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

Modélisation statistique de la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura

Irankunda, Egide; Sous la direction de : Pr Barankanira Emmanuel

2024-05
UB, FS
https://repository.ub.edu.bi/handle/123456789/2029

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 statistique de la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura

Par:

IRANKUNDA Egide

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

Sous la direction de:
Directeur : Pr BARANKANIRA Emmanuel
Co-directeur: Pr NYANDWI Servat

Bujumbura, Mai 2024

Composition du Jury

Dr NIYUKURI David (Président)

Dr NKURUNZIZA Menus (Secrétaire)

Pr BARANKANIRA Emmanuel (Directeur)

Pr NYANDWI Servat (Co-Directeur)

Dédicace

A mes parents;
A mes frères et sœurs;
A toutes mes connaissances;
Je dédie ce travail.

Remerciements

Au terme de tout le parcours de cette recherche, je tiens à témoigner toute ma reconnaissance et ma gratitude à:
au département de Mathématiques de la faculté des sciences qui m'a offert cette formation de qualité; à mon promoteur et directeur de ce mémoire, Pr BARANKANIRA Emmanuel. Pour toutes les modifications et les éclaircissements apportés à l'amélioration de cette recherche ainsi que l'appui méthodologique qu'il n'a cessé de m'apporter tout au long de ce travail, sa rigueur pour le travail bien fait et pour tous les conseils qu'il m'a apportés pour mon amélioration dans la recherche, pour aussi ses précieux conseils et orientations pendant la phase de collecte des données dans la Mairie de Bujumbura; à mon co-directeur de ce mémoire et en même temps responsable de master, Pr NYANDWI Servat, pour sa grande contribution dans le processus de demande d'une autorisation de mener l'enquête, pour ses précieux conseils et orientations pendant la phase de rédaction de ce mémoire; aux trois administrateurs communaux de la Mairie, pour avoir donné l'autorisation de mener l'enquête dans les communes leur confiées; au très cher Professeur NKUNZIMANA Athanase de la Faculté des Lettres et Sciences Humaines département de Géographie, pour avoir accepté de lire ce travail et apporter tant de commentaires malgré ses multiples occupations,pour toutes les connaissances transmises particulièrement sur la Revue de la littérature; à 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 vers tous ceux qui ont contribué à ma formation depuis le cycle primaire jusqu'à l'université et plus particulièrement tous les professeurs de la Faculté des Sciences, surtout ceux du département de Mathématiques et de physique. Qu'ils trouvent ici l'expression de ma profonde gratitude pour m'avoir doté d'une formation tant humaine que scientifique de haut niveau.
Je ne saurais terminer sans adresser mes vifs remerciements à Msc BUKURU

Thierry et à NIYOMUGAVYI Sadock Kévin pour leurs inébranlables soutiens et encouragements. J'exprime ma vive reconnaissance envers mes parents pour leur support tant matériel que moral et leurs encouragements depuis l'école primaire à nos jours, qu'ils trouvent ici le fruit de leurs efforts. Que tous ceux dont les noms n'apparaissent pas ici, mais qui ont contribué de près ou de loin à l'édification et à la réalisation de ce mémoire trouvent ici un mobile pour leur satisfaction.

Résumé

Cette étude s'intéresse à l'analyse de la perception du risque hydro-érosif des ravins qui traversent la ville de Bujumbura. Elle porte sur un échantillon aléatoire de 405 répondants réparti avec une probabilité égale dans les trois communes (Ntahangwa, Mukaza et Muha) de la Mairie de Bujumbura. La variable dépendante est binaire et indique si le risque hydro-érosif des ravins dans le milieu de vie des répondants est élevé. Les variables indépendantes sont des facteurs sociodémographiques à savoir la commune, le sexe, le niveau d'instruction, le statut matrimonial, la profession et le mode de vie et des facteurs associés au milieu de vie à savoir la distance au ravin, l'interruption du sommeil, le degré de la pente, les matériaux de construction, le fait d'avoir été victime de l'écroulement des ravins, le constat des phénomènes hydro-érosifs, le sol sablonneux, le sol insuffisamment protégé par le couvert végétal, le niveau de dégradation de la végétation, les événements dommageables multiples et le fait que l'homme ne peut pas maîtriser le risque. La proportion de répondants qui trouvent que le risque hydro-érosif des ravins dans leurs milieux est élevé a été calculée globalement et selon les facteurs sociodémographiques et les facteurs associés au milieu de vie. Le test du chi-deux d'indépendance de Pearson, le coefficient V de Cramer et le modèle logistique binaire ont été utilisés pour examiner les facteurs susceptibles d'influencer la perception du risque hydro-érosif des ravins. Les données ont été analysées grâce au logiciel R, version 4.3.1. Les résultats montrent qu'environ la moitié (49,6%) des habitants enquêtés estime que le risque hydro-érosif dans leur milieu est élevé. En outre, le niveau de dégradation de la végétation, les événements dommageables multiples, l'interruption du sommeil, le degré de la pente, le fait que l'homme ne peut pas maîtriser le risque et les sols sablonneux influencent significativement la perception du risque hydro-érosif après ajustement sur la profession et le sol insuffisamment protégé par le couvert végétal.

Mots clés: Perception, ravins, test du chi-deux, modèle logistique binaire, ville de Bujumbura.

Abstract

Statistical modeling of the hydrological risk perception of ravines crossing Bujumbura Municipality

This study aims at analyzing the hydrological risk perception of ravines crossing Bujumbura Municipality. A random sample of 405 respondents was selected and distributed in the three communes (Ntahangwa, Mukaza and Muha) using an equal probability. The binary dependent variable showed whether the hydrological risk of ravines in the respondents' living environment was high. The independent variables were sociodemographic factors such the municipality, sex, education level, marital status, occupation and lifestyle, and factors associated with the living environment namely distance from the ravine, interruption sleep, degree of the slope, construction materials, the fact of having been a victim of the collapse of the ravines, the observation of hydrological phenomena, sandy soil, soil insufficiently protected by the plant cover, level of degradation of vegetation, multiple damaging events and the fact that humans cannot control the risk. The proportion of respondents who find that the hydrological risk of ravines in their areas was high was computed overall and according to socio-demographic factors and factors associated with the living environment. Pearson's chi-square test of independence, Cramer's V coefficient and the binary logistic model were used to examine the factors likely to influence the perception of the hydrological risk of ravines. Data were analyzed using R software, version 4.3.1. Findings show that around half (49.6%) of the residents surveyed believe that the hydrological risk in their area is high. Furthermore, level of vegetation degradation, multiple damaging events, sleep interruption, degree of slope, the fact that humans cannot control the risk and sandy soils significantly influence the hydrological risk perception after adjustment for occupation and soil insufficiently protected by the plant cover.

Key words: Perception, ravines, chi-square test, binary logistic model, city of Bujumbura.

Table des matières

Composition du Jury i

Dédicace ii

Remerciements iii

Résumé v

Abstract vi

Table des Matières vii

Liste des Figures x

Liste des Tableaux xi

Liste des sigles et abréviations xii

Avant-propos xiii

Introduction générale 1

0.1 Contexte et problématique 2

0.2 Hypothèses de recherche 3

0.3 Objectifs de travail 4

0.4 Intérêt du sujet 4

0.5 Délimitation du travail 5

1 Revue de la littérature 6

1.1 Introduction 6

1.2 Lien entre le phénomène hydro-érosif des ravins et sa perception 7

1.2.1 Définition de l'érosion hydro-érosif, risques associés et ravin 7

1.2.2 La perception du risque hydro-érosif d'un ravin 8

1.3 Méthodes d'analyses identifiées d'autres études 9

1.4 Perception des risques et l'attitude liée aux causes du ravin . . 10

1.4.1 La perception des risques 10

1.4.2 L'attitude face aux risques 10

1.5 L'étude des inégalités économiques et environnementales 11

1.6 La relation entre décideurs et citoyens 11

1.7 Sites d'étude 13

1.7.1 Sites de la Commune Ntahangwa 13

1.7.2 Sites de la Commune Mukaza 14

1.7.3 Sites de la Commune Muha 15

1.7.4 Cadre conceptuel de la Recherche: Représentation du risque 15

2 Matériel et méthodes 18

2.1 Calcul de taille de l'échantillon 18

2.2 Source et types de données 19

2.3 Variable dépendante et variables indépendantes 19

2.4 Analyse statistique 21

2.4.1 Test du chi-deux d'indépendance 21

2.4.2 Intensité de l'association 22

2.4.3 Loi de Bernouilli 23

2.4.4 Régression logistique simple 23

2.4.5 Modèle logistique multiple 26

2.4.6 Modèle logistique optimal 28

2.4.7 Le pseudo-R^2 de Mc Fadden 29

2.4.8 Courbe ROC 29

2.4.9 Adéquation et validation du modèle 30
2.4.10 Analyse des résidus 31

3 RÉSULTATS, INTERPRÉTATION ET DISCUSSION **33**
3.1 Description de l'échantillon 33
3.2 Tests du chi-deux et coefficient de corrélation 36
3.3 Modèle logistique simple 37
3.4 Modèle logistique multiple 40
3.5 Modèle optimal ou final 43
3.6 Analyse des résidus 45

DISCUSSION DES RÉSULTATS **48**

Conclusion générale **50**

Perspectives **51**

Bibliographie **53**

Liste des figures

1.1 Site de Gikungu-rural 13
1.2 Site de Taba 13
1.3 Site de Cari 13
1.4 Site de Kigobe 13
1.5 Site de Mugoboka I 14
1.6 Site de Mugoboka II 14
1.7 Site de Mutanga-Sud 14
1.8 Site de Jabe 14
1.9 Site de Gatoke 14
1.10 Site de Buyenzi 14
1.11 Site de Kanyosha 15
1.12 Site de Ruziba 15
1.13 Site de Gisyo 15
1.14 Site de Gasekebuye 15
1.15 Composantes de la représentation du risque. 17
1.16 Conceptualisation de la représentation du risque. 17

3.1 Résidus studentisé 46
3.2 Distance de Cook 46
3.3 Leviers 46
3.4 Résidus versus leviers 46

Liste des tableaux

2.1 **Répartition des rivières et des sites selon la commune** 20
2.2 **Qualification selon la valeur du V de Cramer** 23
2.3 **Différents niveaux de discrimination** 30

3.1 **Répartition de l'échantillon selon les facteurs sociodémographiques** 34
3.2 **Répartition de l'échantillon selon les facteurs relatifs aux milieux de vie** 35
3.3 **Statistiques descriptives des variables quantitatives** 36
3.4 **Tests du chi-deux et coefficient de corrélation** 37
3.5 **Modèle logistique pour les variables sociodémographiques** . . . 38
3.6 **Modèle logistique pour les variables liées au milieu de vie** . . . 39
3.7 **Modèle logistique multiple** 42
3.8 **Modèle final** 44

Liste des sigles et abréviations

1 . AUC : Surface sous la Courbe (Area Under Curve)

2 . BIC : Le critère d'information bayésien (de l'anglais, bayesian information criterion)

3 . IC : Intervalle de Confiance

4 . N : Taille de l'échantillon

5 . N^+ : Effectif des répondants qui perçoivent que le risque hydro- érosif est élevé

6 . OR : L'odds ratio, également appelé rapport des chances, rapport des cotes ou risque relatif rapproché

7 . ROC : (de l'anglais, Receiver Operating Characteristic) Caractéristique d'exploitation du receveur

Avant-propos

L'analyse de la perception des risques des catastrophes naturels passe par la mise en place et la définition d'une technique d'enquête. Cette enquête, basée sur un questionnaire et des entretiens individuels, permet de collecter des informations relatives à l'identification des milieux à risque et au degré d'appropriation de ces milieux par les populations. Elle doit également faire ressortir la façon dont les différents acteurs approchent et gèrent ces espaces tout en montrant les variabilités de conception ou de perception des risques hydro-érosifs.

L'environnement est perçu différemment selon que les acteurs sont directement concernés ou non par les catastrophes naturelles (ici, les glissements de terrains (les ravins)) et selon leur utilisation du territoire. De plus, les acteurs de proximité que sont les agriculteurs, les habitants, les élus locaux sont prêts à modifier leurs conduites dans le but de réduire les risques des ravins. Qui plus est, la détermination des risques et des enjeux sur une zone à risque modifie les comportements des décideurs locaux face à l'aménagement des lieux habitables.

En effet, cette étude s'intéresse à l'analyse de la perception du risque hydro-érosif des ravins qui traversent la ville de Bujumbura. Elle porte sur un échantillon aléatoire de 405 répondants réparti avec une probabilité égale dans les trois communes (Ntahangwa, Mukaza et Muha) de la Mairie de Bujumbura.

L'objectif général de cette recherche est d'étudier si la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura pour les habitants est élevé afin de prévenir les risques dus aux ravins.

Introduction générale

Historiquement, l'étude des modèles décrivant les modalités prises par une ou plusieurs variables qualitatives date des années 1940-1950 [Hurlin, 2003].
Les travaux les plus importants de cette époque sont sans équivoque ceux de Berkson (1944, 1950) en Biostatistique consacrés notamment aux modèles dichotomiques simples (modèles logit et probit). Les premières applications ont alors essentiellement été menées dans le domaine de la biologie, de la sociologie et de la psychologie. Ainsi, ces modèles ont été utilisés pour décrire des données économiques avec notamment les travaux de McFadden (1973) et de James J. Heckman (1974) [Altman, 1990]. Parmi ces modèles figure la régression logistique. La régression logistique est un modèle statistique permettant d'étudier les relations entre une variable qualitative binaire y dite expliquée et une variable x dite explicative et qui est de nature qualitative ou quantitative. Contrairement à la régression linéaire multiple qui utilise une combinaison linéaire des variables explicatives pour prédire une variable expliquée avec la fonction de lien identité, la régression logistique prédit la transformation logit de la probabilité que la variable expliquée prenne la valeur 1 à l'aide d'une combinaison linéaire des variables explicatives [Chikhi et Chavance, 2012]. Autrement dit, cette transformation s'effectue sur la proportion d'individus présentant un caractère d'intérêt ou la probabilité qu'un individu présente ce caractère. Il s'agit d'un modèle linéaire généralisé selon lequel la relation liant la variable endogène (ou son espérance mathématique) aux variables explicatives fait intervenir une fonction logistique comme fonction de lien qui correspond à une transformation non linéaire de la variable endogène (en termes de probabilités). S'agissant du nombre de variables explicatives, il existe le cas particulier de régression logistique qui est la régression logistique simple (une seule variable explicative) et la régression logistique multiple (plusieurs variables explicatives). Il est possible aussi de distinguer les trois types de régression logistique que sont la régression logistique binaire (la variable dépendante a deux modalités codées 1 et 0), la régression logistique ordinale (la variable dépendante est ordinale) et

la régression logistique multinomiale ou polytomique (la variable dépendante est nominale). La notion de modélisation statistique est une méthode qui sert à traduire un phénomène en langage statistique en choisissant un modèle adéquat suivant les variables d'intérêt [Gourieroux, 1984]. L'objectif des modèles dichotomiques consiste alors à expliquer la survenue de l'événement considéré en fonction d'un certain nombre de caractéristiques observées pour les individus de l'échantillon. Un des domaines d'application consiste en l'étude des choix d'éducation [Van Erkel et Peter, 1998]. Ainsi, parmi les premiers travaux utilisant les modèles à réponses qualitatives, plusieurs s'intéressaient aux comportements des étudiants que ce soit en termes de choix de filières ou en termes de choix d'établissements [Harari-Kermadec, 2009]. Il s'agissait alors de modéliser ces comportements en fonction d'un certain nombre de caractéristiques propres aux universités (par exemple la présence de campus ou des débouchés professionnels) ou aux étudiants. Du fait de l'organisation privée des études aux États-Unis, de telles modélisations ont connu un grand intérêt, que ce soit dans une perspective purement académique ou dans une perspective appliquée par exemple l'étude de Radner et Miller [Ja, 1983]. Un autre domaine d'application consiste en la modélisation des risques de défaillance dans une relation de prêt ou dans tout autre forme de contrat d'engagement à savoir le contrat d'abonnement téléphonique et le contrat d'assistance [Lusted, 1960].

0.1 Contexte et problématique

Du point de vue de la démarche, l'analyse de cette perception des risques de catastrophes naturels passe par la mise en place et la définition d'une technique d'enquête. Cette enquête, basée sur un questionnaire et des entretiens individuels, permet de collecter des informations relatives à l'identification des milieux à risque et au degré d'appropriation de ces milieux par les populations. Elle doit également faire ressortir la façon dont les différents acteurs approchent et gèrent ces espaces tout en montrant les variabilités de conception ou de perception des risques hydro-érosifs. L'Afrique est la région du monde la plus vulnérable aux risques hydro-climatiques
[Dutozia et Voiron-Canicio, 2019]. Cette vulnérabilité est découplée par la combinaison de plusieurs facteurs tels que la pauvreté endémique, les limites d'une mauvaise gouvernance, l'accès aux capitaux, l'insuffisance des infrastructures et des technologies, la dégradation des écosystèmes, les catastrophes complexes et les conflits [Ahouangan *et al.*, 2014]. L'inefficacité des systèmes

de collecte des déchets solides, l'extension du bâti en zones inondables et les réseaux d'assainissement modestes sont également des facteurs qui aggravent l'aléa. Ces derniers renforcent l'effet prévisible des risques d'inondation sur les enjeux (populations urbaines fortement paupérisées) [Issaka, 2010]. Au Burundi, en 2022, lors d'une visite effectuée par la plateforme nationale chargée de la prévention des risques et de gestion des catastrophes, il a été constaté des gens qui font l'extraction anarchique dans des rivières qui traversent la Mairie de Bujumbura comme par exemple la rivière Kanyosha. Dans la saison pluvieuse où la plaine de l'Imbo enregistre de fortes pluies, les ravins deviennent de plus en plus un problème sérieux pour les différentes zones de la Mairie de Bujumbura. En effet, les localités comme Busoro, Nkenga-Busoro, le quartier Gisyo, Mugoboka I et II, Gikungu rural,...etc et certaines maisons construites tout près des rivières qui y traversent sont en danger suite aux glissements de terrains. En plus, les bâtiments construits tout près de certaines rivières traversant la Mairie de Bujumbura sont parfois menacés par les ravins, ce qui fait qu'ils ne peuvent plus être utilisables suite aux glissements de terrains. En outre, selon Ir Mjr Léonidas Nibigira, expert environnementaliste, beaucoup de facteurs sont à l'origine desdits glissements des terrains : la nature du sol, les travaux anthropiques comme l'extraction anarchique des moellons, la démographie galopante et autres facteurs sont les causes principales de ces catastrophes naturelles.
Deux questions guident notre réflexion :

- **De quelle manière peut-on aborder la perception des risques ?**

- **Qu'apporte la connaissance de la perception des risques dans la compréhension des comportements des populations face à une catastrophe naturelle de type glissement de terrains (ravin)**

0.2 Hypothèses de recherche

L'environnement est perçu différemment selon que les acteurs sont directement concernés ou non par les catastrophes naturelles (ici, les glissements de terrains (les ravins)) et selon leur utilisation du territoire. De plus, les acteurs de proximité que sont les agriculteurs, les habitants, les élus locaux sont prêts à modifier leurs conduites dans le but de réduire les risques des ravins. Qui plus est, la détermination des risques et des enjeux sur une zone à risque modifie les comportements des décideurs locaux face à l'aménagement des lieux habitables.

Nous formulons cinq hypothèses de recherche :
H_1 : Le degré de la pente influence la perception du risque hydro-érosif des ravins ;
H_2 : Les événements dommageables multiples agissent sur la perception du risque hydro-érosif des ravins ;
H_3 : Le fait d'avoir été victime de l'écroulement des ravins a un effet sur la perception du risque hydro-érosif des ravins ;
H_4 : La distance au ravin influence la perception du risque hydro-érosif des ravins ;
H_5 : Le sol sablonneux influence la perception du risque hydro-érosif des ravins.

0.3 Objectifs de travail

L'objectif général de cette recherche est d'étudier si la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura pour les habitants est élevé afin de prévenir les risques dus aux ravins.
Les objectifs spécifiques à ce travail consistent alors à :

- **Tester l'indépendance entre la perception du risque du hydro-érosif des ravins et les facteurs sociodémographiques d'une part et les facteurs liés au milieu de vie d'autre part ;**

- **Déterminer les facteurs sociodémographiques et les facteurs liés au milieu de vie qui influencent la perception des risques hydro-érosifs ;**

0.4 Intérêt du sujet

La recherche a été menée tant dans toute la Mairie de Bujumbura qu'à l'échelle communale. Elle ne prend donc pas en compte certaines caractéristiques locales et individuelles. Par ailleurs, elle souligne qu'en valeur absolue, les personnes les plus vulnérables sont celles qui produisent le plus de risques et en subissent les conséquences y relatives. En même temps, les individus les plus éduqués sont les plus favorablement sensibles aux politiques environnementales mises en place.
L'intérêt de cette recherche est qu'elle porte sur les différentes manières dont les habitants perçoivent le risque hydro-érosif des ravins dans leurs milieux

et les conséquences qui en découlent, en tenant compte des variables sociodémographiques et celles liées au milieu de résidence des répondants. En outre, cette étude combine la statistique descriptive et la statistique inférentielle.

0.5 Délimitation du travail

Cette étude a été subdivisée en trois chapitres : Dans le premier chapitre, il sera question d'expliquer ce qu'est le risque hydro-érosif, ce que sont les causes et les conséquences de l'érosion hydro-érosif, la manière dont elles se manifestent à cause des ravins issus des rivières qui traversent la Mairie de Bujumbura. Dans ce même chapitre, l'historique de l'apparition pour la première fois de cette manifestation du phénomène ainsi que la réponse du Gouvernement Burundais sur cette manifestation seront montrés.
Dans le deuxième chapitre, le test du chi-deux et ses conditions de validité, le coefficient V de Cramer, la notion de régression logistique, l'estimation et le test sur les paramètres du modèle ainsi le théorème de Bayes, la notion de la courbe ROC, l'aire sous la courbe ROC et la contribution des variables explicatives seront introduits point par point.
Dans le troisième chapitre, il s'agira de montrer et d'interpréter les résultats issus des analyses statistiques des données provenant d'une enquête que nous avons menée dans la Mairie de Bujumbura.
Enfin, nous terminons par une conclusion générale et des suggestions.

Chapitre 1

Revue de la littérature

1.1 Introduction

Cette partie va rassembler, analyser et résumer ce qui a été trouvé dans la littérature scientifique afin de proposer une vue globale des avancées scientifiques du rôle de la perception du risque hydro-érosif des ravins par les habitants de la capitale économique du Burundi et ses impacts probables.
En effet, un risque représente également un construit social [Klinke et Renn, 2010].
Le lien existant entre la représentation mentale subjective de chaque individu et la réalité objective se forge au travers des expériences de dangers réels et de leurs conséquences ainsi que sur la base de savoirs, concepts, valeurs, besoins et intérêts.
La perception du risque désigne toutes les représentations mentales et les façons dont les gens appréhendent, comprennent et évaluent un risque, et ce, à partir de leur propre perspective et de leurs propres expériences.
Les facteurs qui modulent les perceptions sont très variés.
Ils peuvent être
[Farley et Gravel, 2013] :

- **liés à l'individu : expérience personnelle et connaissance du risque, âge, niveau de scolarité, caractéristiques ou bénéfices personnels, etc ;**

- **socioculturels : valeurs collectives, lois et normes sociales, communication par les média, par les groupes en présence, mobilisation sociale, etc ;**

- **externes : nature et importance du risque, lieu, écosystème, impacts économiques, etc.**

Ainsi, la perception d'un même risque peut grandement varier d'un individu à un autre ou entre les groupes d'individus. De plus, les perceptions évolueront dans le temps. Or, celles-ci influencent significativement le comportement des individus [Klinke et Renn, 2010]. Il est donc important de les comprendre et de les prendre en compte pour favoriser l'implication des parties prenantes.

1.2 Lien entre le phénomène hydro-érosif des ravins et sa perception

1.2.1 Définition de l'érosion hydro-érosif, risques associés et ravin

L'érosion hydrique est un phénomène naturel qui évolue parallèlement avec l'évolution humaine et la sévérité du climat. Ce phénomène se définit comme étant le processus de détachement des particules du sol par les précipitations et le ruissellement, leur transport par les eaux superficielles qui ruissellent et leur dépôt le long du trajet suivi par les eaux superficielles sur les versants et dans les cours d'eau [Foster, 1972]. Facteur essentiel de l'évolution des paysages, l'érosion hydrique englobe toutes les formes d'usure affectant la couche superficielle de l'écorce terrestre en contact avec l'atmosphère. En réalité, l'érosion est un phénomène qui, certes, abaisse toutes les dd montagnes (d'où le terme de « dénudation rate », vitesse d'abaissement du sol des géographes anglophones) mais en même temps l'érosion engraisse les vallées, forme des riches plaines qui nourrissent une bonne partie de l'humanité. Il n'est donc pas forcément souhaitable d'arrêter toute érosion, mais de la réduire à un niveau tolérable [Roose *et al.*, 1994], d'où le terme « Soil loss tolerance ». Lorsque l'intensité des pluies dépasse la capacité d'infiltration de la surface du sol, il se forme d'abord des flaques. Ces flaques communiquent ensuite par des filets d'eau et lorsque ces filets d'eau ont atteint une certaine vitesse, 25 cm par seconde d'après Hjulström [Hjulström, 1935], ils acquièrent une énergie propre qui va créer une érosion limitée dans l'espace par des lignes d'écoulement. Cette énergie de l'eau n'est plus dispersée sur l'ensemble de la surface du sol, mais elle se concentre sur des lignes de plus forte pente. L'érosion linéaire indique donc que le ruissellement s'est organisé, qu'il a pris de la vitesse et acquis une énergie cinétique capable d'entailler le sol et d'emporter des particules de plus en plus grosses : non seulement des argiles et des limons comme l'érosion en nappe sélective, mais des graviers ou des cailloux et même des blocs rocheux dans les ravines [Roose *et al.*, 1994]. Irréversibilité du phénomène justifie qu'une attention particulière lui est ac-

cordée.
En effet, ce phénomène entraine une perte du potentiel de production du sol là où il est décapé et, au-delà, des dégâts nombreux : coulées boueuses, parfois importantes, détérioration de la voirie par glissement de terrain ou des ouvrages d'art par érosion fluviale. À l'échelle du bassin versant, l'érosion hydrique peut entrainer en amont une dégradation des sols qui engendre une perte de productivité de terrains agricoles et en aval une pollution physique qui touche les cours d'eau, les ouvrages d'art et les infrastructures.
Le mot « risque » est apparu dès le XVIème siècle dans le domaine des assurances maritimes. Dès lors, l'utilisation de cette notion bouleverse le traitement juridique des accidents, notamment les accidents du travail à partir du XIXème siècle. Le risque est alors utilisé dans tous les domaines : il devient social, économique, technologique, politique et écologique. La notion de risque possède une dimension probabiliste et elle est traduite sous la forme d'une fonction : risque=aléa×vulnérabilité Dépendant à la fois de l'aléa et de la vulnérabilité, le risque est apprécié de façon variable et souvent sous-estimé car toujours de l'ordre d'une probabilité et non d'une certitude. Selon November (2000) [MUJIA, 2016], le risque apparaît lors de la mise en danger d'un assemblage. Autrement dit, un événement suffit pour qu'émerge une réalité mal contrôlée dans un monde où tout est à priori stable et maîtrisable. Les situations à risque font entrer de nouvelles composantes inconnues dans un système équilibré pour le déstabiliser et le modifier. Il est possible de définir des zones à risques, des périodes à risques sans jamais savoir où et quand une catastrophe surviendra. Et bien souvent, le risque perdure même après une catastrophe, ce qui implique la nécessité de mettre en place des politiques de prévention.

1.2.2 La perception du risque hydro-érosif d'un ravin

La perception du risque, c'est-à-dire le modèle que s'en font les acteurs intervenant dans sa gestion, doit être suffisamment intense dans la société pour que les précautions pour le limiter puissent être acceptées et justifiées. Une fois la menace identifiée, les gestionnaires doivent arbitrer entre les différentes manières d'appréhender et de réduire le risque. La lutte contre le risque devient alors un enjeu économique, politique et social, écologique majeur et se traduit par des enjeux d'appropriation des territoires [Rode, 2009].
Mais, la perception du risque varie selon les époques et les cultures. Certains risques sont relativement bien acceptés alors qu'ils sont très importants et inversement. De plus, les média arrivent à banaliser des risques graves ou

au contraire à surestimer de moins importants, impliquant une perturbation dans la perception qu'il est possible d'en avoir et sur les politiques de gestion qui en découlent.
Spatialement, le risque se traduit par la délimitation d'aires, c'est-à-dire des zones à risques relativement vastes où les probabilités d'apparition d'une catastrophe sont élevées, ou par la détermination des réseaux que pourrait emprunter la catastrophe. Il est alors possible de cartographier les risques. L'étude d'économie de l'écologie est la jonction du comportement individuel et de l'environnement permettant d'analyser cette perception
[Amanejieu *et al.*, 2018]. En effet, une meilleure perception de risque d'un ravin permettrait aux citoyens de vivre dans des environnements plus protégés. Les études réalisées sur ce sujet identifient deux manières d'appréhender ce phénomène : soit de manière sensorielle, soit par des facteurs socioéconomiques et culturels. L'expérience perceptive directe est fondamentale afin d'identifier les niveaux de causes de l'érosion hydro-érosif et de risques de vivre autour d'un ravin. Une relation positive entre la perception de risque hydro-érosif des ravins et des facteurs sensoriels est mise en avant.

1.3 Méthodes d'analyses identifiées d'autres études

L'analyse quantitative permet d'abord d'identifier les schémas, les corrélations entre des facteurs des facteurs sociaux, du cadre de vie, et la perception différenciée de son environnement. Cependant, ces méthodes ne sont pas très adaptées pour comprendre des mécanismes spécifiques qui peuvent rendre un environnement plus ou moins favorable [Carter et Horwitz, 2014, Lalloué, 2013, Morelli, 2016a].
Ces auteurs soulignent que pour évaluer les impacts environnementaux des effets d'un ravin à l'échelle de l'individu, en un effet à l'échelle de la population, il faut d'abord caractériser le risque, et, ensuite évaluer le risque relatif. Ce dernier est étudié tant au niveau du ravin (étude de vulnérabilité du sol) qu'au niveau du statut (étude de vulnérabilité sociale).
Une analyse qualitative permet ensuite de comprendre les mécanismes sous-jacents (c'est-à-dire les étapes de démarche d'analyse)des schémas observés. Cela passe par une revue de la littérature internationale sur les usages et la perception de l'environnement, mais aussi l'étude du risque. Enfin, l'analyse prend la forme de questionnaires individuels [Bonnefoy *et al.*, 2003, Khazen, 2019, Sheppard *et al.*, 2012].

1.4 Perception des risques et l'attitude liée aux causes du ravin

1.4.1 La perception des risques

Les études relatives aux ravins attestent que les réactions et les réponses face au risque sont influencées à la fois par des caractéristiques intrinsèques telles que l'âge, le genre, le niveau d'instruction, mais aussi par des facteurs extrinsèques tels que les lieux de vies et les proximités avec les rivières causant le phénomène [Dozzi *et al.*, 2008, Khazen, 2019, Noël *et al.*, 2020].
La plupart des études soulignent que les individus perçoivent bien le risque de vivre dans un endroit où il y a un ravin. En revanche, ils sont moins bien informés quant aux risques relatifs d'exposition à ce ravin. Afin de pouvoir mesurer l'impact d'un ravin et prendre des mesures adéquates pour l'empêcher, il faut, d'une part, comprendre la distribution d'exposition de celui-ci, afin de comprendre son impact sur la vulnérabilité des individus ; mais aussi identifier le risque, ce qui permet de quantifier le dommage et donc de mesurer plus tard cette vulnérabilité [Eggerickx, 2016, Morelli, 2016b, Stephenne *et al.*, 2017].
La connaissance des causes d'un ravin est un facteur qui joue sur la sensibilité de l'homme vis-à-vis de son environnement. En d'autres termes, les personnes convaincues d'avoir une bonne connaissance des causes d'un ravin sont aussi celles qui montrent le plus de sensibilité à l'égard de celui-ci. Cependant, les études soulignent que cette sensibilité diffère selon les individus différents, les personnes aisées ont tendance à être davantage informées des problèmes que les personnes plus précarisées [Melard *et al.*, 2013, Carter et Horwitz, 2014].

1.4.2 L'attitude face aux risques

L'analyse qualitative met en évidence des résultats divergents quant à l'attitude face au risque : les catastrophes naturelles menacent différents points de la capitale économique du Burundi. Des glissements de terrains s'observent dans toutes les communes de la Mairie de Bujumbura. Dans certains endroits, les gens qui y habitent ont été contraints de vider les lieux. La pression anthropique est parmi les causes des ravins. L'observation montre que les personnes de plus de 35 ans ont une prise de conscience plus importante des problèmes environnementaux actuels que les nouvelles générations, l'éducation devrait jouer un rôle prépondérant dans la sensibilisation aux risques de ravins [Niyongabo, 2021].

Les individus sont catégorisés selon quatre typologies de réponses ou stratégies d'adaptation : les cultivateurs, les fonctionnaires, les commerçants et les autres catégories. Ces différentes stratégies permettent à l'être humain de se défendre et influent sur ses représentations, sa perception et son évaluation face aux risques des ravins. Le manque de sensibilité face aux risques de ravin laisse supposer qu'il faut incorporer certains fondements éthiques dans la politique environnementale afin que les citoyens perçoivent mieux les politiques mises en place. En effet, si les résidents savent quel est le niveau d'exposition, le type de personne affecté, quel type de personne est plus vulnérable, cela peut encourager la participation sociale des populations locales dans la lutte contre le risque hydro-érosif des ravins et améliorer les relations de confiance entre les différentes institutions et le citoyen [Bickerstaff et Walker, 2001, Melard *et al.*, 2013, Brunet, 2008].

1.5 L'étude des inégalités économiques et environnementales

En économie, le niveau d'inégalités au sein d'une société a un impact négatif sur l'ensemble des individus et est corrélé négativement au bien-être des individus [Gruber, 2005] . Étudier les différentiations d'expositions environnementales et de différentiation de vulnérabilité permet d'étudier les inégalités. Pour ce faire, il faut identifier et quantifier les inégalités d'exposition et de vulnérabilité mais aussi communiquer les dangers. Lalloué (2014) observe un gradient de la santé tout au long de l'échelle sociale et dans tous les pays, ce qui engendre des inégalités sociales. En Mairie de Bujumbura, les ménages les plus touchés sont ceux dont les 25 mètres prévus par la loi n'ont pas été respectés [Ndabashinze, 2024].

1.6 La relation entre décideurs et citoyens

La manière dont les différentes entités vont aborder le problème va dépendre de l'information disponible, fournie par les média, et les différentes communautés de recherches. Soulignons qu'en théorie néo-classique, lorsque quelqu'un jouit d'un bien sans en payer le prix réel, il risque de se confronter à des problèmes de « passagers clandestins » lorsqu'il s'agira d'établir des politiques publiques pour améliorer sa qualité ou analyser ses sources. Le rôle

des sciences sociales et économiques dans ce cadre est d'étudier les représentations et les décisions prises par ces individus. Étudier le comportement des agents économiques, la manière dont ceux-ci perçoivent les risques et réagissent permet de mettre en place des politiques ciblées [Melard *et al.*, 2013, Brunet, 2008, Khazen, 2019].

L'éco-baromètre établi par les gouvernements locaux de Madrid et d'Andalousie en Espagne est un parfait exemple de la prise en considération des perceptions des individus. Ces gouvernements estiment périodiquement les perceptions et les attitudes des citoyens, leur comportement individuel et collectif, ainsi que leur opinion sur les actions des institutions et des organismes publics responsables de la politique environnementale. L'information et la communication sont des éléments clés. D'un côté, l'information joue un rôle primordial dans la rationalité des choix des individus : dans certains pays où les gouvernements sont moins concernés, les média restent un moyen fiable pour la communication du risque à la population. D'autre part, l'attitude des individus concernés conditionne la médiatisation des nuisances [Melard *et al.*, 2013, Charvolin *et al.*, 2017, Khazen, 2019, Morelli, 2016a]. Finalement, la perception est subjective, mais isole un ressenti, dont les décideurs devront tenir compte dans la prise de décision afin de lutter au mieux contre les sources d'un ravin. En démocratie, un citoyen impliqué est un citoyen (pro)actif.

1.7 Sites d'étude

Les figures 1.1 à 1.14 montrent les images des sites d'étude. Ces sites sont répartis selon la commune. Ici, les images concernent les sites de Gikungu-rural, Taba, Cari et Kigobe en commune Ntahangwa, les sites de Mugoboka I, Mugoboka II, Mutanga-Sud, Jabe, Gatoke et Buyenzi en commune Mukaza et les sites de Kanyosha, Ruziba, Gisyo et Gasekebuye en commune Muha.

1.7.1 Sites de la Commune Ntahangwa



Figure 1.1: Site de Gikungu-rural



Figure 1.2: Site de Taba



Figure 1.3: Site de Cari



Figure 1.4: Site de Kigobe

1.7.2 Sites de la Commune Mukaza



Figure 1.5: Site de Mugoboka I



Figure 1.6: Site de Mugoboka II



Figure 1.7: Site de Mutanga-Sud



Figure 1.8: Site de Jabe



Figure 1.9: Site de Gatoke



Figure 1.10: Site de Buyenzi

1.7.3 Sites de la Commune Muha



Figure 1.11: Site de Kanyosha



Figure 1.12: Site de Ruziba



Figure 1.13: Site de Gisyo



Figure 1.14: Site de Gasekebuye

1.7.4 Cadre conceptuel de la Recherche: Représentation du risque

Dans ce cadre, pour caractériser cette représentation, nous pouvons retenir deux dimensions à savoir la dimension spatio-temporelle et la dimension socio-temporelle. En effet, la plupart des auteurs entendent la représentation du risque comme étant la perception et le comportement des acteurs face au risque [Harrison, 1988]. Nous définissons aussi la représentation sociale du risque comme étant la manière dont le risque et l'endroit à risque sont perçus par les communautés, la façon dont ces communautés vivent dans les endroits à risque à travers les modes d'occupation du sol et les stratégies d'aménagements actuels et futurs. Cette représentation peut provenir aussi bien des experts, de par leurs calculs, que des communautés, fruit de leur ressenti, de leur vécu et

de leur système de valeurs [Kermisch, 2012]. Il convient ainsi d'individualiser trois variables clés liées à la notion de représentation. Ces dernières nous permettent de mieux appréhender la représentation dans le temps. Il s'agit donc de la perception du risque hydro-érosif dans le milieu et les endroits sensibles au risque, du vécu et de la projection.

1.7.4.1 La perception du risque et de l'endroit à risque

La perception du risque désigne la compréhension et l'interprétation du risque ou de l'endroit sensible au risque par une communauté. Cette perception est fortement liée à la connaissance du risque, des enjeux (volonté de s'exposer au risque) et du contexte économique, social et culturel (familiarité au risque).

1.7.4.2 Vécu

Ce terme est utilisé ici pour caractériser l'attitude et le comportement d'une communauté donnée dans une zone précise face au risque. En règle générale, le vécu se traduit par des aménagements et la mise en place de stratégies ponctuelles de gestion du risque. Il dépend entièrement de la perception.

1.7.4.3 La projection

La projection est l'appréhension du risque et de l'endroit sensible au risque par une communauté donnée sur le long terme. C'est la dimension prospective du risque et de l'endroit sensible au risque ainsi que les modes de stratégies individuels et collectifs visant à limiter voire stopper le risque sur le long terme.

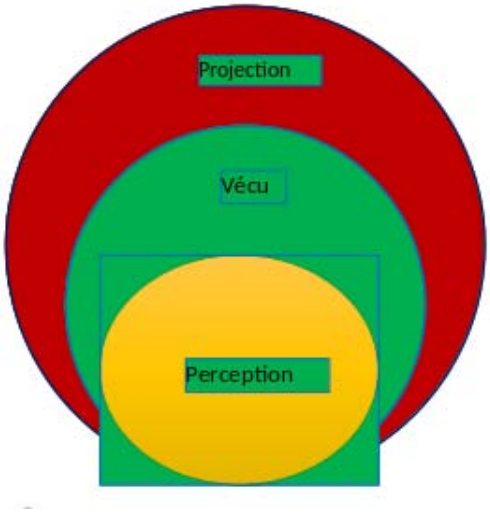


Figure 1.15: Composantes de la représentation du risque.

[**Amanejieu *et al.*, 2018**]

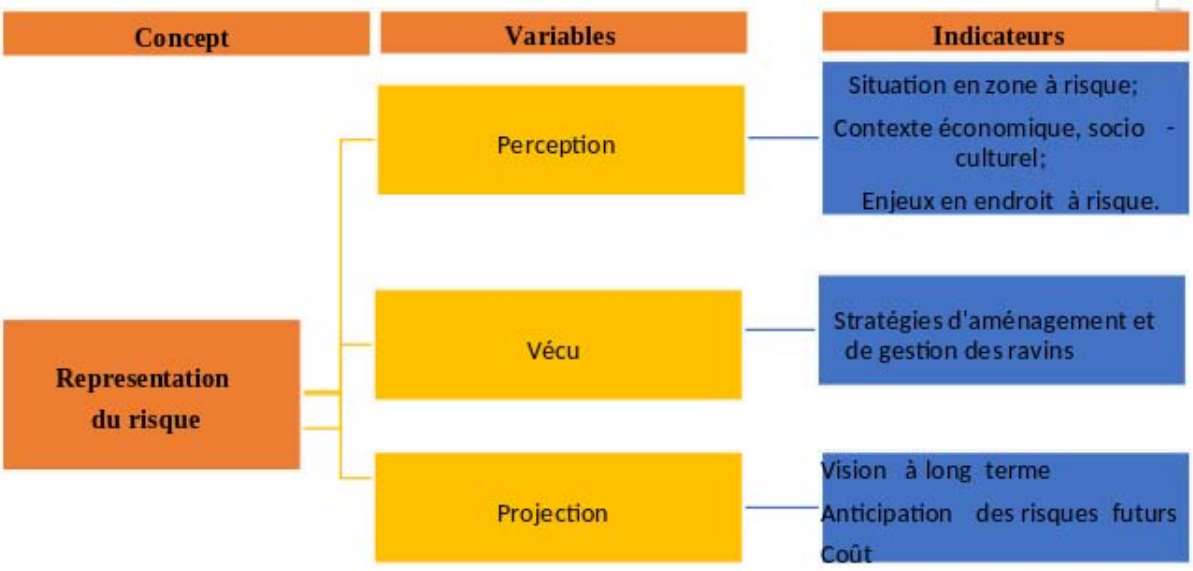


Figure 1.16: Conceptualisation de la représentation du risque.

[**Amanejieu *et al.*, 2018**]

Chapitre 2

Matériel et méthodes

Dans ce chapitre, il sera question de décrire la manière dont l'enquête-ménages sur les perceptions du risque hydro-érosif a été menée dans la Mairie de Bujumbura et les méthodes d'analyse des données. L'analyse des données est l'ensemble des méthodes qui permettent de faire une étude approfondie d'une base de données quantitatives et/ou qualitatives. L'une des méthodes qui permettent de modéliser des données de type qualitatif est la régression logistique. Dans cette étude, la variable dépendante représente la perception du risque hydro-érosif des ravins par les habitants de la ville de Bujumbura. Pour cette étude, deux méthodes statistiques à savoir le test du chi-deux d'indépendance et la régression logistique binaire sont utilisées.

2.1 Calcul de taille de l'échantillon

La proportion p de sujets qui estiment qu'il y a un risque hydro-érosif (glissement ou écroulement des terrains) élevé des ravins dans leur milieu de résidence et la taille de la population ne sont pas connues. Cette proportion a été supposée égale à 0,50. Ainsi, le calcul de la taille de l'échantillon de base a été faite selon la formule [Wolf et Cochran, 1997]

$$n = \frac{z^2}{4e^2} \tag{2.1}$$

où n est la taille de l'échantillon, z le quantile de la loi normale centrée-réduite avec une probabilité de 0,975 et e la marge d'erreur (0,05), ce qui conduit à n=385. De plus, 0,05 d'impondérables ont été ajoutés à cette taille de l'échantillon de base, ce qui aboutit à n=404. Comme ce nombre n'est pas divisible par le nombre de communes (il y en a trois), cette taille a été

portée à n=405. Cette taille d'échantillon a été répartie de manière équitable dans les strates, les strates étant les 3 communes de la Mairie de Bujumbura à savoir Ntahangwa, Mukaza et Muha.

2.2 Source et types de données

La première étape a été d'abord de faire une première descente sur terrain afin d'avoir un entretien avec les chefs des trois communes de la Mairie de Bujumbura qui ont donné l'autorisation de mener l'enquête. Toutes les rivières et tous les sites sur lesquels l'enquête a été réalisée ont été ensuite visités. À cette fin, un questionnaire a été élaboré au regard des facteurs trouvés dans la littérature. Ce questionnaire a enfin permis, pendant la collecte des données, de recueillir des informations nécessaires à la réalisation de cette étude.
Les données utilisées dans cette étude proviennent donc d'une enquête-ménages que nous avons menée aux mois de fin mars et début avril 2023 en Mairie de Bujumbura sur les perceptions du risque hydro-érosif auprès des personnes qui habitent aux environs des ravins qui traversent la ville de Bujumbura. Cette enquête, qui était basée sur un questionnaire administré à l'aide de l'application KoBoCollect auprès de la population vivant aux alentours des ravins, a été menée du 25 mars au 9 avril 2023. Elle a été réalisée sur un échantillon de 405 habitants. De plus, elle était basée sur un échantillonnage stratifié et cette stratification a été faite selon la commune (Ntahangwa, Mukaza, Muha). Une autorisation de faire l'enquête a été obtenue des Administrateurs des communes. Il s'agit d'une enquête transversale, à visée descriptive et analytique. Des informations sur les caractéristiques sociodémographiques de chaque enquêté telles que l'âge, la religion, le sexe, le niveau d'instruction, le statut matrimonial et la profession et les caractéristiques relatives au milieu de vie ont été collectées. Les données ont été téléchargées sous SPSS pour faciliter le nettoyage des données puis importées dans R pour une analyse statistique proprement dite.

2.3 Variable dépendante et variables indépendantes

Les variables utilisées dans ce mémoire sont les suivantes :

La variable dépendante est « Trouvez-vous que le risque hydro-érosif dans votre milieu est élevé ? ». Elle est dichotomique et comprend donc deux modalités : 1 si l'enquêté trouve que le risque hydro-érosif dans son milieu est

élevé et 0 sinon.

Les variables explicatives retenues dans cette étude sont la commune (1=Ntahangwa, 2= Mukaza, 3=Muha), le sexe (1=Homme, 2=Femme), le niveau d'instruction (1=Aucun, 2=Primaire, 3=Secondaire, 4=Universitaire), le statut matrimonial (1=Célibataire, 2=Marié, 3=Divorcé/veuf), la profession (1=Cultivateur, 2=Fonctionnaire, 3=Commerçant, 4=Autre), le mode de vie (1=Locataire, 2=Propriétaire, 3=Autre), la distance au ravin (m) (1=20+, 2=<6, 3=6-9, 4=10-19), l'interruption du sommeil (0=Non, 1=Oui), le degré de la pente (1=Faible, 2=Moyen, 3=Elevé), les matériaux de construction (1=Matériaux durables, 2=Matériaux semi-durables, 3=Briques adobes), le fait d'avoir été victime de l'écroulement des ravins (0=Non, 1=Oui), le constat des phénomènes hydro-érosifs (0=Non, 1=Oui), le sol sablonneux (0=Non, 1=Oui), Sol insuffisamment protégé par le couvert végétal (1=Oui, 2=Non), le niveau de dégradation de la végétation (1=Faible, 2=Moyen, 3=Elevé), le fait d'avoir eu des événements dommageables multiples (0=Non, 1=Oui), le fait que l'homme peut maîtriser le risque (1=Oui, 2=Non). La variable quantitative qui représente la distance entre le ravin et le ménage a été découpée en classes du fait que la relation entre cette variable et la variable dépendante n'est pas linéaire.

Le tableau 2.1 montre la répartition des rivières et des sites selon la commune de la mairie de Bujumbura.

Tableau 2.1: Répartition des rivières et des sites selon la commune

Commune	Rivière	Site
Ntahangwa	Cari	Winterekwa
	Nyabagere	Taba
	Nyanzari	Gikungu-Rural
	Ntahangwa	Kigobe
Mukaza	Ntahangwa	Mugoboka I, Mugoboka II, Mutanga-Sud, Jabe, Buyenzi
	Muha	Gatoke
Muha	Muha	Gasekeb uye
	Kanyosha	Musaga-Kinanira, Kanyosha
	Mugere	Ruziba
	Kizingwe	Gisyo

2.4 Analyse statistique

La statistique descriptive est une branche des statistiques qui regroupe les nombreuses techniques utilisées pour décrire un ensemble relativement important de données. Son but est de synthétiser l'information contenue dans les données à l'aide des tableaux et des graphiques. Toute analyse statistique commence par la description des données. Cette branche de la statistique a permis de calculer la proportion de personnes qui estiment qu'il y a un risque hydro-érosif élevé des ravins dans leur milieu de résidence globalement et selon les modalités des variables explicatives. Elle a été complétée par la construction des intervalles de confiance à 95 % de ces proportions, le calcul de la statistique du test du chi-deux d'indépendance de Pearson, le coefficient de corrélation V de Cramer et les rapports des cotes (modèle logistique).

2.4.1 Test du chi-deux d'indépendance

Le test du chi-deux d'indépendance permet de déterminer si deux variables catégorielles sont susceptibles d'être liées ou pas. Autrement dit, il sert à apprécier l'existence ou non d'une relation entre deux caractères qualitatifs au sein d'une population.

Les hypothèses de test sont:
H_0**: X et Y sont indépendantes**
H_1**: X et Y sont dépendantes**

La statistique de test est la suivante [Gourieroux, 1984]

$$\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{J} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \tag{2.2}$$

à (I-1)(J-1) degrés de liberté avec les effectifs observés O_{ij}(ou valeurs empiriques), e_{ij} les effectifs attendus (ou valeurs théoriques), I le nombre de modalités de la première variable qualitative, J le nombre de modalités de la deuxième variable qualitative. Les effectifs attendus sont obtenus en divisant le produit de l'effectif marginal de ligne et l'effectif marginal de colonne par le nombre total d'individus. Le seuil de significativité de $0,05$ a été retenu. Une p-value inférieure à ce seuil permet de rejeter l'hypothèse nulle et sera interprétée comme une absence de relation (ou une indépendance) entre ces deux variables. Lorsque l'un des effectifs théoriques est inférieur à 5 ou lorsque les sommes marginales du jeu de données réel sont très déséquilibrées, il est

préférable de faire des corrections de continuité ou de se fier au test exact de Fisher [Harari-Kermadec, 2009]. Le test exact de Fisher est un test qui a les mêmes hypothèses que le test du chi-deux d'indépendance pour une table de contingence ayant deux lignes et deux colonnes. Il calcule la probabilité d'obtenir des données observées en utilisant une distribution hypergéométrique. Ces probabilités sont utilisées pour calculer la p-value.
Les hypothèses de test sont :
$H_0 : OR = 1$
$H_1 : OR \neq 1$
La statistique du test exact de Fisher est le rapport des cotes et présenté selon la formule suivante : $OR = \frac{ad}{bc}$
où a, b, c, d (en complétant par colonne) sont des valeurs de la table de contingence. Soient L_1 **et** L_2 **les sommes des effectifs de chacune des 2 lignes d'une table de contingence et,** C_1 **et** C_2 **les sommes des effectifs de chacune des 2 colonnes. Des tables de contingence sont construites de façon à ne pas changer les effectifs marginaux. Sous l'hypothèse nulle d'indépendance, la distribution de probabilité d'une cellule est une loi hypergéométrique.**

2.4.2 Intensité de l'association

Le test d'indépendance du chi-deux de Pearson permettra de déterminer la liaison entre deux variables qualitatives mais sans préciser ni le degré de liaison, ni le sens de la relation. Il sera important de faire recours aux coefficients de corrélation entre autres le coefficient V de Cramer. Le coefficient V de Cramer est la racine carrée de la statistique du χ^2 **divisé par le** χ^2 **maximal. Il mesure l'intensité d'une association entre deux variables qualitatives.**
Sa forme mathématique est [5] :

$$V = \sqrt{\frac{\chi^2}{\chi^2_{max}}} = \sqrt{\frac{\chi^2}{n \times min(I-1, J-1)}} \tag{2.3}$$

où I et J sont respectivement le nombre de lignes et de colonnes de la table de contingence. En principe, il prend une valeur comprise entre 0 et 1. Plus la valeur de V est proche de 0, plus il y a indépendance entre les deux variables. Lorsqu'il vaut 1, alors il y a la dépendance complète. Le V de Cramer s'interprète comme le montre le tableau 2.2 [Van Erkel et Peter, 1998].

Tableau 2.2: Qualification selon la valeur du V de Cramer

Valeur du V de Cramer	Intensité de la relation entre les variables
Inférieure à 0.10	Relation nulle ou très faible
$[0.10; 0.20[$	Relation faible
$[0.20; 0.30[$	Relation moyenne
Supérieure ou égale à 0.30	Relation forte

2.4.3 Loi de Bernouilli

Par un modèle de régression logistique binaire, la variable expliquée ne peut prendre que deux modalités (variable dichotomique $y \in \{0, 1\}$). Par définition, cette variable suit une loi de Bernoulli de paramètre p:

$$y \sim Bern(p) \tag{2.4}$$

La probabilité d'une loi de Bernoulli de paramètre p peut s'écrire :

$$P(y = k) = p^k(1-p)^{1-k} = \begin{cases} p^1(1-p)^0 = p \ si \ k = 1 \\ p^0(1-p)^1 = 1-p \ si \ k = 0 \end{cases} \tag{2.5}$$

Il est aisé de démontrer que les moments de cette variable y sont :

$$\begin{cases} E(y) = p \\ Var(y) = p(1-p) \end{cases} \tag{2.6}$$

2.4.4 Régression logistique simple

La régression logistique binaire simple est une technique permettant de modéliser une variable dépendante qualitative binaire à l'aide d'une seule variable explicative (quantitative ou qualitative). Elle permet de déterminer ou de donner une idée de la contribution de chaque variable explicative à l'explication de la variable dépendante grâce à la statistique du test de Wald.
Soient Y une variable binaire (1=oui/ 0=non par exemple) qui représente le fait de trouver que le risque hydro-érosif dans le milieu de résidence est élevé et X une variable indépendante (le sexe par exemple) concourant à l'explication de Y. La variable Y prend la valeur 1 avec une probabilité $p = P(Y = 1/X = x)$ si la personne trouve que le risque hydro-érosif dans son milieu de résidence est élevé et 0 avec la probabilité $1 - p = P(Y = 0/X = x)$ sinon. La régression logistique a plusieurs applications et chaque facteur potentiel fera l'objet d'une hypothèse à confirmer ou infirmer. Il s'agira d'appliquer la théorie du

modèle logistique à l'étude de la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura.

Le modèle logistique simple s'écrit comme suit :

$$logit(p) = log\frac{p}{1-p} = \beta_0 + \beta_1 x + \varepsilon \tag{2.7}$$

où β_0 est l'intercept, β_1 le coefficient correspondant à la variable explicative X et ϵ un terme d'erreur.

Un calcul simple donne la fonction logistique :

$$p(x) = F(X) = P(Y = 1|X = x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} = \frac{1}{1 + e^{-\beta_0-\beta_1 x}} \tag{2.8}$$

La vraisemblance des données de l'échantillon ou densité jointe ou encore la probabilité d'observer les données $((x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ..., (x_n, y_n))$ pour la régression logistique binaire simple s'écrit :

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} P(Y = y_i|X = x_i) = \prod_{i=1}^{n} [F(X)]^{y_i} [1 - F(X)]^{1-y_i} \tag{2.9}$$

soit

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \left[\frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}}\right]^{y_i} \left[\frac{1}{1 + e^{\beta_0+\beta_1 x_i}}\right]^{1-y_i} = \prod_{i=1}^{n} \frac{(e^{\beta_0+\beta_1 x_i})^{y_i}}{1 + e^{\beta_0+\beta_1 x_i}} \tag{2.10}$$

La log-vraisemblance vaut :

$$l(\beta_0, \beta_1) = \sum_{i=1}^{n} ln\frac{(e^{\beta_0+\beta_1 x_i})^{y_i}}{1 + e^{\beta_0+\beta_1 x_i}} = \sum_{i=1}^{n} \left[y_i(\beta_0 + \beta_1 x_i) - ln(1 + e^{\beta_0+\beta_1 x_i})\right] \tag{2.11}$$

En utilisant la méthode du maximum de vraisemblance pour trouver $\hat{\beta}_0$ et $\hat{\beta}_1$ l'idée principale est de maximiser la log-vraisemblance par rapport à $(\hat{\beta}_0, \hat{\beta}_1)$. La dérivée de la log-vraisemblance par rapport aux $\hat{\beta}_j$ donne respectivement les fonctions S_1 et S_2 des paramètres :

$$S_1(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}}$$

$$S_2(\beta_0, \beta_1) = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \frac{x_i e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}}$$

Il suffit d'égaler chaque dérivée partielle à zéro et de construire le système

d'équations, lequel système n'admet pas de solution analytique.
Il se résout par des procédures de calcul numérique : la méthode itérative.
La matrice de variances-covariances des paramètres s'écrit :

$$Var(\hat{\beta}) = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \tag{2.12}$$

et est estimée par la matrice

$$\left[\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^t}\right]^{-1}_{\beta=\hat{\beta}} \tag{2.13}$$

L'odds ratio (OR), également appelé rapport des chances, rapport des cotes ou risque relatif rapproché, est une mesure statistique exprimant le degré de dépendance entre des variables aléatoires qualitatives [Tamhane *et al.*, 2016]. L'odds ratio est toujours supérieur ou égal à zéro. Ainsi soit p une probabilité.

Son odds (ou cote) est défini par :

$$Odds(p) = \frac{p}{1-p} \tag{2.14}$$

La cote d'un événement représente donc le rapport de la probabilité de l'événement à celle de son complémentaire. Ce risque peut se calculer pour chaque variable indépendante x_i. Le rapport de cotes, quant à lui, elle est calculée de la manière suivante :

L'odds ratio (OR), également appelé rapport des chances, rapport des cotes ou risque relatif rapproché, est une mesure statistique exprimant le degré de dépendance entre des variables aléatoires qualitatives. L'odds ratio est toujours supérieur ou égal à zéro. Ainsi soit p une probabilité.
Si X=1, $log\frac{p_1}{1-p_1} = \beta_0 + \beta_1$.
De même, si X=0, $log\frac{p_0}{1-p_0} = \beta_0$.
Alors $log\frac{p_1}{1-p_1} - log\frac{p_0}{1-p_0}$=$\beta_1$.
Suivant la formule de la régression, l'exponentielle d'un paramètre correspond donc à l'odds ratio. D'où

$$OR = e^{\beta_1} \tag{2.15}$$

En effet, le rapport de cotes est calculé de la manière suivante :

$$Odds(p) = \frac{\frac{P(y_i=1/X=1)}{1-P(y_i=1/X=1)}}{\frac{P(y_i=0/X=0)}{1-P(y_i=0/X=0)}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \tag{2.16}$$

L'exponentielle indique le changement de proportion (Odds Ratio) lorsque le

prédicteur X augmente d'une unité.
L'intervalle de confiance montre la fourchette dans laquelle appartient le paramètre à estimer.
L'intervalle de confiance de l'OR de niveau $1-\alpha$ est :

$$IC_{1-\alpha}(OR) = \exp\left[\beta_1 \pm u_\alpha \sqrt{Var(\beta_1)}\right] \tag{2.17}$$

où u_α est le quantile de la loi normale centrée-réduite avec une probabilité de 0,975. Si $1 \in IC$, alors il n'y a pas de relation et si $1 \notin IC$, il y a une relation.
L'interprétation de l'OR est la suivante :
Si l'OR est égal à 1, l'événement dépendant (y) et la caractéristique (x_i) sont indépendants. Si l'OR est supérieur à 1, le lien entre y et x est positif. Si l'OR est inférieur à 1, le lien entre la variable dépendante et indépendante est négatif [Chang *et al.*, 2016].
En régression logistique simple, l'interprétation des OR s'avère la plus importante, d'où leur test de significativité.
Les hypothèses sont les suivantes :

$$\begin{aligned} H_0 &: OR = 1 \\ H_1 &: OR \neq 1 \end{aligned} \tag{2.18}$$

H_0 est rejetée si $P-valeur < \alpha$. Dans cette analyse, les variables qui sont sorties significatives au seuil de 10% ont été introduites dans le modèle logistique multiple. Le choix de ce seuil a été motivé par le fait d'éviter de perdre des variables explicatives ayant une pertinence dans l'explication du fait de trouver que le risque hydro-érosif dans le milieu de résidence est élevé.

2.4.5 Modèle logistique multiple

Le modèle logistique multiple permet de modéliser une variable dépendante qualitative binaire à l'aide de plus de deux variables explicatives (quantitatives ou qualitatives). Le modèle logistique multiple s'écrit comme suit [Ja, 1983] :

$$logit(p) = log\frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon \tag{2.19}$$

où β_0 est l'intercept; $\beta_1 ... \beta_k$ sont les paramètres du modèle, $X_1 ... X_k$ les variables explicatives et ε le terme d'erreur.

Un calcul simple donne :

$$F(x) = p(x) = \frac{e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}} = \frac{1}{1+e^{-\beta_0-\beta_1 x_1-...-\beta_k x_k}} \tag{2.20}$$

La vraisemblance des données ou la probabilité d'observer les données s'écrit :

$$L(\beta_0,\beta_1,...,\beta_k) = \prod_{i=1}^{n} P(Y=y_i|X=x_i) = \prod_{i=1}^{n} [F(X)]^{y_i} [1-F(x)]^{1-y_i} \tag{2.21}$$

soit :

$$\begin{aligned} L(\beta_0,\beta_1,...,\beta_k) &= \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}}\right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}}\right)^{1-y_i} \\ &= \prod_{i=1}^{n} \frac{(e^{\beta_0+\beta_1 x_1+...+\beta_k x_k})^{y_i}}{1+e^{\beta_0+\beta_1 x_1+...+\beta_k x_k}} \end{aligned} \tag{2.22}$$

ou

$$L(\beta_0,\beta_1,...,\beta_k) = \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}}\right)^{1-y_i} = \prod_{i=1}^{n} \frac{(e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i})^{y_i}}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}} \tag{2.23}$$

Alors, la log-vraisemblance s'écrira :

$$l(\beta_0,\beta_1,...,\beta_k) = \sum_{i=1}^{n} ln \frac{\left(e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}\right)^{y_i}}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}} = \sum_{i=1}^{n} \left[y_i \left(\beta_0 + \sum_{i=1}^{k}\beta_i x_i\right) - ln\left(1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}\right)\right] \tag{2.24}$$

Par la méthode du maximum de vraisemblance, les estimateurs $\hat{\beta}_0$,$\hat{\beta}_1$,...,$\hat{\beta}_k$ s'obtiennent en maximisant la log-vraisemblance par rapport à $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$. Faisons donc les dérivées partielles par rapport à chaque paramètre $\hat{\beta}_{j,0\leq j\leq k}$ qui s'écrivent de la manière suivante :

$$\begin{aligned} S_0(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k) &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}} \\ S_j(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k) &= \sum_{i=1}^{n}\sum_{j=1}^{k} x_j y_i - \sum_{i=1}^{n}\sum_{j=1}^{k} \frac{x_j e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{k}\beta_i x_i}} \end{aligned} \tag{2.25}$$

La matrice de variances-covariances des k paramètres est donnée par :

$$\hat{v}(\hat{\beta}) = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & \cdots & Cov(\hat{\beta}_0, \hat{\beta}_j) & \cdots & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_j) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ Cov(\hat{\beta}_0, \hat{\beta}_j) & Cov(\hat{\beta}_1, \hat{\beta}_j) & \cdots & Var(\hat{\beta}_j) & \cdots & Cov(\hat{\beta}_j, \hat{\beta}_k) \\ \vdots & \vdots & & \vdots & & \ddots \\ Cov(\hat{\beta}_0, \hat{\beta}_k) & Cov(\hat{\beta}_1, \hat{\beta}_k) & \cdots & Cov(\hat{\beta}_j, \hat{\beta}_k) & \cdots & Var(\hat{\beta}_k) \end{bmatrix} \tag{2.26}$$

L'odds ratio s'obtient en appliquant la fonction exponentielle à chaque paramètre :

$$OR_j = e^{\beta_j} \tag{2.27}$$

L'intervalle de confiance de l'OR de niveau $1-\alpha$ est :

$$IC_{1-\alpha}(OR_j) = \exp\left[\beta_j \pm u_\alpha \sqrt{Var(\beta_j)}\right] \tag{2.28}$$

2.4.6 Modèle logistique optimal

Comme le but est de développer un modèle explicatif, certaines procédures ont été mises en œuvre pour sélectionner les variables à garder dans le modèle optimal. Ces procédures ont pour but de vérifier le principe de l'explication d'un phénomène par un minimum de causes élémentaires qui consiste à n'utiliser que le minimum de causes élémentaires pour expliquer un phénomène. En d'autres termes, il sera question de sélectionner le modèle apportant le maximum d'information sur la variable dépendante à partir du plus petit nombre de variables explicatives possible (X_i).

Pour le choix des variables qui peuvent être retenues dans le modèle optimal (final), la variable à expliquer a été associée aux variables explicatives qui ont été retenues dans le modèle logistique complet par élimination progressive grâce à la méthode manuelle pas-à-pas descendante, et puis grâce au test de Wald utilisé, la variable dont le degré de significativité est très élevé par rapport aux autres variables explicatives doit être retirée du modèle et le reste des variables est associé à la variable à expliquer.

Avec la méthode de sélection manuelle pas-à-pas descendante, le modèle optimal est le modèle qui minimise le critère de l'information bayésienne (Bayesian Information Criterion : BIC). Les variables liées de façon suffisamment forte à la variable à expliquer sont alors conservées dans le modèle.

D'une manière générale, ce sont toutes les variables dont le degré de significativité est inférieur à 5 %.

2.4.7 Le pseudo-R^2 de Mc Fadden

Le pseudo-R^2 de Mc Fadden est un indicateur permettant d'évaluer la performance d'un modèle de régression. Il est basé sur les noyaux du logarithme de la vraisemblance du modèle trivial et du modèle optimal.
Sa formule mathématique est [Lusted, 1960] :

$$\mathbf{R^2_{MF} = 1 - \frac{L_M}{L_0}} \tag{2.29}$$

où L_M est la log-vraisemblance du modèle optimal et L_0 la log-vraisemblance du modèle trivial (sans paramètres autre que la constante). Lorsque le modèle n'explique rien, l'indicateur vaut 0 tandis que lorsque le modèle est parfait, l'indicateur vaut 1.

2.4.8 Courbe ROC

La courbe ROC (Receiver Operating Characteristic) a été initialement développée dans les années 1950 à des fins militaires, c'est-à-dire pour exploiter des données Radar. Son intérêt dans le domaine médical a été souligné dès 1960 par Lee Lusted [Klinke et Renn, 2010]. Depuis, cet outil statistique a été utilisé notamment dans le domaine pharmaceutique [Rode, 2009], en radiologie [Amanejieu *et al.*, 2018] et en biologie [MUJIA, 2016]. La courbe de ROC est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test, calculées pour toutes les valeurs seuils possibles [Dutozia et Voiron-Canicio, 2019]. La courbe ROC est un outil statistique qui présente surtout les caractéristiques pour l'évaluation et la comparaison des performances des modèles.
Lors de l'évaluation du modèle à travers la courbe ROC, il est possible de caractériser numériquement la courbe ROC en calculant la surface située sous la courbe. C'est le critère AUC (Area Under Curve). Elle exprime la probabilité de placer un individu positif devant un négatif. Ainsi, dans le cas d'une discrimination parfaite, nous avons $AUC = 1$. Lorsque la courbe ROC se confond avec la première bissectrice, nous avons AUC = 0,5. C'est la situation de référence. Généralement, différents paliers sont proposés pour donner un ordre d'idées sur la qualité de la discrimination.
Le tableau 2.3 montre les différents niveaux de discrimination.

Tableau 2.3: Différents niveaux de discrimination

Valeur de l'AUC	Commentaires
AUC = 0.5	Pas de discimination
$0.5 < \text{AUC} \leq 0.8$	Discrimination acceptable
$0.8 < \text{AUC} \leq 0.9$	Discrimination excellente
AUC > 0.9	Discrimination exceptionnelle

2.4.9 Adéquation et validation du modèle

Dans un modèle logistique, le test de Wald est utilisé pour voir si le modèle est bon ou pas.
Deux hypothèses sont posées [Ja, 1983] :
H_0**: Le modèle est bon**
H_1 **: Le modèle n'est pas bon**

H_0 **est rejetée si la P-valeur est inférieure au seuil de 5%.**

La statistique de test est alors :

$$W_i = \frac{\hat{\beta}_i^2}{var(\hat{\beta}_i)} \tag{2.30}$$

Avec $\hat{\beta}_i^2$ **le carré du paramètre estimé et** $var(\hat{\beta}_i)$ **sa variance.**
La P-valeur est calculée en utilisant la loi du chi-deux à un degré de liberté.

Dans le modèle logistique, le test de Hosmer-Lemeshow a été utilisé pour valider le modèle afin de faire des prédictions, mais il est peu puissant.
Deux hypothèses ont été posées [Ahouangan *et al.*, 2014] :
H_0**: Le modèle est bon**
H_1**: Le modèle n'est pas bon**

H_0 **est rejetée si la P-valeur est inférieure à** α**.**
La statistique du test est :

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - m_k\mu_k)^2}{m_k\mu_k(1-\mu_k)} \tag{2.31}$$

où m_k **sont les effectifs du groupe** K**;** O_k **le nombre de succès observé dans le groupe** K**;** μ_k **la moyenne des** $\hat{p}_\beta(\chi_i)$ **qui est la probabilité et le modèle.**

Pour cette étude, 10 groupes ont été considérés, d'où $K = 10$**.**
Le link test a été utilisé à la place du test de Hosmer-Lemeshow.

2.4.10 Analyse des résidus

Le résidu de Pearson standardisé indique la bonne ou mauvaise modélisation d'un point et est donné, pour l'observation ω, par [Ja, 1983] :

$$r_s(\omega) = \frac{r(\omega)}{\sqrt{1 - h(\omega)}} \tag{2.32}$$

où $r_s(\omega)$ est le résidu de Pearson et $h(\omega)$ l'influence de l'observation ω.

Considérons une observation ω de l'échantillon. Elle a participé à la construction du modèle. Par la suite, la question qui se pose est de savoir si elle est bien modélisée ou pas. Pour peu que ω pèse énormément dans la régression, elle peut être elle-même très bien modélisée, tout en faussant la prédiction des autres.
La solution est de corriger le résidu selon l'influence du point.
La distance de Cook permet, quant à elle, de quantifier l'écart entre les vecteurs de paramètres estimés en présence et en l'absence du point ω (ω étant un individu quelconque de l'échantillon).
Il est possible de le voir sous l'angle d'un test d'hypothèses [Ja, 1983] :
H_0 : Les coefficients sont identiques
H_1 : Un des coefficients au moins est différent La distance de Cook peut être écrite à partir du résidu déviance standardisé ou du résidu de Pearson standardisé.
En prenant la seconde définition, il vient [Ja, 1983] :

$$C(\omega) = \frac{r_s^2(\omega)}{J+1} \times \frac{h(\omega)}{1 - h(\omega)} \tag{2.33}$$

La distance de Cook est une mesure pour l'influence d'une observation sur l'estimation d'une régression. Le seuil critique pour la distance de Cook à partir duquel l'observation est considérée trop influente est $4/(n-p)$ où n est le nombre d'observations et p le nombre de paramètres.

L'analyse des résidus permet de diagnostiquer la qualité de la régression. Plusieurs questions se posent à l'issue du processus de modélisation. Des réponses à apporter concernent la détermination des points qui « clochent » dans les données (données atypiques), qui s'écartent fortement des autres dans l'espace de représentation, celle des points qui sont mal modélisés (mal expliqués) par la régression logistique (résidus), celle des points qui pèsent fortement dans la régression (points leviers) et celle des points qui pèsent exagérément sur les résultats. Si ces points sont retirés de l'échantillon, le

modèle obtenu serait très différent (points influents).

Le traitement informatique (fusion des modalités, découpage des variables) et l'analyse proprement dite des données (calcul des proportions de sujets qui trouvent que le risque hydro-érosif dans leur milieu est élevé globalement et selon les caractéristiques sociodémographiques et des caractéristiques du milieu de résidence, test du chi-deux, régression logistique) ont été réalisés grâce au logiciel R version 4.3.1.

Chapitre 3

RÉSULTATS, INTERPRÉTATION ET DISCUSSION

3.1 Description de l'échantillon

Une description de l'échantillon à travers les fréquences absolues (effectifs) et les fréquences relatives des modalités des variables catégorielles d'une part ainsi que certaines statistiques descriptives (effectif, minimum, moyenne, écart-type, maximum) des variables quantitatives d'autre part a été réalisée.
Le tableau 3.1 montre les fréquences des modalités des variables sociodémographiques.

Tableau 3.1: Répartition de l'échantillon selon les facteurs sociodémographiques

Variables	Modalités	N	N^+	%	IC à 95%
COMMUNE	NTAHANGWA	135	66	48,9	[40,4 ; 57,4]
	MUKAZA	135	68	50,4	[41,9 ; 58,9]
	MUHA	135	67	49,6	[41,1 ; 58,1]
SEXE	Homme	180	80	44,4	[37,1 ; 51,7]
	Femme	225	121	53,8	[47,2 ; 60,3]
Niveau d'instruction	Aucun	59	36	61	[48,4 ; 73,6]
	Primaire	161	86	53,4	[45,7 ; 61,2]
	Secondaire	112	48	42,9	[33,6 ; 52,1]
	Universitaire	73	31	42,5	[31 ; 53,9]
Statut matrimonial	Célibataire	127	53	41,7	[33,1 ; 50,4]
	Marié	257	135	52,5	[46,4 ; 58,7]
	Divorcé et Veuf	21	13	61,9	[40,6 ; 83,3]
Profession	Cultivateur	122	77	63,1	[54,5 ; 71,7]
	Fonctionnaire	49	17	34,7	[21,2 ; 48,2]
	Commerçant	73	28	38,4	[27,1 ; 49,6]
	Autre	161	79	49,1	[41,3 ; 56,8]
Mode de vie	Propriétaire	201	124	61,7	[54,9 ; 68,4]
	Locataire	134	51	38,1	[29,8 ; 46,3]
	Autre	69	26	37,7	[25,7 ; 48,6]
Ensemble		405	201	49,6	[44,7 ; 54,5]

Sur un échantillon de 405 personnes enquêtées, 204 trouvent que le risque hydro-érosif des ravins dans leur milieu est élevé, soit 49,6 %. La proportion la plus élevée (63,1 %) de personnes qui estiment que le risque hydro-érosif dans leur milieu est élevé est observée chez les cultivateurs alors que la plus faible (38,1 %) a été observée chez les locataires.
Le tableau 3.2 montre les fréquences des modalités des variables relatives au milieu dans lequel l'enquêté évolue.

Tableau 3.2: Répartition de l'échantillon selon les facteurs relatifs aux milieux de vie

Variables	Modalités	N	N^+	%	IC à 95%
Distance du ravin au ménage (en mètres)	<6	82	53	64.6	[54,2 ;75,1]
	[6-9]	60	33	55	[42,3 ; 67,7]
	[10-19]	127	57	44.9	[36,2 ; 53,6]
	[20 et plus	136	58	42.6	[34,3; 51]
Interruption du sommeil quand il pleut	Oui	209	145	69.4	[63,1 ; 75,7]
	Non	196	56	28.6	[22,2; 34,9]
Appréciation du degré de la pente	Faible	106	45	42.5	[33 ; 51,9]
	Moyen	134	42	31.3	[23,4 ; 39,3]
	Elevé	165	114	69.1	[62; 76,2]
Matériaux de construction utilisés	Matériaux durables	164	56	34.2	[26,8; 41,4]
	Matériaux semi-durables	139	76	54.7	[46,3; 63]
	Briques adobes	102	69	67.6	[58,5 ; 76,8]
Avoir été victime de l'écroulement	Oui	150	101	67.3	[59,8 ; 74,9]
	Non	255	100	39.2	[33,2 ; 45,2]
Constant du phénomène d'érosion hydrologique	Oui	188	116	61.7	[54,7 ; 68,7]
	Non	217	85	39.2	[32,6 ; 45,7]
Sols sablonneux	Oui	124	73	58.9	[50,1; 67,6]
	Non	281	128	45.6	[39,7; 51,4]
Sols insuffisamment protégés par le couvert végétal	Oui	259	144	55.6	[49,5 ; 61,7]
	Non	146	57	39	[31,1 ; 47]
Niveau de dégradation de la végétation	Faible	65	12	18.5	[8,9 ; 28]
	Moyen	174	57	32.8	[25,7 ; 39,8]
	Elevé	166	132	79.5	[73,3 ; 85,7]
Evénements dommageables multiples	Oui	169	133	78.7	[72,5 ; 84,9]
	Non	236	68	28.8	[23 ; 34,6]
L'homme ne peut-il pas maitriser le risque ?	Oui	263	160	60.8	[54,9 ; 66,8]
	Non	142	41	28.9	[21,4 ; 36,4]
Ensemble		405	201	49.6	[44,7 ; 54,5]

Pour les caractéristiques relatives au milieu de vie, la proportion la plus élevée (79,5 %) de personnes qui estiment que le risque hydro-érosif dans leur milieu est élevé est observée chez les personnes qui trouvent que le niveau de dégradation de la végétation est élevé alors que la plus faible (28,8 %) a été observée chez les personnes qui n'ont pas observé des événements dommageables multiples.

Le tableau 3.3 montre les statistiques descriptives (effectif, minimum, moyenne, déviation standard, maximum) des variables quantitatives à savoir l'âge (en années) de l'enquêté, la distance (en m) qui le séparé du ravin, le temps (en mois) qu'il vient de passer dans le milieu de résidence (Temps) et la taille du ménage.

Tableau 3.3: Statistiques descriptives des variables quantitatives

Variable	Effectif	Minimum	Moyenne	Déviation standard	Maximum
Age (années)	405	15	34.13	10.87	74
Distance (m)	405	0	15.97	13.42	80
Temps (mois)	405	0	6.68	6.70	53
Taille	405	1	5.69	2.51	15

La distance du ravin varie entre 0 et 80 m tandis que le temps que l'enquêté vient de passer dans le milieu varie de 0 à 53 mois.

3.2 Tests du chi-deux et coefficient de corrélation

Le tableau 3.4 montre les résultats des tests du chi-deux d'indépendance de Pearson et les coefficients de corrélation entre la variable dépendante et chacune des variables explicatives.

Tableau 3.4: Tests du chi-deux et coefficient de corrélation

Variables	Statistique	dll	p-value	V Clamer
Commune	0,06	2	0.971	0,01
Sexe	3.48	1	0.062	0,09
Niveau d'instruction	7,54	3	0.057	0,14
Statut matrimonial	5,30	2	0.071	0,11
Profession	16,98	3	0.0007	0,20
Mode de vie	23,24	2	<0,001	0,24
Distance du ravin au ménage(en mètres)	11.87	3	0.008	0,17
Interruption du sommeil quand il pleut	65.37	1	<0,001	0,41
Appréciation du degré de la pente	45.11	2	<0,001	0,33
Matériaux de construction utilisés	30.39	2	<0,001	0,27
Avoir été victime de l'écroulement	29.87	1	<0,001	0,27
Constant du phénomène d'érosion hyd	20.46	1	<0,001	0,22
Sols sablonneux	6,11	1	0.013	0,12
Sols insuffisamment protégés par le couvert végétal	10,24	1	0.001	0,16
Niveau de dégradation de la végétation	104.39	2	<0,001	0,51
Evénements dommageables multiples	98,03	1	<0,001	0,49
L'homme ne peut-il pas maitriser le risque ?	37,69	1	<0,001	0,31

Au seuil de 5 %, l'hypothèse nulle d'indépendance entre la variable dépendante et les variables Commune, Sexe, Niveau d'instruction et Statut matrimonial n'est pas rejetée, ce qui montre une indépendance entre ces variables absence et le fait de trouver que le risque hydro-érosif dans le milieu est élevé. Les coefficients V de Cramer respectifs sont faibles, ce qui montre une relation nulle ou très faible (V<0,10) ou faible (V compris entre 0,10 et 0,20). Les autres variables sont significativement liés à la variable dépendante car leur p-valeur est inférieure à 5 %.

3.3 Modèle logistique simple

Le modèle logistique simple sert à établir la relation entre la variable dépendante et chaque variable explicative. Elle permet de déterminer ou de donner une idée de la contribution de chaque variable explicative d'une façon globale à l'explication de la variable dépendante grâce à la statistique du test de Wald qui est la p-value.
Le tableau 3.5 montre les résultats du modèle logistique simple pour les variables sociodémographiques.

Tableau 3.5: Modèle logistique pour les variables sociodémographiques

Variable	Modalités	OR	IC à 95%	P-value
Commune				0.971
	Ntahangwa	1.00		
	Mukaza	1.06	[0.66;1.71]	0.808
	Muha	1.03	[0.64;1.66]	0.903
Sexe				0.062
	Homme	1.00		
	Femme	1.45	[0.98;2.16]	0.062
Niveau d'instruction				0.059
	Aucun	1.00		
	Primaire	0.73	[0.40;1.35]	0.316
	Secondaire	0.48	[0.25;0.91]	0.025
	Universitaire	0.47	[0.23;0.95]	0.035
Status matrimonial				0.073
	Célibataire	1.00		
	Marié	1.55	[1.01;2.37]	0.047
	Divorcé/veuf	2.27	[0.88;5.86]	0.091
Profession				0.001
	Cultivateur	1.00		
	Fonctionnaire	0.31	[0.16;0.62]	0.001
	Commercant	0.36	[0.20;0.66]	0.001
	Autre	0.56	[0.35;0.91]	0.019
Mode de vie				<0.001
	Locataire	1.00		
	Propiétaire	2.62	[1.67;4.11]	<0.001
	Autre	0.96	[0.53;1.75]	0.898

Le tableau 3.6 montre les résultats du modèle logistique simple pour les variables relatives aux milieux de résidence des répondants.

Tableau 3.6: Modèle logistique pour les variables liées au milieu de vie

Variable	Modalités	OR	IC à 95%	P-value
Distance au ravin (m)				0.009
	20+	1.00		
	<6	2.46	[1.40;4.33]	0.002
	6-9	1.64	[0.89;3.03]	0.111
	10-19	1.10	[0.67;1.78]	0.715
Interruption du sommeil				0
	Non	1.00		
	Oui	5.66	[3.69;8.68]	<0.001
Degré de la pente				<0.001
	Faible	1.00		
	Moyen	0.62	[0.36;1.05]	0.076
	Elevé	3.03	[1.82;5.03]	<0.001
Matériaux de construction				<0.001
	Matériaux durables	1.00		
	Matériaux semi-durables	2.33	[1.46;3.7]	<0.001
	Briques adobes	4.03	[2.38;6.82]	<0.001
Victimes de l'écoulement				<0.001
des ravins	Non	1.00		
	Oui	3.19	[2.09;4.88]	<0.001
Constat des phénomènes				<0.001
hydro-érosifs	Non	1.00		
	Oui	2.50	[1.68;3.74]	<0.001
Sols sablonneux				<0.014
	Non	1.00		
	Oui	1.71	[1.12;2.62]	<0.014

Variable	Modalités	OR	IC à 95%	P-value
Sol insuffisamment protégé par				**0.001**
le couvert végétal	**Non**	**1.00**		
	oui	**1.96**	**[1.29;2.96]**	**0.001**
Niveau de dégradation de la				**<0.001**
végétation	**Faible**	**1.00**		
	Moyen	**2.15**	**[1.07;4.34]**	**<0.032**
	Elevé	**17.15**	**[8.25;35.63]**	**<0.001**
Événements dommageables multiples				**<0.001**
	Non	**1.00**		
	oui	**9.13**	**[5.74;14.51]**	**<0.001**
L'homme peut-il maitriser le				**<0.001**
risque?	**Non**	**1.00**		
	oui	**3.83**	**[2.47;5.94]**	**<0.001**
Victimes de l'écoulement des				**<0.001**
ravins	**Non**	**1.00**		
	Oui	**3.19**	**[2.09;4.88]**	**<0.001**
Constat des phénomènes				**<0.001**
hydro-érosifs	**Non**	**1.00**		
	Oui	**2.50**	**[1.68;3.74]**	**<0.001**
Sols sablonneux				**<0.014**
	Non	**1.00**		
	Oui	**1.71**	**[1.12;2.62]**	**<0.014**

Au seuil de 10 %, le modèle logistique simple montre que les variables : le sexe, le niveau d'instruction, le statut matrimonial, la profession, le mode de vie, la distance entre le ménage et le ravin, le fait d'interrompre le sommeil pendant la nuit, le degré de la pente, les matériaux de construction, le fait d'avoir été victime de l'écroulement des ravins, le constat des phénomènes hydro-érosifs, le sol sablonneux, le sol insuffisamment protégé par le couvert végétal, le niveau de dégradation de la végétation, le fait d'avoir eu des événements dommageables multiples et le fait que l'homme peut maîtriser le risque sont significatives et seront donc utilisées dans le modèle multiple, car leurs p-values sont inférieures à 10 %.

3.4 Modèle logistique multiple

Ce modèle, dit multiple, met en évidence toutes les variables pertinentes (celles sélectionnées dans le modèle logistique simple, à cause de leurs p-values

inférieures à 10%) à pouvoir expliquer la variable dépendante.

Le tableau 3.7 montre les résultats du modèle logistique multiple.

Tableau 3.7: Modèle logistique multiple

Variable	Modalités	OR	IC à 95%	P-value
Sexe				0.062
	Homme	1.00		
	Femme	1.32	[0.72;2.40]	0.367
Niveau d'instruction				0.059
	Aucun	1.00		
	Primaire	1.27	[0.48;3.33]	0.630
	Secondaire	1.45	[0.50;4.22]	0.495
	Universitaire	1.48	[0.42;5.29]	0.544
Status matrimonial				0.900
	Célibataire	1.00		
	Marié	1.16	[0.52;2.59]	0.719
	Divorcé/veuf	0.92	[0.20;4.21]	0.917
Profession				0.001
	Cultivateur	1.00		
	Fonctionnaire	0.47	[0.13;1.69]	0.248
	Commercant	0.32	[0.13;0.79]	0.014
	Autre	1.08	[0.45;2.61]	0.863
Mode de vie				0.065
	Locataire	1.00		
	Propiétaire	2.41	[1.15;5.04]	0.020
	Autre	1.32	[0.49;3.54]	0.579
Distance au ravin (m)				0.829
	20+	1.00		
	<6	1.48	[0.64;3.42]	0.361
	6-9	1.28	[0.51;3.21]	0.595
	10-19	1.14	[0.55;2.35]	0.724
Interruption du sommeil				0.002
	Non	1.00		
	Oui	2.95	[1.49;5.88]	0.002
Degré de la pente				0.048
	Faible	1.00		
	Moyen	0.39	[0.17;0.87]	0.022
	Elevé	0.75	[0.33;1.69]	0.487
Matériaux de construction				0.513
	Matériaux durables	1.00		
	Matériaux semi-durables	1.37	[0.66;2.86]	0.402
	Briques adobes	1.65	[0.69;3.95]	0.263
Victimes de l'écoulement				0.501
des ravins	Non	1.00		
	Oui	0.78	[0.38;1.60]	0.501

Constat des phénomènes				**0.865**
hydro-érosifs	**Non**	**1.00**		
	Oui	**0.95**	**[0.51;1.77]**	**0.865**
Sols sablonneux				**0.001**
	Non	**1.00**		
	Oui	**2.85**	**[1.5;5.42]**	**0.001**
Niveau de dégradation de la végétation				**<0.001**
	Faible	**1.00**		
	Moyen	**2.07**	**[0.76;5.66]**	**0.154**
	Elevé	**9.99**	**[3.58;27.89]**	**<0.001**
Evénement dommageables multiples				**<0.001**
	Non	**1.00**		
	Oui	**4.10**	**[2.21;7.60]**	**<0.001**
Sol insuffisamment protégé par				**0.002**
le couvert végétal	**Non**	**1.00**		
	Oui	**2.75**	**[1.45;5.19]**	**0.002**
Evénements dommageables multiples				**<0.001**
	Non	**1.00**		
	Oui	**4.10**	**[2.21;7.60]**	**<0.001**
L'homme peut-il maitriser le risque?				**0.004**
	Oui	**1.00**		
	Non	**2.59**	**[1.35;4.97]**	**0.004**

3.5 Modèle optimal ou final

Le modèle final est obtenu par la technique de sélection manuelle descendante pas-à-pas. Les variables retenues pour le modèle final ont une p-value inférieure à 5 % par le test de significativité de Wald.
Le tableau 3.8 montre les variables retenues dans le modèle final.

Au seuil de 5 %, les variables comme la profession commerçante, le fait d'interrompre le sommeil pendant la nuit, le degré de la pente, le sol sablonneux, le sol insuffisamment protégé par le couvert végétal, le niveau élevé de dégradation de la végétation, le fait d'avoir eu des événements dommageables multiples et le fait que l'homme ne peut pas maîtriser le risque influencent significativement le fait de trouver que le risque hydro-érosif dans le milieu de vie est élevé car leurs p-values sont inférieures à 5 %. Le pseudo coefficient de détermination de McFadden 0,41. Cette valeur montre que le modèle final n'est pas mal ajusté car il est différent de 0.

Tableau 3.8: Modèle final

Variable	Modalités	OR	IC à 95%	P-value
Profession				0.018
	Cultivateur	100		
	Fonctionnaire	0.48	[0.19;1.24]	0.130
	Commercant	0.29	[0.13;0.65]	0.002
	Autre	0.73	[0.37;1.45]	0.3733
Interruption du sommeil				0.001
	Non	1.00		
	Oui	3.20	[1.77;5.78]	0.001
Degré de la pente				0.017
	Faible	1.00		
	Moyen	0.35	[0.16;0.77]	0.009
	Elevé	0.75	[0.35;1.61]	0.455
Sols sablonneux				0.001
	Non	1.00		
	Oui	2.88	[1.55;5.33]	0.001
Sol insuffisamment protégé par le couvert végétal				0.002
	Non	1.00		
	Oui	2.67	[1.45;4.92]	0.002
Niveau de dégradation de la végétation				<0.001
	Faible	1.00		
	Moyen	2.46	[0.97;6.25]	0.058
	Elevé	10.9	[4.23;28.1]	<0.001
Evénements dommageables multiples				<0.001
	Non	1.00		
	Oui	3.90	[2.16;7.04]	<0.001
L'homme peut-il maitriser le risque?				0.0004
	Oui	1.00		
	Non	2.94	[1.61;5.36]	0.0004

Les résultats du test de Wald utilisé pour le test d'adéquation du modèle final montre que ce modèle est adéquat ($\chi^2 = 115,21$; $ddl = 12$; $p-value < 0.001$)**, ce qui montre que tous les paramètres ne sont pas simultanément nuls ou qu'il existe au moins une variable qui explique le modèle. Pour le test de Wald, la p-value (<0,001) obtenue est inférieure au seuil de 5 %. Ainsi, l'hypothèse nulle qui stipule que le modèle n'est pas adéquat ou qu'il n'y a pas de variables qui expliquent le modèle n'est pas rejetée. Il existe au moins une variable qui explique le modèle. Le test du chi-deux de Pearson donne des résultats qui montrent que l'hypothèse nulle qui stipule que le modèle est bon ou que les valeurs observées et les valeurs prédites de la probabilité de trouver que le risque hydro-érosif dans le milieu de vie est élevé n'est pas rejetée car la p-value (0,090) est supérieure à 5 %. Le modèle est donc bon. Le link test**

ne rejette pas l'hypothèse nulle qui dit que le carré du terme additionnel est nul, ce qui montre que le modèle est bon. La courbe ROC montre que le modèle final a une performance et un pouvoir prédictif puissant, avec une discrimination acceptable (AUC=0,89).

3.6 Analyse des résidus

La distance de Cook est une mesure de l'influence d'une observation sur l'équation de régression. Les données aberrantes et/ou avec un fort effet de levier sont les principales causes de la non fiabilité des résultats et de l'imprécision d'une régression. Six observations (6/405) et dix observations (10/405) ont des résidus studentisés respectivement inférieurs à -2 et supérieurs à 2, soit un total de 16 observations (3,95 %) aberrantes. La figure 3.1 montre l'impact de chaque observation sur le modèle final. Les résidus sont bien séparés par la ligne horizontale passant par l'origine des axes. La figure 3.2 montre la distance de Cook. Aucune distance de Cook n'est supérieure à 1,74, ce qui montre qu'il n'y a d'observation ayant une influence notable. Les données ne présentent donc aucun point influent. Cela signifie qu'il n'y a aucun point ayant une distance de Cook élevée pouvant biaisé les résultats du modèle final. La figure 3.3 montre les leviers. Trente-et-une observations (31/405) ont une valeur du levier qui dépasse la valeur de 0,064.
La figure 3.4 montre la distribution des résidus de Pearson standardisé en fonction des leviers. Les lignes de distance de Cook (en pointillées rouges) ne sont pas affichées sur le graphique des résidus versus leviers car tous les points sont bien à l'intérieur des lignes de distance de Cook.

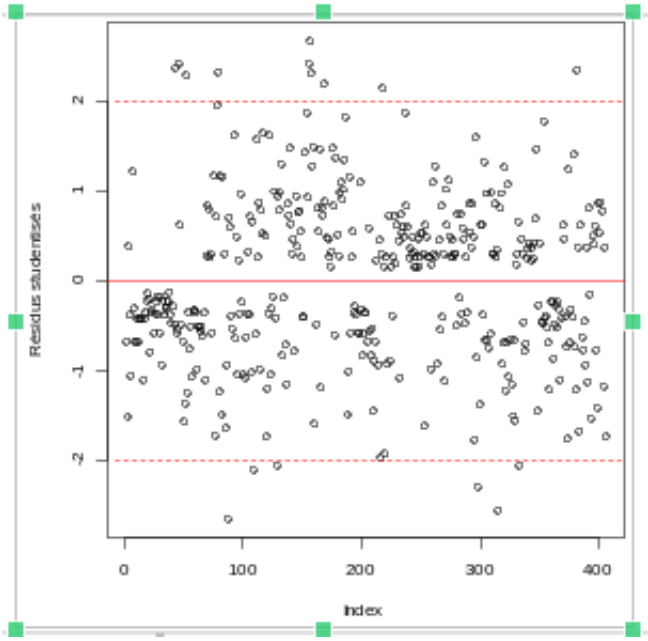


Figure 3.1: Résidus studentisé

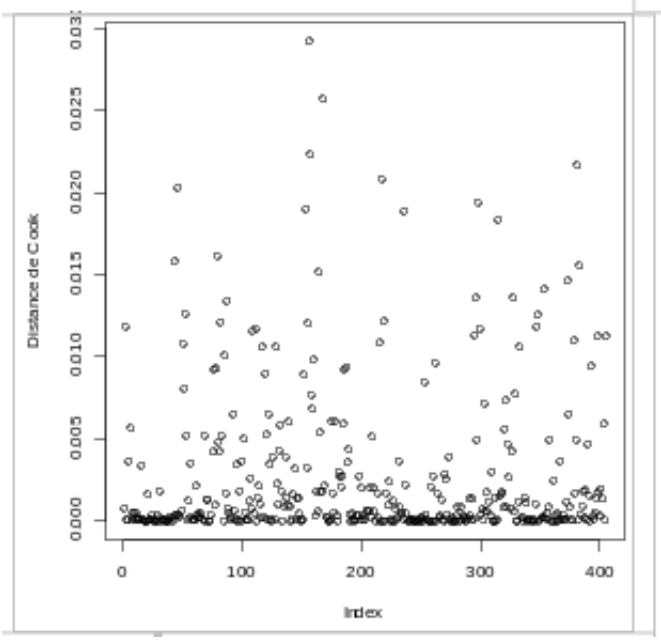


Figure 3.2: Distance de Cook

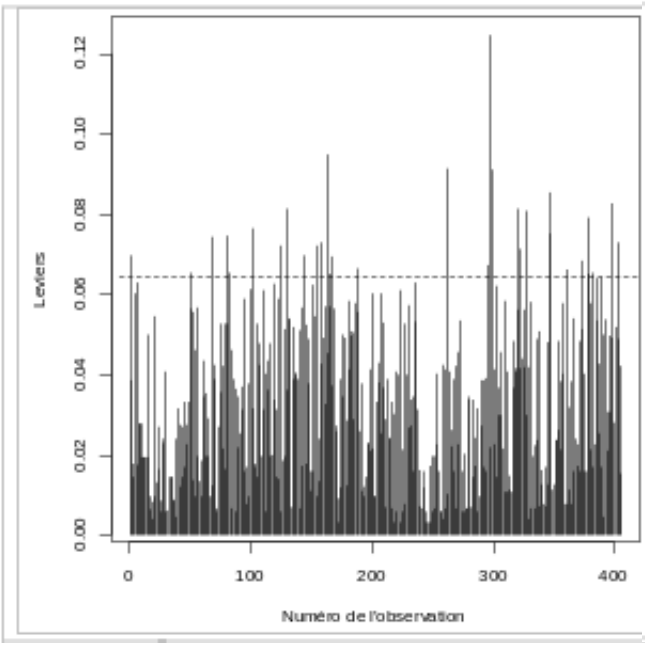


Figure 3.3: Leviers

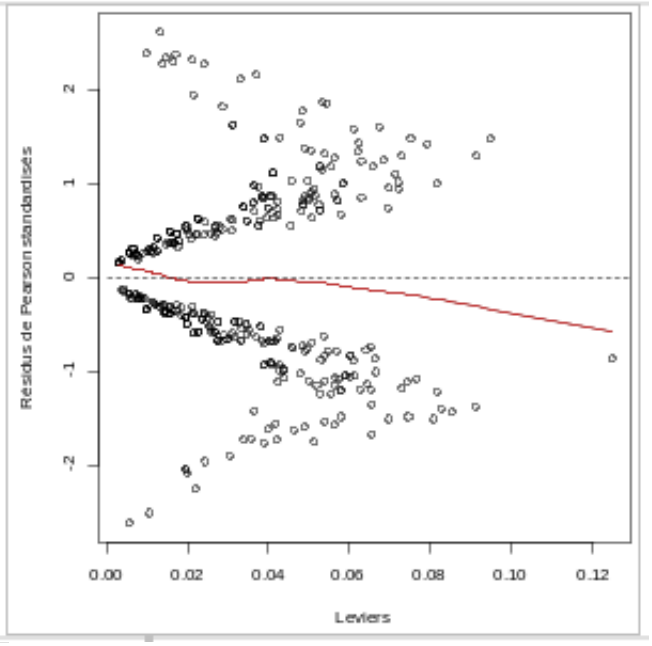


Figure 3.4: Résidus versus leviers

$$\begin{aligned} logit(p) = \hat{\beta}_0 &+ \hat{\beta}_{11} Fonctionnaire + \hat{\beta}_{12} Commerçant + \hat{\beta}_{13} Autre + \hat{\beta}_2 Sommeil \\ &+ \hat{\beta}_{31} Moyen + \hat{\beta}_{32} Elevé + \hat{\beta}_4 Solssablonneux + \hat{\beta}_5 Couvert \\ &+ \hat{\beta}_{61} Moyen + \hat{\beta}_{62} Elevé + \hat{\beta}_7 Evénements + \hat{\beta}_8 Maitriser + \epsilon \end{aligned} \tag{3.1}$$

$$p = P(y = 1) \tag{3.2}$$

avec

$$\hat{\beta}_0 = -3,30 \tag{3.3}$$
$$\hat{\beta}_{11} = -0,73 \tag{3.4}$$
$$\hat{\beta}_{12} = -1,22 \tag{3.5}$$
$$\hat{\beta}_{13} = -0,31 \tag{3.6}$$
$$\hat{\beta}_2 = 1,16 \tag{3.7}$$
$$\hat{\beta}_{31} = -1,04 \tag{3.8}$$
$$\hat{\beta}_{32} = -0,29 \tag{3.9}$$
$$\hat{\beta}_4 = 1,06 \tag{3.10}$$
$$\hat{\beta}_5 = 0,98 \tag{3.11}$$
$$\hat{\beta}_{61} = 0,90 \tag{3.12}$$
$$\hat{\beta}_{62} = 2,39 \tag{3.13}$$
$$\hat{\beta}_7 = 1,36 \tag{3.14}$$
$$\hat{\beta}_8 = 1,08 \tag{3.15}$$

y=le fait de sentir que le risque hydro-érosif est élevé

DISCUSSION DES RÉSULTATS

Cette étude s'est intéressée à la manière dont les habitants de la Mairie de Bujumbura perçoivent le risque hydro-érosif des ravins qui les environnent et les causes(avec les conséquences) qui en découlent.Les résultats de cette étude montrent que la proportion des habitants qui trouvent que le risque hydro-érosif est élevé dans leur milieu est de 49,6 %. Cette proportion est faible du point de vue des autres recherches menées à ce sujet. En effet, selon Louis Yembi-Yembi et Stéphane Costa (2022) [Yembi-Yembi *et al.*, 2022], le résultat du modèle final obtenu explique que le risque élevé se situe entre 55% et 62% de la variation du sentiment d'exposition pour l'ensemble des variables. C'était le seul modèle sur quatre traités qui avait une performance supérieure à 50%. Une autre étude réalisée au Burkina-Faso (La revue électronique en sciences de l'environnement, 2019)[Kabore *et al.*, 2019] a montré que 76,7% des personnes sondées ont constaté une baisse des précipitations annuelles et 50,7% ont trouvé des pluies irrégulières. De même, nos résultats montrent que le fait d'être propriétaire est associé à une perception élevé (61,7%) du risque hydro-érosif du ravin que les locataires (38,1%). Ces résultats sont en accord avec ceux de cette même étude réalisée par Louis Yembi-Yembi et Stéphane Costa (2022)[Yembi-Yembi *et al.*, 2022] qui affirmait que le fait d'être propriétaire (72,75%) du logement est associée à une représentation plus élevée du sentiment d'exposition que les locataires (27,25 %). Une autre recherche effectuée par FRANCIS DUHAMEL (2021)[DUHAMEL, 2021] a montré également que le fait d'être propriétaire (87,7 %) du logement représente une perception très élevée du sentiment d'exposition à un risque environnemental que les locataires (12,3 %).

L'analyse multivariable a montré que le niveau de dégradation de la végétation, les événements dommageables multiples, l'interruption du sommeil, le degré de la pente, le fait que l'homme ne peut pas maîtriser le risque et les sols sablonneux ont une influence significative sur la perception du risque hydro-érosif après ajustement sur la profession et le sol insuffisamment protégé par le couvert végétal.Selon une étude réalisée en 2021 sur la gestion

des inondations fluviales à Sussex et Sussex Corner à l'Université du Québec à Montréal, 75 bâtiments résidentiels considérés dans leur étude avaient un niveau de risque élevé et plus de 75% de ces bâtiments se trouvaient à Sussex dans les zones A et B (FRANCIS DUHAMEL, 2021)[DUHAMEL, 2021]. Ce risque élevé s'explique par leur vulnérabilité socio-économique forte. Les auteurs affirment que la zone D est moins à risque puisque ses habitants ont adopté considérablement plus de mesures de préparation face aux inondations.La même recherche affirme que les autorités devraient donc leur attribuer une attention particulière étant donné qu'ils sont à la fois très exposés et vulnérables (FRANCIS DUHAMEL, 2021)[DUHAMEL, 2021]. En effet, nos résultats trouvent que pour les matériaux de construction utilisés, les maisons construites en matériaux durables représentent 34,2%, celles construites avec des matériaux semi-durables représentent 54,7% et celles avec des briques adobes représentent 67,6%. Avec nos résultats, comparativement aux études de FRANCIS DUHAMEL sur le risque élevé des bâtiments, puisque l'enquête a visé les habitants proches des ravins, ceux avec des maisons construites avec des briques adobes ont le plus de risque que ceux avec des maisons en matériaux semi-durables (respectivement avec ceux en matériaux durables). Une étude publiée très récemment par le Journal OpenEdition au Burkina-Faso a montré que le niveau d'éducation influence significativement la perception paysanne du changement des précipitations (Louis Yembi-Yembi et Stéphane Costa, 2022)[Yembi-Yembi *et al.*, 2022]. Ces résultats sont en accord avec nos résultats puisque le niveau d'instruction est parmi les facteurs qui influencent la perception du risque hydro-érosif des ravins. Dans cette même étude, les résultats montrent que les perceptions du changement climatique par les populations locales du Centre-Nord sont en adéquation avec les tendances climatiques, ce qui est aussi en accord avec nos résultats par exemple en considérant les facteurs comme les sols insuffisamment protégés par le couvert végétal et le niveau de dégradation de la végétation. L'intérêt de notre étude est qu'une analyse de perception du risque, avec toute l'information qu'elle peut contenir, servira d'outil d'aide à la décision ou de complément à une approche qui traite les habitants des sites à risque et les propriétaires comme des acteurs responsables de la gestion des risques des ravins. L'instauration d'une culture du risque nécessite une compréhension et une prise en compte des facteurs sociaux qui influencent l'interaction entre ces acteurs.

Conclusion générale

La proportion d'habitants qui trouvent que le risque hydro-érosif dans leur milieu de résidence est élevé est de 49,6 %. Le sentiment envers le risque hydro-érosif est un élément indispensable qui permet de gagner la confiance en soi-même et la satisfaction des habitants. Plusieurs facteurs seraient liés à la perception du risque hydro-érosif des ravins traversant la ville de Bujumbura.

Pour cette étude, au seuil de 5 %, le niveau de dégradation de la végétation, les événements dommageables multiples, l'interruption du sommeil, le degré de la pente, le fait que l'homme ne peut pas maîtriser le risque et les sols sablonneux influencent significativement la perception du risque hydro-érosif après ajustement sur la profession et le sol insuffisamment protégé par le couvert végétal.

Ces résultats permettent de confirmer les hypothèses :
H_1 **: Le degré de la pente influence la perception du risque hydro-érosif des ravins ;**
H_2 **: Les événements dommageables multiples agissent sur la perception du risque hydro-érosif des ravins ;**
H_4 **: La distance au ravin influence la perception du risque hydro-érosif des ravins ;**
H_5**: Le sol sablonneux influence la perception du risque hydro-érosif des ravins.**

L'hypothèse suivante est infirmée :
H_3 **: Le fait d'avoir été victime de l'écroulement des ravins a un effet sur la perception du risque hydro-érosif des ravins.**

Perspectives

Dans l'objectif de mieux évaluer la perception du risque hydro-érosif des ravins et de développer des mesures d'adaptations spécifiques et locales, l'hypothèse est de voir si l'environnement est perçu différemment selon que les acteurs sont directement concernés ou non par les catastrophes naturelles (ici, les glissements de terrains (les ravins) et selon leur utilisation du territoire et que les acteurs de proximité que sont les agriculteurs, les habitants, les élus locaux sont prêts à modifier leurs conduites dans le but de réduire les risques des ravins, voir si la détermination des risques et des enjeux sur une zone à risque modifie les comportements des décideurs locaux face à l'aménagement des lieux habitables. Il serait donc avantageux d'intégrer les perceptions et les connaissances ainsi que l'héritage culturel de la population dans la prise de décision afin de mettre à disposition des acteurs des informations suffisantes et crédibles qui leur permettront de mieux anticiper, de prévoir ou de planifier des réponses adéquates aux impacts potentiels de la variabilité climatique, hydro-érosif et géologique dans les secteurs déterminés. Le travail d'enquête est alors un moyen efficace de comprendre comment la population locale perçoit et parvient à connaître le rapport entre leurs installations et les différents types de risques, de menaces et de défis dont ils auront à affronter en cas de catastrophes naturelles. Ceci permettrait éventuellement la construction d'une base des données (cognitive référentielle, statistique et cartographique) qui viendra compléter les travaux scientifiques dans le but d'atténuer l'impact des désastres, et l'amélioration de la protection des personnes et des biens contre les ravins, les écroulements des terrains ou même les glissements de terrains ; et non pas un moyen d'éliminer complètement les dégâts. En effet, ces résultats soulèvent également d'autres pistes de recherche qui méritent d'être approfondies. Par exemple, l'identification des raisons pour lesquelles les populations s'installent dans des zones considérées comme à risque, et la hiérarchisation des groupes sociaux en fonction de leur vulnérabilité en termes de représentation du risque. Dans le cadre de la connaissance du risque d'un ravin par les habitants, il s'agira question de voir comment

les populations spatialisent l'aléa et les enjeux, dans un objectif de comparabilité de la spatialisation de l'aléa, modélisée par des experts en le domaine et représentée par la population. En perspectives, il sera aussi question de cartographier le risque des ravins, de modéliser spatialement la perception du risque et de calculer les probabilités prédites.

Bibliographie

[Ahouangan *et al.*, 2014] Ahouangan, M. B. D., Djaby, B., Ozer, P., Hountondji, Y.-C., Thiry, A. et De Longueville, F. (2014). Adaptation et résilience des populations rurales face aux catastrophes naturelles en afrique subsaharienne. cas des inondations de 2010 dans la commune de zagnanado, bénin.

[Altman, 1990] Altman, D. G. (1990). *Practical statistics for medical research*. Chapman and Hall/CRC.

[Amanejieu *et al.*, 2018] Amanejieu, A. *et al.* (2018). Analyse temporelle de la représentation du risque d'inondation de 1980 à 2018 à douala-cameroun.

[Bickerstaff et Walker, 2001] Bickerstaff, K. et Walker, G. (2001). Public understandings of air pollution: the 'localisation'of environmental risk. *Global environmental change*, 11(2):133–145.

[Bonnefoy *et al.*, 2003] Bonnefoy, B., Frère, S., Maramotti, I., Moch, A. et Roussel, I. (2003). La perception de la pollution atmosphérique d'origine industrielle par les habitants de l'agglomération de dunkerque. *Pollution atmosphérique*, 45(178):251–259.

[Brunet, 2008] Brunet, P. (2008). De l'usage raisonné de la notion de «concernement»: mobilisations locales à propos de l'industrie nucléaire. *Natures sciences sociétés*, 16(4):317–325.

[Carter et Horwitz, 2014] Carter, M. et Horwitz, P. (2014). Beyond proximity: The importance of green space useability to self-reported health. *EcoHealth*, 11:322–332.

[Chang *et al.*, 2016] Chang, H. S., Donohoe, M. et Sougiannis, T. (2016). Do analysts understand the economic and reporting complexities of derivatives? *Journal of Accounting and Economics*, 61(2-3):584–604.

[Charvolin *et al.*, 2017] Charvolin, F., Frioux, S., Kamoun, L., Mélard, F. et Roussel, I. (2017). ***Un air familier?: sociohistoire des pollutions atmosphériques.*** Presses des Mines via OpenEdition.

[Chikhi et Chavance, 2012] Chikhi, M. et Chavance, M. (2012). Estimation du modèle linéaire généralisé et application. ***Sciences & Technology. A, exactes sciences***, pages 13–21.

[Dozzi *et al.*, 2008] Dozzi, J., Lennert, M. et Wallenborn, G. (2008). Inégalités écologiques: analyse spatiale des impacts générés et subis par les ménages belges. ***Espace populations sociétés. Space populations societies***, (2008/1): 127–143.

[DUHAMEL, 2021] DUHAMEL, F. (2021). Gestion des inondations fluviales à sussex et sussex corner, nouveau-brunswick: Une approche par le risque.

[Dutozia et Voiron-Canicio, 2019] Dutozia, J. et Voiron-Canicio, C. (2019). La résilience urbaine face aux risques, en contexte de changement climatique-articuler la gestion des évènements extrêmes et les stratégies de développements des territoires pour opérationnaliser la résilience des villes d'aujourd'hui et de demain-février 2018.

[Eggerickx, 2016] Eggerickx, T. (2016). Les inégalités sociales et spatiales de mortalité en belgique. ***In «Riches et pauvres, à la vie à la mort», CEPAG (Centre d'éducation populaire André Genot).***

[Farley et Gravel, 2013] Farley, D. M.-C. B.-C. et Gravel, F. (2013). Competency framework for scientific processes underlying the cross-cutting functions of the institut national de santé publique du québec.

[Foster, 1972] Foster, G. (1972). A closed-form soil erosion equation for upland areas. ***Sedimentation***, pages 12–1.

[Gourieroux, 1984] Gourieroux, C. (1984). Econométrie des variables qualitatives. ***(No Title).***

[Gruber, 2005] Gruber, J. (2005). ***Public finance and public policy.*** Macmillan.

[Harari-Kermadec, 2009] Harari-Kermadec, H. (2009). Econométrie 2: données qualitatives, probit et logit. ***Université Paris***, 1.

[Harrison, 1988] Harrison, D. (1988). La représentation du risque professionnel et l'autonomie ouvrière. ***Sciences sociales et santé***, 6(3):75–98.

[Hjulström, 1935] Hjulström, F. (1935). ***Studies of the morphological activity of rivers as illustrated by the River Fyris.*** Thèse de doctorat, The Geological institution of the University of Upsala.

[Hurlin, 2003] Hurlin, C. (2003). Econométrie des variables qualitatives, polycopié de cours. ***Maîtrise d'Econométrie, Université d'Orléans.***

[Issaka, 2010] Issaka, H. (2010). ***Mise en carte et gestion territoriale des risques en milieu urbain sahélien à travers l'exemple de Niamey (Niger).*** Thèse de doctorat, Strasbourg.

[Ja, 1983] Ja, H. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. ***Radiology***, 148:839–843.

[Kabore *et al.*, 2019] Kabore, P. N., Barbier, B., Ouoba, P., Kiema, A., Some, L. et Ouedraogo, A. (2019). Perceptions du changement climatique, impacts environnementaux et stratégies endogènes d'adaptation par les producteurs du centre-nord du burkina faso. ***VertigO***, 19(1).

[Kermisch, 2012] Kermisch, C. (2012). Vers une définition multidimensionnelle du risque. ***VertigO-la revue électronique en sciences de l'environnement***, 12(2).

[Khazen, 2019] Khazen, M. (2019). Perception des risques liés à la pollution de l'air à beyrouth (liban).

[Klinke et Renn, 2010] Klinke, A. et Renn, O. (2010). Risk governance: Contemporary and future challenges. ***Regulating chemical risks: European and global challenges***, pages 9–27.

[Lalloué, 2013] Lalloué, B. (2013). ***Méthodes d'analyse de données et modèles bayésiens appliqués au contexte des inégalités socio-territoriales de santé et des expositions environnementales.*** Thèse de doctorat, Université de Lorraine.

[Lusted, 1960] Lusted, L. B. (1960). Logical analysis in roentgen diagnosis: memorial fund lecture. ***Radiology***, 74(2):178–193.

[Melard *et al.*, 2013] Melard, F., Charvolin, F., Blavier, G., Frioux, S., Kamoun, L. et Roussel, I. (2013). Percevoir et objectiver la pollution de l'air. une exploration historique et sociologique de ses «mesures» et de ses débordements. ***Rapport final du projet «HIstoire de la Concertation et***

de la régulation de la Qualité de l'AiR»(HICQUAR) pour le programme Concertation, Décision, Environnement.

[Morelli, 2016a] Morelli, X. (2016a). ***Air pollution, health and social deprivation in an urban setting.*** Thèse de doctorat, Université Grenoble Alpes.

[Morelli, 2016b] Morelli, X. (2016b). ***Air pollution, health and social deprivation in an urban setting.*** Thèse de doctorat, Université Grenoble Alpes.

[MUJIA, 2016] MUJIA, G. G. F. (2016). ***Exposition aux risques morpho-hydrologiques dans deux secteurs périurbains de la ville de Yaoundé. Cas des monts Akok-Ndoué et Mvog-betsi au sud-ouest de la ville.*** UNIVERSITÉ DE YAOUNDÉ. https://www.memoireonline.com, consulté le 28/02/2024.

[Ndabashinze, 2024] Ndabashinze, R. (2024). ***Changement climatique au Burundi : Vers un système d'alerte précoce pour tous.*** Iwacu. https://www.iwacu-burundi.org/changement-climatique-au-burundi-vers-un-systeme-dalerte-precoce-pour-tous/, consulté le 04/03/2024.

[Niyongabo, 2021] Niyongabo, J. M. V. (2021). ***Mairie de Bujumbura : 1ère province exposée aux catastrophes naturelles.*** Burundi Eco. https://burundi-eco.com/mairie-bujumbura-1ere-province-exposee-aux-catastrophes-naturelles/, consulté le 28/02/2024.

[Noël *et al.*, 2020] Noël, C., Rodríguez Loureiro, L., Gadeyne, S., Vanroelen, C. et Casas, L. (2020). Environmental injustice in outdoor air pollution.

[Rode, 2009] Rode, S. (2009). ***Au risque du fleuve. La territorialisation de la politique de prévention du risque d'inondation en Loire moyenne.*** Thèse de doctorat, Université de Nanterre-Paris X.

[Roose *et al.*, 1994] Roose, E. *et al.* (1994). ***Introduction à la gestion conservatoire de l'eau, de la biomasse et de la fertilité des sols (GCES)***, volume 70. Organisation des Nations Unies pour l'alimentation et l'agriculture.

[Sheppard *et al.*, 2012] Sheppard, A. J., Salmon, C., Balasubramaniam, P., Parsons, J., Singh, G., Jabbar, A., Zaidi, Q., Scott, A., Nisenbaum, R., Dunn, J. *et al.* (2012). Are residents of downtown toronto influenced by their urban neighbourhoods? using concept mapping to examine neighbourhood characteristics and their perceived impact on self-rated mental well-being. ***International journal of health geographics***, 11:1–11.

[Stephenne *et al.*, 2017] Stephenne, N., Beaumont, B., Hallot, E., Lenartz, F., Lefebre, F., Lauwaet, D., Poelmans, L. et Wolff, E. (2017). Exposure and vulnerability geospatial analysis using earth observation data in the city of liege, belgium. ***ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences***, 4:149–156.

[Tamhane *et al.*, 2016] Tamhane, A. R., Westfall, A. O., Burkholder, G. A. et Cutter, G. R. (2016). Prevalence odds ratio versus prevalence ratio: choice comes with consequences. ***Statistics in medicine***, 35(30):5730–5735.

[Van Erkel et Peter, 1998] Van Erkel, A. R. et Peter, M. (1998). Receiver operating characteristic (roc) analysis: basic principles and applications in radiology. ***European Journal of radiology***, 27(2):88–94.

[Wolf et Cochran, 1997] Wolf, B. L. et Cochran, K. R. (1997). Floating patterns of metered dose inhalers. ***Journal of Asthma***, 34(5):433–436.

[Yembi-Yembi *et al.*, 2022] Yembi-Yembi, L., Costa, S. et Bonnet, E. (2022). Perceptions et représentations socio-spatiales des risques dans les communes multirisques: le cas de la place du risque d'inondation par la mer chez les habitants de port-gentil au gabon. ***Cybergeo: European Journal of Geography***.