONS EN EPIDEMIOLOGIE.

[43] de la SAIBU, C. (2023). REVUE DES CARREFOURS SCIENTIFIQUES. Revue des Carrefours Scientifiques, 5, 50011-50025.

[44] EL BETTIOUI, R., Aziz, O. U. I. A. (2018). L'impact du financement par la dette extérieure publique sur la croissance économique marocaine : investigation quantitative. Finance, Finance Internationale, (10).

[45] BOTOMANOVATSARA, F., RATIARISON, A. A., RAKOTOVAO, N. A. (2018). Étude du débit d'eau dans le barrage de Mantasoa. Afrique SCIENCE, 14(3), 244-256.

[46] Le Gallo, J. (2004). Hétérogénéité spatiale. Economie prevision, 162(1), 151-172.

[47] Bourbonnais, R. (2015). Econométrie. Paris : Dunod.