THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

École des Ponts ParisTech

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

CIRED

# Climate Change, Droughts, and Water Pollution in Sub-Saharan Africa

Doctoral School ED 465, Économie Panthéon Sorbonne

Discipline : Economics

Thesis produced at Paris School of Economics and CIRED

Date of defense : January, 10th, 2023

**Mélanie GITTARD**

**Jury :**

*Referees*

**Eric, STROBL**
Professor at University of Bern

**Christelle, DUMAS**
Professor at University of Fribourg

*Examiners*

**Raja, CHAKIR**
Research Director at INRAE, Paris-Saclay Applied Economics

**Liam, WREN-LEWIS**
Professor at PSE, Research Fellow at INRAE

*Advisors*

**Denis, COGNEAU**
Professor at PSE, Director of Studies at EHESS, Research Director at IRD

**Philippe, QUIRION**
Research Director at CNRS, CIRED

THÈSE DE DOCTORAT

de l'École des Ponts ParisTech

# Changement climatique, sécheresses et pollution de l'eau en Afrique subsaharienne

École doctorale N° ED465, Économie Panthéon Sorbonne

Spécialité du doctorat : Économie

Thèse préparée au sein de Paris School of Economics et CIRED

Thèse soutenue le 10 Janvier 2023, par :

**Mélanie GITTARD**

**Composition du jury :**

***Rapporteurs***

**Eric, STROBL**
Professeur à l'Université de Berne

**Christelle, DUMAS**
Professeur à l'Université de Fribourg

***Examinateurs***

**Raja, CHAKIR**
Directrice de Recherche à INRAE, Paris-Saclay Applied Economics

**Liam, WREN-LEWIS**
Professeur à PSE, Chargé de Recherche à INRAE

***Directeurs***

**Denis, COGNEAU**
Professeur à PSE, Directeur d'études à l'EHESS, Directeur de Recherche IRD

**Philippe, QUIRION**
Directeur de Recherche au CNRS, CIRED

# Acknowledgement - Remerciements

This thesis would not have been possible without the support of many people whom I would like to warmly thank here.

First, I would like to express my deepest gratitude to my supervisors, Denis Cogneau and Philippe Quirion.
I thank Denis Cogneau for believing in my project from the start and for his involvement in each step of the journey. His tremendous support, starting from the Master thesis in 2017, has been a driving force behind this dissertation. I realize the progress made since the first time I approached him with my desire to work on climate migration. I am very grateful for his availability and for all the very stimulating discussions and feedback. His intellectual rigor and academic excellence have taught me a lot and made me the researcher I have become.
I would also like to thank Philippe Quirion for his generosity in terms of time for discussions and help. I thank Philippe Quirion for meeting with me once a week for several months during a period devoid of social interaction and punctuated by lockdowns. I am very grateful for all the exchanges we had over the Ph.D., and for his intellectual rigor in pushing my projects toward his academic excellence.
Denis Cogneau and Philippe Quirion are true inspirations to me by the high quality of their academic research and their commitment to their work. They have constantly helped me to grow as a researcher in economics, and I will treasure what they taught me throughout my career.

I am grateful to Eric Strobl and Christelle Dumas for accepting to be the referees of this dissertation. I thank Raja Chakir and Liam Wren-Lewis for accepting to be the examiners. I thank all the jury members for their time and dedication and for thoroughly reading the different chapters of my dissertation. I would like to thank Eric Strobl, Christelle Dumas, and Liam Wren-Lewis for the precious feedback during my pre-defense.

The second chapter of this dissertation would not exist without organizations that funded the data purchase. For this, we are indebted to the CEPREMAP, EUR-PJSE, and the GPET thematic group of PSE. We thank the support of the EUR grant ANR-17-EURE-0001.

I am highly grateful to Lara Tobin for providing me with the censuses used in the first chapter of my Ph.D.

Je remercie ma co-auteure Irène Hu, qui m'a accompagnée tout au long de cette thèse et m'a permis de partager cette expérience à deux. Travailler à tes côtés sur un sujet aussi stimulant m'a beaucoup appris, et j'espère que notre collaboration continuera. Au- delà de l'aspect scientifique, merci d'avoir été un binôme ces quatre dernières années, merci pour ton soutien constant, et pour ton amitié.

Throughout my thesis, I had the chance to evolve within great work environments, both at PSE and CIRED.

For PSE, I would like to thank particularly Sylvie Lambert, who was in charge of the doctoral program at PSE and was ever-ready to help any Ph.D. student.

I am very grateful to the research members from the development group for their intensive feedback during the CFDS internal seminars and the interaction during the development drinks. I am grateful to Sylvie Lambert and Karen Macours for creating such a nice atmosphere within the development group, and to the new heads of the department, Liam Wren-Lewis and Oliver Vanden Eynde. I would like also to thank Akiko Suwa-Eisenmann, Jérémie Gignoux, François Libois, Luc Behagel, David Margiolis, and Katrin Millock. I want to thank Katrin Millock, who was on the jury of my Master's thesis. I am deeply grateful for the doctoral course of the development group, from which I learned a lot and which gave me essential tools at the beginning of my Ph.D.
I also had the chance to follow the doctoral course in economic history and follow the seminars. For this, I would like to thank the whole team and, in particular, Jérome Bourdieu and Eric Monnet for the organization.

J'aimerais également remercier l'ensemble de l'équipe administrative de PSE. J'aimerais remercier Lucia-Roxana pour son soutien indéfectible aux doctorants.

Ph.D. life would be a solitary journey without Ph.D. friends! Julieta, my dear co-organizer of the CFDS, my dear friend, thanks for all your support and for listening to me each time I needed it. Thanks also to the third musketeer Victor, who is always cool when things heat up and there to calm the madness. Our little trio has helped me navigate through the Ph.D. journey, knowing I would always have a friend to have my back. Dave, Christian, and my former PPD mates, thanks for your unshakable friendship. Thanks also to my friends from the official *development office R6.01*. Merci à l'ancienne génération, Juliette and Sarah qui ont été comme des grandes soeurs pour nous. Merci à Andrea et Charlotte, pour nos diners et nos verres. Thanks to all the friends from the development group, Kelsey, Matthew, Ismael, Julia, Zhexun, Marion, Duncan, Eric, and Louise. Merci à Justine pour tous ses conseils et pour son amitié.

Merci également à tous mes amis du bureau R5.68. Yajna, merci pour tes conseils et de m'avoir poussé à m'inscrire à des conférences ! Nitin, you have been a huge support, even when you were at Namur. Jeanne, nos cafés vont me manquer. Thanks to Sebastian, the new generation who takes over in the office. Merci à tous mes autres amis de PSE, Adrien, Etienne, Antonela, Jean-Laurent. J'aimerais en particulier remercier Caroline, mon amie depuis le master qui m'apporte depuis son soutien constant. J'admire la docteure que tu es devenue, ton engagement dans les causes que tu défends et sur lesquelles tu travailles, et t'avoir pour amie dans ce métier me pousse encore plus loin.

Côté CIRED, je souhaite remercier l'ensemble de l'équipe du laboratoire. Le CIRED est un environnement à la fois stimulant et chaleureux, en grande partie grâce à l'ensemble des chercheurs et doctorants. Je souhaiterais remercier ceux qui m'ont aidée dans mes recherches, merci à Tristan Le Cotty pour sa lecture méticuleuse de mon premier chapitre, à Tamara Ben Ari et à Louis Gaëtan Giraudet. Merci à Aurélie Méjean et Antoine Missemer pour l'organisation des séminaires du Mardi. Merci à l'équipe des doctorants, Behrang, Emilien, Auriane, Simon, Anne. Marion, tu as été une amie formidable tout le long de cette thèse. Claire, j'ai hâte de poursuivre toutes nos discussions autour de la pollution de l'eau !

Je remercie tous mes autres amis en dehors du monde académique. Merci à mes amies fidèles, Christelle, Laura, Myrto, Sarah G, Selma, Sarah F, Sarah H, Laetitia, Romane, Alix, Roxane, Sophie, Luiza, Jennifer. Merci à Alexandre. Merci à tous mes amis des Mines, Oscar, Arnaud, Gregoire, Yvon, Jubin, Félix, Ilan, Alexis, LB et toute la bande. Thanks to the OECD cool kids team. Merci à ma bande du théâtre pour ces moments d'évasion.

Cette thèse n'aurait pas été possible sans le soutien de ma famille. Je remercie mes parents de m'avoir toujours poussée vers le haut et le maximum de mes compétences, et d'avoir toujours cru en moi. Merci à mon frère et à mes grands-parents.

Enfin, je remercie Antoine pour son soutien constant. Merci d'être présent dans les bons comme les mauvais moments, de savoir me remonter le moral, me faire rire, d'être là tout simplement.

# Summary

This Ph.D. focuses on three applied econometric studies on the consequences of repeated droughts on migration strategies and agricultural vulnerability, as well as the consequences of water resource degradation on health in sub-Saharan Africa.

The **first chapter**, *Droughts, migration, and population in Kenya* looks at the effects of repetitive droughts on internal migration in Kenya over twenty years. Since 2000, East Africa has been facing a decrease in the agricultural season length and an increase in dry conditions. Migration is a possible strategy for adaptation, which can act as a complement or a substitute to on and off-farm adaptations. The Kenyan rural setting has diverse agroecological zones and livelihood systems. Most of the labor force is involved in agriculture and pastoralism, which are highly dependent on rainfall occurrence. Understanding the characteristics of climate-induced migration in rural settings is critical to reducing vulnerability to climate events. Using a unique combination of exhaustive administrative censuses and satellite-based data on daily rainfalls, I exploit the spatial variation in the increase in droughts. I use a difference-in-difference strategy to look at the effects of dry agricultural seasons on population growth and migration at the scale of sublocations over two decades. I show evidence that the increase in drought frequency triggers out-migration, mainly from rural areas where pastoralism is the main economic activity. The main contribution of this paper is the heterogeneity analysis, which investigates the migration response across several socio-economic outcomes. I find that migration forms differ across livelihoods. Within pastoralist systems, the results display little heterogeneity, which suggests the migration of entire households. Herders leave pastoral areas hit by several droughts and migrate towards agriculture-oriented areas with favorable rainfall conditions, which is in line with a rural-rural migration of last resort. The out-migration of agricultural systems is characterized by the migration of the most skilled individuals, in the age of working, while the unskilled individuals are trapped in affected areas. This form of migration is in line with a strategy of income diversification.

The **second chapter**, *MiningLeaks: Water Pollution and Child Mortality in Africa*, is co-authored with Irène Hu. It looks at the effects of mining-induced water pollution on child mortality in Africa. The increase in commodity prices since 2000 has intensified industrial mining activity, and Africa is facing a mining boom attracting foreign investments. Individuals living nearby industrial mines are exposed to high concentrations of heavy metals released during ore extraction processes. The minerals' separation from significant volumes of rocks generates wastes stored in retention ponds where they can leak within the local environment and contaminate local water resources. This paper uses a unique dataset on the location and timing of industrial mining activity in Africa, as we retrieved by intensive hand work the opening dates. Our estimation combines geo-coded information on 2,016 mines with the Demographic Health Survey (DHS) micro-data, which gives information on health outcomes from 1986 to 2018 for 26 countries. Through a staggered difference-in-difference strategy, we exploit the variation in the opening date of mines and the relative topographic position of surrounding DHS villages. We look indirectly at the water-induced pollution as we compare the health outcomes of villages downstream to those upstream of a mining site before and after its opening. Compared to upstream individuals, we find that being downstream of an open mine increases the 24-month mortality rate by 25%. We find an increase in the 12-month mortality only for children that were not exclusively breastfed and consumed plain water, which corroborates the mechanism of water pollution. The effect on mortality is higher in areas with high mining density, fades out with distance, and occurs mainly when the mine is active. The heterogeneity analysis suggests that the increase in mortality is mainly driven by the pollution of open-pit mines, in line with an intensive extractive process, and by foreign-owned only mines. To exclude other mechanisms, we show that the effect is not driven by a change in women's fertility, improved facilities, or the in-migration of the labor force.

The **third chapter** is *Impacts of repetitive droughts and the key role of experience: evidence from Nigeria*. It investigates whether the experience of past dry events can reduce the vulnerability of rural households to short-term shocks in Nigeria. In the 1980s, both the Sahelian and Gulf Guinean regions faced long-lasting severe droughts, which affected rural households who had to find ways to adapt. This paper uses satellite-based data on daily rainfall with a high spatial resolution to analyze the long-term trends of rainfall and how they interact with more recent climate events. I refute the recovery of rains in the Sahel, as I observe less frequent and more intense precipitation in the long run, which is expected to increase the likelihood of

extreme events in recent years. The center and southern parts of the country, mainly agricultural-oriented, were severely hit by dry shocks in 2013 and 2015. Looking at agricultural outcomes, I use three waves of panel household surveys from 2010 to 2016. Through a two-way fixed effect strategy, I exploit the variation in the exposure to the recent droughts and show that droughts decrease crop production by 14% and household food diversity by 1%. Results from a heterogeneity analysis suggest that the effects are attenuated for individuals in charge of the same plot during the intense dry events of the 1980s. This is suggestive evidence ( but not causal) that having a long-lasting experience under dry events on cultivated land reduces vulnerability to rainfall shocks.

# Résumé

Cette thèse se compose de trois études d'économétrie appliquée sur les conséquences de la répétition des sécheresses sur les stratégies de migration et la vulnérabilité agricole, ainsi que les conséquences de la dégradation des ressources en eau sur la santé en Afrique subsaharienne.

Le **premier chapitre**, *Sécheresses, migration et population au Kenya* examine les effets des sécheresses répétitives sur la migration interne au Kenya sur une période de vingt ans. Depuis 2000, l'Afrique de l'Est est touchée par une diminution de la durée de la saison agricole et par une augmentation des conditions sèches. La migration est une stratégie d'adaptation, qui peut servir à la fois de complément ou de substitut aux adaptations sur place et en dehors de l'exploitation agricole. Le Kenya est un pays majoritairement rural, et présente une diversification de zones agroécologiques et de systèmes de subsistance. La majorité de la main-d'œuvre est impliquée dans l'agriculture et le pastoralisme, qui dépendent fortement de la fréquence des précipitations. Comprendre les caractéristiques de la migration climatique est essentiel pour réduire la vulnérabilité des ménages ruraux face aux événements climatiques. À l'aide d'une combinaison unique de recensements administratifs exhaustifs et de données satellitaires sur les précipitations quotidiennes, j'exploite la variation spatiale de l'augmentation de la fréquence des sécheresses. J'utilise une stratégie empirique de Difference-in-Difference (DiD), regardant l'écart des différences, pour examiner les effets des saisons agricoles sèches sur la croissance démographique et la migration à l'échelle des sous locations sur deux décennies. Je montre que l'augmentation de l'occurrence des sécheresses induit une migration de sortie, principalement en dehors des zones rurales à dominante pastorale. La principale contribution de cet article est l'analyse de l'hétérogénéité, qui étudie la réponse migratoire à travers plusieurs dimensions socio-économiques. Je constate que la forme de la migration diffère selon l'activité économique principale. La migration en dehors des zones pastorales présente peu d'hétérogénéité, ce qui suggère une

migration des ménages entiers. Les éleveurs quittent les zones pastorales touchées par des sécheresses répétées et semblent migrer vers des zones à dominance agricole avec des conditions pluviométriques favorables, ce qui suggère une migration rurale-rurale de dernier recours. La migration en dehors des systèmes agricoles présente une forte hétérogénéité, et est dominée par la migration des individus les plus qualifiés, ayant été scolarisés à l'école primaire au minimum, et qui sont jeunes, en âge de travailler. Les individus non qualifiés restent piégés dans les zones touchées par les sécheresses. Cette forme de migration est cohérente avec une stratégie de diversification des revenus.

Le **deuxième chapitre** est *MiningLeaks : pollution de l'eau et mortalité infantile en Afrique* et est co-écrit avec Irène Hu. Il étudie les effets de la pollution de l'eau induite par l'exploitation minière industrielle sur la mortalité infantile en Afrique. Depuis les années 2000, l'augmentation des prix des matières premières a intensifié l'activité minière industrielle, en particulier en Afrique qui fait face à un boom minier attirant de nombreux investisseurs étrangers. Les personnes vivant à proximité des mines industrielles sont exposées à de fortes concentrations de métaux lourds libérés lors des processus d'extraction du minerai. Séparer les minéraux d'importants volumes de roches génère des déchets oxydés, qui, stockés dans des bassins de rétention, peuvent fuir dans l'environnement proche et contaminer les ressources en eau. Cet article utilise un ensemble de données unique sur la localisation et la temporalité de l'activité minière industrielle en Afrique, grâce à un travail manuel intensif afin de récupérer les dates d'ouverture. Notre estimation combine des informations géocodées sur 2 016 mines avec les micro-données d'enquêtes Demographic Health Survey (DHS), qui fournissent des informations de santé au niveau village de 1986 à 2018 pour 26 pays Africains. Utilisant une stratégie de différence dans la différence avec date de traitement variable, nous exploitons la variation dans la date d'ouverture des mines et la position topographique relative des villages DHS environnants. La pollution de l'eau induite est examinée indirectement, car nous comparons les effets sur la santé des villages en aval à ceux en amont du site minier, avant et après son ouverture. Les résultats montrent que vivre en aval d'une mine qui a ouvert augmente la probabilité de mourir dans les 24 mois de 25%, en comparaison aux enfants vivant en amont. Nous constatons une augmentation de la mortalité à 12 mois uniquement pour les enfants non exclusivement allaités et ayant consommé de l'eau, ce qui corrobore le mécanisme de pollution de l'eau. L'effet sur la mortalité est plus élevé dans les zones à forte densité minière, s'estompe avec la distance et se produit principalement pendant la durée d'activité de la mine. L'analyse d'hétérogénéité suggère que l'augmentation de la mortalité est principalement due à la pollution des mines à ciel ouvert, ce qui est cohérent avec un processus d'extraction intensive, et aux mines détenues uniquement par des compagnies étrangères. Pour exclure d'autres mécanismes potentiels, nous montrons que l'effet n'est pas lié par un changement dans la fécondité des femmes, ni à l'amélioration des infrastructures publiques et privées, ou à l'immigration de la main-d'œuvre.

Le **troisième chapitre** est intitulé *Impacts des sécheresses répétitives et rôle clé de l'expérience : témoignages du Nigeria*. Il s'interroge sur le rôle que joue l'expérience dans la capacité d'adaptation des ménages face aux sécheresses dans le cas du Nigeria. Dans les années 1980, à la fois le Sahel et le Golfe de Guinée ont été frappés par des sécheresses sévères et durables, qui ont affecté les ménages ruraux, forcés de trouver des stratégies d'adaptation. Cet article utilise des données satellitaires sur les précipitations journalières à fine résolution spatiale pour analyser les tendances à long terme des précipitations et leur interaction avec les événements climatiques plus récents. Je réfute la théorie de retour à des pluies normales au Sahel, observant des précipitations moins fréquentes et plus intenses sur le long-terme, ce qui augmente la probabilité d'événements extrêmes dans les années récentes. Le centre et le sud du pays, dominés par des systèmes agricoles, ont été durement touchés par les sécheresses de 2013 et 2015. Pour les données agricoles, j'utilise trois vagues de panel d'enquêtes auprès des ménages de 2010 à 2016. Grâce à une stratégie à doubles effets fixes, j'exploite la variation de l'exposition aux sécheresses récentes et montre que les sécheresses diminuent la production agricole de 14% et la diversité alimentaire des ménages de 1%. Un exercice d'hétérogénéité suggère que les effets sont atténués pour les individus qui étaient en charge de la même parcelle lors des sécheresse intense des années 1980. Ceci suggère, de manière non causale, que le fait d'avoir une expérience de longue durée du changement climatique réduit la vulnérabilité aux chocs de pluie récents.

# Table of Contents

# General Introduction

### Overview

Climate change and water pollution are two major environmental issues.

Anthropogenic climate change and its adverse impacts on ecosystems are now unequivocal. The changes in the climate system, especially intensive extreme events, have affected natural and human systems unequally around the world (IPCC, 2022a), contributing to the loss and degradation of ecosystems, reduced water and food security, and damaging livelihoods. The relationship between climate and economic outcomes is multifaceted. Climate events impact income per capita and growth rates (Dell, Jones, and Olken, 2012; Hsiang, 2010), labor productivity and labor supply, consumption (Dercon, 2004), health and mortality (Deschenes, Greenstone, and Guryan, 2009), civil conflict (Miguel, Satyanath, and Sergenti, 2004; M. Burke et al., 2009) and political stability (P. Burke and Leigh, 2010) through a variety of channels. In particular, climate change has adverse impacts on agricultural productivity and farm income (Fisher et al., 2012; Schlenker and Lobell, 2010; Feng, Krueger, and Oppenheimer, 2010; Barrios, Ouattara, and Strobl, 2008; Dell, Jones, and Olken, 2014).

These adverse effects are unevenly distributed across systems and regions. While Sub-Saharan Africa has contributed negligibly to anthropogenic change (IPCC, 2022b), it is one of the most vulnerable regions to climate change. African economies are highly dependent on the agricultural sector and rain-fed systems, which is combined with low levels of socioeconomic growth. Adaptation to climate change defined by the (IPCC, 2007) as *the adjustment in natural or human systems in response to actual or expected climatic stimuli or their effects* plays a key role in reducing risks and vulnerability from climate change in developing countries, vulnerability being defined as *the degree to which a system is susceptible to, or unable to cope with, adverse effects of climate change, including climate variability and extremes.*

Agricultural households can rely on on-farm adaptation strategies, such as crop diversification (Veljanoska, 2018; Komowski et al., 2015), technology adoptions, water management (Fishman, 2013; Schlenker, Hanemann, and Fisher, 2005) or changes in crop choices (Kurukulasuriya and Mendelsohn, 2008) and planting dates. The off-farm strategy mainly consists of access to credit and insurance products and income diversification. Migration is also an adaptation strategy, being a substitute or a complement to on-farm strategies.

Climate change scenarios predict that changes in climate patterns will strongly affect agricultural activities in Africa and lead to large migration flows within and out of high-risk countries (IPCC, 2014; Missirian and Schlenker, 2017). In developed countries, these projections raise concern about the increase in asylum applications, which are translated into alarming newspaper headlines. However, the economic literature has moderated these large figures, showing that in the poorest economies, climate change has reduced international migration by affecting individuals' credit constraints and ability to finance a costly migration (Beine and Parsons, 2015; Cattaneo and Peri, 2016). Evidence points towards local adaptation and within-country migration (Dallmann and Millock, 2017; Sedova and Kalkuhl, 2020; Strobl and Valfort, 2015; Gray and Mueller, 2012) implying significant changes in local demographic patterns (McGuirk and Nunn, 2020). Understanding local adaptation and vulnerability to past climate events is key to reducing future risks.

Climate change occurs and interacts with other significant societal changes. Rapid technological change has led to an increased demand for minerals leading to soaring commodity prices and foreign investments in mining activities. The geologically richly endowed continent of Africa is particularly affected and suffers from a natural resource curse as industrial mining sites compete with the local water resource demand and contaminate the environment. In particular, industrial wastes and toxic metals, such as arsenic, copper, mercury, and nickel, leak into the local environment and contaminate water resources.

Water pollution is a major environmental issue in Africa, where universal access to improved drinking water and sanitation is still a key challenge, especially in water-stressed areas. Polluted water associated with poor sanitation increases the transmission of infectious diseases. Anthropic pollution of trusted water resources is also a major leading cause of child mortality in Africa.

**Thesis structure and contributions**

This dissertation is organized around three chapters that analyze the adverse impacts of climate events and industrial-induced water pollution in Sub-Saharan Africa, using a variety of data sources and methodologies. Chapter 1 and Chapter 3 focus on local adaptation to climate shocks in various settings and intend to capture the heterogeneity of the different strategies. In the first chapter, I look at the effects of the recent increase in drought frequency on internal migration in Kenya. In particular, I investigate the profile of migrants across several socioeconomic characteristics to understand how migration occurs. An open question that arises from this paper is whether migrants anticipate climate change and chose to adapt by migration. The third chapter displays descriptive insights related to this question, looking at on-farm outcomes. It investigates the impacts of short-term droughts on agricultural production and food security in Nigeria and attempts to identify whether the past experience of climate events plays a role in reducing household vulnerability in recent years. Eventually, the second chapter, co-authored with Irène Hu, focuses on water pollution and looks at the effects of mining-induced water pollution on child mortality in Sub-Saharan Africa.

In order to tackle these questions, this thesis uses a variety of data sources, from exhaustive censuses to panel surveys and satellite-based products. It takes advantage of microdata with a high spatial resolution to capture small-magnitude effects and look at heterogeneous effects. In the first chapter, I use three exhaustive administrative census data over twenty years, providing information at the individual level across several dimensions, such as gender, age, main economic activity, and educational level. From this data, I create a unique panel of sublocations from intensive hand-work, which is matched with high-resolution satellite-based data on daily rainfall, daily temperature, and land cover. In the second paper, we use a panel of industrial mining sites from a privately owned and on-license database. We build a unique dataset on the location and timing of sites as we retrieved by hand the opening dates from companies' activity reports and satellite images. We combine this mining database with cross-repeated surveys on health outcomes across 26 African countries over thirty years. Finally, the third paper combines three waves of panel household surveys focusing on agriculture to high-resolution satellite-based data on daily rainfall.

Another essential feature of this thesis is the wide geographical scope and the different scales of analysis. Both chapters 1 and 3 are country-level case studies. Chapter 1 focuses on Kenya and investigates climate patterns associated with the El Nino phenomenon. Chapter 2 looks at Nigeria and studies the evolution of rainfall to explore the lasting effects of the severe droughts of the 1980s. Chapter 2 has more external validity and covers half of the African continent to look at the effects of mining-induced water pollution.

The different chapters also exploit several analytical tools. First, they establish causal inferences mainly using different types of difference-in-differences (DiD) approaches. In the first chapter, I use a two-way fixed effect estimation, exploiting the spatial variation in the increase of droughts' frequency since 2000 to capture induced migration. Chapter 2 adopts a staggered DiD strategy exploiting the variation in the opening of a mine and the relative topographic position of villages (upstream or downstream). Both strategies from Chapters 1 and 2 are based on DiD strategy with heterogeneous treatment effects and test for recent development in the econometric literature (de Chaisemartin and d'Haultfœuille, 2020). Chapter 3 relies on a two-way fixed effect estimation as well, with binary and homogenous treatment, as I exploit the spatial variation in the exposition to a particular drought. Finally, all chapters rely on spatial econometrics to investigate the impacts and characteristics of rainfall evolution (Chapters 1 and 3) or to look at effects according to topographic position (Chapter 2).

### Contributions by Chapter

**Chapter 1**, *Droughts, Migration and Population in Kenya*, studies internal migration induced by the increase in the frequency of droughts in recent decades. It contributes to the literature as it improves the understanding of the heterogeneity of migration responses across several dimensions. The paper takes advantage of exhaustive censuses giving detailed information at the individual level and high spatial resolution to capture small-size effects that have so far been overlooked in the literature.

Since 2000, Kenya has faced erratic rainfall and an increase in extreme events. As the majority of the labor force relies on rain-fed activities, including both agriculture and pastoralism, households are highly vulnerable to repetitive droughts. The literature has identified migration as a possible strategy for adaptation in rural settings (Joseph and Wodon, 2013; Dallmann and Millock, 2017; Bertoli et al., 2021). It can take a full range of forms, from rural-urban (Henderson, Storeygard, and Deichmann, 2017; Sedova and Kalkuhl, 2020) to rural-rural mobility (McGuirk and Nunn, 2020), from temporal (Gray and Mueller, 2012) to permanent reallocation. It can be a deliberate choice or an option of last resort. However, few papers in the economic literature capture rural-rural migration. There is a need to estimate the multi-dimensional heterogeneity of migration and distinguish the different types of migration according to the local context.

Using high-resolution satellite-based products on daily rainfall data (CHIRPS), I first document rainfall patterns over 1983-2013. I observe a significant decrease in the length of the agricultural period associated with an increase in drought frequency. I use three waves of exhaustive administrative censuses to look at demographic outcomes at the sublocation level. I look at the changes in decadal population growth rate (DPGR) between two periods, 1989-1999 and 1999-2009 on a panel of 2516 sublocations. Using a two-period difference-in-difference strategy, I compare the demographic growth of sublocations according to the number of dry rainy-season over each period. The increase in drought frequency since 2000 is captured by the number of dry rainy seasons over 1999-2009. Migration is proxy by the DPGR, and the heterogeneity analysis rules out any effect on fertility, infant, and old age mortality.

The results show that exposition to drought increases out-migration, mainly from rural sublocations where pastoralism is the dominant livelihood. The results suggest herders out-migrate with their entire household towards agriculture-oriented rural areas with more favorable rainfall conditions. Migration out of affected agricultural areas displays more heterogeneity. The results suggest that the out-migration of farmers is driven by educated individuals in the working age, while uneducated individuals are trapped in dry sublocations. This form of migration is in line with a strategy of income diversification.

Overall, this paper underlines the importance of looking at the heterogeneity of migration to capture the effects of climate shocks on livelihoods. Hence, policies aiming at reducing household vulnerability to climate change should take into account these different forms of migrations.

**Chapter 2**, *MiningLeaks : Water Pollution and Child Mortality in Africa*, looks at the adverse impacts of water pollution on local populations' health. Its main contribution is to nuance the results from the literature, which finds beneficial effects of industrial mining on health, and to give large-scale and indirect evidence of the effects of water pollution.

Since 2000, Africa is facing a mining boom. The health-wealth trade-off of industrial mining activity has been intensively studied in the literature (Mamo, Bhattacharyya, and Moradi, 2019; Aragón and Rud, 2016; Dietler et al., 2021; Berman et al., 2017; Benshaul-Tolonen, 2018). However, few papers focus on the negative externalities that mining activity might create on the environment (Bialetti et al., 2018; Von der Goltz and Barnwal, 2019). In particular, not much has been analyzed and quantified in terms of water pollution, and there is no order of magnitude on the number of individuals it might affect at a continent-level scale. Yet, the ore extraction processes release toxic metals and pollutants prone to contaminate surrounding water sources, including arsenic, cadmium, copper, lead, mercury, and nickel.

In this paper, we build a novel database on the location and opening dates of an African industrial mining site from the SNL Mining and Metals database. We retrieved information on the start of production by hand from companies' activity reports and satellite images. We match 2016 industrial mines to the Demographic Health Survey (DHS) over 26 African countries. We conduct a staggered difference-in-difference strategy, comparing the health outcomes of individuals living upstream to those living downstream of the mine, before and after its opening. We indirectly isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison.

Results show that living downstream of a mine that has opened increases the likelihood of dying before 24 months by 25 %. The 12-month mortality rate increases only for children non-exclusively breastfed and who consume plain water, which corroborates the fact that the results are driven by water pollution. The adverse effect is particularly severe during mining activity, production peaks, and in the vicinity of mines, as it fades out with distance. Heterogeneity analysis shows that adverse effects are higher for open-pit mines, in line with an intensive extractive process, and foreign-owned only mines. We show that neither changes in women's fertility nor, increased access to better infrastructure nor migration flows drive the results.

In this paper, by creating a unique dataset and implementing a topographic comparison, we shed light on a natural resource curse. We give insights on the order of magnitude of mining-induced water pollution on child mortality at the scale of the African continent.

Can the experience of past dry events reduce the vulnerability of households to short-term shocks? **Chapter 3**, *Impacts of repetitive droughts and the key rôle of experience: evidence from Nigeria* investigates the vulnerability of agricultural households to exposure to climate shocks according to the experience of long-lasting events.

The adverse impacts of climate shocks on agricultural outcomes have been well identified in the literature (Schlenker, Hanemann, and Fisher, 2005; Blanc and Schlenker, 2017; Dell, Jones, and Olken, 2012; Deschênes and Greenstone, 2007; Schlenker and Lobell, 2010). However, few papers have tried to distinguish the different impacts of short-term from long-term exposition under adaptation behaviors. If some studies have looked at the link between perception and adaptation (Maddison, 2007; Tambo and Abdoulaye, 2013; Silvestri et al., 2012; Komowski et al., 2015), adaptation might happen under collective behaviors rather than only individual perceptions.

Western Africa was hit by severe and long-lasting droughts at the beginning of the 1980s. This paper attempts to assess whether exposure to these past dry conditions might affect recent responses to rainfall shortages in the case of Nigeria. I combine three waves of panel surveys on agricultural households from the GHS, to a satellite-based product on daily rainfall CHIRPS. I investigate the evolution of rainfall in the long run and give evidence of the dramatic fall in rainfall in the 1980s. In the long run, I observe less frequent and shorter rainy seasons, which is linked to the increase in droughts in recent decades. First, I look at the effects of the 2015 drought on agricultural production and food security. Then, I exploit the variation in the timing of the year of land acquisition to see whether individuals managing the same plot as during the 1980s droughts are differently impacted by short-term drought. As this raises endogeneity concerns, this analysis is only a piece of descriptive evidence.

First, results show that a short-term drought increases household vulnerability, as it decreases yields by 14% and food diversity by 1 % . Then, the heterogeneity analysis suggests that the adverse effects are attenuated for individuals in charge of the same plot as in the early 1980s, especially for those who were severely affected.

Overall, this paper displays a piece of descriptive evidence that having a long-lasting experience under dry events might reduce vulnerability to rainfall shocks. As I can not rule out endogeneity issues and self-selection occurring after the 1980s droughts, the causal estimation is an open question for future research.

# Bibliography


Aragón, Fernando and Juan Pablo Rud (2016). “Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana”. In: *The Economic Journal* 126.597, pp. 1980–2011.

Barrios, Salvador, Bazoumana Ouattara, and Eric Strobl (2008). “The impact of climate change on agricultural production : Is it different for Africa ?” In: *Food Policy* 33.4, pp. 287–298.

Beine, Michel and Christopher Parsons (2015). “Climatic Factors as Determinants of International Migration”. In: *The Scandinavian Journal of Economics* 117.2, pp. 723–767.

Benshaul-Tolonen, Anja (Aug. 2018). “Local Industrial Shocks and Infant Mortality”. In: *The Economic Journal* 129.620, pp. 1561–1592.

Berman, Nicolas et al. (June 2017). “This Mine Is Mine! How Minerals Fuel Conflicts in Africa”. In: *American Economic Review* 107.6, pp. 1564–1610.

Bertoli, Simone et al. (Nov. 2021). “Weather shocks and migration intentions in Western Africa: insights from a multilevel analysis”. In: *Journal of Economic Geography* 22.2, pp. 289–323.

Bialetti, Anca et al. (2018). “Lease Splitting and Dirty Entrants: The Unintended Deforestation Consequences of India’s Environmental Clearance Process Reform”. In: *PEDL Research Papers*.

Blanc, Elodie and Wolfram Schlenker (2017). “The Use of Panel Models in Assessments of Climate Impacts on Agriculture”. In: *Review of Environmental Economics and Policy* 11.2, pp. 258–279.

Burke, Marshall et al. (2009). “Warming Increases the Risk of Civil War in Africa”. In: *Pro-ceedings of the National Academy of Sciences* 106.49, pp. 20670–74.

Burke, Paul and Andrew Leigh (Oct. 2010). “Do Output Contractions Trigger Democratic Change?” In: *American Economic Journal: Macroeconomics* 2.4, pp. 124–57.

Cattaneo, Cristina and Giovanni Peri (2016). “The migration response to increasing temperatures”. In: *Journal of Development Economics* 122.Supplement C, pp. 127–146.

Dallmann, Ingrid and Katrin Millock (Sept. 2017). “Climate Variability and Inter-State Migration in India”. In: *CESifo Economic Studies* 63.4, pp. 560–594.

de Chaisemartin, Clément and Xavier d’Haultfœuille (Sept. 2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. In: *American Economic Review* 110.9, pp. 2964–96.

Dell, Melissa, Benjamin Jones, and Benjamin Olken (July 2012). “Temperature Shocks and Economic Growth: Evidence from the Last Half Century”. In: *American Economic Journal: Macroeconomics* 4.3, pp. 66–95.

– (2014). “What Do We Learn from the Weather? The New Climate–Economy Literature”. In: *Journal of Economic Literature* 52.3, pp. 740–798.

Dercon, Stefan (2004). “Growth and shocks: evidence from rural Ethiopia”. In: *Journal of Development Economics* 74.2, pp. 309–329.

Deschenes, Olivier, Michael Greenstone, and Jonathan Guryan (May 2009). “Climate Change and Birth Weight”. In: *American Economic Review* 99.2, pp. 211–17.

Deschênes, Olivier and Michael Greenstone (Mar. 2007). “The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather”. In: *American Economic Review* 97.1, pp. 354–385.

Dietler, Dominik et al. (2021). “Impact of mining projects on water and sanitation infrastructures and associated child health outcomes: a multi-country analysis of Demographic and Health Surveys (DHS) in Sub-Saharan Africa”. In: *Global Health.*

Feng, Shuaizhang, Alan Krueger, and Michael Oppenheimer (2010). “Linkages among climate change, crop yields and Mexico–US cross-border migration”. In: *Proceedings of the National Academy of Sciences* 107.32, pp. 14257–14262.

Fisher, Anthony et al. (2012). “The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment”. In: *The American Economic Review* 102.7, pp. 3749–3760.

Fishman, Ram (Dec. 2013). “Rainfall Variability, Adaptation through Irrigation, and Sustainable Management of Water Resources in India”. In: *AGU Fall Meeting Abstracts*, GC32A-07, GC32A–07.

Gray, Clark and Valerie Mueller (2012). “Natural Disasters and Population Mobility in Bangladesh”. In: *Proceedings of the National Academy of Sciences* 109.16, pp. 6000–6005.

Henderson, Vernon, Adam Storeygard, and Uwe Deichmann (2017). "Has climate change driven urbanization in Africa?" In: *Journal of Development Economics* 124.C, pp. 60–82.

Hsiang, Solomon (Oct. 2010). "Temperatures and Cyclones Strongly Associated with Economic Pro- duction in the Caribbean and Central America". In: *Proceedings of the National Academy of Sciences* 107.35, pp. 15367–72.

IPCC (2007). *Climate Change 2007: Impacts, Adaptation and Vulnerability.* Cambridge Univ Press, Cambridge, UK.

– (2014). "Fifth Assessment Report : Impacts, Adaptations, and Vulnerabillity". In: *UNEP, Washington DC.*

– (2022a). *Climate Change 2022: Impacts, Adaptation, and Vulnerability.* Tech. rep. Cambridge University Press.

– (2022b). *Climate Change 2022: Mitigation of Climate Change.* Tech. rep. Cambridge University Press.

Joseph, George and Quentin Wodon (2013). "Is Internal Migration in Yemen Driven by Climate or Socio-economic Factors?" In: *Review of International Economics* 21.2, pp. 295–310.

Komowski, Frédéric et al. (2015). *Les sociétés rurales face aux changements climatiques en Afrique de l'Ouest. Innover en milieu rural Ouest-Africain : Quels changements dans les pratiques agricoles des exploitants ?*

Kurukulasuriya, Pradeep and Robert Mendelsohn (Mar. 2008). "Crop switching as a strategy for adapting to climate change". In: *African Journal of Agricultural and Resource Economics* 2.1, pp. 1–22.

Maddison, David (Jan. 2007). "The Perception of and Adaptation to Climate Change in Africa". In: *The World Bank, Policy Research Working Paper Series.*

Mamo, Nemera, Sambit Bhattacharyya, and Alexander Moradi (2019). "Intensive and extensive margins of mining and development: Evidence from Sub-Saharan Africa". In: *Journal of Development Economics* 139, pp. 28–49.

McGuirk, Eoin and Nathan Nunn (Dec. 2020). *Transhumant Pastoralism, Climate Change, and Conflict in Africa.* NBER Working Papers 28243. National Bureau of Economic Research, Inc.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti (2004). "Economic Shocks and Civil Conflict: An Instrumental Variables Approach". In: *Journal of Political Economy* 112.4, pp. 725–753.

Missirian, Anouch and Wolfram Schlenker (2017). "Asylum applications respond to temperature fluctuations". In: *Science* 358.6370.

Schlenker, Wolfram, Michael Hanemann, and Anthony Fisher (Mar. 2005). “Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach”. In: *American Economic Review* 95.1, pp. 395–406.

Schlenker, Wolfram and David Lobell (2010). “Robust negative impacts of climate change on African agriculture”. In: *Environmental Research Letters* 5.1, p. 014010.

Sedova, Barbora and Matthias Kalkuhl (2020). “Who are the climate migrants and where do they go? Evidence from rural India”. In: *World Development* 129.C.

Silvestri, Silvia et al. (Dec. 2012). “Climate change perception and adaptation of agro-pastoral communities in Kenya”. In: *Regional Environmental Change* 12.

Strobl, Eric and Marie-Anne Valfort (2015). “The Effect of Weather-Induced Internal Migration on Local Labor Markets. Evidence from Uganda”. In: *World Bank Economic Review* 29.2, pp. 385–412.

Tambo, Justice and Tahirou Abdoulaye (2013). “Smallholder farmers’ perceptions of and adaptations to climate change in the Nigerian savanna”. In: *Reg Environ Chang* 13.375.

Veljanoska, Stefanija (2018). “Can Land Fragmentation Reduce the Exposure of Rural Households to Weather Variability?” In: *Ecological Economics* 154, pp. 42–51.

Von der Goltz, Jan and Prabhat Barnwal (2019). “Mines: The local wealth and health effects of mineral mining in developing countries”. In: *Journal of Development Economics* 139, pp. 1–16.

# Chapter 1

# Droughts, Migration and Population in Kenya


## Abstract

This paper studies the effects of the increase in droughts occurrence on internal migration in Kenya. I build a panel of 2,518 sublocations using three exhaustive censuses in 1989, 1999, and 2009, matched with high-resolution rainfall data (CHIRPS). Using a two-way fixed effect model, I compare the demographic growth of sublocations according to the number of dry-rainy seasons over each decade. An additional drought decreases the demographic growth rate by 1.7 p.p, equivalent to a 1% population loss. The result holds for the [15; 65] years old cohort, which rules out other demographic effects and shows that the result is driven by migration. Migration occurs mainly in rural areas dominated by pastoralist activities. The main contribution is the heterogeneity analysis across gender, age, educational level, and economic activity, which identifies different forms of migration across livelihoods. It suggests a rural-rural migration of entire households of herders with little heterogeneity, in line with migration being a last resort solution. Agriculture-oriented rural areas are less vulnerable to droughts and display significant heterogeneity. Results show the migration of the most educated individuals in the working age, while non-educated individuals are trapped in affected areas. This paper is in line with rural-rural migration, as results suggest out-migration from rural pastoralist to rural agriculture-oriented areas with humid conditions. The results are robust to using binary treatment, correcting for negative weights, accounting for spurious correlation, and to randomization inference tests.

## 1.1 Introduction

Over the past decades, East Africa has faced repetitive climate extremes, changes in temperature averages, and precipitation trends (Gebrechorkos, Hülsmann, and Bernhofer, 2019), driving significant modifications of human behaviors and demographic patterns. The effects of climate shocks are unevenly distributed, affecting the most vulnerable farmers in the poorest countries, who lack adaptation resources. In particular, climate is a major determinant of agricultural production in Sub-Saharan Africa. The economies are highly dependent on the agricultural sector that encompasses most of the labor force, and the soils suffer from aridity and are vulnerable to rainfall shortages.

The relationship between climate and economic outcomes is multifaceted in developing countries, and adaptation strategies exist both on and off-farm. Households dependent on agricultural income can adapt through land fragmentation and crop diversification (Veljanoska, 2018), through changes in crop choices, in livestock species, or technology adoptions such as irrigation (Fishman, 2013). Off-farm adaptation takes place through income diversification by seeking employment in the non-agricultural sector, or through informal credit and insurance products, which have low take-up in developing countries (Cole et al., 2013; Banerjee et al., 2015).

Migration is another possible strategy for adaptation to climate events. It can be a substitute for on-farm adaptation, but can also occur in addition to on-farm strategies, as remittances relax local liquidity constraints and boost local adaptation and innovation adoptions. Migration might also occur when local strategies have failed and be the strategy of last resort. Eventually, it is a costly option that might be impossible for liquidity-constrained households. Migration can take a full range of forms, from rural-urban to rural-rural mobility, from temporal to permanent reallocation, and can be a choice, or an option of last resort.

The majority of the Kenyan labor force relies on agricultural and pastoralist activities as their main livelihoods, making them all the more vulnerable to climate shocks. The lack of infrastructure and facilities in the country, 2% of the cultivated areas being irrigated, - make the occurrence of rainfall a crucial determinant of crop production and animal husbandry (Bryan et al., 2010).Since 2000, Kenya has faced a decrease in the length of the agricultural period and an increase in extreme events, especially droughts. The country has been hit by major dry shocks, with national coverage in 2000 and with more spatial variation in 2004 and 2007, which shows the increase in the frequency of dry conditions, especially in the center of the country. The increase in erratic rainfall and dry conditions are detrimental to rural households and make adaptation even more necessary.

In this paper, I look at the effects of the increase in drought occurrence on internal migration in Kenya. To analyze rainfall patterns over the 1983-2013 period, I use high-resolution satellite-based rainfall and temperature data, the CHIRPS and CHIRTS product from the Climate Hazard Center (CHC) (Dinku et al., 2018). Socio-economic variables are built using three waves of exhaustive censuses from the Kenya National Bureau of Statistics (KNBS) giving information at the individual level about gender, age, educational level, and main economic activity. This study relies on a panel of 2518 sublocations, the smallest administrative unit in Kenya, from 1989 to 2009. Migration is proxied by changes in the decadal population growth rate (DPGR) over two decadal periods, 1989-1999 and 1999-2009. Restrictions of the DPGR to the [15,65] years old cohort and heterogeneity across age brackets rule out any effects on fertility, as well as old-age and infant mortality. I estimate a two-way fixed effects strategy exploiting the spatial variation in the increase of drought repetition since 2000 across Kenyan sublocations. Using a two-period difference-in-difference strategy with continuous treatment, I compare the demographic growth of sublocations according to the number of dry rainy-season over each decadal period. A heterogeneity analysis is conducted in order to determine the profile of migrants across sublocation types. I build a demographic record of the migration across socio-economic characteristics and look at the heterogeneity of the migration responses according to the type and main livelihood of each sublocation.

The main result of this paper gives evidence of out-migration in response to repetitive droughts in Kenya. An additional dry rainy season over a decade decreases the demographic growth rate by 1.7 percentage points, which corresponds to a 6% reduction of the DPGR compared to normal rainfall conditions. This effect is mainly driven by the out-migration from rural areas, especially those where pastoralism is the main economic activity, showing a critical effect on animal husbandry activities.

Overall, the heterogeneity analysis shows that migrants are mainly young individuals of working age and with minimum education, as non-skilled individuals are trapped in treated sublocations. This heterogeneity is mainly observed within agriculture-oriented rural sublocations, and different migrant profiles are identified depending on the dominant livelihood of the area. Within rural areas dominated by pastoralist activity, the results show little heterogeneity, which is in line with nomadic livelihoods and the displacement of entire households or villages due to rainfall shortages. This suggests that the increase in drought frequency accelerates and intensifies the short-distance and rural-rural migration of herders in Kenya. On the other hand, agriculture-oriented areas display high heterogeneity, as the out-migration is driven by young and skilled individuals, and illiterate individuals are trapped in affected rural areas, which is in line with individual migration as a response to climate shocks. Finally, I also find a structural change in the labor market in treated urban areas, as business owners seem to fall into unemployment.

I find no significant results of the effects of an additional flood during the main agricultural season, which shows the determinant impact of droughts. However, when looking at the intensive margin of extreme events, the results suggest that herder migrants relocate into rural agriculture-oriented areas with favorable rainfall conditions. This suggests that the response to repetitive droughts in Kenya results in rural-rural migration.

The main result is robust in controlling for average temperature and evapotranspiration, and temperature anomalies. It is also robust to the de Chaisemartin and d'Haultfœuille, 2020 estimator, to using a difference-in-difference relying on a binary treatment, testing for spurious correlation such as spatial correlation and spatially dependent trends, to running spatial and temporal randomization inference tests. I address additional threats to the identification by testing for the common trend assumption and by correcting for the contamination of the control group.

The major contributions of this paper are twofold. First, it gives evidence of small-magnitude rural-rural movements as a response to repetitive droughts, mainly driven by households involved in pastoralist systems. The paper relates to the literature using long-term and exhaustive information at the individual level, which tackles the limitations of macro-oriented studies that estimate aggregate flows and neglect intra-country heterogeneity. The results display relatively small magnitude effects, hard to capture at a bigger scale, showing the necessity to use exhaustive and local demographic data to capture internal climate-induced migration. High-resolution samples have an important comparative advantage giving enough power to detect the small magnitude effects found. The second contribution is the heterogeneity analysis which identifies different types of migrant profiles and forms of migration. To my knowledge, it is the first study to build a demographic record of the migratory response to repetitive droughts across four socio-economic characteristics simultaneously at the local scale.

The remainder of the paper is structured as follows. Section 1.2 reviews the literature and presents the contributions. Section 3.3 details the data used in the paper, and Section 2.3.2 the context as well as spatial and temporal changes in precipitation in Kenya. Section 3.4 presents the main empirical strategy. Section 2.5 lays out the main results, while Section 2.7 investigates the heterogeneity. Section 2.9 looks at the intensive margin of the results, and Section 2.10 proposes a list of robustness checks and tests. Section 1.10 concludes.

## 1.2 Literature review and contributions

If migration is a response to climate change increasingly investigated in the literature, there is still debate on the characteristics of the relationship, as the literature highlights heterogeneity across the types of climate-induced adaptation. The assessment of the causal effect of climate on economic outcomes depends on the definition of the event. The response to natural disasters and short-term variations differ from the response to climate variability, to slow-onset events, and to climate change in the longer-run (Dell, Jones, and Olken, 2014; Auffhammer et al., 2013). The induced migration can take several forms as well, from international to internal displacement, from rural-urban to rural-rural resettlement, and from temporary to permanent movements. This paper is in line with the micro-oriented literature, which looks at the effects of slow-onset events and identifies rural-rural movements and heterogeneous migration.

This section first displays the literature on the response to fast-onset events. It then describes the literature on longer-term changes, places this paper within this literature, and gives its contribution.

### 1.2.1 Short-term events

Natural disasters and fast-onset events, such as landslides, hurricanes, or floods, trigger temporary and short-distance migration of the labor force toward cities. Looking at aggregated outcomes at the scale of several developing countries, Beine and Parsons, 2015 show that natural disasters spur migration to neighboring countries and increase internal rural-urban migration, proxied by the rate of urbanization, in both poor and middle-income economies.

This pattern is also observed by looking at case studies, such as in developed countries with the case of the famous American Dust Bowls that occurred in the 1930s in Texas and Kansas. If Hornbeck, 2012 observes short-run population decreases at the county level, interpreted as an-out migration towards California, the paper suffers from omitted variables bias and reported issues. Long and Siu, 2018 refute the exodus towards California, showing that the population decline was a consequence of pre-region characteristics and a fall in the flow of in-migrants in these counties. It finds that land erosion led to intra-county short-distance movements, linked to a decrease in agricultural productivity. Lynham, Noy, and Page, 2017 looks at demographic and economic damages of the Hilo Tsunami of 1960 on Hawaii Island and finds a civilian population decline explained by a decrease in the number of employers and population moving away. Gray and Mueller, 2012 fails to find any effects of flooding in Bangladesh, but shows a strong effect of crop failures associated with droughts on short-distance migration of low-income households. Findley, 1994, uses a small sample of household survey data to analyze the effects of the 1983-1985 drought in Mali and blames famine as a potential driver for migration. The paper finds that even if the global migration rate did not change, there was a massive migration of women and children and a shift to short-cycle circulation. Famines linked to natural disasters are also distress factors leading to temporary rural-urban mobility, followed by return migration to the origin region, as happened in the case of the Irish Great Famine (Gráda and O'Rourke, 1997).

### 1.2.2 Slow-onset events

In this paper, I mainly contribute to two strands of the literature. First, I contribute to the literature that measures the effects of slow-onset events on the magnitude of migration, as I look at the effect of repetitive droughts over twenty years. The literature on climate-induced migration identifies several types of migration. If macroeconomic studies mainly focus on international and urbanization due to the nature of aggregated outcomes, this paper is in line with case studies that are able to distinguish between rural-urban and rural-rural movements.

#### 1.2.2.1 International migration and urbanization rates

First, the macroeconomic-oriented literature looks at the impact of climate trends on international migration using country-level panel data (Özden, Parsons, and Schiff, 2011). It displays mixed evidence of climate-induced migration according to the dependence on the agricultural sector (Cai et al., 2016; Feng, Krueger, and Oppenheimer, 2010), income distribution, and the destination of the migrants. While Reuveny and Moore, 2009 finds that slow environmental degradation plays a significant role in out-migration and Missirian and Schlenker, 2017 in asylum applications towards the EU, Beine and Parsons, 2015 fail to discern a direct impact on international migration mainly due to the income shock. The paper argues that changes in climate patterns do not only impact the incentives to migrate but also reduce wages and consequently the capacity to finance a costly migration. It shows that in the poorest countries, rainfall and temperature deviations weaken the ability to migrate, which is evidence of the poverty trap story (Piguet, Pécoud, and Guchteneire, 2011). The longer-term process of climate change seems to affect individual credit constraints more than their incentives to migrate.

Another strand of macroeconomic studies focuses on urbanization to investigate effects on rural-urban internal migration. Barrios, Bertinelli, and Strobl, 2006 use urbanization rates to assess the long-term rural-urban migration caused by changes in average rainfall for 78 Sub-Saharan countries, compared to the rest of the developing world. They found larger effects of the decrease in precipitation after decolonization, explained by more freedom in legislation to move. Cattaneo and Peri, 2016 display evidence of the poverty trap story by finding more impact of temperature increases on urbanization rates in middle income than in poor economies. Marchiori, Maystadt, and Schumacher, 2012, use theoretical and empirical evidence showing that migration occurs in two steps, first rural-urban and then international. The paper finds that both temperature and rainfall anomalies increase internal and international migrations as it leads to wage gaps, first between rural and urban areas, and then in comparison to international wages, intensified by living condition deterioration within cities due to population inflows. Looking deeper into the rural-urban migration, Henderson, Storeygard, and Deichmann, 2017 compare different types of cities. They find that soil moisture deficiency induces urbanization for cities with the capacity to integrate leaving farmers into the labor force, being cities with manufacturing.

#### 1.2.2.2 Internal movements

However, macro studies use aggregated migration flows and climate patterns, neglect within-country heterogeneity, and miss country-specific responses with smaller-magnitude effects, such as rural-rural displacements. If empirical case studies are fewer due to lack of data, they disentangle migration types and heterogeneity and show that response to slow-onset events is not limited to urbanization (Mueller, Gray, and Hopping, 2020). Overcoming the drawbacks of local and small sample size surveys, they use exhaustive population census data (Joseph and Wodon, 2013; Strobl and Valfort, 2015; Dallmann and Millock, 2017; Long and Siu, 2018), historical archive records (Hornbeck, 2012; Lynham, Noy, and Page, 2017), or panel survey (Sedova and Kalkuhl, 2020) and pulled repeated cross-sections (Bertoli et al., 2021).

The literature displays mixed evidence, showing that climate-induced migration is highly dependent on the context. Dallmann and Millock, 2017 use two Indian censuses (1991 and 2001) and find evidence of inter-district migration due to drought frequency (based on the SPI), which is attenuated for districts with high irrigation rates. Mexican migration to the US is higher in rain-fed agriculture communities which had lower rainfall (Munshi, 2003) and with temperature shocks on crop yields (Feng, Krueger, and Oppenheimer, 2010). Using one census in Yemen Joseph and Wodon, 2013 find that socio-economic and cost factors affect much more migration than climate variability. In the Malian context, Defrance, Delesalle, and Gubert, 2022 use administrative censuses as well and finds evidence of net outflows in response to dry shocks driven from the SPEI Vicente-Serrano, Beguería, and López-Moreno, 2010, which fades in localities with more diversified crops. Strobl and Valfort, 2015 instrument the net-in-migration rate by the weather-predicted determinants to study the effect of climate migration on local labor markets and find negative effects on the employment probability of the non-migrants in the destination areas. Looking at road density, they find higher results for regions less favorable to capital mobility.

#### 1.2.2.3 Heterogeneity and Contributions

This paper also relates to the micro-oriented literature that examines the heterogeneity of climate-induced migration.

First, it looks at the heterogeneity of the response according to sublocation characteristics, such as density. I also distinguish rural areas where agriculture prevails from those where pastoralism prevails, to identify the different types of migration. McGuirk and Nunn, 2020 focuses on pastoralism migration and shows that droughts in western Africa modify the timing of pastoral groups' migration, which triggers conflict with sedentary farmers. This paper is in line with a rural-rural response, as it gives evidence of the out-migration of herders, who seem to move towards rural areas with normal rainfall conditions in the case of Kenya.

Eventually, the main contribution of this paper is to understand precisely who are the migrants. Sedova and Kalkuhl, 2020 investigate the characteristics of the migrants according to their level of schooling and dependence on agriculture using a panel survey in India. The paper corroborates the urbanization response in the case of India, as it shows that weather anomalies push people into faraway cities and more prosperous states but decreases rural-rural movements. The heterogeneity shows that climate migrants are likely from the lower end of the skill distribution and households highly dependent on agricultural production. If my paper is in line with the agricultural channel showing that individuals with the age of working and involved in agricultural activity out-migrate, it shows that it is mainly triggered by the educated population.

This paper contributes to the literature through a multi-dimensional heterogeneity analysis and improves the understanding of the heterogeneous migration responses in different contexts (Cattaneo, Beine, et al., 2019). To my knowledge, it is the first analysis to look at the climate-induced response through so many socio-economic characteristics simultaneously and comprehensively. It takes advantage of exhaustive censuses giving precise information on age, gender, economic activity, and education to capture small-size effects. It helps improve the understanding of heterogeneous migration responses.

## 1.3 Data

This paper matches socio-economic data from exhaustive censuses provided by the KNBS [1] and temperature and rainfall data from the CHIRPS and CHIRTS products of the CHC [2].

### 1.3.1 Climate Data

This paper uses the CHIRPS product from the CHC, which combines a satellite-based rainfall product (CHIRP [3]) with station observations data. It gives a good spatial (0.05 lat/long) and temporal (daily, decadal, and monthly) resolution for historical (1981-2019) mean maximum and minimum precipitations. It has been validated over Eastern Africa and assessed as the best satellite-based product Dinku et al., 2018. For temperature, we use the CHIRTS product, also from the CHC, which also combines satellite and station-based estimates of maximal, minimal, and mean temperature ($T_{max}$, $T_{min}$ and $T_{mean}$), with the same spatial and temporal resolution as CHIRPS (Funk et al., 2019).

### 1.3.2 Population Data

The demographic variables come from three waves of exhaustive censuses conducted each decade in Kenya. Since independence, five complete censuses have been conducted in Kenya, in August 1969, August/September 1979, August/September 1989, August 1999, and August 2009. As the magnetic reels on which the 1969 and 1979 censuses were stored got wet and part of the data were lost, only three censuses can be used in the present study: 1989, 1999, and 2009 [4]. Table A.1 in Section A.1.1 compares the number of observations in each province between the data files and the census reports per Province for the 1989, and 1999 and 2009 censuses. This comparison gives the rate of missing information from the censuses. A discussion in the Appendix Section A.1.1 is made about the quality and reliability of the data, and as justifies the exclusion of the Nyanza and North-Eastern Provinces from the analysis, which is mainly due to data inconsistency in 1989 and 1999.

[1] Kenya National Bureau of Statistics

[2] Climate Hazard Center

[3] Climate Hazards Group Infrared Precipitation

[4] I am highly grateful to Lara Tobin for providing me with the censuses (Tobin, 2017), which were granted by KNBS.

The population universe of the 1989,1999 and 2009 Kenyan censuses is composed of all persons living in the national territory. They are exhaustive, both at the housing and individual levels, giving information about the relationship with the head of the household, the age, the gender, the tribe/nationality (only in 1989), the marital status, the previous economic activity, the years of schooling and the type of the sublocation (whether it is rural or urban area). There are five scales of administrative boundaries in Kenya: Provinces, Districts, Divisions Locations, and Sublocations. Aside from provinces, they all have been reorganized and redrawn over the years. The analysis at the sublocation level has required precise work of matching sublocations over the years, as some changes were geometrically chaotic. Explanation and descriptive statistics of this work can be found in Appendix Section A.1.2.

Precise information about migration is only available at the district levels in the censuses. As districts are large areas, looking at the effects at the scale of districts does not capture local effects and might be biased by omitted variables and concomitance. The main analysis of this paper focuses on intra-district population variation and proxies the migration using population growth outcomes. Net migration rates over each 10 year-period ([1989,1999] and [1999,2009]) are proxied by the Decadal Population Growth Rate (DPGR) at the sublocation level. More formally, the DPGR can be written as follows :

$$DPGR_{i,[t-10,t]} = \frac{\Delta pop_{i,[t-10,t]}}{pop_{t-10}} = \frac{pop_{i,t} - pop_{i,t-10}}{pop_{t-10}}$$

Where i indicates the sublocality and t the last year of each census (1999 for the first census, 2009 for the second one). As the DPGR captures fertility and mortality as well [5], I define the $DPGR_{[15,65]}$, which is the DPGR of the population aged between [15,65] years old at the beginning of the decade and between [25,75] years old, as I follow the cohort. The $DPGR_{[15,65]}$ rules out any effects on fertility, old age, and infant mortality, and shows that the effects are mainly driven by migration. As one goal of this paper is to investigate the heterogeneity of migration, the heterogeneity analysis according to age brackets and vulnerable population groups intends to verify

[5]The DPGR is the sum of the new births, deaths, and the net migration: $DPGR_{i,[t-10,t]} = \frac{\sum_{T=t-10}^{t} Birth_T - \sum_{T=t-10}^{t} Death_T + \sum_{T=t-10}^{t} Inmigration_T - Outmigration_T}{pop_{t-10}}$

as well that the effect is mainly driven by migration.

The censuses make it possible to distinguish sublocations according to their types. I use the 1989 classification given in the administrative census to class sublocations as rural and urban. A main comparative advantage of the 2009 census is that it gives information on whether individuals are involved in farming or pastoralism activity. In this paper, a rural sublocation is defined as highly pastoralist if the total share of the population working in pastoralism in 2009 is amongst the highest, and scarcely pastoralist if it is amongst the lowest [6]. Thus, within rural areas, I create the High Pastoralism and Low Pastoralism classes.

## 1.4 Context and descriptive statistics

### 1.4.1 Setting

#### 1.4.1.1 Climatology

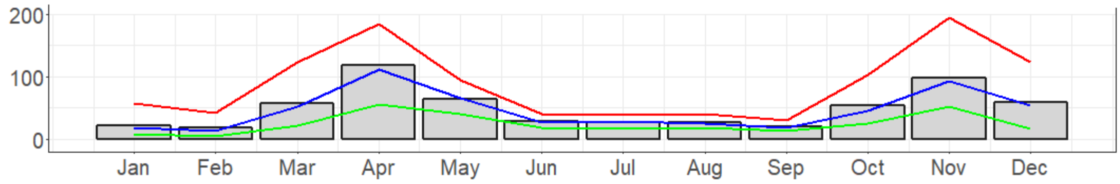
The study focuses on Kenya, ranging from equatorial (West), tropical (East Coast), semi-arid and arid (North), and temperate (inland) climatology. Rainfall patterns are influenced by heterogeneous and multiple local factors, including topography, land surface, monsoon systems, Rift Valley lakes, and large-scale factors such as global forcing mechanisms (El Nino-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD))[7] (Endris et al., 2013; Hoerling et al., 2006). Kenya is part of the eastern Horn of Africa, separated from the rest of the continent by high elevations and the basin of the Turkana Lake.

Lying across the equator, Kenya experiences a bimodal seasonal pattern, as shown in Figure 1.1. The two wet seasons are the long rains and the short rains and are separated by two other seasons with little rainfall. The long rainy season is the primary agricultural season, and extends from March to June (MAMJ), with a peak centered around March/May, and is modulated by local factors rather than global scale ones

[6]It is possible to enumerate the pastoralists in the 2009 census thanks to a question about the type of employer of each working individual. Individuals that are *self pastoralist* and *pastoralist employed* are classified as pastoralists. When looking at the distribution of the share of the working population involved in pastoralism, we consider a sublocation with high pastoralism if the share is above the 60th decile, and with low pastoralism, if it is under the 40th decile.

[7]The El Nino-Southern Oscillation (ENSO) is a recurring climate pattern defined by the change in temperature gradients across the central and eastern tropical Pacific Ocean, while the Indian Ocean Dipole (IOD) across the equatorial Indian Ocean.

Figure 1.1: Long-term average of monthly precipitation




**Notes:** The Figures represent the long-term average of the monthly precipitation (1983-2013) (mm) over Kenya. Red lines plot the 95th percentile of rainfall distribution, blue lines the 50th percentile, and green lines the 5th percentile.
*Sources:* Author's elaboration on CHIRPS data.

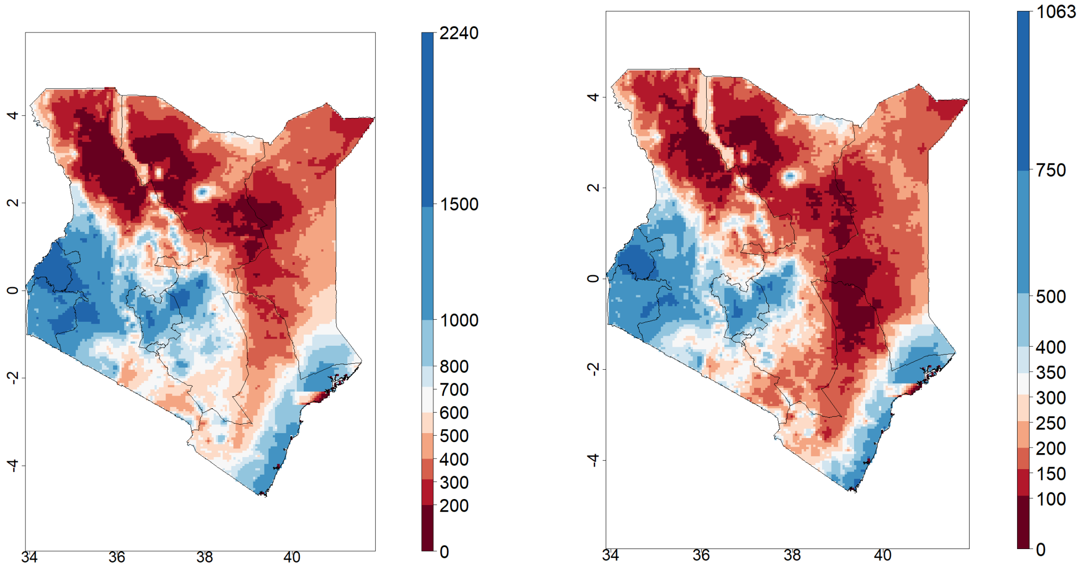
(Omondi et al., 2013). The short-rainy season occurs over October-December (OND). As the short rains are influenced by the ENSO and IOD global factors (Nicholson, 2015; Liebmann et al., 2014), they are less reliable than the MAMJ season and less determinant for agricultural systems. Eventually, January/February (JF) and July/August/September (JAS) are the driest months of the year, with low precipitations and high temperatures. Figure A.2 in Section A.2.1 displays the seasonal patterns across provinces. If some heterogeneity is observed, all provinces show a bi-modal pattern, and I define the MAMJ as the long-rainy season overall in the country.

As local rains are the dominant source of water for Kenyan agriculture (limited groundwater and reservoir storage), the variation and long-term evolution of the long rains MAMJ are a major cause of concern.

Figure 1.2 plots the spatial variation of the long-term average of annual rains (Figure 1.2a) and the long-rainy season (MAMJ) cumulative precipitations (Figure 1.2b), over the [1983-2013] period. In this paper, the long-term period refers to the 30 years [1983-2013]. Annual precipitation is lower than in other parts of equatorial Africa, as the long-term average of annual rainfall is only 593 mm, and only 268 mm over MAMJ [8]. Figure 1.2 shows high heterogeneity, as the highest rainfall amount is registered in the western part of the country (up to 2000 mm per year), while the minimum amount of precipitation is observed in the North East, at the frontier with Ethiopia/Somalia (less than 150mm) Figure 1.2. Figure A.3 in Section A.2.1 plots the long-term mean of rainfall characteristics during the long-rainy season, including the number of wet days (R1plus), the length of the wet and dry spells (CWD and

[8]For instance, in Tanzania, the long-term average of annual precipitations is 976mm over 1981-2016 Gebrechorkos, Hülsmann, and Bernhofer, 2019

Figure 1.2: Spatial variation of rainfall long-term average




(a) Annual precipitation (mm)

(b) MAMJ precipitation (mm)

**Notes:** Figures plot the spatial distribution of the long-term annual average (a) and the long-term average of the long-rainy season of cumulative precipitations over [1983-2013].
*Sources:* Author's elaboration on CHIRPS data .

CDD), and the daily intensity of rainfalls (SDII). It shows that the Center and Northern areas display shorter agricultural seasons, with fewer wet days but more intense daily rains.

#### 1.4.1.2 Land-cover and livelihoods

Beyond the heterogeneity in climate contexts, Kenya is an interesting setting because it has diversified agroecological zones and rural livelihood systems. Figure 1.3a plots the land cover classes, indicates the main urban centers, and shows that Kenya contains mainly pastoral areas. Overall, 27% of the country is made of croplands, 61% of pastures, 9% of bare areas, 2% of waterbodies, and 0.1 % of urban areas. Amongst croplands, the majority are rainfed as only 2% of cultivated areas are equipped for irrigation (Bryan et al., 2010). Within pastoral areas, 10% are forests, 33% grasslands, 7% shrubs, and 50% herbaceous categories [9].

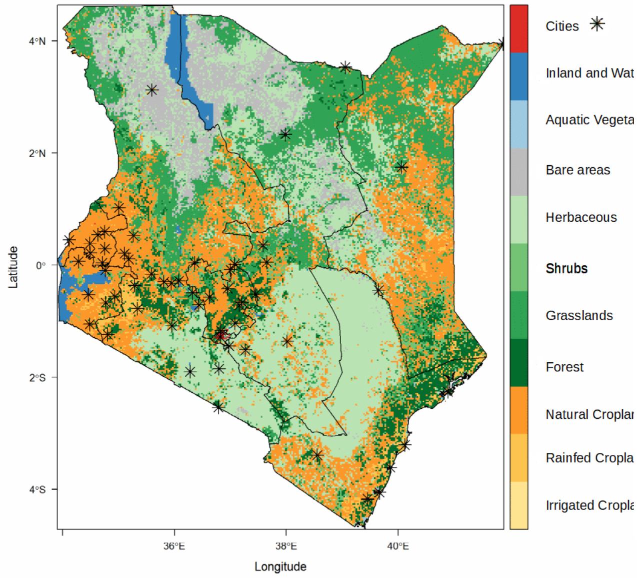
[9] Percentages are computed using the ESA GlobCover data

The main cultivated food crops in Kenya are maize (up to 60% of arable lands), sorghum, and sweet potatoes. The inlands, including the center of the country, the south of the Rift Valley, as well as the region of the Capital Nairobi, and the snow-covered Mount Kenya, are the most important agricultural regions thanks to a tropical savanna climate, less warm than the rest of the country. Tea and coffee are also cultivated, in particular in the center of Kenya, which has an ideal amount of precipitations and volcanic red soils. A cut flower industry has developed in the south of the Rift Valley as well, well-known for its exportation of roses. The West of the country borders Lake Victoria and is divided between Western and Nyanza provinces. Thanks to an equatorial climate, and being the most humid part of the country, it is also mainly agricultural, made of croplands and several major cities. The southeast coastal zone of the country has tropical and humid weather also prone to agriculture, with important food crops being cassava, sweet potatoes, and maize.

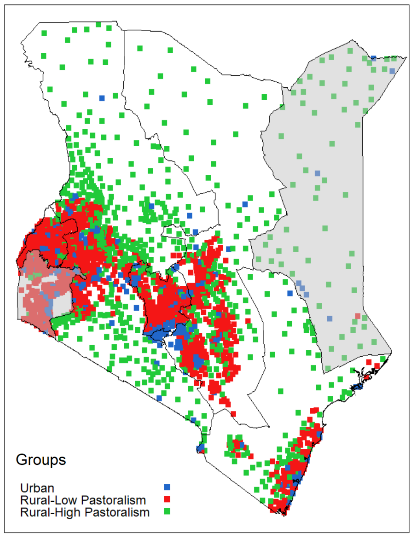
Eventually, 80% of Kenya is covered by arid and semi-arid lands (ASALs), in the northern and eastern parts of the country. Figures 1.2 and 1.3a show that the northwest is the aridest region with desert landscapes, highly hot and dry, while the rest of the North has a warm semi-arid climate. Figures A.2 A.3 in Section A.2.1 show that the ASALs have shorter and less intense agricultural seasons. Within the ASALs, pastoralism is the main source of livelihood, which accounts for 90% of employment and more than 95% of household incomes (Nyariki and Amwata, 2019). It is an economic activity based on livestock production systems, fitting with dryland environments where resources are scarce and unstable. In particular, it is well adapted to generate income in the ASALs despite the instability of precipitations, and despite the fact that water is available over short spans and unpredictable concentrations.

Pastoral systems in Kenya are complex and diverse, and there is not a unique definition of pastoralism in the country (Hesse and MacGregor, 2006). Each specific area is characterized by the varied composition of the herds, and the organization of the economic system, according to its ecological characteristics and available resources. The Maasai pastoralists from the south of the Rift Valley are sedentary and rely on diversified livelihood strategies, not only husbandry. The proximity of riverine areas enables herders to practice more restricted movements. The herds are mainly composed of cattle and sheep, with few camels which reflects the favorable ecological conditions. The drier northern part of the country is characterized by different pastoral systems, mainly nomadic and transhumant pastoralism, as resources are

Figure 1.3: Spatial variation of land cover and main economic activity




(a) Land Cover




(b) Main economic activity

**Notes:** Figure (a) plots the land cover classes across croplands, pastures, bare areas, inlands, and main cities. Figure (b) plots the sublocations types, and classes of rural sublocations according to whether pastoralism is a main economic activity. Each dot corresponds to the sublocation centroid. *Sources:* Author's elaboration on ESA Globcover and KNBS data.

scarce and rainfall unstable (WFP, 2018). To maintain access to water and grazing resources, herders are forced to move regularly (Campbell and Axinn, 1980). The herds are mainly camels and goats, rather than sheep and cattle. If the different pastoral systems display diversified characteristics, most of them rely on herd mobility and migration as a strategy to cope with climate events (Hesse and MacGregor, 2006).

Thus, the majority of the labor force is involved in agricultural activity including both farming and husbandry. Figure 1.3b displays the classification of each sublocation according to their type and the dominant livelihood strategy, using the information provided in the 2009 census. It plots the urban sublocations, and within rural sublocations distinguishes those for which pastoralism is the main economic activity from those where it is not. Figure 1.3 shows the correlation between the dummy indicating pastoralist activity and the presence of grasslands and pastures [10].

[10]The dummy for high pastoralism correlates at $46^{***}\%$ with the presence of grasslands

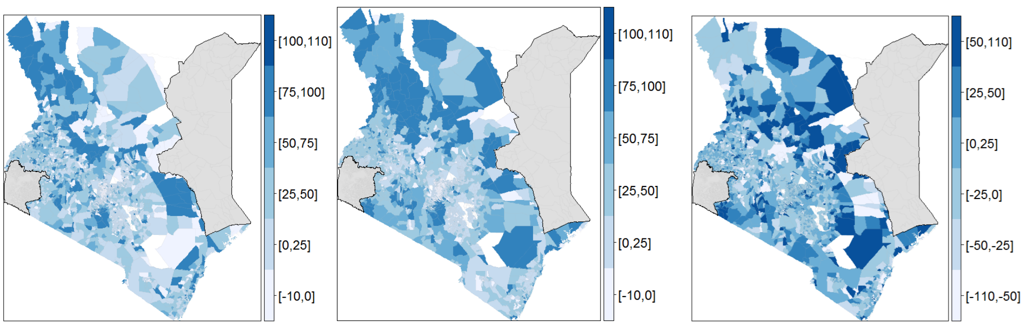
#### 1.4.1.3 Long-term rainfall trends

Section A.2.2 describes the long-term trends of rainfall characteristics, and their statistical significance, over Kenya across several indicators. It shows that the semi-arid and arid regions (ASALS) are facing more erratic rainy seasons, which are becoming shorter with more intense daily rains. In the long run, patterns display significant downward trends in the number of rainy days and the length of wet spells during the long-rainy season, associated with higher intensity of wet days. This shows a decrease in the length of the agricultural period and an increase in extreme events, in a region highly vulnerable because dependent on the agricultural sector. This suggests evidence that the increase in the recurrence of droughts since 2000 in Kenya, which is exploited in this paper, is a consequence of climate change in the long run.

### 1.4.2 Temporal and spatial variation

#### 1.4.2.1 Population and migration

Figure 1.4: Spatial variation of the DPGR across periods




(a) DPGR [1989-1999] (b) DPGR [1999-2009] (c) Long Differences

**Notes:** Figures (a) and (b) plot the spatial distribution of the DPGR over (a) period 1 [1989,1999] and (b) period 2 [1999,2009]. Figure (c) shows the spatial distribution of the long-difference of the DPGR in percentage points.

The Kenyan population is highly dependent on agricultural and livestock income, thus vulnerable to climate variability. Over the 1991-2007 period, 0.72% of the total population was an inter-district migrant, on average (Section A.6). Kenya faces high population growth, and Central and Western provinces are the most populated and dense areas (aside from Nairobi) (Tables A.6 A.7).

Figure 1.4 maps the spatial and temporal variations of the population growth, as it plots the sublocation DPGR over both periods of the analysis, and the DPGR long-difference. It shows increasing trends of the DPGR in the Eastern part of the country, while attenuated and decreasing trends in the center and western areas. Figure1.4 shows spatially clustered population growth trends, which will be discussed in Section 1.9.3. Figure A.8 and A.7 in Section A.2.3 plot spatial patterns of population size and density for each sublocation.

#### 1.4.2.2 Temporal and spatial variation of rainfall

Kenya has a large interannual and intraseasonal variability of total precipitation and extremes (Nicholson, 2015). Extreme events, mainly droughts, and floods, are recurrent, occurring once every three to four years (Herrero et al., 2010) and generally attributed to the ENSO, even though the causes of droughts in Eastern Africa are hardly understood by the climatological literature (Lyon and DeWitt, 2012; Nicholson, 2017).

When looking at longer time scales, long rains amount have decreased in East Africa, mainly due to a recent increasing trend in the sea-surface temperatures (SSTs) in the Indian Ocean. The change in rainfall characteristics over the long run explained in the previous Section 1.4.1.3, shows the decreasing trends of the length of the long-rainy season. This dramatic decline in precipitations since the 1980s is linked with a more abrupt decrease in rainfall during the rainy season since 2000 (Lyon and DeWitt, 2012). During the 1983-2013 period, the most important droughts that Kenya faced occurred in 1983,1993, 1999-2000, 2004-2005, and 2009-2011 (Nicholson, 2015), impacting more and more people (Herrero et al., 2010), especially in the vulnerable ASALs [11]. Since 2000, dry events have alternated with excessive rainfall as well, but have been shown to have less economic impacts than droughts (Mogaka et al., 2006).

Figure A.9 from Section A.2.4 displays the time series of CHIRPS annual precipitation departures [12] aggregated over the country. It shows high interannual variability and shows that droughts were particularly severe in 2000 and 2004, with rainfall decreasing up to 25% in comparison to the long-term mean. Figure A.9 shows evidence of climate variability, displaying the alternation of dry and wet conditions, as the 2000 and 2009 droughts are both preceded by excessive rainfall, up to 60% above the mean.

[11] Arid and semi-arid region

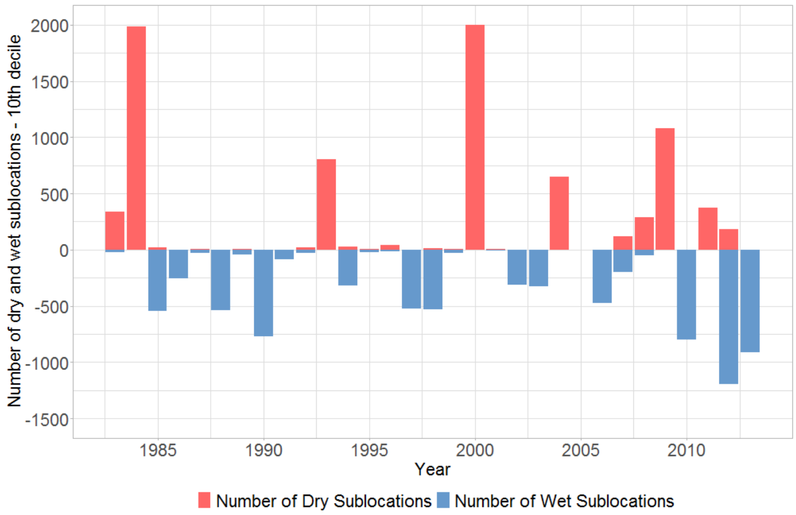
[12] rainfall departures are percentage above or below 1983-2013 long-term mean

Since 2000, droughts have occurred mainly during the boreal summer (rainy season MAMJ) and have become more frequent and severe, longer and more intense, with persistence though several rainy seasons (Nicholson, 2015; Nicholson, 2017). Figure A.10 plots times series of seasonal anomalies, and displays high intraseasonal variability. It shows that the decline in precipitation in Kenya is mainly borne by the decrease during the rainy season MAMJ, while the increase in excessive rains is mainly borne by an increase in wet conditions during the short rainy season OND. As this paper focuses on the effect of drought repetition, the main analysis focuses on the occurrence of rainfall over the long-rainy season MAMJ.

Figure 1.6 shows both the spatial and temporal distribution of rainfall over the 1983-2013 period. For each year, it plots the departure from the long-term mean during the long rainy season across the country. It displays the main droughts and floods, their pattern and persistence across years, as well as the areas the most impacted. Figure 1.6 identifies 1984, 1992-1993, 1999-2001, 2005-2006, and 2010-2011 dry periods and displays their spatial distribution. If droughts are regionally clustered, some impacting most parts of the country, their severity and extent over the rainy seasons differ across Kenya. The 1983-1984 drought's greatest deficits were over the North of the North-Eastern region, the center of Eastern Kenya, and the length of the Rift Valley, as precipitation were $50 - 75\%$ below the long-term mean. The dry period seems to extend over two years, with a complete recovery of the rains in 1984. Figure 1.6 shows the 2000 drought which impacted the majority of the country, with partial recovery in 2001 and the spatial variation of the 2004-2005 and 2007 droughts. As an intra-district variation of the rainfall shortages is observed, it underlines the advantage of doing the main analysis at the sublocation level. Figure 1.6 displays as well the occurrence of excessive rains over MAMJ, which are more distributed over the years and less intense than droughts. Figure A.11 in Section A.2.4 displays the same map for the short-rainy season and shows that, if critical droughts mainly occur over MAMJ, severe floods are mainly born by the OND season.

This paper looks at the effects of the increase in the repetition of droughts during the long-rainy season since 2000 on demographic movements. For each year, I construct a dummy based on the long-term mean of rainfall during MAMJ for each sublocation, which equals 1 for dry rainy seasons, and 0 otherwise. We define a season as dry if the cumulative rains over the MAMJ are lower than the 10th percentile of each

Figure 1.5: Number of dry and wet sublocations across years - 10th decile



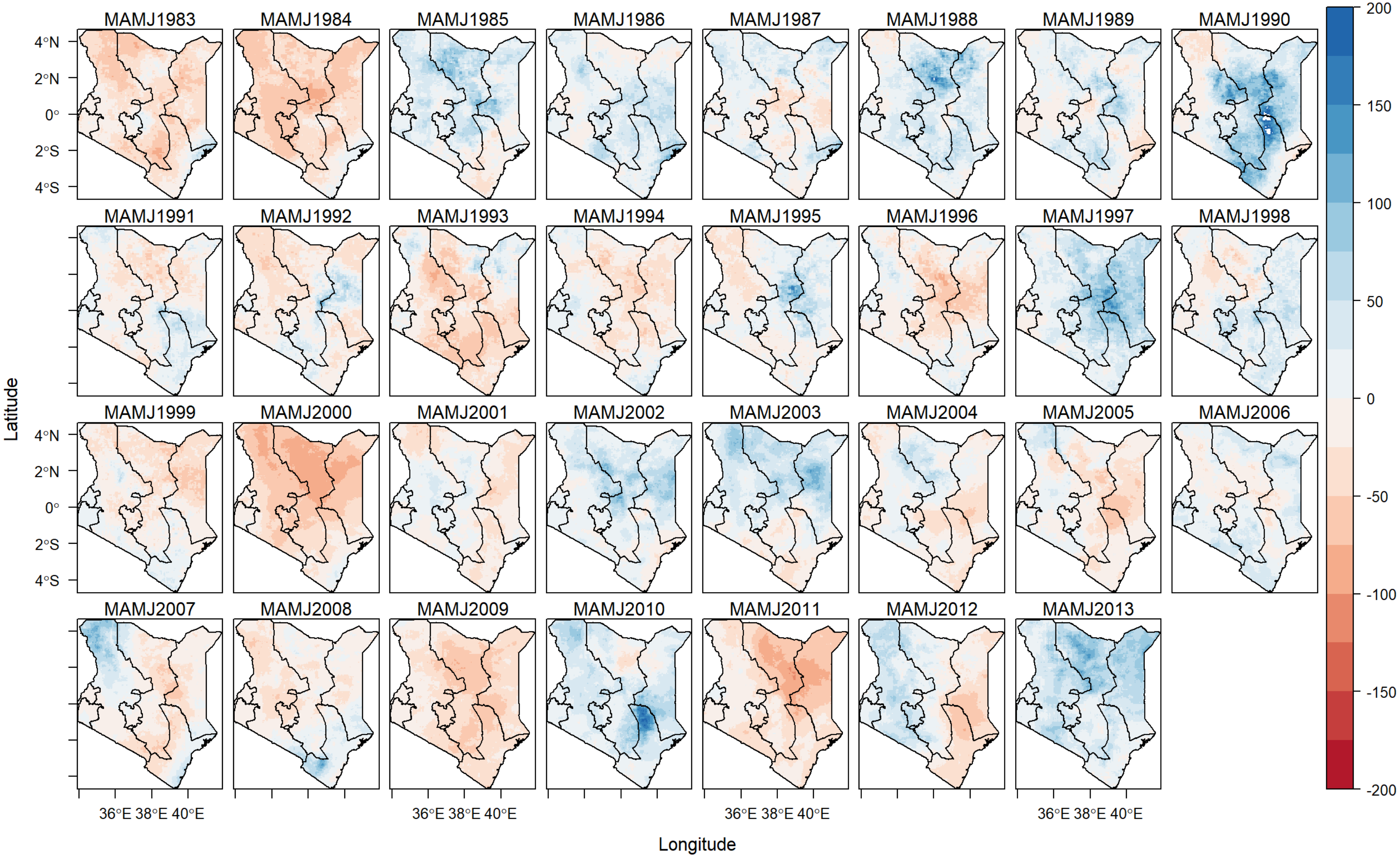

**Notes:** This Figure plots in red the number of sublocations for which the rainy season for a specific year is dry, meaning the cumulative rains are below the 10th decile of the sublocation distribution. Accordingly, it plots in blue the number of sublocations for which the rainy season is wet. For instance, around 1000 sublocations are wet in 1988, while 2000 sublocations are dry in 2000. *Sources:* Author's elaboration on CHIRPS and KNBS data.

sublocation cumulative MAMJ rains over the 1983-2013 period. The independent variable used in the analysis is the number of dry rainy seasons, based on the previous definition, over each decadal period. More formally, the number of dry years is written as follows: $z_{i,t} = \sum_{j=1}^{10} \mathbb{1}_{i,t-j}$, such as:

$$\mathbb{1}_{i,t-j} = \begin{cases} \mathbf{1} & \text{if } \sum_{m=March}^{June} Rain_{i,m,t} < 10^{th} \text{ percentile of } Rain_{i,[1983-2013]} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

Figure 1.5 plots for each year the total number of sublocations for which the dummy $\mathbb{1}_{i,t}$ equals 1. Again, we observe that 2000 and 1984 were national droughts, while 1993, 2004, and 2007-2008 droughts impacted unevenly the country. The Figure shows the increase in the occurrence of droughts since 2000. I exploit the spatial variation of the increase in droughts since 2000 by comparing the effects of the number of droughts over 1989-1999 to the effects of the number of droughts over 1999-2009. If Figure 1.5 shows that 2009 was particularly intense, its effects on Kenyan demography will not be analyzed due to the availability of population data. Figure 1.5 plots also the number of sublocations for the excessively wet years, defined as the cumulative rains being over the 90th percentile of the distribution. As shown in Figure A.10, Figure 1.5 shows that floods are more distributed and less severe over the period, and display no increasing trends of floods born by the MAMJ season.

Figure 1.6: Rainfall percent departures of the long-rainy season (MAMJ) from 1983-2013 mean




**Notes:** The Figure plots the percent departure from the long-term mean of the main rainy season (1983-2013) for each pixel.
*Sources :* Author's elaboration on CHIRPS data.

## 1.5 Empirical strategy

### 1.5.1 Main estimation

The main empirical strategy of this paper exploits the time and spatial variation of rainfall in Kenya, as described in Section 1.4.2.2. It looks at the effects of the number of dry years over a decade on the migration, proxied by the DPGR, for each sublocation over two time periods. I compare the evolution of the DPGR over [1989,1999] and [1999,2009], according to the number of dry rainy seasons per decade. I use a two-way fixed-effect regression, with both sublocation and period fixed effects, on a panel of sublocations. This estimation relies on a Difference-in-Difference (DiD) comparison with heterogeneous treatment, as there are treated sublocations at both periods, with various treatment intensities (see Section 1.9.1 for further discussion). More formally, the empirical strategy can be written as follows :

$$DPGR_{i,t} = \alpha_0 + \alpha_1 dens_{i,t_0} + \alpha_2 z_{i,t} + \alpha_3 dens_{i,t_0} \times z_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t} \tag{1.1}$$

With $z_{i,t} = \sum_{j=1}^{10} \mathbb{1}_{i,t-j}$, the number of years considered as dry over the decade (definition based on the long-rainy season, cf Section 1.4.2.2 ). $\gamma_i$ and $\gamma_t$ are sublocation and time-fixed effects adjusting for spatial and period-specific confounders. $t_0$ indicates the first year of analysis (1989), and $dens_{i,t_0}$ is the baseline density of the sublocation, centered around the median in 1989 (190 $p.km^2$ ) [13]. The independent variable is interacted with the initial density of each sublocation (centered around the median) to investigate the heterogeneity of the effect according to the density distribution and control for the type of sublocation (Figure A.7 maps the density distribution). Using density as a proxy for urbanization levels, this accounts for the first distinction between rural and more urbanized areas.

Equation 3.1 is a two-way fixed effects model, capturing the heterogeneity according to the baseline population density of the sublocation, using interaction terms. The estimator of interest, $\alpha_2$, gives the effect of an additional dry rainy season over the period, for a fictive sublocation of median density.

[13]This value is way above the national mean given by WB numbers, around 100 $p.km^2$. This is because the means and medians here are calculated over sublocations and not for all the country.

### 1.5.2 Identification assumption

The key assumption of a DiD is that the demographic growth of the treated areas would have evolved as the demographic growth of control areas in the absence of repetitive droughts. As I can not test that treated and control sublocations would have followed the same time trends, I test in Section 1.9.2 the common trend assumption using pre-treatment data. A main issue in this test is that the identification relies on two periods only, the [1989-1999] and the [1999-2009], making the pre-treatment data hard to observe. KNBS provided the 1979 census as well, but the census was damaged and part of the data was completely lost [14]. Section 1.9.2 relies on the hypothesis that the damage was random, and uses a sample of the census to show that the common trend assumption holds in this paper.

However, the fact that pre-treatment data are parallel is neither a necessary nor a sufficient condition for the identification.

First, past trends can be identical but the control group may be affected by a group-specific shock during the period of the treatment. Omitted variables bias is exacerbated in the presence of spatial dependency. Thus, other threats to the identification are spurious correlations linked to the spatial dependency of the dependent and independent variables. This is discussed in Section 1.9.3, which shows that the result is robust to correcting for spatial correlation of the shocks, as well as spatially dependent trends.

Second, the contamination of the control group raises concern about the fact that the trends of the control group would be the ones that would have prevailed in absence of dry events in the treated sublocations. The contamination of the control group is directly linked to the nature of the dependent variable. As the results show out-migration from the treated sublocations, individuals migrate somewhere within the control group, which is *de facto* contaminated by the treatment group. Section 1.9.4 discusses this issue and proposes a robustness check to test for this threat.

Another concern is that the paper proxies migration by demographic growth. A way to show that the results are driven by migration effects is to restrict the sample to the [15,65] years old cohort. This rules out any effects from fertility, and mortality of vulnerable groups of the population, which is reinforced by the heterogeneity analysis per age bracket.

[14] magnetic reels of the censuses were stored but got wet and part of the data was lost, including the reels of Nyanza province in 1989 which explains the discrepancies in the data

Eventually, a threat to the heterogeneity analysis would be that economic status and education are endogenous to rainfall shocks. Section 1.7.2 looks at the effects according to the skill distributions of the [20,70] cohort, which means that I look at the demographic growth of individuals aged between [20,70] in 1989 and between [30,80] in 1989. This rules out any effects of early-life climate shock on school participation. Indeed, I exclude all potential students and *de facto* any endogenous effect on educational attainment, which is highly correlated to climate shocks (Randell and Gray, 2016). The endogeneity remains for the economic activity, and as discussed in Section 1.7.3, I can not rule out the fact that the results suggest changes in the labor market rather than pure migration effects.

## 1.6 Main results

### 1.6.1 Drought intensity

This section gives the main result of the paper estimated from equation 3.1. Table 2.2 displays the effects of the number of dry rainy seasons on the DPGR, including period and sublocation fixed effects. Columns (1) to (3) show the effects overall of Kenya, while Column (4) focuses on the urban sample and Column (5) on the rural sample. Column (6) gives the DiD estimator for rural sublocations where pastoralism is low, and Column (7) for rural sublocations where pastoralism is the predominant livelihood. Section 1.3.2 explains how urban, rural, low, and high pastoralism are built. Columns (2) to (6) include the interaction with the 1989 density centered around the median of each sublocation [15], while Column (3) controls for decadal mean temperature and potential evapotranspiration (PET) for each sublocation over MAMJ.

The results show that one additional dry agricultural season over a decade decreases the DPGR by 1.7 percentage points (p.p), which corresponds to a 6% reduction of the DPGR. An average sublocation loses 110 individuals due to an additional dry

[15] For all the regressions, demographic outcomes such as the DPGR, RDPGR, and density are winsorized at the 5% threshold to deal with extreme values

year, which corresponds to losing 1.3% of its population over a decade [16]. Column (2) shows that the effect is significantly attenuated with the density, which is in line with the hypothesis that the effect fades out in highly dense areas, a proxy for being more urbanized.

Table 1.1: Effects of the number of dry rainy seasons on the DPGR

| | All Kenya | | | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Nb of dry years | -1.727*** | -1.920*** | -1.769*** | -0.230 | -3.016*** | -1.342 | -4.153*** |
| | [0.549] | [0.568] | [0.603] | [1.519] | [0.680] | [1.044] | [1.132] |
| × density | | 0.00116* | 0.000874 | 0.000292 | 0.00738*** | 0.00537** | 0.0143*** |
| | | [0.000676] | [0.000669] | [0.000520] | [0.00162] | [0.00210] | [0.00414] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | No | Yes | No | No | No | No |
| N | 5036 | 5036 | 5036 | 756 | 4280 | 1626 | 1800 |
| R2 | 0.674 | 0.674 | 0.676 | 0.746 | 0.661 | 0.703 | 0.613 |
| Mean DPGR (%) | 27.75 | 27.75 | 27.75 | 31.55 | 27.08 | 20.09 | 34.15 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Columns (1) to (3) display results for all Kenya, Column (4) focuses on urban sublocations, Column (5) (6) and (7) on rural sublocations. Column (6) focuses on rural sublocations where agriculture is the main activity, while Column (7) those where it is pastoralism. Each regressions includes year and sublocation fixed effects. Variable number of dry years gives the number of years with dry rainy seasons over each decade. The variable density is the density in 1989 for each observation, centered around the median of the 1989 density (194 p./$km^2$). Column (3) controls for the mean temperature and Potential Evapotranspiration (PET) over MAMJ over the period for each sublocation. Nyanza and North Eastern provinces are excluded due to missing variables. Each demographic variable is winsorized at the 5% threshold, including the DPGR and the centered density.

Column (5) shows no effect for urban areas. The main effect is mainly driven by the comparison within rural areas, as one additional drought reduces the DPGR by 3 p.p, which corresponds to an 11% decrease (Column (5)). Within rural areas, the decrease of the DPGR seems to be concentrated in areas where pastoralism activity is high (Column (7)), where the DPGR decreases by 4.15 p.p (12% decrease). Column (3) shows the robustness of the results when controlling for decadal mean temperature and PET over the rainy season [17].

[16] On average, a sublocation is made of 6443 persons in 1989. As the mean DPGR over 1989-1999 is 28%, without any drought, an average sublocation size in 1999 should be 8247 ($6443 \times (1 + 0.28)$). Being hit by one additional drought over a decade implies a decrease of the DPGR by 1.7 p.p, which results in population size in 1999 of 8137 ($6443 \times (1 + 0.28 - 0.017)$). On average, a sublocation hit by a drought loses 110 persons over a decade, which means that the population size is reduced by 1.3%.

[17] As the result is robust to controlling to mean temperature and PET, the rest of the paper no longer includes these controls, to avoid multicollinearity issues with the main independent variable *number of dry years.*

Table 1.2: Effects of the number of dry rainy seasons on the DPGR

| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|
| **[15,65]** | (1) | (2) | (3) | (4) | (5) |
| Number of dry years | -1.788*** | -1.168 | -2.465*** | -1.849*** | -3.034*** |
| | [0.396] | [1.107] | [0.480] | [0.697] | [0.793] |
| Number of dry years × density | 0.000952* | 0.000182 | 0.00575*** | 0.00434*** | 0.00969*** |
| | [0.000517] | [0.000366] | [0.00114] | [0.00144] | [0.00274] |
| Period FE | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| N | 5036 | 756 | 4280 | 1626 | 1800 |
| R2 | 0.603 | 0.703 | 0.578 | 0.592 | 0.561 |
| Mean DPGR (%) | -11.74 | -7.534 | -12.48 | -15.16 | -9.709 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Table 1.2 gives the same results as Table 2.2, using as dependent variable the $DPGR_{[15,65]}$ for all individuals aged from 15 to 65 years old, which represents 46% of the total population. The $DPGR_{[15,65]}$ follows the age cohort as it captures the demographic growth of individuals aged between 15 to 65 years old at t-10, and aged between 25 to 75 years old at t [18]. The analysis on the $DPGR[15, 65]$ rules out any effects on fertility, infant, and old age mortality and shows that the effect is mainly driven by migration. An additional dry rainy season decreases the $DPGR_{[15,65]}$ by 1.78 p.p, which is, again, mainly driven by rural areas (Column (3)), in particular those where pastoralism is the main economic activity (Column (5)). We observe a 1.8 p.p decrease within rural sublocation where pastoralism is low (Column (4)), which can be interpreted as a proxy for high intensity of agricultural activity. This result suggests that within agricultural sublocations, one additional drought implies the migration of individuals aged between 16-65 years old, which corresponds to individuals in their working age.

These results show that the majority of the effect overall in Kenya is driven by induced migration within rural areas, and more specifically pastoralist sublocations.

[18]More formally: $DPGR_{[15,65[,i,[t-10,t]} = \frac{\Delta pop_{[15,65[,i,[t-10,t]}}{pop_{[15,65[,t-10}} = \frac{pop_{[25,75[,i,t} - pop_{(15,65[,i,t-10}}{pop_{[15,65[,t-10}}$.

This is in favor of an agricultural channel for induced climate migration and is in line with the literature. Based on the livelihoods of nomadic pastoralists, this suggests a story of short-distance movement, rural-rural migration, of entire households/villages depending on husbandry activity (McGuirk and Nunn, 2020). Migration within agricultural sublocations occurs only for individuals in the age of working, which is in line with individual migration. Section 2.7 investigates the heterogeneity of the migration across sublocation characteristics to better understand these different forms of migration.

### 1.6.2 Flood intensity

Table 1.3 shows no significant effect of the number of highly wet long-rainy seasons on the DPGR, both overall Kenya and across sublocation types. As floods are mainly born by short-rainy season OND, Table A.13 in Section A.5.4 replicates the analysis on the number of wet short-rainy season and still, show no effect of an additional flood occurring over OND. This shows that the repetition of dry conditions plays a major role in internal migration in Kenya. In Section A.3.1, I attempt to look at the effect of being hit by both droughts and floods. The results suggest that being hit by at least one drought attenuates the out-migration in response to increasing droughts.

Table 1.3: Effects of the number of wet rainy seasons on the DPGR

| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| (6) | | | | | |
| Number of wet years | 0.841<br>[0.520] | 0.454<br>[1.201] | 0.811<br>[0.606] | 1.423<br>[0.912] | 1.564<br>[1.259] |
| Number of wet years × density | 0.000182<br>[0.000240] | 0.000149<br>[0.000244] | 0.00153<br>[0.00220] | -0.000956<br>[0.00254] | 0.00186<br>[0.00735] |
| Period FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| N | 5036 | 756 | 4280 | 1626 | 1800 |
| R2 | 0.673 | 0.746 | 0.657 | 0.703 | 0.605 |
| Mean DPGR (%) | 27.75 | 31.55 | 27.08 | 20.09 | 34.15 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

## 1.7 Heterogeneity

In this section, I investigate the heterogeneity of migration across individuals' socioeconomic characteristics. A demographic account is built using the Relative Decadal Population Growth Rate (RDPGR) [19], which gives the contribution of each population subgroup to the total migration effect. Let's call $C = (c_1, ...c_i, c_n)$ an exact partition of the total population (for instance females and males). The effect on the DPGR of the total population is equal to the sum of the effect on the RDGR for each subgroup of the partition ($\beta_{DPGR,tot} = \sum_{i=1}^{n} \beta_{RDPGR,c_i}$).

Section 1.7.1 gives the heterogeneity of the migration across gender and age brackets, Section 1.7.2 according to the educational level (defined according to past schooling attendance of adults), and Section 1.7.3 according to the economic activity.

### 1.7.1 Age and Gender

#### 1.7.1.1 Gender

Table 1.4 displays the effect of one additional dry rainy season on the RDPGR of males (odd Columns) and females (even Columns). It gives the heterogeneity across gender and location. Columns (1) and (2) give the results overall Kenya, Columns (3) to (8) within rural sublocations, Columns (5) and (6) within rural sublocations where pastoralism is low while Columns (7) and (8) where it is high.

An additional dry year decreases the RDPGR of males by 1 p.p, which corresponds to a 7% reduction. It decreases the RDPGR of females by 0.9 p.p, which corresponds to a 6.5% reduction. As males and females are an exact partition of the total population, the effect on the DPGR of the total population (Table 2.2 Column (2)) is exactly the sum of the effect on males and females: $-1.920 = -1.013 - 0.907$. This implies that the migration is slightly more masculine as 53% of migrants are males, a result which holds in rural sublocations. This is attenuated within pastoralist rural areas, which display less heterogeneity, as the migration is 51% masculine.

Table A.9 in Section A.4.1 reproduces the same heterogeneity analysis across gender following the cohort of individuals aged from 15 to 65 years old as in Table 1.2.

[19] RDPGR rate is: $RDPGR_{c,i,[t-10,t]} = \frac{\Delta pop_{c,i,[t-10,t]}}{pop_{t-10}} = \frac{pop_{c,i,t} - pop_{c,i,t-10}}{pop_{t-10}}$, while $DPGR_{c,i,[t-10,t]} = \frac{\Delta pop_{c,i,[t-10,t]}}{pop_{c,t-10}} = \frac{pop_{c,i,t} - pop_{c,i,t-10}}{pop_{c,t-10}}$

Again, this allows us to rule out effects on fertility, infant, and old-age mortality. The results on the $RDPGR_{[15-65]}$ display similar patterns as in Table 1.4. As in Table 1.2, the effect is significant in rural areas with low pastoralism, in line with the hypothesis of the working population.

Table 1.4: Effects of the number of dry rainy seasons across gender and location

| Sample | All Kenya | | Rural | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | All | | Low Pastoralism | | High Pastoralism | |
| **RDPGR** | Males | Females | Males | Females | Males | Females | Males | Females |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Nb of dry years | -1.013*** | -0.907*** | -1.606*** | -1.411*** | -0.750 | -0.591 | -2.131*** | -2.022*** |
| | [0.289] | [0.287] | [0.345] | [0.344] | [0.522] | [0.533] | [0.575] | [0.572] |
| × density | 0.000457 | 0.000701** | 0.00350*** | 0.00388*** | 0.00283*** | 0.00254** | 0.00648*** | 0.00787*** |
| | [0.000338] | [0.000341] | [0.000824] | [0.000824] | [0.00107] | [0.00107] | [0.00206] | [0.00213] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 5036 | 5036 | 4280 | 4280 | 1626 | 1626 | 1800 | 1800 |
| R2 | 0.663 | 0.677 | 0.652 | 0.661 | 0.694 | 0.703 | 0.608 | 0.612 |
| Mean RDGR (%) | 13.86 | 13.89 | 13.57 | 13.51 | 10.04 | 10.04 | 17.16 | 16.99 |
| Share (%) | 48.82 | 51.18 | 48.59 | 51.41 | 48.4 | 51.6 | 48.85 | 51.15 |

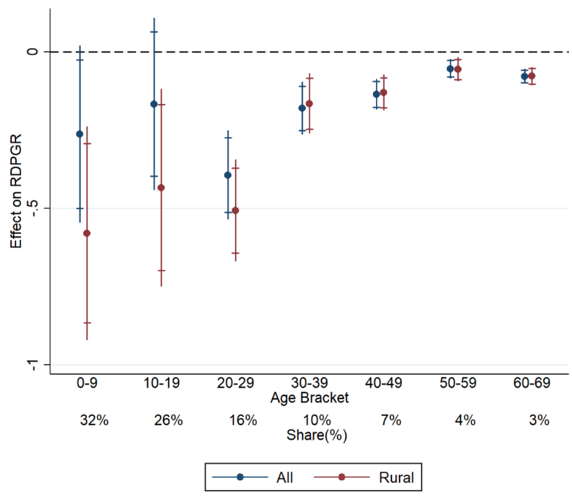
**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

#### 1.7.1.2 Age brackets

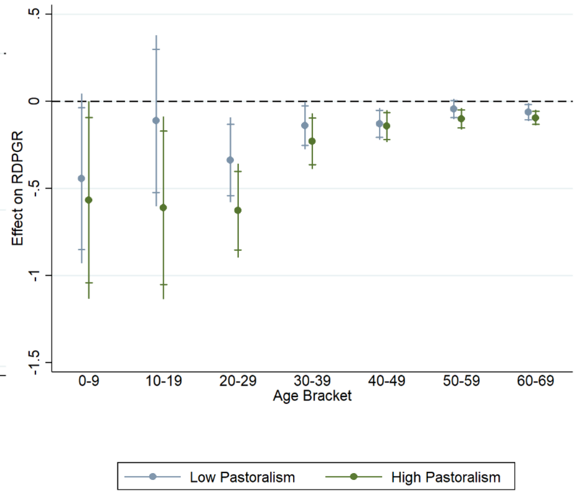
This section displays the heterogeneity analysis according to age brackets and location types. It follows the different cohorts of individuals aged between 0-69 years old at the beginning of the period (t-10) and aged between 10-79 years old at the end (t). Figure 1.7 breaks down the effect on the RDPGR of each 10-year bracket within [0-69], each bracket following cohorts. Table A.10 in section A.4.1 gives the effect on all the [0-69] cohort, which is the sum of each regression dot per location type. Figure 1.7a plots the effect for all the sublocations and rural sublocations, and Figure 1.7b distinguishes between low pastoralist and high pastoralist rural sublocations. The size of the effect on each RDPGR *de facto* depends on the size of each bracket: for instance, the [0-9] age brackets represent 32% of the total population while the [60-69] 2.9 %, as displayed in Figure 1.7a.

Overall Kenya, Figure 1.7 shows no effect on the RDPGR of children under 10 years old, which suggests no effect of droughts on infant mortality. The Figure shows an effect on those under 10 years old within rural high pastoralist areas. This can be

Figure 1.7: Effect of the number of dry rainy seasons across age brackets and location



(a) All Kenya



(b) Rural areas

**Notes:** Figure (a) plots the main result of the number across age brackets of dry years overall in Kenya and within rural sublocations. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

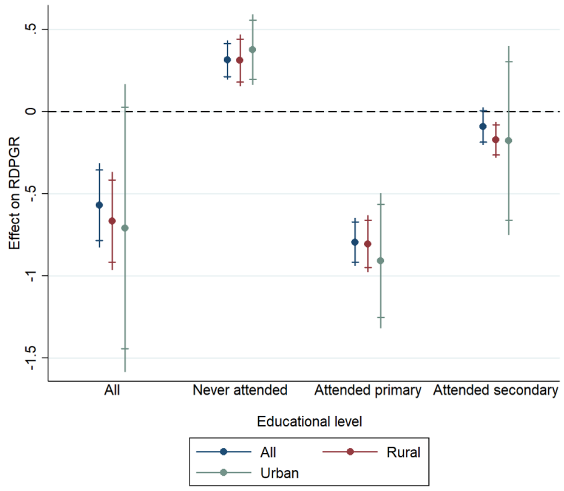
interpreted as evidence of some infant mortality, as well as homogeneous migration across age brackets, in line with the migration of entire herder households.

Figure 1.7 shows that the effect is mainly driven by the migration of young individuals in the age of working ( [10-19] and [20-29] brackets), across all types of sublocation types. This is especially the case in low pastoralist areas, as the effect is only significant for individuals aged between [20-29], while it is more homogeneous across age brackets in high pastoralist areas (Figure 1.7b). This result is in line with an out-migration of the young working population within agricultural areas.

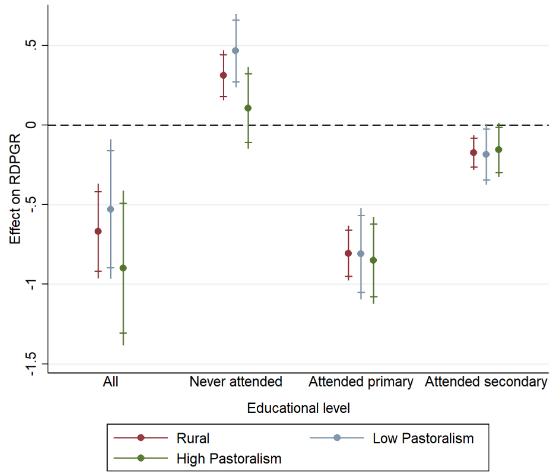
### 1.7.2 Education

This section looks at the effect of the number of dry rainy seasons across the skill distribution of adults. It follows the cohort of individuals aged between [21,69] years to omit potential students and endogeneity linked to school attendance. Figure 1.8 distinguishes individuals that never attended schooling, that at least have attended

Figure 1.8: Effect of the number of dry rainy seasons across educational level and location ([21,69])




(a) All Kenya




(b) Rural areas

**Notes:** Figure (a) plots the main result of the number of dry years across the educational level of individuals aged between 21 and 70 years in the first year of the decade. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

primary school [20], and those who have at least attended secondary education [21].

The migration is mainly driven by adults that are from the middle of the skill distribution, and who have at least attended primary education. This is consistent across all types of locations. Figure1.8 shows a reverse effect of individuals that never went to school, which can be interpreted as a proxy for the illiterate population. Figure 1.8a shows that people from the low end of the skill distribution significantly stay in affected areas. An additional dry rainy season implies an increase of the RDPGR of the illiterate population by 0.3 p.p, both within rural and urban sublocations. Within rural sublocations, this result holds only within low pastoralist sublocations but is no longer significant within sublocations where pastoralism prevails.

This result is in line with two mechanisms that are illustrated in the literature. First, it can be in line with a poverty trap story. If illiteracy is considered as a

[20] attended or completed primary education

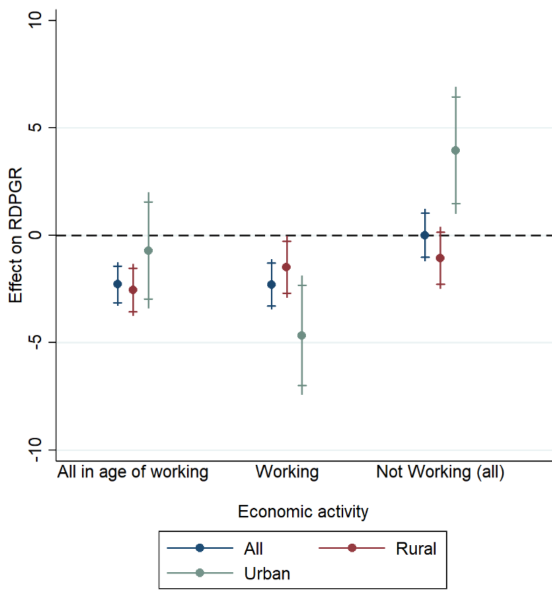
[21] attended or completed secondary education. This includes individuals that went to university. As they only represent 0.72% of the population, this subgroup could not be distinguished

proxy for richness, this result can be explained by the fact that individuals are too credit-constrained to finance a costly migration, even for short-distance movements. A second explanation would be that individuals that never attended school will have fewer professional opportunities in other places, be less able to diversify their economic activity, and have no interest to adapt by migrating. This result is mainly driven by agricultural areas, which supports the mechanism of the migration of the most skilled young individual within the household, as an adaptative response to climate variability.

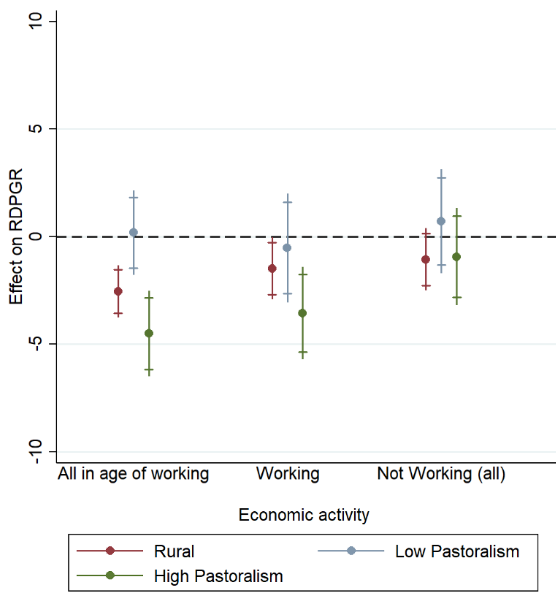
Figure A.12 in section A.4.2 displays the results on education according to age brackets, to verify that these results are not driven by any age effects.

### 1.7.3 Economic Activity

Figure 1.9: Effect of the number of dry rainy seasons across economic activity and location




(a) All Kenya




(b) Rural areas

**Notes:** Figure (a) plots the main result of the number of dry years across the economic activity of individuals in the age of working in the first year of the decade. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

Information on the economic activity of individuals allows me to understand the effect of droughts according to the type of livelihoods, and on structural transformation patterns and labor allocation. Figure 1.9 displays the demographic record of the migration according to economic activity. It compares the effect of the number of droughts on the individuals working [22] to those not working. Please note that the total effect on *All in the age of working* equals the sum of the effect on those *Working* and *Not Working (all)*.

Figure 1.9 shows that the majority of the effect on the population of working age is driven by a drop in the working population. Within rural areas, this is especially the case in high pastoralist sublocations in line with an agricultural channel [23].

As the RDPGR is an indirect measure for migration, we can not rule out the possibility that Figure 1.9 translates an effect of climate variability on labor allocation rather than migration. An important result from Figure 1.9 is the effect that appears in urban areas. In line with previous results, the effect on the total population of working age in cities is null. However, an additional drought decreases the RDPGR of the working population by 4.6 p.p while it increases the RDPGR of individuals not working by 4 p.p, within urban sublocations. Rather than evidence of out-migration, this equilibrium suggests a change in the labor allocation within cities as a consequence of droughts. This suggests that in urban areas, business owners lose their job to become unemployed because of climate variability. This story is strengthened by the results in Figure A.13 from section A.4.3, which breaks down the *Not working (all)* variable into subcategories and shows a significant effect of droughts on the increase in the share of people seeking work in urban areas.

As the results in rural areas are not balanced, we argue that it suggests evidence of the out-migration of business owners involved in agricultural practices, in line with an agricultural channel. This section displays evidence of a change in the labor allocation in urban sublocation, as business owners seem to lose their job and fall in unemployment due to climate variability.

[22]individuals working embraces those working for a pay/profit, those working for their own business and/or family holding.

[23]As we can not directly identify farming and herding as economic activities, we consider the working population within rural areas as a proxy for agricultural activity. Please note that within the working population in rural areas, 32% work for a profit while 63% work for their own business/family holding

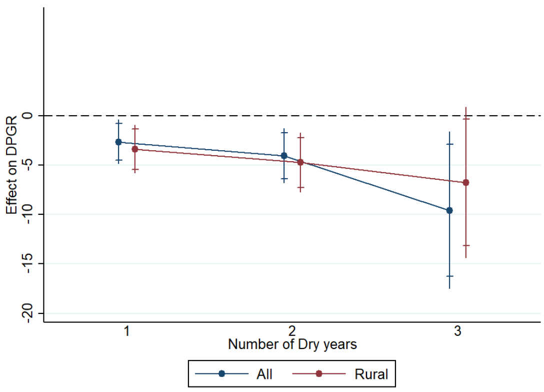
## 1.8 Intensive Margin

### 1.8.1 Droughts intensity

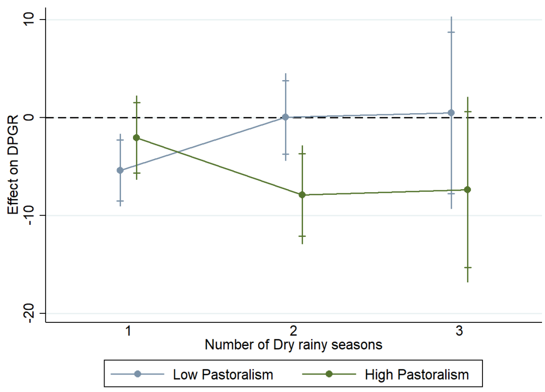
This section explores the intensive margin of the result according to the number of dry years occurring over each period. Figures 1.14a and 1.14b plot the spatial variation of the number of droughts for each period. Few sublocations were hit by more than 2 droughts within the first period. Being hit by 2 or 3 droughts occurs mainly in the second period, which corresponds to the sublocations hit by droughts in 2000, 2004, and 2007 as illustrated in Figure 1.5.

Figure 1.10 plots the statistical difference between being hit by 1, 2, and 3 droughts over the period, across location types. Overall, the effect increases when the number of droughts increases. Being hit by one drought decreases the DPGR by 3 p.p, while being hit by three droughts decreases the DPRG by 9.5 p.p, in comparison to having zero droughts over the period. Figure 1.10b shows that the effect on rural areas where pastoralism prevails is mainly driven by sublocations that have been hit by 2 droughts.

Figure 1.10: Intensive margin - Effect of the number of dry years on the DPGR




(a) All Kenya



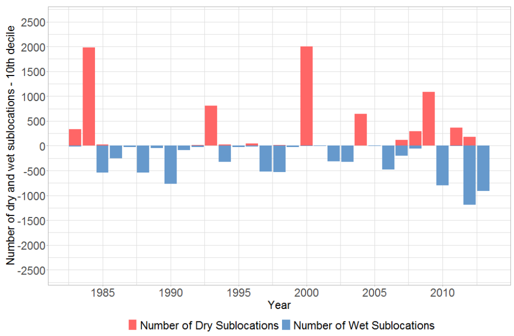

(b) Rural areas

**Notes:** Figure (a) plots the main result according to the number of dry years overall in Kenya and within rural sublocations. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

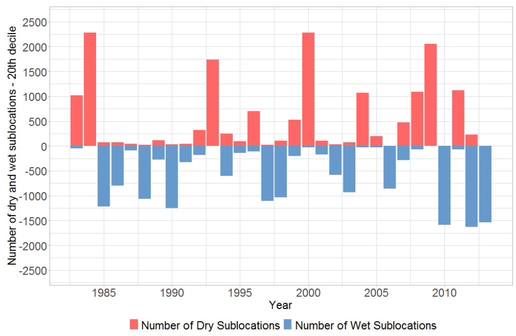
### 1.8.2 Threshold intensity

In the main result, the number of droughts over each decade is based on a dummy variable, which specifies for each sublocation whether a year is dry if the cumulative rains are below the 10th percentile of the sublocation distribution over 1983-2013. In this section, I investigate an intensive margin of the intensity of the rainfall shock, as I look at other percentiles to define the rainfall shock. Figure 1.11 replicates Figure 1.5 for the 20th, 30th and 40th deciles. For each decile, it gives the number of sublocation for which the dummy variable is dry per year. It gives as well the number of sublocations for which a particular year corresponds to excessive rains, i.e for which the cumulative rains exceed the 80th, 70th, and 60th deciles of the 1983-2013 distribution.

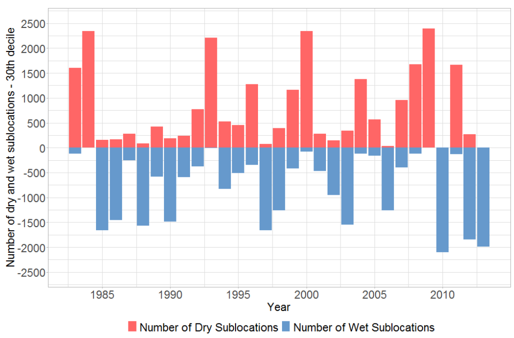
Figure 1.11: Number of dry and wet sublocations across years and deciles



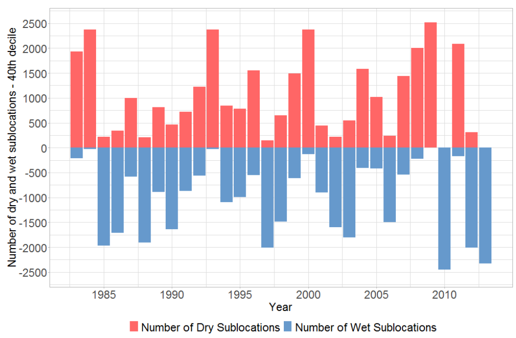

(a) 10th decile




(b) 20th decile




(c) 30th decile



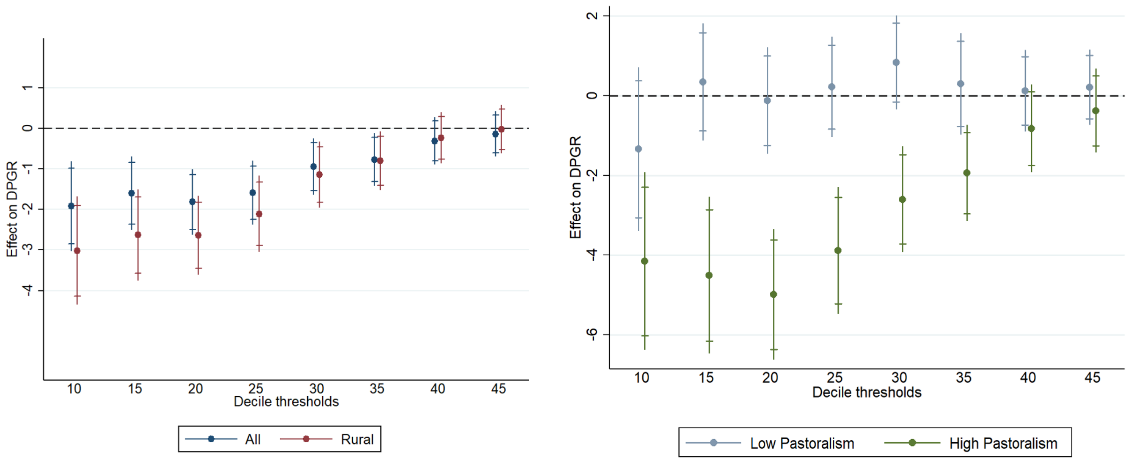

(d) 40th decile

**Notes:** These Figures in red are the number of sublocations for which the rainy season is dry and wet across years, for four different decile thresholds.

Figure 1.12 plots the DiD estimation for each regression when changing the percentile to define the rainfall shock, across each sublocation type. Effects on the 10th percentile correspond to the main results from Table 2.2. Figure 1.12a distinguishes the effects on the DPGR for all sublocations and rural ones. It shows an attenuation of the effect when increasing the thresholds, which shows that the effect is alleviated when the drought intensity reduces. Within rural areas where pastoralism prevails (Figure 1.12b), the decrease of the DPGR is still high up to the 20th decile and then reduces as the threshold increases.

Figure 1.13 replicates the same analysis for excessive rains. Effects on the 90th decile correspond to the main results from Table 1.3. The results show that when the treatment is defined based on the 85th decile, the number of wet rainy seasons increases the DPGR by 1.3 p.p. This result holds up to a treatment based on the 60th decile. This suggests that moderate rainfalls attract individuals and that the effect no longer holds when normal conditions are reached (55th decile), and is not significant for excessive rains either (90th decile). As this effect is mainly driven by rural areas (no effect within urban sublocations), this suggests a rural-rural migration, individuals leaving dry rural-areas for rural areas more humid. Figure 1.13b shows that this attraction effect is mainly borne by sublocations where agriculture is the main livelihood strategy, and pastoralism is not predominant.

Figure 1.12: Intensive margin - Effect of the number of dry years on the DPGR

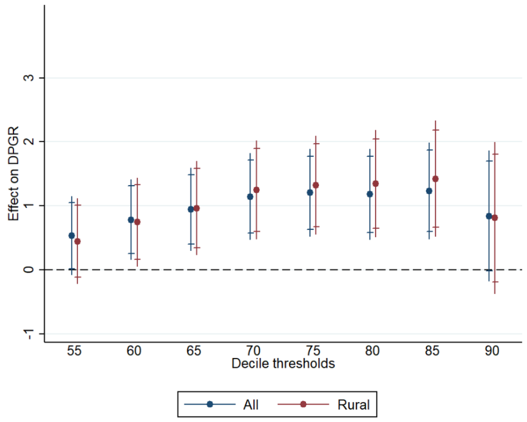



(a) All Kenya

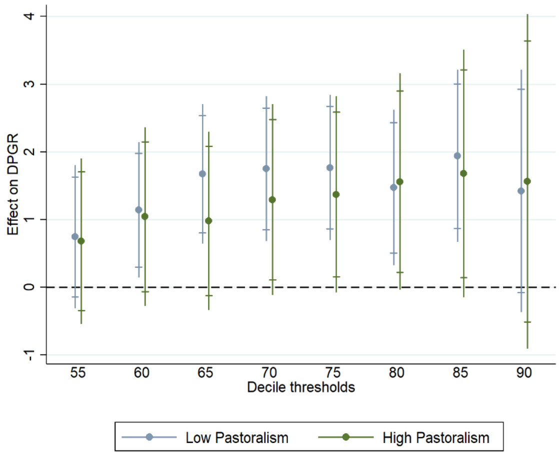
(b) Rural areas

**Notes:** Figure (a) plots the main result of changing the thresholds for being treated. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

Figure 1.13: Intensive margin - Effect of the number of wet years on the DPGR




(a) All Kenya




(b) Rural areas

**Notes:** Figure (a) plots the effect of the number of wet years on the DPGR changing the thresholds for being treated. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

### 1.8.3 Land Cover and Agricultural activity

Table 1.5 replicates the main analysis according to a land-use classification of sublocations, based on the ESA GlobCover data, as illustrated in Figure 1.3a. The land cover outcomes are computed from satellite images dated 2009. GlobCover is an ESA initiative in partnership with JRC, EEA, FAO, UNEP, GOFC-GOLD, and IGBP which provides land cover maps using input observations from the 300m MERIS sensor.

If 27% of the territory is made of croplands, 51% of sublocations are mainly composed of croplands and 46% of pastures, as sublocation size within cropland regions are smaller. Cropland areas include irrigated, rainfed, and natural croplands [24], and are a proxy for agricultural activity. Pastoral areas encompass forest, grasslands, shrubs, and herbaceous [25], and are more in line with husbandry activities.

Table 1.5 Columns (1) and (2) look at the effects of the number of droughts within areas where cropland is the main land-cover, while Columns (3) and (4) where it

[24] Unfortunately, most of the croplands are classified as natural croplands, and it is impossible to distinguish any effect between rainfed and irrigated croplands

[25] accordingly, grasslands is the dominant category

is mainly pastoral areas. It shows no effect within cropland regions, and that the overall effects are mainly driven by a decrease in the DPGR within pastoral areas [26]. This result is in line with the fact that the out-migration is triggered by herders. Column (2) interacts the independent variable with a dummy indicating the presence of pastures and shows that an additional dry year decreases the DPGR by 2 p.p within mixed areas in comparison to those where croplands are the only land-cover. Column (4) replicates this interaction and shows that within pastoral areas, the effect is significantly driven by mixed areas as well, which corresponds to pastoral areas with the presence of croplands.

Table 1.5: Effects of the number of dry rainy seasons on the DPGR across land cover categories

| | Cropland areas | | Pastoral areas | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Number of dry years | -1.148<br>[0.714] | -0.106<br>[0.721] | -2.165**<br>[0.954] | 3.251<br>[2.484] |
| Number of dry years × Pasture presence | | -2.054***<br>[0.762] | | |
| Number of dry years × Cropland presence | | | | -5.567**<br>[2.580] |
| Period FE | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes |
| N | 2594 | 2594 | 2304 | 2304 |
| R2 | 0.649 | 0.650 | 0.685 | 0.685 |
| Mean DPGR (%) | 28.38 | 28.38 | 26.42 | 26.42 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Section 1.9.1.1identifies that the treated areas are located in the Western and the Central provinces mainly agricultural, and the South of the Rift Valley is made of relatively sedentary herders. The results are driven by mixed areas, containing both pastoral areas and croplands, this suggests that the out-migration applies to herders within the Rift Valley, depending both on livestock and agricultural outcomes, and with less nomadic livelihoods. This suggests that repetitive droughts change the livelihoods of relatively sedentarised pastoralist, coping with climate events by

[26]Please note that the interaction between the independent variable and the classification dummy is not significant

migrating towards agriculture-oriented rural areas, and maybe changing their main economic activity.

## 1.9 Robustness checks

### 1.9.1 Binary treatment and de Chaisemartin and d'Haultfœuille, 2020

#### 1.9.1.1 Binary treatment

The main estimation of this paper relies on a two-period comparison of the number of droughts per sublocations. The Two-Way Fixed Effects (TWFE) estimator is driven by changes in the demographic growth of switchers, which are sublocations that change treatment status, in comparison to those that do not change status [27]. The treatment is heterogeneous, as three groups can be distinguished: Group 1 gathers sublocations for which the treatment increases, Group 2 for which it decreases, and Group 3 for which it remains stable between the two periods. Group 1 and Group 2 are sublocations that switch treatment status and who drive the main result.

However, there might be a discrepancy between the actual treatment and Groups 1, 2, and 3. As a year is defined as dry if the cumulative rains are below the 10th percentile of a 30-year period (1983-2013), each sublocation is *de facto* hit by three droughts over 1983-2013. By definition, a dry year every 10 years is not a shock, but a natural decadal drought. Let's consider some examples showing that Groups 1, 2, and 3 might be bad predictors for treatment. A sublocation hit in 1983, 1993, and 2009 (Figure 1.5) belongs to Group 2, as it switches treatment: 1 drought in period 1 [1989-1998] and 0 in period 2 [1999-2008]. However, as the three droughts are spaced out over at least ten years, their occurrence does not illustrate any increase in climate variability in the area but a normal variation in rains. This sublocation is wrongly attributed to a treatment group. Accordingly, a sublocation for which the dry years occur in 1984, 2000, and 2012 belongs to Group 1 and is wrongly allocated to a treatment group.

This section proposes another identification strategy that corrects this misallocation of treatment and control groups. I use a binary treatment which accounts for the

[27] The setting of this paper verifies the existence of stable groups in the DiD Assumption 10 from de Chaisemartin and d'Haultfœuille, 2020

increase in the occurrence of dry rainy seasons since 2000. As Figures 1.5 and 1.6 show, the second period has been hit by a national drought in 2000 (the El-Nino event) and two other shocks in 2004 and 2007-2008 that impacted the country unevenly. I define a treatment dummy $D_s$ such as [28]:

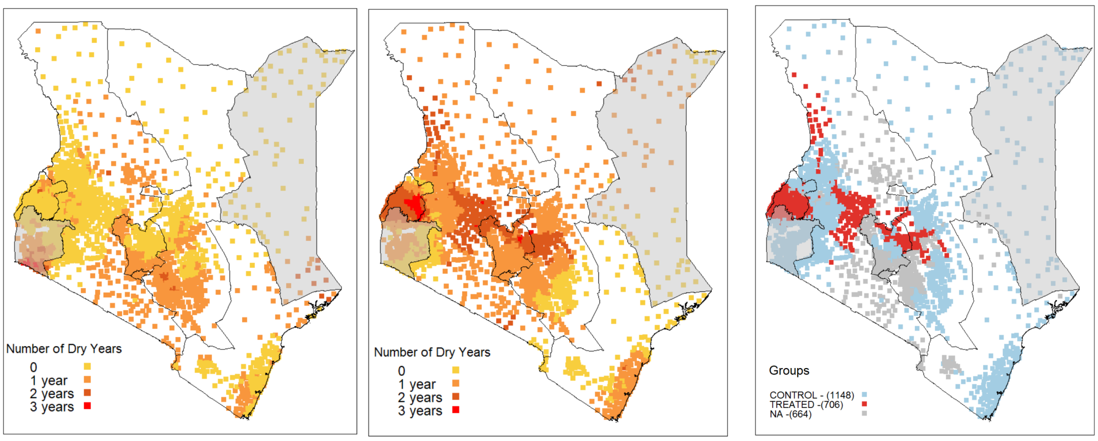
$$D_i = \begin{cases} \mathbf{1} & \text{if sublocation i was hit by at least 2 droughts in period 2} \\ & \text{(2000 and 2004 and/or 2007/2008) and no more than one in period 1 (in 1993)} \\ \mathbf{0} & \text{if sublocation i has either known a drought in 1993, either 2000, or none} \\ . & \text{if sublocation i has been hit both in 1993 and 2000} \end{cases}$$

The empirical strategy can be formally written as follows :

$$DPGR_{i,t} = \frac{\Delta pop_{i,[t-10,t]}}{pop_{t-10}} = \alpha_0 + \alpha_1 D_i + \alpha_2 P_t + \alpha_3 D_i \times P_t + \gamma_i + \epsilon_{i,t} \qquad (1.2)$$

Where $D_i$ is the treatment dummy, $P_t$ a dummy for the period, and $\gamma_i$ sublocation fixed effects.

Figure 1.14: Number of dry rainy seasons across periods and DiD groups




(a) Period 1 [1989-1998] (b) Period 2 [1999-2008] (c) DiD Groups

**Notes:** Figures (a) and (b) map the number of dry rainy seasons per period for each sublocation, each dot being the centroid of a sublocation. Figure (c) shows the DiD groups.

Figure 1.14 maps the number of dry rainy seasons over the two periods and identifies

[28] I exclude from the analysis sublocations that have been impacted both in 1993 and 2000. As the duration between the two droughts is smaller than 10 years, this represents an anomaly from the rain distribution. However, we can not capture the effect of the repetition of these two droughts as they straddle two censuses

the treated and control sublocations. Treated sublocations are clustered, located in the Western and Central provinces, as well as the center of the Rift Valley. The South of the Western province is highly urban, with high-quality lands, with no clear decreasing pattern of the DPGR (Figure 1.4 ). This is not the case for the Central and Rift Valley, with hard geographic conditions, a high share of pastoralism (1.3b) and decreasing DPGR trends. As treated sublocation are clustered, Section 1.9.3.1 adjusts standard errors for both spatial and serial auto-correlation.

Table 1.6: Balance Table - Double Difference with Binary Treatment - Descriptive Statistics

| | Period 1 [1999-1989] | | | | | Period 2 [2009-1999] | | | Within Control | Within Treated |
|---|---|---|---|---|---|---|---|---|---|---|
| | Control | | Treated | | Diff | Control | Treated | Diff | | |
| | N | Mean /(SD) | N | Mean /(SD) | (4-2) /(p.value) | Mean /(SD) | Mean /(SD) | (7-6) /(p.value) | (6-2) /(p.value) | (7-4) /(p.value) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **DPGR** | 1148 | 29.05 | 706 | 23.88 | -5.17 | 31.6 | 23.24 | -8.36 | 2.56 | -0.63 |
| | | (27.16) | | (23.27) | (0) | (25.17) | (21.39) | (0) | (0.02) | (0.59) |
| **p.a** | 1148 | 2.37 | 706 | 2 | -0.36 | 2.6 | 1.97 | -0.62 | 0.23 | -0.03 |
| | | (2.05) | | (1.76) | (0) | (1.86 ) | (1.65) | (0) | (0) | (0.74) |
| **Pop(t)** | 1148 | 6237.85 | 706 | 6018.51 | -219.34 | 8273.21 | 7651.45 | -621.76 | 2035.36 | 1632.94 |
| | | (6905.35) | | (4647) | (0.41) | (10430.32) | (6971.89) | (0.12) | (0) | (0) |
| **Dens.(t)** | 1148 | 743.27 | 706 | 560.71 | -182.57 | 479.56 | 542.52 | 62.95 | -263.71 | -18.19 |
| | | (1859.59) | | (978.39) | (0.01) | (1570.48) | (778.45) | (0.25) | (0) | (0.7) |

**Notes:** Standard errors in parentheses, p-values in brackets. *p<0.1; **p<5e-02; ***p<1e-02. Outcomes descriptive statistics of sub locations during both 1999-1989 and 2009-1999 periods,for sublocations in the control and treated groups. Outcomes are : the population size and the density at the initial year of the period ( Pop(t) and Density(t)),the Ratio giving the percentage evolution of the population (DPGR, in %) and the per annum growth rate (p.a, in %).

Balance Table 1.6 compares the changes in the demographic growth between treated and control sublocations. It displays also the size of each group, with 706 sublocations being treated and 1148 being in the control group. On average, the increase in the DPGR is higher for the control group than for the treated one, for which the DPGR is quite stable (Column (9) *vs* (10)). For control units, the DPGR trend is 3.19 p.p higher than for treated units (Column (9)-(10) or (8)-(5) [29], and is mainly explained by the fact that the population growth of the treated has less accelerated than the one of the control group. In both periods, the p.value correctly rejects the null hypothesis, that both sublocation groups have similar DPGR distribution, the difference being higher in the second period. The same observations are made for the per annum population growth (p.a). The table displays changes for the average population and density at the beginning of each period (in 1989 and 1999).

Table 1.7 displays the results of the binary treatment from equation 1.2 and shows that the main result is robust to using a binary treatment. The treatment decreases the DPGR by 3.19 p.p, which is mainly driven by rural sublocations, where pastoralism is the main economic activity. Table A.11 in Section A.5.1 gives the same results for the $DPGR_{[15,56]}$.The binary treatment analysis is not affected by the bias of negative weights, as discussed in the next Section 1.9.1.2.

Table 1.7: Effects of the increase in droughts on the DPGR - Binary treatment

| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Dummy treatment × Period | -3.189** | -0.439 | -3.549*** | -0.646 | -8.479*** |
| | [1.275] | [3.484] | [1.368] | [1.732] | [2.829] |
| Dummy Period | 2.556*** | -0.209 | 2.919*** | 0.488 | 6.991*** |
| | [0.949] | [2.499] | [1.022] | [1.385] | [1.781] |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| N | 3708 | 436 | 3272 | 1248 | 1316 |
| R2 | 0.667 | 0.701 | 0.663 | 0.727 | 0.606 |
| Size Control Group | 1148 | 133 | 1015 | 336 | 460 |
| Size Treatment Group | 706 | 85 | 621 | 288 | 198 |
| Mean DPGR Control | 30.33 | 30.89 | 30.25 | 22.62 | 36.07 |
| Mean DPGR Treated | 23.56 | 23.72 | 23.54 | 20.07 | 29.20 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

#### 1.9.1.2 de Chaisemartin and d'Haultfœuille, 2020

The main analysis of this paper uses a two-period Difference-in-Difference setting with non-binary treatment, to estimate the effect of the number of droughts on the DPGR. Recent literature shows that under heterogeneous treatment, the ATT is a weighted sum of different ATTs with weights that may be negative (de Chaisemartin

[29]This is equivalent to the DiD coefficient, $\alpha_3$ from 1.2

and d'Haultfœuille, 2020). The negative weights are an issue when the treatment effect is heterogeneous between groups over time, as one could have the treatment coefficient in those regressions as negative while the true average treatment effect is positive. In this setting, the treatment is heterogeneous over time as the control Group 3 is compared to two groups of sublocations switching treatment status, Group 1 for which the treatment increases, and Group 2 for which it decreases. The binary treatment from the previous Section 1.9.1.1 is a setting with homogeneous treatment and is not affected by negative weights, and its DiD estimator is not biased. In this section, I use the de Chaisemartin and d'Haultfœuille, 2020 estimator which deals with the issue of negative weights in a heterogeneous and non-binary treatment effect.

Table 1.8: Effects of the number of dry rainy seasons de Chaisemartin and d'Haultfœuille, 2020

| Sample | Sample 1 | | Sample 2 | |
|---|---|---|---|---|
| | TWFE | dCDH | TWFE | dCDH |
| | (1) | (2) | (3) | (4) |
| Number of dry years | -1.132* | 0.339 | -2.65*** | -3.356*** |
| | [0.639] | [1.984] | [0.644] | [0.619] |
| Sublocation FE | Yes | Yes | Yes | Yes |
| Period FE | Yes | Yes | Yes | Yes |
| N | 4,682 | 4,682 | 3708 | 3708 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Column (1) and (2) give the results on the sample comparing Group 1 and Group 3, excluding Group 3 (all sublocations for which the treatment decreases). Column (2) and (3) display the results on the sample comparing Group A and Group B, according to the binary treatment defined in Section 1.9.1.1. Columns (1) and (3) give the TWFE estimator, while Columns (2) and (4) give the de Chaisemartin and d'Haultfœuille, 2020 estimator.

Table 1.8 displays the effects of the non-binary treatment (i.e the number of droughts) for the TWFE estimation and the dCDH estimator[30]. Using the fuzzy did command and following the procedure from de Chaisemartin, d'Haultfœuille, and Guyonvarch, 2019, sublocations where the number of droughts decreased are excluded from the analysis (Group 2) [31]. Columns (1) and (2) display the results when comparing

[30]also called the time-corrected Wald ratio (Wald-TC) which relies on common trends assumptions within subgroups of units sharing the same treatment at the first date

[31]As the setting of this paper has only two periods, it is not possible to correct weights for sublocations for which the treatment decreases.

Group 1 and Group 3, for which the stable assumption holds [32]. Column (1) shows that, when excluding Group 2 from the analysis, being hit by an additional drought decreases the DPGR by 1.1 p.p , which is significant at the 5% level. The dCDH estimator is positive and non-significant (Column 2), and the STATA command showed that the two estimators are not significantly different. However, as discussed in the previous Section 1.9.1.1, the comparison between Group 1 and Group 2 is not an exact predictor of abnormal rainfall. Columns (4-6) replicate the same exercise comparing sublocations for which $D_i = 1$ (Group A) to those for which $D_i = 0$ (Group B), as plotted in Figure 1.14c. Column (3) shows that, in this sample, being hit by one additional drought decreases the DPGR by 2 p.p, which is significant at the 5% level. The dCDH estimator is larger, as an additional drought decreases the DPGR by 3.3 p.p, and is also significant at the 5% level. The two estimators are not significantly different.

### 1.9.2 Common trend assumption

The key assumption of the DiD strategy is that the dependent variable would follow the same time trends in the absence of droughts both in treated and control groups. To test for the common trends assumption, one can observe the pre-treatment data and the evolution of the DPGR before the two periods.

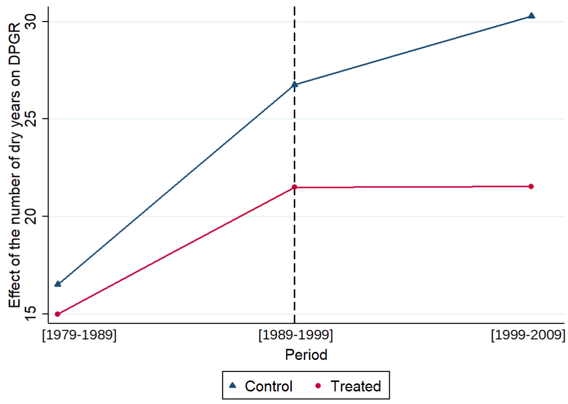
However, the main estimation of this paper relies on a two-period DiD and does not include pre-treatment data. This is mainly because the 1979 administrative data was damaged when magnetic reels got wet. To test for common trend assumption, I will consider in this section that the damages were random and did not affect the treated and control groups differently. Another issue that arises when using the 1979 census is that administrative frontiers have changed compared to the ones in 1989, and I have built a panel of sublocation starting only in 1989. Thus, I restrict this analysis to sublocations for which administrative frontiers were unchanged between 1979 and 1989, which I assume to be evenly distributed across treated and control groups. I exclude the Nyanza and North Eastern provinces as in the main analysis.

Eventually, I have a panel of 668 sublocations from 1979-2009 and for which I build the DPGR over three periods. As part of the data has been damaged, the 1979 census

[32] between the two periods, there are sublocations for which the treatment is stable, i.e the number of droughts does not change, Assumption 10 from de Chaisemartin and d'Haultfœuille, 2020 : it is Group 3

only gives a subsample of individuals for each sublocation, and the $DPGR_{[1979,1989]}$ displays very high and unrealistic numbers. As I am only interested in looking at the difference in pre-treatment trends between treated and control sublocations, I divide the$DPGR_{[1979,1989]}$ by 100 to look at averages across groups in Figure 3.6[33]. Table A.12 in Appendix A.5.2 shows that the main results over 1989-2009 are robust when restricting the sample to the 668 sublocations used in this section to test for parallel trends in pre-treatment observations.

Figure 1.15: Linear trends of the DPGR across DiD groups - three periods restricted sample




**Notes:** This Figure plots the linear trends of the DPGR across periods, averaged overt treated, and control groups defined in the binary treatment Section 1.9.1.1.

Figure 3.6 plots the linear trends of the DPGR across the three periods [1979,1989], [1989-1999], and [1999,2009], and distinguishes between treated (Group A) and control sublocations (Group B). For each period, it plots the DPGR averaged over each group, with no control nor fixed effects. Figure 3.6 shows almost parallel trends between the [1979,1989] and [1989,1999] periods, suggesting that the treated and control sublocations follow a similar pattern of demographic growth. However, with so many data discrepancies and hypotheses made on the 1979 data, the test is only indicative and has to be read carefully [34].

Following Table 1.7, Figure 3.6 shows that the DiD estimators are driven by a deceleration of the demographic growth for treated sublocations, in comparison to

[33]I divided by 100 to have coherent numbers. Besides, following the main analysis, I winsorize the $DPGR_{[1979,1989]}$ at the 5% level

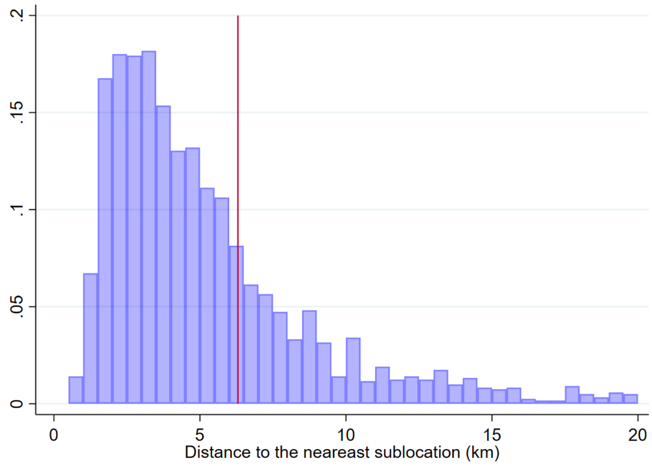
[34]A potential improvement of the test would be to randomly extract the same percentage of missing data in the 1989 and 2009 censuses

control sublocations.

### 1.9.3 Spurious correlation

#### 1.9.3.1 Spatial correlation

Figure 1.16: Distance to the nearest sublocation




**Notes:** This Figure gives the distribution of the distance between each sublocation and its closest sublocation, giving insights into how far the sublocations are located from each other. In red plots the mean distance (6.3km) (the maximum distance is up to 68 km).
*Sources:* Author's elaboration on KNBS data

The main result has two sources of spatial correlation, for both the rainfall shocks and the migration, as shown in Figures 1.14 and 1.4. This section tests for spatial correlation among sublocations that fall within different distances of each other. It accounts for the spatial pattern by using Conley, 1999 standard errors.

Figure 1.16 gives the distribution of the distance to the nearest sublocation, showing that on average a sublocation is located at 6 kilometers of its closest sublocations [35]. Standard errors are re-estimated with a spatial HAC correction following the method developed by Conley, 1999, using the Stata command introduced by Colella et al., 2019. Table 1.9 shows the stability of the significance of the main result (Column (2) Table 2.2) for difference cut-off distances of spatial correlation (from 0.5 km to 200 km).

[35] Distances are computed using the distances between each sublocation centroids. For each sublocation, the centroid location is calculated using the geometric center method

Table 1.9: Effect of the number of dry rainy seasons on the DPGR, Conley spatial correction

| Outcome | DPGR | | | | | | |
|---|---|---|---|---|---|---|---|
| Conley spatial correction threshold | 0 km | 1 km | 10 km | 20 km | 50 km | 100 km | 200 km |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Number of dry years | -1.920*** | -1.920*** | -1.920*** | -1.920*** | -1.920** | -1.920* | -1.920* |
| | [0.568] | [0.571] | [0.612] | [0.701] | [0.845] | [1.007] | [0.996] |
| Number of dry years × density | 0.00116* | 0.00116 | 0.00116 | 0.00116 | 0.00116 | 0.00116 | 0.00116 |
| | [0.000676] | [0.000755] | [0.000903] | [0.000919] | [0.000940] | [0.000948] | [0.00101] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 5036 | 5036 | 5036 | 5036 | 5036 | 5036 | 5036 |
| R2 | 0.674 | 0.00450 | 0.00450 | 0.00450 | 0.00450 | 0.00450 | 0.00450 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

#### 1.9.3.2 Placebo tests

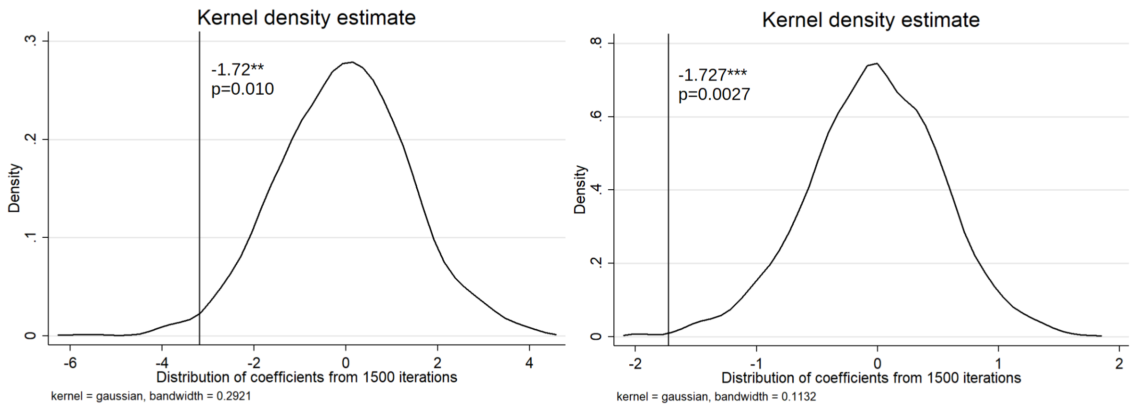
This section runs randomization tests to verify the statistical significance of the treatment effect, checking whether it is unlikely to be observed by chance. I draw 1500 permutations and compute the precise p-value based on the distribution of the 1500 counterfactual treatment effects, under the sharp null hypothesis of no effect [36]. Figure 3.7a runs spatial counterfactuals as sublocations are assigned rainfall shocks from a randomly selected sublocation. This maintains the distribution of the independent variable and removes spatial patterns. Figure 3.7b randomly changes the timing (and thus the total number) of the rainfall shocks for each sublocation.

Both simulations show that the distribution of the treatment effects are shifted around zero, and are almost perfect replication of the standard normal distribution. The vertical lines indicate the location of the estimates under the implemented treatment assignment (Table 2.2 Column 2), indicating the rejection regions, and gives the new estimated p-value. I am sure at the 1% level (Figure 3.7b) and 5% level (Figure 3.7a) that the model is not misspecified. Figure A.14 replicates this inference test, changing randomly the treated sublocations for the binary treatment over 1500 permutations, and shows that the binary treatment model is not misspecified at the

[36] The test is done using the *ritest* STATA command.

1% level.

Figure 1.17: Temporal randomization inference tests - Continuous treatment




(a) Random sublocation (b) Random timing

**Notes:** The two figures represent the distribution of the treatment effects of the number of dry years when conducting 1,500 permutations. Figure (a) randomly changes the sublocations allocation to droughts while Figure (b) randomly changes the timing/number of droughts for each sublocation. The vertical line indicates the location of the estimate under the implemented treatment assignment (Table 2.2 Column 2), and gives the new estimated p-value.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

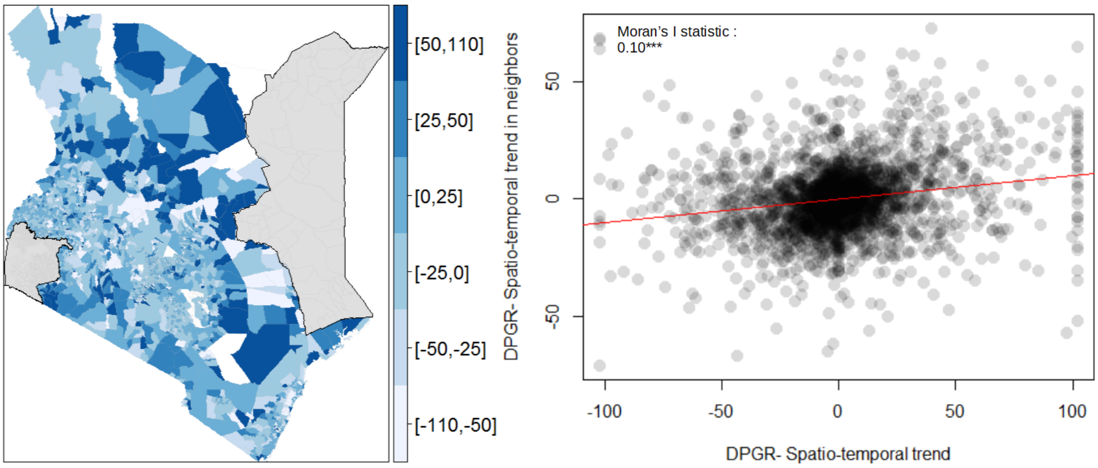
#### 1.9.3.3 Spatially dependent trends

The occurrence of rainfall is, by nature, spatially and temporally correlated, which can generate a spurious relationship between rainfall and other spatially correlated outcomes. Issues linked to omitted variables are exacerbated in the presence of spatial dependency. Sublocation fixed effects, as well as (Conley, 1999) standard errors (previous Section 1.9.3.1) control for spatial patterns, as long as the dependent variable only exhibits spatial correlation. However, spurious correlations remain in the case of spatially dependent trends.

Section 1.9.3.2 represents a first solution to the problem of spurious rainfall effects when temporal trends are spatially correlated. It shows that my estimate is robust to testing for the null hypothesis rejection when both changing the spatial and temporal allocation of rainfall shocks. This section goes beyond placebo tests in solving the problem of spurious correlation, following the procedure of Lind, 2019.

First, Figure 1.18 illustrates the spatial dependency of the dependent variable. Figure

Figure 1.18: Spatiotemporal patterns of DPGR




(a) DPGR trends (b) Moran's plot - DPGR trends per neighboring

**Notes:** Figure (a) plots the spatial distribution of the DPGR trends (or long-difference). Figure (b) plots Moran's scatterplot.
*Sources:* Author's elaboration on KNBS data.

1.18a plots the geographical distribution of sublocation temporal trends, as it plots the DPGR long-difference for each sublocation. It displays clear positive trends in the north-eastern part of the country (excluding the north-eastern province), while more attenuated, decreasing trends in the west and the center. Figure 1.18b tests for Moran's I statistics. It shows a Moran plot, plotting the sublocation DPGR long-difference against the average trends in adjacent municipalities [37]. The slope of the line is Moran's I coefficient, which equals 0.10***. The positive slope suggests that when the DPGR of a sublocation increases, so does those of its neighboring sublocations. Moran's test for no spatial dependency is rejected with a very small p-value ($2.2 \times 10^{-16}$), and shows the spatial correlation of the migration outcome. Figure 1.14 shows the spatial clustering of the trends of rainfall shocks as well.

As spatial pattern is found, Table 1.10 proposes several tests. Columns (2) to (4) control for spatial trends and show the robustness of the main estimation. Column (2) controls for province trends and Column (3) for district trends. Column (4) control for the tensor product of Legendre polynomials with $1 \times 6$ terms to control for spatiotemporal trends as proposed in Lind, 2019, and show that my estimation is

[37]Moran's scatterplot and test have been conducted using the moran.test from R spdep package

robust [38].

Table 1.10: Effects of the number of dry rainy seasons on the DPGR -Controling for Spatiotemporal trends

| | Without trend | With trend | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Number of dry years | -1.920*** | -2.789*** | -3.566*** | -1.881*** |
| | [0.568] | [0.674] | [0.965] | [0.582] |
| Number of dry years × density | 0.00116* | 0.00169** | 0.00124* | 0.00111 |
| | [0.000676] | [0.000783] | [0.000730] | [0.000687] |
| Period FE | Yes | No | No | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes |
| Province trend | No | Yes | No | No |
| District trend | No | No | Yes | No |
| Tensor product of Legendre Polynomials | No | No | No | Yes |
| N | 5036 | 5036 | 5036 | 5036 |
| R2 | 0.674 | 0.679 | 0.695 | 0.674 |
| Mean DPGR (%) | 27.75 | 27.75 | 27.75 | 27.75 |

**Notes:**Standard errors clustered at the sublocation level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

### 1.9.4 Contamination of the Control group

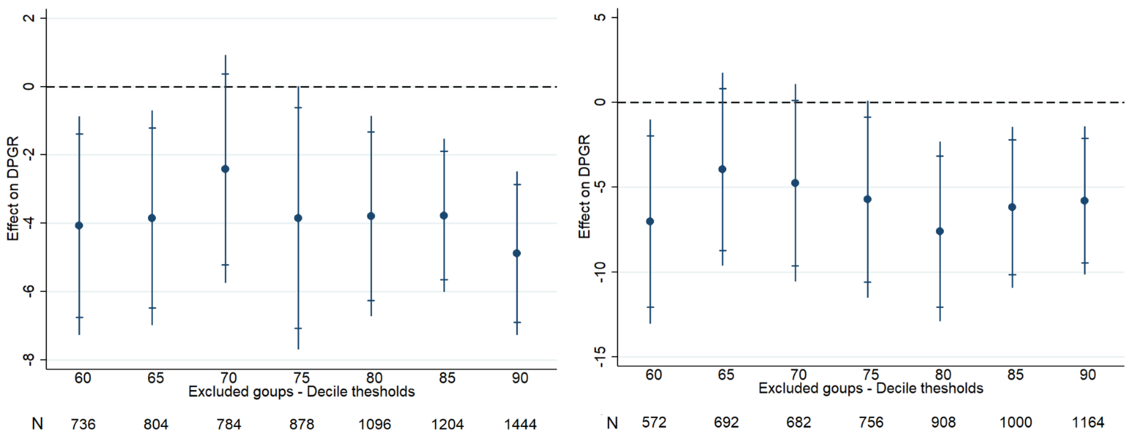
One concern in the main analysis is the overestimation of the effect due to the contamination of the control group. The main result shows that being hit by an additional dry year increases the likelihood of out-migration. As the results suggest a rural-rural migration mainly born by herders, there is a low probability of international migration in the setting. The universe of individuals remaining stable, out-migrants are likely to migrate towards control sublocations, which biases the DiD estimation. The main concern is that, as the DPGR of treated sublocations decreases due to droughts, the DPGR of control sublocations *de facto* increases, being the recipient of out-migrants. As the DiD estimates the difference in demographic growth between the treated and control areas, this gives an upward bias of the result.

Section 1.8.2 suggests that the individuals migrate within sublocations with humid conditions. Figure 1.13 shows that being under an additional moderate wet year

[38]The tensor is a function of longitude, latitude and time, such as $T(x, y, t) = U(x, y)t = t\sum_{k=0}^{K}\sum_{l=0}^{L}{}_{kl}P_k(x)P_l(y)$, where $P_i(.)$ is the ith-order Legendre polynomial. Legendre polynomials are defined recursively with $P_0(x) = 1$, $P_1(x) = x$, and for $i \geq 2$, $P_i(x) = [(2i-1)xP_{i-1}(x) - (i-1)P_{i-1}(x)]/i$

increases the DPGR by 1.5 p.p . These results suggest that within control areas, migrants favor sublocations becoming wetter (but not too much wet). As the goal of this Section is to test the contamination of the control group, I check whether the result remains stable when excluding those sublocations for which the DPGR increases.

Figure 1.19: Effects of the number of dry years on the DPGR - changing samples for estimation




(a) TWFE Continuous treatment

(b) DiD Binary treatment

**Notes:** Figures plot the effect of the number of dry years excluding sublocations for which the number of years the rains exceed a certain threshold are excluded. Figure (a) plots the effect of several regressions for the TWFE continuous treatment, while Figure (b) plots the treatment effect for the binary treatment. For instance, the coefficient 60 in Figure (a) gives the estimator when excluding all sublocations for which the number of wet years is not stable across periods, defining wet years according to years for which cumulative rains over MAMJ exceed the 60th threshold. N gives the number of observations for each regression.

Figure 1.19a plots the results of the main estimation, looking at the effect of the number of dry years changing the sample of the control group. The coefficient 60 excludes all sublocations for which the number of wet years is not stable across periods, defining wet years as being years for which cumulative rains over MAMJ exceed the 60th threshold. It shows that for sublocations for which the number of humid years remained stable across periods, one additional dry year decreases the DPGR by 4 p.p, showing that the main result of this paper is not overestimated. The other dots correspond to different levels of exclusions. Figure 1.19b replicates the same exercise for the binary treatment as defined in Section 1.9.1.1.

### 1.9.5 Other climate indicators

#### 1.9.5.1 The role of temperature and evapotranspiration

Table 1.11 Column (1) shows that the main result of this paper is robust when controlling for mean temperature and PET over the rainy season across periods. Climate shocks being multidimensional (Auffhammer et al., 2013), this section goes beyond looking at the effect of extremely warm temperatures. I define the number of hot years per period as the number of years for which the mean temperature over MAMJ is over the 90th decile of the temperature distribution of each sublocation. Table 1.11 Columns (2) displays no significant effect of extreme temperature on the DPGR, and shows that the effect of the number of droughts is robust when controlling for temperature anomalies.

Table 1.11: Effects of the number of dry rainy seasons on the DPGR - Control for temperature and evapotranspiration

| | All Kenya | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of dry years (10th decile) | -1.769*** | -1.881*** | -1.930*** | | | |
| | [0.603] | [0.580] | [0.672] | | | |
| Number of dry years × density | 0.000874 | 0.00136* | 0.00130* | | | |
| | [0.000669] | [0.000710] | [0.000708] | | | |
| Number of hot years (90th decile) | | 1.491 | | | | |
| | | [1.048] | | | | |
| Number of hot years (80th decile) | | | 0.0988 | | | |
| | | | [0.755] | | | |
| Number of dry years (SPEI - 2months) | | | | -1.110* | | |
| | | | | [0.635] | | |
| Number of dry years (SPEI - 3months) | | | | | -0.383 | |
| | | | | | [0.579] | |
| Number of dry years (SPEI - 4months) | | | | | | 0.436 |
| | | | | | | [0.606] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Temp and PET Controls | Yes | No | No | No | No | No |
| N | 5036 | 5034 | 5034 | 5036 | 5036 | 5036 |
| R2 | 0.674 | 0.675 | 0.674 | 0.673 | 0.673 | 0.673 |
| Mean DPGR (%) | 27.75 | 27.74 | 27.74 | 27.75 | 27.75 | 27.75 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Columns (2) to (6) display the result of the number of dry years defined according to the SPEI (Vicente-Serrano, Beguería, and López-Moreno, 2010), which controls for both temperature and PET (Section A.5.4 gives the definition of the SPEI). *Number* *of dry years (SPEI -2months)* is the number of years for which at least two months over the MAMJ season are under a drought. *Number of dry years (SPEI -4months)* indicates a dry year if the entire rainy season is under droughts according to the SPEI definition. The results show almost no effect of the SPEI, only the *Number of dry years (SPEI -2months)* decreases the DPGR by 1.1 p.p and is hardly significant at the 10% level. This is mainly explained by the fact that, as the SPEI looks at anomalies of water balance, it captures temperature trends and identifies less critical droughts than my estimator relying on rainfall shortages directly. Figure A.15b in Section A.5.4 plots the number of sublocation under a dry year according to the SPEI definition, and shows that the dry years are more temporally and spatially distributed.

#### 1.9.5.2 Other indicators

Table 1.12 shows the effect of the number of years under abnormal climate conditions, defined according to other climate indicators than cumulative rains. All indicators are built over the long-rainy season MAMJ. Column (1) looks at the effect of the number of years for which the number of wet days is very low (R1plus under its 10th decile), Column (2) for which the length of the longest dry spell is high (CDD over its 90th decile), Column (3) for which the length of the longest wet spell is low (CWD¡10th decile). Column (4) and (5) gives the effects of the number of years for which the daily intensity of rains is low (under the 10th decile) and high (above the 90th decile).

Table 1.12: Effects of other climatic indicators on the DPGR

| | All Kenya | | | | |
|---|---|---|---|---|---|
| Indicator | R1plus | CDD | CWD | SDII | SDII |
| Threshold | 10th decile (1) | 90th decile (2) | 10th decile (3) | 10th decile (4) | 90th decile (5) |
| Number years under/above threshold | -0.325<br>[0.530] | -0.0120<br>[0.409] | -0.546<br>[0.694] | -1.261**<br>[0.493] | -0.210<br>[0.479] |
| Period FE Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE Yes | Yes | Yes | Yes | Yes | Yes |
| N | 5036 | 5036 | 5036 | 5036 | 5036 |
| R2 | 0.673 | 0.673 | 0.673 | 0.674 | 0.673 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Table 1.12 shows no significant effect for each indicator, aside from the decrease in the daily intensity of rains. An additional year with a low daily intensity of rains decreases the DPGR by 1.26 p.p. As the SDII gives the cumulative precipitations divided by the number of wet days (R1plus), it is the closest to the main independent variable and explains why it is significant. Table 1.12 shows that it is mainly the drop in cumulative rains over MAMJ that play a key role in migration.

Table A.13 in Section A.5.4 shows no effects of droughts nor floods occurring during the short rainy season OND.

### 1.9.6 Replication at the district level

Table 1.13: Effects of the number of dry rainy seasons on district DPGR

| | All Kenya | | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of dry years | -5.035*<br>[2.844] | -5.526*<br>[3.124] | 1.691<br>[2.905] | -5.858*<br>[3.304] | 2.722<br>[3.094] | -7.594**<br>[3.472] |
| Number of dry years × density | | 0.00748<br>[0.00968] | -0.00348<br>[0.00727] | 0.00828<br>[0.0105] | 0.00204<br>[0.00714] | 0.0114<br>[0.0111] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 76 | 76 | 66 | 76 | 58 | 76 |
| R2 | 0.870 | 0.871 | 0.868 | 0.871 | 0.904 | 0.863 |
| Mean DPGR (%) | 43.68 | 43.68 | 46.85 | 42.10 | 29.34 | 43.82 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Table 1.13 replicates the main analysis at the district levels. It looks at the effects of the number of dry years on the demographic growth of districts. For each district, I build the DPGR by looking at the total population per district and year. Results from Table 1.13 show that an additional dry year decreases the district DPGR by 5p.p, which is again mainly born by rural areas, where pastoralism is the main economic activity [39]. The results being hardly significant at the 10% levels, it shows the comparative advantage to look at the effect at a less aggregated level. A main contribution of this paper is to look at the effects at the sublocation level, which

[39]The magnitude size seems to differ from the main estimation, but this is mainly driven by higher DPGR mean at the district levels. A 5 p.p decrease in the DPGR at the district corresponds to a 8% increase. The mean DPGR is higher because the winsorization has been done on the population sizes at the sublocation level before aggregating at the district level. As there are only 41 districts, the DPGR could not be winsorized at the aggregated level.

makes it possible to capture the small magnitude effects found in the heterogeneity analysis, which are not found when looking at the district level.

Section A.6 proposes an other estimation, looking at the effects of yearly droughts on the inter-district bilateral migration flows using a PPML estimation. Again, as the results fail to find a significant out-migration at the district level, it shows the necessity to look at the effects at the scale of sublocation in order to capture small magnitude effects.

## 1.10 Conclusion

This paper estimates the effects of past climatic conditions on migration movements, in the long run at a micro-level in Kenya. In recent decades, Kenya has faced downward trends in the number of rainy days and the length of wet spells during the long-rainy season, associated with higher intensity of wet days. This decrease in the length of the agricultural period is associated with an increase in the occurrence of droughts in a country highly vulnerable because dependent on agricultural and livestock incomes. In the most recent period, several droughts are identified in Kenya with different spatial coverage (2000, 2004, 2008-2009), and this increase in the repetition of droughts since 2000 makes Kenya an interesting setting to analyze the effects of climate shocks on local migration.

This paper exploits the spatial variation of the intensification of dry events since the 2000 regional El Nino drought and investigates the migration response at a local level in the long-run. I match exhaustive administrative census data provided by the KNBS with high accuracy, spatial and temporal resolution precipitation and temperature data from the CHC over 20 years. I use the decadal population growth rate (DPGR) of 2518 sublocations over two periods, [1989-1999] and [1999-2009] as a proxy for migration rates. I propose a two-way fixed effects strategy that estimates the effects of an additional dry rainy season over each 10-year period on the DPGR.

The results show that an additional dry rainy season decreases the DPGR by 1.7 p.p, which corresponds to a 6% reduction of the DPGR. The effect is mainly driven by the out-migration of rural-areas, especially those where pastoralism plays a key role in livelihood strategies. The result is robust to restricting the sample to the [15-65] cohort, which shows that it is not driven by a change in fertility outcomes, old age, or infant mortality rates. I find no effects of the number of floods, showing that the migration is mainly triggered by the repetition of slow-onset events such as droughts rather than rainfall extreme disasters.

The main contribution of the paper is a multi-dimensional heterogeneity analysis, which identifies different types of migrations. Within rural areas where pastoralism prevails, I find little heterogeneity in migration across gender, age brackets, and educational levels. This suggests that herders out-migrate with their entire households. The results suggest that repetitive droughts change the livelihoods of relatively sedentarised pastoralists located within the South of the Rift Valley, coping with climate events by migrating towards agriculture-oriented rural areas. This result relates to rural-rural migration being a solution of last resort for herders.

Overall, agriculture-oriented rural areas are less vulnerable to droughts. Within these sublocations, I observe important heterogeneity, as the out-migration is mainly driven by skilled and young individuals who reached the age of working, in line with an individual migration. Results display a trap effect for the illiterate population, which can be interpreted as the consequence of liquidity constraints, or limited job-opportunity in the destination. These results suggest that farmers' households adapt to recurrent droughts with the out-migration of the most skilled individuals, who are the more able to work. This relates to an off-farm adaptation strategy of income diversification.

This paper is in-line with a rural-rural migration in response to the repetition of several droughts occurring over a short span. Changing the thresholds of treatments, for both dry and wet events, the results suggest that individuals out-migrate from rural areas where pastoralism prevails to agriculture-oriented rural areas with normal and humid conditions.

I run manyfold robustness checks and show that the results hold when using a simple difference-in-difference strategy with binary treatment and controlling for the de Chaisemartin and d'Haultfœuille, 2020 estimator. I show that the results are not biased by spurious correlation, as they are robust to correcting for spatial auto-correlation (Conley, 1999), spatial and temporal randomization inference tests, and correcting for spatially dependent trends (Lind, 2019). I find no effects on other climate indicators such as highly hot rainy seasons and dry short-rainy seasons and propose a test to correct for the contamination of the control group.

# Bibliography


Auffhammer, Maximilian et al. (May 2013). “Using Weather Data and Climate Model Output in Economic Analyses of Climate Change”. In: *Review of Environmental Economics and Policy* 7.2, pp. 181–198.

Banerjee, Abhijit et al. (2015). “The Miracle of Microfinance? Evidence from a Randomized Evaluation”. In: *American Economic Journal: Applied Economics* 7.1, pp. 22–53.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl (2006). “Climatic change and rural–urban migration: The case of sub-Saharan Africa”. In: *Journal of Urban Economics* 60.3, pp. 357–371.

Beine, Michel and Christopher Parsons (2015). “Climatic Factors as Determinants of International Migration”. In: *The Scandinavian Journal of Economics* 117.2, pp. 723–767.

Bertoli, Simone et al. (Nov. 2021). “Weather shocks and migration intentions in Western Africa: insights from a multilevel analysis”. In: *Journal of Economic Geography* 22.2, pp. 289–323.

Bryan, Elizabeth et al. (Jan. 2010). *Coping with Climate Variability and Adapting to Climate Change in Kenya: Household and Community Strategies and Determinants.* Report3a. Report to the Worl Bank.

Cai, Ruohong et al. (2016). “Climate variability and international migration: The importance of the agricultural linkage”. In: *Journal of Environmental Economics and Management* 79, pp. 135–151.

Campbell, David J and George Axinn (1980). *Pastoralism in Kenya.* Report. American Universities Field Staff.

Cattaneo, Cristina, Michel Beine, et al. (2019). “Human Migration in the Era of Climate Change”. In: *Review of Environmental Economics and Policy* 13.2, pp. 189–206.

Cattaneo, Cristina and Giovanni Peri (2016). “The migration response to increasing temperatures”. In: *Journal of Development Economics* 122.Supplement C, pp. 127–146.

Cole, Shawn et al. (Jan. 2013). “Barriers to Household Risk Management: Evidence from India”. In: *American Economic Journal: Applied Economics* 5.1, pp. 104–35.

Colella, Fabrizio et al. (2019). “Inference with Arbitrary Clustering”. IZA Discussion Paper n. 12584.

Conley, Timothy (Sept. 1999). “GMM estimation with cross sectional dependence”. In: *Journal of Econometrics* 92.1, pp. 1–45.

Dallmann, Ingrid and Katrin Millock (Sept. 2017). “Climate Variability and Inter-State Migration in India”. In: *CESifo Economic Studies* 63.4, pp. 560–594.

de Chaisemartin, Clément and Xavier d’Haultfœuille (Sept. 2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. In: *American Economic Review* 110.9, pp. 2964–96.

de Chaisemartin, Clément, Xavier d’Haultfœuille, and Yannick Guyonvarch (2019). “Fuzzy differences-in-differences with Stata”. In: *The Stata Journal* 19.2, pp. 435–458.

Defrance, Dimitri, Esther Delesalle, and Flore Gubert (2022). “Migration response to drought in Mali. An analysis using panel data on Malian localities over the 1987-2009 period”. In: *Environment and Development Economics*, pp. 1–20.

Dell, Melissa, Benjamin Jones, and Benjamin Olken (2014). “What Do We Learn from the Weather? The New Climate–Economy Literature”. In: *Journal of Economic Literature* 52.3, pp. 740–798.

Dinku, Tufa et al. (2018). “Validation of the CHIRPS satellite rainfall estimates over eastern Africa”. In: *Quarterly Journal of the Royal Meteorological Society* 144.S1, pp. 292–312.

Endris, Hussen Seid et al. (Oct. 2013). “Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall”. In: *Journal of Climate* 26.21, pp. 8453–8475.

Feng, Shuaizhang, Alan Krueger, and Michael Oppenheimer (2010). “Linkages among climate change, crop yields and Mexico–US cross-border migration”. In: *Proceedings of the National Academy of Sciences* 107.32, pp. 14257–14262.

Findley, Sally (1994). “Does drought increase migration? A study of migration from rural Mali during the 1983-1985 drought”. In: *International Migration Review* 28.3, pp. 539–553.

Fishman, Ram (Dec. 2013). “Rainfall Variability, Adaptation through Irrigation, and Sustainable Management of Water Resources in India”. In: *AGU Fall Meeting Abstracts*, GC32A-07, GC32A–07.

Funk, Chris et al. (Aug. 2019). “A High-Resolution 1983–2016 Tmax Climate Data Record Based on Infrared Temperatures and Stations by the Climate Hazard Center”. In: *Journal of Climate* 32.17, pp. 5639–5658.

Gebrechorkos, Solomon, Stephan Hülsmann, and Christian Bernhofer (2019). “Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania”. In: *International Journal of Climatology* 39.1, pp. 18–30.

Gráda, Cormac and Kevin O’Rourke (1997). “Migration as disaster relief: Lessons from the Great Irish Famine”. In: *European Review of Economic History* 1.1, pp. 3–25.

Gray, Clark and Valerie Mueller (2012). “Natural Disasters and Population Mobility in Bangladesh”. In: *Proceedings of the National Academy of Sciences* 109.16, pp. 6000–6005.

Henderson, Vernon, Adam Storeygard, and Uwe Deichmann (2017). “Has climate change driven urbanization in Africa?” In: *Journal of Development Economics* 124.C, pp. 60–82.

Herrero, Mario et al. (2010). “Climate variability and climate change and their impacts on Kenya’s agricultural sector”. In: *International Livestock Research Institute ILRI Report*.

Hesse, Ced and James MacGregor (2006). *Pastoralism: drylands’ invisible asset? Pastoral Civil Society in East Africa*. Tech. rep. 142. IIED Issue Pape.

Hoerling, Martin et al. (2006). “Detection and Attribution of Twentieth-Century Northern and Southern African Rainfall Change”. In: *Journal of Climate* 19.16, pp. 3989–4008.

Hornbeck, Richard (June 2012). “The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe”. In: *American Economic Review* 102 (4).

Joseph, George and Quentin Wodon (2013). “Is Internal Migration in Yemen Driven by Climate or Socio-economic Factors?” In: *Review of International Economics* 21.2, pp. 295–310.

Liebmann, Brant et al. (Dec. 2014). “Understanding Recent Eastern Horn of Africa Rainfall Variability and Change”. In: *Journal of Climate* 27.23, pp. 8630–8645.

Lind, Jo Thori (2019). “Spurious weather effects”. In: *Journal of Regional Science* 59.2, pp. 322–354.

Long, Jason and Henry Siu (2018). “Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants”. In: *The Journal of Economic History* 78.4, pp. 1001–1033.

Lynham, John, Ilan Noy, and Jonathan Page (2017). "The 1960 Tsunami in Hawaii: Long-Term Consequences of a Coastal Disaster". In: *World Development* 94, pp. 106–118.

Lyon, Bradfield and David DeWitt (Jan. 2012). "A recent and abrupt decline in the East Africa long rains". In: *Geophysical Research Letters* 39, pp. 2702–.

Marchiori, Luca, Jean-Francois Maystadt, and Ingmar Schumacher (2012). "The impact of weather anomalies on migration in sub-Saharan Africa". In: *Journal of Environmental Economics and Management* 63.3, pp. 355–374.

McGuirk, Eoin and Nathan Nunn (Dec. 2020). *Transhumant Pastoralism, Climate Change, and Conflict in Africa.* NBER Working Papers 28243. National Bureau of Economic Research, Inc.

Missirian, Anouch and Wolfram Schlenker (2017). "Asylum applications respond to temperature fluctuations". In: *Science* 358.6370.

Mogaka, Hezron et al. (Jan. 2006). "Climate variability and water resources degradation in Kenya: Improving water resources development and management". In: *World Bank Working Paper*, pp. 1–105.

Mueller, Valerie, Clark Gray, and Douglas Hopping (2020). "Climate-Induced migration and unemployment in middle-income Africa". In: *Global Environmental Change* 65, p. 102183.

Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market". In: *The Quarterly Journal of Economics* 118.2, pp. 549–599.

Nicholson, Sharon (May 2015). "An analysis of recent rainfall conditions in eastern Africa". In: *International Journal of Climatology* 36.

– (2017). "Climate and climatic variability of rainfall over eastern Africa". In: *Reviews of Geophysics* 55.3, pp. 590–635.

Nyariki, DM and DA Amwata (2019). "The value of pastoralism in Kenya: Application of total economic value approach". In: *Pastoralism 9* 9.

Omondi, Philip et al. (2013). "Linkages between global sea surface temperatures and decadal rainfall variability over Eastern Africa region". In: *International Journal of Climatology* 33.8, pp. 2082–2104.

Özden, Caglar, Christopher Parsons, and Maurice Schiff (2011). "Where on Earth is Everybody?" In: *World Bank Economic Review* 25.1, pp. 12–56.

Piguet, Etienne, Antoine Pécoud, and Paul de Guchteneire (2011). "Migration and Climate Change: An Overview". In: *Refugee Survey Quarterly* 30.3, pp. 1–23.

Randell, Heather and Clark Gray (2016). "Climate variability and educational attainment: Evidence from rural Ethiopia". In: *Global Environmental Change* 41, pp. 111–123.

Reuveny, Rafael and Will Moore (Sept. 2009). "Does environmental degradation influence migration? Emigration to developed countries in the late 1980s and 1990s". In: *Social Science Quarterly* 90.3, pp. 461–479.

Sedova, Barbora and Matthias Kalkuhl (2020). "Who are the climate migrants and where do they go? Evidence from rural India". In: *World Development* 129.C.

Strobl, Eric and Marie-Anne Valfort (2015). "The Effect of Weather-Induced Internal Migration on Local Labor Markets. Evidence from Uganda". In: *World Bank Economic Review* 29.2, pp. 385–412.

Tobin, Lara (2017). *Essais sur l'urbanisation en Afrique Subsaharienne*. Tech. rep.

Veljanoska, Stefanija (2018). "Can Land Fragmentation Reduce the Exposure of Rural Households to Weather Variability?" In: *Ecological Economics* 154, pp. 42–51.

Vicente-Serrano, Sergio, Santiago Beguería, and Juan López-Moreno (Apr. 2010). "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index". In: *Journal of Climate* 23, pp. 1696–1718.

WFP (2018). *Pastoral and Agro-pastoral Production System in the Arid and Semi-Ari Areas*. Field Practitioners Guide No. 3.

# Chapter 2

# MiningLeaks: Water Pollution and Child Mortality in Africa


## Abstract

Industrial mining can be a boon or a bane for communities living in the vicinity of production sites. We assess the effects of mining-induced pollution on health outcomes in Sub-Saharan Africa using the DHS micro-data from 1986-2018 in 26 countries matched with geocoded data of industrial mining sites. Through a staggered difference-in-difference strategy, we exploit the variation of the opening of a mine and the relative topographic position of surrounding villages, comparing upstream and downstream villages. Being downstream of an open mine increases by 2.18 percentage points the 24-month mortality rate, corresponding to a 25% increase. This effect is mainly driven by the consumption of plain water, corroborating the mechanism of water pollution. The effect on mortality is not driven by a change in women's fertility, nor by a change in the access to piped water or other facilities, nor by in-migration. The effect is concentrated while the mine is active and when international mineral prices are high, is larger in areas with high mining density, and fades out with distance. It is robust to the estimator of de Chaisemartin and d'Haultfœuille, 2020, to a restriction to a balanced sample, to accounting for measurement errors, and to spatial and temporal randomization inference tests.

## 2.1 Introduction

The increase in commodity prices since 2000, especially in the extractive sector, has intensified investments in areas with abundant resources, from hydrocarbons to minerals. The African geology, which is richly endowed containing 30% of the world's mineral reserves (Chuhan-Pole, Dabalen, and Land, 2017), remains largely unexplored due to inhospitable terrains and the lack of infrastructure (UN Environment Program, 2022; Africa Bank, 2022) and represents an opportunity for mining investors. Africa is facing a mining boom since the 2000s, attracting foreign investment mainly from China, Canada, Australia, Brazil, and Russia, which raises concerns about environmental degradation (Taylor et al., 2009; Edwards et al., 2013) and health impact on local populations. Human exposure to heavy metals through the consumption of contaminated water is of prior concern in Africa and Sub-Saharan Africa in particular, where only 24% of the population have access to safe drinking water (UNESCO, 2019).

Throughout each stage of a mine's life cycle, its activity can produce and release chemical and mineral pollutants prone to contaminate the surrounding air, water, and soil. Moreover, the ore extraction processes are water-demanding and need access to a water source that very often competes with the local demand, which is all the more alarming in water-stressed areas. Mining activity mostly consists in extracting small concentrations of minerals from huge volumes of rocks and therefore creates a lot of waste, which leaking is hard to avoid. For industrial mines, these wastes are diluted into water and then stored in retention ponds, where they can leak within the local environment and contaminate soil and water bodies. If low concentration levels of heavy metals can be essential for human health, the abnormal quantities found in the environment within the mine's vicinity can cause several health problems. Individuals living nearby industrial mining are exposed to high concentrations of heavy metals through ingestion, dermal contact, and inhalation of soil particles. We mainly focus on the absorption mechanism as we identify the effects of mining activity through water pollution. Exposure to heavy metals plays detrimental effects on human health in general and child health in particular, especially during their first months of development, both in and ex-utero (Coelho and Texeira, 2011). Children are the most sensitive, even at low concentrations, as they are at a stage of rapid biological development (Dike et al., 2020), but also as they are more exposed, through higher blood concentration linked to incidental ingestion of urban soil and unclean water (He et al., 2020).

In this paper, we focus on Africa and investigate the local impacts of industrial mining activity on health through water pollution using geocoded micro-data. We examine the 12-month and 24-month mortality as a primary health outcome, as effects on children are the most dramatic, and to capture the effects of heavy metal absorption on early-age biological development. Child mortality is a short-term measure (Greenstone and Hanna, 2014; Do, Joshi, and Stolper, 2018), and is available over a long-time span of four decades and across the majority of African countries. We also look at the effects on other children's health outcomes such as anthropometric measures and anemia, as well as women's health and fertility outcomes. We match socio-economic and health data from the Demographic Health Surveys (DHS) with state-of-the-art geolocalized data on industrial mineral resource exploitation from the SNL Metals and Mining database, which provides information on opening dates and mineral types. Our study spans 26 out of the 54 African countries, from 1986 to 2018. We conduct a staggered Difference-in-Difference strategy exploiting the variation of the opening of a mine and the relative topographic position of surrounding villages. We indirectly isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison, which is used as a proxy for the exposure to water pollution linked to mining activity.

Our main result shows that being born in a village located downstream of an open mine increases the mortality rates under 24 months by 2.18 percentage points, which corresponds to an increase by 25% of the mortality rate. We find no significant result for the 12-month mortality rate, which suggests a lag in the effect of water pollution on early-childhood health and in the absorption of toxic elements. Our analysis suggests that this could be explained by the protection provided by breastfeeding (VanDerSlice, Popkin, and Briscoe, 1994; Fängström et al., 2008), as we find a significant increase in the 12-month mortality rate among children who were given plain water, in comparison to those who were not. We find effects neither on other children's health outcomes nor on women's health outcomes and fertility. We exclude many potential mechanisms that could explain our main result on infant mortality and show the robustness to controlling for households' access to water and facilities such as health facilities and electricity as well as in-migration flows. Our main effect is mainly driven by the mortality of boys, and by individuals living downstream of an open-pit mine that has opened. The effect fades out with distance and increases with surrounding mine density and the intensity of production, proxied by yearly international mineral prices. It seems to be mainly occurring during the mining activity status, as it is robust to restricting the analysis to mines for which we have the closure year. Last but not least, we conduct a battery of robustness checks: the de Chaisemartin and d'Haultfœuille, 2020 estimator, using a balanced sub-sample of DHS repeated cross-sections, correcting for DHS random displacement, restricting to the sample of mines for which we have the exact coordinates to account for measurement errors, testing for spatial spillovers, and running spatial and temporal randomization inference tests.

The major contributions of this paper are twofold. First, it lies in the construction of the industrial mines dataset, as we checked over 1,700 mines by hand to complete their opening date [1]. Our second contribution is to introduce the topographic heterogeneity of the effects of mining activity on health and to identify the negative effects induced by water pollution, and how they outweigh the positive effects. It nuances the results from the literature looking at average effects by using geographical distance to a mine as a proxy for exposure to mining activity, which finds a reduction in mortality rates (Benshaul-Tolonen, 2018; Cossa et al., 2022).

The remainder of the paper is organized as follows. Section 3.2 reviews the literature and presents our contribution. Section 3.3 describes the context and the data. Section 3.4 details the methodology and the main empirical strategy. Section 2.5 introduces the main results, while section 2.6 investigates the mechanisms, and section 2.7 the heterogeneity of the results. Section 2.8 looks at the dynamic effects, and section 2.9 at the intensive margins, digging into the heterogeneity of the results according to the distance of the mine, the mining density, and the production intensity. Section 2.10 proposes a list of robustness checks and placebo tests. Section 2.11 discusses the limits of the study and section 2.12 proposes a policy discussion. Eventually, section 2.13 concludes.

## 2.2 Literature review and contributions

This section first displays the literature on the trade-off of mining activity in developing countries. It then describes the mining-induced pollution literature and the economic literature on the health effects of mines. Thirdly, we discuss the issues emerging from this literature and the solutions we propose to tackle them.

[1] Opening dates indicate when production first began. Data available in the SNL database was gathered by SNL from the mining companies' reports, and the hand-check work we made has completed this database by going deeper into archival mining reports

#### 2.2.0.1 Trade-off of mining activity

Our work is related to the strand of literature analysing the health-wealth trade-off of industrial mining activity in developing countries, which results are still under debate. If mining can improve health and well-being through local industrial development, it can also damage health through negative externalities such as conflicts, massive migration waves, and exposure to harmful pollution. Determining which of these effects is predominant is still debated in the literature studying the relevance of a natural resource curse (Ploeg, 2011; Cust and Poelhekke, 2015; Venables, 2016).

At a broad scale of analysis, Mamo, Bhattacharyya, and Moradi, 2019 look at the effects of the discoveries of industrial mining deposits in Sub-Saharan Africa. They find an increase in district-level night light emissions but no significant effects on household wealth [2].They find temporary positive effects on public service provisions but a degradation of the sewerage system and piped water supply in the medium and long run. Mining creates also negative effects on the environment and agricultural productivity. Aragón and Rud, 2016 find that the expansion of large-scale gold mining in Ghana (1997-2005) is responsible for the agricultural total factor productivity decrease in the vicinity of mines. The use of cross-sectional satellite imagery of NO2 concentration suggests that air pollution is the main explaining factor. Dietler et al., 2021 analyze a panel of 52 mines in Sub-Saharan Africa using the same DHS and SNL databases. They find improvements in access to modern water and sanitation infrastructures after a mine opens when comparing individuals living within 50 kilometers of an isolated mine. Yet, proxying exposure to mining activity with distance and focusing on areas with low mining density raises many identification issues, that will be largely discussed in Section 2.2.3.1. Our paper deals with these issues and encompasses a wider sample of mines. Other negative externalities of mining activity are the increase of rapacity and corruption and the trigger of insecurity and conflicts (Berman et al., 2017), migration flows of mine workers fueling the spread of infectious diseases such as HIV (Corno and Walque, 2012), and discouragement of educational attainment among children (Atkin, 2016; Ahlerup, Baskaran, and Bigsten, 2020; Malpede, 2021).

Our paper focuses on industrial mining and does not encompass artisanal and small-scale mining (ASM). Few papers have looked at the effects of artisanal and

[2]Household wealth was measured through the dimensions of access to electricity, wealth index, urbanization, mortality, and education.

small-scale mining (ASM), mainly due to data limitations. Bazillier and Girard, 2020 compare the local spillovers between artisanal and industrial mining sites in Burkina-Faso. They find positive impacts of artisanal mining (labor intensive and managed in common) and an absence of industrial mines' opening (capital intensive and privatized) on household consumption. Our paper focuses on the effects of industrial mining pollution. If ASM has severe effects on miners' health due to hazardous working conditions it is likely to be of smaller magnitude than industrial mining which extracts and treats larger volumes [3]. If ASM is often accused of generating more severe pollution than the industrial sector because of their illegal use of mercury [4], the latter often use cyanide instead. Both chemicals being highly toxic pollutants, focusing on industrial mining only is a lower-bound analysis of the impacts of mining activity on local populations' health.

### 2.2.1 Mining-induced pollutions

Each stage of industrial mining activity produces chemicals and minerals likely to pollute the surrounding air, water, and soil (Coelho and Texeira, 2011). The exploration and prospecting stage can last several years before a mine is considered economically viable and worthwhile to open. Meanwhile, mining companies conduct mapping and sampling, as well as drilling, boreholes, and excavation that require both physical and chemical measurement methods likely to pollute at the surface and underground, depending on the nature of the deposit in the targeted area. If found financially viable, the company launches the discovery phase where the design and planning of the construction are undertaken, and the feasibility study of the project requires further exploration and engineering studies. Subsequently, the development stage takes place and the mine's infrastructures and processing facilities are constructed. It is only after all these stages that production can start. Once the deposit is exhausted comes the closure and reclamation stage, where the company is supposed to clean, stabilize and rehabilitate the land and isolate contaminated material. Yet, it is common that waste, tailings, or retention dams are just left abandoned without care and maintenance, and this constitutes a potential disaster if the hazardous materials are leaked and discharged into the environment. Figure B.9 in the Appendix proposes a scheme to explain the life cycle of a mine. Figure B.8 displays satellite images of the different stages of the Essakane mine, an open-pit gold mine in Burkina-Faso.

[3] Industrial mines are responsible for 80% of the gold production and 75% of the diamond production McQuilken and Perks, 2020.

[4] Mercury has been officially banned in over 140 countries (Minamata Convention on Mercury, adopted in 2013).

Throughout all these stages, different types of pollution can be engendered. Air pollutants can be carried by dust over long distances by ore transportation and the wind, they can damage surrounding soils and crops, and be inhaled mostly by mine workers but also by the local population. The leakage of pollutants in the air can also affect water through acid mine drainage that ends up polluting the surface and then groundwater. During the digging and processing to extract the targeted ore from waste rocks, rocks are crushed and then go through either heap leaching, froth flotation, or smelting. These techniques require the addition of chemicals such as cyanide or acid, that can separate the targeted minerals from waste. Moreover, these processes are water-demanding and need access to a water source in competition with the local demand. Last but not least, even without the use of these chemicals, leaching happens through the contact of water and oxygen with sulfide minerals contained in the extracted rocks , which accelerates the acidification process and modifies the pH levels of water bodies. Pollutants can be released into the environment during the process by spills or after by leaks of humid waste stored in retention dams but also through the erosion and the sedimentation of solid wastes that are piled in the tailings around the mining site and that drain to the soil with rain. The wastes actively pollute during the whole life cycle of the mine, starting from its opening and during production, but also can continue to pollute when a mine closes and is left without maintenance. This is the case when retention ponds are not covered and dry, letting these wastes go directly through the environment.

Few papers have managed to show to what extent industrial mining activity creates negative externalities on the environment. Bialetti et al., 2018 look at the effects of mining industries on deforestation in India, Von der Goltz and Barnwal, 2019 have suggested the mechanism of water pollution but without strong empirical evidence (looking at anemia). Yet, in-situ measurements have shown the contamination of water drinking sources by harmful levels of nitrate, turbidity, iron, cadmium, manganese, and arsenic by industrial mining sites (Cobbina, Kumi, and Myilla, 2013). To our knowledge, we are the first to provide indirect, systematic, and large-scale evidence of the mechanism of water pollution by industrial mining activity.

The main toxic metals released by mining sites are arsenic, cadmium, copper, lead, mercury, and nickel. Depending on their blood level concentration, they can be essential or non-essential for human health (El-Kady and Abdel-Wahhab, 2018).

However, heavy metals released by mining activity are non-biodegradable, have long-term impacts on the environment, and are found at abnormally high concentrations in the vicinity of mines, within the soil, water resources, vegetation, and crops (Oje et al., 2010; Dike et al., 2020). People living in that environment are exposed to high quantities of heavy metals through ingestion, dermal contact, and inhalation of soil particles, which can cause several implications for their health. High blood metal concentrations are associated with neurological effects (which induce behavioral problems, learning deficits, and memory losses, especially among children) (Dike et al., 2020), neurodegenerative diseases, cardiovascular effects, gastrointestinal hemorrhages (Obasi et al., 2020), organ dysfunction (kidney, decrease of the production of red and white blood cells, lung irritation) (Briffa, Sinagra, and Renald, 2020), higher probability of cancer development (Madilonga et al., 2021; Obasi et al., 2020), but also a higher probability of infertility, miscarriages for women, and malformation of newborns (Briffa, Sinagra, and Renald, 2020). Thus, exposure to heavy metals plays detrimental effects on human health in general and child health in particular, especially during their first months of development, both in and ex-utero (Coelho and Texeira, 2011). Besides, children at an early age are the most sensitive, even at low concentration, as they are at a stage of rapid biological development, but also as they are more exposed, through higher blood concentration linked to incidental ingestion of urban soil and dirty water (less conscious of their environment and danger, playing with polluted soil, eating and drinking without care (He et al., 2020)).

### 2.2.2 Health effects of mining activity

The empirical economic literature on the local effects of mining on local communities has been growing during the past decade, yet the debate remains on the costs and benefits, and the positive and negative impacts of industrial mining activity in developing countries. Diverse results have been found on the effects on health, and there is still uncertainty on the direction and the magnitude of the impacts of mines on the local population's health. Besides, if geographical proximity to a mining site is usually used as a proxy for pollution exposure, few papers observe the negative externalities on the environment and its consequences on health.

Papers studying the effects of industrial mines on health proxy the exposition to mining activity by the distance to the mine and can be found in the literature different thresholds and mixed results. Using cross-section data in the state of Orissa in India, Shubhayu et al., 2011 uses the distance to the mine as a proxy to measure environmental effects, and finds that individuals living near a mine report higher respiratory illness and more work days lost due to malaria. Cross-sectional data prevent identifying a clear causal relationship and from adjusting to specific time and spatial confounders. Benshaul-Tolonen, 2018 uses a Difference-in-Difference strategy, comparing individuals living within 10 kilometers to those living between 10-100 kilometers of a mine, before and after its opening. The paper finds that large-scale gold mining in nine countries of Sub-Saharan Africa [5] decreases infant mortality within 10 km during the opening and operating phases, with no effect on further communities (10-100km). Cossa et al., 2022 uses a similar methodology studying a broader set of countries and finding a decrease in child mortality as well.

Von der Goltz and Barnwal, 2019 assess the effects of industrial mines in 44 developing countries from 1988 to 2012. The paper also relies on a Difference-in-Difference strategy, comparing households living within 0 to 5 km to households living between 5 and 20 km before and after the opening of a mine. They find gains in asset wealth, increased anemia among women, and stunting in young children. As anemia and growth deficits are argued to be mainly the consequences of exposure to lead, the observed effects on health are interpreted to be the results of pollution due to metal contamination and lead toxicity. They find that women in mining communities show depressed blood hemoglobin, recover more slowly from blood loss during pregnancy and delivery, and that children in mining communities suffer some important adverse growth outcomes from in-utero exposure (stunting).

### 2.2.3 Challenges and contributions

The most common way to proxy exposure to mining activity is to rely on the distance to an active or open mine, however, there is no clear consensus on which threshold to use, and the treatment allocation seems arbitrary. The disparities in the results from the literature could be explained by these differences in terms of empirical strategies and distance choices. Beyond this, using the Euclidian distance to a mine as treatment raises endogeneity concerns. This subsection discusses the main issues arising when studying the local impacts of industrial mining activity on health.

[5] Burkina Faso, Ivory Coast, the Democratic Republic of the Congo, Ghana, Guinea, Ethiopia, Mali, Senegal, and Tanzania between 1987 and 2012

#### 2.2.3.1 Endogeneity issues

In this section, two challenges are discussed: the endogeneity issues that arise (i) when using the Euclidian distance as a proxy for exposure to mining activity, and those when using (ii) repeated cross-sectional data such as the DHS.

Using the interaction between being close to a mine and the mine's activity status raises endogeneity concerns. For instance, Von der Goltz and Barnwal, 2019 use a mine panel and pairs each DHS village to its closest mine. This creates unbalanced treatment and control groups, and such imbalance might be endogenous to socio-economic outcomes or polluting behaviors. As each village is paired to its closest mine, this *de facto* excludes from the control group villages that are in both distance categories (within 5 km of mine A but 5-20 km of mine B). Thus, there is a higher probability to be treated in areas with high mining density, which is not a random allocation. As a mine fixed-effect identification relies on a within-mine buffer-area comparison, the estimator is driven by mines that have been paired to villages both in the treated and control areas, which is correlated to the mining density of the region. The estimation endogenously selects mines from regions of low or middle mining importance, which might be correlated with the intensity or the type of pollution, the socio-characteristics of the neighboring population, and thus the way health is affected by pollution. To reduce endogeneity issues, Von der Goltz and Barnwal, 2019 instrument the mine location with mineral deposit information from S&P, which are *deposits that are being explored or prepared for exploitation*. However, mining exploration is not a random allocation and raises the same concerns as it is directly correlated to mining density. Benshaul-Tolonen, 2018 reduces endogeneity issues linked to the pairing by using an administrative district fixed-effect panel and extending the distances (10-100km), but the same concern remains.

A second concern is linked to the nature of the DHS data, which are repeated cross-section surveys. The literature argues that the conditions for an industrial mine to settle are the presence of mineral deposits, which is considered random. However, the presence of a mine and of a declared mineral deposit is correlated to the population density. As mining exploration is labor intensive, it is more likely to occur in dense areas, where DHS is more likely to have surveyed individuals. A treatment allocation based on geographic proximity to the mine is endogenous: treatment groups close to the mine might not be comparable to control groups located further. As district fixed-effect relies on a within-district comparison, the estimation is driven by districts with both control and treated groups, before and after a mine opening, which is correlated to the probability of being surveyed. As DHS renews the surveyed villages at each wave, and as the probability to be surveyed is determined by the population density, the estimation is driven by specific areas. The regression *de facto* and endogenously selects districts that were already dense before the opening and remained after. This might be areas that are more stable, well-off, and where individuals might be less affected by pollution. This might bias the estimation upward (i.e less mortality linked to mining activity), and explain the positive effect of mines that Benshaul-Tolonen, 2018 finds on mortality in Africa.

In Appendix section B.6, we propose a replication analysis of Benshaul-Tolonen, 2018, taking advantage of our handwork which extends the SNL database. We find similar results as Benshaul-Tolonen, 2018 using the same set of countries and our extended sample of mines (only gold mines as in the paper). However, we find no longer significant results when applying to our more comprehensive sample, meaning when including other African countries and industrial mines, which suggests that the effects are context and regional-dependent.

#### 2.2.3.2 Upstream-downstream analyses

Using geographic distance to a mine as treatment allocation raises endogeneity concerns. An upstream-downstream analysis, which relies on a topographic comparison, reduces these concerns as individuals are compared from similar distances.

Few papers have dealt with upstream and downstream at the scale of a continent, since it requires much more computational capacity and a complex pairing methodology. Duflo and Pande, 2007 study the productivity and distributional effects of large irrigation dams in India and use river networks and calculate gradients computed from digital elevation maps for India. Do, Joshi, and Stolper, 2018 use river networks and pollution monitoring stations data in India to conduct their upstream-downstream analysis. Unfortunately, it is impossible in our case study due to the absence of water quality data at the scale of Africa. Garg et al., 2018 use river networks in Indonesia and re-calculate the upstream-downstream relationship between village pairs using a 30m resolution Digital Elevation Model. Their very refined level of study is not likely to be undertaken at the scale of the African continent in our case, so we choose secondary data computed by hydrologists (HydroSheds). We use systematic and highly disaggregated data on water sub-basins that enable us to encompass a wider set of countries, overcoming the issue of pairing a mine or a village to the closest river, since there is uncertainty about whether this point is located above or below in altitude compared to the level of the river. E. Strobl and R. Strobl, 2011 studied the distributional effects of large dams on agricultural productivity at the scale of the African continent, using Pfafstetter level 6 with an average area of 4200 km$^2$. Our study takes into account sub-basins at the Pfafstetter level 12, with an average area of 100 km$^2$.

## 2.3 Data and Context

This section describes the data used for our empirical strategy, and some descriptive statistics on the context of industrial mining and child mortality in Africa.

### 2.3.1 Data

In this paper, we match socio-economic data from the Demographic Health Surveys to an industrial mining database provided by SNL Mining and Metals.

#### 2.3.1.1 Health and socio-economic data

We use all available survey rounds from the Demographic Health Surveys that contain GPS coordinates, from 1986 to 2018, covering 36 out of the 54 African countries. We then select all the countries which have at least two survey waves to be able to implement our Difference-in-Difference strategy with a sufficient time length before and after the opening of a mine and end up with 26 countries[6], 12,442 clusters and 240,431 children under the age of five. We consider that doing a Difference-in-Difference strategy on the sample of countries that have only one round of the survey, hence a maximum of five years period, will not enable us to capture the longer-term effects of mining activity[7]. Table B.2 in the Appendix displays the DHS survey years and countries that we use for the analysis.

We construct the variables of child mortality based on the DHS child recode database which has information on the age and death of children under five years old, whose

[6]The list of countries within our sample are: Benin, Burkina Faso, RDC, Burundi, Cote d'Ivoire, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Liberia, Lesotho, Madagascar, Mali, Malawi, Nigeria, Niger, Namibia, Rwanda, Sierra Leone, Senegal, Togo, Tanzania Zambia, and Zimbabwe

[7]Please note that our final sample does not include Egypt which has 7 DHS waves and is a well-known mining country. This is explained by the fact that the SNL database characterized Egypt within the Middle East rather than in Africa and thus was dropped from our sample.

mothers are aged between 15 and 49 years old. Our dependent variable is the probability of 12-month and 24-month mortality for each DHS cluster (i.e. for each child, we build a dummy equal to 1 if she or he is alive and 0 if not, conditional on having reached 12 and 24 months respectively). We also estimate the effects of mining activity on biomarker variables and other indicators of occurrences of illness (diarrhea, fever, and cough) within two weeks preceding the day of the interview among young children. We extend our analysis to women's fertility behavior and health: current pregnancy, total lifetime fertility, miscarriage, and anemia. Finally, as the aim of this article is to isolate the mechanism of water pollution, we use the questions from the DHS on the main source of drinking water, the presence of flushed toilets, electricity, and the access to health facilities to control for households' sanitary and economic environment.

#### 2.3.1.2 Mineral resource exploitation data

The industrial mining variables come from the SNL Metals and Mining database, which is privately owned by *S&PGlobal* and on license [8]. The SNL database is the best existing panel of mine production, providing information on the location, the dates of opening and closure, the commodity type, and the yearly production (for some mines). This is a non-exhaustive panel of industrial mines in Africa, yet to our knowledge, it constitutes the most comprehensive sample of mines giving the timing of the industrial activity. This dataset has been intensively used in the literature and argued to be the best product available (Aragón and Rud, 2016; Berman et al., 2017; Kotsadam and Tolonen, 2016; Benshaul-Tolonen, 2018; Von der Goltz and Barnwal, 2019; Mamo, Bhattacharyya, and Moradi, 2019). We emphasize here that this paper focuses on the effects of industrial mining, and that we do not include artisanal mining (ASM) that are not available in the SNL database.

Overall, the SNL database gathers 3,815 industrial mines in Africa from 1981 to 2021, and 2,016 were located within 100 km of a DHS cluster from a country with at least two surveys. For our difference-in-difference strategy, we need information on the timing of the beginning of the mining production. The SNL database gives this information for 278 mines and we retrieved from handwork the start-up year for the 1,738 remaining mines. The hand-check was realized using the information on the mining history available in the SNL database and mine reports (cross-checked with Google Maps and aerial images). We describe this handwork more extensively in the

[8]We are grateful to CEPREMAP, PjSE, EHESS, and the GPET thematic group of PSE, for their financial support and their help in purchasing the access to the data.

Appendix B.1.2.

We build three main variables from the SNL Mining and Metals database, relying on the geocoded information and the time of opening. According to the estimation strategy, we will use a variable of proximity (distance to the closest mine), position (whether individual i is upstream or downstream), and a dummy for being open or not. Opening dates that were available in the SNL database were computed by the SNL team, and indicated the actual start-up year of the mine, i.e when production first began. We used the same criteria for our handwork. Finally, we restrict the main analysis that is associated with heavy metal mines (metals with density higher than $5g/cm3$ (Briffa, Sinagra, and Renald, 2020), which are the metals listed in Table B.6 in Section B.2.2 of the Appendix. We also include coal mines, as their extraction is associated with mercury and arsenic which are highly toxic heavy metals.

#### 2.3.1.3 Water basins

We consider the topographic relationship of water basins where mines and villages are located. A water basin is an area where all the surface water converges towards the same point. We use the HydroBASINS sub-basins geographic information provided by HydroSHEDS, which delineates water basins consistently and subdivides sub-basins into multiple tributary basins to the network of nested sub-basins at different scales. Following the topological concept of the Pfafstetter coding system, each polygon of the sub-basin has a unique direction flow and provides information on the up-and down-stream connectivity. We take the finest Pfafstetter level (12 out of 12) that breaks down sub-basins at an average area of 100 $km^2$. See Figure 2.6a for an example. We conduct our analysis taking into consideration the three closest sub-basins to each industrial mine, meaning that we take each mine's sub-basin A and tag the one just downstream that we call B, the one just downstream of B that we call C, and then the one just downstream of C that we call D. Thus, B, C, and D are the three closest sub-basins of A.

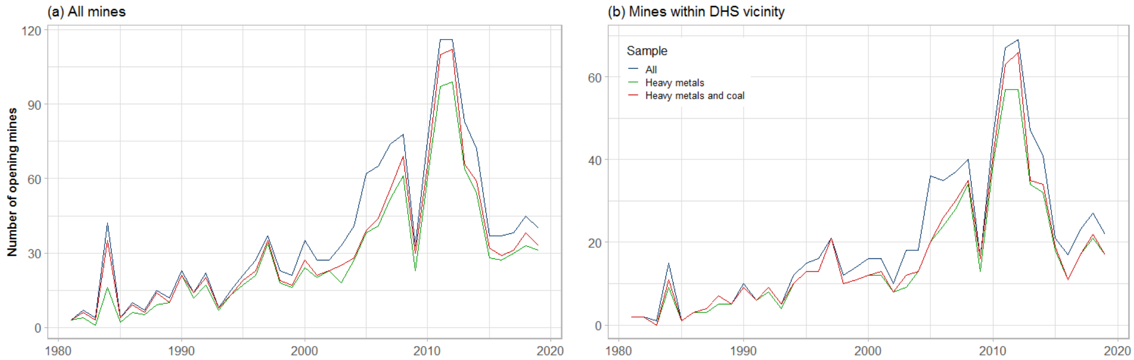
### 2.3.2 Descriptive statistics

#### 2.3.2.1 Mining in Africa

**Temporal and spatial variation**

Figure 2.1 shows the evolution of the yearly number of mines that opened in Africa over the 1981-2019 period, Figure 2.1 (a) for the entire mining sample while Figure

Figure 2.1: Temporal evolution of mine opening




**Notes:** The Figures plot the number of mines opening each year over the 1981-2019 period, for all mines, heavy metal mines including coal (sample of the main analysis), and only heavy metal mines. Figure (a) displays the temporal evolution of the total mine sample, while Figure (b) of mines that are within the sample of the main analysis, meaning mines that have DHS clusters upstream at most at 100km and DHS clusters downstream within the three closest sub-basins. *Sources:* Authors' elaboration on DHS and SNL data.

2.1 (b) for the mines that are in the sample of the main analysis. The mining boom since 2000 is captured in the Figures, with the first peak in 2007, in line with the peak in exploration activity that occurred in 2003 (Taylor et al., 2009) (as the exploration phase is on average a couple of years before a mine opens), and the second one in 2012. For instance, around 120 industrial mines opened in 2012 (based on the non-exhaustive SNL database). The Figures also distinguish the evolution according to the mines' characteristics: it distinguishes the pattern for all mines, heavy metal mines, and heavy metals including coal mines. We observe no differences in timing patterns between Figure 2.1 (a) and (b), neither between mine types.

What is striking in Figure 2.1 is that the evolution of mine openings follows the same pattern as the evolution of industrial metal prices, as plotted in Figure B.10 from Section B.2.2 in the Appendix. The mining boom since 2000 follows the increase in real prices of Copper, Tin, Lead, Aluminum, Zinc, Nickel, and other heavy metals, while the sharp fall around 2008/2009 corresponds to the financial crisis. Again, the local minimum around 2016 corresponds to the drop in commodity prices in June 2014 (Khan, Nguyen, and Ohnsorge, 2016; Glöser et al., 2017). This similar evolution suggests that heavy metal prices are good Instrument Variables for the variable year of mine opening, such as Berman et al., 2017; Bazillier and Girard, 2020 used in their analysis. In Section 2.9.3 we will use it as a proxy for production intensity.

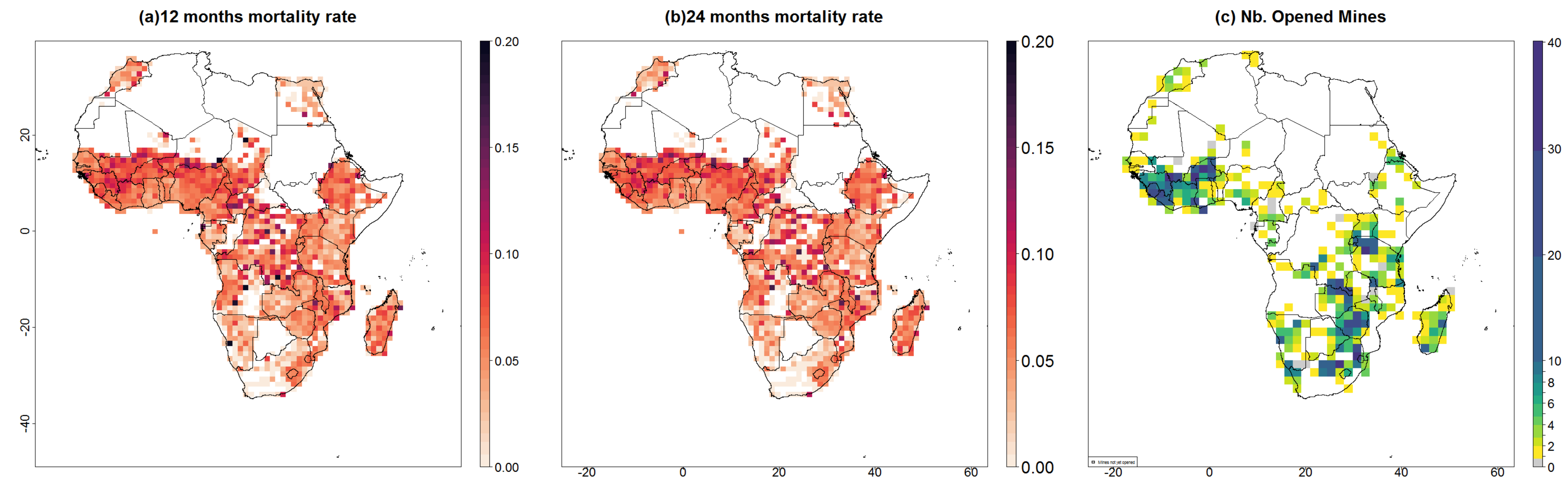
Figure 2.2 (c) shows the map of the number of mines that have opened before 2019, including mines that opened before 1986, averaged at the cell level (160 km cells). Cells in grey represent areas where no mine opened before 2019, but where at least one will open in the future (whether we know from the data that it has opened between 2019-2021, or if the opening is planned further). The main mining countries in the SNL database are Guinea, Sierra Leone, Ivory Coast, Ghana, Niger, Burkina Faso, Zimbabwe, Tanzania, Zambia, and the north of South Africa. Please note that, as we exclude countries with only one DHS wave in our main analysis' sample (cf Tables B.1 and B.2), to avoid comparing areas with too many differences in terms of temporal variations, we did not undertake the hand work for these countries, which explains why South Africa (which is not in the final sample) does not appear as a major mining country in Figure 2.2 (c). Figure 2.3 shows both the temporal and spatial variation of mine opening in Africa (for all the mines sample, and not the restricted one for our main analysis), as it plots the number of mines that opened over different periods of our analysis per grid cell. The cells in red are areas where no mines opened during the period, but where at least one mine has opened before, whereas cells in grey are areas where no mines have ever opened while at least one will open in the future. We observe that the increase in mine opening was higher during the third period 2008-2019 (which is in coherence with Figure 2.1), and was particularly important in West Africa.

#### 2.3.2.2 Health risks

Africa faces high infant mortality rates, as the average 12 months mortality rate is 6.4 % and the average 24 months mortality rate is 8.3% according to DHS data (cf Table B.3). Figures 2.2 (a) and (b) plot the average mortality rates for all DHS from 1986-2019 averaged at the grid level, and show the spatial variation of mortality rates [9]. Figures 2.4 and 2.5 map both spatial and temporal variation of mortality rates as it shows the average mortality rates for the three main periods of our DHS sample. We can observe the global reduction of mortality over the period and also the DHS cluster distribution. Figures B.12, B.13 and B.14 plot the same maps for the sample restricted to the one used in the main regression.

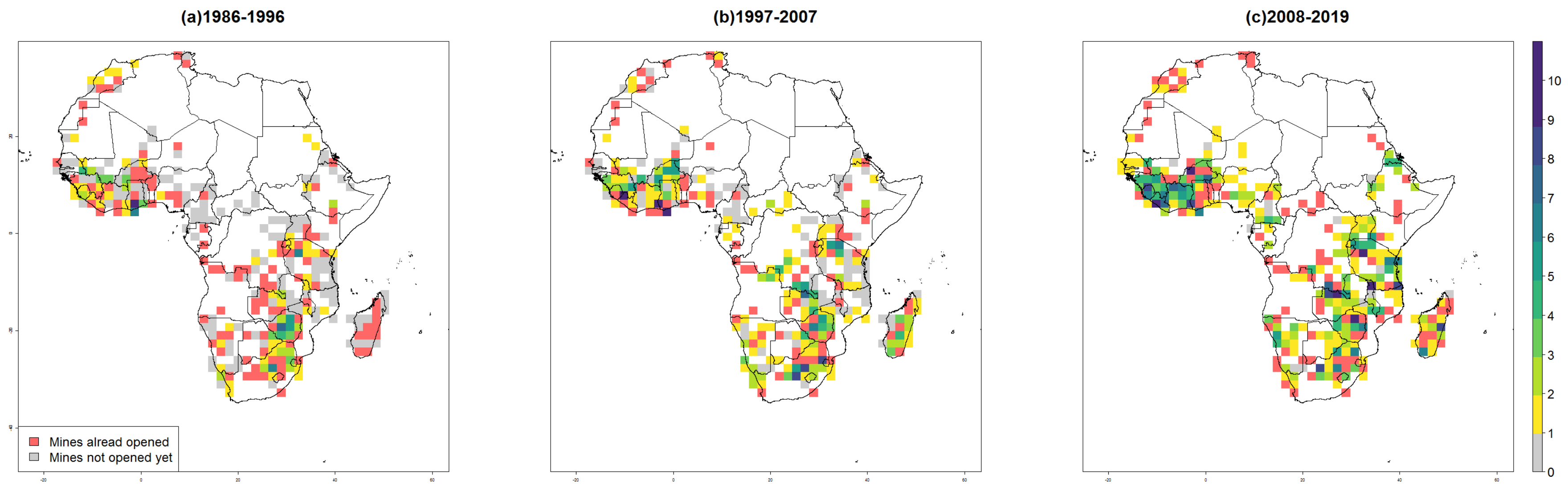
[9]Please note that the higher the DHS cluster density, the more accurate the average. The spatial variation is endogenous to the DHS sample.

Figure 2.2: Outcomes spatial distribution



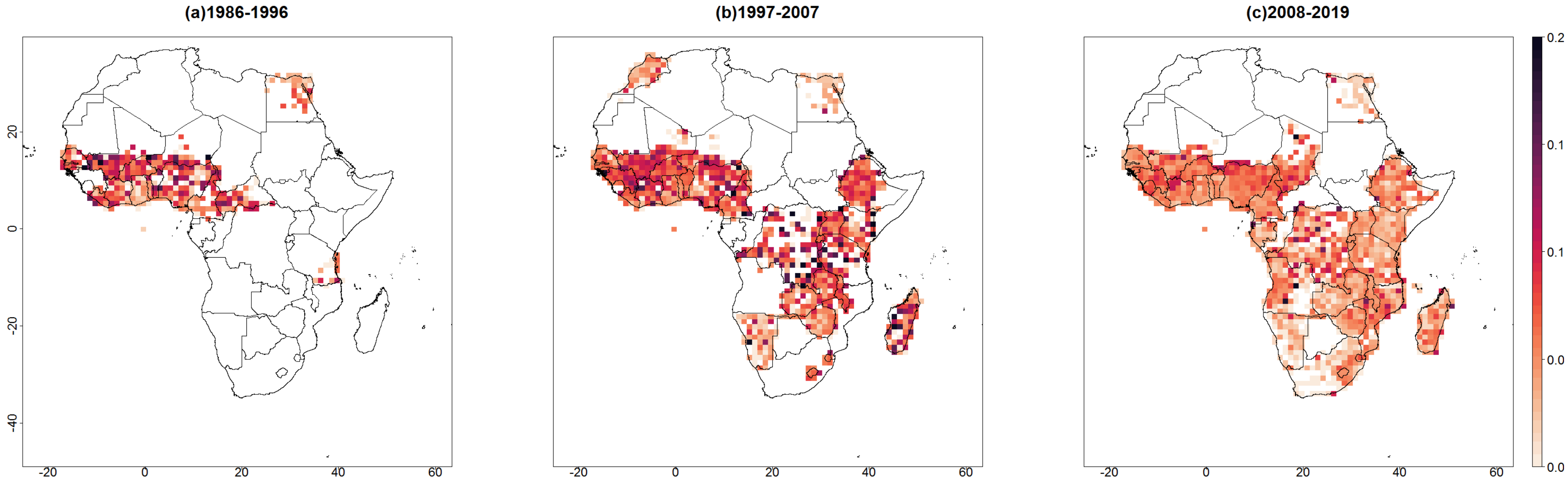

**Notes:** Figures (a) and (b) represent the means of 12- and 24-month mortality rates for each DHS wave available (listed in table B.2), from 1986 to 2019. Means are computed at the grid level (100km mean size). The mortality rates are estimated without the children that did not reach 12/24 months at the time of the survey. Figure (c) displays the stock of mines that opened before 2019 (including mines that opened before 1986). Means are computed at the grid level (100km mean size).
*Sources:* Authors' elaboration on DHS and SNL data.

Figure 2.3: Spatial variation of mine opening per period



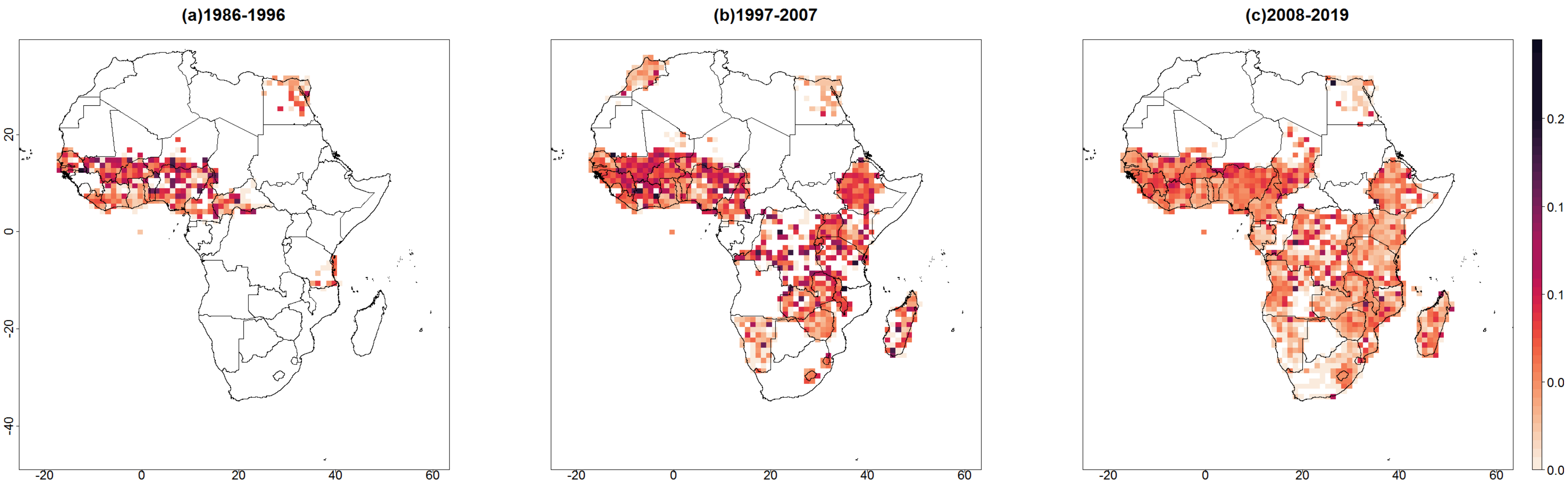

**Notes:** The figures represent the number of mines that opened during the periods over the grid area (160 km on average). A red grid cell represents an area where no mine opened over the period, but where at least one mine open before the period. A grey cell represents an area where no mine opened over the period, but where at least one mine will open in the future.
*Sources:* Authors' elaboration on SNL data.

Figure 2.4: Spatial variation of 12-month mortality rates per period




**Notes:** The figures represent the means of 12-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008, and (c) 2008-2019. The mortality rates are estimated without the children that did not reach 12 months at the time of the survey.
*Sources:* Authors' elaboration on DHS data.

Figure 2.5: Spatial variation of 24-month mortality rates per period




**Notes:** The figures represent the means of 24-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008, and (c) 2008-2019. The mortality rates are estimated without the children that did not reach 24 months at the time of the survey.
*Sources:* Authors' elaboration on DHS data.

## 2.4 Empirical strategy

The main empirical strategy of this paper uses the relative topographic position of sub-basins as a proxy for exposure to mining activity pollution. It compares the effects on the health of individuals living downstream to those living upstream of a mine, before and after the opening of at least one site. It is a staggered design Difference-in-Difference analysis with two-way fixed effects at the mine's sub-basin and birth year level. This upstream-downstream strategy intends to identify the mechanism of water pollution.

As seen in Section 2.2.3.1, this strategy alleviates some endogeneity issues raised by treatments using the Euclidian distance as a proxy for exposure to the mine. First, it reduces the bias linked to unbalanced samples due to endogenous pairing. Second, it breaks the average effects based on distance buffers and highlights the heterogeneity of the effects of mining activity on health, and isolates the negative externalities linked to water degradation.

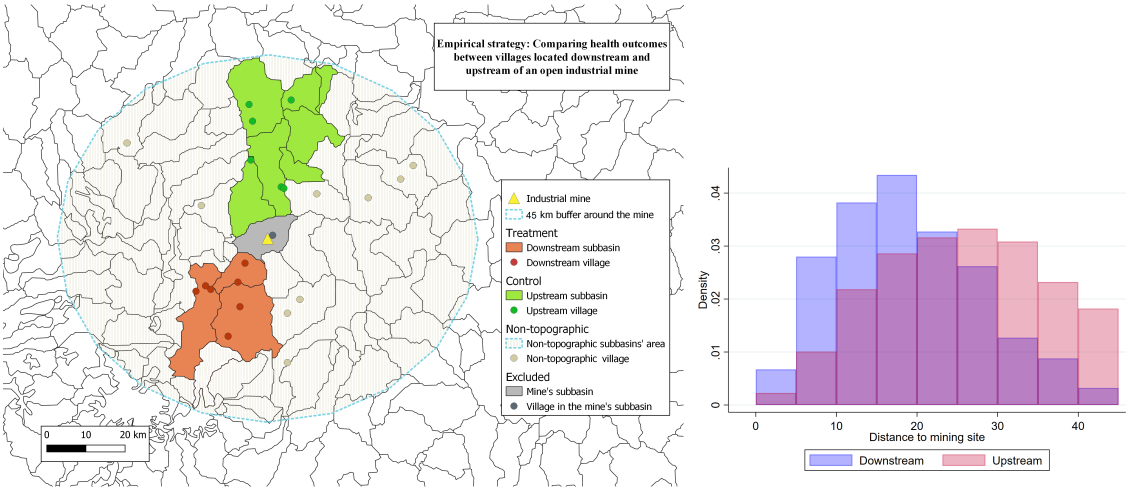
### 2.4.1 Measuring exposure to pollution

#### 2.4.1.1 Pairing strategy

The pairing of DHS clusters to mines represents a significant challenge, as each DHS cluster can be downstream of and close to several industrial sites in major mining areas. It introduces endogeneity in the sample selection and raises the issue of unbalanced samples. In this analysis, we propose the following pairing to overcome this issue and thus be able to measure the exposure to pollution of each DHS cluster.

First, we construct a 100 km buffer around each DHS cluster and register all mines within this buffer (independently of their activity status). We then categorize the topographic position of the DHS cluster relative to the industrial site, using a dummy equal to 1 if the cluster is downstream of the mine and 0 if it is located upstream. This topographic position is defined using the relative position of each sub-basin. As each cluster and sites have GPS coordinates, they lie in a specific sub-basin, and we used the relative position of each sub-basin to classify the DHS according to the paired mine. Through such a process, we also have pairs that are located in the same sub-basin, and for which it is impossible to say exactly whether the cluster is downstream or upstream of the mine. At this stage, for these specific couples,

Figure 2.6: Pairing Strategy




(a) Scheme for pairing strategy (b) Density of distance to the mine

**Notes:** Figure (a) is a scheme that illustrates the pairing, giving the example of a mine, its main sub-basin, its three closest downstream and upstream sub-basins, and DHS clusters that are in the treatment and control areas within 45 kilometers. Figure (b) plots the density of the distance (in km) to the mining site for DHS clusters across their upstream-downstream position.
*Sources:* Authors' elaboration on DHS, SNL, and HydroSheds data.

we consider the DHS to be downstream. Please note that, as explained in section 2.3.1.3, we used the finest Pfasftetter level 12 that breaks down sub-basins at an average area of $100km^2$ (the size of the sub-basin varies according to their shape, cf Figure 2.6a). At this stage, some villages can be paired with several mines and can have more than one occurrence in the sample. The difficulty of the strategy lies in choosing the mine that will be paired with the cluster.

Second, we restrict the group of downstream DHS clusters to the ones that lie within one of the three closest sub-basins downstream of the mine's sub-basin, to focus on the potentially most contaminated areas. Third, we pair each cluster with only one mine, proceeding as follows. If a DHS was in both groups (i.e downstream a mine A and upstream a mine B), then it is automatically assigned to the downstream group, and it is paired to the mine from which it is downstream (i.e it is paired to mine A), regardless of its activity status. At this stage, some clusters may still be counted twice, as they can be upstream of several mines, or in the three closest sub-basins downstream of several mines. To complete the uniqueness of the pairing, we paired each cluster to the nearest mine, regardless of its activity status as well.

In conclusion, the DHS clusters are attached to the nearest mine from which they are downstream up to the third sub-basin level, or else attached to the nearest mine upstream up to a radius of 100km. The final remaining problem relates to the clusters that are in the mine's same sub-basin, which we have so far identified as being downstream. We eliminated from the main analysis all DHS villages which are located in the same sub-basin of the mines from which they were paired. Also, this reduces the noise linked to the random displacement of DHS villages (cf Section 2.10.3.1) and avoids allocating villages as being downstream whereas they are upstream due to the displacement, as it drops the closest areas around the mine.

Once the pairing is done, we restrict the control group to upstream villages that are within 45 kilometers of the mine, to ensure the comparability of upstream and downstream villages. To choose this distance cut-off, we have calculated the mean of the maximal distance between a mine and the furthest extremity of its third downstream sub-basin, which was 44.7 kilometers. Figure 2.6b plots the distribution of the distance to the mine for both upstream and downstream villages. As downstream villages are prioritized in the pairing strategy, they are slightly closer to the mine, but the two distributions are comparable.

The pairing is illustrated in Figure 2.6a. It gives the example of a mine, its main sub-basin (grey), the downstream sub-basins (orange), and upstream sub-basins (green) up to 45 kilometers. The dashed area displays the sub-basins within 45 kilometers, with no topographic relationship to the mine, meaning they are neither downstream nor upstream. In the main strategy, we compare the villages within the green area to those in the orange area. In section 2.9 we run robustness tests checking whether the results hold allowing for further sub-basins and heterogeneity effects by distance to the mining site. In section 2.10.2.1 we discuss the results when including the non-topographic sub-basins.

### 2.4.2 Identification Strategy

#### 2.4.2.1 Main estimation

The main analysis relies on a Difference-in-Differences strategy using the topographic position of a DHS cluster relative to a mine deposit to indirectly identify the channel of water pollution. We propose a staggered Difference-in-Difference specification (DiD), with a sub-basin fixed effect panel for each mine. We isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison. We restrict our analysis using the pairing strategy explained in the previous section A.1. We compare health outcomes in upstream-downstream areas, both before and after the opening of the paired mine. The empirical strategy can be formally written as follows:

$$\begin{aligned} Death_{i,v,c,SB} = & \alpha_0 + \alpha_1 Opened_{birthyear,i,v} + \alpha_2 Downstream_{v,SB} \\ & + \alpha_3 Opened_{birthyear,i,v} \times Downstream_{v,SB} + \alpha_4 X_i \\ & + \gamma_S B + \gamma_{SB-trend} + \gamma_{c,birthyear} + \epsilon_v \end{aligned} \tag{2.1}$$

With $Death_{i,v,c,SB}$ a dummy equal to one if child $i$ from DHS village $v$ of country $c$, has reached the $n^{th}$ month and has died ($n$ being 12 for the 12-month old mortality, same for 24 months). $Opened_{birthyear,i,v}$ is a dummy equal to 1 if the mine, which is located in sub-basin $SB$, has opened before child $i$'s year of birth. $Downstream_{v,SB}$ is a dummy of relative position (equal to 1 if village DHS $v$ is located in a sub-basin downstream of the mine sub-basin $SB$, and 0 if it is upstream), $X_i$ a vector of child and mother level controls (mother's age, age square, years of education, urban residency). Finally, $\gamma_{SB}$ is a mine sub-basin fixed effect, $\gamma_{SB-trend}$ a mine sub-basin linear birthyear trend and $\gamma_{c,birthyear}$ a country-birthyear fixed effect. This analysis is a staggered design as the treatment shock (mine opening) does not occur at the same time for each DHS cluster.

The main regression is run without the DHS clusters that lie within the same sub-basin as the mine they are coupled with, as discussed in the previous section. The list of countries and survey years used in the main regression are given in Table B.2, and the list of metals in Table B.6.

#### 2.4.2.2 Identification assumption

The key assumption of a DiD is that the downstream group would have evolved as the upstream group in the absence of the opening of a mine. As we cannot test those upstream and downstream areas would have followed the same time trends, we test in Section 2.8 the common trend assumption using pre-treatment data.

However, the fact that pre-treatment data are parallel is neither a necessary nor a sufficient condition for the identification. Past trends can be identical but the upstream group may be affected by a group-specific shock during the period of the treatment. The estimation of this paper relies on the fact that the comparison between downstream and upstream villages is a proxy for exposure to water pollution. The major identification assumption is that the opening of a mine affects differently upstream and downstream areas only through the decrease in water quality. Throughout the paper, we will try to address the concerns of unobservable factors that might not be orthogonal to our treatment and Section 2.11 displays a final general discussion on the threats to the identifying assumption, and how they have been solved in the analysis.

### 2.4.3 Descriptive statistics

In this section, we describe the balance tables for the outcomes that play a key role in our analysis, out of parsimony.

Table 2.1: Balance Table

| | Before Mine Opening | | | | | After Mine Opening | | | | | Within Up. | Within Dwn. | Within |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Upstream | | Downstream | | Diff | Upstream | | Downstream | | Diff | | | |
| | N | Mean /(SD) | N | Mean /(SD) | (4-2) /(p.v) | N | Mean /(SD) | N | Mean /(SD) | (9-7) /(p.v) | (7-2) /(p.v) | (9-4) /(p.v) | (12-11) /(p.v) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| **Dth<12** | | | | | | | | | | | | | |
| All | 23,547 | 0.073 | 7,875 | 0.074 | 0.001 | 12,319 | 0.055 | 4,738 | 0.051 | -0.004 | -0.018 | -0.023 | -0.005 |
| | | (0.261) | | (0.262) | (0.83) | | (0.228) | | (0.219) | (0.256) | (0) | (0) | (0.468) |
| Mines | 244 | | 237 | | | 179 | | 183 | | | | | |
| **Dth<24** | | | | | | | | | | | | | |
| All | 17,726 | 0.096 | 5,928 | 0.098 | 0.002 | 8,664 | 0.068 | 3,330 | 0.072 | 0.004 | -0.028 | -0.026 | 0.002 |
| | | (0.294) | | (0.297) | (0.618) | | (0.252) | | (0.259) | (0.428) | (0) | (0) | (0.671) |
| Mines | 244 | | 236 | | | 168 | | 168 | | | | | |

**Notes:** Standard errors and p-values in parentheses. Descriptive statistics of 12- and 24-month mortality outcomes, for villages located upstream and downstream of mining sites, for individuals born before and after the opening of the mine.

Balance Table 2.1 compares the changes in infant mortality before and after the opening of a mine, for places upstream *vs* downstream of the mining site, following the pairing strategy. It displays also the number of individuals and paired mines in each group of the analysis. On average, upstream and downstream areas have non-significant differences in terms of 12 and 24-month mortality (columns 5 and 10). For both upstream and downstream clusters, the opening of a mine significantly decreases the mortality probability (columns 11 and 12), which is in line with the result of Benshaul-Tolonen, 2018, and with the fact that mortality rates decrease over time in Africa, as trends are not included (Figures B.12 and B.13). Table 2.1 shows that this reduction is overall slightly more important in upstream areas than in downstream areas for under 24-month mortality (0.002), while it is the contrary for under 12-month mortality (-0.005) but they remain not significant differences (column 13). Table 2.1 does not include any controls and is only descriptive. Table B.7 from Section B.3.1 in the Appendix replicates this exercise for control variables.

Figure B.11 in the Appendix identifies the country with the biggest stock of open mines in our sample (Ghana, Zimbabwe, and Tanzania with the highest density of open mines nearby DHS), as well as insights on the variation in mine opening over the period per country.

## 2.5 Main results

This section displays the results of our main analysis. The first section describes the overall effects of mining opening on child mortality among the villages living downstream compared to those living upstream. The second section displays the effects of being downstream of an open mine on other child's health outcomes, while the third section focuses on women's outcomes.

### 2.5.1 Child mortality

This section displays the main results of this paper from equation 2.1 for the 12-month mortality rate and the 24-month mortality rate. Table 2.2 gathers our main results with mine sub-basin and country-birth-year fixed effects. We also include mine sub-basin and birthyear linear trends, adjusting for spatial and period-specific cofounders and trends, and commodity fixed effects. Columns (1) to (4) give the results for the 12-month mortality rate, while columns (5) to (8) for the 24-month mortality rate. Columns (1), (2), (5), and (6) display the results for the total population while columns (3), (4), (7), and (8) focus on the rural population. Control variables are birth order, mother's age, mother's age square, the mother's years of education, urban, and the intensity of the river network. [10]. Even columns include the number of open mines within 45km of the DHS cluster as control, which controls for the mining density.

The results show that being downstream of an open mine increases by 2.18 percentage points (p.p) the 24-month mortality rate [11]. This corresponds to an increase by 25% as the average 24-month mortality increases from 8.7% to 10.9%. The results are higher in terms of magnitude in rural areas, as being downstream an open mine increases by 3.8 p.p the 24-month mortality rate, which is associated with an increase by 40%, as the mortality increases from 9.4% to 13.2%. This is in line with the fact that rural populations have less access to facilities and infrastructure and are more exposed to unsafe water. The results are not significant concerning the 12-month mortality rate, and are very close to zero, showing no difference between individuals leaving upstream to those leaving downstream. This lag in the effect of water pollution on children's health may be explained by the higher probability of children under 12 months to be breastfed compared to children under 24-month, hence their decreased exposure to contaminated water and limitation of direct ingestion (VanDerSlice, Popkin, and Briscoe, 1994; Fängström et al., 2008). This mechanism explaining the different results on the 12-month mortality *vs* 24-month mortality is explored in Section 2.6.3.

[10] The variable intensity of the river network using the HydroRIVERS product. It is a continuous variable, which takes into account the area of the catchment that contributes directly to a river reach, and the Strahler order of the specific river segment. In our sample, the Strahler spans from 3 to 10.

[11] 95 % Confidence interval: [0.000595; 0.042993]

Table 2.2: Effects of industrial mining opening on child mortality

| | 12-month mortality | | | | 24-month mortality | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Population | | Rural Population | | Total Population | | Rural Population | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Down×Open | -0.00352 | -0.00506 | 0.00517 | 0.00510 | 0.0231** | 0.0218** | 0.0379*** | 0.0380*** |
| | [0.00824] | [0.00831] | [0.0102] | [0.0102] | [0.0105] | [0.0108] | [0.0130] | [0.0130] |
| Downstream | -0.0140** | -0.0152** | -0.0203*** | -0.0204*** | -0.0202*** | -0.0211*** | -0.0287*** | -0.0284*** |
| | [0.00655] | [0.00665] | [0.00743] | [0.00762] | [0.00731] | [0.00739] | [0.00795] | [0.00810] |
| Open | 0.0121* | 0.00963 | 0.0106 | 0.0102 | -0.00302 | -0.00496 | -0.00650 | -0.00588 |
| | [0.00722] | [0.00754] | [0.00858] | [0.00952] | [0.00986] | [0.0101] | [0.0115] | [0.0122] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Nb open mines | No | Yes | No | Yes | No | Yes | No | Yes |
| Birthmth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ctry-bthyr FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| MineSB-bthyr trd | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 48,472 | 48,472 | 33,231 | 33,231 | 35,638 | 35,638 | 24,544 | 24544 |
| R2 | 0.0378 | 0.0378 | 0.0476 | 0.0476 | 0.0511 | 0.0511 | 0.0633 | 0.0633 |
| Outcome Mean | 0.0666 | 0.0666 | 0.0716 | 0.0716 | 0.0873 | 0.0873 | 0.0945 | 0.0945 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. The variables Downstream and Opened are dummies that indicate whether an individual lives in a village downstream of at least one mining site and whether the site opened before the year of birth. Each DHS village is paired to only one mining site so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Columns (1-3) give the results for the total population while columns (4-6) display the results for rural villages. Columns (2, 4, 6, 8) control for the number of open mines within 45 km. Control variables are birth order number, mother's age, mother's age square, mother's years of education and urban, number of open mines, and a continuous variable indicating the presence of rivers and their order.

### 2.5.2 Other health effects

Table 2.3 represents the effect of industrial mining on other children's health outcomes than mortality. Columns (1-3) display the results on anthropometric measures of children who were still living at the time of the survey and are measured at the time of the survey. A child is affected by stunting if her height-for-age z-score is below minus 2 standard deviations below the mean on the World Health Organization Child Growth Standards. The same definition applies to underweight (weight-for-age) and wasting (weight-for-height). We find a negative and significant effect of industrial mining on underweight but not on stunting and wasting: living downstream of an open mine decreases wasting by 3.9 pp. This result could potentially be explained by the death of the most vulnerable children and the survival of the heaviest ones. We find no results on other diseases among living children: anemia (measured), diarrhea, cough, or fever (reported within the two weeks preceding the interview). We find no effect either of industrial mining on low weight at birth (below 2.5 kg) nor reported size at birth (reported as small or very small by the mother).

Table 2.3: Effects of industrial mining opening on other child health outcomes

| | Surviving children | | | | | | | All births | |
|---|---|---|---|---|---|---|---|---|---|
| | Stunting | Underweight | Wasting | Anemia | Diarrhea | Cough | Fever | $< 2.5kg$ | Small |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Down×Open | -0.0167 | -0.0389** | -0.00479 | -0.0265 | 0.00146 | -0.00824 | -0.00219 | -0.00925 | -0.00337 |
| | [0.0199] | [0.0169] | [0.0109] | [0.0277] | [0.0130] | [0.0176] | [0.0154] | [0.0180] | [0.0120] |
| Downstream | -0.0126 | -0.00293 | 0.00330 | 0.0428** | 0.00287 | -0.0115 | 0.0141 | -0.00951 | 0.00673 |
| | [0.0172] | [0.0152] | [0.00897] | [0.0188] | [0.0108] | [0.0138] | [0.0144] | [0.0163] | [0.0107] |
| Open | -0.00618 | 0.0257* | 0.00769 | 0.00582 | -0.00545 | 0.00547 | -0.00571 | 0.00571 | 0.0191 |
| | [0.0162] | [0.0146] | [0.0106] | [0.0246] | [0.0111] | [0.0124] | [0.0127] | [0.00954] | [0.0159] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ctry-bthyr FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| MineSB-bthyr trd | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 37,393 | 37,043 | 37,903 | 19,331 | 55,162 | 54,958 | 54,955 | 29,162 | 58,338 |
| R2 | 0.155 | 0.124 | 0.0895 | 0.215 | 0.0900 | 0.104 | 0.111 | 0.0608 | 0.0532 |
| Outcome mean | 0.308 | 0.246 | 0.0893 | 0.660 | 0.165 | 0.237 | 0.246 | 0.171 | 0.157 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. Columns (1-3) focus only on surviving children (due to variable construction in DHS), while the others encompass all children, including those who died before the survey. The same sample and controls as Table 2.2 Column 2 apply.

### 2.5.3 Women's outcomes

We make sure that the results found on child mortality are not due to a change of fertility among women [12]. We find no significant effect of industrial mining neither on whether women ever had a child (Table 2.4 column 1) nor on the total number of children she had (column 2). We find no effect either on whether women were pregnant during the time of the survey (column 3). Table 2.4 also displays results on women's other health outcomes: we find no effect of industrial mining on neither whether women ever had a miscarriage or on their anemia status.

Table 2.4: Effects of industrial mining opening on women outcomes

| | Fertility | | | Health | |
|---|---|---|---|---|---|
| Outcome | Ever had a child | Total lifetime fertility | Currently pregnant | Ever had a miscarriage | Anemia |
| | (1) | (2) | (3) | (4) | (5) |
| Down × Open | 0.0156<br>[0.00977] | -0.0164<br>[0.0720] | -0.0171<br>[0.0111] | -0.00507<br>[0.0138] | 0.000806<br>[0.0261] |
| Downstream | 0.00886<br>[0.00952] | 0.0841<br>[0.0731] | 0.0160<br>[0.0118] | 0.00894<br>[0.0134] | -0.00928<br>[0.0233] |
| Open | -0.00161<br>[0.00916] | 0.0663<br>[0.0595] | 0.00607<br>[0.00833] | -0.00136<br>[0.0119] | -0.00417<br>[0.0285] |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Ctry-survey year FE | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes |
| MineSB-svey year trd | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes |
| N | 82,406 | 82,406 | 82,373 | 72,423 | 31,587 |
| R2 | 0.510 | 0.659 | 0.0422 | 0.0906 | 0.122 |
| Outcome mean | 0.737 | 2.912 | 0.0939 | 0.136 | 0.396 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Control variables are birth order number, woman's age, woman's age square, woman's years of education, urban, number of open mines and presence of rivers.

[12] The analysis has been made using DHS Women Recode, which population sample is all women aged 15-49 years old.

## 2.6 Mechanisms

### 2.6.1 Households' access to water and facilities

We deepen our analysis by studying whether the effects found on children's mortality are indeed due to water pollution downstream of mines and not driven by improved access to water and sanitation or facilities upstream. Under 24-month mortality is still increased significantly by 2 p.p. when adding the triple interaction with several facilities variables: whether a household has piped water as the main drinking source (Table 2.5 column 1), whether a household has a flushed toilet (column 2), whether it has access to electricity (column 3), and whether the mother had visited health facilities during the 12 months preceding the survey (column 4). We find no significant heterogeneity across the four facilities, which suggests that our result is not explained by an improvement of facilities upstream, which contradicts the findings of Dietler et al., 2021. Table B.8 in Section B.3.1 of the Appendix looks at the DiD estimator using the access to piped water and electricity as dependent variables and shows no difference after the opening of a mine between upstream and downstream villages.

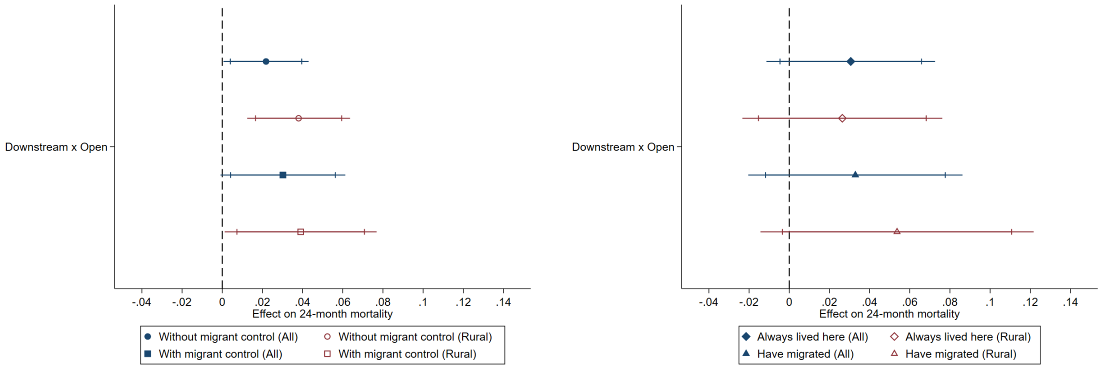
Table 2.5: Effects of industrial mining opening on access to water, sanitation and facilities

| Outcome | 24-month mortality | | | |
|---|---|---|---|---|
| Var. | Has piped water | Has flushed toilet | Has electricity | Visited health facilities |
| | (1) | (2) | (3) | (4) |
| Downstream×Open × Var. | 0.000283 | -0.0356 | -0.0114 | -0.0140 |
| | [0.0195] | [0.0263] | [0.0192] | [0.0165] |
| Downstream×Open | 0.0210* | 0.0236** | 0.0243** | 0.0292* |
| | [0.0118] | [0.0110] | [0.0115] | [0.0159] |
| Var. | -0.00266 | -0.0165* | -0.0157** | -0.00491 |
| | [0.00637] | [0.00920] | [0.00695] | [0.00541] |
| Downstream | -0.0229*** | -0.0215*** | -0.0223*** | -0.0170* |
| | [0.00770] | [0.00744] | [0.00764] | [0.00983] |
| Open | -0.000958 | -0.00482 | -0.00715 | -0.00543 |
| | [0.0103] | [0.0101] | [0.0103] | [0.0122] |
| Controls | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes |
| N | 35,638 | 35,536 | 35,423 | 32,018 |
| R2 | 0.0512 | 0.0513 | 0.0512 | 0.0512 |
| Outcome mean | 0.0873 | 0.0873 | 0.0873 | 0.0857 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The same sample and controls as Table 2.2 Column 2 apply.

### 2.6.2 Migration

Figure 2.7: Migration analysis




(a) With and without migrant control

(b) Migrant/Always lived here samples

**Notes:** Figure (a) plots the coefficients associated with being downstream of an open mine on under 24-month mortality when controlling for mothers' migration status or not, across the whole rural sample. Figure (b) plots the coefficients of the same interaction but across the sample of mothers who have migrated or always lived here. *Sources:* Authors' elaboration on DHS and SNL data.

We pursue our analysis by making sure that our results on child mortality are not due to migration and do not suffer from selection bias. The migration information is retrieved from the variable indicating whether mothers have ever migrated to the actual place of residency, or if they have always lived there. The information is available among 60% of our sample (cf. Table B.4) and controls for in-migration, which is an important effect of the opening of a mine that attracts new working populations (cf Section B.3.1 for more discussion on bias linked to migration). Figure 2.7 displays the coefficient associated with the interaction term of being downstream of an open mine. The top coefficient in Figure 2.7a is our main specification when we do not control for migration, across the whole. We plot the same focusing on the rural sample. The bottom two coefficients are when we control for migration, across our whole and rural sample. We find that all are statistically positive and significant. We further our analysis by splitting the sample across mothers who have ever migrated and mothers who have always lived here (Figure 2.7b). The estimation suffers from a lack of statistical power (see Appendix Table 2.6 for the drop of observations) but suggests no differential effect of industrial mining across the two samples.

Table 2.6: Effects of industrial mining activity, migration analysis

| Outcome | Mortality under 24 months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Spec. | Without migrant control | | With migrant control | | Migrant sample | | Always lived here | |
| Sample | All | Rural | All | Rural | All | Rural | All | Rural |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Downstream x Open | 0.0218** | 0.0380*** | 0.0302* | 0.0390** | 0.0329 | 0.0537 | 0.0306 | 0.0264 |
| | [0.0108] | [0.0130] | [0.0158] | [0.0193] | [0.0272] | [0.0347] | [0.0214] | [0.0254] |
| Downstream | -0.0211*** | -0.0284*** | -0.0249** | -0.0332*** | -0.0229 | -0.0304 | -0.0301** | -0.0418*** |
| | [0.00739] | [0.00810] | [0.0106] | [0.0119] | [0.0190] | [0.0226] | [0.0139] | [0.0147] |
| Open | -0.00496 | -0.00588 | -0.0255 | -0.0235 | 0.0116 | 0.0187 | -0.0409** | -0.0425* |
| | [0.0101] | [0.0122] | [0.0159] | [0.0194] | [0.0260] | [0.0335] | [0.0198] | [0.0256] |
| Migrant | | | 0.00850* | 0.00348 | | | | |
| | | | [0.00449] | [0.00594] | | | | |
| N | 35638 | 24544 | 22231 | 15060 | 8658 | 6007 | 13503 | 8982 |
| R2 | 0.0511 | 0.0633 | 0.0634 | 0.0770 | 0.112 | 0.132 | 0.0797 | 0.107 |
| Outcome mean | 0.0873 | 0.0945 | 0.0946 | 0.104 | 0.0892 | 0.102 | 0.0983 | 0.107 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. The same sample and controls as Table 2.2 Column 2 apply.

### 2.6.3 Early life characteristics

We want to better understand the mechanism behind the results on 12- and 24-month mortality, and focus on children's nutrition and access to health care as the main potential explaining factors. Table 2.7 gives the results of several triple interactions looking at the heterogeneity of the effect according to early life characteristics.

We find a significant increase in mortality among children living downstream of an open mine and who were given plain water 24 hours before the survey for both the 12-month mortality and 24-month mortality. For individuals living downstream of a mine that opened and who consumed plain water, the 12-month mortality increases by 5.3 p.p. and the 24-month mortality increases by 6.5 p.p. (Columns (1) and (2)). Unfortunately, the DHS variable does not specify the source of plain water, and we cannot show that the given plain water is more polluted downstream than upstream. We find no significant effect of the triple interaction with breastfeeding behaviors on mortality: whether the child was ever breastfed (columns 3 and 4) or number of months during which the child was breastfed (columns 5 and 6). This absence of result can be explained by the low variability as 98 % of the children in the sample were breastfed. There is a larger variability of plain water consumption, as it was the case for 18.7% of children. We thus interpret the fact that drinking plain water as a child is a proxy for having non-exclusive breastfeeding [13].

We find no significant effect in either of the triple interaction with access to health care: whether the mother received prenatal care (columns 7 and 8) or whether the child was ever vaccinated (columns 9 and 10).

[13] DHS questionnaire asks whether the children have been breastfed and have consumed plain water during the 24 hours before the survey

Table 2.7: Effects of industrial mining opening on explaining factors 12 vs. 24 months

| | Child's nutrition | | | | | | Child's access to health care | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Var. | Was given plain water | | Ever breastfed | | Breastfeed months | | No prenatal care | | Ever vaccinated | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Mortality outcome | 12m | 24m | 12m | 24m | 12m | 24m | 12m | 24m | 12m | 24m |
| Downstream× Open × Var. | 0.0530** | 0.0650* | 0.000597 | -0.0450 | 0.00106 | 0.00106 | 0.0165 | 0.0288 | 0.0114 | 0.0134 |
| | [0.0244] | [0.0350] | [0.0519] | [0.0481] | [0.00176] | [0.00210] | [0.0272] | [0.0402] | [0.0191] | [0.0231] |
| Downstream | -0.0000982 | -0.00344 | -0.0104 | -0.0161 | -0.0125 | -0.0293 | -0.0140** | -0.0200** | 0.0117 | 0.00258 |
| | [0.00760] | [0.0138] | [0.0250] | [0.0305] | [0.0237] | [0.0300] | [0.00613] | [0.00871] | [0.0127] | [0.0126] |
| Open | 0.0104 | -0.00951 | 0.0172 | 0.0163 | -0.0785*** | -0.134*** | 0.00842 | 0.000268 | 0.00433 | 0.00929 |
| | [0.00861] | [0.0189] | [0.0266] | [0.0345] | [0.0231] | [0.0324] | [0.00770] | [0.0117] | [0.0105] | [0.0175] |
| Downstream×Open | -0.0131 | 0.0193 | 0.00423 | 0.0702 | -0.0213 | -0.000897 | -0.0114 | -0.000663 | -0.0119 | -0.0105 |
| | [0.0103] | [0.0214] | [0.0523] | [0.0488] | [0.0373] | [0.0516] | [0.00823] | [0.0121] | [0.0178] | [0.0211] |
| Var. | 0.0338*** | 0.0277** | -0.921*** | -0.880*** | -0.0179*** | -0.0201*** | 0.0299*** | 0.0398*** | -0.0201*** | -0.0257*** |
| | [0.0102] | [0.0137] | [0.0115] | [0.0134] | [0.000595] | [0.000657] | [0.00774] | [0.0110] | [0.00728] | [0.00828] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 19,671 | 10,797 | 45,168 | 33,022 | 29,015 | 18,323 | 31,656 | 19,543 | 17,372 | 13,638 |
| R2 | 0.0735 | 0.102 | 0.208 | 0.174 | 0.330 | 0.355 | 0.0558 | 0.0822 | 0.239 | 0.306 |
| Outcome mean | 0.0396 | 0.0758 | 0.0493 | 0.0694 | 0.0466 | 0.0768 | 0.0479 | 0.0639 | 0.00835 | 0.0121 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$. The same sample and controls as Table 2.2 Column 2 apply.

## 2.7 Heterogeneity

### 2.7.1 Individual characteristics

Table 2.8: Effects of industrial mining opening across children's location and gender

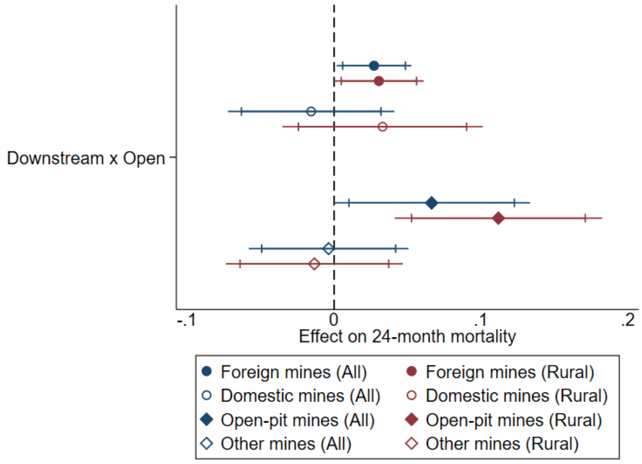
| Outcome | 24-month mortality | | | | | |
|---|---|---|---|---|---|---|
| Sample | All: urban and rural | | | Rural | | |
| | All | Girls | Boys | All | Girls | Boys |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Downstream× Open | 0.0218** | 0.0120 | 0.0334** | 0.0380*** | 0.0204 | 0.0677*** |
| | [0.0108] | [0.0151] | [0.0167] | [0.0130] | [0.0186] | [0.0199] |
| Downstream | -0.0211*** | -0.0203* | -0.0206* | -0.0284*** | -0.0292** | -0.0292** |
| | [0.00739] | [0.0114] | [0.0111] | [0.00810] | [0.0126] | [0.0117] |
| Open | -0.00496 | 0.00364 | -0.0178 | -0.00588 | 0.00419 | -0.0200 |
| | [0.0101] | [0.0155] | [0.0143] | [0.0122] | [0.0186] | [0.0181] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 35,638 | 17,452 | 18,142 | 24,544 | 12,009 | 12,481 |
| R2 | 0.0511 | 0.0758 | 0.0762 | 0.0633 | 0.0942 | 0.0972 |
| Outcome mean | 0.0873 | 0.0805 | 0.0938 | 0.0945 | 0.0883 | 0.101 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$. The same sample and controls as Table 2.2 Column 2 apply.

We conduct a heterogeneity analysis across children's location and gender (Table 2.8). We find that being downstream of an open mine is more critical in rural areas, (Column 4) as it increases the 24-month mortality by 3.8 p.p, which corresponds to a 40 % increase in the mortality rates. The heterogeneity by gender shows that our results are mainly driven by the mortality of males (columns 3 and 6), which remains consistent in rural areas.

### 2.7.2 Mining activity's characteristics

Figure 2.8: Heterogeneity across mines' characteristics




**Notes:** This graph represents the coefficients associated with the interaction of living downstream of an open mine when splitting the sample between mines that are owned by foreign companies and mines that are owned by at least one domestic company (in blue) and between mines that are open-pit and not open-pit (underground, placer, and in-situ leach) (in red).
*Sources:* Authors' elaboration.

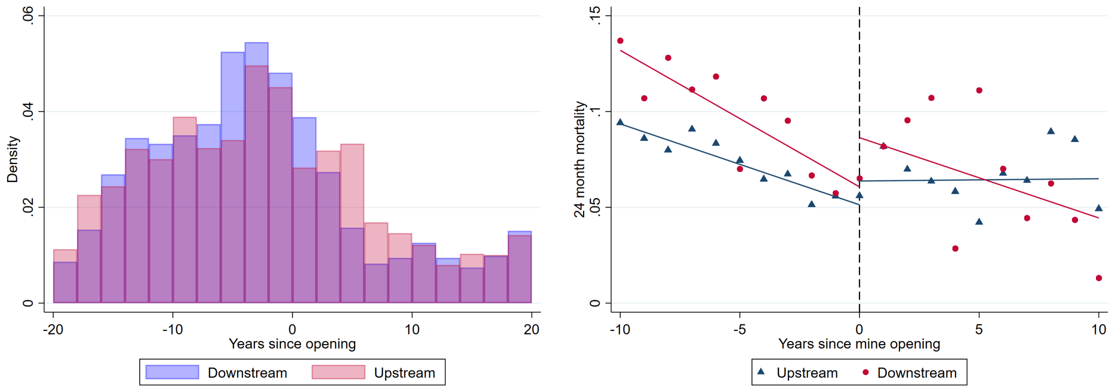
We pursue the heterogeneity analysis across mines' ownership and extraction methods. A mine was considered domestic if at least one of the owning companies is from the same country as the country of location, and they represent 17.8 percent of our mine sample. Figure 2.8 represents the coefficients associated with the interaction term of living downstream of an open mine. We find no effect of mine opening when we restrict the sample to domestically owned mines whereas our results hold when we restrict to the foreign-owned only mines (in blue). This could potentially be explained by improved management of a mine or a better consideration of the surrounding populations if a national company is involved. We then look at the open-pit nature of the industrial site, which concerns 21.6 percent of our mine sample. We find that our results hold when restricting to open-pit mines but not when we only look at the sample of other extraction methods (underground, placer, and in-situ leach) (in red). This is consistent with the fact that open-pit mines are the most polluting mines due to the generation of large amounts of waste kept in tailing storage facilities.

## 2.8 Dynamic effects

In this section, we investigate the dynamic effects of the opening of an industrial mine, looking at pre-trends and at whether the effects on 24-month mortality occur within a short or long time, and during the mining activity.

### 2.8.1 Pre-trends and event-study

Figure 2.9: Linear trends of 24-month mortality



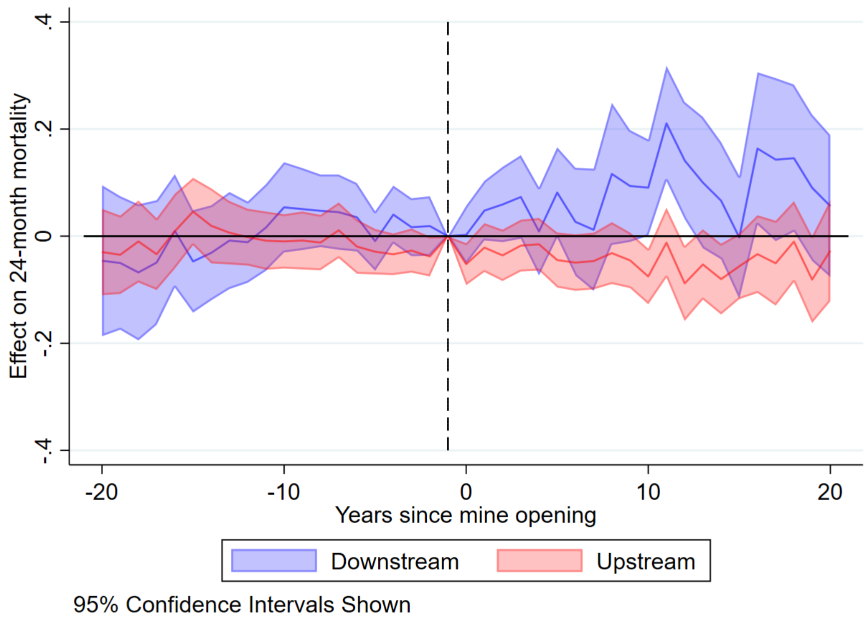

(a) All sample (b) All sample

**Notes:** Figure (a) gives the distribution of the number of observations per opening year. Figure (b) plots the trends of the 24-month mortality rates according to the year of opening. The figures are made for the whole sample and include neither control variables nor fixed effects. The reference point is -1, the year before the mine opening.
*Sources:* Authors' elaboration on DHS and SNL data.

The key assumption of the DiD strategy is that the outcome - the 24-month mortality - would follow the same time trend in the absence of the mine opening both in upstream and downstream areas. The common trends assumption cannot be tested. However, we can observe the pre-treatment data and the evolution of mortality rates before each mine opening according to the topographic position. Figure 2.9b plots the linear trends of the 24-month mortality rates and distinguishes between upstream and downstream DHS clusters, before and after the opening of the paired mine. For each year, it plots the average mortality rates over the sample, with no control nor fixed effect. Figure 2.9a plots the distribution of the years before and after the mine opening. Figure 2.9b shows non-exact parallel trends but is only descriptive, and we can see looking at the scatter plot that the downstream and upstream areas seem to follow a similar pattern of decreasing mortality before a mine opens. This decreasing pattern is triggered by temporal trends, as the years closest to the mine opening are more likely to be recent years, and the mortality rates are overall decreasing over the recent decades (cf Figure 2.5). This pattern is corrected in Figure 2.10.

Figure 2.10 plots the event study of the effect of mine opening, for both the upstream and downstream samples. It includes the same controls and fixed effects as the main analysis (cf Table 2.2), and thus corrects for previous trends as we include country and mine sub-basin trends. Both upstream and downstream villages do not display any pre-trends, which suggests that the common trend assumption is verified. We observe almost no effect of a mine opening on the mortality rates upstream, a slight decrease which is significant 10 years after the opening. Figure 2.10 shows that the infant mortality downstream increases once the mine opens. This effect is significant in the medium run, around a decade after the mine opening.

Figure 2.10: Event study - dynamic effect of mine opening on 24-month mortality



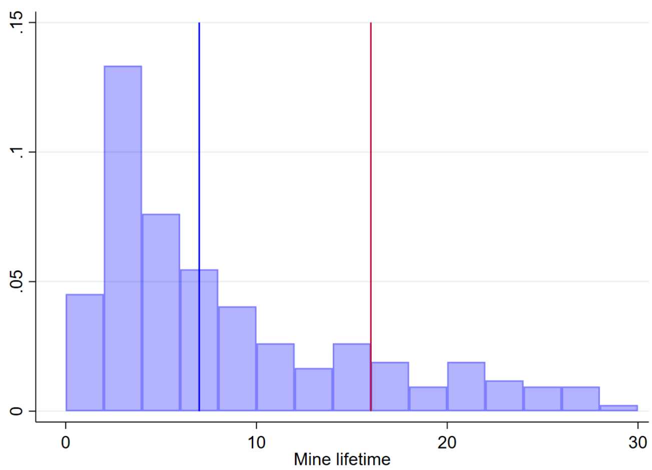

**Notes:** This Figure plots the event study of the effect of mine opening for downstream and upstream DHS villages, 10 years before the mine opening and 10 years after. The year before the mine opening, -1, is taken as the reference point.
*Sources:* Authors' elaboration on DHS and SNL data.

### 2.8.2 Mine closure

In the main analysis, we focus on mine opening without taking into account the closure year, since it is a piece of information harder to retrieve by hand. In this section, we focus on a restricted sample of mines for which the SNL database provides directly this information, to understand whether the closing date plays a role in our main effect.

Figure 2.11 gives the distribution of mines' lifetime, for the restricted sample of mines for which closing years are available. On average, a mine lasts 16 years, but the distribution is skewed to the right, and the majority of mines close before 10 years. Please note that the closing date available in the SNL database for this restricted sample is not exact. Over its lifetime, a mine can be put on hold several times, for political or economic reasons. In Table 2.9 we look at the effects of being downstream of a mine that is active at the year of birth of the child. Columns (1) and (2) show no effect on the 12-month mortality rates. Column (3) shows that being downstream of a mine that is active the year of birth increases the mortality rate by 4 p.p, which corresponds to an increase of 40% in the mortality rate. This result suggests that the harmful effects of mining activity on the individuals living downstream are mainly critical while the mine is active.

Figure 2.11: Distribution of a mine lifetime




**Notes:** This figure gives the distribution of mines' lifetime. The red line y=16 plots the mean of a mine lifetime, while the blue line y=7 plots the median. The maximum of a mine lifetime in our sample is 138 years.
*Sources:* Authors' elaboration on SNL data.

Table 2.9: Average effects of mine activity on infantile mortality

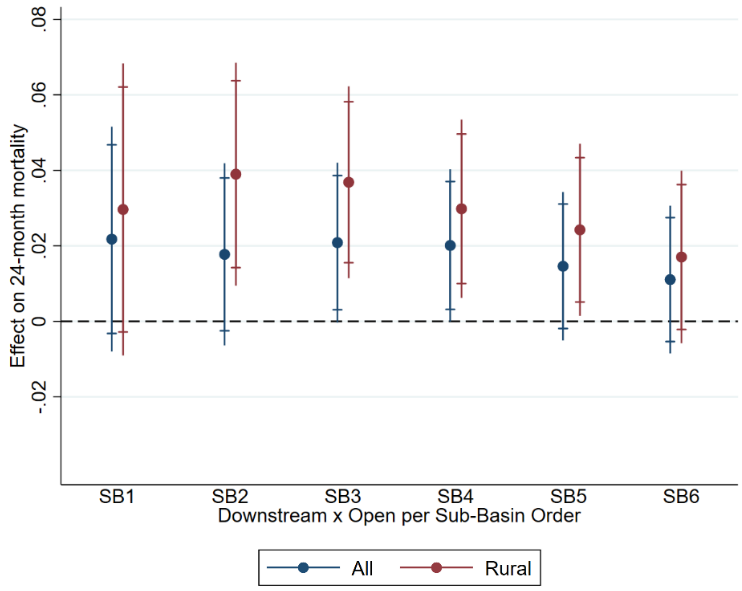
| | Mortality under 12 months | | Mortality under 24 months | |
|---|---|---|---|---|
| | All | Rural | All | Rural |
| | (1) | (2) | (3) | (4) |
| Downstream×Active | 0.0112 | 0.0321 | 0.0409* | 0.0488** |
| | [0.0222] | [0.0254] | [0.0248] | [0.0242] |
| Downstream | -0.0264* | -0.0434** | 0.00877 | -0.00721 |
| | [0.0139] | [0.0184] | [0.0145] | [0.0184] |
| Active | 0.000600 | 0.00420 | -0.00842 | -0.00345 |
| | [0.0140] | [0.0173] | [0.0172] | [0.0193] |
| Controls | Yes | Yes | Yes | Yes |
| Nb open mines | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes |
| N | 7,231 | 5,589 | 5,270 | 4,082 |
| R2 | 0.0899 | 0.0960 | 0.0984 | 0.108 |
| Outcome Mean | 0.0756 | 0.0825 | 0.0981 | 0.104 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. The variables Downstream and Active are dummies that indicate whether the individual lives in a village downstream of at least one mining site and whether the site is active during the year of birth. The same sample and controls as Table 2.2 Column 2 apply.

## 2.9 Intensive Margin

### 2.9.1 Spatial intensive margin

Figure 2.12: Effect of industrial mine opening according to the downstream sub-basin order

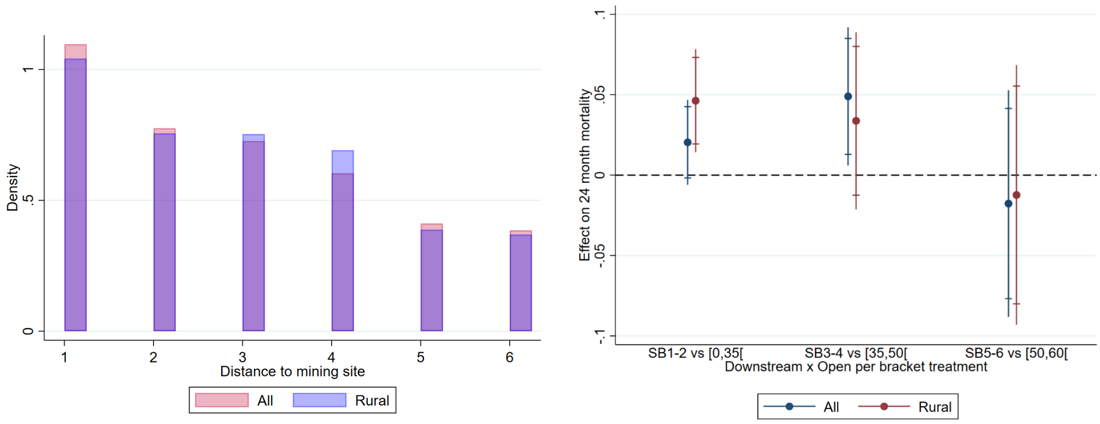



**Notes:** This Figure plots the treatment effect when changing the treatment group. SB 1 gives the DiD estimator when the control group includes DHS villages up to the first sub-basin, SB 2 up to the second, and then SB6 up to the sixth sub-basin. SB 3 gives the main result of the paper.
*Sources:* Authors' elaboration on DHS and SNL data.

In this section, we change the cut-off for being treated and test the effect on different orders of downstream sub-basins. In Figure 2.12, we test whether the effect holds when allowing for further sub-basins downstream. It plots the coefficient on *Downstream* $\times$ *Open* for six different models. SB1 corresponds to the model where the treatment group includes only the first neighboring downstream sub-basin, SB2 up to the second, and SB6 up to the sixth. SB3 gives the main results from Column (6) in Table 2.2. The Figure shows an attenuation of the magnitude of the effect when including further sub-basins. For all the individuals, the results are significant at the 5% level up to the third sub-basin and up to the fourth, while for all rural areas, it is significant from the second up to the fifth sub-basin. We interpret the non-significance of the result in the first sub-basins as being the consequence of statistical power (as the sample size is relatively low up to SB1 and SB2).

Figure 2.13 looks at the effect per distance brackets. To build the control group for each sub-basin, we determined which distances correspond to which sub-basin order.

Figure 2.13: Intensive margin - Effect of the number of mine opening on under 24 months mortality




(a) (b)

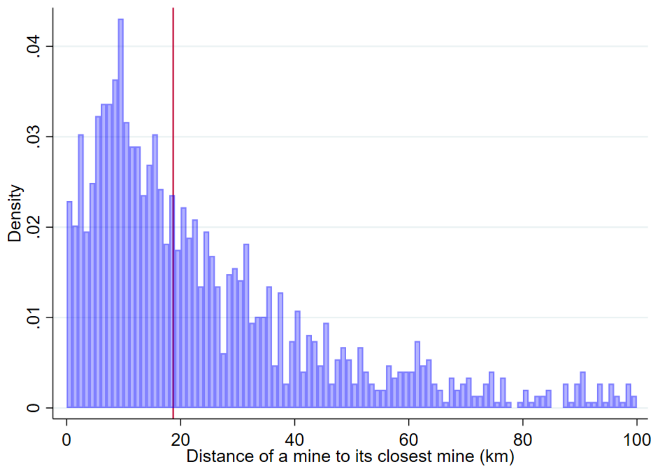
**Notes:** Figure (a) plots the distribution of observations downstream per sub-basin order. Figure (b) plots the interaction term on the 24-month mortality rates per distance brackets. The first coefficient on the left gives the effect for individuals living within the first and second sub-basins compared to individuals living up to 35 kilometers. The second coefficient gives the effect for individuals living downstream within the third and fourth sub-basins compared to those living upstream between 35 and 50 kilometers. The third coefficient gives the effect for individuals living downstream within the fifth and sixth sub-basins compared to those living upstream between 50 and 60 kilometers. The distances are chosen based on the mean of the distance between the mine and the extremity of the XXth sub-basin. [a]
*Sources:* Authors' elaboration on DHS and SNL data.

[a]The mean of the distance between the mine and the extremity of the first sub-basin is 27 km. It is 37 km for SB2, 45km for SB3, 46km for SB4, 58km for SB5 and 59km for SB6

We calculated the mean of the maximal distance between the mine and the furthest extremity of each sub-basin order. On average, a mine is at 27 kilometers of the furthest extremity of its first sub-basin, at 37 kilometers of its second sub-basin, 45 kilometers of sub-basin 3, 46 kilometers of sub-basin 4, 58 kilometers of sub-basin 5 and 59 kilometers of sub-basin 6. Following this indicator, we compare in Figure 2.13 the individuals living downstream within the first and second sub-basin to those living upstream within 35 kilometers of the mine (coefficient SB1-2). Then, we compare individuals living downstream within the third and fourth sub-basins to those living upstream within 35 to 50 kilometers of the mine (coefficient SB3-4). Finally, we compare individuals living downstream within the fifth and sixth sub-basin to those living within 50 to 60 kilometers of the mine. Figure 2.13 shows that our result is mainly driven by the effect within the third and fourth sub-basins, while in rural areas the effect is only significant within the closest sub-basins. This shows that the effect is critical close to the mine, where the pollution is supposed to be the highest. The difference between the whole and rural samples might be explained by the fact that the location of mines close to urban areas suffers from lower precision (cf section 2.10.3.2).

### 2.9.2 Mine density

Figure 2.14: Distance of mines to each other

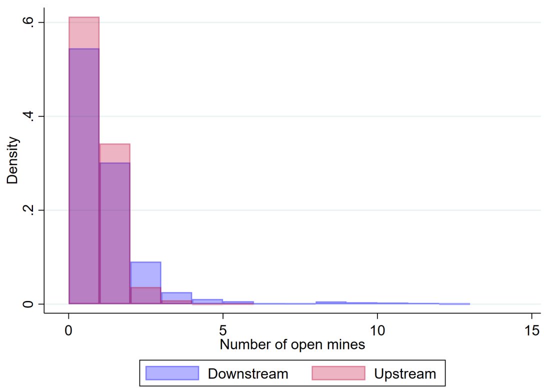



**Notes:** This Figure gives the distribution of the distance between each mine and its closest industrial mining site, giving insights into how far these mines are located from each other. In red is plotted the median distance (18km), and the graph is given for distances under 100 km (please note that the maximum distance is up to 474 km).
*Sources:* Authors' elaboration on DHS and SNL data.

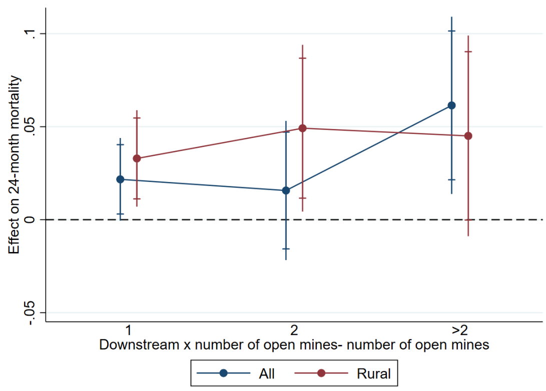
In this section, we explore the intensive margin of our result according to the mine density. Figure 2.14 gives insights into how far these mines are located from each other. It shows the distribution graph of the distance to the closest mine for each mine. It is made with no regard for the activity status of the mining site. On average, a mine is located at least 31 kilometers from its closest mining site, while the median is 18 kilometers. The distribution is skewed to the right, showing that the majority of the mining sites are located in areas with a high density of mining activity. Few mines are isolated, up to 100 kilometers from the closest mining site. This graph shows the necessity first to control for the number of open mines within the area in the main analysis (Table 2.2), and the necessity to look at heterogeneous effects according to the intensity of the mining activity within the area.

Figure 2.15a plots the frequency of the number of open mines within our main sample, both for upstream and downstream areas, within 45 kilometers and up to the third sub-basin. As the number of observations falls starting at 3 open mines, we investigate in Figure 2.15b and Table 2.10 the statistical difference between being downstream of one opening site, two or more than two. For the whole sample, being downstream of one open mine increases the 24-month mortality by 2 p.p. The effect increases when the number of open mines increases, as being downstream of more than 2 open mines increases by 6 p.p the mortality rate in comparison to being downstream of only one open mine. Table 2.10 column (1) gives the DiD interaction term when open becomes a continuous variable and not a dummy, being the number of open mines. It shows that being downstream one additional mine that opens increases the mortality by 1.3 p.p, and by 2 p.p within rural areas.

Figure 2.15: Intensive margin - Effect of the number of mine openings on 24-month mortality




(a)




(b)

**Notes:** Figure (a) plots the distribution of the number of open mines across downstream and upstream villages. Figure (b) plots the interaction variable on the 24-month mortality rates. It gives the average treatment effects of the number of mines open on 24-month mortality.
*Sources:* Authors' elaboration on DHS and SNL data.

Table 2.10: Effects of the number mine opening on infantile mortality according to the number of open mine

| | 24-month mortality | | | |
|---|---|---|---|---|
| | All | | Rural | |
| | (1) | (2) | (3) | (4) |
| Downstream×Nb open | 0.0130*** | | 0.0205*** | |
| | [0.00484] | | [0.00744] | |
| Downstream×Nb open=1 | | 0.0217* | | 0.0329** |
| | | [0.0113] | | [0.0132] |
| Downstream×Nb open=2 | | 0.0157 | | 0.0492** |
| | | [0.0191] | | [0.0228] |
| Downstream×Nb open >2 | | 0.0614** | | 0.0451 |
| | | [0.0243] | | [0.0275] |
| Downstream | -0.0214*** | -0.0208*** | -0.0295*** | -0.0308*** |
| | [0.00723] | [0.00767] | [0.00823] | [0.00837] |
| Nb Open | -0.00419 | | -0.00842 | |
| | [0.00526] | | [0.00613] | |
| Nb Open =1 | | -0.0000134 | | -0.00328 |
| | | [0.00785] | | [0.00895] |
| Nb Open =2 | | -0.00850 | | -0.0220 |
| | | [0.0129] | | [0.0141] |
| Nb Open >2 | | -0.0407** | | -0.0329 |
| | | [0.0190] | | [0.0219] |
| Controls | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes |
| N | 35,638 | 35,638 | 24,544 | 24,544 |
| R2 | 0.0511 | 0.0511 | 0.0633 | 0.0634 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$. Columns (1-3) give the results for the total population while columns (4-6) display the results for rural villages. Columns (2, 4, 6, 8) control for the number of open mines within 45 km. The same sample as Table 2.2 Column 2 applies. Control variables are birth order number, mother's age, mothers' age square, mother's years of education and urban, and a continuous variable indicating the presence of rivers and their order.

### 2.9.3 Production intensive margin

Table 2.11: Effects of industrial mining opening, across each commodity's price evolution.

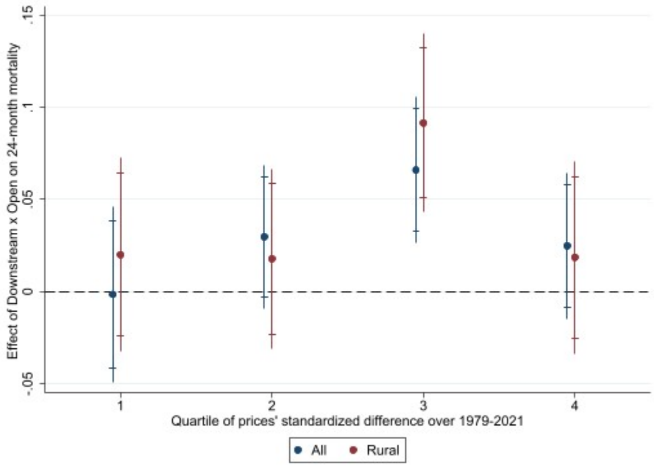
| Outcome | 24-month mortality | |
|---|---|---|
| Price var. | (1) Standardized difference | (2) Z-score |
| Downstream × Open× Price var | 0.0160* | 0.0101** |
| | [0.00823] | [0.00463] |
| Downstream | -0.00244 | 0.00102 |
| | [0.0104] | [0.0110] |
| Controls | Yes | Yes |
| Country-survey year FE | Yes | Yes |
| Mine SB FE | Yes | Yes |
| Mine SB-survey year trend | Yes | Yes |
| Commodity FE | Yes | Yes |
| N | 31,517 | 31,517 |
| R2 | 0.0509 | 0.0509 |
| Outcome mean | 0.0907 | 0.0907 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$. Standardized difference and Z-score of each commodity's price calculated over 1979-2021. The same sample and controls as Table 2.2 Column 2 apply.

We proxy the production intensity of each mine by using the global price of each mine's primary commodity as done in Berman et al., 2017 and Girard, Molina-Millan, and Vic, 2022. We presume the higher the prices, the more intense the production. Annual prices were retrieved from the SNL data (coal, gold, lead, nickel, platinum, silver, and zinc) and the World bank pink sheet (copper). For each commodity, we calculate the average price over 1971-2021, and for each year we calculate the standardized difference ($\frac{Price_t - \overline{Price}_{[1971-2021]}}{\overline{Price}_{[1971-2021]}}$) and Z-score ($\frac{Price_t - \overline{Price}_{[1971-2021]}}{\sigma_{Price_{[1971-2021]}}}$). We find a positive and significant effect of the triple interaction (Table 2.11) with both price variables (column 1 for the standardized difference and column 2 for the Z-score).

We then plot the coefficients associated with the interaction term of being downstream of an open mine, across the quartiles of the change in the z-score (Figure 2.16). For both total and rural samples, we find that an increase in prices (going from the second to the third quartile) leads to an even higher effect of industrial mining on the 24-month mortality.

Figure 2.16: Effect of living downstream of an open mine across the evolution of the mine's primary commodity's price




**Notes:** This graph represents the coefficients associated with being downstream of an open mine across the evolution of each mine's primary commodity's price (available for coal, copper, gold, lead, nickel, platinum, silver and zinc). For each commodity, the standardized difference to the mean over the 1979-2021 period was calculated and then split across quartiles to grasp the relative price evolution specific to each type of commodity.
*Sources:* Authors' elaboration using DHS, SNL, and World Bank pink sheet data.

## 2.10 Robustness checks

### 2.10.1 Balanced sample and de Chaisemartin and d'Haultfœuille, 2020

#### 2.10.1.1 Balanced sample

One issue of working with DHS data is dealing with repetitive cross-sections instead of an exact panel. In this section, we define a balanced sample as a restricted sample for which each mine has observations before and after its opening, both upstream and downstream. In this sense, it is a balanced panel of mine, if we consider only two points in time which are (1) the period before the mine opens and (2) the period after its opening. Please note that, in this paper, it is possible to restrict the analysis to a balanced sample by the extension of the mines' sample size, and it underlines the limits of analyses looking at dynamic effects while using a few mines. In this section, we first define the balanced sample and then we replicate the main analysis on this restricted sample.

First, let's define the construction of the balanced sample. A staggered DiD is driven by changes in mortality rates of switchers, which are observations that change treatment status, in comparison to those that do not change treatment status. In the case of a balanced sample, the design of this paper distinguishes three groups of observations :

- Group 1 "the switchers": the subgroup for which a mine has opened between two different years (for which there are DHS observations) and thus for which the treatment status changes from 0 to 1.
- Group 2 "the always treated": the subgroup of areas for which the mine has always been opened and are thus always treated, i.e. the treatment variable which is an interaction is always equal to 1.
- Group 3 "the never treated": the subgroup of areas where mines have not yet opened. The treatment variable is equal to 0. The third group is made of subgroups for which the mine has not opened yet in 2022 but the opening is planned in the future (the mine is projected to open), but also of mines where no DHS cluster was surveyed after it opened. This group is called "the never treated", but it includes both DHS villages that will never be treated or are not yet treated because they will be treated in the future.

The balanced sample makes it possible to identify the three groups. Formally, it is defined as the following, for each group:

Let's consider observations that can be divided into G groups and T periods, for every $(g, t) \in \{1, ..., G\} \times \{1, ...T\}$, let $N_{g,t}$ denote the number of observations in the group $g$ and period $t$, and let $N = \sum_{g,t} N_{g,t}$ be the total number of observations. For all $(g, t) \in \{1, ..., G\} \times \{1, ..., T\}$, let's call $D_{g,t}$ the $Downstream_{g,t}$ variable and $O_{g,t}$ the $Opened_{g,t}$ variable.

**Definition 1** (Group 1- Balanced sample of "switchers"). *Let's call* $G_1 = \{g_0, ..., g_{n_1}\}$ *the set of Group 1. Group 1 is defined as the following:*
*For all* $g \in G_1, \exists (v_1, v_2, v_3, v_4) \times (t_1, t_2, t_3, t_4) \in g \times T$ *such as:*
*(i)* $N_{v_1,t_1} > 0 \wedge D_{v_1,t_1} = 0 \wedge O_{v_1,t_1} = 0$
*(ii)* $N_{v_2,t_2} > 0 \wedge D_{v_2,t_2} = 1 \wedge O_{v_2,t_2} = 0$
*(iii)* $N_{v_3,t_3} > 0 \wedge D_{v_3,t_3} = 0 \wedge O_{v_3,t_3} = 1$
*(iv)* $N_{v_4,t_4} > 0 \wedge D_{v_4,t_4} = 1 \wedge O_{v_4,t_4} = 1$
*In our setting, g is the whole area associated with a mine, including both upstream and downstream observations, and is made of* $k \in N$ *DHS clusters such as* $g = \{v_1, ..., v_k\}$.

**Definition 2** (Group 2- Balanced sample of "always treated" ). *Let's call* $G_2 =$

$\{g_0, ..., g_{n_2}\}$ *the set of Group 2. Group 2 is defined as the following:*
*For all* $g \in G_2, \exists (v_1, v_2) \times (t_1, t_2) \in g \times T$ *such as:*
*(i)* $N_{v_1,t_1} > 0 \wedge D_{v_1,t_1} = 0 \wedge O_{v_1,t_1} = 1$
*(ii)* $N_{v_2,t_2} > 0 \wedge D_{v_2,t_2} = 1 \wedge O_{v_2,t_2} = 1$

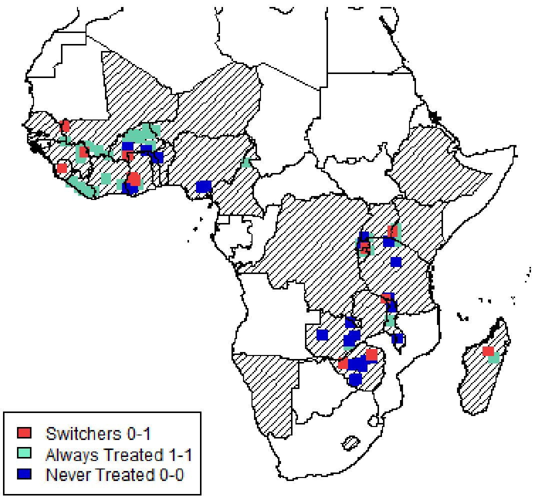
**Definition 3** (Group 3- Balanced sample of "never treated" )**.** *Let's call* $G_3 = \{g_0, ..., g_{n_3}\}$ *the set of Group 3. Group 3 is defined as the following:*
*For all* $g \in G_3, \exists (v_1, v_2) \times (t_1, t_2) \in g \times T$ *such as:*
*(i)* $N_{v_1,t_1} > 0 \wedge D_{v_1,t_1} = 0 \wedge O_{v_1,t_1} = 0$
*(ii)* $N_{v_2,t_2} > 0 \wedge D_{v_2,t_2} = 1 \wedge O_{v_2,t_2} = 0$

Definitions 1, 2, and 3 define the treatment group (Group 1) and the control groups (Group 2+3) in the setting of the mine-balanced sample. For Group 1 switchers, we restrict the sample to areas that have DHS clusters surveyed both downstream and upstream both before and after the opening of the mine. This means that we select areas that have been surveyed at least in four different locations by DHS. For groups 2 and 3, meaning that the surveys have occurred only after the opening (Group 2) or before the opening (Group 3), we restrict to areas that have observations both upstream and downstream.

Figure 2.17 plots the different groups from the balanced sample and displays the group of switchers, the always-treated and the never treated groups. It displays the areas that are key to the main estimation, which are located in Western Africa, Zimbabwe, Western Kenya, Rwanda, Tanzania, and Madagascar. Table B.9 from Appendix plots our main estimator across the African sub-regions and shows that our results are mainly driven by Western Africa, and remain significant in Eastern Africa. Table 2.12 gives the size of the three groups in the balanced sample, as well as the associated number of mines. It shows that Group 1 switchers account for around 12% of the total balanced sample, and corresponds to the neighborhood of 13 mines. In total, the control groups gather 75 mines. We observe that the average mortality rates have decreased over time, before and after the opening of the mining site in the Switcher Group, linked to the decrease of infant mortality in Africa over time, which is in coherence with the Balance Table 2.1 and Figures 2.4 and 2.5, and which highlights the importance of controlling for trends.

Table 2.13 displays the balance table for the restricted sample. This table shows the within comparison before and after a mine opens both downstream and upstream. It is only descriptive statistics and neither accounts for control variables nor account for fixed effects. The table shows that there is a significant difference between upstream and downstream areas after a mine opens concerning both the 12 and 24-month mortality rates. From a descriptive point of view, being downstream of a mine increases the 12-month mortality by 2.7 p.p, and the 24-month mortality by 2.4 p.p (column (13)). This difference is explained by a significant decrease in mortality rates within upstream areas after a mine opening (column (11)).

Figure 2.17: Balanced Panel - Group identification




**Notes:** The Figure plots the groups' areas across the three groups of the balanced panel, for the 24-month mortality rate.
*Sources:* Authors' elaboration on DHS and SNL data.

Table 2.12: Balanced Sample - Descriptive Statistics

| | Group 1 : Switchers 0-1 | | | | | | Groups 2+3 | | Group 2 : 1-1 | | Group 3 : 0-0 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | | Before Opening | | After Opening | | | | | | | |
| | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) | N | Mean (SD) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| **Dth<12** | | | | | | | | | | | | |
| All | 1,191 | 0.07 | 894 | 0.069 | 297 | 0.071 | 8,423 | 0.07 | 2,368 | 0.056 | 6,055 | 0.076 |
| | | (0.255) | | (0.254) | | (0.257) | | (0.256) | | (0.23) | | (0.265) |
| Mines | 13 | | 13 | | 13 | | 75 | | 31 | | 44 | |
| **Dth<24** | | | | | | | | | | | | |
| All | 1,191 | 0.089 | 894 | 0.091 | 297 | 0.084 | 8,423 | 0.091 | 2,368 | 0.072 | 6,055 | 0.099 |
| | | (0.285) | | (0.287) | | (0.278) | | (0.288) | | (0.258) | | (0.298) |
| Mines | 13 | | 13 | | 13 | | 75 | | 31 | | 44 | |

**Notes:** Standard errors and p-values are in parentheses. Outcomes' descriptive statistics of under 12-and 24-month mortality, for villages within the Group 1 Switchers for individuals born before and after the opening of the mine, then Group 2 always treated and Group 3 never treated.

Table 2.13: Balance Table

| | Before Mine Opening | | | | | After Mine Opening | | | | | Within Up. | Within Dwn. | Within |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Upstream | | Downstream | | Diff | Upstream | | Downstream. | | Diff | | | |
| | N | Mean /(SD) | N | Mean /(SD) | (4-2) /(p.v) | N | Mean /(SD) | N | Mean /(SD) | (9-7) /(p.v) | (7-2) /(p.v) | (9-4) /(p.v) | (12-11) /(p.v) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| **Dth<12** | | | | | | | | | | | | | |
| All | 5,272 | 0.079 | 1677 | 0.063 | -0.016 | 1,812 | 0.054 | 853 | 0.066 | 0.012 | -0.025 | 0.002 | 0.027 |
| | | (0.27) | | (0.243) | (0.025) | | (0.226) | | (0.248) | (0.248) | (0) | (0.814) | (0.005) |
| Mines | 54 | | 56 | | | 38 | | 37 | | | | | |
| **Dth<24** | | | | | | | | | | | | | |
| All | 5,272 | 0.101 | 1677 | 0.088 | -0.012 | 1,812 | 0.07 | 853 | 0.081 | 0.011 | -0.031 | -0.007 | 0.024 |
| | | (0.301) | | (0.284) | (0.123) | | (0.254) | | (0.273) | (0.306) | (0) | (0.527) | (0.03) |
| Mines | 54 | | 56 | | | 38 | | 37 | | | | | |

**Notes:** Standard errors and p-values in parentheses. Descriptive statistics of 12-month and 24-month mortality outcomes, for villages upstream and downstream of mining sites, for individuals born before and after the opening of a mine, over the balanced sample.

#### 2.10.1.2 Heterogeneous treatment effects with two-way fixed effects: de Chaisemartin and d'Haultfœuille, 2020

The main result of this paper estimates the effect of being downstream of an open mine by using standard difference-in-difference designs. However, recent developments in the estimation of difference-in-differences in staggered adoption designs (Borusyak, Jaravel, and Spiess, 2021; Goodman-Bacon, 2018; Callaway and Sant'Anna, 2021; de Chaisemartin and d'Haultfœuille, 2020) show that the estimated ATT[14] is a weighted sum of different ATTs with weights that may be negative. The negative weights are an issue when the treatment effect is heterogeneous between groups over time, as one could have the treatment coefficient in those regressions as negative while the treatment effect is positive in every group and time period. Using treated observations as controls creates these negative weights. In our design, the effect on Group 1 Switchers is compared to two control groups, Group 2 always treated and Group 3 never treated. The negative weights might come from the comparison of the effect of the Group 1 switchers to the Group 2 always treated. This biases the DiD estimator as it is an average of local treatment effects. In this section, we use the de Chaisemartin and d'Haultfœuille, 2020 estimator which deals with the issue of negative weights in a staggered adoption design.

Table 2.14 compares the two-way fixed effects (TWFE) used in the main result (odd columns), to the de Chaisemartin and d'Haultfœuille, 2020 estimator (dCDH) (even columns) [15]. Columns (1, 2, 5, 6) give the results for the entire sample, while columns (3, 4, 7, 8) for the balanced sample, defined in previous Section 2.10.1.1. Columns (1-4) give the results for the 12-month mortality rates, while columns (5-8) for the 24-month mortality rates.

First, let's look at the 24-month mortality rates. When looking at the TWFE estimator, we see that the results are stable on the balanced sample, even though it only represents 27% of the whole sample. This is coherent with the fact that the balanced sample keeps the villages that drive the main estimation's results. When focusing on the balanced sample, being downstream of an opened mine increases the 24 month-mortality rates by 3.19 p.p, which represents an increase of 36% of the mortality. Column (6) gives the dCDH estimator for the whole sample, while column (4) is for the balanced sample. We observe an increase in terms of the magnitude of

[14]Average Treatment on the Treated

[15]The Stata command *did_multiplegt* is used to run the dCDH estimator.

the effect when correcting for negative weights, as being downstream of an open mine increases the 24-month mortality rates by 11 p.p, which represents an increase of 129% of the mortality. If these magnitudes seem high, it is reassuring to observe the stability of the direction and significance of our main effect when using the dCDH estimator. Regarding the 12-month mortality rates, we observe that the restriction to the balanced sample displays a 2.8 p.p increase.

Table 2.14: Effects of industrial mining opening on 24-month mortality de Chaisemartin and d'Haultfœuille, 2020

| | 12-month mortality | | | | 24-month mortality | | | |
|---|---|---|---|---|---|---|---|---|
| | Whole Sample | | Balanced Sample | | Whole Sample | | Balanced Sample | |
| | TWFE | dCDH | TWFE | dCDH | TWFE | dCDH | TWFE | dCDH |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Downstream×Open | -0.00506 | 0.0249 | 0.0286* | 0.1457 | 0.0218** | 0.1109** | 0.0319** | 0.1667* |
| | [0.00831] | [0.0262] | [0.00222] | [0.1132] | [0.0108] | [0.0405] | [0.0162] | [0.1112] |
| Downstream | -0.0152** | | -0.0242*** | | -0.0211*** | | -0.0283*** | |
| | [0.00665] | | [0.00826] | | [0.00739] | | [0.00866] | |
| Open | 0.00963 | | 0.0347 | | -0.00496 | | 0.0238 | |
| | [0.00754] | | [0.0320] | | [0.0101] | | [0.0347] | |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 48, 472 | 48, 472 | 9, 606 | 9, 606 | 35, 638 | 35, 638 | 9, 606 | 9, 606 |
| R2 | 0.0378 | | 0.0523 | | 0.0511 | | 0.0599 | |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Column (1) gives the result of the main anlysis, the Two Way Fixed Effect (TWFE) for the whole sample, while Column (2) gives the de Chaisemartin and d'Haultfœuille, 2020 estimator. Columns (3) and (4) give the TWFE and de Chaisemartin and d'Haultfœuille, 2020 estimators for the balanced sample.

### 2.10.2 Sensitivity analysis

#### 2.10.2.1 Including non-topographic sub-basins

In this section, we replicate the main analysis from Table 2.2, adding within the control group, individuals living in a sub-basin with no topographic relation to the mine sub-basin, within 45 kilometers. This test can have several readings.

First, it strengthens the control for income effects linked to mining activity and enables to isolate more precisely of the channel of water pollution, and excludes other potential mechanisms. Indeed, villages close to the mine but located in a sub-basin with no topographic relationship with the mine, are allegedly less exposed to mining-induced water pollution and would be as exposed to income or labor effects, conflicts, or migration.

Yet, it also leads to the comparison of villages that do not necessarily share the same water resources, and this could blur the interpretation of our estimation. For example, other activities such as more intensive agriculture or livestock farming could aggregate around the mining site and could be responsible for other types of pollution. If these activities are located in a sub-basin with no topographic relationship to the mine, the estimated comparison would display the difference between the pollution of the mine and the pollution of these activities, rather than the pollution of the mine only, and this would lead to a downward bias to our analysis. Moreover, as mining activity is water intensive, the location of these activities might also be endogenous to the location of the mine, and this could induce an even larger downward bias.

Table 2.15 displays the results when including the non-topographic sub-basins within the control group. As expected, the table suggests that the main results of the 24-month mortality rates are downward biased and only significant at the 10% level for the rural population.

Table 2.15: Effects of industrial mining opening on infantile mortality - including DHS with non-topographic relationship

| | 12-month mortality | | | | 24-month mortality | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Population | | Rural Population | | Total Population | | Rural Population | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Downstream×Open | -0.00266 | -0.00227 | 0.00341 | 0.00381 | 0.00746 | 0.00804 | 0.0172* | 0.0165* |
| | [0.00543] | [0.00544] | [0.00632] | [0.00632] | [0.00752] | [0.00753] | [0.00887] | [0.00888] |
| Downstream | -0.00490 | -0.00472 | -0.00904* | -0.00884* | -0.00486 | -0.00455 | -0.0103 | -0.0106* |
| | [0.00445] | [0.00446] | [0.00506] | [0.00507] | [0.00561] | [0.00561] | [0.00637] | [0.00636] |
| Open | 0.00364 | 0.00488 | 0.00318 | 0.00466 | 0.00115 | 0.00308 | 0.00283 | 0.000259 |
| | [0.00292] | [0.00302] | [0.00345] | [0.00359] | [0.00377] | [0.00391] | [0.00457] | [0.00440] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Nb open mines | No | Yes | No | Yes | No | Yes | No | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-bthyr FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-bthyr trd | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 168,931 | 168,931 | 123,413 | 123,413 | 124,670 | 124,670 | 91,395 | 91395 |
| R2 | 0.0214 | 0.0215 | 0.0252 | 0.0252 | 0.0305 | 0.0305 | 0.0356 | 0.0355 |
| Outcome Mean | 0.0638 | 0.0638 | 0.0670 | 0.0670 | 0.0824 | 0.0824 | 0.0872 | 0.0872 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Columns (1-3) give the results for the total population while columns (4-6) display the results for rural villages. The same controls as Table 2.2 apply. Columns (2, 4, 6, 8) control for the number of open mines within 45 km. The sample includes individuals living in non-topographic sub-basins within 45km.

#### 2.10.2.2 Dropping fixed effects and other tests

Table 2.16: Effects of industrial mining opening on 24 months mortality, while dropping fixed-effects.

| Outcome | 24-month mortality | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Downstream×Open | 0.0218** | 0.0216** | 0.0179* | 0.0177* |
| | [0.0108] | [0.0108] | [0.0105] | [0.0104] |
| Downstream | -0.0211*** | -0.0211*** | -0.0218*** | -0.0219*** |
| | [0.00739] | [0.00738] | [0.00734] | [0.00733] |
| Open | -0.00496 | -0.00494 | -0.00459 | -0.00466 |
| | [0.0101] | [0.0101] | [0.00972] | [0.00962] |
| Controls | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | No | No | No |
| Commodity FE | Yes | Yes | No | No |
| Mine SB-birthyear trend | Yes | Yes | Yes | No |
| N | 35,638 | 35,638 | 35,638 | 35,638 |
| R2 | 0.0511 | 0.0504 | 0.0503 | 0.0491 |
| Outcome mean | 0.0873 | 0.0873 | 0.0873 | 0.0873 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The same sample as Table 2.2 Column 2 apply.

We find stability in our results when dropping fixed effects one by one: birth month primary commodity, and sub-basins birth year trend (Table 2.16) until keeping the two-way fixed effects (i.e keeping the mine sub-basin fixed effect and the country-birthyear fixed effect).

Section B.5.1 runs other tests. Table B.10 shows that our result is stable when controlling for the hand work. Figure B.18 shows that the main results are stable when dropping countries one by one, and Figure B.17 when dropping metals one by one.

#### 2.10.2.3 Spatial correlation

As an additional robustness check, we run our main result's specification while taking into account the spatial correlation of DHS clusters. We estimate the standard errors with a spatial HAC correction following the method developed by Conley, 1999 and using the Stata command introduced by Colella et al., 2019. Table 2.17 shows the stability of our results for different cut-off distances of spatial correlation (from 20 km to 200 km). We did not include directly the Conley, 1999 test in the main analysis as it does not allow for several fixed effects. Table 2.17 corrects for spatial correlation for the results when using only Mine-Subbasin and country-birthyear fixed effects (result from Table 2.16 Column (4)).

Table 2.17: Effects of industrial mining activity, Conley spatial correction (acreg)

| Outcome | Mortality under 24 months | | | | | |
|---|---|---|---|---|---|---|
| Conley spatial correction threshold | 20 km | 45 km | 60 km | 80 km | 100 km | 200 km |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Downstream×Open | 0.0177* | 0.0177* | 0.0177* | 0.0177* | 0.0177* | 0.0177* |
| | [0.0100] | [0.00999] | [0.0101] | [0.0101] | [0.00996] | [0.00916] |
| Downstream | -0.0219*** | -0.0219*** | -0.0219*** | -0.0219*** | -0.0219** | -0.0219** |
| | [0.00709] | [0.00746] | [0.00789] | [0.00847] | [0.00913] | [0.0106] |
| Open | -0.00466 | -0.00466 | -0.00466 | -0.00466 | -0.00466 | -0.00466 |
| | [0.00941] | [0.00932] | [0.00943] | [0.00960] | [0.00955] | [0.00938] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 35,648 | 35,648 | 35,648 | 35,648 | 35,648 | 35,648 |
| R2 | 0.00262 | 0.00262 | 0.00262 | 0.00262 | 0.00262 | 0.00262 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. The variables Downstream and Opened are dummies that indicate whether the individual lives in a village downstream of at least one mining site and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines. Control variables are birth order number, mother's age, mother's age square, mother's years of education, urban, number of open mines, and presence of rivers.

### 2.10.3 Measurement errors

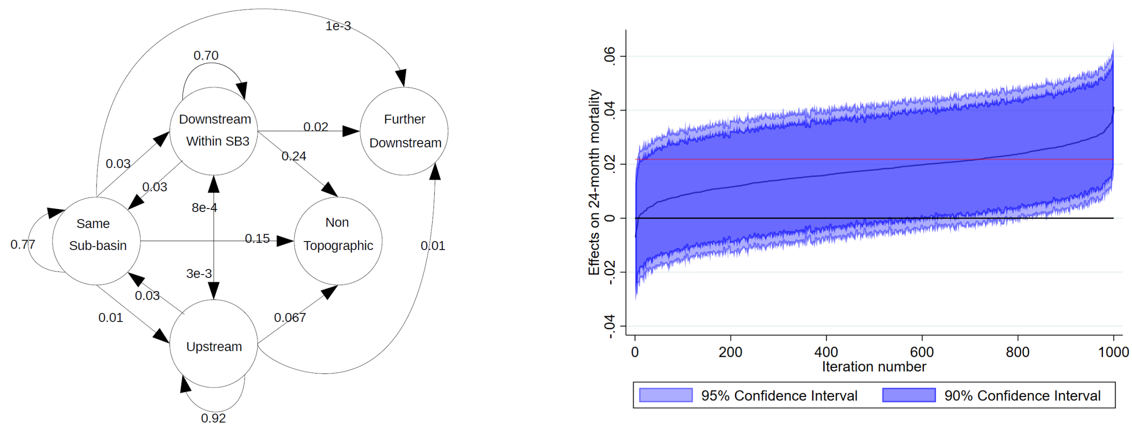
In this section, we deal with measurement errors that come from the nature of the data. Section 2.10.3.1 tests for the random displacement of DHS villages and section 2.10.3.2 tests our result according to the precision of the mine location.

#### 2.10.3.1 DHS random displacement

DHS randomly displaces the GPS coordinates of each village to protect the confidentiality of respondents. Urban locations are displaced within 2 kilometers, while rural clusters are displaced within 5 kilometers, with 1% of rural clusters moved up to 10 kilometers. Displacements are made within administrative districts. This random reshuffling of DHS villages introduces measurement errors in our main estimation, all the more important as our treatment allocation depends on the relative position of the DHS villages to the mine.

First, we randomly displace 1,000 times each DHS village within a buffer of 2 kilometers for urban clusters and 5 kilometers for rural clusters. Thus, each displaced village is located in a new sub-basin, which can be the initial sub-basin or not. Then, we determine the topographic relation of this sub-basin to the sub-basin of the mine, which gives the treatment of the DHS village: whether it falls into a sub-basin upstream, downstream, in the same sub-basin as the mine or in a sub-basin with no topographic relationship with the one of the mine. The topographic relation of the new sub-basin gives the new treatment status of the DHS village. We only reshuffled the position of DHS villages that have a topographic relation with the mine initially, and that were up to the third sub-basin downstream. This means that we can have some DHS villages that exit the main sample, for instance, if their newly assigned sub-basin has no topographic relation, or is downstream in the fourth sub-basin, or if it falls into the same sub-basin as the mine, as these cases are excluded from the main result. The only new observations that come within the sample are DHS villages that were initially within the same sub-basin as the mine and fall upstream, or downstream with the new iteration of the random displacement of their location. Please note that it is possible as well, but very rare, that a DHS village falls into the ocean.

Figure 2.18: DHS random displacement - 1,000 iterations




(a) (b)

**Notes:** Figure (a) plots the transition probability graph for 1,000 random displacements of DHS clusters. Figure (b) plots the interaction term for 1,000 different regressions, each done for a new sample where DHS GPS coordinates have been randomly displaced. The red line y=0.0218 plots the coefficient from our main result. The coefficients are ordered, and we plot the 95% and 90% intervals.
*Sources:* Authors' elaboration.

Figure 2.18a gives the probability graph showing the transition probabilities of changing treatment status. For instance, after 1,000 iterations, a DHS village initially downstream within the third sub-basin has 70% chances to remain downstream up to sub-basin three, has 0.3% chances to be upstream the mine, 24% chances to fall into a sub-basin with no topographic relation (and be out of the sample), 3.5% chances to be in the same sub-basin of the mine and finally 2% chances to be downstream further than the third sub-basin (and be out of the sample). In the end, a DHS village treated in our initial sample has 25% chances to leave the sample. Please note that this random reshuffling is not perfect, as DHS villages should be reshuffled within administrative level 2 boundaries as made in the DHS procedure.

Figure 2.18b plots the interaction term $Donwstream \times Open$ of our main estimation for 1,000 random displacements of DHS GPS coordinates. The coefficients are ordered, and we plot the 95% and 90% intervals. As there is a higher probability that a DHS cluster leaves the sample rather than a new enters it, the number of observations varies for each iteration and is likely to be smaller than our main estimation.

#### 2.10.3.2 Accuracy of mine location

We further test for potential measurement errors by looking at the precision of the mines' location. The SNL database provides information on the accuracy levels of each mine's GPS coordinates enables us to restrict the analysis to mines with exact coordinates, precise at 1 km. Our main results are positive but no longer significant when restricting to the mines with exact coordinates but hold when focusing on rural households. This hints towards a higher effect of industrial mining activity on child mortality among rural households, and a lack of precision in the location of mines close to urban areas.

Table 2.18: Effects of industrial mining opening, restriction to exact GPS coordinates.

| Outcome | 24-month mortality | | |
|---|---|---|---|
| Accuracy level | All | Exact coordinates | |
| Sample | Urban and rural | Urban and rural | Rural |
| | (1) | (2) | (3) |
| Downstream×Open | 0.0218** | 0.0116 | 0.0294** |
| | [0.0108] | [0.0116] | [0.0143] |
| Downstream | -0.0211*** | -0.0204** | -0.0239*** |
| | [0.00739] | [0.00807] | [0.00888] |
| Open | -0.00496 | 0.00532 | 0.00805 |
| | [0.0101] | [0.0104] | [0.0130] |
| Controls | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes |
| N | 35,638 | 29,195 | 20,172 |
| R2 | 0.0511 | 0.0517 | 0.0626 |
| Outcome mean | 0.0873 | 0.0858 | 0.0920 |

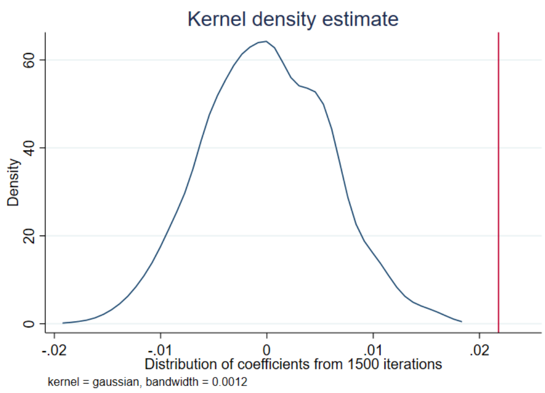
**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The same controls as Table 2.2 Column 2 apply.

### 2.10.4 Placebo tests

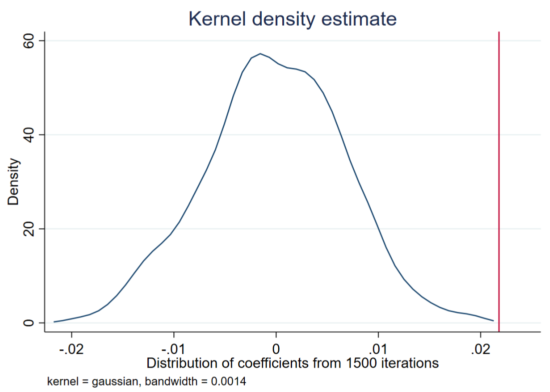
#### 2.10.4.1 Randomization inference

To make sure that the assignment of each village to its topographic position relative to the mine is indeed what drives our result on child mortality, we run a randomization inference test. We draw randomly 1,500 permutations of the "Downstream" variable without changing the start-up year and 1,500 permutations of the "Open" variable without changing the downstream position [16]. The simulations show that the distribution of treatment effects (Downstream × Open) are shifted around zero (Figure 2.19). The red line represents the initial treatment effect using our main specification: we are sure at the 1 percent level that our main model is not misspecified.

Figure 2.19: Spatial and temporal randomization inference tests




(a) Randomization of the "Downstream" variable




(b) Randomization of the "Open" variable

**Notes:** The two figures represent the distribution of coefficients associated with the interaction term of being downstream of an open mine and its effect on under 24-month mortality when conducting 1,500 permutations of the "Downstream" position of each DHS sub-basin (Figure (a)) and 1,500 permutations of the "Open" variable (Figure (b)). The red line represents the initial treatment effect using our main specification.
*Sources:* Authors' elaboration using the Stata *ritest* command.

#### 2.10.4.2 Placebo diseases

We conduct a placebo test on other potential diseases that could affect women's and thus children's mortality. We do not find significant industrial mining on the infection of any sexually transmitted disease among women living downstream of an open mine (Table 2.19 column 1) or among awareness of tuberculosis (column 2). This

[16]The randomization inference of the "Downstream" and "Open" treatment are within the sub-basin level, and are clustered at the DHS village level.

absence of differential results on women's health across upstream and downstream villages is reassuring for our identification of the water pollution channel.

Table 2.19: Effects of industrial mining opening on women, placebo diseases.

| | (1) | (2) |
|---|---|---|
| Outcome | Any sexually transmitted infection | Heard of tuberculosis |
| Downstream × Open | 0.00332<br>[0.00923] | -0.0387<br>[0.0259] |
| Downstream | 0.00751<br>[0.00772] | 0.0277<br>[0.0210] |
| Open | 0.00766<br>[0.00913] | 0.00469<br>[0.0314] |
| Controls | Yes | Yes |
| Country-survey year FE | Yes | Yes |
| Mine SB FE | Yes | Yes |
| Mine SB-survey year trend | Yes | Yes |
| Commodity FE | Yes | Yes |
| N | 66,653 | 14,750 |
| R2 | 0.0888 | 0.186 |
| Outcome mean | 0.0501 | 0.938 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The same sample and controls as Table 2.4 apply.

## 2.11 Discussion and limits

### 2.11.1 Selection issues

Our study includes the best available data on mining and child health we have at the scale of a continent, but of course, is neither exhaustive nor represents the whole African continent. First, as we limited our sample to countries with at least two waves of DHS, we had to drop many countries with only one wave among which South Africa which has intense mining activity. Future work will be possible once other survey waves will have been conducted. Our 26 countries represent about two-thirds of the total population of the continent[17].

Second, our mining data is also limited to industrial sites, and do not include artisanal or small-scale mining which information is much harder to retrieve at the scale of the continent. One possibility for another project is to use data on the suitability of artisanal gold mining (Girard, Molina-Millan, and Vic, 2022) and to compare their

[17]The proportion is stable between 1981 and 2020.

environmental impacts to industrial mining.

One remaining concern is about the exhaustivity of the SNL, and the heterogeneity of the sampling selection across countries. We tried our best to evaluate the exhaustivity of the SNL data by comparing it with other mining data and social sources (Ministries, USGS, mining website. . . ) but it is out of our feasible means to get an exact proportion of representativity and to compete with the business-oriented activity of SNL.

### 2.11.2 Threats to the identifying assumption

#### 2.11.2.1 Type of pollution

First, our study mainly focuses on water pollution through the lens of water subbasins while controlling for rivers, i.e. surface water. Yet, further work could also control for groundwater as their pollution may follow different dynamics than surface water: villages located in areas with low-depth groundwater could pump the water more easily and with more affordable water pumps than in areas with deeper groundwater. These former villages could therefore be more exposed to mining-induced water pollution than the latter ones. Moreover, groundwater could take more time to be contaminated by mining-induced pollution than surface water, but its contamination could also last more permanently.

This paper does not directly examine mining-induced air pollution. The main hypothesis is that wind direction is less correlated to the topographic position of the village than water pollution and that the comparison between upstream and downstream villages should exclude the effect of air pollution. Besides, as discussed in Section 3.2, the effects of air pollution seem to concern the mine workers more than the surrounding population, even though fine particles can be displaced over long distances. However, our main result is not entirely net off the impacts of air pollution. The best control included so far is adding the sub-basins with no topographic relationship to the mine, as they are allegedly exposed to air pollution only, while sub-basins with a topographic relationship would be exposed to both water and air pollution. It is beyond the scope of the current paper to take into account the direction of the wind to disentangle both sources of pollution, but empirically feasible for another paper. Our study is most likely an underestimation of the total pollution induced by industrial mining activity.

The same concern remains for soil pollution. An additional heterogeneity analysis would be to take into account areas prone to subsistence agriculture or livestock, as mining-induced water pollution could also contaminate soils and cattle in the long run and the subsequent food produced. We assume more harmful effects of industrial mining on local population health if both water and food are polluted. In another paper, one could study the heterogeneous effects across the global agroecological zones (GAEZ) and crop suitability, and test whether villages located near industrial mining sites in high-yield crop areas are more affected than villages with less suitable soils.

#### 2.11.2.2 Threats to identification

A major threat to identification is that the opening of a mine may not be orthogonal to unobservable factors that affect health and water quality, in different ways for downstream and upstream areas.

Migration is a major methodological concern, as we show in Table B.8 of Appendix B.3.1 that migrants significantly settle downstream after a mine opening. Section 2.6.2 shows that our main result is robust when controlling for in-migration. A main violation of the identifying assumption would be if downstream villages anticipate the mine opening and strategically out-migrate within upstream areas to avoid pollution. In this case, there would be a selection bias, as the individuals surveyed downstream after the mine opening would be those that were not able to migrate or anticipate the pollution. Controlling for in-migration in DHS villages, we show that our result is robust to this specific strategic behavior. However, we cannot control for strategic out-migration outside of the study area, meaning individuals out-migrating to avoid pollution elsewhere than the upstream area. In this paper, we made the choice not to use mother fixed-effects and retrospective questions on birth history, to limit endogenous selection due to out-migration, and to account for children born up to five years prior to the year of the survey.

Accordingly, a threat to the identification would be a differed improved access to infrastructure associated with the opening of a mine between upstream and downstream areas. Table B.8 shows no difference in terms of access to electricity and piped water between downstream and upstream areas after a mine opening, and Section 2.6 shows that our result is robust controlling for improved access to facilities.

An important omitted variable in our current study is the increased presence of conflicts and violence around areas with mining activity, as shown by Berman et al., 2017 and that could also explain the increase in child mortality in the vicinity of mines. We do not directly control for conflict, but there would be an upward bias of our estimation only if conflicts systematically happen more downstream than upstream. As our results hold when including non-topographic subbasins in rural areas, it is a first-step approximation that water pollution is indeed the main explaining factor of increased child mortality. Further work could include the ACLED data to exclude this mechanism.

Another concern is that other industries could aggregate around the mining industry and be partly responsible for the pollution. More than a bias, this could be a threat to identification if the location of the industry is correlated to the topographic position of the mine. In another paper, we could look at the correlation between mining activity and other industry implementations. Controlling for them could enable us to isolate the pollution linked to the mining activity solely.

## 2.12 Policy discussion

In this section, we first try to compute how many deaths were related to the water pollution linked to industrial mining activity in the 26 countries of our sample. Then, we try to assess whether the Extractive Industries Transparency Initiative, a global standard for good governance in the extractive sector, has been successful to reduce this mortality.

### 2.12.1 Back-of-the-envelope calculation

In this section, we compute a back-of-the-envelope calculation to grasp how many deaths could have been averted had there been policies implemented to limit water pollution, over the 1981-2020 period and within the 26 Sub-Saharan countries of our analysis.

First, we consider that as DHS is representative at the national level, it is feasible to calculate the proportion of individuals living within 45 kilometers of a mine, the proportion of those living downstream, etc. Here are the probabilities computed using the DHS database:

- $x = 28\%$: Proportion of individuals living within 45 kilometers of a mine [18]
- $x = x_d + x_u + x_{nt} + x_{sb}$, with
  - $x_d = 1.94\%$ : Proportion of individuals living downstream [19]
  - $x_u = 5.25\%$: Proportion of individuals living upstream
  - $x_{nt} = 17.93\%$: Proportion of individuals living with no topographic relation
  - $x_{sb} = 2.92\%$: Proportion of individuals living in the same sub-basin as a mine

Our analysis leads to an estimation of $e = 2.18\%$ the increased mortality rate because of industrial mining-induced water pollution. In our sample, 9,258 individuals live downstream of a mine and we count 822 deaths among them. The total number of additional deaths due to mining-induced water pollution is 9,258 $\times$ 2.18% = 202 deaths. We now look at the 880 million children who were aged 0-2 years over 1981-2020 in our 26 countries [20]. As we assume the representativity of the DHS surveys and the stable proportion of the population living in the vicinity of mines, this would mean that 1.94% $\times$ 880 million = 17 million children lived within 45 km downstream of a mine. This leads to 2.18% $\times$ 17 million = 370,600 deaths due to mining-induced water pollution over 1981-2020 in our 26 countries, i.e. 9,265 deaths per year, or 16 deaths per mine per year.[21] To grasp a better sense of the magnitude of this figure, there are on average 840,000 births per year and per country (average within the 26 countries over 1981-2020), which means that the number of deaths caused by mining-induced water pollution over 26 countries represents 1.1% of the number of births per country[22].

### 2.12.2 Extractive Industries Transparency Initiative members

We look at whether there is a significant difference across countries that have signed the Extractive Industries Transparency Initiative, launched in 2002 and which cur-

[18]Please note that, exactly, this is the proportion of individuals living within 45 kilometers of a mine and downstream up to the third sub-basin.
[19]up to the third sub-basin
[20]Source: World Bank data.
[21]There are 604 mines in total in our main results' regressions.
[22]As sampling weights are not considered in the calculation, we do not give a number per country.

Table 2.20: Effects of industrial mining opening, across EITI membership.

| Outcome | 24-month mortality | | | | | |
|---|---|---|---|---|---|---|
| Sample | All | | | Rural | | |
| | Not an EITI member | EITI member | | Not an EITI member— | EITI member | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Downstream× Open | 0.0179 | 0.0259** | 0.0133 | -0.0428 | -0.0221 | -0.0557 |
| | [0.0210] | [0.0127] | [0.0190] | [0.0518] | [0.0416] | [0.0669] |
| Surveyed after joining EITI | | | 0.0237 | | | -0.0259 |
| | | | [0.0303] | | | [0.0315] |
| D× O× Surv. after joining EITI | | | 0.0243 | | | 0.0221 |
| | | | [0.0236] | | | [0.0705] |
| Downstream | -0.0552*** | -0.00934 | -0.00446 | 0.00399 | 0.0356 | 0.0118 |
| | [0.0140] | [0.00876] | [0.0107] | [0.0500] | [0.0381] | [0.0641] |
| Open | -0.00576 | -0.00588 | -0.00656 | -0.0456 | -0.00345 | -0.0114 |
| | [0.0246] | [0.0112] | [0.0163] | [0.0651] | [0.0268] | [0.0443] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 8,434 | 26,810 | 26,810 | 2,251 | 8,685 | 8,685 |
| R2 | 0.0373 | 0.0548 | 0.0548 | 0.0838 | 0.0677 | 0.0679 |
| Outcome mean | 0.0716 | 0.0920 | 0.0920 | 0.0604 | 0.0738 | 0.0738 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. The same sample and controls as Table 2.2 Column 2 apply.

rently gathers 55 countries. Member countries commit to disclose information along the production value chain of oil, gas, and mining extraction and respect a common set of governing standards. We want to see if there is an effect of the EITI Rules signed by the member countries on the effect of industrial mining on child mortality. 18 out of the 26 countries included in our sample signed the EITI[23] which gather 76 percent of our sample of children.

We estimate our main specification across the sample of countries that are members of the EITI or not. We find that our results hold even for countries who committed to improved governance of their extractive industries (Table 2.20), but which are also countries heavily relying on this activity in their national economy. We find no significant effect of our results when looking at whether surveys were conducted before or after their country signed the EITI Standards (triple interaction Downstream × Open × Surveyed after joining EITI in columns (3) and (6)).

## 2.13 Conclusion

This paper identifies a negative externality of industrial mining on local population living standards, as we show that industrial mining sites increase infant mortality in surrounding villages, indirectly through the contamination of water resources. We match geocoded repeated-cross sectional household surveys to geocoded data on industrial mine openings obtained through intensive handwork. We propose a staggered Difference-in-Difference strategy and isolate the mechanism of water pollution by building the treatment and control groups using an upstream-downstream comparison. We compare the effect on the health of villages located upstream and downstream of a mine deposit, before and after its opening. We are the first, to the best of our knowledge, to take into account the topography of mining areas using an upstream-downstream comparison and to empirically quantify this effect at the scale of 604 mines in 26 countries of Sub-Saharan Africa over 1981-2020.

We find that the opening of industrial mines increases by 25% the 24-month mortality rate among villages located downstream compared to villages located upstream, and thus indirectly isolates the channel of water pollution. We find almost no effects on

[23] Burkina Faso, Cote d'Ivoire, Democratic Republic of the Congo, Ethiopia, Ghana, Guinea, Liberia, Madagascar, Malawi, Mali, Niger, Nigeria, Senegal, Sierra Leone, Tanzania, Togo, Uganda, and Zambia are EITI members. Benin, Burundi, Kenya, Namibia, Rwanda, and Zimbabwe have not joined the EITI.

other children's health outcomes, such as anthropometric measures, cough, fever, diarrhea, or anemia. We exploit the variation of the opening of a mine and show that our results are not driven by a change in women's fertility behavior, differential access to piped water, electricity, or health facilities but mainly by mining-induced water pollution, as children who were given plain water show increased mortality. The heterogeneity in the consumption of plain water seems to explain the null result on the 12-month mortality rates, as we observe a significant increase in the 12-month mortality rates exclusively for those who consume plain water. This can be interpreted as a proxy for having non-exclusive breastfeeding.

In an additional heterogeneity analysis, we show that our results are mainly driven by the pollution occurring during the time of mining activity. We find that the effects are even more harmful in rural areas, for open-pit and foreign-owned mines, and in places with a high density of mines. We also find that the effects increase with productivity intensity (proxied by international commodity prices). We run manyfold robustness checks and find that our results hold when controlling for in-migration, and when restricting to a balanced sample which deals with the issue of repeated cross-section surveys. Our results are also robust to the heterogeneous treatment effects estimator of de Chaisemartin and d'Haultfœuille, 2020, to measurement error tests, and a battery of placebo tests such as spatial and temporal randomization inference tests.

# Bibliography


Africa Bank (2022). *Mining Industry Prospects in Africa*. https://blogs.afdb.org/fr/afdb-championing-inclusive-growth-across-africa/post. Accessed: 2022-04-28.

Ahlerup, Pelle, Thushyanthan Baskaran, and Arne Bigsten (2020). "Gold Mining and Education: A Long-run Resource Curse in Africa?" In: *The Journal of Development Studies* 56.9, pp. 1745–1762.

Aragón, Fernando and Juan Pablo Rud (2016). "Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana". In: *The Economic Journal* 126.597, pp. 1980–2011.

Atkin, David (Aug. 2016). "Endogenous Skill Acquisition and Export Manufacturing in Mexico". In: *American Economic Review* 106.8, pp. 2046–85.

Bazillier, Remi and Victoire Girard (2020). "The gold digger and the machine. Evidence on the distributive effect of the artisanal and industrial gold rushes in Burkina Faso". In: *Journal of Development Economics* 143, p. 102411.

Benshaul-Tolonen, Anja (Aug. 2018). "Local Industrial Shocks and Infant Mortality". In: *The Economic Journal* 129.620, pp. 1561–1592.

Berman, Nicolas et al. (June 2017). "This Mine Is Mine! How Minerals Fuel Conflicts in Africa". In: *American Economic Review* 107.6, pp. 1564–1610.

Bialetti, Anca et al. (2018). "Lease Splitting and Dirty Entrants: The Unintended Deforestation Consequences of India's Environmental Clearance Process Reform". In: *PEDL Research Papers*.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". In: *Revisions Requested from Review of Economic studies*. arXiv: 2108.12419 [econ.EM].

Briffa, Jessica, Emmanuel Sinagra, and Blundell Renald (Sept. 2020). "Heavy metal pollution in the environment and their toxicological effects on humans". In: *Heliyon*.

Callaway, Brantly and Pedro Sant'Anna (2021). "Difference-in-Differences with multiple time periods". In: *Journal of Econometrics* 225.2. Themed Issue: Treatment Effect 1, pp. 200–230.

Chuhan-Pole, Puna, Andrew L Dabalen, and Bryan Christopher Land (2017). "Mining in Africa : Are Local Communities Better Off?" In: *Africa Development Forum* Washington, DC: World Bank and Agence Francaise de developpement.

Cobbina, Kumi, and Myilla (2013). "Small Scale Gold Mining And Heavy Metal Pollution: Assessment of Drinking Water Sources In Datuku In The Talensi-Nabdam District." In: *International Journal of Scientific and Technology Research* 2, pp. 96–100.

Coelho, Patricia and Joao Paulo Texeira (2011). "Mining Activities: Health Impacts". In: *Elsevier.*

Colella, Fabrizio et al. (2019). "Inference with Arbitrary Clustering". IZA Discussion Paper n. 12584.

Conley, Timothy (Sept. 1999). "GMM estimation with cross sectional dependence". In: *Journal of Econometrics* 92.1, pp. 1–45.

Corno, Lucia and Damien de Walque (Feb. 2012). "Mines, migration and HIV/AIDS in southern Africa". In: *Journal of African Economies* 21.

Cossa, Hermínio et al. (2022). "Assessing the effects of mining projects on child health in sub-Saharan Africa: a multi-country analysis". In: *Globalization and Health.*

Cust, James and Steven Poelhekke (2015). *The Local Economic Impacts of Natural Resource Extraction.* OxCarre Working Papers 156. Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

de Chaisemartin, Clément and Xavier d'Haultfœuille (Sept. 2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110.9, pp. 2964–96.

Dietler, Dominik et al. (2021). "Impact of mining projects on water and sanitation infrastructures and associated child health outcomes: a multi-country analysis of Demographic and Health Surveys (DHS) in Sub-Saharan Africa". In: *Global Health.*

Dike, Ibiwari et al. (Jan. 2020). "Evaluation of Pb concentrations in selected vegetables and portable drinking water, and intelligent quotients of school children in Ishiagu-a Pb mining community: health risk assessment using predictive modeling". In: *Environ Monit Assess* 192(2):126.

Do, Quy-Toan, Shareen Joshi, and Samuel Stolper (2018). "Can environmental policy reduce infant mortality? Evidence from the Ganga Pollution Cases". In: *Journal of Development Economics* 133, pp. 306–325.

Duflo, Esther and Rohini Pande (May 2007). "Dams*". In: *The Quarterly Journal of Economics* 122.2, pp. 601–646.

Edwards, David et al. (Dec. 2013). “Mining and the African Environment”. In: *Conservation Letters* 7.

Fängström, Bob et al. (2008). “Breast-feeding protects against arsenic exposure in Bangladeshi infants”. In: *Environmental health perspectives,*

Garg, Teevrat et al. (2018). “(Not so) gently down the stream: River pollution and health in Indonesia”. In: *Journal of Environmental Economics and Management* 92, pp. 35–53.

Girard, Victoire, Teresa Molina-Millan, and Guillaume Vic (2022). “Artisanal mining in Africa”.

Glöser, Simon et al. (Apr. 2017). “The cobweb theorem and delays in adjusting supply in metals’ markets”. In: *System Dynamics Review* 32.

Goodman-Bacon, Andrew (Sept. 2018). *Difference-in-Differences with Variation in Treatment Timing.* Working Paper 25018. National Bureau of Economic Research.

Greenstone, Michael and Rema Hanna (Oct. 2014). “Environmental Regulations, Air and Water Pollution, and Infant Mortality in India”. In: *American Economic Review* 104.10, pp. 3038–72.

He, A et al. (Aug. 2020). “Potentially toxic metals and the risk to children’s health in a coal mining city: An investigation of soil and dust levels, bioaccessibility and blood lead levels”. In: *Environ Int* 141:105788.

El-Kady, Ahmed A and Mosaad Attia Abdel-Wahhab (2018). “Occurrence of trace metals in foodstuffs and their health impact”. In: *Trends in Food Science and Technology* 75, pp. 36–45.

Khan, Tehmina, Trang Nguyen, and Richard Ohnsorge Franziska an Schodde (2016). “From Commodity Discovery to Production”. In: *Policy Research Working Paper* World Bank, Washington, DC.7823.

Kotsadam, Andreas and Anja Tolonen (2016). “African Mining, Gender, and Local Employment”. In: *World Development* 83, pp. 325–339.

Madilonga, Rofhiwa et al. (June 2021). “Water Quality Assessment and Evaluation of Human Health Risk in Mutangwi River, Limpopo Province, South Africa”. In: *Int J Environ Res Public Health* 18(13):6765.

Malpede, Maurizio (Apr. 2021). “The Dark Side of Batteries: Education and Cobalt Mining in the Democratic Republic of the Congo”. In: *SSRN.*

Mamo, Nemera, Sambit Bhattacharyya, and Alexander Moradi (2019). “Intensive and extensive margins of mining and development: Evidence from Sub-Saharan Africa”. In: *Journal of Development Economics* 139, pp. 28–49.

McQuilken, James and Rachel Perks (2020). *2020 State of the Artisanal and SmallScale Mining Sector. Washington, D.C.: World Bank.* Tech. rep. World Bank.

Obasi, NA et al. (Apr. 2020). "Metal pollution and human health risk assessment of soils and edible plants in farmlands around Enyigba lead-zinc mining site, Ebonyi State, Nigeria". In: *Environ Monit Assess* 192(5):292.

Oje, OA et al. (Sept. 2010). "Environmental pollution levels of lead and zinc in Ishiagu and Uburu communities of Ebonyi State, Nigeria". In: *Bull Environ Contam Toxicol* 85(3):313-7.

Ploeg, Frederick van der (June 2011). "Natural Resources: Curse or Blessing?" In: *Journal of Economic Literature* 49.2, pp. 366–420.

Shubhayu, Saha et al. (2011). "Under-mining health: Environmental justice and mining in India". In: *Health  Place* 17.1. Health Geographies of Voluntarism, pp. 140–148.

Strobl, Eric and Robert Strobl (2011). "The distributional impact of large dams: Evidence from cropland productivity in Africa". In: *Journal of Development Economics* 96.2, pp. 432–450.

Taylor, Cliff et al. (2009). "Geology and Nonfuel Mineral Deposits of Africa and the Middle East". In: *U.S. Geological Survey*.

UN Environment Program (2022). *Our work in Africa*. https://www.unep.org/regions/africa. Accessed: 2022-04-28.

UNESCO (2019). *The United Nations world water development report 2019: leaving no one behind*. UNESCO World Water Assessment Programme.

VanDerSlice, John, Bob Popkin, and John Briscoe (1994). "Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health." In: *Bulletin of the World Health Organization, 72(4), 589–601.*

Venables, Anthony (Feb. 2016). "Using Natural Resources for Development: Why Has It Proven So Difficult?" In: *Journal of Economic Perspectives* 30.1, pp. 161–84.

Von der Goltz, Jan and Prabhat Barnwal (2019). "Mines: The local wealth and health effects of mineral mining in developing countries". In: *Journal of Development Economics* 139, pp. 1–16.

# Chapter 3

# Impacts of repetitive droughts and the key role of experience : evidence from Nigeria

## Abstract

Western African Sahel faced severe droughts in the 1980s, affecting agricultural production and food security. In recent decades, farmers have faced uncertainty in the timing and amount of rainy seasons and are confronted with erratic rainfall with high interannual variations. Can the experience of past dry events reduce the vulnerability of households to short-term rainfall shocks? In this paper, I match three waves of panel household surveys focusing on agriculture in Nigeria (GHS, from 2010-2016) and high temporal resolution precipitation data set from the Climate Hazard Center (CHIRPS). I show evidence of the extreme importance of the long-dry period of the 1980s and identify more recent droughts in 2013/2015, which are in line with a change in the characteristics of the rainfall trends. Through a two-way-fixed effect strategy, I exploit the spatial variation of the exposition to the 2015 drought. First, I look at the short-term effects of being hit by a drought on agricultural production and food security indicators. I show that being hit by a drought decreases yields by 14%, and decreases the food diversity of households by around 1%. Second, I look at the impacts' heterogeneity according to the plot's experience, using the timing of the year of acquisition of the plot. I compare short-term droughts' effects on households that acquired their first plot before the 1980s dry period to those that acquired it after. Results suggest that acquiring the land before 1985 attenuates the

harmful effects of a climate shock, as these particular households have only a 3% reduction in their yields due to the 2015 drought. This is especially the case when households were severely hit in the 1980s. This result is only descriptive and can not lead to any causal interpretation. It might suggest that having a long-lasting experience under extreme dry events on cultivated land reduces vulnerability to rainfall variability.

## 3.1 Introduction

Reducing the sensitivity of agricultural production to climate shocks is a key factor in tackling food insecurity. In particular for Sub-Saharan countries, where most of the population is rural, involved in predominant rainfed agriculture, soils suffer from aridity and are vulnerable to droughts (Benson and Clay, 1998) and smallholder farmers face malnutrition and limited resources (Lobell et al., 2008). Historical higher temperature and rainfall fluctuations have reduced economic output in Africa, and especially agricultural productivity [1] and farm income, increasing the gap with developed countries (Barrios, Ouattara, and Strobl, 2008; Dell, Jones, and Olken, 2012; Nordhaus, 2006) [2].

If rainfall strongly declined during the prolonged dry period of the 1970s-1980s in the Sahel, climatologists have observed a partial recovery of seasonal precipitation in most recent decades (Nicholson, 2005), leading to the re-greening of the Sahel (Brandt et al., 2015; Fensholt et al., 2012). Satellite-based analysis of vegetation greenness in semi-arid areas has found an increase in the vegetation index, the NDVI, used as a proxy for vegetation production (Fensholt et al., 2012). However, the NDVI signal combines leaf biomass of woody species and herb biomass, and Brandt et al., 2015 have shown that the greening phenomenon was born by tree species, which are more resilient to droughts than herbs. On the contrary, the dynamics of biodiversity were found to decline over the study period. There is also a debate about the return of normal precipitations (Michela Biasutti, 2019). Since the 1990s, Sahelian farmers have reported changes in rainfall characteristics, noticing fluctuations in the timing, amount, and pattern of the rains (decreases during shorter rainy seasons) (Tambo and Abdoulaye, 2013). Analyses of rainfall trends from gauges find that the recovery results from increases in daily rainfall intensity rather than in frequency, rains being concentrated in the late rainy season and away from the west coast (Giannini et al., 2013; Panthou et al., 2018). Unlike temperature trends, predictive models for Sahel rainfall changes due to climate change and its impact on yields are also uncertain but point towards more variation in precipitations (Biasutti et al., 2008). Schlenker and Lobell, 2010 predict serious future damages of temperature increase for maize production in Sub-Saharan Africa. Sultan et al., 2013 have realized several projec-

[1] case of Millet in Niger

[2] Under the assumption of climate-economy equilibrium, (Nordhaus, 2006) finds in a global cross-section analysis that 20 percent of the income differences in Africa relative to high-income regions can be explained by geographic variables, including temperature and precipitation, but also elevation, soil quality. Looking at annual changes in historical temperature, Dell, Jones, and Olken, 2012 show that being 1°C warmer reduces per capita income by 1.4 %, but only in poor countries.

tions to quantify the yield responses of varieties of mil and sorghum over Africa to a pair of temperature and rainfall anomalies. They show that future responses and patterns will be very different from historical ones, as past mean yields were mainly vulnerable to rainfall anomalies, while higher temperatures will shape future ones.

Such uncertainty about seasonal precipitations is an important challenge for agriculture. Smallholder farmers in developing countries, who lack credit (Cole et al., 2013; Banerjee et al., 2015), and lack information about suitable measures and knowledge of climate change, might be able to mitigate negative impacts using adaptive strategies. However, farmers' perceptions of climate change and variability are hard to assess directly or even proxy. Besides, understanding the long-run effects of climate on agricultural production and other socio-economic outcomes requires distinguishing the historical multiple responses and to link inter-annual and longer-term patterns of precipitations. This project's first goal is to contribute to the literature by linking the recent rainfall variability to the longer-term evolution of the rains. Particular attention is devoted to the analysis of rainfall variability and its relation to changes in precipitation trends over the country. Refuting the theory of the recovery of rainfall over Nigeria, this analysis identifies important droughts over the more recent decades, from 2013-2015, which are linked to less frequent and more intense rains in the long-rain in the Gulf-Guinean part of the country since the dramatic dry year period of the 1980s.

I match socio-economic data from a three-round panel survey, the Nigerian General Household Survey Panel (GHS) to high-resolution historical precipitation data, the CHIRPS product. This paper's research question is to assess the effects of short-term droughts on agricultural outputs and food security. I use a two-way fixed effect strategy over a three years survey panel in 2010, 2012, and 2015. I exploit the spatial variation of droughts occurring before/during the last wave. Second, the main goal is to understand the heterogeneity of the impacts using retrospective questions. Ideally, I aim to test whether past exposure to the severe dry period of the 1980s explains the capacity to adapt to recent rainfall shocks and to reduce the negative impacts on yields and food security. For the moment, I only test a reduced form and run a heterogeneity analysis according to the timing of acquisition of the first plot of the household. I compare the impact of recent rainfall shortages for households that acquired their first plot before the 1980s dry year to those who acquired it later, the hypothesis being that having experience of your own plot under drastic dry conditions might increase your knowledge of climate change, good practices and how to adapt, and thus reduce the effect of recent droughts. The main research question is: can the experience of past dry events reduce the vulnerability of households to rainfall shocks?

The first main result of the paper shows that facing a dry year in the recent decade decreases yields by 14%, and implies a reduction in the diversity food score of households, losing 0.14 food group over 12 (1.2% reduction). The heterogeneity analysis shows that the results are mainly driven by households who acquired their land after the 1980s dry period, especially those that were severely hit by the 1980s droughts. The heterogeneity analysis suggests that being hit by a drought decreases the yields on average by 19% for households that acquired their first land after 1985, in comparison to those who acquired it previously, for which the effect is attenuated by 16%, facing a 3% decrease. This result suggests that working on the same land that was hit by the intense droughts from the 1980s reduces the vulnerability to the 2015 drought. As the year of land acquisition is an endogenous variable, this is only descriptive evidence, and these results can not be interpreted in a causal way. This is a suggestion of the role that plays experience, knowledge, and past exposure to intense dry years. Based on a reduced form, I can not directly conclude whether this means that land acquisition before past extreme events results in better adaptive strategies and a better perception or knowledge. However, this result suggests that having a long-lasting experience of the cultivated land reduces vulnerability to rainfall shocks, especially when having experience of the land under past extreme dry conditions.

The remainder of the paper is organized as follows. Section 3.2 presents the context in light of the literature. Section 3.3 describes the data, the context, as well and the statistical analysis of long trends of climate change and how they can be linked to recent rainfall variability. Section 3.4 details the main empirical strategy. Section 3.5 introduces the results and Section 3.6 the heterogeneity analysis, while Section 3.7 proposes a list of robustness checks. Section 3.8 concludes.

## 3.2 Literature review

Assessing the historical and future effects of climate change and variability on agricultural production requires understanding how farmers adapt or, if not possible, account for it. This is the main limitation of agronomic studies, which construct crop models based on plant physiology to predict how climate change will directly affect yields (R. Adams, 1989; R. Adams et al., 1995; Antle and Stöckle, 2017; Asseng et al., 2015) [3]. These studies usually ignore adaptation strategies due to the lack of information on farmers' behaviors and practices and then overestimate the impact of climate change on yields.

The Ricardian method is based on Ricardo's approach that land values reflect land productivity and allow for adaptation (land is put to best use) (Kurukulasuriya, Mendelsohn, et al., 2006; Seo, Mendelsohn, et al., 2009; Seo and Mendelsohn, 2007; Mendelsohn and Dinar, 2003; Sanghi and Mendelsohn, 2008; Wood and Mendelsohn, 2014; Fleischer, Lichtman, and Mendelsohn, 2008; Kurukulasuriya and Ajwad, 2006). Net revenue or profit from farms are used as a proxy for land values and are regressed on temperature and precipitation, with environmental and socioeconomic controls such as soil quality, latitude for day length, population density for access to market, and opportunity costs of the land. Large-scale studies (Mendelsohn and Dinar, 2003; Schlenker, Hanemann, and Fisher, 2005; Kurukulasuriya, Mendelsohn, et al., 2006; Seo, Mendelsohn, et al., 2009; Seo and Mendelsohn, 2007) have applied the Ricardian model to large countries such as the US (Schlenker, Hanemann, and Fisher, 2005; Mendelsohn and Dinar, 2003), India and Brazil (Sanghi and Mendelsohn, 2008), and at continental level, and compare values across climatic regions. In a cross-sectional analysis of the two latter countries, Sanghi and Mendelsohn, 2008 find lower climate sensitivity of agriculture than agronomic models based on yields, because of adaptation mechanisms [4]. Continental cross-sectional Ricardian studies showed the heterogeneity of the sensitivity to climate change within Africa. Kurukulasuriya and Mendelsohn, 2008 emphasize the importance of crop switching as an adaptation strategy for farmers. Looking at primary crop choices and production for 11 African countries in 2003, they look at the marginal effects of climate on conditional net revenue, taking into account the probability of crop switching. Results show that

[3] Asseng et al., 2015 even shows that crops models are less accurate at higher temperatures

[4] When comparing simulations based on their Ricardian analysis to findings from the agronomic literature, the authors find lower climate sensitivity of agriculture (net revenue reduction between 7-17 % in India for a warming of 3.5 with a 7 % precipitation increases, *vs* yields losses between 30-40 %)

farmers adapt their crop choices endogenously to the climate they face, leading to smaller losses.

Smaller scale studies restrict to more homogeneous areas displaying important climate variations (Fleischer, Lichtman, and Mendelsohn, 2008; Kurukulasuriya and Ajwad, 2006; Wood and Mendelsohn, 2014; Ouedraogo, 2012; Molua and Lambi, 2007), Deressa, Hassan, and Poonyth, 2005. In the Fouta Djallon area (Northern Guinea and Southern Senegal) Wood and Mendelsohn, 2014 show that the effect of temperature increases and precipitation variations on net revenue depends on the season considered, and that production losses from the summer and rainy seasons can be balanced by benefits in the winter.

Ricardian approaches argue that the net revenue reductions are lower than predicted losses in yields because farmers' adaptation strategies and potential adjustments are taken into account. However, the Ricardian approach is a partial equilibrium analysis. As agricultural prices are assumed to remain constant, the comparability between negative effects on net revenue and yields does not hold if markets are not integrated, and Ricardian analysis might underestimate the negative effects of climate on crops. Ricardian methods might be biased due to omitted variables acting as confounders of climatic variables in a cross-section. One concern is the forgotten role of irrigation, implicitly relying on a cost fee adaptation, and cannot be used to estimate dynamic adjustments costs (Cline, 1996; Schlenker, Hanemann, and Fisher, 2005) [5]. As positive effects of irrigation water access are higher in hotter areas, cross-sectional estimates of the effect of temperature/rains on land values are biased Schlenker, Hanemann, and Fisher, 2005. [6].

In contrast, panel studies (Blanc and Schlenker, 2017; Dell, Jones, and Olken, 2012; Deschênes and Greenstone, 2007) look at exogenous year-to-year variations in temperature and precipitation and use location-fixed effects to absorb time-invariant factors. They rule out confounding variation, accounting for the fact that areas might differ in other variables correlated to climate. They estimate the effects of inter-annual variations in temperature and rainfall on yields or profits from deviations from location-specific means. Barrios, Ouattara, and Strobl, 2008 and Schlenker and

[5] Cline, 1996 reproaches to Ricardian analysis to implicitly assume the infinitely elastic supply of irrigation water at today's prices, potentially wrong for Sub-Saharan countries.

[6] Schlenker, Hanemann, and Fisher, 2005 empirically shows that land values vary for dry land versus irrigated American counties.

Lobell, 2010 use panel analyses to assess the response of yields to climate change for specific Sub-Saharan crops and suggest that well-fertilized modern seed varieties are more sensitive to climate variations.

However, effects and adaptive responses from short-term variations are likely to differ from the ones from climate change in the longer run (Dell, Jones, and Olken, 2014; Auffhammer et al., 2013). Panel studies look at weather shocks, and because they only take into account coping strategies, they are not informative enough to predict the future evolution of agricultural output. It is still unknown whether short-run responses to weather will increase or decrease damages from long-run global warming. The sign of the bias introduced by estimations from panel data is still debatable. It is commonly assumed that short-run coefficients overestimate the negative effects of longer-run changes on agricultural outcomes because, in the short run, farmers might not have the time to find available adaptive strategies. Otherwise, Schlenker, Hanemann, and Fisher, 2005 give the example of pumping groundwater for irrigation during punctual drought as a short-run adaptation to weather anomalies, which is not tenable in the long run because of limited resources. This is an example where short-run analysis might underestimate the long-run impacts of climate on yields under adaptation. Trying to give an answer to this debate for US trends, Burke and Emerick, 2016 compare estimates of the impacts of temperature and precipitation of long difference versus panel strategies. They find negative responses of productivity to decadal changes but cannot distinguish them from responses to annual variation in extreme heat in the same period.

Thus, it is hard to distinguish the different impacts of short-run from long-run climatic factors under adaptation behaviors. Besides, if farmers can choose strategies that will increase yields, such as changing planting dates or crop variety, risk-averse ones might decide to shift to activities less dependent on rainfall and temperature, such as migration, tree planting, or diversification of business activities. If that is the case, studies will misinterpret the impact of climate on agricultural productivity because it does not distinguish between different adaptive responses. One way to investigate this further is to rely on a two steps analysis, taking into account the link between perception of climate change, weather, and adaptation strategies (Maddison, 2007; Tambo and Abdoulaye, 2013; Silvestri et al., 2012; Komowski et al., 2015). In a case study in the region of Djougou in Benin, Komowski et al., 2015 find that the most used strategies are tree plantation, the shift of planting dates, and the use of new crop choices (new crops or mixed crops). Farmers do not directly identify climate change and variability as the main reason for changes in practices and seem to respond to favor short-run food security. In the Nigerian Savannah, the majority of respondents have noticed both a decrease in rainfall and changes in the timing of the rains (Tambo and Abdoulaye, 2013). The most common adaptation is the use of drought tolerant and early maturing varieties but also changes in dates, irrigation, afforestation, and off-farm income diversification. Limitations of this two steps literature are that self-reported perceptions might be biased and that perceptions are not enough to generate adaptive behaviors (credit constraints for poverty-trapped households). Adaptation might happen under collective behaviors rather than based on individual perceptions, and adaptation can reduce vulnerability to climate change without being made under this purpose.

Despite debate about the recovery of rainfall and the re-greening of Sahel since the 1990s, evidence points towards more erratic precipitations (Michela Biasutti, 2019) and decline of biodiversity (Brandt et al., 2015). Farmers face uncertainty in the timing, amount, and pattern of precipitations. Understanding the long-run effects of climate on agricultural production requires distinguishing the multiple historical responses. If Ricardian methods look at long-run impacts and take into account farmers' adaptation by using land values (often proxied by net revenue), they are biased due to omitted variables acting as confounders of climatic variables in a cross-section. Using exogenous inter-annual changes in rain and temperature, panel data rule out confounding variations. However, they take into account coping strategies in response to short-run shocks, that might differ from adaptation to long-run global warming. The sign of the bias introduced by short-run estimates is not direct. It is not clear whether estimation from panel data relying on year-to-year variations over or underestimates future impacts of longer-run climate change. This is especially the case for Sub-Saharan countries, highly dependent on agricultural activity and with smallholder farmers that might be credit-constrained, lacking information, or risk-averse. As there is a lack of studies differentiating between the effects of different types of adaptive behaviors, this research will contribute by comparing responses and damages linked to inter-annual and longer-run fluctuations. An important part of the paper is to link recent climate variability to long-term patterns and to assess whether exposure to past dry conditions might affect recent responses to rainfall shortages.

## 3.3 Data and Context

In this paper, I match socio-economic data from the Nigerian General Household Survey-Panel (GHS), which is a four waves panel survey with a strong focus on agriculture, to the CHIRPS product for rainfall.

### 3.3.1 Socio-economic data

The socioeconomic variables are built from the GHS, a panel survey conducted by the Nigerian Bureau of Statistics (NBS, 2012), and the World Bank as part of the Living Standards Measurement Survey - Integrated Surveys on Agriculture (LSMS-ISA). The survey is a stratified two-stage sample design. Within each of the six Nigerian geopolitical zones, the Enumeration Areas (EAs, also mentioned as villages in this paper) were firstly selected with a probability proportional to the size, and then a random sampling procedure was used to select surveyed households within each EAs. The four waves of the LSMS-ISA are a subsample of the GHS-Panel Sample, which is initially made of 5000 households from 500 EAs, each contributing 10 households. The survey is nationally representative, as well as representative of the Nigerian geopolitical zones. The GHS has been conducted through four waves in 2010/2011, 2012/2013, 2015/2016, and 2018/2019. In each wave, households are visited twice over a 12-month period in order to collect detailed information on agricultural activities. Both post-planting, from September to November, [7] and post-harvest, from February to April, data were collected with Agriculture, household, and community questionnaires.

The GHS-Panel has been conducted over four waves, however, there has been a partial refresh of the sample for the last wave During the fourth wave, 3600 households have been refreshed, added to a subsample of the original panel from 2010. The long panel, including the four waves, includes only 1447 households from 157 EAs. For this reason, this paper focuses only on the three waves, 2010/2011, 2012/2013, and 2015/2016 in order to build a 5 years panel, which is described in Table C.1 in Section C.0.1 in Appendix. The final sample used from the three waves includes 4162 households from 463 EAs, households that have stayed within their 2010 village. Please note that 189 households migrated and were tracked in the second and/or third waves but were not included in the final analysis. Table C.2 displays attrition rates for the second and third waves and shows the levels of attrition overall in

[7] For the three last waves, the timing of the post-planting survey was the same, from September to November. In the first round, the post-planting occurred in August-October 2010 instead

Nigeria at the household level (2.7% for the second wave *vs* 13% for the third wave). Table C.2 identifies higher levels of attrition in the North East and the South West of the country for the 2015/2016 wave.

The coordinates of the EAs have been modified to keep the data anonymous, and the displacement procedure relies on a random offset of cluster center coordinates (the average of household GPS within each EAs) within specific ranges (0-2km for urban areas, while 0-5km offset for rural areas). As the distance between each EAs is higher, there is no mismatch between villages, and I am able to match EAs with their respective climate characteristics as the clusters of climate data are around 5km.

The GHS survey collects rich information on household on-farm and off-farm livelihoods, total agricultural production, agricultural practices, food security outcomes, and welfare variables. The panel dimension makes it possible to control for omitted variables and to adjust for time and spatial-specific confounders. I will exploit a balanced panel in order to capture the heterogeneity in household outcomes and choices when facing rainfall variability.

#### 3.3.1.1 Mains Variables

The main variables of interest are measured at the household level for each GHS wave. This analysis focus on agricultural production and food security variables. Work-in-progress is made in order to capture the impact of droughts on livelihood strategies - including off and on-farm activities, income diversification (Fowowe, 2020), short-term migration (Ghebru et al., 2019)- as well as other agricultural choices - such as technology adoption (Fadare, Akerele, and Toritseju, 2014), farm diversification (Ayenew et al., 2018), land fragmentation (Veljanoska, 2018). Another variable of interest in this paper, which is used in the heterogeneity analysis, is the year of acquisition of the plot by the household. This section displays some context and descriptive statistics for the main variables.

### Agricultural Production

The main variables computed to account for agricultural production are yields. Crop yields are the total crop production per land area planted [8] and rely on self-reported information, both on crop production and cultivated area. Total crop production is computed for each crop, on each plot of each household per survey rounds. The quantities are computed in kilograms (kg), measured using the conversion factors given by the World Bank, as self-reported crop production is displayed in non-standard measurement units [9]. The planted area for each crop is given in hectares (ha) [10] and is self-reported by the household (same for the harvested area).

Self-reported crop yields suffer from measurement errors and are subject to non-classical measurement errors (over-estimated on smaller plots) (Yacoubou Djima and Kilic, 2021; Carletto, Gourlay, and Winters, 2015). The source of bias is twofold, as both the numerator (total crop production) and the denominator (cultivated land area) face complications (Yacoubou Djima and Kilic, 2021). Self-reported crop productions suffer from potential recall bias, a high probability of rounding the numbers (Wollburg, Tiberti, and Zezza, 2021), and the noise introduced by the use of non-standard measurement units and conversion factors. Accordingly, self-reported land areas suffer from the use of conversion factors and rounding numbers (Carletto, Gourlay, and Winters, 2015) and display discrepancies from GPS-based measures (Yacoubou Djima and Kilic, 2021). Literature tends to conclude that farmers tend to over-report land areas, all the more at, the lower end of the plot area distribution. Overall, self-reported yields tend to be over-estimated for smaller plots when compared to objective measures, which suggests non-classical measurement errors linked to these data (Yacoubou Djima and Kilic, 2021). Table C.4 from Section C.0.1 in the Appendix displays descriptive statistics of the main variables of interest and shows the discrepancies between GPS measures and Self-reported measures of land areas. In this paper, I try to correct self-reported yields by treating outliers according to different techniques. [11]. Another solution will be to look at the correlation of yields

[8] Robustness checks will be done to compare yields per land area planted *vs* per land area harvested

[9] Households report harvested quantities in kg/gram/liter for standardized measures, but also in number of bags, baskets, basins, bundles, wheelbarrows.

[10] The land area is measured using conversion factors as well, as the planted area is displayed in the LSMS-ISA in heaps, ridges, stands, plots acres, hectares, and sqmeters

[11] For the main analysis, the outliers are imputed at the median. For now, outliers are simply defined as the 10% and 90% percentiles of the distribution, work-in-progress is made to change the identification of the outliers, such as data points whose z-score is below the third standard deviation. Robustness checks are made, and yields are winsorized and trimmed at 10% and 5%.

(at aggregated levels) with satellite image products and to correct yields in places with the lowest rates of correlation. Work-in-progress ongoing, using the NDVI and the Global Dataset of Historical yields for major crops, the GDHY (Iizumi and Sakai, 2020; Wing, De Cian, and Mistry, 2021).

Work-in-progress is made in order to capture variations in agricultural production using another measure than self-reported yields, such as agricultural income (which is, unfortunately, a noisy measure as well). This is also the reason why I also address the effects of rainfall shocks on food security outcomes.

**Food Security**

Food insecurity is mainly driven by four dimensions, including food availability, food access, utilization, and stability (Bertelli, 2019). In this paper, I will focus on the two first dimensions, which are food availability and food access. Food availability is captured using a food insufficiency measure, which is a dummy indicating whether, in the past 12 months, the household have been faced with a situation when it did not have enough food to feed the household [12]. Food access is measured using two indicators. Firstly, I use the food security scale score, also named the Food Insecurity Experience Scale (FIES), which captures the level of food insecurity based on 8 questions on the experience of the last 7 days. These experience questions are listed in Table C.3, and the variable displays the number of days when the household faced the particular situation (ranges from 0, never occurred to 7, occurred every day). The FIES is built by summing up all the responses. We follow the strategy from Bertelli, 2019, and in the main analysis, I reverse the score so that the higher the FIES, the more food secure the household, and we standardize the indicator. Finally, I build the Household Dietary Diversity Score HDDS which captures food diversity and is measured as the number of food groups that the households have consumed during the seven days preceding the survey. The HDDS has been computed by Swindale and Bilinsky, 2006, and gathers 12 different food groups : (1) cereals, (2) root and tubers, (3) vegetables, (4) fruits, (5) meat/poultry and offal, (6) eggs, (7) fish/seafood, (8) pulses/legumes/nuts, (9) milk and milk products, (10) oil/fats, (11) sugar/honey, (12) miscellaneous. The main difference between the HDDS from Swindale and Bilinsky, 2006 is that it is computed based on the household consumption from the last 7 days, *vs* the last day from Swindale and Bilinsky, 2006. Descriptive statistics

[12]this indicator can be also used as a continuous variable, which indicated the number of months of critical situation

of the indicators are displayed in Table C.4 from Section C.0.1 in Appendix.

#### Year of Land Acquisition

Another variable of interest used in the heterogeneity analysis is the year of acquisition of the land. This is a single variable fixed over waves for each household. For households cultivating several plots, I define the year of land acquisition as the year of acquisition of their first plot. Figure 3.9b plots the distribution of this variable. If discrepancies existed for the year of acquisition of a particular plot across the three waves, I favored the year that was given by the manager of the plot. If discrepancies still persisted, I took the minimum amongst the year given by GHS. More than the timing of acquisitions, the survey gives insights into how the plot was acquired by the household. Overall, 5.7% acquired it via *"outright purchase"*, 7.5% because *"rented for cash or in-kind goods from"* an outside person. 8.6% respond having acquired the land *"free of charge"*, while 78% report acquiring it because it was *"distributed by community or family, or family inheritance"*. Only the last wave disentangles between *"distributed by community or family"* and *"Family inheritance"*, and shows that the majority of households (70.5% *vs* 7.5) inherited the plot. Inheritance might play an important role in the mechanisms of results from Section 3.6, and work in progress is done in order to better understand the social and cultural norms of land inheritance in Nigeria.

### 3.3.2 Climate data

I use the CHIRPS product by the Climate Hazard Center (CHC), which combines a satellite-based rainfall product (CHIRP [13]) with station observations data. It gives a good spatial (0.05 lat/long), and temporal (daily, decadal, and monthly) resolution for historical (1981-2019) mean, maximum and minimum precipitations. It has been validated over Africa and assessed as the best satellite-based product (Dinku et al., 2018). For temperature, I use the CHIRTS product, also from the CHC, which also combines satellite and station-based estimates of maximal temperature ($T_{max}$), with the same spatial and temporal resolution as CHIRPS (Funk et al., 2019).

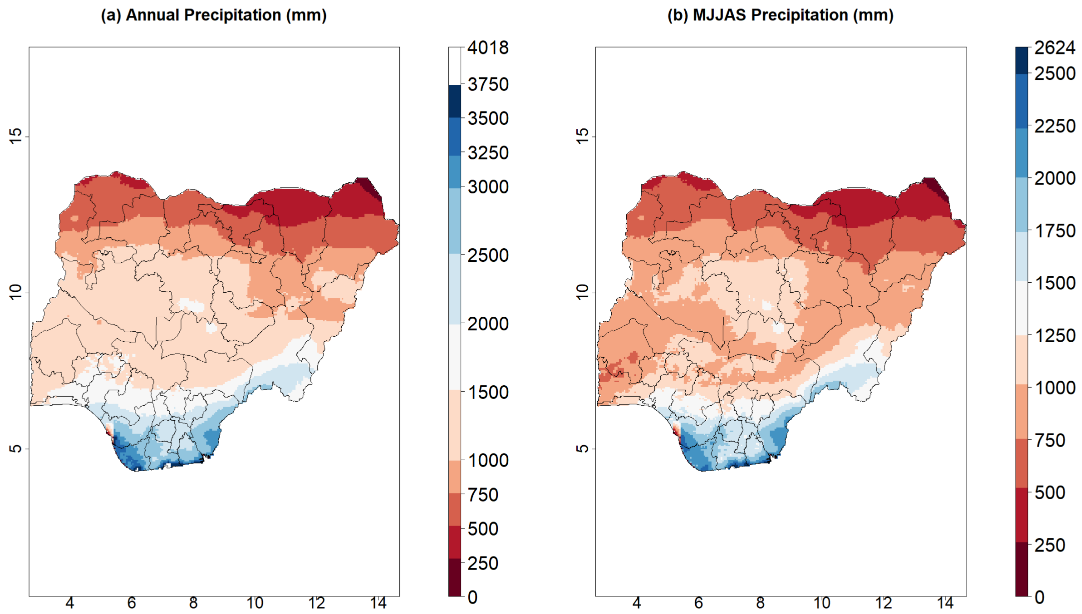
[13] Climate Hazards Group Infrared Precipitation

#### 3.3.2.1 Context and climatology of Nigeria

Nigeria is a context highly dependent on agriculture and rain-fed activities, as 70% of households are engaged in crop farming activities, and 47% own or raise livestock, which makes it highly vulnerable to climate events. Nigeria has an impressive population (the most populated African country, around 200 million), but lacks of adaptive capacity due to low financial and technological tools, weak institutions, and low knowledge of climate change. Nigeria includes an important part of Western African farmers that have faced climate change and food security issues and had to adapt to rainfall changes over time. Amongst the adaptive strategy described in the literature, portfolio diversification, changing dates of planting, planting trees, and use of irrigation have been identified in the southern Nigeria rainforest zone (Onyeneke and Madukwe, 2010; Sofoluwe, Tijani, and Baruwa, 2011). Crop diversification, but also the change of crop varieties to drought/early mature resistant varieties, and farm relocation are used as adaptive strategies in the northern part of the country (Dabi et al., 2008). Based on a field survey within the Nigerian savanna, Tambo and Abdoulaye, 2013 shows that most of the farmers have noticed changes in rainfall patterns and that those who lack information on climate change are facing limitations in adapting. This shows the key role of the perceptions and experiences of climate change in the adaptive capacity of Nigerian farmers.

Nigeria is a diverse setting, with high heterogeneity of livelihood zones and climatology, from tropical rainforest and tropical monsoon in the south to tropical savanna and Sahel climate in the North. Overall, it has a tropical climate with two seasons, the wet season being from May to September (MJJAS) (cf Figure C.1). Annual precipitation is amongst the highest in Western Africa - especially in the Gulf of Guinea - with a long-term average (1981-2019) of annual rainfall being 1491mm, and of cumulative rains over the MJJAS wet season of 1101 mm. The country displays important differences in terms of climatology, which are shown in Figure 3.1, which maps the long-term rainfall average of annual precipitation (3.1 (a)) and of cumulative rains over the long-rainy season, MJJAS (3.1 (b)). The means are computed from the long-term period from 1981-2019. Nigeria is dominated by four climate types, changing in the meridional direction, from south to north. Monsoons from the Atlantic Ocean influence the south tropical monsoon climate (AM), which is the most humid region of the country, experiencing abundant rainfall (up to 3750mm long-term annual mean). The geography of the southern part of Nigeria is dominated by the Niger Delta, an important river area composed of deltas and humid mangrove

Figure 3.1: Rainfall long-term average




**Notes:**The Figures represent the spatial distribution of (a) the long-term annual average, based on the 1981-2019 long-term period, and (b) the long-term average of the long rainy season MJAS of cumulative precipitation. Units of long-term averages are in mm.
*Sources:* Authors' elaboration on CHIRPS

swamps. The tropical savanna climate (Aw) covers the major part of the country, and annual rainfall varies from 1000mm (lowlands) to 1500mm (southwestern). The zone is made of the Guinean forest-savanna (plains and tall grass/trees), then the Sudan savannah is arider (short grass/trees). The northern part of Nigeria lies within the Sahel and experiences a semi-arid climate (BSh), and has dramatic low rainfall, with annual means varying from 500mm to lower than 250mm. The Sahel savanna is mostly composed of grass and sand and is the aridest area o the country.

Across the country, seasonal rainfall patterns also vary sharply in the meridional direction. When averaging rainfall at the country level, the main rainy season, which corresponds to the primary agricultural season, extends from May to September (MJJAS), with a peak in August, as we can see in Figure C.1. However, the south of Nigeria has two rainy seasons (cf Figure C.2). The first rainy season starts from March to July (peak in June), which is followed by a short dry season (2/3 weeks) in August. The second wet season lasts from September to October. In the northern part of the country, the unique wet season is shorter, lasting from June to September. For the purpose of this paper, we define the main rainy season over the country, the MJJAS period, from May to September. Work-in-progress is made to refine the main rainy seasons according to the agroecological zones.

Going further, defining the main rainy season according to the geography, work-in-progress is ongoing to define the main agricultural season according to the main crop (following Wichern et al., 2019, we can obtain crop-specific parameters on rainfall using the R package dismo and the Ecocrop database from the FAO). Indeed, the main cultivated crops differ between the main zones, as shown in Figure C.4, mainly maize cassava and yam in the south and maize, guinea corn, and beans in the north.

#### 3.3.2.2 Long-term changes in rainfall patterns and characteristics

**The Sahelian Droughts**

Nigerian rainfall characteristics from the north, which is part of the Sahel, differ from the characteristics of the rainfall along the coast of the Gulf of Guinea, in the southern part of the country. However, the very dry decades of 1970-1980 hit both of the regions and had devastating impacts on food and livestock supply. Farmers and pastoralists of the zone had to find adaptive strategies, such as changing the timing of the planting, weeding, and harvesting using different types of crops and varieties, diversifying the livelihoods (Mortimore and W. Adams, 2001).

Western Africa Sahel faced severe droughts in the 1970s and in the 1980s. The 1980s decade included some of the most extreme droughts years on record (Dai et al., 2004; Nicholson, 2005; Nicholson, 2018), which were mostly enforced by sea surface temperature anomalies (Michela Biasutti, 2019). The dry events of the 1980s called the 'Sahelian droughts,' are said to be *among the most undisputed and largest recent climate changes recognized by the climate research community* (Dai et al., 2004). If these droughts are more associated with the Sahel because they were more pronounced there, they were also dramatic in the Gulf Guinea area. Figure 3.2 displays the departure of the rains for each pixel from the long-term mean (1981-2019) for each year of the period over Nigeria. It gives the spatial distribution of precipitation extremes for both deficits and intensive rains and displays as well the comparability of the magnitudes of the different events. In line with climatologic literature, Figure 3.2 clearly identifies the dry period of the 1980s and shows that both the Sahelian and Gulf Guinean parts were severely impacted mostly in the early 80s, 1982, and

1983 being intense dry years with national coverage. We observe the persistence of the dry period up to 1987, mostly in the Sahelian part of the region. Figure 3.2 also shows the high magnitudes of the 1980s dry events in comparison to the rest of the time period.

#### Debate on the recovery of the rains in the recent decades - Sahel region

Since the 1980s heavy rains over the Sahel and Guinea Gulf, the climatologic literature observes rainfall trends going upward, referring to the most recent period from 1990 to be a period of rainfall recovery for the Sahel (Dai et al., 2004; Nicholson, 2005) and the Guinea Gulf (Sanogo et al., 2015). However, the rainfall recovery is debated in the literature, both for the Sahelian (Michela Biasutti, 2019) and Gulf Guinean region (Bichet and Diedhiou, 2018).

More evidence is found in the literature concerning the Sahel, in particular, thanks to long-term observations such as the International African Monsoon Multidisciplinary Analysis (AMMA - CATCH) program (Panthou et al., 2018). First, evidence from field surveys show that farmers have noticed recent changes in climate, such as changes in the characteristics of the rain season and more erratic rains (Tambo and Abdoulaye, 2013). Second, statistical studies of rainfall data, both from station gauges and satellite observations, corroborate farmers' perceptions and refute the theory of the return to normal conditions. Giannini et al., 2013 analyze rainfall gauges in Burkina Faso and Senegal and find that the recovery of the rains is mainly born by daily rainfall intensity. Salack et al., 2014 look at interannual rains and intra-seasonal droughts episodes using stations in Senegal and Niger and show that if cumulative rains seem to have reached pre-1970s normal conditions, seasonal rainfall amounts are susceptible to an extreme that implies delayed start and cessation of cropping seasons. Accordingly, Panthou, Vischel, and Lebel, 2014 observe an increased probability of extreme daily rainfall looking at gauges from Benin, Burkina-Faso, and Niger.

Figures C.5, C.7 and C.8 from Section C.0.2 display the long-term trends of climatic indicators based on the CHIRPS product over the 1981-2019 period. Figure C.5 shows the trends of yearly precipitations (Figure C.5 (a)), and of wet season MJJAS cumulative precipitations (Figure C.5 (b)) over the long-term period of 1981-2019. Both annual and MJJAS precipitations display a significant increasing trend in the

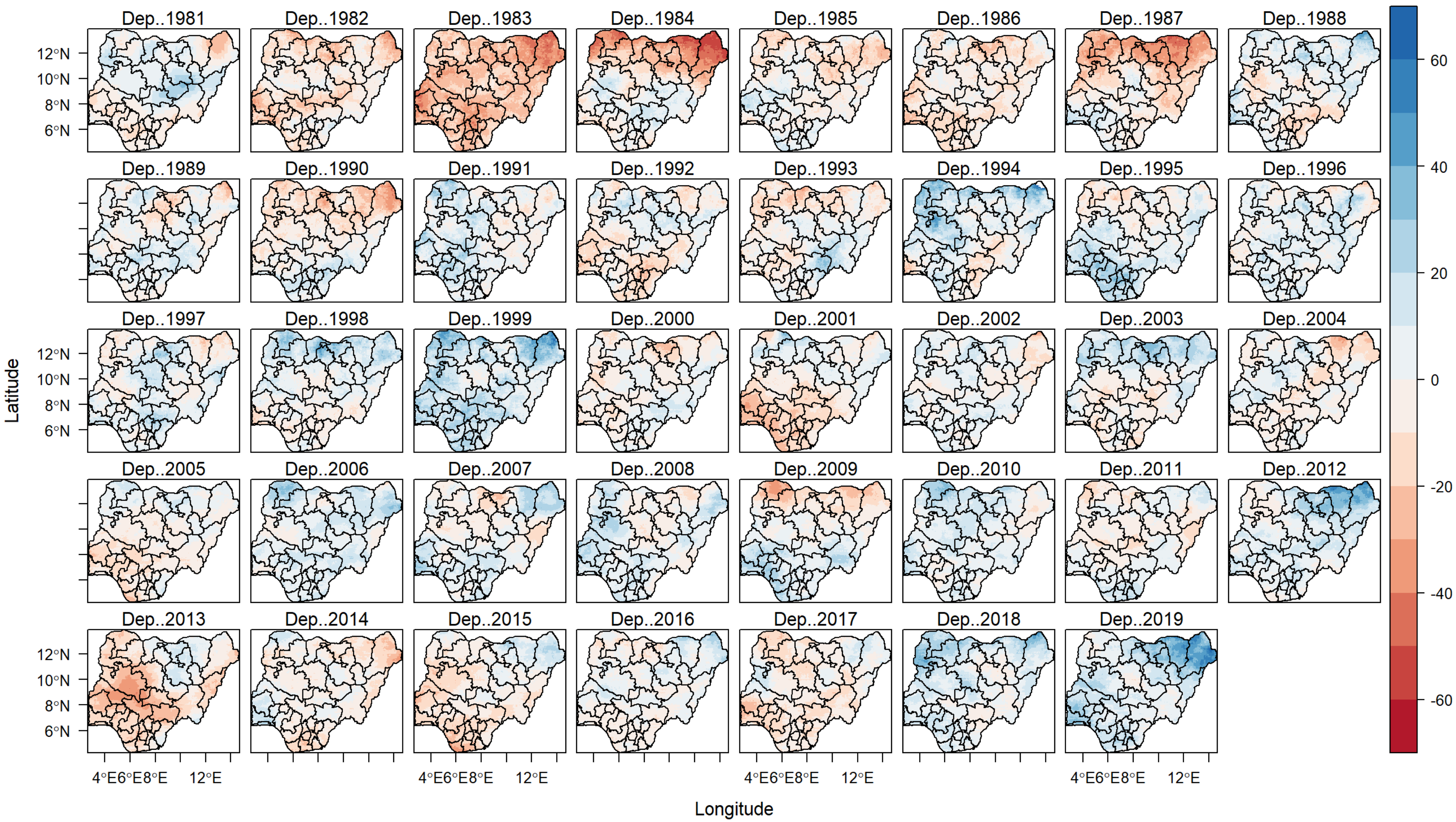
Sahelian part of the country, with up to an 11mm increase in the North East. This result is in line with the recovery of the cumulative rains after the 1980s dry decade in the Sahel, a phenomenon that I observe less in the south of Nigeria. Still, in the North, I observe significant changes in the characteristics of the rains. I observe significantly increasing trends in the MJJAS numbers of wet days (Figure C.7 (a)), in the number of extreme rains as well (C.7 (c)), and in the intensity of daily rains (Figure C.8 (a)). Trends in patterns and characteristics of the rains entirely depend on the long-term period chosen to describe the evolution. When taking the 1981-2019 long-term period, the severe droughts from the 1980s account for the evolution and mainly explain the partial recovery of cumulative rains. Figures C.6 reproduce the exact same figures as Figures C.5, but rely on the 30-year period from 1989 to 2019, instead of 1981-2019, excluding the extreme droughts from the 1980 decade. If these figures show that the cumulative rains seem to be increasing on average in the Sahelian region, the trends are no longer significant, suggesting interannual variability occurring in the more recent decades instead of steady trends.

#### Debate on the recovery of the rains in the recent decades - Gulf -Guinea region and Central Nigeria

Despite the importance of the dry 1980s period in the Gulf of Guinea region, less analysis is made in comparison to the Sahel. However, there is also debate in the literature about the recent decades being a period of partial recovery of the rains of the region (Sanogo et al., 2015; Bichet and Diedhiou, 2018). Using the CHIRPS product over 1981-2014 overall, the Gullf-Guinea region, Bichet and Diedhiou, 2018 find an absence of significant trends of rains during the wet season but find trends towards less frequent but more intense rainfall. The results from our study are in line with these results. In Figures C.5, C.7 and C.8, I display trends of rainfall indicators using the CHIRPS data over the 1981-2019 long-term period. Figure C.5 shows no significant trend evolution of the annual and wet season rains in the southern and central parts of Nigeria. However, Figure C.7 displays a significant decrease in the number of wet days over the main rainy season (Figure C.7 (a)), in the south and central regions. This is associated with a significant increase in the trends of the daily intensity of rains (Figure C.8 (a)), up to 0.22 mm/day, and a significant decrease in the length of the wet spells (Figure C.8 (b)) Consecutive Wet Days Index CWD.

These evolution in the characteristics of the rains are in line with the literature on the climate evolution of rains in the region of Gulf Guinea, and refutes the fact that recent decades represent a recovery period for Nigeria. We show that the absence of significant decreasing or increasing trends of cumulative rains in the Southern part of Nigeria hides a change in the characteristics of the rains. Rains over the wet season are becoming less frequent, more intense, and more concentrated, which is expected to increase the likelihood of extreme events such as droughts (Bichet and Diedhiou, 2018).

Figure 3.2: Rainfall percent departures of the annual rains from the 1981-2019 mean




**Notes:** The Figure plots the percent departure from the long-term mean of the main rainy season (1981-2019) for each pixel.
*Sources:* Author's elaboration on CHIRPS data.

#### 3.3.2.3 Main shock of interest

As explained previously, Nigeria faced a severe dry period over the 1980 decade (especially over the early years). Long-term trends evolution suggest less frequent but more intense rains in the south/central region in recent decades, which can cause extreme events such as droughts. If the long-term trends evolution of the Sahelian rains seem to increase significantly, it is entirely driven by the intensity of the dry years over the 1980s. The evolution of the 30 years period from 1981 to 2019 shows no significant trends, and evidence from the literature points to more erratic rains, which increases the probability of droughts as well. Section 3.3.2.2 shows the key role of the long-term period chosen when looking at rain patterns, which will play a key role in the definition of our main shock of interest.

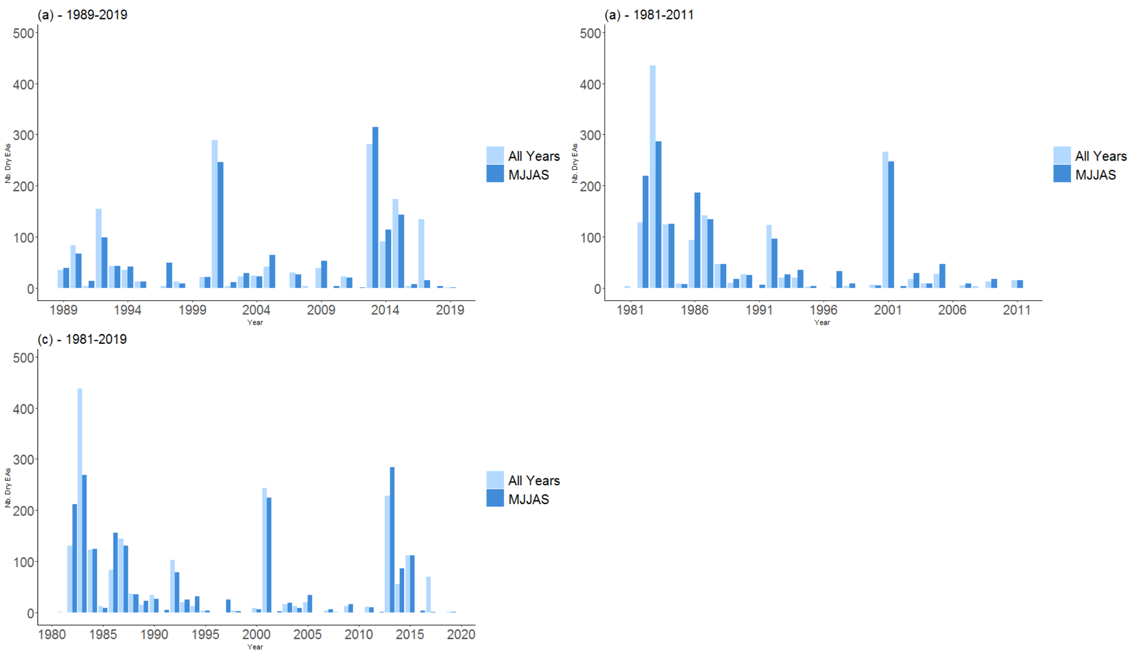
The first part of this paper intends to look at the effects of short-term rainfall variation over agricultural outputs, the GHS panel survey covering the 2010-2016 period. The main dependent variable for the main analysis is defined as follows. I construct a dummy for each year, based on the long-term mean (that will be defined) of each GHS EAs (also called villages in the paper), which equals 1 when the year/MJJAS is dry, 0 otherwise. I define a year/MJJAS season as dry if the cumulative rains over the year/MJJAS are lower than the 15th percentile of the cumulative rains for the EAs over the chosen long-term period. The climatic indicator I will use for the main analysis is whether the year of the survey is dry or if the agricultural season MJJAS is dry the year of the survey. The choice of the threshold is discussed in Section 3.7.3.1.

Each GHS survey has two rounds. For instance, for the second wave, the post-planting occurs from August to September 2012, while the post-harvest occurs from February to April 2013. The critical timing of the rains for this cropping season is the wet rainy season in 2012, which is defined from May to September, which is why I focus on the rains occurring in 2012. Figure C.3 plots this timeline between the cropping seasons and the post-planting and post-harvest GHS rounds.

Figure C.10 plots for each year the total number of EAs for which the dummy dry $D_{i,t}$ equals 1. The three figures plot the dry variable for three long-term periods, each of thirty years at least in order to capture climate change (Auffhammer et al., 2013). Figure C.10 (a) plots over the 1989-2019 period, in order to emphasize recent droughts. Figure C.10 (b) plots over 1981-2011 to capture the dry decade of 1980, while Figure C.10 (c) plots over the 1981-2019 period, which makes it possible the comparison between recent extreme events and the 1980s dry events. Figure C.10 (b) identifies the droughts from the 1980s, and underlines the importance of 1982, 1983, and 1984 as dry years (we see that 1983 has national coverage, as all EAs are dry). Figure C.10 (a) shows the 2001 dry year (El Nino phenomenon), and the increase in the frequency of dry events since 2010, in particular for the 2013, 2014, and 2015 years. Over the thirty-year period, normal conditions would imply a dry year occuring every decade. Having more than one dry year over the timing of the GHS panel is the shock that I exploit in order to look at the effect of rainfall variability on socioeconomic outcomes. Finally, Figure C.10 (c) compares the intensity of the three main dry periods over 1981-2019 (which are 1982-1984, 2001, and 2013-2015).

The main shock used in this analysis will be based on the thirty long-term mean from 1989-2019. I intentionally do not take into account the 1980s dry year in the construction of the independent variable, as they would have influenced the treated group from our Difference-in-Difference strategy (EAs particularly shocked

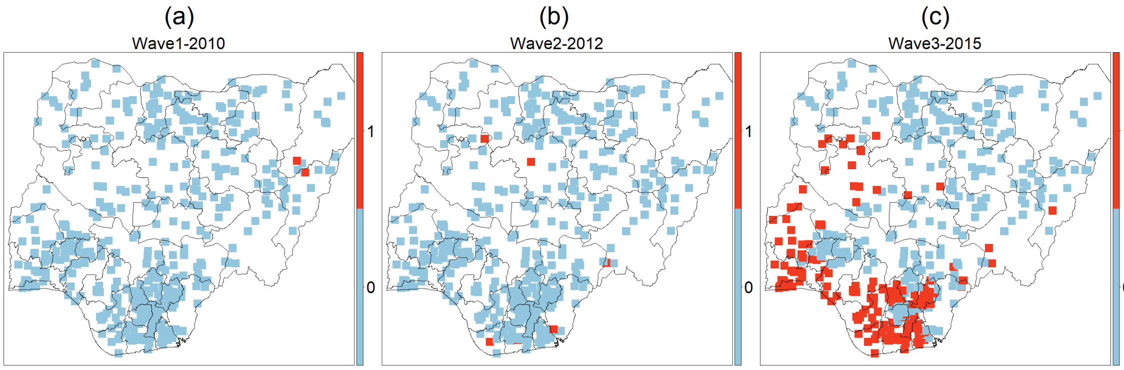
Figure 3.3: Number of dry EAs according to the long-term average




**Notes:** The Figures plot the number of GHS villages (over 483 EAs) for which the year is dry when the dry dummy is built according to the different long-term mean. Figure (a) plots the number of dry EAs per year, over the 1989-2019 long-term mean, Figure (b) according to the 1981-2011 long-term mean and Figure (c) according to the 1981-2019 long-term mean. Figure (a) points to the importance of the dry period from 2013-2016, Figure (b) to the dry period from 1981-1989, while Figure (c) makes it possible the comparison between the two dry spells. The Figure is made for the EAs from the GHS panel.
*Sources:* Authors' elaboration on CHIRPS and GHS data.

in the 1980s would have been mechanically counted as control, as the main dry years would have been concentrated in the 1980s). As the second goal of this paper is to understand whether the past experience of the 1980s changes the impacts of short-term rainfall variability on agricultural outcomes, I must avoid the independent variable to be *de facto* correlated to the exposition to the 1980s droughts. I must avoid accounting as treated EAs as those that were the least affected in the 1980s decade.

Figure 3.4: Spatial distribution of the number of dry rainy seasons -Binary treatment- current years

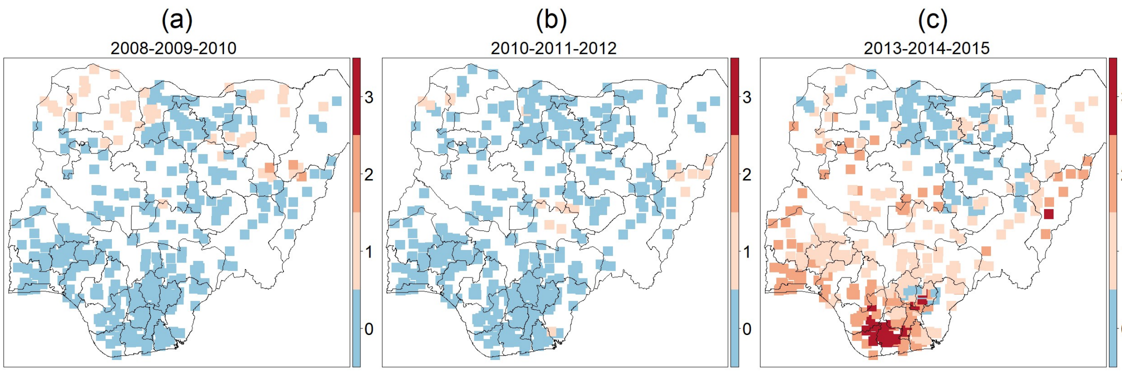



**Notes:**The Figures plot the dry rainy seasons for each GHS village from the panel, per GHS waves. Figure (a) plots the 2010 dry rainy seasons, Figure (b) in 2012, while Figure (c) in 2015. Dry years are defined according to the 1989-2019 long-term average.
*Sources:* Authors' elaboration on CHIRPS and GHS data.

Figures 3.4, 3.5, as well as C.9 and C.10 in Appendix display the spatial distributions of the shocks observed in Figure C.10. Figure 3.4 plots the spatial variation of the main independent variable and indicates for each survey year which villages are hit by a drought occurring over MJJAS. It shows the treated and controlled EAs that will play a key role in the two-way fixed effects analysis. The dry years are mainly concentrated in the last wave (2015-2016) and have spatial variation. The Southern and western parts of the country were mainly affected in 2015. Figure 3.4 shows that the 2015 drought is spatially clustered, which is discussed in Section 3.7.2.1. Figure 3.5 plots the number of dry years during the GHS survey year and the two years before for each wave, based on the 1989-2019 long-term mean. It shows that some areas were hit by cumulative droughts over the 2013-2015 period, which might have critical effects on agricultural production. The Southern part of the country is mainly affected, with several villages impacted in the three years, while the Central and North Eastern parts of the country shocks vary between two and one dry years.

Figure C.9 from C.0.3 maps the spatial distribution of the C.10 (b) and shows the spatial distribution of the 1980s dry period. Figure C.10 shows the spatial distribution of Figure C.10 (c), and underlines the relative importance of the shocks in the 1980s to the ones occurring over 2011-2019.

Figure 3.5: Spatial distribution of the number of dry rainy seasons - current years




**Notes:** The Figures plot the number of dry rainy seasons for each GHS village from the panel, per GHS waves. Figure (a) plots the number of dry rainy seasons from 2008 to 2010 (including), Figure (b) from 2010 to 2012, while Figure (c) from 2013 to 2015, the maximum being three dry rainy seasons. Dry years are defined according to the 1989-2019 long-term average.
*Sources:* Authors' elaboration on CHIRPS and GHS data.

## 3.4 Empirical strategy

### 3.4.1 Identification Strategy

The main identification strategy relies on a two-way fixed effects (TWFE) with a binary treatment. I estimate the effect of being hit by a drought the year of the survey on socio-economic outcomes including yields and food security indicators. The TWFE regression is made over the three first waves of the GHS, in 2010, 2012, and 2015, with both household and year-fixed effects, on a balanced panel of households. The treatment is defined at the EA level, as defined in Section 3.3.2.3: it is the dummy being under a dry year, defined according to the 1989-2019 long-term mean period. GHS villages are mainly treated in wave 3, as shown in Figure 3.5. The group of switchers is EAs that have experienced a change in their treatment status over the three waves. More formally, the empirical strategy can be formally written as follows :

$$Y_{h,i,t} = \alpha_0 + \alpha_1 \times D_{i,t} + \alpha_2 X_{h,t} + \gamma_h + \gamma_t + \epsilon_{i,t} \tag{3.1}$$

Where $X_{h,t}$ account for socio-economic characteristics of the households, including the age, gender, and level of education [14] of the household head, as well as the number of adults (aged over 15), which can be used as a proxy for labor endowment. $\gamma_h$ and $\gamma_t$ are household and time-fixed effects adjusting for spatial and period-specific confounders. $D_{i,t}$ is the dummy indicating a drought during the year of the survey round, which is displayed for each EAs in Figure 3.5. Errors $\epsilon_{i,t}$ are clustered at the EA level. Finally, $Y_{h,i,t}$ represents a socio-economic outcome. For agricultural production, $Y_{h,i,t} = log(yields_{h,i,t})$ where $yields_{h,i,t}$ are the yields of the household at wave t, defined as $yields_{h,i,t} = \frac{Q_{h,t}}{A_{h,t}}$, with $Q_{h,t}$ equals to the self-reported total crop production of the household at wave t (in kg) and $A_{h,t}$ the self-reported total planted land holding (h). For food security indicators, $Y_{h,i,t}$=$FIED_{h,t}$ or $HDDS_{h,t}$, as defined in Section 3.3.1.1.

The second aim of this paper is to understand whether experience plays a role in the capacity of farmers to face rainfall shocks. I run a heterogeneity analysis in order to assess whether the year of land acquisition can explain the results from regression 3.1. The year of land acquisition (defined and described in Section 3.3.1.1) is a dummy, which equals 0 if the household cultivates at least one plot that he has acquired before 1985, and 1 if after. The Year of land acquisition is a variable that proxies the experience of the household on his plot. The threshold of 1985 is used in order to account for the fact that the household has faced the main dry years of the 1980s (1982/1983/1984) working the same land that he works in recent years. The choice of the threshold year is discussed in Section 3.6. More formally, the role of the year of land acquisition on the impacts of rainfall variability is estimated following the equation :

$$Y_{h,i,t} = \alpha_0 + \alpha_1 \times D_{i,t} + \alpha_2 \times D_{i,t} \times L_h + \alpha_3 X_{h,t} + \gamma_h + \gamma_t + \epsilon_{i,t} \tag{3.2}$$

[14] The level of education of the household head is defined as the following : 0 if the head has no diplomas, 1 if he/she completed primary school 2 if he/she completed secondary school

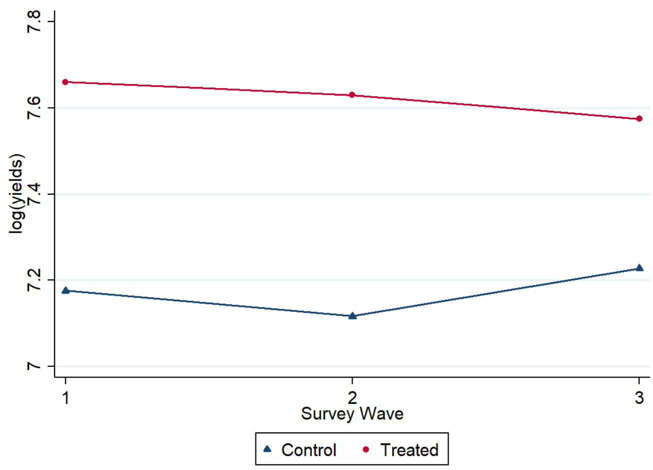
Where $L_h$ is the dummy of land acquisition.

### 3.4.2 Common trend assumption

The key assumption of a difference-in-difference (DiD) strategy is that the dependent variable would follow the same trends in the absence of droughts both in treated and control villages. In this section, I test for the common trend assumption using pre-treatment data. As Figure 3.4 shows, the drought mainly occurred during the last wave in 2015. For this test, I implemented a simple DiD design where I select villages that were not hit by any drought over wave 1 and wave 2. Treated villages were hit by the 2015 drought, while control villages were not. With this design, the evolution of yields over wave 1 and wave 2 is pre-treatment observation data that I use to test for the common trend assumption.

Figure 3.6 plots the linear trends of the logarithm of yields across the three survey rounds and distinguishes between treated and control villages. Figure 3.6 shows that yields follow similar trends over waves 1 and 2, suggesting that the treated and control villages follow a similar pattern of agricultural production. This test is only descriptive as it does not take into account any controls or fixed effects.

Figure 3.6: Linear trends of agricultural production across treatments




**Notes:** The Figure plots the linear trends of the log(yields) across survey rounds, averaged over-treated and control groups, defined as being hit by the drought in 2015.
*Sources:* Authors' elaboration on CHIRPS and GHS data.

## 3.5 Results

### 3.5.1 Agricutural Productivity

Table 3.1: Effects of short-term droughts on yields

| | Annual drought | | MJJAS drought | |
|---|---|---|---|---|
| | All crops | Main crops | All crops | Main crops |
| Yield (log) | (1) | (2) | (3) | (4) |
| Drought | -0.140** | -0.128** | -0.148** | -0.111* |
| | [0.0588] | [0.0598] | [0.0624] | [0.0625] |
| Nb. Adults | -0.0185 | -0.00523 | -0.0179 | -0.00449 |
| | [0.0148] | [0.0153] | [0.0148] | [0.0153] |
| Gender head | -0.00713 | 0.0252 | -0.00520 | 0.0258 |
| | [0.106] | [0.111] | [0.107] | [0.111] |
| Age head | -0.00271 | -0.00441** | -0.00261 | -0.00430** |
| | [0.00216] | [0.00219] | [0.00215] | [0.00216] |
| Education head | -0.0180 | -0.0298 | -0.0196 | -0.0314 |
| | [0.0431] | [0.0437] | [0.0424] | [0.0433] |
| Observations | 5274 | 5183 | 5274 | 5183 |
| R2 | 0.523 | 0.537 | 0.523 | 0.537 |
| log(yields) Mean | 7.305 | 7.344 | 7.305 | 7.344 |
| Yields Mean | 2369 | 2491 | 2369 | 2491 |
| Household FE | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 3.1 displays results from estimation 3.1, showing the effect of droughts on agricultural outputs [15]. Columns (1) and (3) give the results on the yields for all types of crops, while columns (2) and (4) for the main crops cultivated in Nigeria [16]. Columns (1) and (2) display the effects of a drought defined on an annual basis, while columns (3) and (4) during the main rainy season MJJAS. All estimations are made on a balanced panel of crops, meaning that each household intervenes three times in the regression.

The results show that being hit by drought during the main rainy season decreases

[15] I have to account for the fact that I am estimating a semi-log functional form, the yields being measured in terms of the log of the yields, while the independent variable is a dummy).

[16] The list of the main crops is: cassava, maize, sorghum, cowpeas, yam, millet, groundnut, rice, cocoyam and oil palm tree.

yields by around 14%, which is significant at 5%. On average, this corresponds to a decrease in yields by 333 kg/ha. This magnitude is in line with results from the literature, Veljanoska, 2018 finding, for instance, that one rain deviation more reduces yields by 6.6%.

### 3.5.2 Food Security

Table 3.2 displays the results for the household security indicators. Columns (1) and (2) give the results of an annual drought, while columns (3) and (4) of a drought occurring during the agricultural season MMJAS. The gender of the household head is determinant in terms of food insecurity. I observe no statistically significant result regarding the FIES indicator. Regarding the food diversity score, I observe that being hit by drought during MJJAS implies that the household loses around 0.14 food group over 12, which corresponds to a 1.2% decrease. This result suggests that droughts decrease the food diversity of households, which is only significant at the 10% level.

Table 3.2: Effects of short-term droughts on food security

| | Annual drought | | MJJAS drought | |
|---|---|---|---|---|
| | FIES | HDDS | FIES | HDDS |
| | (1) | (2) | (3) | (4) |
| Drought | -0.318<br>[0.268] | -0.0149<br>[0.0834] | -0.264<br>[0.243] | -0.140*<br>[0.0806] |
| Nb. Adults | -0.0526<br>[0.0645] | -0.0180<br>[0.0201] | -0.0521<br>[0.0643] | -0.0209<br>[0.0200] |
| Gender head | -1.622***<br>[0.464] | -0.240**<br>[0.110] | -1.620***<br>[0.464] | -0.236**<br>[0.110] |
| Age head | 0.0155<br>[0.0115] | 0.000662<br>[0.00320] | 0.0158<br>[0.0115] | 0.000693<br>[0.00320] |
| Education head | 0.231<br>[0.149] | 0.0737<br>[0.0486] | 0.230<br>[0.149] | 0.0739<br>[0.0486] |
| Observations | 11721 | 12123 | 11721 | 12123 |
| R2 | 0.570 | 0.631 | 0.570 | 0.632 |
| Outcome mean | -3.210 | 8.218 | -3.210 | 8.218 |
| Household FE | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$.

## 3.6 Heterogeneity analysis

The main research question of this paper is to understand the role of experience and of exposure to the 1980s droughts on the impacts of short-term droughts. We proceed to a heterogeneity analysis, looking at the heterogeneity of the results from Section 3.5 according to the year of acquisition of the first plot purchased by the household and whether this plot was acquired before or after the main droughts of 1980s. As the year of land acquisition and the exposure to past droughts are highly endogenous variables, this section is mainly descriptive and intends to give some suggestive insights into the role that plays exposure to past climate events on households' vulnerability. Results can not lead to any causal interpretation.

### 3.6.1 Year of land acquisition

Table 3.3 gives the results for the yields of the main crops, while Table 3.4 for the HDDS. All estimations are made of a balanced panel for each dependent variable. From now on, I will only look at the effects of droughts occurring during the rainy MJJAS season, as it captures critical shocks of the cropping season. I control for the number of adults in the household and for the age, gender, and especially the educational level of the household head.

Columns (1) and (2) give the results of regression 3.2, when the short-term drought interacted with the dummy of the year of land acquisition. Column (2) control for the effect of the shock interacted with the household head's age, which is another measure of experience. Table 3.3 column (1) shows that short-term drought decreases yields by 19% for households that acquired their first land after 1985, in comparison to those who acquired it previously. The interaction term suggests that having acquired the land before 1985 attenuates the negative effect of being hit by a drought on yields. For households that acquired the land before, being hit by a drought decreases yields by 3%: the decrease is attenuated by 16 p.p in comparison to the decrease faced by households who acquired the land after. This result suggests that working on the same land that was hit by the intense droughts from the 1980s reduces the vulnerability to the 2015 drought. This is a suggestion of the role that plays experience, knowledge, and past exposure to intense dry years.

As the year of land acquisition is an endogenous variable, the interaction term does not directly capture exposure to the 1980s drought. Column (2) shows that the result does not longer hold when controlling for the dummy drought interacted with the age of the household head. Thus, columns (3) and (4) rely on another dummy, called *Dummy exposure*, which equals 1 if the household has acquired its land before 1985 and has been exposed to severe droughts over 1982/1983/1984, 0 otherwise [17].

I consider that a household has been exposed to severe droughts in the 1980s if it has been hit by at least two droughts. In order to limit endogeneity issues, I construct for the 1980s the dry dummy year in comparison to the long-term period from 1981 to 2011. In this sense, I avoid the fact that being hit by droughts in the 1980s decreases *de facto* the likelihood of being hit by a drought in more recent years (as the dummy is defined based on the 1989-2019 mean). However, endogeneity between recent and 1980s droughts still remains, as being severely hit in the past might still increase the probability of being hit by intense droughts in recent years, through environmental degradation, for instance. Again, this result is only descriptive and can not be interpreted in a causal way.

Column (3) suggests that the negative effects of the recent drought on yields are mainly driven by individuals for which the dummy *exposure* is null. The interaction term suggests that individuals exposed to the 1980s droughts have an attenuation of the yield decrease by 38 p.p, compared to the yield decrease of households that were not exposed. This suggests an over-reaction of these households, for which being hit by drought increases yields by 20% (-0.156+0.38). This might be explained by the fact that households implement adaptation strategies, whose benefits outweigh the negative effects of droughts. Again, this is only a suggestive insight, not a causal one, and the magnitude effects are relatively large. Column (4) controls for the interaction of the recent shock and the age of the household.

Table 3.4 displays the same analysis for the HHDS and shows little effect on food security.

[17] I do not run a triple interaction, mean that I do not directly interact the year of land acquisition and exposure to the 1980s drought because of endogeneity issues

Table 3.3: Effects of short-term droughts on yields - Heterogeneity according to year of land acquisition

| | Main crops | | | |
|---|---|---|---|---|
| Yield (log) | (1) | (2) | (3) | (4) |
| Drought (MJJAS) | -0.198** | -0.222 | -0.156** | -0.250 |
| | [0.0863] | [0.192] | [0.0657] | [0.194] |
| Drought × Acquired Land Before 1985 | 0.164* | 0.160 | | |
| | [0.0991] | [0.105] | | |
| Drought × Dummy exposure | | | 0.380*** | 0.377*** |
| | | | [0.111] | [0.111] |
| Drought × Age | | 0.000472 | | 0.00174 |
| | | [0.00325] | | [0.00309] |
| Nb. Adults | -0.00289 | -0.00281 | -0.00240 | -0.00199 |
| | [0.0154] | [0.0153] | [0.0153] | [0.0152] |
| Gender head | 0.0160 | 0.0160 | 0.0313 | 0.0304 |
| | [0.112] | [0.112] | [0.112] | [0.112] |
| Age head | -0.00408* | -0.00414* | -0.00441** | -0.00460** |
| | [0.00215] | [0.00216] | [0.00215] | [0.00217] |
| Education head | -0.0271 | -0.0273 | -0.0294 | -0.0296 |
| | [0.0429] | [0.0428] | [0.0432] | [0.0432] |
| Observations | 5183 | 5183 | 5183 | 5183 |
| R2 | 0.538 | 0.538 | 0.538 | 0.538 |
| log(yields) Mean | 7.344 | 7.344 | 7.344 | 7.344 |
| Household FE | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, ${}^{*}p < 0.1,{}^{**}p < 0.05,{}^{***}p < 0.01$.

Table 3.4: Effects of short-term droughts on HDDS - Heterogeneity according to year of land acquisition

| | HDDS | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Drought (MJJAS) | -0.144 | -0.241* | -0.122 | -0.230 |
| | [0.107] | [0.144] | [0.100] | [0.147] |
| Drought × Acquired Land Before 1985 | 0.122 | 0.106 | | |
| | [0.120] | [0.123] | | |
| Drought × Dummy exposure | | | 0.304 | 0.298 |
| | | | [0.231] | [0.230] |
| Drought × Age | | 0.00221 | | 0.00231 |
| | | [0.00220] | | [0.00217] |
| Nb. Adults | -0.0408* | -0.0397* | -0.0401* | -0.0389* |
| | [0.0232] | [0.0231] | [0.0233] | [0.0233] |
| Gender head | -0.133 | -0.130 | -0.130 | -0.126 |
| | [0.133] | [0.133] | [0.132] | [0.132] |
| Age head | -0.000516 | -0.000895 | -0.000602 | -0.000980 |
| | [0.00383] | [0.00388] | [0.00383] | [0.00388] |
| Education head | 0.0821 | 0.0817 | 0.0802 | 0.0801 |
| | [0.0567] | [0.0566] | [0.0567] | [0.0566] |
| Observations | 8793 | 8793 | 8793 | 8793 |
| R2 | 0.613 | 0.613 | 0.613 | 0.613 |
| Outcome Mean | 8.006 | 8.006 | 8.006 | 8.006 |
| Household FE | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

### 3.6.2 Exposure to the 1980s droughts

As the likelihood of being hit by droughts in the 1980s is endogenous to the likelihood of being hit in more recent years, the interaction with the *Dummy exposure* raises endogeneity issues. To deal with this issue, I investigate in this section the role of the year of land acquisition for villages that were severely hit in 1980 on one side and for villages that were less hit in the 1980s on the other side.

Table 3.5 gives the results for the yields of the main crops. Column (4) to (6) focuses on the sample of villages that were hit by two droughts in the early 1980s, while Column (1) to (3) focuses on the other less exposed villages. Columns (1) and (4) give the effects of being hit by a short-term drought on yields. It shows that for villages that were not highly affected in the 1980s, being hit by a dry rainy season decreases yields by 15.6%. This effect does not hold for villages hit in the 1980s, which suggests that the negative effects on yields are mainly driven by households who were not exposed to the 1980s droughts.

Columns (2) and (5) look at the interaction with the dummy of the year of land acquisition. Column (4) shows that for the comparison across households less hit in the 1980s, the decrease is only significant for households who acquired their land later. However, the year of land acquisition does not seem to play a key role as the interaction term is not significant. Column (5) shows that, for villages that were highly affected by the 1980s droughts, there is no longer a significant decrease in agricultural production. On top of that, the year of land acquisition seems to play a significant role, as, for households who acquired their land before 1985, the recent drought increases yields by 35%, in comparison to those who acquired it later. This is in line with the results found in Table 3.3, with large magnitude effects, raising endogeneity issues. Columns (6) and (7) control for another measure of experience, the interaction of the shock with the age of the household age.

Table 3.6 displays the same analysis for the HDDS indicator. Column (1) shows that being hit by a short-term drought decreases the food diversity by 0.3 food group, which corresponds to a 2.5% decrease. As column (4) displays no significant effects, it shows that the negative effects of short-term drought on food diversity are mainly driven by households that were less affected in the 1980s.

Table 3.5: Effects of short-term droughts on yields - Heterogeneity according to exposure to the 1980s droughts

| | Main crops | | | | | |
|---|---|---|---|---|---|---|
| | Not hit by 1980s droughts | | | Hit by 1980s droughts | | |
| Yield (log) | (1) | (2) | (3) | (4) | (5) | (6) |
| Drought (MJJAS) | -0.156** | -0.227** | -0.458* | 0.0241 | -0.130 | 0.202 |
| | [0.0787] | [0.0987] | [0.239] | [0.122] | [0.188] | [0.282] |
| Drought × Acquired Land Before 1985 | | 0.126 | 0.0935 | | 0.351* | 0.433** |
| | | [0.107] | [0.113] | | [0.200] | [0.212] |
| Drought × Age | | | 0.00447 | | | -0.00737 |
| | | | [0.00402] | | | [0.00553] |
| Nb. Adults | 0.00153 | 0.00284 | 0.00404 | -0.0136 | -0.00996 | -0.0101 |
| | [0.0207] | [0.0208] | [0.0207] | [0.0224] | [0.0224] | [0.0226] |
| Gender head | 0.0410 | 0.0327 | 0.0324 | -0.0527 | -0.0517 | -0.0505 |
| | [0.125] | [0.126] | [0.127] | [0.262] | [0.267] | [0.266] |
| Age head | -0.00760*** | -0.00736*** | -0.00808*** | 0.000302 | 0.000351 | 0.000805 |
| | [0.00262] | [0.00260] | [0.00262] | [0.00349] | [0.00346] | [0.00344] |
| Education head | -0.0411 | -0.0370 | -0.0394 | -0.0193 | -0.0137 | -0.0127 |
| | [0.0561] | [0.0556] | [0.0553] | [0.0693] | [0.0678] | [0.0668] |
| Observations | 3185 | 3185 | 3185 | 1998 | 1998 | 1998 |
| R2 | 0.515 | 0.515 | 0.516 | 0.523 | 0.525 | 0.525 |
| log(yields) Mean | 7.500 | 7.500 | 7.500 | 7.094 | 7.094 | 7.094 |
| Household FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.6: Effects of short-term droughts variability on HDDS - Heterogeneity according to exposure to the 1980s droughts

| | HDDS | | | | | |
|---|---|---|---|---|---|---|
| | Not hit by 1980s droughts | | | Hit by 1980s droughts | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Drought (MJJAS) | -0.281*** | -0.215 | -0.305* | 0.0117 | -0.120 | -0.173 |
| | [0.108] | [0.137] | [0.172] | [0.128] | [0.200] | [0.279] |
| Drought × Acquired Land Before 1985 | | 0.0554 | 0.0384 | | 0.358 | 0.351 |
| | | [0.133] | [0.138] | | [0.257] | [0.259] |
| Drought × Age | | | 0.00226 | | | 0.00119 |
| | | | [0.00270] | | | [0.00403] |
| Nb. Adults | -0.0402 | -0.0464* | -0.0449 | 0.00616 | -0.0299 | -0.0297 |
| | [0.0247] | [0.0280] | [0.0279] | [0.0334] | [0.0406] | [0.0406] |
| Gender head | -0.231* | -0.151 | -0.146 | -0.255 | -0.0938 | -0.0941 |
| | [0.133] | [0.155] | [0.155] | [0.195] | [0.257] | [0.257] |
| Age head | -0.00206 | -0.00245 | -0.00292 | 0.00387 | 0.00195 | 0.00183 |
| | [0.00409] | [0.00486] | [0.00496] | [0.00505] | [0.00620] | [0.00623] |
| Education head | 0.0151 | 0.0451 | 0.0448 | 0.141** | 0.124 | 0.124 |
| | [0.0658] | [0.0761] | [0.0760] | [0.0708] | [0.0844] | [0.0845] |
| Observations | 7230 | 5400 | 5400 | 4893 | 3393 | 3393 |
| R2 | 0.643 | 0.617 | 0.617 | 0.590 | 0.574 | 0.574 |
| Outcome Mean | 8.492 | 8.281 | 8.281 | 7.815 | 7.568 | 7.568 |
| Household FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Results from the heterogeneity analysis are based on a reduced form. It is not possible to conclude whether this means that the acquisition of the land before past extreme events results in better adaptive strategies and a better knowledge of perception. However, I show that having a long-lasting experience of the cultivated land reduces vulnerability to rainfall shocks, especially when having experience of the land under past extreme dry conditions. This is only a piece of descriptive evidence.

### 3.6.3 Endogeneity and discussion

The year of land acquisition is a highly endogenous variable. It might be endogenous to climate shocks, both those occurring in recent years and those that occurred in the 1980s, as farmers make their choice according to the accumulation of droughts. Besides, it is, of course, correlated to other measures of experience such as the age of the household head or the educational level. Even if the heterogeneity analysis is not a causal exercise, it might still be interesting to investigate to which variables the year of land acquisition is correlated.

Table 3.7: Table of correlation -land of year acquisition

| | Acquisition before 1985 | Year of land acquisition |
|---|---|---|
| | (1) | (2) |
| Drought 2015 (MJJAS) | 0.0247 | -1.740 |
| | [0.0347] | [1.448] |
| Drought 1980 (MJJAS) | -0.0357 | 1.446 |
| | [0.0299] | [1.211] |
| Nb. Adults | 0.00656 | 0.0383 |
| | [0.00501] | [0.201] |
| Gender head | 0.0317 | -1.863 |
| | [0.0256] | [1.142] |
| Age head | 0.0106*** | -0.429*** |
| | [0.000739] | [0.0318] |
| Education head | -0.0402** | 1.624*** |
| | [0.0164] | [0.567] |
| Observations | 2931 | 2931 |
| Outcome mean | 0.525 | 1982.9 |
| Region FE | Yes | Yes |

**Notes:** Standard errors clustered at the village level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 3.7 runs two simple OLS regressions with region-fixed effects. Column (1) looks at the correlation of the dummy variable, which indicates whether a household acquired the land before 1985, while column (2) looks at the continuous variable, which is the year of land acquisition. As both variables are fixed over time, I only look at the third wave, 2015/2016, and at the correlation with the drought occurring in 2015, which is the most important. Table 3.7 shows that both variables do not seem to be correlated with the climate shock in 2015 nor in 1980. The timing of the land acquisition is, as expected, mainly correlated with the age of the household head and his educational level. The earlier the land has been acquired, the older and less educated the household head is.

There is no way to rule out entirely the possibility that these heterogeneity effects might be driven by endogenous self-selection occurring after the 1980s droughts. One main omitted variable in this paper is wealth, which might be highly correlated to the year of land acquisition. The previous analysis can not rule out that the observed heterogeneity is driven by differences in the wealth of the households rather than experience. The way the land was acquired might play a key role as well as if the land was acquired through inheritance, the person in charge of the plot would have had experience on the plot. Eventually, as the heterogeneity is mainly driven by agricultural production, it would be insightful to verify if adaptation occurred and through which method. Thus, work-in-progress is ongoing to include a variety of asset wealth indicators, the effect of inheritance, and identify adaptation strategies such as planting trees, changing planting dates, and crop diversification of innovation adoption.

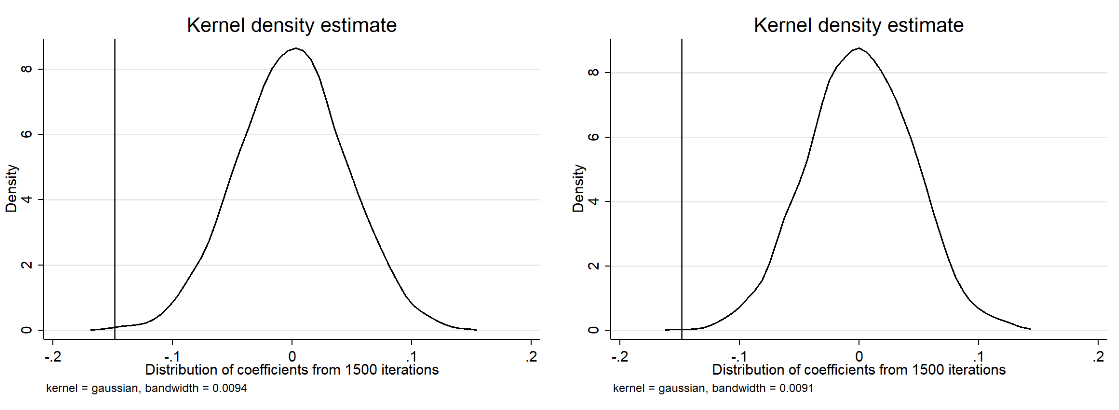
## 3.7 Robustness and Sensitivity analysis

### 3.7.1 Placebo tests

This section runs inference tests to check whether the effect of short-term droughts on agricultural production is unlikely to be observed by chance. I test the main result from Table 3.1 column (3), which indicates that being hit by a drought decreases yields by 14.8%. I draw 1500 permutations and compute the precise p-value based on the distribution of the 1500 counterfactual treatment effects, under the sharp null hypothesis of no effect [18].

[18]The test is done using the *ritest* STATA command.

Figure 3.7: Effects of short-term droughts on yields - Temporal randomization inference tests




(a) Random sublocation (b) Random timing

**Notes:** The two figures represent the distribution of the treatment effects of being hit by drought when conducting 1,500 permutations. Figure (a) randomly changes the village's allocation to droughts while Figure (b) randomly changes the timing of a drought for each village. The vertical line indicates the location of the estimate under the implemented treatment assignment (Table 3.1 Column (3)), and displays the new estimated p-value.
*Sources:* Authors' elaboration on CHIRPS and GHS data.

Figure 3.7a runs spatial counterfactuals as villages are assigned rainfall shocks from randomly selected villages. This maintains the distribution of the independent variable and removes spatial patterns. This inference test accounts also for spurious correlation linked to spatially dependent trends (Lind, 2019). Figure 3.7b randomly changes the timing of the rainfall shocks for each village.

Both simulations show that the model is not misspecified at the 1% level. The distribution of treatment effects drawn from permutations is shifted around zero and has the shape of a standard normal distribution. The vertical lines indicate the location of the main result from Table 3.1 $-0.148^{***}$.

### 3.7.2 Correlations

#### 3.7.2.1 Spatial correlation

This section accounts for the spatial correlation within 100km using Conley, 1999 standard errors, focusing on the results obtained for agricultural production. Table 3.8 shows that the results from Table 3.1 are robust (Column (1)), as well as those from Table 3.3 (Columns (2) and (3)) and those from Table 3.5 (Columns (4) to (7)).

Table 3.8: Effects of short-term droughts on yields - Spatial correction (100km)

| | Main crops | | | | | | |
|---|---|---|---|---|---|---|---|
| | All Observations | | | Not hit in 1980s | | Hit in 1980s | |
| Yield (log) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Drought (MJJAS) | -0.111* | -0.198** | -0.156** | -0.156** | -0.227** | 0.0241 | -0.130 |
| | [0.0652] | [0.0996] | [0.0686] | [0.0781] | [0.106] | [0.129] | [0.204] |
| Drought × Acquired before 1985 | | 0.164 | | | 0.126 | | 0.351* |
| | | [0.122] | | | [0.137] | | [0.204] |
| Drought × Dummy exposure | | | 0.380*** | | | | |
| | | | [0.105] | | | | |
| Observations | 5192 | 5192 | 5192 | 3194 | 3194 | 1998 | 1998 |
| R2 | 0.00273 | 0.00384 | 0.00491 | 0.00582 | 0.00658 | 0.000433 | 0.00374 |
| log(yields) Mean | 7.344 | 7.344 | 7.344 | 7.501 | 7.501 | 7.094 | 7.094 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Household FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Balanced Panel | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** Conley, 1999 standard errors correcting for spatial correlation at 100km , $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Each estimation controls for the number of adults in the household, the gender, age and educational level of the household head.

### 3.7.3 Changing thresholds

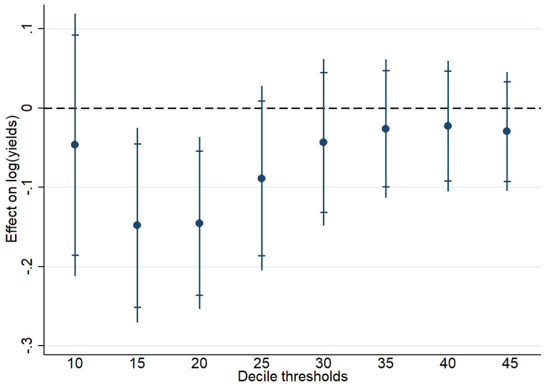
#### 3.7.3.1 Rainfall threshold

In this section, I discuss the choice of the threshold to define a dry year. In the main analysis, a year is defined as dry if the cumulative rains over MJJAS are under the 15th percentile of the rainfall distribution of each village over the 1989-2019 long-term mean. Figure 3.8b gives the number and percentages of treated and control villages changing for different thresholds defining the treatment, which is mainly being hit by the 2015 drought. For instance, I could not choose the 10th threshold as very few villages were treated (4%). I choose the 15th decile as the best trade-off between being hit by a severe drought and having enough treated observations.

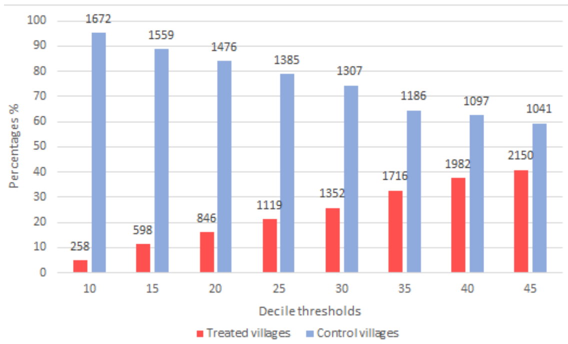
Figure 3.8a plots different estimations of the effect of short-term droughts on yields from Table 3.1 column (3), changing the thresholds of treatment. The coefficient under the 15th is the main estimation from Table 3.1. As expected, there is no effect of the 10th decile, mainly due to a limitation in the number of treated observations.

The negative effect remains significant up to the 20th threshold and is then attenuated when the severity of the dry shock decreases.

Figure 3.8: Effect of droughts on yields - changing drought threshold




(a)




(b)

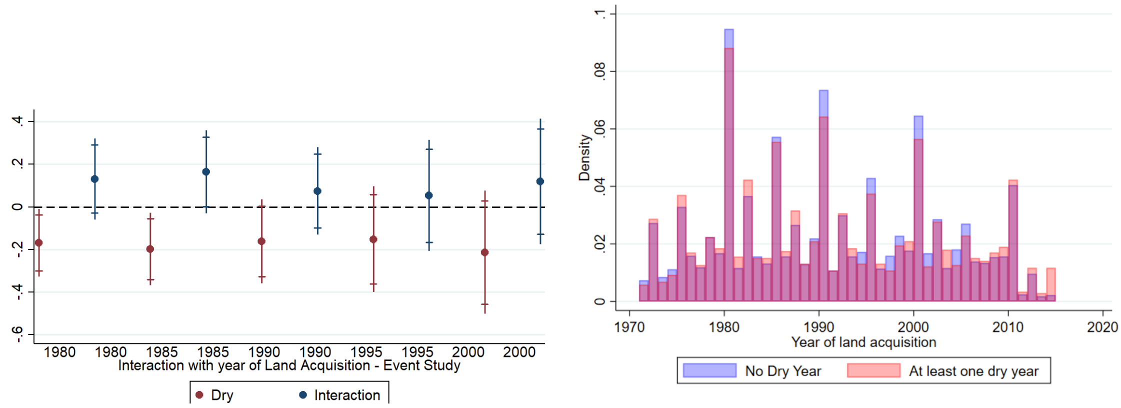
**Notes:**Figure (a) plots the effect of being hit by a drought on yields changing the threshold for being treated. Figure (b) gives, for each threshold, the percentage of treated and control villages, as well as the number of villages in each group.
*Sources:* Author's elaboration on GHS and CHIRPS data.

#### 3.7.3.2 Timing of land acquisition

The threshold used to define the dummy of year acquisition of the land, 1985, is chosen in order to capture both experiences of the land under dry conditions and to compare households that have acquired their land after the intense Sahelian droughts or before. In this Section, I discuss the choice of this threshold to convince the reader that it is the past experience of the early 1980s that plays a key role.

Figure 3.9a plots the results of the estimation 3.2 on the log of yields for different thresholds for the dummy of year acquisition. In red is plot $\alpha_1$, the estimator for households that acquired their first land after the indicated year, and in blue is plot the $\alpha_2$, meaning the interaction term. For instance, the two first dots on the left give the two estimators from the same regression, where the dummy drought in recent years interacted with a dummy, which equals 0 if the household has acquired the land after 1980, 1 if after. The five pairs of dots correspond to five distinct regressions.

Figure 3.9: Rainfall variability effects according to the year of land acquisition




(a) (b)

**Notes:** Figure (a) plots the results of the estimation 3.2 on the log of yields for a different threshold for the dummy of year acquisition. 5% and 10% significance are given. Figure (c) plots the distribution of the distance.
*Sources:* Author's elaboration on GHS and CHIRPS data.

## 3.8 Conclusion

This paper looks at whether the experience of past dry events can reduce the vulnerability of households to current rainfall variability. First, I analyze the statistical trends of rainfall over the 1981-2019 period in Nigeria, using the satellite and stations based rainfall product CHIRPS. If I show evidence of the long and severe dry period of the 1980s both in the Sahelian and Gulf Guinean regions, I find different patterns. This paper displays statistics that refute the recovery of rains in the Sahel and observe in the south that if the cumulative rains show no significant patterns, rains are becoming significantly less frequent and more intense, which increases the probability of extreme events in the more recent decades. In line with this result, I give evidence of a period of the high occurrence of droughts over the 2013-2015 period.

This paper matches a three-wave panel survey from 2010-2016 from the GHS, with a strong focus on agricultural outcomes, to the CHIRPS product. I use a two-way fixed effect strategy and exploit the variation in the occurrence of droughts, mainly over the last wave. First, I look at the short-term effects of droughts on agricultural production and food security indicators. I show that being hit by a drought decreases yields by 14%, and decreases the food diversity of households by around 1%. Second, I try to assess the role of experience in the capacity to find adaptive strategies and cope with rainfall variation. I look at the heterogeneity of the impacts according to the experience of the plot, using the timing of the year of acquisition of the first plot of the household. I compare the impacts of rainfall shocks on households that acquired their first plot before the 1980s dry period to those that acquired it after. Results show that one additional dry year decreases the yields on average by 19% for households that acquired their first land after 1985, in comparison to those who acquired it previously. This result suggests that having a long-lasting experience under extreme dry events on cultivated land reduces vulnerability to rainfall variability. This is only a piece of descriptive evidence, which can not lead to causal interpretation.

Important further work will be to understand in more detail the rules of land acquisition in Nigeria and check the role of acquisition through inheritance. Refining the definition of the rainy season according to the geographical context and the main crop is also the next step in the paper. A main limitation is the reduced form analysis. A key question is to assess the differences in terms of agricultural practices between households that experienced the 1980s with their land and the others. Work is still needed in order to understand what are the good agricultural practices facing rainfall variability. Additional data, based on satellite images, can also be used to look at correlations with our measure of yields.

# Bibliography


Adams, Richard (1989). “Global Climate Change and Agriculture: An Economic Perspective”. In: *American Journal of Agricultural Economics* 71.5, pp. 1272–1279.

Adams, Richard et al. (June 1995). “A reassessment of the economic effects of global climate change on U.S. agriculture”. In: *Climatic Change* 30.2, pp. 147–167.

Antle, John and Claudio Stöckle (2017). “Climate Impacts on Agriculture: Insights from Agronomic-Economic Analysis”. In: *Review of Environmental Economics and Policy* 11.2, pp. 299–318.

Asseng, Senthold et al. (Feb. 2015). “Rising temperatures reduce global wheat production”. In: *Nature Climate Change* 5, pp. 143–147.

Auffhammer, Maximilian et al. (May 2013). “Using Weather Data and Climate Model Output in Economic Analyses of Climate Change”. In: *Review of Environmental Economics and Policy* 7.2, pp. 181–198.

Ayenew, Habtamu Yesigat et al. (2018). “Production diversification, dietary diversity and consumption seasonality: panel data evidence from Nigeria”. In: *BMC Public Health*.

Banerjee, Abhijit et al. (2015). “The Miracle of Microfinance? Evidence from a Randomized Evaluation”. In: *American Economic Journal: Applied Economics* 7.1, pp. 22–53.

Barrios, Salvador, Bazoumana Ouattara, and Eric Strobl (2008). “The impact of climate change on agricultural production : Is it different for Africa ?” In: *Food Policy* 33.4, pp. 287–298.

Benson, Charlotte and Edward Clay (1998). *The impact of drought on sub-saharan African economies*. The World Bank.

Bertelli, Olivia (Apr. 2019). “Food security measures in Sub-Saharan Africa. A validation of the LSMS-ISA scale”. In: *Journal of African Economies*.

Biasutti, M et al. (2008). “SST Forcings and Sahel Rainfall Variability in Simulations of the Twentieth and Twenty-First Centuries”. In: *Journal of Climate* 21.14, pp. 3471–3486.

Biasutti, Michela (2019). “Rainfall trends in the African Sahel: Characteristics, processes, and causes”. In: *WIREs Climate Change* 10.4, e591.

Bichet, Adeline and Arona Diedhiou (2018). “Less frequent and more intense rainfall along the coast of the Gulf of Guinea in West and Central Africa (1981-2014)”. In: *Climate Research* 76, pp. 191–201. URL: https://www.documentation.ird.fr/hor/%7BPAR%7D00018747.

Blanc, Elodie and Wolfram Schlenker (2017). “The Use of Panel Models in Assessments of Climate Impacts on Agriculture”. In: *Review of Environmental Economics and Policy* 11.2, pp. 258–279.

Brandt, Martin et al. (2015). “Ground- and satellite-based evidence of the biophysical mechanisms behind the greening Sahel”. In: *Global Change Biology* 21.4, pp. 1610–1620.

Burke, Marshall and Kyle Emerick (Aug. 2016). “Adaptation to Climate Change: Evidence from US Agriculture”. In: *American Economic Journal: Economic Policy* 8.3, pp. 106–40.

Carletto, Calogero, Sydney Gourlay, and Paul Winters (May 2015). “From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis”. In: *Journal of African Economies* 24.5, pp. 593–628.

Cline, William (Dec. 1996). “The Impact of Global Warming of Agriculture: Comment”. In: *The American Economic Review* 68.5, pp. 3–21.

Cole, Shawn et al. (Jan. 2013). “Barriers to Household Risk Management: Evidence from India”. In: *American Economic Journal: Applied Economics* 5.1, pp. 104–35.

Conley, Timothy (Sept. 1999). “GMM estimation with cross sectional dependence”. In: *Journal of Econometrics* 92.1, pp. 1–45.

Dabi, DD et al. (2008). “Past, present and future adaptation by rural households of northern Nigeria”. In: *In: Leary N, Adejuwon J, Barros V, Burton I, Kulkarni J, Lasco R (eds) Climate change and adaptation, Chapter 8*, pp. 147–162.

Dai, Aiguo et al. (2004). “The recent Sahel drought is real”. In: *International Journal of Climatology* 24.11, pp. 1323–1331.

Dell, Melissa, Benjamin Jones, and Benjamin Olken (July 2012). “Temperature Shocks and Economic Growth: Evidence from the Last Half Century”. In: *American Economic Journal: Macroeconomics* 4.3, pp. 66–95.

– (2014). “What Do We Learn from the Weather? The New Climate–Economy Literature”. In: *Journal of Economic Literature* 52.3, pp. 740–798.

Deressa, T, R Hassan, and D Poonyth (2005). “Measuring the impact of climate change on South African agriculture: the case of sugarcane growing region”. In: *Agrekon* 44.4, pp. 524–542.

Deschênes, Olivier and Michael Greenstone (Mar. 2007). “The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather”. In: *American Economic Review* 97.1, pp. 354–385.

Dinku, Tufa et al. (2018). “Validation of the CHIRPS satellite rainfall estimates over eastern Africa”. In: *Quarterly Journal of the Royal Meteorological Society* 144.S1, pp. 292–312.

Fadare, Olusegun Ayodeji, Dare Akerele, and Begho Toritseju (2014). “Factors influencing adoption decisions of maize farmers in Nigeria”. In: *International Journal of Food and Agricultural Economics (IJFAEC).* Vol 2 1128-2016-92049, p. 10.

Fensholt, Rasmus et al. (June 2012). “Greenness in semi-arid areas across the globe 1981–2007 — an Earth Observing Satellite based analysis of trends and drivers”. In: *Remote Sensing of Environment* 121, pp. 144–158.

Fleischer, Aliza, Ivgenia Lichtman, and Robert Mendelsohn (Apr. 2008). “Climate change, irrigation, and Israeli agriculture: Will warming be harmful?” In: *Ecological Economics* 65.3, pp. 508–515.

Fowowe, Babajide (2020). “The effects of financial inclusion on agricultural productivity in Nigeria”. In: *Journal of Economics and Development.*

Funk, Chris et al. (Aug. 2019). “A High-Resolution 1983–2016 Tmax Climate Data Record Based on Infrared Temperatures and Stations by the Climate Hazard Center”. In: *Journal of Climate* 32.17, pp. 5639–5658.

Ghebru, Hosaena et al. (Feb. 2019). *Role of Land Access in Youth Migration and Youth Employment Decisions: Empirical Evidence from Rural Nigeria.* Feed the Future Innovation Lab for Food Security Policy Research Papers 303019. Michigan State University, Department of Agricultural, Food, and Resource Economics, Feed the Future Innovation Lab for Food Security (FSP).

Giannini, Alessandra et al. (Apr. 2013). “A unifying view of climate change in the Sahel linking intra-seasonal, interannual and longer time scales”. In: *Environmental Research Letters* 8, p. 024010.

Iizumi, Toshichika and Toru Sakai (2020). “The global dataset of historical yields for major crops 1981–2016”. In: *Scientific Data* 97.7.

Komowski, Frédéric et al. (2015). *Les sociétés rurales face aux changements climatiques en Afrique de l’Ouest. Innover en milieu rural Ouest-Africain : Quels changements dans les pratiques agricoles des exploitants ?*

Kurukulasuriya, Pradeep and Mohamed Ihsan Ajwad (Mar. 2006). “Application of the Ricardian Technique to Estimate the Impact of Climate Change on Smallholder Farming in Sri Lanka”. In: *Climatic Change* 81.1, pp. 39–59.

Kurukulasuriya, Pradeep and Robert Mendelsohn (Mar. 2008). “Crop switching as a strategy for adapting to climate change”. In: *African Journal of Agricultural and Resource Economics* 2.1, pp. 1–22.

Kurukulasuriya, Pradeep, Robert Mendelsohn, et al. (2006). “Will African Agriculture Survive Climate Change?” In: *The World Bank Economic Review* 20.3, pp. 367–388.

Lind, Jo Thori (2019). “Spurious weather effects”. In: *Journal of Regional Science* 59.2, pp. 322–354.

Lobell, David et al. (2008). “Prioritizing Climate Change Adaptation Needs for Food Security in 2030”. In: *Science* 319.5863, pp. 607–610.

Maddison, David (Jan. 2007). “The Perception of and Adaptation to Climate Change in Africa”. In: *The World Bank, Policy Research Working Paper Series.*

Mendelsohn, Robert and Ariel Dinar (2003). “Climate, Water, and Agriculture”. In: *Land Economics* 79.3, pp. 328–341.

Molua, Ernest and Cornelius Lambi (2007). *The Economic Impact Of Climate Change On Agriculture In Cameroon, Volume 1of 1.* The World Bank.

Mortimore, Michael and William Adams (2001). “Farmer adaptation, change and crisis’ in the Sahel”. In: *Global Environmental Change* 11, pp. 49–57.

NBS (2012). “Basic information document”. In: *Nigeria general household survey-panel* Abuja: National Bureau of Statistics.

Nicholson, Sharon (2005). “On the question of the “recovery” of the rains in the West African Sahel”. In: *Journal of Arid Environments* 63.3. Special Issue on The ”Greening” of the Sahel, pp. 615–641.

– (2018). “The ITCZ and the Seasonal Cycle over Equatorial Africa”. In: *Bulletin of the American Meteorological Society* 99.2, pp. 337–348.

Nordhaus, William (2006). “Geography and macroeconomics: New data and new findings”. In: *Proceedings of the National Academy of Sciences* 103.10, pp. 3510–3517.

Onyeneke, RU and DK Madukwe (2010). “Adaptation measures by crop farmers in the southeast rainforest zone of Nigeria to climate change”. In: *Science World Journal.*

Ouedraogo, Mathieu (June 2012). “Impact des changements climatiques sur les revenus agricoles au Burkina Faso”. In: *Journal of Agriculture and Environment for International Development* 106, pp. 3–21.

Panthou, G, T Vischel, and T Lebel (2014). “Recent trends in the regime of extreme rainfall in the Central Sahel”. In: *International Journal of Climatology* 34.15, pp. 3998–4006.

Panthou, G et al. (2018). “Rainfall intensification in tropical semi-arid regions : the Sahelian case”. ENG. In: *Environmental Research Letters* 13, p. art. 064013 [10 p.]

Salack, Seyni et al. (2014). “Oceanic influence on the sub-seasonal to interannual timing and frequency of extreme dry spells over the West African Sahel”. In: *Climate Dynamics.*

Sanghi, Apurva and Robert Mendelsohn (Oct. 2008). “The impacts of global warming on farmers in Brazil and India”. In: *Global Environmental Change* 18.4, pp. 655–665.

Sanogo, Souleymane et al. (2015). “Spatio-temporal characteristics of the recent rainfall recovery in West Africa”. In: *International Journal of Climatology* 35.15, pp. 4589–4605.

Schlenker, Wolfram, Michael Hanemann, and Anthony Fisher (Mar. 2005). “Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach”. In: *American Economic Review* 95.1, pp. 395–406.

Schlenker, Wolfram and David Lobell (2010). “Robust negative impacts of climate change on African agriculture”. In: *Environmental Research Letters* 5.1, p. 014010.

Seo, Niggol and Robert Mendelsohn (2007). *Climate Change Impacts On Animal Husbandry In Africa : A Ricardian Analysis.* The World Bank.

Seo, Niggol, Robert Mendelsohn, et al. (Mar. 2009). “A Ricardian Analysis of the Distribution of Climate Change Impacts on Agriculture across Agro-Ecological Zones in Africa”. In: *Environmental and Resource Economics* 43.3, pp. 313–332.

Silvestri, Silvia et al. (Dec. 2012). “Climate change perception and adaptation of agro-pastoral communities in Kenya”. In: *Regional Environmental Change* 12.

Sofoluwe, Nurudeen, Akeem Tijani, and O Baruwa (Sept. 2011). “Farmers’ perception and adaptation to climate change in Osun State, Nigeria”. In: *African Journal of Agricultural Research* 6.

Sultan, Benjamin et al. (Mar. 2013). “Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa”. In: *Environmental research letters* 8.

Swindale, Anne and Paula Bilinsky (Feb. 2006). *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide.* Tech. rep. Food and Nutrition Technical Assistance FANTA III.

Tambo, Justice and Tahirou Abdoulaye (2013). “Smallholder farmers’ perceptions of and adaptations to climate change in the Nigerian savanna”. In: *Reg Environ Chang* 13.375.

Veljanoska, Stefanija (2018). "Can Land Fragmentation Reduce the Exposure of Rural Households to Weather Variability?" In: *Ecological Economics* 154, pp. 42–51.

Wichern, Jannike et al. (2019). "Vulnerability and adaptation options to climate change for rural livelihoods – A country-wide analysis for Uganda". In: *Agricultural Systems* 176, p. 102663.

Wing, Ian Sue, Enrica De Cian, and Malcolm Mistry (2021). "Global vulnerability of crop yields to climate change". In: *Journal of Environmental Economics and Management* 109, p. 102462.

Wollburg, Philip, Marco Tiberti, and Alberto Zezza (2021). "Recall length and measurement error in agricultural surveys". In: *Food Policy*.

Wood, Stephen and Robert Mendelsohn (Mar. 2014). "The impact of climate change on agricultural net revenue: a case study in the Fouta Djallon, West Africa". In: *Environment and Development Economics* 20.01, pp. 20–36.

Yacoubou Djima, Ismael and Talip Kilic (2021). *Survey Measurement Errors and the Assessment of the Relationship between Yields and Inputs in Smallholder Farming Systems: Evidence from Mali*. The World Bank.

# Appendix A

# Appendix to Chapter 1 : Droughts, Migration and Population in Kenya

## A.1 Data

### A.1.1 Population censuses

Table A.1 gives the number of individuals for each census in each province of Kenya. It compares the number of observations calculated from the data to the one stated in the official report censuses. If there is little difference between the dataset and the report for all of Kenya, the numbers differ across provinces. The missing observations are reduced across the years (and much smaller in 2009), however, we can observe an important difference between the data and reports for Nyanza province in 1989, with more than 90% of the observations missing. Another high discrepancy is for the North Eastern province in 1999, with 16% of the total population missing according to the report. Tables A.2 and A.3 show in more detail that it is necessary to exclude the Nyanza and North Eastern provinces from the analysis.

Table A.2 (resp. A.3) shows descriptive statistics of the population distribution within Nyanza (resp. North Eastern), at the scale of districts [1]. Column(1) gives the total number of individuals in each district, while Column (2) is the mean (standard deviation) of the sublocation sizes within the district. Column (3) displays the number of individuals living in the smallest (resp. biggest) sublocation of the district. Nyanza province is made up of four districts, Kisii, Kisumu, Siaya, and South Nyanza. While Kisii is totally absent from the dataset in 1989, the three other districts have abnormally low numbers in 1989 (Column (3)), compared to the two other censuses. Thus, these irregular numbers exhibit that the missing observations in Nyanza in 1989 are not only driven by the absence of Kisii in the dataset, nor the total absence of some sublocations, but missing observations distributed within sublocations. Thus, there is no district/sublocations that can be kept in the analysis.

The same issue is observed in Table A.3 for North Eastern in 1999, with very low minimum values for the sublocations population in comparison to 1989, especially in Mandera. If the overall population has increased between 1989 and 1999, the sublocation populations have decreased, such as the size of the smallest sublocation (and biggest, apart from Wajir). This is in accordance with the information from Table A.6 which shows that 16% of the total observations are missing in the censuses. Such as Nyanza province, I can not conclude that these missing data are driven by some districts or sublocations and exclude North Eastern province from the analysis.

[1] districts matched over the censuses to the districts in 1989

Besides, Table A.3 displays high standard values for the population of sublocations, which illustrates skewed population distributions with spread-out differences.

The other provinces have comparable missing observations from what is stated in the reports.

Table A.1: Returns from the census report and micro data, 1989, 1999 and 2009

| | | 1989 | | | 1999 | | | 2009 | | 1999 | -1989 | p.a (%) | 1999 | -1989 | p.a (%) | 2009 | -1989 | p.a (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| province | Data | Report | Missing | Data | Report | Missing | Data | Report | Missing | Data | Report | Diff. (p.p) | Data | Report | Diff. (p.p) | Data | Report | Diff. (p.p) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| Nairobi | 1,242,424 | 1,324,570 | 6.2 | 2,004,116 | 2,143,254 | 6.5 | 3,109,424 | 3,138,369 | 0.9 | 4.9 | 4.9 | 0 | 4.5 | 3.9 | 0.6 | 4.7 | 4.4 | 0.3 |
| Central | 3,040,580 | 3,116,703 | 2.4 | 3,593,417 | 3,724,159 | 3.5 | 4,368,888 | 4,383,742 | 0.3 | 1.7 | 1.8 | -0.1 | 2 | 1.6 | 0.3 | 1.8 | 1.7 | 0.1 |
| Coast | 1,770,088 | 1,829,191 | 3.2 | 2,369,247 | 2,487,264 | 4.7 | 3,290,292 | 3,325,307 | 1.1 | 3 | 3.1 | -0.2 | 3.3 | 2.9 | 0.4 | 3.1 | 3 | 0.1 |
| Eastern | 3,701,017 | 3,768,677 | 1.8 | 4,525,518 | 4,631,779 | 2.3 | 5,636,311 | 5,668,123 | 0.6 | 2 | 2.1 | -0.1 | 2.2 | 2 | 0.2 | 2.1 | 2.1 | 0.1 |
| North Eastern | 364,923 | 371,391 | 1.7 | 807,198 | 962,143 | 16.1 | 2,301,746 | 2,310,757 | 0.4 | 8.3 | 10 | -1.7 | 11 | 9.2 | 1.9 | 9.6 | 9.6 | 0.1 |
| Nyanza | 117,160 | 3,507,162 | 96.7 | 4,263,934 | 4,392,196 | 2.9 | 5,416,670 | 5,442,711 | 0.5 | 43.3 | 2.3 | 41 | 2.4 | 2.2 | 0.3 | 21.1 | 2.2 | 18.9 |
| Rift Valley | 4,789,367 | 4,981,613 | 3.9 | 6,723,765 | 6,987,036 | 3.8 | 9,949,727 | 10,006,805 | 0.6 | 3.5 | 3.4 | 0 | 4 | 3.7 | 0.3 | 3.7 | 3.5 | 0.2 |
| Western | 2,555,504 | 2,544,329 | -0.4 | 3,299,835 | 3,358,776 | 1.8 | 4,317,466 | 4,334,282 | 0.4 | 2.6 | 2.8 | -0.2 | 2.7 | 2.6 | 0.1 | 2.7 | 2.7 | 0 |
| Kenya | 17,581,063 | 21,443,636 | 0.2 | 27,587,030 | 28,686,607 | 0.0 | 38,390,524 | 38,610,096 | 0 | 4.6 | 3 | 1.7 | 3.4 | 3 | 0.3 | 4 | 3 | 1 |

Table A.2: Returns from the micro data in Nyanza province, 1989, 1999 and 2009

| | Nyanza 1989 | | | Nyanza 1999 | | | Nyanza 2009 | | |
|---|---|---|---|---|---|---|---|---|---|
| | N(indiv.) | Mean(SD) | Min(Max) | N(indiv.) | Mean(SD) | Min(Max) | N(indiv.) | Mean(SD) | Min(Max) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Nyanza | 117160 | 248.75 (166.66) | 10 (1854) | 4263934 | 4575.04 (3198.62) | 321 (40334) | 5416670 | 5624.79 (4036.91) | 353 (47412) |
| Kisii | 0 (0) | 0 (0) | 0 (0) | 1416096 | 6525.79 (3316.63) | 1694 (28372) | 1744961 | 7240.5 (4283.91) | 767 (43388) |
| Kisumu | 32589 | 285.87 (254.56) | 26 (1854) | 770861 | 4729.21 (4597.06) | 667 (40334) | 959027 | 5708.49 (5780.06) | 353 (47412) |
| Siaya | 31768 | 209 (78.06) | 43 (535) | 699686 | 3953.03 (1885.86) | 972 (13571) | 838779 | 4685.92 (2503.46) | 1187 (18069) |
| South Nyanza | 52803 | 257.58 (146.85) | 10 (827) | 1377291 | 3672.78 (2251.74) | 321 (17148) | 1873903 | 4997.07 (3120.65) | 373 (23155) |

Table A.3: Returns from the micro data in North Eastern province, 1989, 1999 and 2009

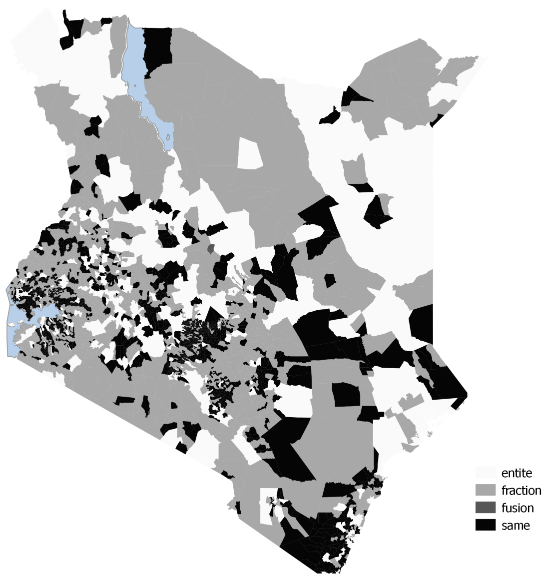
| | North Eastern 1989 | | | North Eastern 1999 | | | North Eastern 2009 | | |
|---|---|---|---|---|---|---|---|---|---|
| | N(indiv.) | Mean(SD) | Min(Max) | N(indiv.) | Mean(SD) | Min(Max) | N(indiv.) | Mean(SD) | Min(Max) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| North Eastern | 364923 | 3119 (2509.88) | 236 (14869) | 807198 | 2603.86 (2378.21) | 112 (19718) | 2301746 | 7238.19 (7477.14) | 373 (65386) |
| Garissa | 121857 | 2972.12 (2589.19) | 422 (14869) | 253028 | 2750.3 (3011.6) | 283 (19718) | 619497 | 6733.66 (10606.52) | 786 (65386) |
| Mandera | 121050 | 3904.84 (2945.56) | 936 (12619) | 245110 | 2113.02 (1908.56) | 112 (8581) | 1023653 | 8602.13 (6805.94) | 373 (33636) |
| Wajir | 122016 | 2711.47 (1993.12) | 236 (8548) | 309060 | 3030 (2122.44) | 215 (14247) | 658596 | 6155.1 (3944.39) | 713 (23349) |

### A.1.2 Construction of the sublocations panel

One main challenge of this study was to build a panel of sublocations over the 20 years, as administrative frontiers of sublocations have changed over decades. Sublocations frontiers in 1989 differ from those in 1999, which differ from those in 2009. If some transformations are geometrically coherent (fusion/division of previous frontiers), some transformations had no geometrical logic. To have a panel of sublocations, I have created new units coherent over time, both by hand and via coding (for recognizable geometric transformations).

Figure A.1 plots the transformation and displays each unit design used in the paper. *same* means that the sublocation has never changed during the three censuses, *fraction* that it has been divided at least once, and *fusion* merged with other sublocation (at least once). *Entite* indicates new units that I have created by hand, merging sublocation together so that I can have a coherent and stable population panel per sublocations over the 20 years.

Figure A.1: Matching of sublocation - type of transformation




**Notes:** The Figure maps the transformation that has been done to build the sublocation panel. *Sources :* author's elaboration on KNBS data.

Table A.4 provides descriptive statistics on the matching of sublocations between the three Kenyan censuses 1989, 1999, and 2009 (based on whether the frontier is built from 1989, or the 1999 or the 2009 frontiers). The comparison is made on the areas in $km^2$, calculated from the map created based on 1989/1999 sublocations, in comparison to the one used in the analysis, made based on 2009 sub localities. We observe that the difference between the 1999 and 2009 maps is smaller. The paired samples t-test does not reject the null hypothesis of the mean of the average areas, which confirms that the three matchings are similar.

Table A.4: Descriptive Statistics : Sub locations matching between censuses

| | | | Area ($km^2$) Diff [1989 -2009] | | | Area ($km^2$) Diff [1999 -2009] | | |
|---|---|---|---|---|---|---|---|---|
| | N. | perc. | Mean | S.D | P-Value | Mean | S.D | P-Value |
| Entity | 244 | 7.76 | -4.55 | 109.7 | 0.51 | 0.92 | 86.89 | 0.87 |
| Fraction | 1424 | 45.26 | 1.95 | 91.25 | 0.42 | -0.19 | 22.41 | 0.74 |
| Fusion | 2 | 0.06 | 22.53 | 32.61 | 0.51 | 14.34 | 20.18 | 0.50 |
| Same | 1476 | 46.9 | 21.5 | 43.97 | 0.06 | 0.014 | 20.63 | 0.98 |
| Total | 3146 | 100 | 1.55 | 74.9 | 0.25 | -0.00 | 31.8 | 0.99 |

## A.2 Setting

### A.2.1 Climatogy - descriptive statistics

Figure A.2 plots the long-term average of monthly precipitations across provinces. It shows that all Kenyan provinces follow a bimodal seasonal pattern. Figure A.3 plots the long-term mean of rainfall characteristics over the MAMJ season across several indicators. Table A.5 gives the definition of each indicator.

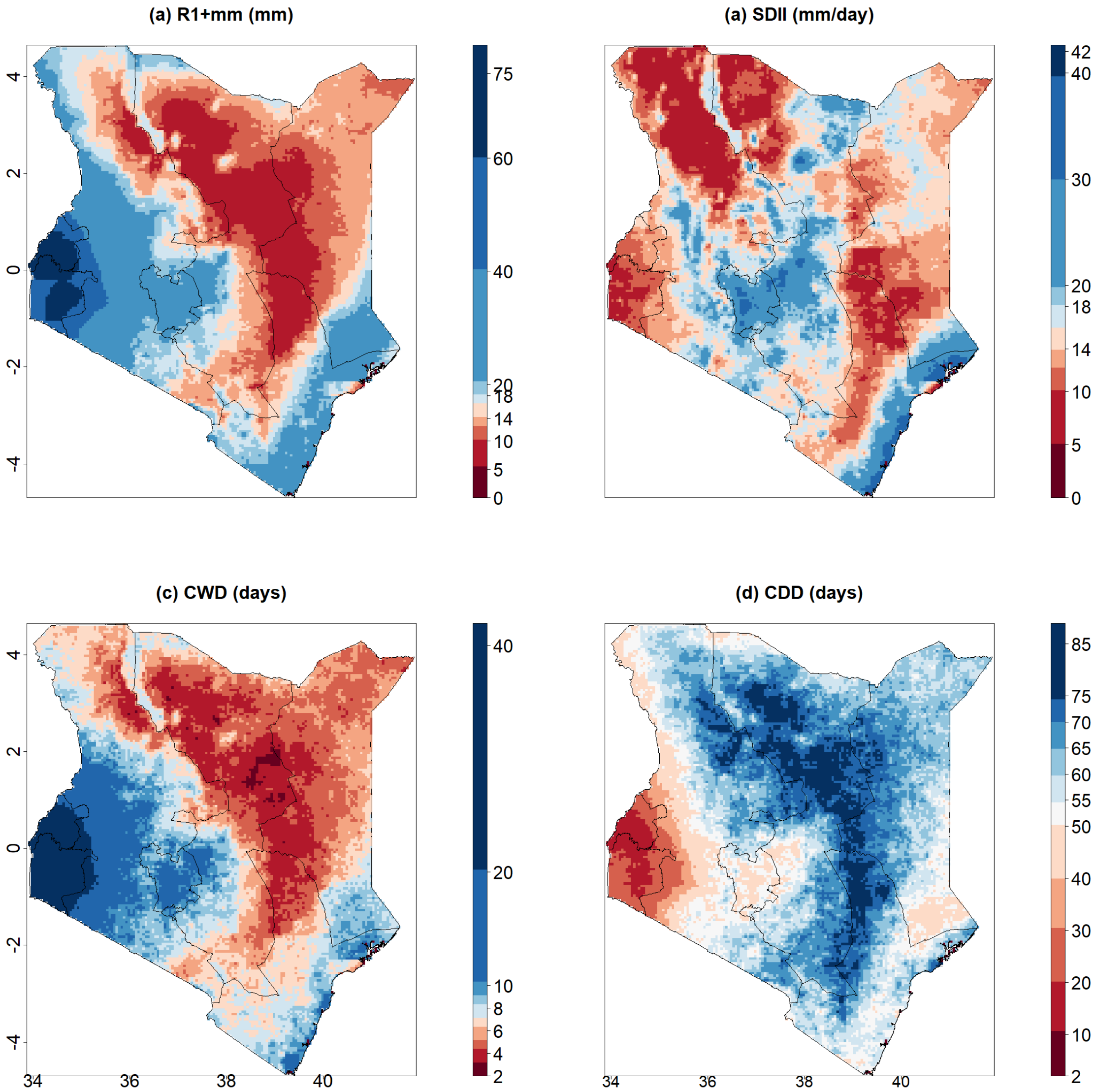
Table A.5: Extreme precipitation indices and their ETCCDI and ECA definitions (from Gebrechorkos, Hülsmann, and Bernhofer, 2019)

| Indicator | Definition | Unit |
|---|---|---|
| **R1+mm** | Number of wet days : Sum of days where daily precipitation is $\geq$ 1mm over the long rainy season | Days |
| **R1-mm** | Number of dry days : Sum of days where daily precipitation is $<$ 1mm over the long rainy season | Days |
| **R20mm** | Number of heavy rains : Sum of days where daily precipitation is $>$ 20mm over the long rainy season | Days |
| **CWD** | Consecutive wet day index : maximum number of consecutive days with precipitation above 1 mm over the long rainy season | Days |
| **CDD** | Consecutive dry day index : maximum number of consecutive days with precipitation below 1 mm over the long rainy season | Days |
| **SDII** | Simple Daily intensity index : total precipitation divided by R1+mm over the long rainy season | mm/Day |

Figure A.2: Long-term average of monthly precipitation across provinces


**Notes:** The Figures represent the long-term average of the monthly precipitation (1983-2013) (mm) over Kenya (a) and among all the Provinces (b-i). Red lines plot the 95th percentile of rainfall distribution, blue lines the 50th percentile, and green lines the 5th percentile.
*Sources :* author's elaboration on CHIRPS data.

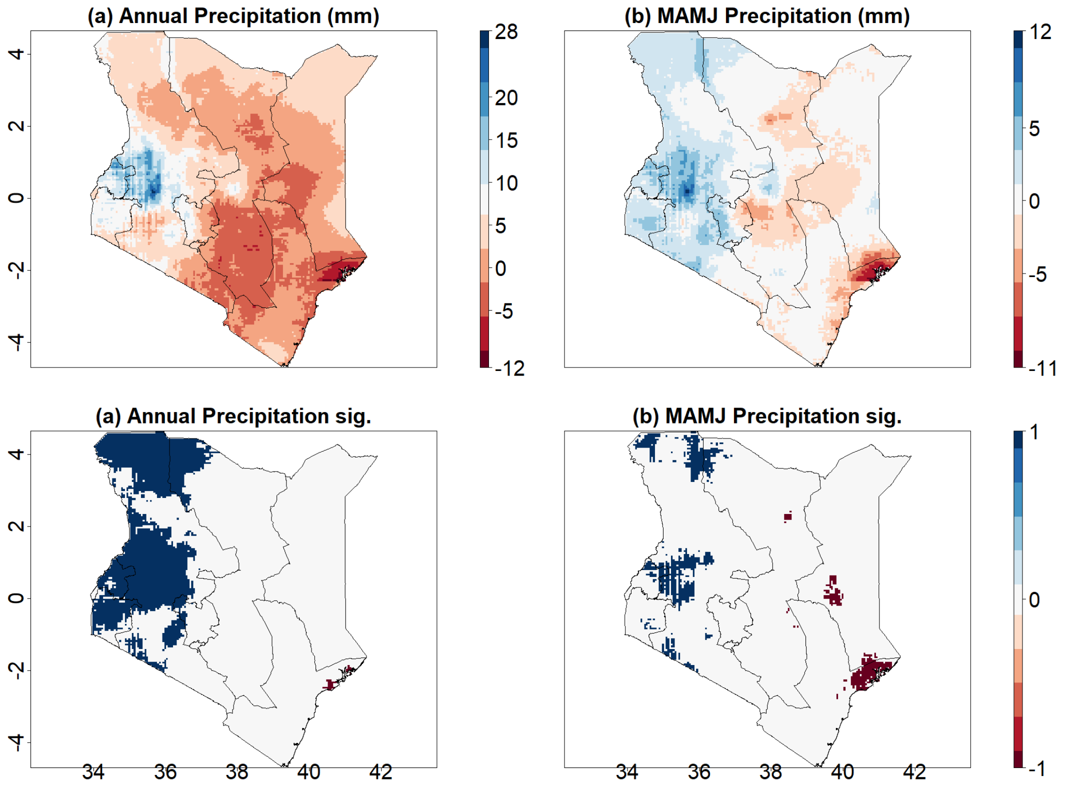
Figure A.3: Spatial distribution of the long-term average of climate indicators




**Notes:** The Figure plots the spatial distribution of long-term averages over 1983-2013 of (a) R1+ (mm) (b) SDII (mm/day) (c) CWD (days) (d) CDD
*Sources :* author's elaboration on CHIRPS data.

## A.2.2 Long-term evolution of rainfall

Figure A.4: Long-term trends of climate indicators

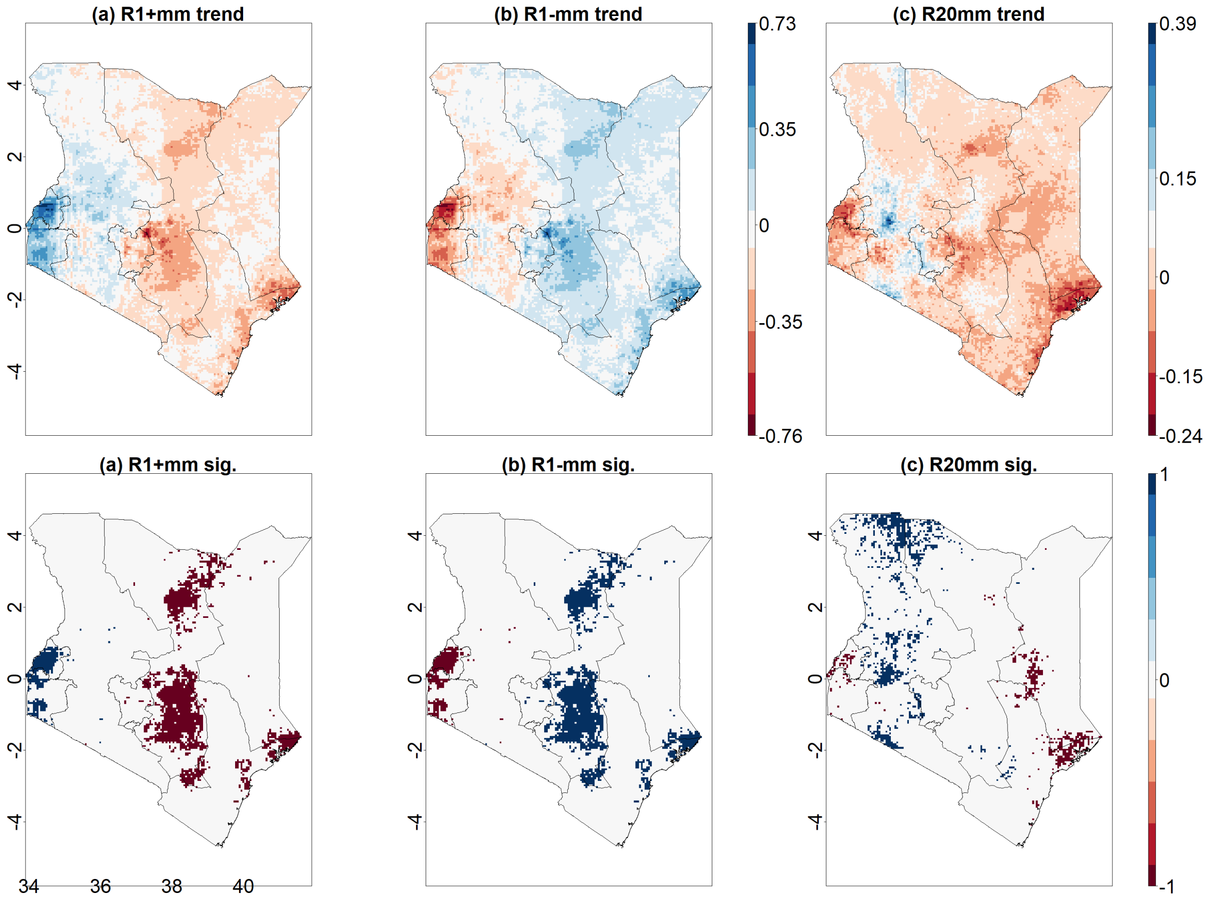



**Notes:** The Figure plots the annual (a) and long-rainy season trends (b) of precipitation amounts (mm) during the long-term period 1983-2013, based on CHIRPS data. The bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays significant increasing trends, while red (-1) is a significant decreasing one and 0 non-significant changes.
*Sources:* Author's elaboration on CHIRPS data.

The long-term trend analysis of annual precipitation and long-rain amounts shows little significant changes. Figure C.5 plots the trend values of yearly cumulative rains (a) and cumulative rains over the long-rainy season (b) over the 1983-2013 period. Annual precipitations display a significant increasing trend in the western part of the country, with up to 28mm increase at the frontier between the Western region and the Rift Valley. The increasing trend in the northwest part of the Rift Valley is lower, but still significant, ranging from 5 to 10 mm increase per year. The only significant decreasing trends in annual precipitations are found along the Coast, around the city of Lamu, and the highest decreasing trend is about -12mm per year. The observed decreasing trends in the South of the Eastern region are not significant, such as the increasing trends in the northeast of the country. Trends over the long-rainy season are in coherence with the annual trends, significantly increasing in the west and north part of the Rift Valley (up to +12 mm) and significantly decreasing around

Lamu (up to -11 mm) and in some parts of the Eastern and North Eastern regions.

Figure A.5: Long-term trends of climate indicators (2)




**Notes:** The Figure plots (a)R1+mm(b)R1+mm and (c) R20mm trends over the long-rainy season (days), during the long-term period 1983-2013, based on CHIRPS data. The bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays significant increasing trends, while red (-1) is a significant decreasing one and 0 non-significant changes.
*Sources :* author's elaboration on CHIRPS data.

Figure C.7 and Figure C.8 show a long-term modification in precipitation characteristics, each indicator being described in Table A.5. Figure C.7 (a) and (b) are symmetric *de facto* [2]. The number of wet days during MAMJ (R1+mm) (resp. dry days (R1-mm)) has significantly decreased (resp. increased) along the coast and in large part of the Eastern Region, from -0.35 days to -0.76 days (resp +0.35 to +0.73 days). In these areas (and especially the Eastern region), this comes with significant increases in the Simple Daily Intensity Index (SDII) (Figure C.8 (a), between +0.3 and +0.78 mm per day), and significant decreases (resp. increases) in the Consecutive

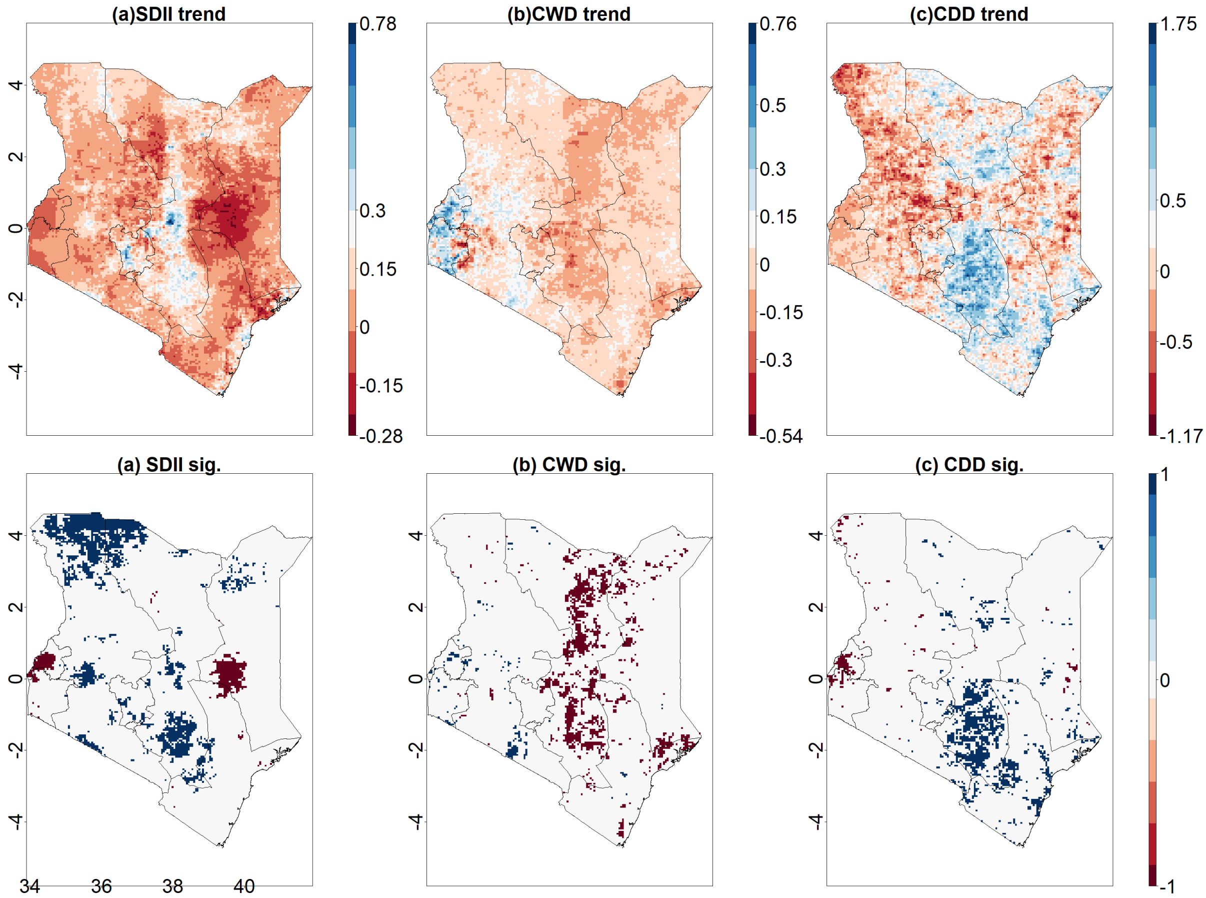
[2]By construction, R1+mm + R1-mm =122 days, i.e equals the total number of days during MAMJ

Wet Day index (CWD) (resp. Dry Day index CDD), up to -0.54 days (resp. +1.75). These results show that rains over the long-rainy season are becoming more concentrated over shorter periods and days, and that daily rainfall intensity has increased.

The number of days of very heavy precipitation days (R20mm) displays significant increasing trends over the Rift Valley (up to +0.39 days), which coincides with significant increases in the north of the daily intensity (SDII) and precipitation amounts (Figure C.5 (b)). This suggests that increasing trends of rainfall magnitudes in the northern part of the Rift Valley are due to an increase in heavy precipitation days. R20 mm displays significant decreasing trends along the Coast and the Western region of Kenya.

This long-term trends analysis shows that the ASALs region, and particularly across the Eastern region, are facing downward trends in the number of rainy days and the length of wet spells during the long-rainy season, associated with higher intensity of wet days. This suggests a decrease in the length of the agricultural period and an increase in extreme events, in a region highly vulnerable because dependent on the agricultural sector (transition from pastoralism systems to more intensive types of production and mixed systems, with increases in livestock production and reductions in lands associated to rangeland systems (Silvestri et al., 2012)).

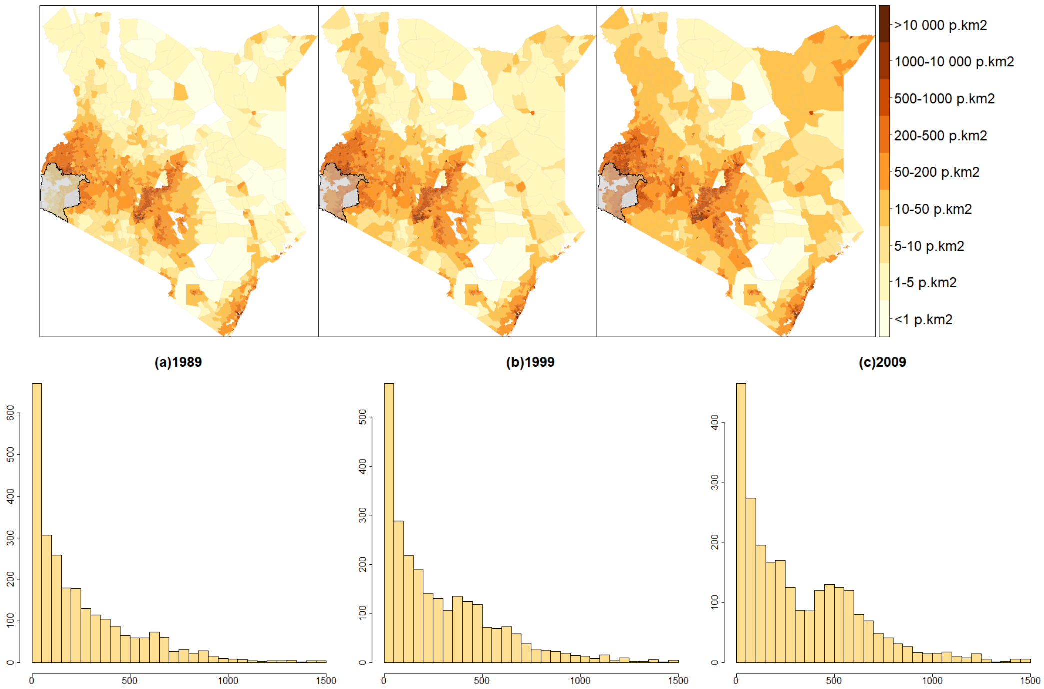
Figure A.6: Long-term trends of climate indicators (3)




**Notes:** The Figure plots (a)SDII ($mm.day^{-1}$)(b)CWD and (c) CDD(days) trends over the long-rainy season (days), during the long-term period 1983-2013, based on CHIRPS data. The bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays significant increasing trends, while red (-1) is significant decreasing one and 0 non-significant changes. *Sources :* author's elaboration on CHIRPS data.

### A.2.3 Temporal and spatial variation of population and migration

Figure A.7: Spatial distribution of the population density across census wave




**Notes:** The Figures plot the population density for each sublocation in 1989 (a), 1999 (b) and 2009 (c).
*Sources :* author's elaboration on KNBS data.

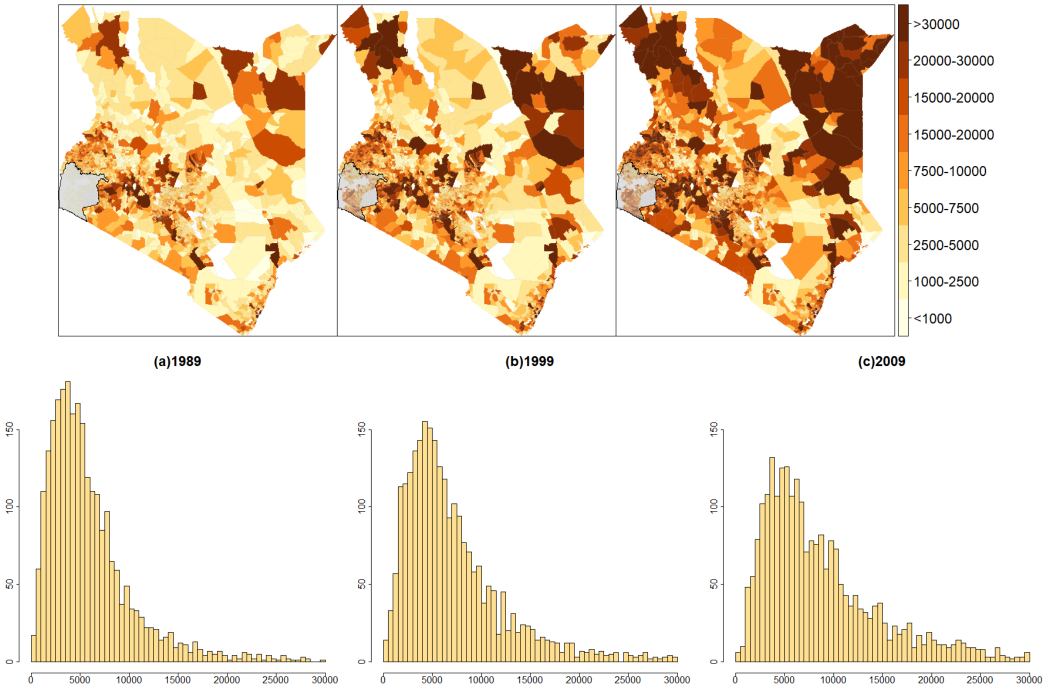
Table A.6 and Table A.7 display descriptive statistics of population, density, and population growth variables across period for each province [3]. Apart from Nyanza and North Eastern, The yearly (p.a) and decadal (DPGR) growth of the population are stable when calculated from different censuses [4]. When looking at DPGR long-differences for Nairobi, Central, Coast, Eastern Rift Valley, and Western provinces, we do not observe any clear pattern for the demographic evolution of the country. Indeed, while the DPGR long-difference is positive in some regions (+11p.p in Eastern, +7p.p in Rift Valley, and + 1p.p in Western) it is negative for others (especially in Nairobi, with -25p.p, as it was already a densely populated area in 1989). Thus, we conclude that the +2.56 p.p long difference of the control group from Table 1.6 (Column (9)) is not erratic and is in line with the descriptive statistics of Table A.7.

[3]The per annum population growth rate is defined as follows: $p.a = exp(\frac{log(\frac{pop(t_2)}{pop(t_1)})}{10}) - 1$

[4]As the Table A.7 gives information about Nyanza and North Eastern provinces, the national average includes both provinces with discrepancies

Figure A.8 plots the spatial distribution of the population size for each census wave, while Figure A.7 plots the spatial distribution of the population density for each census wave.

Figure A.8: Spatial distribution of sublocation's population size across census wave




**Notes:** The Figures represent the population sizes per sublocations over Kenya in (a) 1989 (b) 1999 and (c) 2009. Density is displayed without Nyanza and North-Eastern provinces.
*Sources :* author's elaboration on KNBS data.

Table A.6: Descriptive Statistics of Province Population from microdata 1989, 1999 and 2009

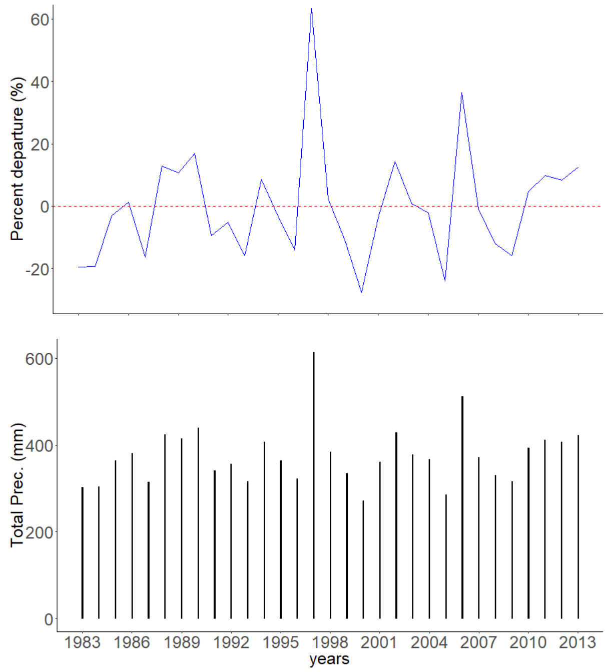
| | 1989 | | | 1999 | | | 2009 | | | [1999-1989] | [2009-1999] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Tot<br>(1) | Mean\(SD)<br>(2) | Min\(Max)<br>(3) | Tot<br>(4) | Mean\(SD)<br>(5) | Min\(Max)<br>(6) | Tot<br>(7) | Mean\(SD)<br>(8) | Min\(Max)<br>(9) | DPGR-1\(p.value)<br>(10) | DPGR-1\(p.value)<br>(11) |
| **Kenya** | | | | | | | | | | | |
| *Population* | 17168480 | 5720.92 | 10 | 25953524 | 8648.29 | 82 | 36394457 | 12127.44 | 31 | 2.59 | -0.61 |
| | | (8139.99) | (187901) | | (13435.42) | (292173) | | (22254.18) | (407124) | (0) | (0) |
| *Densite* | 31.02 | 389.13 | 0.1 | 46.9 | 514.15 | 0.09 | 65.76 | 623.76 | 0.23 | | |
| | | (1789.95) | (49036.32) | | (2109.74) | (54387.98) | | (2590.79) | (78040.3) | | |
| **Nairobi** | | | | | | | | | | | |
| *Population* | 1238130 | 29479.29 | 3589 | 2013992 | 47952.19 | 4793 | 3126064 | 74430.1 | 3423 | -0.32 | -0.57 |
| | | (36192.97) | (187901) | | (55833.19) | (228140) | | (98434.71) | (402314) | (0.05) | (0) |
| *Densite* | 1765.91 | 8242.79 | 128.46 | 2872.5 | 10473.68 | 75.31 | 4458.62 | 13037.51 | 126.71 | | |
| | | (10662.55) | (49036.32) | | (12383.89) | (54387.98) | | (15450.94) | (78040.3) | | |
| **Central** | | | | | | | | | | | |
| *Population* | 3007551 | 5478.23 | 337 | 3570204 | 6503.1 | 188 | 4335680 | 7897.41 | 31 | -0.82 | -0.86 |
| | | (3682.58) | (53101) | | (5612) | (79785) | | (10030.85) | (142038) | (0) | (0) |
| *Densite* | 239.51 | 471.91 | 2.72 | 284.32 | 535.01 | 2.66 | 345.28 | 601.88 | 0.44 | | |
| | | (319.58) | (2861.02) | | (429.11) | (4185.92) | | (583.68) | (6906.88) | | |
| **Coast** | | | | | | | | | | | |
| *Population* | 1761859 | 6936.45 | 107 | 2358618 | 9285.9 | 135 | 3281875 | 12920.77 | 122 | -0.63 | -0.65 |
| | | (10992.72) | (138465) | | (16604.07) | (217622) | | (27581.73) | (364838) | (0) | (0) |
| *Densite* | 21.26 | 715.09 | 0.16 | 28.46 | 782.96 | 0.09 | 39.6 | 926.87 | 0.23 | | |
| | | (2819.09) | (22369.66) | | (3013.73) | (24954.31) | | (3414.65) | (25507.43) | | |
| **Eastern** | | | | | | | | | | | |
| *Population* | 3543644 | 6173.6 | 61 | 4377931 | 7627.06 | 82 | 5425613 | 9452.29 | 444 | -0.78 | -0.67 |
| | | (5577.4) | (73633) | | (7527.84) | (87963) | | (10812.7) | (139172) | (0) | (0) |
| *Densite* | 26.46 | 214.4 | 0.1 | 32.69 | 249.9 | 0.14 | 40.51 | 285.35 | 0.5 | | |
| | | (296) | (3474.96) | | (348.12) | (4614.6) | | (392.62) | (5421.07) | | |
| **North Eastern** | | | | | | | | | | | |
| *Population* | 383489 | 5478.41 | 356 | 848459 | 12120.84 | 1136 | 2401879 | 34312.56 | 2054 | 0.65 | 1.52 |
| | | (6253.4) | (27609) | | (14182.34) | (64918) | | (38484.75) | (179846) | (0.03) | (0) |
| *Densite* | 2.88 | 8.77 | 0.31 | 6.36 | 16.3 | 1.04 | 18.02 | 43.76 | 0.89 | | |
| | | (20.94) | (139.07) | | (41.13) | (304.99) | | (96.13) | (589.3) | | |
| **Nyanza** | | | | | | | | | | | |
| *Population* | 115999 | 280.87 | 10 | 2829841 | 6851.92 | 667 | 3655104 | 8850.13 | 867 | 22.9 | -0.74 |
| | | (235.43) | (2564) | | (6323.28) | (79181) | | (8565.74) | (91743) | (0) | (0) |
| *Densite* | 11.27 | 16.81 | 0.52 | 274.99 | 388.67 | 27.26 | 355.18 | 487.25 | 30.3 | | |
| | | (35.31) | (576.46) | | (660.68) | (7789.81) | | (836.29) | (10700.2) | | |
| **Rift Valley** | | | | | | | | | | | |
| *Population* | 4581581 | 6012.57 | 28 | 6659156 | 8739.05 | 194 | 9858857 | 12938.13 | 301 | -0.53 | -0.46 |
| | | (7585) | (102510) | | (15444.61) | (292173) | | (22992.29) | (407124) | (0) | (0) |
| *Densite* | 26.62 | 107.07 | 0.37 | 38.69 | 139.45 | 0.77 | 57.28 | 189.98 | 2.13 | | |
| | | (126.99) | (1863.25) | | (156.29) | (1635.77) | | (226.38) | (2857.4) | | |
| **Western** | | | | | | | | | | | |
| *Population* | 2536227 | 7525.9 | 2024 | 3295323 | 9778.41 | 2301 | 4309385 | 12787.49 | 2467 | -0.73 | -0.72 |
| | | (3902.76) | (37543) | | (5885.51) | (53084) | | (8183.55) | (62313) | (0) | (0) |
| *Densite* | 331.66 | 500.52 | 35.24 | 430.92 | 590.83 | 92.24 | 563.53 | 728.8 | 104.9 | | |

Table A.7: Descriptive Statistics of Province Population Growth from micro data, 1989, 1999 and 2009

| | [1989-1999] | | | [1999-2009] | | | Diff p.p (5)-(2) | p.value |
|---|---|---|---|---|---|---|---|---|
| | Total | Mean//(SD) | Min//(Max) | Total | Mean//(SD) | Min//(Max) | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Kenya** | | | | | | | | |
| *DPGR* | 0.51 | 3.59 | -0.83 | 0.4 | 0.39 | -0.84 | -3.21 | 0 |
| | | (9.4) | (247.8) | | (1.02) | (36.16) | | |
| *p.a* (%) | 4.22 | 7.24 | -16.08 | 3.44 | 2.67 | -16.49 | -4.57 | 0 |
| | | (12.52) | (73.61) | | (3.26) | (43.55) | | |
| **Nairobi** | | | | | | | | |
| *DPGR* | 0.63 | 0.68 | -0.58 | 0.55 | 0.43 | -0.57 | -0.25 | 0.07 |
| | | (1.03) | (4.19) | | (0.55) | (2.64) | | |
| *p.a* (%) | 4.99 | 4.06 | -8.34 | 4.49 | 3.07 | -8.16 | -0.99 | 0.17 |
| | | (5.34) | (17.89) | | (3.68) | (13.79) | | |
| **Central** | | | | | | | | |
| *DPGR* | 0.19 | 0.18 | -0.54 | 0.21 | 0.14 | -0.84 | -0.04 | 0.03 |
| | | (0.4) | (6.16) | | (0.28) | (2.48) | | |
| *p.a* (%) | 1.73 | 1.36 | -7.48 | 1.96 | 1.07 | -16.49 | -0.29 | 0.01 |
| | | (2.39) | (21.76) | | (2.23) | (13.27) | | |
| **Coast** | | | | | | | | |
| *DPGR* | 0.34 | 0.37 | -0.73 | 0.39 | 0.35 | -0.65 | -0.01 | 0.76 |
| | | (0.51) | (3.75) | | (0.55) | (6.44) | | |
| *p.a* (%) | 2.96 | 2.65 | -12.38 | 3.36 | 2.65 | -9.88 | -0.01 | 0.99 |
| | | (3.39) | (16.87) | | (2.81) | (22.22) | | |
| **Eastern** | | | | | | | | |
| *DPGR* | 0.24 | 0.22 | -0.83 | 0.24 | 0.33 | -0.51 | 0.11 | 0.1 |
| | | (0.31) | (2.94) | | (1.58) | (36.16) | | |
| *p.a* (%) | 2.14 | 1.78 | -16.08 | 2.17 | 2.16 | -6.79 | 0.39 | 0.03 |
| | | (2.42) | (14.7) | | (3) | (43.55) | | |
| **North Eastern** | | | | | | | | |
| *DPGR* | 1.21 | 1.65 | -0.65 | 1.83 | 2.52 | -0.41 | 0.87 | 0.09 |
| | | (2.36) | (14.08) | | (3.12) | (12.05) | | |
| *p.a* (%) | 8.26 | 7.69 | -9.85 | 10.97 | 10.44 | -5.1 | 2.75 | 0.07 |
| | | (7.61) | (31.17) | | (8.26) | (29.29) | | |
| **Nyanza** | | | | | | | | |
| *DPGR* | 23.4 | 23.9 | 8.92 | 0.29 | 0.26 | -0.36 | -23.64 | 0 |
| | | (12.57) | (247.8) | | (0.19) | (1.86) | | |
| *p.a* (%) | 37.64 | 37.39 | 25.79 | 2.59 | 2.27 | -4.38 | -35.12 | 0 |
| | | (3.49) | (73.61) | | (1.42) | (11.08) | | |
| **Rift Valley** | | | | | | | | |
| *DPGR* | 0.45 | 0.47 | -0.7 | 0.48 | 0.54 | -0.58 | 0.07 | 0.21 |
| | | (1.39) | (35.04) | | (0.79) | (9.93) | | |
| *p.a* (%) | 3.81 | 3.11 | -11.36 | 4 | 3.82 | -8.38 | 0.71 | 0 |
| | | (3.79) | (43.11) | | (3.35) | (27.02) | | |
| **Western** | | | | | | | | |
| *DPGR* | 0.3 | 0.27 | -0.45 | 0.31 | 0.28 | -0.24 | 0.01 | 0.41 |
| | | (0.23) | (1.63) | | (0.14) | (0.64) | | |
| *p.a* (%) | 2.65 | 2.27 | -5.77 | 2.72 | 2.42 | -2.66 | 0.15 | 0.07 |
| | | (1.71) | (10.17) | | (1.15) | (5.07) | | |

### A.2.4 Temporal and spatial variation of rainfall

Figure A.9: Long-term of yearly rainfall departures

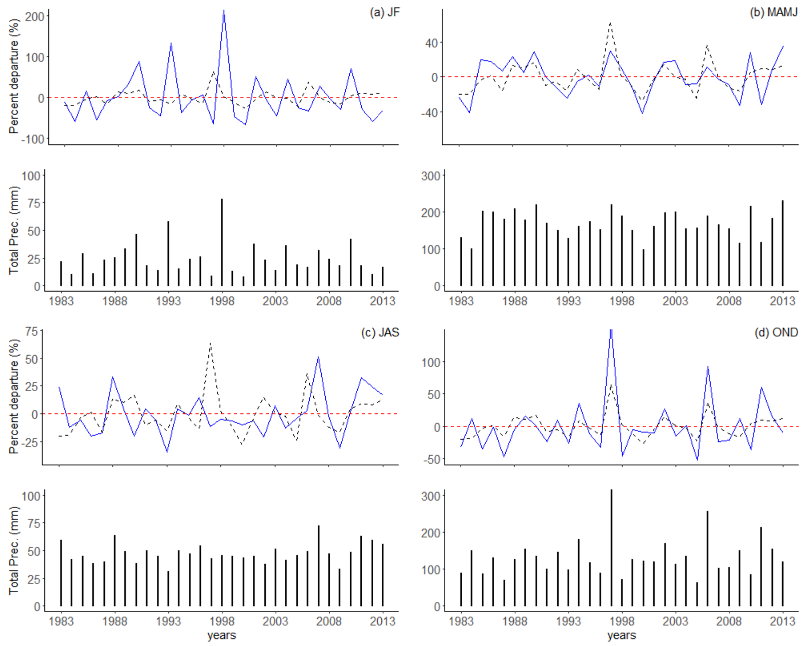



**Notes:** The Figures represent the time series of yearly rainfall departures ( % above or below 1983-2013 mean) and time series of annual precipitations (mm)
*Sources :* author's elaboration on CHIRPS data.

Figure A.9 plots the time series of annual precipitation departures aggregated over the country and interannual rainfall variability. Figure A.10 shows the time series of the seasonal anomalies, compared with the annual departures (in dashed lines), and informs about intraseasonal variability. The 1992-1995, 1999-2001, 2005-2006, and 2009-2010 droughts are all distinct in the four seasons. OND variations are high during the 1983-2013 period, with rainfall on the order of −50% below the mean ( Figure A.10 (d) OND). As expected from the literature, the short rains seem to play a major role in the interannual variability and have the highest correlation with yearly departures over the 1983-2013 period. However, short rain variations, linked to the ENSO, are stable over the decades (same rainfall deficit in 1987 as in 2005). Figure A.10 (b) displays an increase in dry conditions during the long-rains since 1999s, in comparison to the previous decade ( two small dry events of less than −15% deficit over the 1988-1998 decade, while 3 major ones over the 1999-2009 decade, up to −35%). This is in line with the main results of Lyon and DeWitt, 2012, which observed a drastic failure in the long rains after 1999 over East Africa. This suggests that the decline in precipitation in Kenya since the 1980s is mainly borne by the fall in the long rains since 1999s (manifested by longer dry spells, more intense and concentrated rains over the long rains, mainly over Eastern Africa, as shown in long-trends analysis).

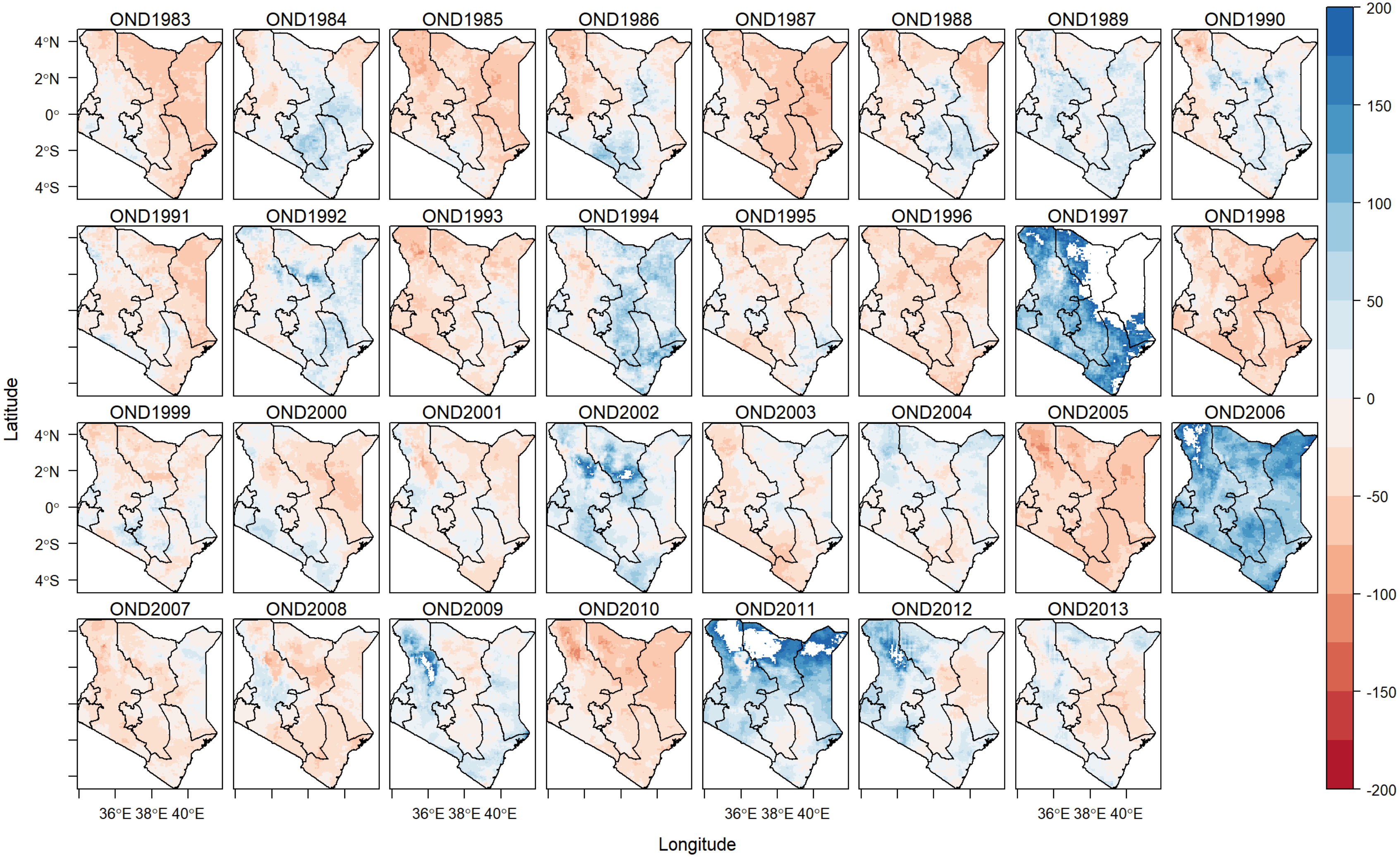
Figure A.11 plots the spatial pattern of the departures of the rains over the short-rainy season from October to December (OND). Severe floods over the short-rainy season can be identified in 1997 and 2006.

Figure A.10: Long-term of seasonal rainfall departures




**Notes:** The Figures represent time series of seasonal rainfall departures ( % above or below the 1983-2013 mean) compared with the annual departures, and total precipitation for (a) January-February (JF) (b) March-April-May-June (MAMJ) (c) July-August-September(JAS) and (d) October-November-December (OND)
*Sources :* author's elaboration on CHIRPS data.

Figure A.11: Rainfall percent departures of the short-rainy season (OND) from the 1983-2013 mean



## A.3 Main result

### A.3.1 Climate variability

Table A.8 looks at the interaction of the number of droughts with a dummy which equals 1 if the sublocation has been hit by at least one flood over the decade. Independent variables are computed based on the 10th and 90th percentile of each sublocation rainfall distribution.

The Table displays the results across sublocation types. First, it shows that the main result is robust to controlling for flood occurrence. Then, it shows that the out-migration is higher for sublocations that were not hit by any flood during the period, as the effect is twice bigger. Under no excessive rains, an additional dry rainy season decreases the DPGR by 3.97 p.p, which corresponds to a 14% decrease. The effect holds within agriculture-oriented rural areas, where the DPGR decreases by 3.5 p.p (17% decrease), and is still higher in rural areas where pastoralism prevails (18.5% decrease). The interaction term shows that the effect of an additional drought is significantly attenuated for sublocations hit by at least one flood by 2.5 p.p. Overall if a sublocation has faced at least one flood, an additional dry year decreases the DPGR by 1.5 p.p. The occurrence of flood cancels out the decrease of the DPGR in rural areas with low pastoralism.

This result can have several readings. First, it might reveal that being hit by both rainfall shortages and excess reduces the financial capacity to migrate for individuals. Second, it could be that excessive rainfall over the rainy season is less severe than droughts, a result that is found in the analysis of the spatial and temporal variation of rainfall (Figure 1.5). In this sense, the rainfall extremes would be beneficial and would attenuate the negative effects of droughts on agricultural outcomes for instance. If the results from Section **??** are more in line with the second interpretation, these results have to be read carefully. Even if the number of dry years over each period is not correlated with the number of wet years [5], as suggested in Figure 1.5, the results might be biased due to multicollinearity.

[5]coefficient correlation: 0.0053

Table A.8: Effects of the number of dry and wet rainy seasons on the DPGR

| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Number of dry years | -3.997*** | -1.527 | -4.304*** | -3.551** | -6.348*** |
| | [0.899] | [2.684] | [0.963] | [1.605] | [1.550] |
| Number of wet years ¿0 | 1.248 | 1.195 | 1.122 | -0.854 | 2.609 |
| | [1.206] | [3.596] | [1.292] | [1.617] | [2.524] |
| Number of dry years × Number of wet years ¿0 | 2.485** | 0.933 | 2.696** | 3.934** | 3.041 |
| | [1.063] | [3.167] | [1.126] | [1.620] | [2.263] |
| Period FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| N | 5036 | 756 | 4280 | 1626 | 1800 |
| R2 | 0.677 | 0.747 | 0.661 | 0.705 | 0.613 |
| Mean DPGR (%) | 27.75 | 31.55 | 27.08 | 20.09 | 34.15 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

# A.4 Heterogeneity

## A.4.1 Gender and age brackets

Table A.9 gives the results across gender and sublocation types restricting to the [15;65] years old cohort. Table A.10 gives the results on the whole [0;69] cohort for comparison purposes.

Table A.9: Effects of the number of dry rainy seasons across gender and location

| Sample | All Kenya | | Rural | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | All | | Low Pastoralism | | High Pastoralism | |
| **RDPGR** | Males | Females | Males | Females | Males | Females | Males | Females |
| **[15,65]** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Nb of dry years | -0.455*** | -0.423*** | -0.579*** | -0.521*** | -0.328** | -0.449*** | -0.792*** | -0.609*** |
| | [0.0989] | [0.0928] | [0.116] | [0.111] | [0.165] | [0.164] | [0.192] | [0.184] |
| Nb dry years × density | 0.000106 | 0.000232* | 0.000763*** | 0.00113*** | 0.000682* | 0.000702** | 0.00125** | 0.00232*** |
| | [0.000122] | [0.000128] | [0.000282] | [0.000273] | [0.000362] | [0.000343] | [0.000633] | [0.000657] |
| Period FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 5036 | 5036 | 4280 | 4280 | 1626 | 1626 | 1800 | 1800 |
| R2 | 0.567 | 0.609 | 0.542 | 0.582 | 0.558 | 0.589 | 0.525 | 0.566 |
| Mean RDGR (%) | -2.449 | -2.789 | -2.601 | -2.915 | -3.210 | -3.550 | -1.971 | -2.249 |
| Share (%) | | | | | | | | 16.99 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

Table A.10: Effects of the number of dry rainy seasons on RDPGR

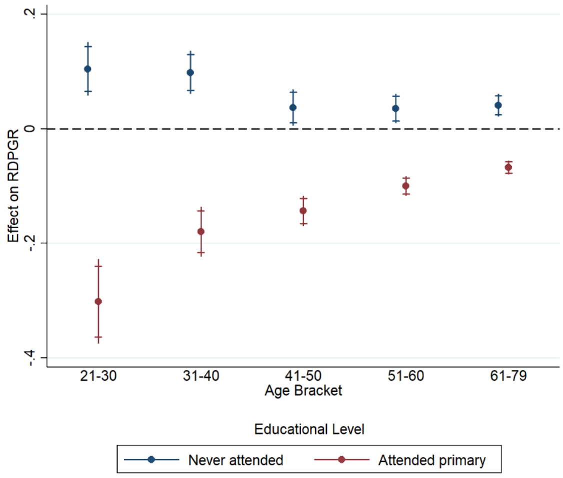
| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
|---|---|---|---|---|---|
| **[0,69]** | (1) | (2) | (3) | (4) | (5) |
| Number of dry years | -1.196*** | -0.0900 | -1.875*** | -1.205* | -2.278*** |
| | [0.382] | [1.065] | [0.459] | [0.669] | [0.761] |
| Number of dry years × density | 0.000758 | 0.0000579 | 0.00494*** | 0.00369*** | 0.00895*** |
| | [0.000491] | [0.000383] | [0.00112] | [0.00142] | [0.00274] |
| Period FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| Yes | | | | | |
| N | 5036 | 756 | 4280 | 1626 | 1800 |
| R2 | 0.645 | 0.749 | 0.607 | 0.619 | 0.594 |
| Mean RDPGR (%) | -8.640 | -1.686 | -9.868 | -12.93 | -6.784 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

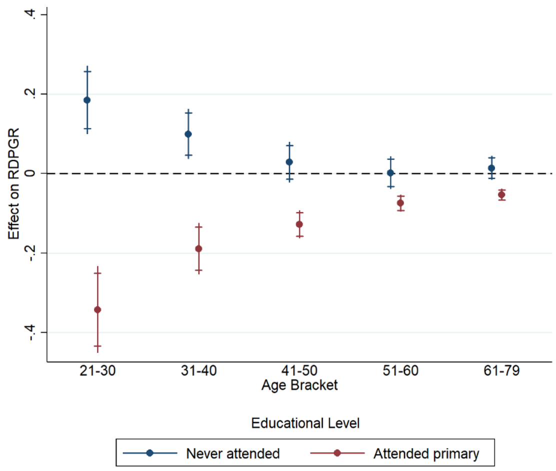
## A.4.2 Education

Figure A.12 displays the heterogeneity across educational levels per age bracket of the adults. As schooling attendance has increasing trends in Kenya, it could be possible that the result on the unskill individuals is driven by older adults. Figure A.12 shows that it is not the case, and the result across the skill distribution is not driven by any age effects.

Figure A.12: Effect of the number of dry rainy seasons across educational level and age brackets




(a) All Kenya



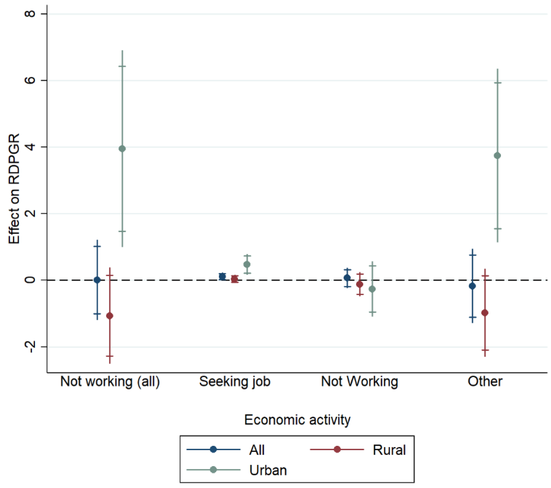

(b) Rural areas - Low-Pastoralism

**Notes:** Figure (a) plots the main result of the number of dry years across age brackets activity of individuals that never attended school and those who attended at least primary education. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is not the main agricultural activity.
*Sources:* Author's elaboration on KNBS and CHIRPS data.

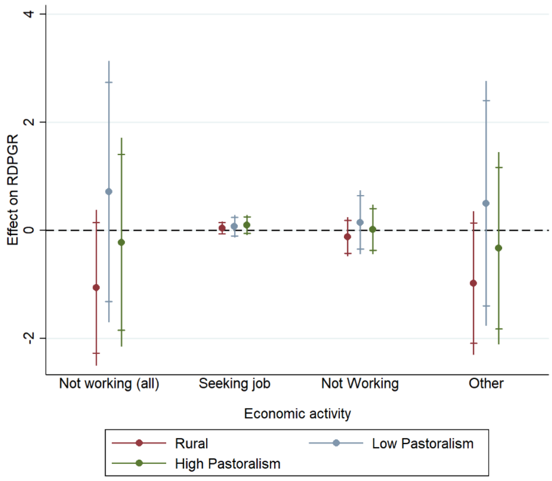
## A.4.3 Economic Activity

Figure A.13 displays the heterogeneity results according to the economic activity. It breaks down the results on the *Not working* individuals. If the results are mainly driven by the class *Other*, the Figure shows a significant increase of the unemployed population within urban areas. The magnitude of the effect on individuals seeking jobs is small as the unemployed accounts for a small proportion of the population in age of working (only 9%).

Figure A.13: Effect of the number of dry rainy seasons across economic activity and location 2




(a) All Kenya




(b) Rural areas

**Notes:** Figure (a) plots the main result of the number of dry years across economic activity of individuals in the age of working in the first year of the decade. Figure (b) plots the same coefficient, focusing on rural sublocations where pastoralism is the main agricultural activity and where it is not.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

## A.5 Robustness

### A.5.1 Binary treatment

Table A.11 gives the results of the binary treatment for the $DPGR_{[15,65]}$.

### A.5.2 Common trend assumption

Table A.12 shows that the results of both the continuous and binary treatments are robust when restricting the sample to the 668 sublocations used in the test of parallel trends in Section 1.9.2.

Table A.11: Effects of the increase in droughts on the DPGR - Binary treatment

| Outcome | DPGR [15,65] | | | | |
|---|---|---|---|---|---|
| | All Kenya | Urban | Rural | Low Pastoralism | High Pastoralism |
| | (1) | (2) | (3) | (4) | (5) |
| Dummy treatment × Period | -1.894** | 0.556 | -2.225** | -1.009 | -5.280*** |
| | [0.899] | [2.698] | [0.953] | [1.185] | [1.989] |
| Dummy Period | -1.585** | -2.429 | -1.475** | -1.810* | 0.239 |
| | [0.652] | [1.730] | [0.702] | [0.936] | [1.235] |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes |
| N | 3708 | 436 | 3272 | 1248 | 1316 |
| R2 | 0.667 | 0.701 | 0.663 | 0.727 | 0.606 |
| Size Control Group | 1148 | 133 | 1015 | 336 | 460 |
| Size Treatment Group | 706 | 85 | 621 | 288 | 198 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

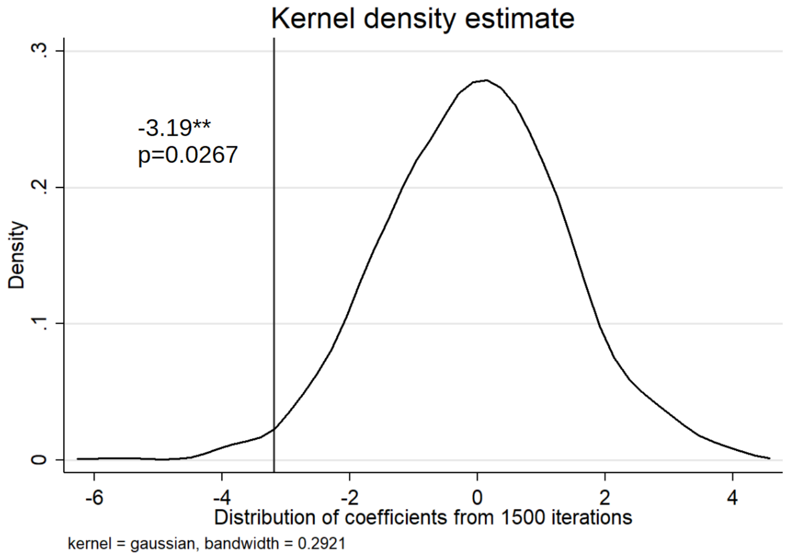
Table A.12: Effects of the number of dry rainy seasons on the DPGR - Continuous and Binary treatment - Restricted Sample

| | TWFE - Continuous treatment | | | DiD - Binary treatment | | |
|---|---|---|---|---|---|---|
| | All Kenya | Rural | High Pastoralism | All Kenya | Rural | High Pastoralism |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Number of dry years | -2.415** | -2.669** | -4.153*** | | | |
| | [1.084] | [1.187] | [1.132] | | | |
| Number of dry years × density | 0.0000396 | -0.00280 | 0.0143*** | | | |
| | [0.000499] | [0.00400] | [0.00414] | | | |
| Dummy treatment × Period | | | | -3.484* | -4.631** | -7.274 |
| | | | | [2.084] | [2.214] | [4.763] |
| Period | | | | 3.518** | 4.509*** | 6.939** |
| | | | | [1.380] | [1.501] | [2.781] |
| Period FE | Yes | Yes | Yes | | | |
| Sublocation FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1336 | 1164 | 1800 | 1336 | 1164 | 466 |
| R2 | 0.705 | 0.700 | 0.613 | 0.704 | 0.698 | 0.611 |
| Mean DPGR (%) | 26.30 | 25.25 | 34.15 | 26.30 | 25.25 | 33.08 |

**Notes:**Standard errors clustered at the sublocation level, $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

### A.5.3 Spurious correlation

Figure A.14: Temporal randomization inference tests - Binary treatment




**Notes:** This Figure represents the distribution of the treatment effects of the binary treatment when conducting 1,500 permutations. Each permutation randomly changes the sublocations allocation to the treatment. As the binary treatment excludes sublocations that have known close droughts straddling the two censuses, the sample change for each estimation, while the sample size remains 3708 for each permutation. The vertical line represents the main treatment effect using my main estimation (Table 1.7 Column 1) and gives the new estimated p-value.
*Sources:* Author's elaboration on CHIRPS and KNBS data.

### A.5.4 Other climate indicators

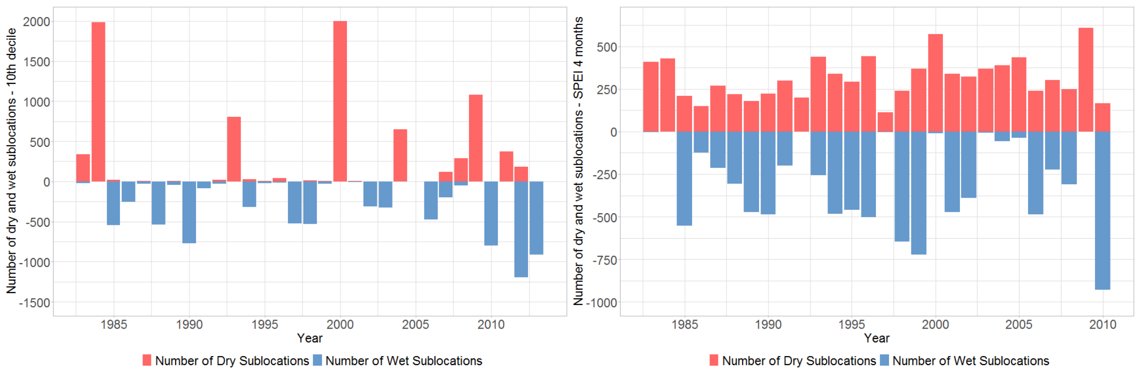
Figure A.15 compares the distribution of the number of sublocations hit by droughts across years, according to the 10th decile and SPEI definitions. Table A.13 displays the effects of the number of dry and wet years, based on cumulative rains over the short rainy season OND being under the 10th decile and over the 90th decile. It shows that extreme events occurring during the short-rainy season have little effect on migration.

The Standardized Precipitation Evaporation Index (SPEI) is a multiscalar index (Vicente-Serrano, Beguería, and López-Moreno, 2010), including the role of precipitation, temperature, and potential evapotranspiration as it captures anomalies of the water balance. For each sublocation, I build the SPEI using the CHIRPS and CHIRTS data on the 1983-2013 historical mean [6], which gives the standard

[6]The SPEI has been computed using the R library SPEI developed by Vicente-Serrano, Beguería, and López-Moreno, 2010

deviations of water balance. SPEI values give the intensity of droughts, as moderate droughts range from [-1,-2] and severe droughts from [-2,-3], and accordingly for excessive rainfall. For each MAMJ season, I calculate the number of months under a drought according to the SPEI values. I define a binary variable by taking the value one if the rainy season in year y was hit by 2 or 4 months, as calculated by the SPEI, and then calculate the number of the dry rainy season for each period.

Figure A.15: Number of dry and wet sublocations across indicators




(a) Number of dry years -10th decile (b) Number of dry years - SPEI 4 months

**Notes:** Figures plot in red the number of sublocations for which the rainy season is dry across year. Accordingly, it plots in blue the number of sublocations for which the rainy season is wet. Figure (a) plots it based on the main independent variable of this paper, relying on the cumulative rains being below the 10th decile. Figure (b) defines a dry year if the entire rainy season is under droughts according to the SPEI definition.
*Sources:* Author's elaboration on CHIRPS data .

Table A.13: Effects of the number of dry short-rainy seasons (OND) on the DPGR

| | All Kenya | | Rural areas | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Number of dry years | -0.222<br>[0.778] | | 0.313<br>[0.815] | |
| Number of wet years | | -0.253<br>[0.755] | | -0.442<br>[0.850] |
| Period FE | Yes | Yes | Yes | Yes |
| Sublocation FE | Yes | Yes | Yes | Yes |
| N | 5036 | 5036 | 4280 | 4280 |
| R2 | 0.673 | 0.673 | 0.657 | 0.657 |
| Mean DPGR (%) | 27.75 | 27.75 | 27.08 | 27.08 |

**Notes:** Standard errors clustered at the sublocation level, $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

## A.6 Bilateral Migration at the District Level

In this section, I propose an analysis at the district level, looking at the effect of droughts on yearly bilateral migration between districts, from 1991 to 2007. This empirical fails at finding significant effects of droughts on bilateral migration at the district level, which shows the comparative advantage of this paper to look at the sublocation level in order to capture small-magnitude effects.

### A.6.1 Data

In the censuses, relevant and precise information about individual movements is only available at the district levels. Each individual is asked about his district of birth and previous residence, which corresponds to the district in which the individual was in August the year before the administrative censuses (1988, 1998, and 2008). The main advantage and uniqueness of the more recent censuses (1999 and 2009) is a retrospective question that allows building a panel at the district level. In both censuses is asked the duration of residence of each individual, indicates the date at which an individual moved to the current district *"When did <NAME> move to the current district ?*.

This specification of the 1999 and 2009 censuses enables to build of a retrospective panel of migration for each district and contributes to the micro-oriented literature which usually has limited years of analysis because uses the question of the place of residence the year preceding the census (1998 and 2008 - the answer to the question *Where was <NAME> living in August 1998/2008 ?*), such as Dallmann and Millock, 2017, using two censuses and thus two-time points. The main caveat about the variable of outmigration from each district is the fact that we assume that the district of birth and the origin districts are the same.

### A.6.2 Empirical analysis

I estimate the effects of yearly droughts on migration behaviors for all 41 districts of Kenya, over the 1991-2007 period, using both the 1999 and 2009 censuses. A panel of bilateral migration is built over 14 years (1991-1997 and 2001-2007, as year t and t-1 of censuses are excluded) using retrospective questions about the year of arrival into the district, and the year/place of birth. I take as the district of origin the place of birth for each individual, which equals the district left in the year $t$ as the place of birth, and thus neglect stages of migration. The in-migration is exact, as I have the year of the arrival of each individual leaving in district d in August 1999 and 2009 [7]. The population staying in the district $d$ at the year $t$ is calculated from the population leaving in district $d$ at the time of the census, such as

$$pop_{d,t} = pop_{d,t=99 \; or \; t=09} - entry_{post(t)} - birth_{post(t)} + exit_{post(t)}$$

Where $entry_{post(t)}$ is the number of individuals living in district d at the time of the census, but that arrived in this district after $t$, $exit_{post(t)}$ the number of individuals not leaving in this district at the time of the census but that left this district (proxied by the district of birth) after the year $t$. Finally, $birth_{post(t)}$ is the number of individuals living in the district $d$ at the time of the census but born after the year $t$. For each district, we miss individuals who died during the year $t$ and the time of the census.

I estimate a gravity equation on bilateral migration rates, controlling for existing migration determinants both in origin and destination districts. I use a Poisson Pseudo Maximum Likelihood (PPML) regression to account for the high proportion of zero flows. The model estimates the effect of short-term droughts occurring during the rainy season on inter-district migration. More formally, the empirical strategy can be written as follows:

$$M_{od,t} = \frac{m_{od,t}}{pop_{oo,t}} = \alpha_0 + \alpha_1 D_{o,t} + \gamma_o + \delta_{d,t} + \beta_{od} + \epsilon_{od,t} \tag{A.1}$$

Where $M_{od,t}$ measures migration rates from the district of origin to the district of destination d. $\gamma_o$,$\delta_{d,t}$ and $\beta_{od}$ are origin, destination $\times$ year fixed effects, taking into account time-varying characteristics of the destination of migration. Characteristics of the migration pair such as the distance traveled, the presence of a common border, are captured by bilateral fixed effect, $\beta_{od}$. $D_{i,t}$ is the dummy for the dry shock, which indicates whether the rainy season of the year t is considered as dry or not.

[7] As the censuses have a more detailed question, which is the place of residence the year preceding the census (so 1998 and 2008 - the answer to the question *Where was ¡NAME¿ living in August 1998/2008 ?*), I do not take these two years into account, to avoid bias of better self-reporting (I indeed observe a peak of migration during these years)

### A.6.3 Results

Table A.14 compares an OLS to a PPML estimator of equation A.1. Columns (1) and (2) present the OLS estimators, and Columns (3) and (4) the PPML ones. A caveat of the use of OLS estimator for bilateral migration is the pairs of districts with no migration for a given year, which is corrected in PPML estimations. Columns (2) and (3) restrict the analysis by excluding the observations with zero bilateral migration flows. All estimations fail to find a significant effect of yearly droughts on inter-district migration. However, all estimators are positive, and can, be read as being hit by a drought increases by 8% the migration rates in the bilateral flows.

Table A.14: Effect of yearly droughts on bilateral migration at the district level

| | OLS | | PPML | |
|---|---|---|---|---|
| | $ln(M_{od,t})$ | $ln(M_{od,t}+1)$ | $M_{od,t}$ | $M_{od,t} > 0$ |
| | (1) | (2) | (3) | (4) |
| Dry year | 0.00145<br>[0.00185] | 0.000308<br>[0.00410] | 0.0789<br>[0.133] | 0.0826<br>[0.133] |
| Origin district FE | Yes | Yes | Yes | Yes |
| Destination-time FE | Yes | Yes | Yes | Yes |
| Origine-destination FE | Yes | Yes | Yes | Yes |
| Zero migration rates excluded | No | Yes | No | Yes |
| N | 23534 | 18312 | 23520 | 18312 |

**Notes:** Standard errors clustered at the origin district level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

# Appendix B

# Appendix to Chapter 2 : MiningLeaks : Water Pollution and Child Mortality in Africa

## B.1 Descriptive Statistics

### B.1.1 Data

Table B.1 displays for each country the number and years of DHS waves, and the total number of DHS clusters and children under 5 years old. Overall, DHS sample gathers 36 countries overall Africa, from 1986 to 2018. In our main empirical analysis, we decided to only keep DHS countries that had at least two survey rounds, in order to have comparable temporal variation across countries. Our final sample accounts for the following countries (cf. Table B.2): Tanzania, Burkina-Faso, Ghana, Zimbabwe, Mali, Democratic Republic of Congo, Guinea, Namibia, Madagascar, Cote d'Ivoire, Sierra Leone, Liberia, Nigeria, Senegal, Ethiopia, Uganda, Botswana, Malawi, Cameroon, Morocco, Niger, Kenya, Mauritania, Rwanda, Burundi, Lesotho, Togo, Eswatini, Algeria, Benin, Eritrea, Republic of the Congo, Guinea-Bissau, Somalia, Sudan, Tunisia, Djibouti, Equatorial Guinea (by order of importance in terms of mining activity according to Figure B.4).

Table B.1: DHS surveys overall across countries

| Countries | Survey Years | Number of clusters | Number of children under 5 |
|---|---|---|---|
| AO | 2015 | 625 | 14,177 |
| BF | 1993, 1999, 2003, 2010 | 1,413 | 36,744 |
| BJ | 1996, 2001, 2012, 2017 | 1,752 | 31,884 |
| BU | 2010, 2016 | 930 | 20,824 |
| CD | 2007, 2013 | 836 | 27,307 |
| CF | 1994 | 230 | 2,639 |
| CI | 1994, 1998, 2012 | 674 | 12,227 |
| CM | 1991, 2004, 2011, 2018 | 1,619 | 31,279 |
| EG | 1992, 1995, 2000, 2003, 2005, 2008, 2014 | 7,741 | 75,394 |
| ET | 2000, 2005, 2010, 2016 | 2,313 | 42,173 |
| GA | 2012 | 334 | 5,911 |
| GH | 1993, 1998, 2003, 2008, 2014 | 2,037 | 17,931 |
| GN | 1999, 2005, 2012, 2018 | 1,289 | 26,588 |
| KE | 2003, 2008, 2014 | 2,391 | 32,235 |
| KM | 2012 | 252 | 3,134 |
| LB | 1986, 2007, 2013 | 776 | 16,224 |
| LS | 2004, 2009, 2014 | 1,199 | 10,269 |
| MA | 2003 | 480 | 6,030 |
| MD | 1997, 2008 | 860 | 15,932 |
| ML | 1996, 2001, 2006, 2012, 2018 | 1,867 | 52,996 |
| MW | 2000, 2004, 2010, 2015 | 2,655 | 56,688 |
| MZ | 2011 | 610 | 10,950 |
| NG | 1990, 2003, 2008, 2013, 2018 | 3,830 | 106,848 |
| NI | 1992, 1998 | 503 | 11,332 |
| NM | 2000, 2006, 2013 | 1,290 | 13,630 |
| RW | 2005, 2008, 2010, 2014 | 1176 | 21,927 |
| SL | 2008, 2013 | 787 | 17,483 |
| SN | 1993, 1997, 2005, 2010, 2012, 2014, 2015, 2016, 2017 | 2,572 | 73,084 |
| SZ | 2006 | 274 | 2,706 |
| TD | 2014 | 624 | 18,441 |
| TG | 1988, 1998, 2013 | 768 | 13,869 |
| TZ | 1999, 2010, 2015 | 1,259 | 20,520 |
| UG | 2000, 2006, 2011, 2016 | 1,765 | 37,603 |
| ZA | 2017 | 671 | 3,397 |
| ZM | 2007, 2013, 2018 | 1,585 | 29,105 |
| ZW | 1999, 2005, 2010, 2015 | 1,431 | 19,847 |

**Notes:** This table gives the sample size of children under five years old overall DHS surveys.

Table B.2: DHS surveys in regression sample across countries

| Countries | Survey Years | Number of clusters | Number of children under 5 |
|---|---|---|---|
| BF | 1993, 1999, 2003, 2010 | 694 | 23,846 |
| BJ | 2001, 2012, 2017 | 62 | 1,911 |
| BU | 2010, 2016 | 317 | 8,280 |
| CD | 2007, 2013 | 82 | 5,092 |
| CI | 1994, 1998, 2012 | 196 | 4,838 |
| CM | 1991, 2004, 2011, 2018 | 90 | 2,513 |
| ET | 2000, 2005, 2010, 2016 | 100 | 2,956 |
| GH | 1993, 1998, 2003, 2008, 2014 | 1,217 | 12,074 |
| GN | 1999, 2005, 2012, 2018 | 360 | 11,775 |
| KE | 2003, 2008, 2014 | 233 | 4,130 |
| LB | 1986, 2007, 2013 | 190 | 7,537 |
| LS | 2004, 2009, 2014 | 336 | 2,810 |
| MD | 1997, 2008 | 131 | 3,301 |
| ML | 1996, 2001, 2006, 2012, 2018 | 570 | 19,147 |
| MW | 2000, 2004, 2010, 2015 | 207 | 6,651 |
| NG | 1990, 2003, 2008, 2013, 2018 | 105 | 3,993 |
| NI | 1992, 1998 | 40 | 1,105 |
| NM | 2000, 2006, 2013 | 138 | 2,175 |
| RW | 2005, 2008, 2010, 2014 | 713 | 14,615 |
| SL | 2008, 2013 | 377 | 13,717 |
| SN | 1993, 1997, 2005, 2010, 2012, 2014, 2015, 2016, 2017 | 363 | 10,111 |
| TG | 1988, 1998, 2013 | 104 | 2,187 |
| TZ | 1999, 2010, 2015 | 325 | 6,866 |
| UG | 2000, 2006, 2011, 2016 | 305 | 9,031 |
| ZM | 2007, 2013, 2018 | 364 | 10,966 |
| ZW | 1999, 2005, 2010, 2015 | 468 | 8,307 |

**Notes:** This table gives the sample size of children under five years old that are in our main analysis, meaning within 100 km of an industrial mine.

Tables B.3, B.4, and B.5 display the descriptive statistics of all our outcome and control variables for the sample of all individuals living within 45 km of an industrial mine, regardless of their topographic position, in the 26 countries of Sub-Saharan Africa with at least 2 waves of DHS and for heavy metals and coal mines. These descriptive figures are important to show that our analysis does not suffer from selection biases across the samples we use for our different regressions.

Table B.3: Descriptive statistics of children's outcomes

| | Mean | SD | Med | Min | Max | N |
|---|---|---|---|---|---|---|
| **Mortality rates** | | | | | | |
| 12-month mortality | 0.064 | .244 | 0 | 0 | 1 | 189,181 |
| 24-month mortality | 0.083 | .275 | 0 | 0 | 1 | 139,683 |
| **Control variables** | | | | | | |
| Birth order number | 3.655 | 2.421 | 3 | 1 | 18 | 240,431 |
| Male | 0.508 | 0.500 | 1 | 0 | 1 | 240,431 |
| **Anthropometric measures** | | | | | | |
| Stunting | 0.319 | 0.466 | 0 | 0 | 1 | 137,834 |
| Underweight | 0.234 | 0.423 | 0 | 0 | 1 | 136,043 |
| Wasting | 0.077 | 0.267 | 0 | 0 | 1 | 138,222 |
| **Weight and size at birth** | | | | | | |
| Less than 2.5 kg | 0.164 | 0.370 | 0 | 0 | 1 | 117,651 |
| Small or very small size | 0.161 | 0.367 | 0 | 0 | 1 | 226,796 |
| **Measured anemia level** | | | | | | |
| Any anemia | 0.633 | 0.482 | 1 | 0 | 1 | 67,567 |
| **Illness in the last 2 weeks** | | | | | | |
| Diarrhea | 0.168 | 0.374 | 0 | 0 | 1 | 216,097 |
| Cough | 0.260 | 0.439 | 0 | 0 | 1 | 214,940 |
| Fever | 0.265 | 0.441 | 0 | 0 | 1 | 214,913 |
| **Nutrition** | | | | | | |
| Given plain water | 0.187 | 0.390 | 0 | 0 | 1 | 122,915 |
| Ever breastfed | 0.980 | 0.140 | 1 | 0 | 1 | 223,039 |
| Months breastfed | 14.788 | 8.917 | 15 | 0 | 59 | 156,011 |
| **Health access** | | | | | | |
| No prenatal care | 0.101 | 0.301 | 0 | 0 | 1 | 169,268 |
| Ever vaccinated | 0.788 | 0.409 | 1 | 0 | 1 | 82,082 |
| **Characteristic of paired mine** | | | | | | |
| Domestic mine | 0.177 | 0.381 | 0 | 0 | 1 | 240,431 |
| Open-pit mine | 0.676 | 0.468 | 1 | 0 | 1 | 103,667 |

**Notes:** We present the mortality rates at $n$ months, conditionnally on having reached $n$ months, for the whole sample of children living within 45 km of an industrial mine and regardless of their topographic position. The sample is restricted to the 26 Sub-Saharan countries with at least two waves of DHS and to heavy metals and coal mines.

Table B.4: Descriptive statistics of mothers' outcomes

| | Mean | SD | Med | Min | Max | N |
|---|---|---|---|---|---|---|
| **Mother's characteristics** | | | | | | |
| Mother's age | 28.918 | 6.979 | 28 | 15 | 49 | 240,431 |
| Years of education | 3.985 | 4.226 | 3 | 0 | 22 | 240,332 |
| Urban | 0.287 | 0.452 | 0 | 0 | 1 | 236,966 |
| Migrant | 0.594 | 0.491 | 1 | 0 | 1 | 161,292 |
| **Access to sanitation and health facilities** | | | | | | |
| Piped water as main drinking water source | 0.261 | 0.439 | 0 | 0 | 1 | 240,431 |
| Has flushed toilet | 0.086 | 0.280 | 0 | 0 | 1 | 239,773 |
| Has electricity | 0.218 | 0.413 | 0 | 0 | 1 | 236,692 |
| Visited health facility in the last 12 months | 0.623 | 0.485 | 1 | 0 | 1 | 218,053 |

**Notes:** The sample is restricted to all mothers of 0-5 years old children living within 45 km of an industrial mine in the 26 Sub-Saharan countries with at least two waves of DHS and to heavy metals and coal mines.

Table B.5: Descriptive statistics of women's outcomes

| | Mean | SD | Med | Min | Max | N |
|---|---|---|---|---|---|---|
| **Fertility behavior and health** | | | | | | |
| Ever had a child | 0.736 | 0.441 | 1 | 0 | 1 | 330,889 |
| Total lifetime fertility | 2.890 | 2.785 | 2 | 0 | 18 | 330,889 |
| Currently pregnant | 0.091 | 0.288 | 0 | 0 | 1 | 330,744 |
| Ever had a miscarriage | 0.127 | 0.333 | 0 | 0 | 1 | 296,235 |
| Any anemia | 0.378 | 0.485 | 0 | 0 | 1 | 115,481 |
| **Placebo disease** | | | | | | |
| Any STD | 0.049 | 0.216 | 0 | 0 | 1 | 276,924 |
| Heard of tuberculosis | 0.935 | 0.246 | 1 | 0 | 1 | 88,438 |

**Notes:** The sample is restricted to all women aged 15-49 living within 45 km of an industrial mine in the 26 Sub-Saharan countries with at least two waves of DHS and to heavy metals and coal mines.

### B.1.2 Handwork

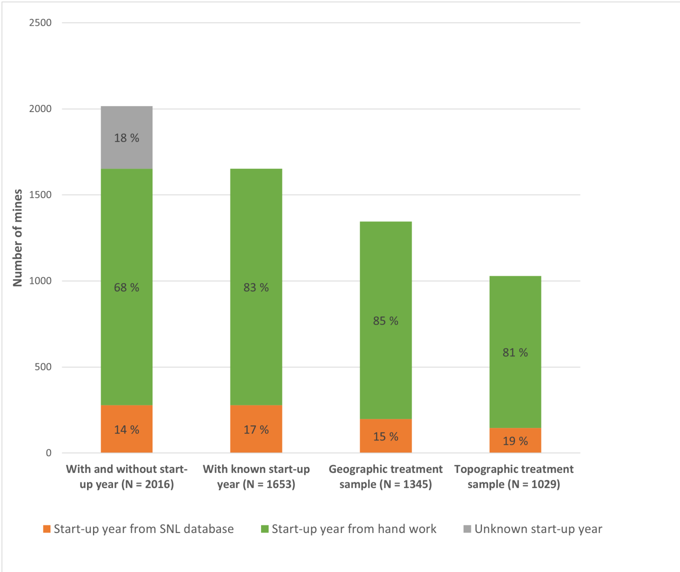
Out of the 3815 industrial mines recorded by the SNL database in Africa, 2016 were located within 100 km of a DHS cluster (with at least 2 waves of DHS). 278 had information on the opening and closing years within the database, and for the 1738 remaining mines, we searched for their years of opening [1]. The handwork consisted in reading the reports (comments and work history) available in the database and browsing through the aerial images available on the SNL platform which provided the exact GPS coordinates and main location labels. This information was corroborated with online research (press releases, mining companies' websites, specialized websites on global mining activities, etc.) as well as Google maps and Google timelapse satellite imagery. A mine opening corresponds to the beginning of the production.

The exact startup year could not be determined for 18 % of our sample (Figure B.1 Bar (1)), and these mines are dropped in our regressions. In total we hand-checked 83% of the mines located within 100 km of a DHS cluster, and for which we know their year of opening (Figure B.1 Bar (2)). Among the sample of mines with startup year, 83.2 % opened after 1981 (first year of birth within the DHS child surveys). For each of the following graphs, we study the whole sample of 2016 mines and plot the percentage of mines that were hand-checked and the percentage of mines that ends up having a startup year and are thus included in our study. We conduct this analysis on all the available mines within 100 km of a DHS cluster to be transparent on the creation of our sample compared to the original one.

The distribution across each mining site's primary commodity of production can be found in Figure B.3. Half of our sample consists of gold mining sites. Figure B.4 represents the distribution across country of location. Ownership information is available for 65 percent of our sample and the main owners are from the USA, UK, Canada, Australia, and China (Figure B.6).

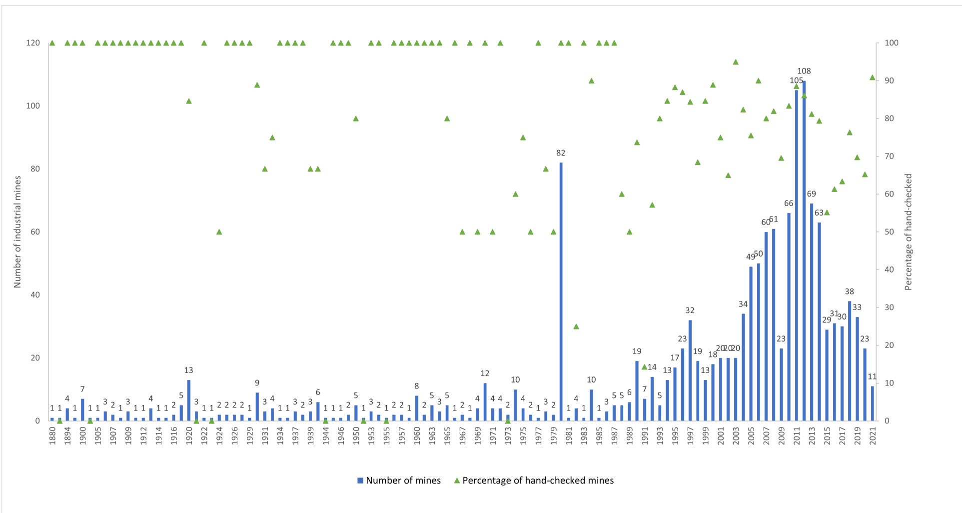
[1] We also looked at their closure date as well as their current activity status, i.e. whether the mining site looked active or inactive. However, this was an information harder to retrieve and finally we focused on the date of opening.

Figure B.1: Description of hand work and industrial mines samples



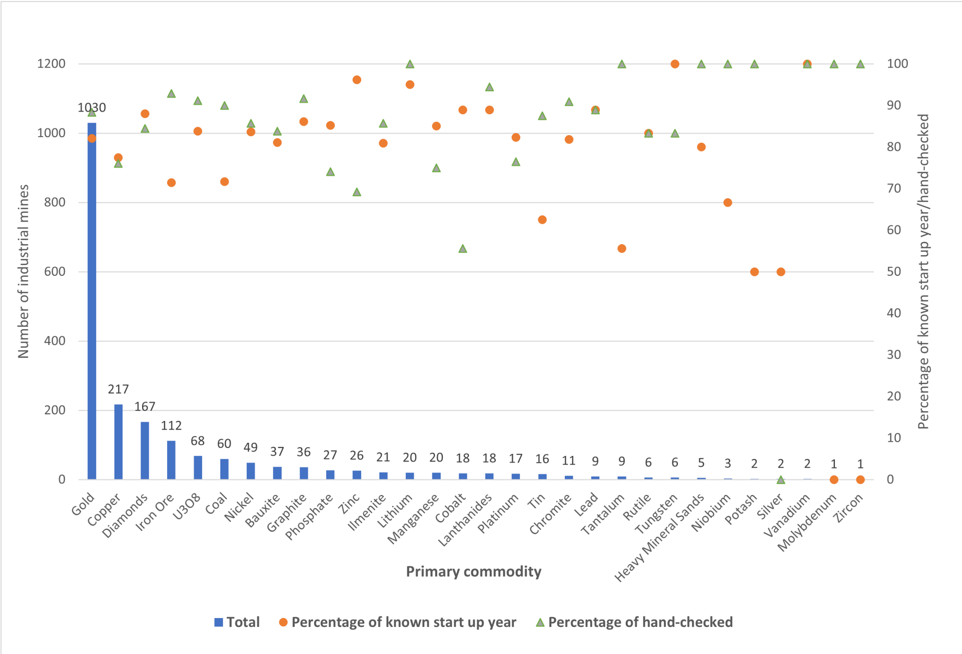

**Notes:** This Figure gives the number and percentage of mines for which we have retrieved the year of opening by hand. Bars (1) and (2) gives it for all the mines located within 100 km of a DHS cluster, Bar (3) for the sample associated to the replication of Benshaul-Tolonen, 2018 Section B.6. Bar (4) corresponds to the main analysis, i.e to mines that have at least one DHS cluster upstream within 100km, and one DHS cluster downstream within the three closest sub-basins (cf pairing strategy SectionA.1) .
*Sources:* Authors' elaboration on DHS and SNL data.

Figure B.2: Mines and percentage of hand-checked across start-up years



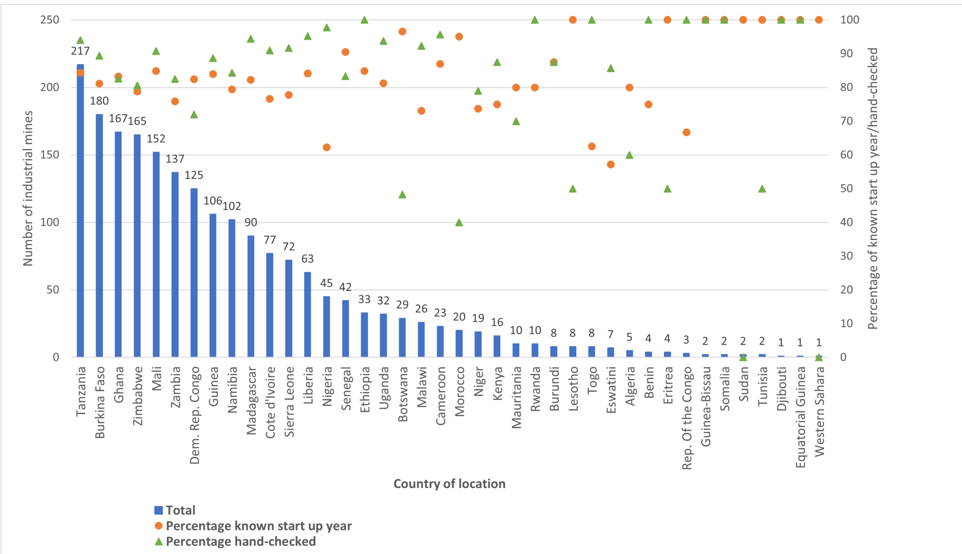

**Notes:** This graph displays the number of mines that opened during a specific year and the percentage of hand checked for the 2016 mines located within 100km of a DHS cluster.

Figure B.3: Mines and percentage of hand-checked across primary commodities



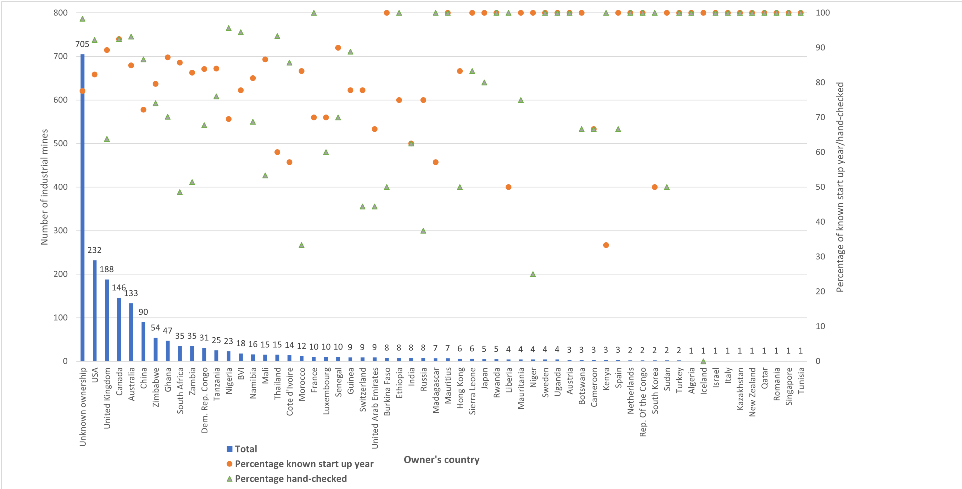

**Notes:** This graph gives the number of mines for each primary commodity, and the percentage of hand-checked, for the 2016 mines located wihin 100 km of a DHS cluster.

Figure B.4: Mines and percentage of hand-checked across country of location



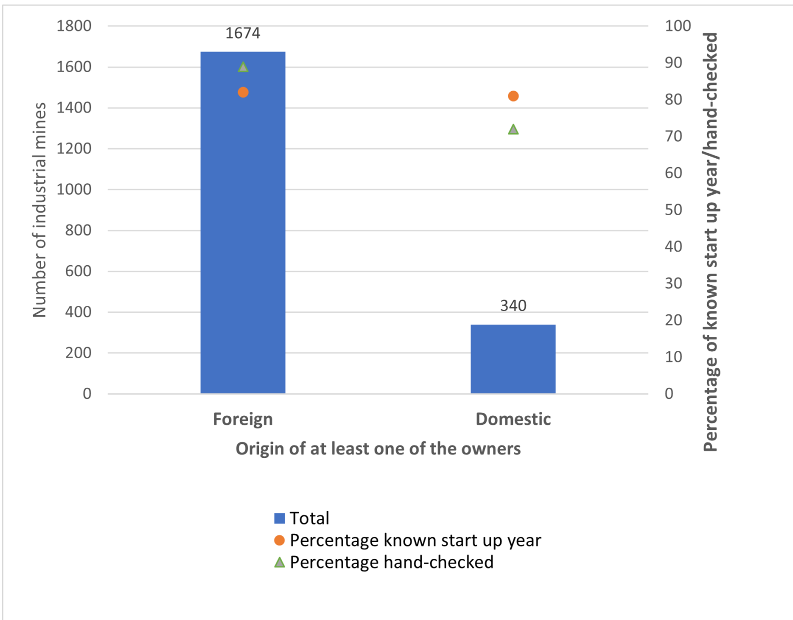

**Notes:** This graph gives the number of mines for each country of location, and the percentage of hand-checked, for the 2016 mines located wihin 100 km of a DHS cluster.

Figure B.5: Mines and percentage of hand-checked across owner's country




**Notes:** This graph gives the number of mines by owning company's registration country, and the percentage of hand-checked, for the 2016 mines located wihin 100 km of a DHS cluster.

Figure B.6: Mines across foreign and domestic ownership




**Notes:** This graph gives the number of mines across domestic and foreign ownership, and the percentage of hand-checked, for the 2016 mines located wihin 100 km of a DHS cluster.

# B.2 Context

## B.2.1 Case study: the Essakane mine

Figure B.7: Satellite image of Essakane Mine in 2019

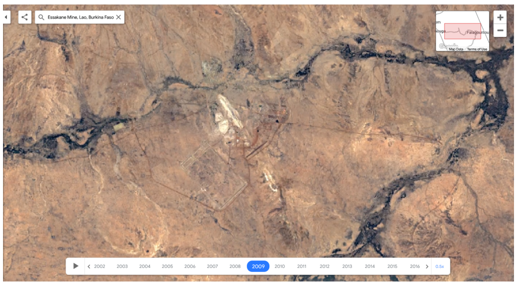

**Notes:** Satellite image of the Essakane Mine in 2019. Retention dams can be seen.
*Sources:* Google Earth.

Figure B.7 shows the satellite image of a mine from our sample, the Essakane mine in 2019, and Figure B.8 shows the different stages of expansion and construction of the mine. Essakane is the most productive gold mine and the second largest in Burkina-Faso, still in activity. It is an open-pit gold mine that extends over a 100 $km^2$ area. It is located in the North-East of Burkina-Faso in the Oudalan province, near the Nigerian and Malian borders, and is hydrologically found in the sub-basin of Gorouol and Feildegasse rivers. It is exploited by the Burkinabé society Iamgold Essakane and belongs to the Canadian investor IamgoldInc (International African Mining Gold Corporation), who obtained the project in 2007. The installation in 2009 of the mine has forcibly displaced five villages, and 16,000 people with no choice, and the promised compensation to the communities for the displacement cost, loss of pastures and common forests have not been fulfilled (Atlas des Conflits pour la Justice Environnementale, 2022). Mining at Essakane has been shown to have negative impacts on the environment and the health of the local population, both indirectly and accidentally.

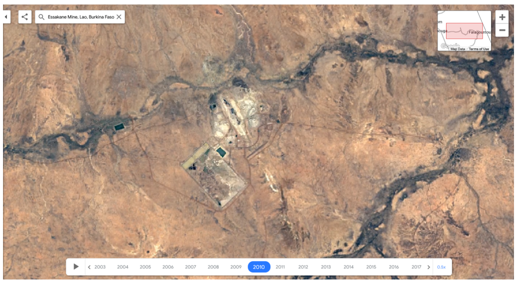
In November 2015 Drechsel, Engels, and Schäfer, 2018 has run qualitative interviews among the inhabitants of local communities of six active mining zones in Burkina-Faso, including the Essakane zone. If the local population admits the benefits of the construction of a primary school and educational establishment, of a health center, of roads and electricity, the interviewed people do not find it sufficient to outweigh the negative aspects. The mine does not display formal employment to the local population, not educated enough to undertake the required skilled-work. On the contrary, due to the loss of agricultural land and the prohibition to practice gold panning, the local population fell into unemployment and poverty, as a farmer from the Essakane area explains: *"Before the mine arrival, we had better lives, we had animals, we were rich"* (Drechsel, Engels, and Schäfer, 2018). In 2010, the tailing storage facility of the mine collapsed, which caused the death of the surrounding livestock poisoned by chemicals used in the mine, and created tensions between the local population and the operator. In 2011 a truck carrying two containers of cyanide fell into the water source of Djibo's dam and led to the death of all the fish in the dam. Tensions regarding water scarcity exist as well, the mine being water-consuming and reinforcing the vulnerability of the local population to droughts. Even if the regional government had prohibited the mine to use the village water, the national government overruled the decision, and the operator directly uses the water originally intended for the village. This led to the protestation of the local population around the mine in 2011, with no success. Finally, miners had major impacts on soil degradation, due to the construction of mining infrastructures, the multiplication of satellite pits and abandoned sterile holes devoid of gold (Porgo and Gokyay, 2016).

Environmental pollution has also degraded living conditions around the Essakane mine. Porgo and Gokyay, 2016 use water sampling and digital calipers to capture water pollution and particle measurement in the Essakane zone and survey the surrounding population to understand the related health conditions. They find high levels of particles at the Essakane site center due to transportation and mining activity, such as the work of perforation, blasting, loading, transportation of ore, crushing, grinding, and energy production based on hydrocarbons. This air pollution mainly concerns mine workers, who develop acute respiratory infections (ARI), uncurable lung diseases caused by prolonged and severe inhalation of fine particles. The drilled well water samples display abnormally high concentrations of arsenic (higher than the WHO standard), which comes from the intensive use of acids (low pH) and the liberation of trace metals. The surrounding population presents diarrhea (13%) and affections of the skin and wounds (11%), reported to be caused by lack of hygiene, and use of drugs and chemicals. The main important health impact associated with the mine is the increase of malaria (20%), as stagnant water from mining dams attracts infected mosquitoes.

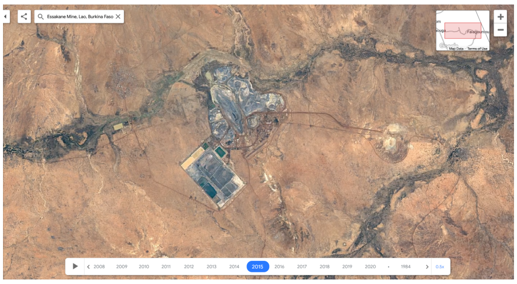
Figure B.8: Expansion of the Essakane Mine, 2005-2019.



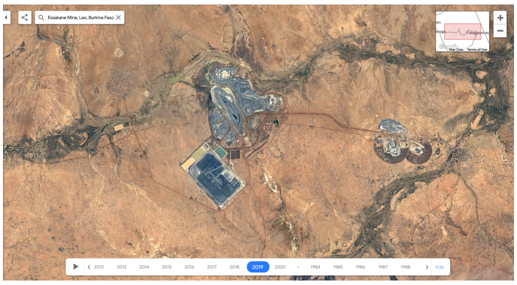
(a) 2009



(b) 2010



(c) 2015



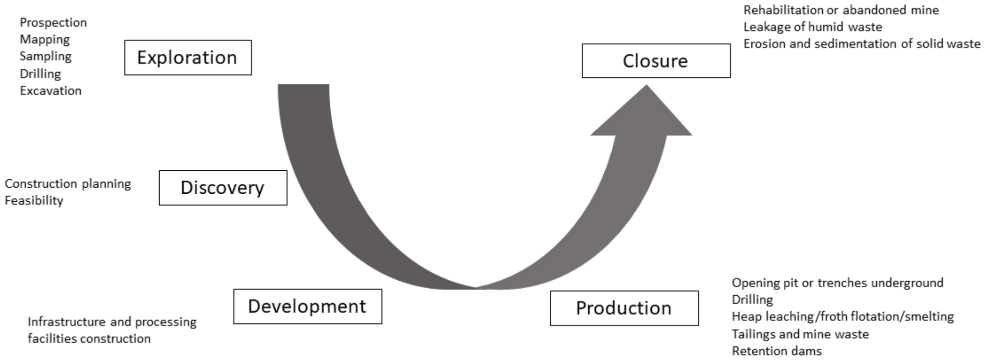
(d) 2019

**Notes:** The six satellite images represent the expansion of the Essakane plant, in Burkina Faso. Retention dams can be seen.
*Sources:* Google Earth engine Timelapse.

### B.2.2 Mine life cycles and types

Figure B.9 gives the main stages of an industrial mining project, from the exploration phase to the start of the production and the closure of the mine. If it is hard to give the average length of each phase, yet the average mine lifetime is 16 years from the start of production to its closure (Figure 2.11). Figure B.10 gives the time evolution of international prices for all commodities used in our main sample.

Figure B.9: Industrial mine's life cycle




**Notes:** The figure schematizes the main stages of an industrial mining project .
*Sources:* Authors' elaboration, largely inspired by Coelho, Teixeira and Goncalves (2011)

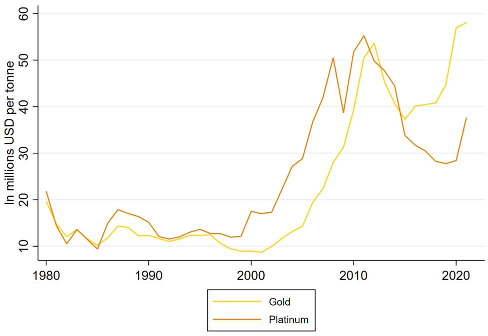
Table B.6 gives the chemical properties of each metal, including their main chemical compounds (Column 1), their density (Column 2), and displays their share in the main estimation sample (in terms of the number of mines Column (3) and Total Individual Sample (Column (4)). Heavy metals are defined according to their density as being greater than $5gcm^{-3}$ (Briffa, Sinagra, and Renald, 2020). If small amounts of heavy metals can be mandatory, a high and abnormal concentration of heavy metals may cause health issues due to chronic toxicity. Heavy metals released in mining activity are toxic elements that degrade the environment and human biology. This is the case as well for heavy metals released during the mining and burning of coal, which is linked to toxic heavy metals such as lead, mercury, arsenic, and nickel (Global Energy Monitor Wiki, 2021). This is the reason why the main regression analysis includes heavy metals and coal mines, to capture the negative externalities linked to the most toxic mines.

Table B.6: Metals, chemical properties and sample distribution

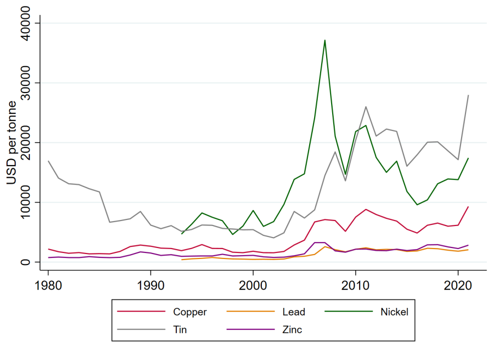
| Metals | Main chemical coumpounds (1) | density ($gcm^{-3}$) (2) | Nb. Mines (3) | Total Individual Sample (%) (4) |
|---|---|---|---|---|
| **Heavy Metals** | | | | |
| Gold | Gold | 19.3 | 581 | 41.88 |
| Copper | Copper | 8.96 | 89 | 5.03 |
| Iron ore | Iron | 7.87 | 54 | 8.72 |
| U308 | Uranium | 8.39 | 36 | 1.60 |
| Nickel | Nickel | 8.9 | 25 | 5.06 |
| Platinum | Platinum | 21.45 | 21 | 0.43 |
| Zinc | Zinc | 7.14 | 19 | 2.46 |
| Chromite | Iron<br>Chromium | [4.5,5.09 ] | 16 | 0.57 |
| Ilmenite | titanium | 4.6 | 14 | 3.67 |
| Lanthanides | Lanthane(57)<br>Lutecium(71) | [6.1,9.8] | 13 | 1.95 |
| Manganese | Manganese | 7.21 | 12 | 0.62 |
| Tin | Tin | [5.7;7.26] | 10 | 4.87 |
| Cobalt | Cobalt | 8.9 | 7 | 0.56 |
| Tungsten | Tugsten | 19.25 | 6 | 1.06 |
| Tantalum | Tantalum | 16.69 | 5 | 0.15 |
| Vanadium | Vanadium | 6.12 | 4 | 0.04 |
| Niobium | Niobium | 8.57 | 3 | 0.39 |
| Heavy Mineral Sands | Zirconium<br>Titanium<br>Tungsten<br>Thorum | [4.5,17.6] | 3 | 0.16 |
| Silver | Silver | 10.49 | 1 | 0.00 |
| Lead | Lead | 11.29 | 1 | 0.06 |
| **Non-Heavy Metals** | | | | |
| Diamonds | Carbon | 3.5 | 115 | 11.73 |
| Coal | Carbon<br>Mercury?<br>Arsenic? | 1.35 | 55 | 2.19 |
| Bauxite | Aluminium | 2.79 | 23 | 1.94 |
| Graphite | Carbon | 2.26 | 21 | 0.82 |
| Phosphate | Phosphate | 1.83 | 14 | 2.78 |
| Lithium | Lithium | 0.53 | 14 | 0.80 |
| Rutile | titanium | 4.23 | 2 | 0.29 |
| Potash(Salt) | Potassium | 0.89 | 1 | 0.17 |

**Notes:** This table gives for each metals the main chemical coumpounds (Column (1)) and their density (Column (2)). Columns (3) gives the number of mines within 100 km of a DHS cluster for which the metal is the main primary commodity, and Column (4) the percentage of children under 5 associated to these mines.

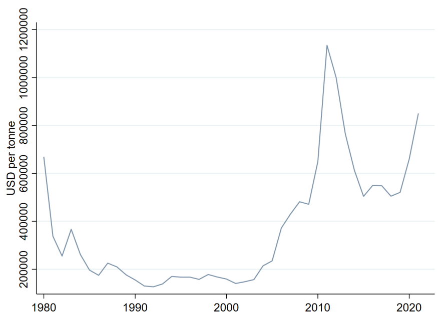
Figure B.10: Time evolution of international commodity prices



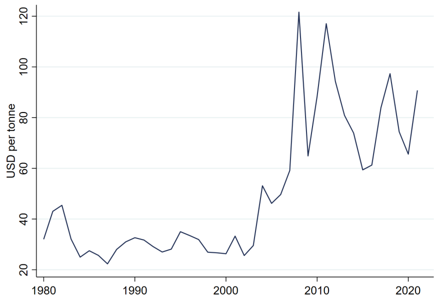

(a) Gold and Platinum




(b) Other metals




(c) Silver




(d) Coal

**Notes:** These Figures plot the evolution of metal prices from 1980 to 2020.
*Sources:* Authors' elaboration from SNL data and World Bank pink sheet data.

## B.3 Empirical Strategy

### B.3.1 Descriptive Statistics

Table B.7 gives the balance table for some household and mother characteristics. This is a descriptive table that accounts neither for controls nor fixed effects. Table B.8 gives the effect of being downstream of a mine for the same variables. We observe no statistical difference in terms of access to piped water, and electricity, the age, and years of education of the mother. However, Table B.8 shows that the proportion of urban households increases by 13 p.p once a mine has opened in downstrea areas, compared to upstream areas. The proportion of mothers that are migrants also increases by 8 p.p.. These results suggest that the in-migrants coming after the mine opening, seeking jobs, for instance, settle down in downstream areas that become more urban. It suggests that the miner villages are located downstream of the mine. This gives the necessity to control for migration and verify that this is not driving our results (cf Section 2.6.2).

We plot in Figure B.11 the distribution of mines opened within 100 km upstream or within the 3 closest sub-basins downstream during a child's birthyear, so as to see which countries gather the highest number of industrial mining activity in the vicinity of surveyed households over 1986-2018. Ghana, Zimbabwe, Tanzania, Zambia, Guinea and Sierra Leone have the highest density of open mines nearby DHS clusters, while Benin, Burundi, Cameroon, Lesotho and Niger have the lowest number of open mining sites. This figure also represents the variation in the number of mines that opened between the first and last year of surveys for each country. We can thus grasp the context of change in industrial mining activity over our period of interest. Ghana, Tanzania, Guinea, Mali, and Burkina-Faso witnessed the highest number of mine openings between 1986 and 2018.

Figures B.12, B.13 and B.14 give the spatial variation of the infant mortality outcomes and the mine openings for the sample restricted to our main analysis.

Table B.7: Balance Table - Double Difference with Topographic Treatment - Descriptive Statistics

| | Before Mine Opening | | | | | After Mine Opening | | | | | Within Up. | Within Dwn. | Within |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Upstream | | Downstream | | Diff | Upstream | | Downstream. | | Diff | | | |
| | N | Mean /(SD) | N | Mean /(SD) | (4-2) /(p.v) | N | Mean /(SD) | N | Mean /(SD) | (9-7) /(p.v) | (7-2) /(p.v) | (9-4) /(p.v) | (12-11) /(p.v) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| **Household Charateristics** | | | | | | | | | | | | | |
| **% Urban Household** | | | | | | | | | | | | | |
| All | 29,399 | 0.33 | 9,835 | 0.175 | -0.155 | 16,285 | 0.385 | 6,174 | 0.291 | -0.094 | 0.055 | 0.116 | 0.061 |
| | | (0.47) | | (0.38) | (0) | | (0.487) | | (0.454) | (0) | (0) | (0) | (0) |
| Mines | 244 | | 237 | | | 190 | | 193 | | | | | |
| **Has piped water** | | | | | | | | | | | | | |
| All | 29,399 | 0.307 | 9,835 | 0.193 | -0.115 | 16,285 | 0.365 | 6,174 | 0.27 | -0.095 | 0.057 | 0.077 | 0.019 |
| | | (0.461) | | (0.395) | (0) | | (0.481) | | (0.444) | (0) | (0) | (0) | (0.002) |
| **Has electricity** | | | | | | | | | | | | | |
| All | 29,399 | 0.211 | 9,835 | 0.14 | -0.071 | 16,285 | 0.356 | 6,174 | 0.226 | -0.13 | 0.145 | 0.086 | -0.059 |
| | | (0.408) | | (0.347) | (0) | | (0.479) | | (0.418) | (0) | (0) | (0) | (0) |
| **Mother Charateristics** | | | | | | | | | | | | | |
| **Age** | | | | | | | | | | | | | |
| All | 29,399 | 29.106 | 9,835 | 29.187 | 0.081 | 16,285 | 28.779 | 6,174 | 28.818 | 0.039 | -0.328 | -0.369 | -0.042 |
| | | (7.065) | | (7.039) | (0.325) | | (6.847) | | (6.986) | (0.707) | (0) | (0.001) | (0.764) |
| **Years of Education** | | | | | | | | | | | | | |
| All | 29,399 | 2.406 | 9,835 | 2.91 | 0.504 | 16,285 | 4.297 | 6,174 | 4.851 | 0.554 | 1.891 | 1.941 | 0.05 |
| | | (3.6) | | (3.741) | (0) | | (4.417) | | (4.2) | (0) | (0) | (0) | (0.001) |
| **% Migrant** | | | | | | | | | | | | | |
| All | 18,509 | 0.615 | 6,593 | 0.578 | -0.037 | 9,773 | 0.597 | 3,962 | 0.589 | -0.007 | -0.019 | 0.011 | 0.029 |
| | | (0.487) | | (0.494) | (0) | | (0.491) | | (0.492) | (0.421) | (0.002) | (0.278) | (0.094) |

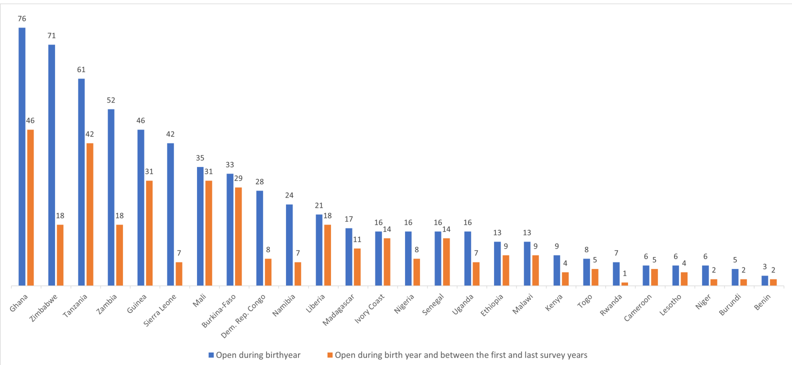
**Notes:** Standard errors and p-values in parentheses.

Table B.8: Average effects of mine opening on control variables

| | Household's characteristics | | | Mother characteristics | | |
|---|---|---|---|---|---|---|
| | % urban households | Has piped water | Has electricity | Age | Yers of education | % migrant |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Downstream×Open | 0.131*** | 0.0205 | -0.00100 | 0.0426 | -0.0121 | 0.0881*** |
| | [0.0422] | [0.0277] | [0.0200] | [0.162] | [0.144] | [0.0314] |
| Downstream | -0.0142 | -0.0411* | 0.00433 | -0.0835 | -0.119 | -0.0327 |
| | [0.0307] | [0.0239] | [0.0143] | [0.132] | [0.108] | [0.0272] |
| Open | -0.0455 | -0.00637 | -0.0104 | -0.136 | -0.126 | -0.0432* |
| | [0.0293] | [0.0209] | [0.0181] | [0.136] | [0.112] | [0.0233] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Nb open mines | Yes | Yes | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 61690 | 61690 | 61179 | 61690 | 61690 | 38834 |
| R2 | 0.608 | 0.489 | 0.547 | 0.681 | 0.463 | 0.185 |

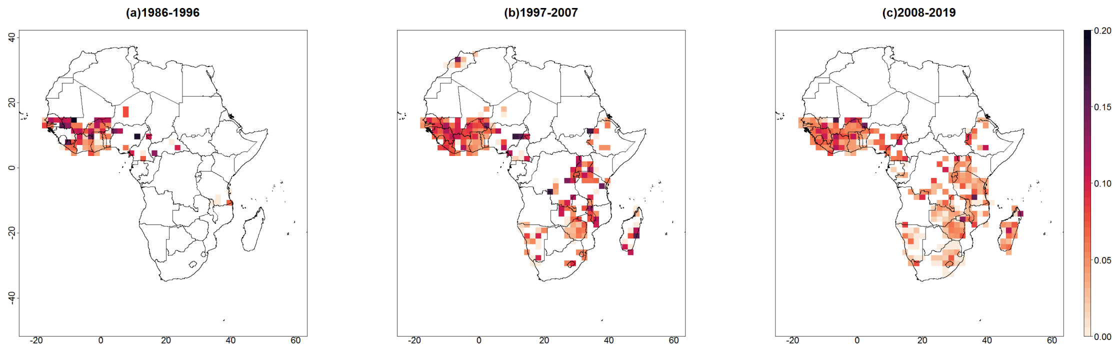
**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$. The variables Downstream and Opened are dummies which indicate whether the individual lives in a village downstream of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines.

Figure B.11: Number of open mines during the birth year and between first and last wave



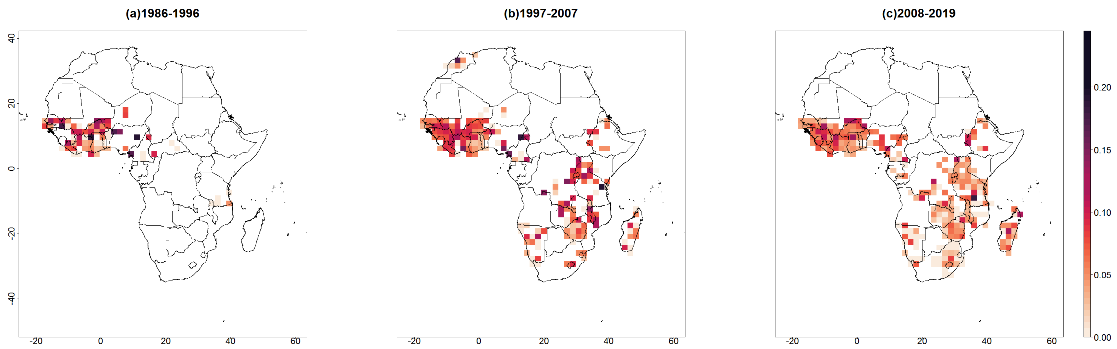

**Notes:** The figure represents the number of mines that were opened during the birth year of children located within our topographic treatment sample by country, and the number of mines that were opened during the birth year of children located within our topographic treatment sample and which opened between the first and last year of survey for each country.
*Sources:* Authors' elaboration on SNL and DHS data.

Figure B.12: Spatial variation of 12-month mortality rates per period - Restricted Sample



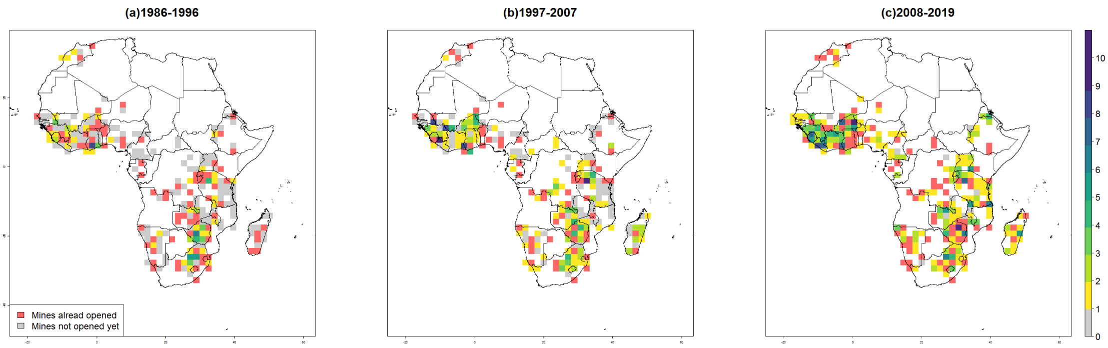

**Notes:** The figures represent the means of 12-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008, and (c) 2008-2019, for the sample of the main analysis. The mortality rates are estimated without the children that did not reach 12 months at the time of the survey.
*Sources:* Authors' elaboration on DHS data.

Figure B.13: Spatial variation of 24 months mortality rates per period - Restricted Sample




**Notes:** The figures represent the means of 24-month mortality rates averaged at the grid level over (a) 1986-1996, (b) 1997-2008, and (c) 2008-2019, for the sample from the main analysis. The mortality rates are estimated without the children that did not reach 24 months at the time of the survey.
*Sources:* Authors' elaboration on DHS data.

Figure B.14: Spatial variation of mine opening per period - Restricted Sample




**Notes:** The figures represent the number of mines that opened during the periods over the grid area (160 km on average). A red grid cell represents an area where no mine opened over the period, but where at least one mine has opened before the period. A grey cell represents an area where no mine opened over the period, but where at least one mine will open in the future.
*Sources:* Authors' elaboration on SNL data.

## B.4 Heterogeneity

Table B.9: Effects of industrial mining activity, across sub-regions

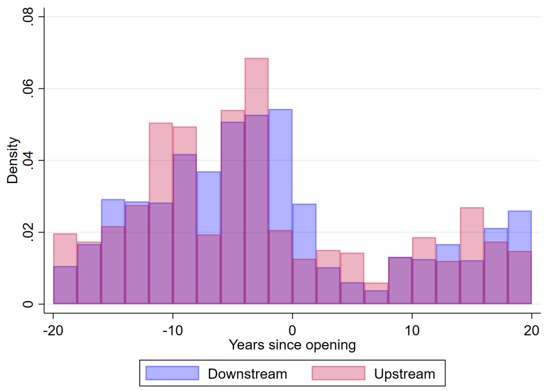
| | Mortality under 24 months | | |
|---|---|---|---|
| | Western Africa | Eastern Africa | Central and Southern Africa |
| | (1) | (2) | (3) |
| Downstream×Open | 0.0443** | 0.0292* | -0.0304 |
| | [0.0176] | [0.0154] | [0.0225] |
| Downstream | -0.0153 | -0.0369*** | 0.0538** |
| | [0.00946] | [0.0116] | [0.0227] |
| Open | 0.00523 | -0.0137 | 0.0120 |
| | [0.0133] | [0.0157] | [0.00857] |
| Controls | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes |
| N | 21,006 | 13,484 | 20,014 |
| R2 | 0.0521 | 0.0457 | 0.0238 |
| Outcome Mean | 0.0981 | 0.0712 | 0.0836 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

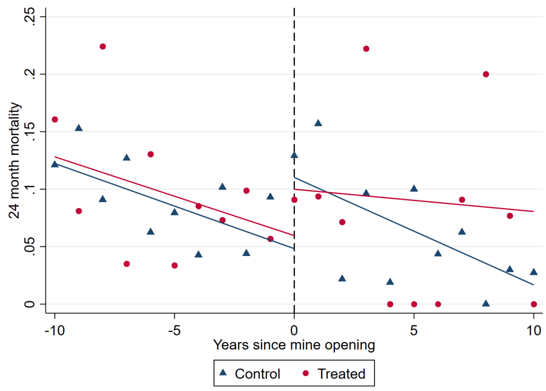
## B.5 Dynamic effects - pre trends and event study

In this section, we display the appendix figures associated with Section 2.8. It gives the same analysis as in the main Section restricted to the balanced sample. Figure B.15 plots the parallel trends, while Figure B.16 plots the event study for the balanced sample.

Figure B.15: Linear trends of 24 month mortality - Balanced Sample



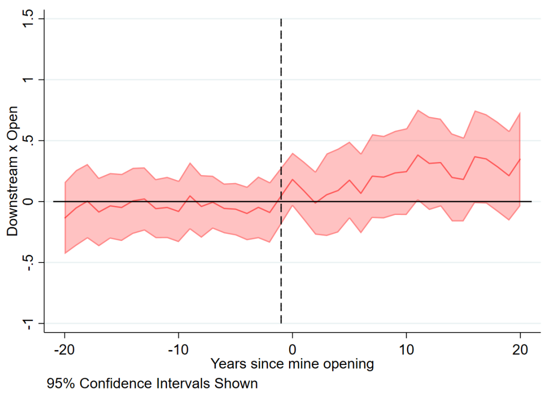

(a) Balanced sample



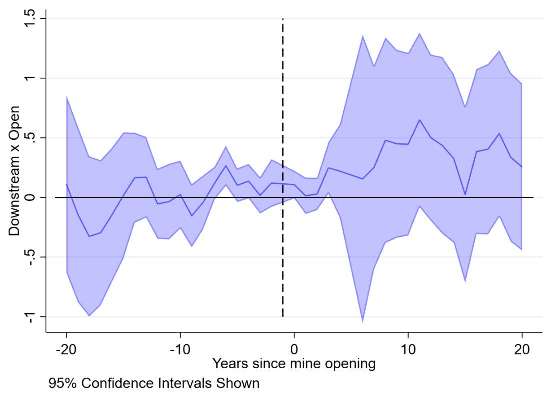

(b) Balanced sample

**Notes:** Figure (a) gives the distribution of the number of observations per opening year. Figure (b) plots the trends of the 24-month mortality rates according to the year of opening. The figures are made for the balanced sample and include neither control variables nor fixed effects.

Figure B.16: Event study - dynamic effect of mine opening on under 24 months mortality - Balanced Sample




(a) Balanced sample - upstream




(b) Balanced sample - downstream

**Notes:** Figure (a) plots the event study for the upstream villages, while Figure (b) plots the event study for the downstream villages for the balanced sample. Controls and fixed effects are the same as in the main analysis (column (4) Table 2.2)

### B.5.1 Sensitivity analysis

Table B.10 shows that our result is stable when controlling for a dummy indicating whether the mine opening year has been found by hand or was given directly in the SNL database (column (2)).

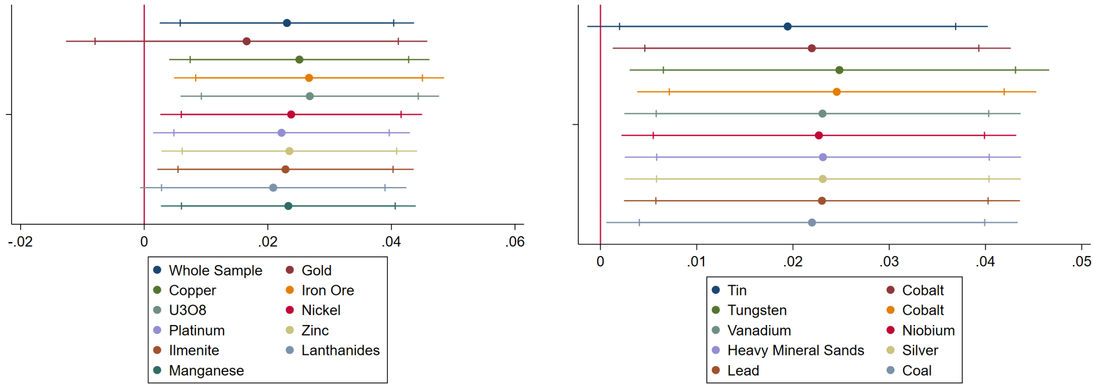
Table B.10: Effects of industrial mining opening, controlling for handwork.

| Outcome | Mortality under 24 months | | | |
|---|---|---|---|---|
| Specification | Main result (1) | Adding control (2) | SNL database (3) | Handwork (4) |
| Downstream × Open | 0.0218** | 0.0218** | 0.0222 | 0.0344** |
| | [0.0108] | [0.0108] | [0.0351] | [0.0138] |
| Dummy handwork | | 0.0254 | | |
| | | [0.0335] | | |
| Downstream | -0.0211*** | -0.0212*** | -0.0167 | -0.0316*** |
| | [0.00739] | [0.00739] | [0.0246] | [0.00844] |
| Open | -0.00496 | -0.00489 | -0.0194 | -0.00973 |
| | [0.0101] | [0.0101] | [0.0476] | [0.0126] |
| Controls | Yes | Yes | Yes | Yes |
| Birthmonth FE | Yes | Yes | Yes | Yes |
| Country-birthyear FE | Yes | Yes | Yes | Yes |
| Mine SB FE | Yes | Yes | Yes | Yes |
| Mine SB-birthyear trend | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes |
| N | 35,638 | 35,638 | 6,702 | 22,017 |
| R2 | 0.0511 | 0.0511 | 0.0615 | 0.0641 |
| Outcome mean | 0.0873 | 0.0873 | 0.0727 | 0.0954 |

**Notes:** Standard errors clustered at the DHS village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Columns (1) and (2) rely on the same sample and controls as Table 2.2 Column 2. Column (2) controls for the hand-work, while Columns (3) and (4) split the samples.

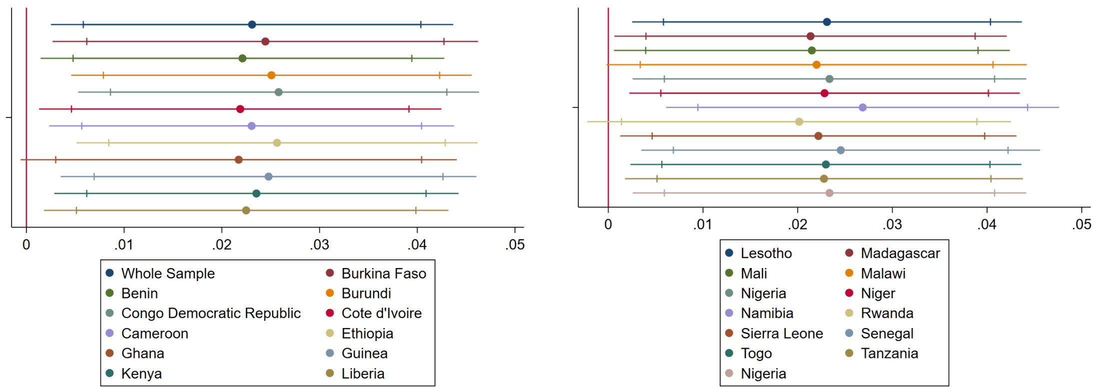
Figure B.17 displays the DiD estimators for different regression with restricted samples, meaning while dropping each metal one by one, using the sample for the 24-month mortality rates, and the heavy metals and coal mine sample. This suggests that our main results are not driven by a specific metal. Accordingly, Figure B.18, plots the DiD estimators while dropping countries one by one and show that our analysis is not driven by a particular country.

Figure B.17: Regression results when dropping commodities one by one




(a) (b)

Figure B.18: Regression results when dropping countries one by one




(a) (b)

*Sources:* Authors' elaboration on DHS and SNL data.

## B.6 Geographic Treatment

In this section, we propose to replicate the empirical strategy of Benshaul-Tolonen, 2018, who finds that a mine opening is associated with a 5.5 p.p decrease in the 12-month mortality. The identification strategy relies on a treatment based on proximity, comparing individuals living nearby to those living further from an industrial mine. In this estimation, geographical proximity is used as a proxy exposure to industrial mining activity, including both positive and negative externalities, such as exposure to mining pollution.

The identification strategy relies on a DiD strategy.It compares within each district, the infant mortality in areas within 10 km of a mine deposit (treatment group) to infant mortality in DHS clusters further away from a mine deposit (10-100km, control group), before and after the opening of the mine deposit. As the strategy is a two-way fixed-effects, including a district-fixed effect, the comparison is made within each district. The identification can be formally written as:

$$\begin{aligned} Death_{i,v,c,m,SB} =& \alpha_0 + \alpha_1 Opened_{birthyear,i,v} + \alpha_2 MineDeposit_{[0;10km]v} \\ & + \alpha_3 Opened_{birthyear,i,v} \times MineDeposit_{[0;10km]v} + \alpha_4 X_i \\ & \gamma_d + \gamma_{d-bthtrend} + \gamma_{c,birthyear} + \epsilon_v \end{aligned} \tag{B.1}$$

With $Death_{i,v,c,district}$ a dummy equals to one if child $i$ from DHS village $v$ (within district $d$) of country $c$, has reached the $n^{th}$ month and has died ($n$ being 12 for the 12-month mortality, 24 and so on). $Opened_{birthyear,i,v}$ is a dummy equal to 1 if at least one mine located within 10 km for the treatment group, or within 100 km for the control group, has opened before child $i$'s year of birth (this cohort comparison can be considered here as a source of tripe difference).$MineDeposit_{[0;10km]v}$ is a dummy of proximity (1 if village DHS $v$ is within 10 km of a mine deposit, 0 if it is within 10-100km), $X_i$ a vector of child/mother level controls (mother's age and age square, years of education, urban status). Finally, $\gamma_d$ is a district fixed effect, $\gamma_{d-bthtrend}$ a district birthyear linear trend, and $\gamma_{c,birthyear}$ a country-birthyear fixed effect. Please note that the matching of DHS clusters to mines relies on the same strategy as in Benshaul-Tolonen, 2018, and assigns a DHS cluster to the closest mine (without consideration of its opening status). Thanks to this pairing, if a DHS cluster is both in the treatment and control groups of two different mines (i.e within 10km of mine A and within 10-100km from Mine B), we assign it mechanically to the treatment group (so linked to mine A). This creates bias explained in Section 2.2.3, which explains the choice for a district fixed effects and reduces the noise linked to DHS random displacements.

Firstly, we give our estimators from the exact replication of Benshaul-Tolonen, 2018 results, using our own calculation, and find similar impacts (Tables B.11 and B.12). Second, we propose the replication of the results using our extended sample, including more countries, DHS waves, types of mines, and mines hand-checked, and show the results from Benshaul-Tolonen, 2018 are mainly determined by the choice of countries.

### B.6.1 Exact replication of Benshaul-Tolonen, 2018

The geographic treatment proxies exposure to mining activity using the distance to the site and follows partly the analysis from Benshaul-Tolonen, 2018, and finds contradictory impacts on infant mortality. To understand better how our results can be compared to the literature, we propose in this section a replication exercise of the main result from Benshaul-Tolonen, 2018[2].

For this replication analysis, we used the same mines and DHS survey rounds as Benshaul-Tolonen, 2018. Please note that we have few differences in terms of the whole sample, as Benshaul-Tolonen, 2018 counts 37,365 children *vs* 41,902 for us, that might be explained by the way we calculated the 100km buffer distance [4]. A main difference between our paper and Benshaul-Tolonen, 2018 is the independent variable, as we use as a shock the opening of the industrial mine whereas Benshaul-Tolonen, 2018 uses the activity status based on production data given by the SNL product. This accounts for interim years, between the opening and final closing of the mine, where the production has been on hold. In this section, we replicate this exact same variable.

Table B.11 displays the replication of the main results from Benshaul-Tolonen, 2018 Table 2. We find that a mine opening within 10 kilometers is associated with a 4.7

[2]Please note that a first difference between the two analyses is the sample, as Benshaul-Tolonen, 2018 uses 43 gold mines that match with 31 DHS surveys from nine countries (Burkina Faso, Cote D'Ivoire, Ethiopia, Ghana, Guinea, Mali, Senegal, Tanzania, and DRC [3]). However, when pairing the DHS cluster to the same industrial mining sites from Benshaul-Tolonen, 2018, no DHS from DRC remained. In the end, the analysis is only on the 8 first countries, in accordance with Figure A6 from Appendix of Benshaul-Tolonen, 2018), for a whole sample of 1-year-old children of 48,151.

[4]In the replication codes of Benshaul-Tolonen, 2018, one can observe that the distance has been determined using the Stata command *nearstat* [...] dband(0,25) which relies on different projections (not specified) as ours from *R* libraries, explaining the small sample differences

Table B.11: Replication Benshaul-Tolonen, 2018 Main Results

| Dependent variable | Infant mortality first 12 months | | | |
|---|---|---|---|---|
| Sample : | Children<br>(1) | Children drop spillover<br>(2) | Boys<br>(3) | Girls<br>(4) |
| Industrial × mine deposit (at birth) | -0.0472** | -0.0474* | -0.0289 | -0.0781*** |
| | [0.0230] | [0.0260] | [0.0320] | [0.0301] |
| Mine deposit [0;10km] | 0.0392** | 0.0546*** | 0.0517** | 0.0561** |
| | [0.0169] | [0.0195] | [0.0229] | [0.0231] |
| Mother's age | -0.0145*** | -0.0154*** | -0.0155*** | -0.0152*** |
| | [0.00190] | [0.00210] | [0.00274] | [0.00297] |
| Mothers's age × Mother's age | 0.000222*** | 0.000236*** | 0.000223*** | 0.000245*** |
| | [0.0000302] | [0.0000335] | [0.0000435] | [0.0000475] |
| Years edu. | -0.00214*** | -0.00230*** | -0.00272*** | -0.00184** |
| | [0.000489] | [0.000547] | [0.000827] | [0.000760] |
| Urban$_h h$ | -0.0125*** | -0.0120** | -0.00710 | -0.0183*** |
| | [0.00428] | [0.00480] | [0.00687] | [0.00659] |
| Birth-month FE | Yes | Yes | Yes | Yes |
| Country birth year FE | Yes | Yes | Yes | Yes |
| District FE | Yes | Yes | Yes | Yes |
| District BirthYear trend | Yes | Yes | Yes | Yes |
| Drop10-30 km away | No | Yes | Yes | Yes |
| Drop investment phase | No | Yes | Yes | Yes |
| Mean of outcome | 0.102 | 0.104 | 0.110 | 0.099 |
| Mean(treatment, pre-treatment) | 0.154 | 0.163 | 0.173 | 0.153 |
| Observations | 41902 | 34228 | 17534 | 16694 |

**Notes:** $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from Benshaul-Tolonen, 2018 and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth. Columns 2-5 drop the two years preceding th opening year, defined as investment phase in Benshaul-Tolonen, 2018 and the individuals living within 10-30km of the closest industrial mine. Mean (treatment, pre-treatment) is the sample for the treatment group before the mine were active.
dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

percentage point decrease in infant mortality rates, while Benshaul-Tolonen, 2018 found 5.5 p.p. Our results is slightly less significant than from Benshaul-Tolonen, 2018, and we identify a different impact according to gender, with a significant reduction of girl mortality rates of 7 p.p *vs* a non-significant reduction for boys, which differs from the previous study. To follow Benshaul-Tolonen, 2018 example, we excluded in Columns 2-5 from Table B.11 individuals born within 10-30 kilometers of the closest industrial mining site and those born the two years before the opening of a mine, which is a proxy for the investment phase according to Benshaul-Tolonen, 2018.

Please note that in accordance with the descriptive statistics from Benshaul-Tolonen, 2018 we have in the sample a very high mean of 12-month mortality rates (from 10 to 17 % according to the groups). These are relatively high numbers, that do not match with World Bank data. This is because Benshaul-Tolonen, 2018 drops all the individuals who are still alive but did not reach the age of 12 months yet to measure the mortality, in order to avoid growing mechanically the mortality rates of these cohorts [5]. For replication purposes, we propose to keep this variable and correct this in Table B.12, where we observe average mortality rates around 7%. Figure B.19 replicates the Figure A6 from Benshaul-Tolonen, 2018, which shows the coefficient estimates of the main regression for *industrial* $\times$ *mine deposit* on infant mortality, each regression excluding the sample from one country as indicated by the country name. The Figure B.19 shows that results are highly sensitive to the presence of Mali, Senegal, and Ghana in the sample (whereas they do not consist for the majority of the sample (5847, 1098 and 5595 respectively).

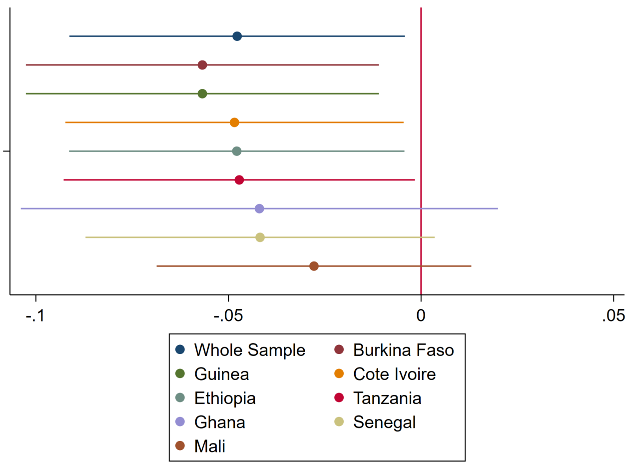
[5]We can read in the codes that if the living individuals were dropped, the children that died before their 12 months from these specific cohorts were not dropped: mechanically, the mortality rates for all the years preceding the survey rounds are 100 %, which explain the high mean of outcomes.

Table B.12: Replication Benshaul-Tolonen, 2018 Main Results

| Dependent variable | Infant mortality first 12 months corrected | | | |
|---|---|---|---|---|
| Sample : | Children<br>(1) | Children drop spillover<br>(2) | Boys<br>(3) | Girls<br>(4) |
| Industrial × mine deposit (at birth) | -0.0494** | -0.0471* | -0.0439 | -0.0631** |
| | [0.0229] | [0.0244] | [0.0317] | [0.0298] |
| Mine deposit [0;10km] | 0.0394** | 0.0587*** | 0.0682*** | 0.0513** |
| | [0.0179] | [0.0198] | [0.0255] | [0.0235] |
| Mother's age | -0.0118*** | -0.0123*** | -0.0120*** | -0.0124*** |
| | [0.00175] | [0.00196] | [0.00256] | [0.00283] |
| Mothers's age × Mother's age | 0.000182*** | 0.000189*** | 0.000172*** | 0.000203*** |
| | [0.0000279] | [0.0000312] | [0.0000405] | [0.0000452] |
| Years edu. | -0.00143*** | -0.00152*** | -0.00204*** | -0.000803 |
| | [0.000455] | [0.000510] | [0.000772] | [0.000715] |
| Urban$_h h$ | -0.0106*** | -0.0113*** | -0.00501 | -0.0196*** |
| | [0.00384] | [0.00436] | [0.00661] | [0.00600] |
| Birth-month FE | Yes | Yes | Yes | Yes |
| Country birth year FE | Yes | Yes | Yes | Yes |
| District FE | Yes | Yes | Yes | Yes |
| District BirthYear trend | Yes | Yes | Yes | Yes |
| Drop10-30 km away | No | Yes | Yes | Yes |
| Drop investment phase | No | Yes | Yes | Yes |
| Mean of outcome | 0.079 | 0.080 | 0.083 | 0.077 |
| Mean(treatment, pre-treatment) | 0.109 | 0.118 | 0.120 | 0.115 |
| Observations | 40386 | 32873 | 16823 | 16050 |

**Notes:** $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from Benshaul-Tolonen, 2018 and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth. Columns 2-5 drop the two years preceding th opening year, defined as investment phase in Benshaul-Tolonen, 2018 and the individuals living within 10-30km of the closest industrial mine. Mean (treatment, pre-treatment) is the sample for the treatment group before the mine were active.
dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

Figure B.19: Regression results when dropping one country at a time




*Sources:* Authors' elaboration on DHS and SNL data.

#### B.6.1.1 Replication using an extended sample

Table B.13 and Table B.14 display the results, replicating Benshaul-Tolonen, 2018 estimation strategy, with our overall sample of mines and DHS surveys. Table B.13 focuses on the 12-month mortality rates and shows that we find a significant reduction of infant mortality by 0.8 p.p only when controlling for migrants (column (2)). Columns (1) and (2) display the results for the whole sample, while columns (3) and (4) while dropping the spillovers effects (areas between [10-30]km and the two years before the mine opening, which represents the investment phase in Benshaul-Tolonen, 2018). Columns (5) and (6) replicate the analysis for the male sample while columns (7) and (8) for the girls.

Table B.14 displays the result for the 12-month mortality rates (Columns (1)-(4)) and 24 months mortality rates (Columns (5)-(8)) and compares the estimators when not including the migrant control variable (Columns (1), (3) (5) and (7)), and when including it (Columns ((2),(4),(6) and (8)). We also display the estimators for the restricted sample of rural areas (Columns (3),(4), (7), and (8)). Again, we observe a significant reduction of 12 months mortality rates in Column (2), i.e for the overall sample while controlling for migrants, and find no results otherwise. This absence of results suggests that using proximity as a proxy for exposure to mining activity averages contradictory effects, including both positive and negative externalities, and shows the importance of our main estimation strategy which relies on topographic position.

Figure B.20 plots the linear trends of the 12 and 24 months mortality rates for the geographic treatment, including our overall mine and DHS sample. We see that the linear trends assumption seems to be validated for the 24-month mortality, but not for the 12-month mortality rates.

Table B.13: Geographic Treatment

| | Infant mortality first 12 months | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All | | Drop spillover | | Boys | | Girls | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Indus. × deposit | -0.00259 | -0.00823** | -0.00189 | -0.00575 | 0.00250 | -0.00302 | -0.00513 | -0.00807 |
| | [0.00329] | [0.00418] | [0.00407] | [0.00537] | [0.00570] | [0.00764] | [0.00522] | [0.00674] |
| Deposit | 0.00130 | 0.00374 | 0.00103 | -0.000128 | 0.00632 | 0.0113 | -0.00366 | -0.0109* |
| | [0.00252] | [0.00317] | [0.00392] | [0.00500] | [0.00546] | [0.00708] | [0.00513] | [0.00628] |
| Birth order | 0.00389*** | 0.00315*** | 0.00360*** | 0.00320*** | 0.00349*** | 0.00304*** | 0.00382*** | 0.00356*** |
| | [0.000345] | [0.000428] | [0.000423] | [0.000518] | [0.000606] | [0.000742] | [0.000549] | [0.000671] |
| Mother's age | -0.0105*** | -0.0107*** | -0.0102*** | -0.0110*** | -0.0116*** | -0.0128*** | -0.00884*** | -0.00924*** |
| | [0.000541] | [0.000668] | [0.000669] | [0.000824] | [0.000953] | [0.00119] | [0.000903] | [0.00111] |
| agesquare | 0.000147*** | 0.000151*** | 0.000142*** | 0.000156*** | 0.000163*** | 0.000183*** | 0.000121*** | 0.000127*** |
| | [0.00000853] | [0.0000106] | [0.0000106] | [0.0000131] | [0.0000150] | [0.0000187] | [0.0000142] | [0.0000175] |
| Years edu. | -0.000877*** | -0.00103*** | -0.000874*** | -0.00101*** | -0.000881*** | -0.00103*** | -0.000873*** | -0.000968*** |
| | [0.000135] | [0.000167] | [0.000164] | [0.000200] | [0.000238] | [0.000290] | [0.000216] | [0.000265] |
| Urban | -0.00610*** | -0.00725*** | -0.00708*** | -0.00906*** | -0.00825*** | -0.0111*** | -0.00563** | -0.00622** |
| | [0.00135] | [0.00172] | [0.00169] | [0.00214] | [0.00235] | [0.00297] | [0.00227] | [0.00289] |
| migrant | | 0.00543*** | | 0.00509*** | | 0.00255 | | 0.00754*** |
| | | [0.00120] | | [0.00145] | | [0.00208] | | [0.00196] |
| Constant | 0.229*** | 0.232*** | 0.226*** | 0.240*** | 0.251*** | 0.273*** | 0.201*** | 0.206*** |
| | [0.00826] | [0.0101] | [0.0103] | [0.0126] | [0.0146] | [0.0181] | [0.0138] | [0.0169] |
| Birth-month FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ctry-bthyr FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Dist-bthyr trend | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Drop10-30 km | No | No | Yes | Yes | No | No | No | No |
| Drop t-2 | No | No | Yes | Yes | No | No | No | No |
| N | 359219 | 243645 | 236573 | 165202 | 119860 | 83570 | 116696 | 81601 |

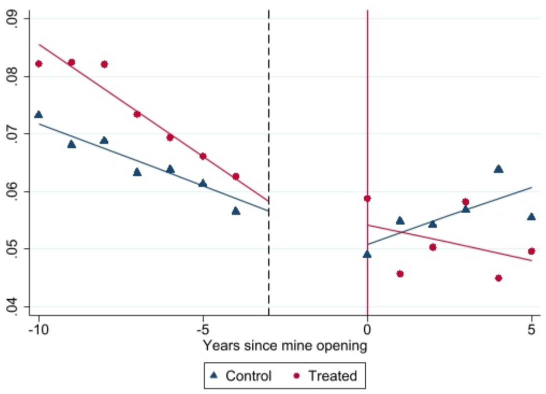
**Notes:** $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$. Standard errors clustered a DHS cluster level. The variables Mine deposit [0;10km] and Industrial × mine deposit (at birth) are a replication from Benshaul-Tolonen, 2018 and indicate whether the child is born within 10km of at least one industrial mining site and whether this site was active at the time of the birth. All regressions control for mother's age, age square, mother's education and whether the household is urban, for district, birth month and country-birth year. The main outcome is infant mortality in the 12 months since birth.
dummies which indicate whether the individual lives in a village within at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to the closest opened mining site, so that each individual appears only once in the regression. Other variables are control variables.

Table B.14: Effects of industrial mining activity on under 12, 24 mortality - Geographic Treatment - All Households

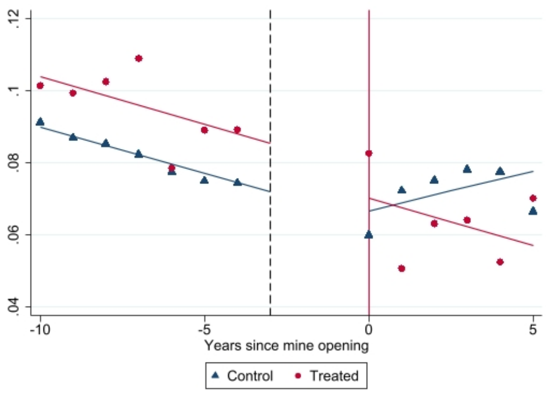
| | Death <12m | | | | Death < 24m | | | |
|---|---|---|---|---|---|---|---|---|
| | All | All | Rural | Rural | All | All | Rural | Rural |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| indus.×deposit | -0.00259 | -0.00823** | -0.00259 | -0.00627 | 0.000248 | -0.00264 | 0.000248 | -0.00248 |
| | [0.00329] | [0.00418] | [0.00329] | [0.00509] | [0.00431] | [0.00535] | [0.00431] | [0.00657] |
| Deposit | 0.00130 | 0.00374 | 0.00130 | 0.00313 | 0.000627 | 0.00121 | 0.000627 | 0.000859 |
| | [0.00252] | [0.00317] | [0.00252] | [0.00368] | [0.00321] | [0.00411] | [0.00321] | [0.00477] |
| Indus. | 0.00131 | 0.00222 | 0.00131 | 0.00340 | 0.00116 | 0.00122 | 0.00116 | 0.00190 |
| | [0.00155] | [0.00200] | [0.00155] | [0.00230] | [0.00201] | [0.00259] | [0.00201] | [0.00297] |
| Birth order | 0.00389*** | 0.00315*** | 0.00389*** | 0.00353*** | 0.00512*** | 0.00401*** | 0.00512*** | 0.00447*** |
| | [0.000345] | [0.000428] | [0.000345] | [0.000500] | [0.000440] | [0.000549] | [0.000440] | [0.000642] |
| Mother's age | -0.0105*** | -0.0107*** | -0.0105*** | -0.0116*** | -0.0115*** | -0.0124*** | -0.0115*** | -0.0140*** |
| | [0.000541] | [0.000668] | [0.000541] | [0.000787] | [0.000704] | [0.000873] | [0.000704] | [0.00103] |
| Age square | 0.000147*** | 0.000151*** | 0.000147*** | 0.000161*** | 0.000151*** | 0.000167*** | 0.000151*** | 0.000187*** |
| | [0.00000853] | [0.0000106] | [0.00000853] | [0.0000122] | [0.0000110] | [0.0000136] | [0.0000110] | [0.0000159] |
| Years edu. | -0.000877*** | -0.00103*** | -0.000877*** | -0.000792*** | -0.00145*** | -0.00157*** | -0.00145*** | -0.00132*** |
| | [0.000135] | [0.000167] | [0.000135] | [0.000219] | [0.000173] | [0.000215] | [0.000173] | [0.000283] |
| Urban | -0.00610*** | -0.00725*** | -0.00610*** | | -0.00940*** | -0.00995*** | -0.00940*** | |
| | [0.00135] | [0.00172] | [0.00135] | | [0.00175] | [0.00222] | [0.00175] | |
| migrant | | 0.00543*** | | 0.00514*** | | 0.00727*** | | 0.00630*** |
| | | [0.00120] | | [0.00144] | | [0.00155] | | [0.00186] |
| Constant | 0.229*** | 0.232*** | 0.229*** | 0.247*** | 0.273*** | 0.286*** | 0.273*** | 0.315*** |
| | [0.00826] | [0.0101] | [0.00826] | [0.0120] | [0.0109] | [0.0134] | [0.0109] | [0.0159] |
| Birthmonth FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cty-Bthyr FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine Bthyr trend | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 359,219 | 243,645 | 359,219 | 179,155 | 265,735 | 179,729 | 265,735 | 132,398 |
| R2 | 0.0195 | 0.0235 | 0.0195 | 0.0281 | 0.0289 | 0.0337 | 0.0289 | 0.0393 |
| Mean | 0.0630 | 0.0653 | 0.0630 | 0.0688 | 0.0816 | 0.0851 | 0.0816 | 0.0903 |

**Notes:**Standard errors clustered at the village level, $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. The variables Proximity and Opened are dummies which indicate whether the individual lives in a DHS village within 10 km of at least one mining site, and whether the site opened before the birth year of the child. Each village DHS is paired to only one mining site, so that each individual appears only once in the regression. Other variables are control variables. The sample focuses on heavy metal mines.

Figure B.20: Linear Trends dropping investment phase - Geographic Treatment




(a) Infant mortality Rate 12 months




(b) Infant mortality Rate 24 months

*Sources:* Authors' elaboration on DHS and SNL data.

# Appendix C

# Appendix to Chapter 3 : Impacts of repetitive droughts and the key role of experience : evidence from Nigeria

### C.0.1 Descriptive Statistics

Table C.1 give the composition of the Original Sample (columns (1) and (2)), and the final panel sample for the three first GHS rounds, including households that have moved and been tracked by the LSMS-ISA teams (column (5)), per Nigerian geopolitical zones. Table C.2 gives insights of the attrition rates per waves and geopolitical zones of Nigeria.

Table C.3 displays the different questions used in the survey to build the FIES (or scale score) food security variable. Table C.4 displays the descriptive statistics of the main variables of interest of the paper. Each variables is driven from a panel sample. Households Characteristics outcomes statistics are given for the whole panel sample, Agricultural production and practices are given for the panel sample for which each household cultivated at least one plot for each wave (or answered), while Food security outcomes panel samples for which we were able to build the indicator for each waves per households. The number observation (column (6)) display the number of households from the sample, the real number of observations being three times this number (as for three waves). The descriptive statistics show discrepancies between GPS-based and self-reported plot areas, as mentionned in the literature Yacoubou Djima and Kilic, 2021, both in terms of total land holding per household and average plot area per household. Please note that the GPS measures are asked for the entire land holdings of the households, while Self-Reported (SR) measures only for plots cultivated by the household at the time of the survey. The descriptive statistics show that SR measure of plot areas overestimate the land areas in comparison to GPS measures, which seem to be mainly explained by outliers and lower/higher end of the plot distribution. Indeed, while means highly differ, median seem to be more comparable. As Self-Reported total crop production can be also mismeasured, we observe that the median highly differ from the mean. Self-Reported Yields are thus twofold noisy, and the variables displayed in Table C.4 have been treated for outliers (winsorized by the median at 10%).

Table C.1: Descriptive Statistics of the 6 years panel per geopolitical Zones

| Zones | Original Sample 2010-2011 | | Panel Sample Original Location | | Panel Sample Moved/Tracked | Total | % Original Sample |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | EAs | HHs | EAs | HHs | HHs | HHs | % |
| **North Central** | | | | | | | |
| Urban | 22 | 217 | 22 | 200 | 1 | 201 | 7.37 |
| Rural | 57 | 577 | 57 | 548 | 11 | 559 | 3.12 |
| Total | 78 | 794 | 78 | 748 | 12 | 760 | 4.28 |
| **North East** | | | | | | | |
| Urban | 14 | 138 | 9 | 79 | 6 | 85 | 38.41 |
| Rural | 63 | 659 | 52 | 521 | 6 | 527 | 20.03 |
| Total | 77 | 797 | 61 | 600 | 12 | 612 | 23.21 |
| **North West** | | | | | | | |
| Urban | 17 | 170 | 16 | 150 | 2 | 152 | 10.59 |
| Rural | 69 | 728 | 69 | 704 | 10 | 714 | 1.92 |
| Total | 85 | 898 | 84 | 854 | 12 | 866 | 3.56 |
| **South East** | | | | | | | |
| Urban | 21 | 204 | 21 | 172 | 4 | 176 | 13.73 |
| Rural | 57 | 590 | 57 | 542 | 10 | 552 | 6.44 |
| Total | 76 | 794 | 76 | 714 | 14 | 728 | 8.31 |
| **South South** | | | | | | | |
| Urban | 23 | 229 | 23 | 186 | 15 | 201 | 12.23 |
| Rural | 55 | 540 | 55 | 460 | 32 | 492 | 8.89 |
| Total | 78 | 769 | 78 | 646 | 47 | 693 | 9.88 |
| **South West** | | | | | | | |
| Urban | 63 | 612 | 62 | 417 | 71 | 488 | 20.26 |
| Rural | 26 | 253 | 26 | 183 | 21 | 204 | 19.37 |
| Total | 87 | 865 | 86 | 600 | 92 | 692 | 20 |
| **Nigeria** | | | | | | | |
| Urban | 160 | 1570 | 153 | 1204 | 99 | 1303 | 17.01 |
| Rural | 327 | 3347 | 316 | 2958 | 90 | 3048 | 8.93 |
| Total | 481 | 4917 | 463 | 4162 | 189 | 4351 | 11.51 |

Table C.2: Attrition rates for wave 2 and 3 per geopolitical zones

| Zones | Original Sample | Location | Moved | Total | Attrition (%) | Location | Moved | Total | Attrition(%) |
|---|---|---|---|---|---|---|---|---|---|
| | Wave1 | Wave2 | Wave2 | Wave2 | Wave2 | Wave3 | Wave3 | Wave3 | Wave3 |
| **North Central** | | | | | | | | | |
| Urban | 217 | 206 | 5 | 211 | 2.76 | 201 | 0 | 201 | 7.37 |
| Rural | 577 | 576 | 7 | 583 | -1.04 | 550 | 4 | 554 | 3.99 |
| Total | 794 | 782 | 12 | 794 | 0 | 751 | 4 | 755 | 4.91 |
| **North East** | | | | | | | | | |
| Urban | 138 | 123 | 2 | 125 | 9.42 | 96 | 0 | 96 | 30.43 |
| Rural | 659 | 643 | 5 | 648 | 1.67 | 523 | 5 | 528 | 19.88 |
| Total | 797 | 766 | 7 | 773 | 3.01 | 619 | 5 | 624 | 21.71 |
| **North West** | | | | | | | | | |
| Urban | 170 | 157 | 0 | 157 | 7.65 | 158 | 2 | 160 | 5.88 |
| Rural | 728 | 714 | 8 | 722 | 0.82 | 705 | 0 | 705 | 3.16 |
| Total | 898 | 871 | 8 | 879 | 2.12 | 863 | 2 | 865 | 3.67 |
| **South East** | | | | | | | | | |
| Urban | 204 | 194 | 8 | 202 | 0.98 | 172 | 1 | 173 | 15.2 |
| Rural | 590 | 572 | 4 | 576 | 2.37 | 545 | 2 | 547 | 7.29 |
| Total | 794 | 766 | 12 | 778 | 2.02 | 717 | 3 | 720 | 9.32 |
| **South South** | | | | | | | | | |
| Urban | 229 | 206 | 18 | 224 | 2.18 | 190 | 6 | 196 | 14.41 |
| Rural | 540 | 510 | 22 | 532 | 1.48 | 461 | 7 | 468 | 13.33 |
| Total | 769 | 716 | 40 | 756 | 1.69 | 651 | 13 | 664 | 13.65 |
| **South West** | | | | | | | | | |
| Urban | 612 | 504 | 67 | 571 | 6.7 | 431 | 13 | 444 | 27.45 |
| Rural | 253 | 219 | 18 | 237 | 6.32 | 188 | 8 | 196 | 22.53 |
| Total | 865 | 723 | 85 | 808 | 6.59 | 619 | 21 | 640 | 26.01 |
| **Nigeria** | | | | | | | | | |
| Urban | 1570 | 1390 | 100 | 1490 | 5.1 | 1248 | 22 | 1270 | 19.11 |
| Rural | 3347 | 3234 | 64 | 3298 | 1.46 | 2972 | 26 | 2998 | 10.43 |
| Total | 4917 | 4624 | 164 | 4788 | 2.62 | 4220 | 48 | 4268 | 13.2 |

Table C.3: FIES descriptive statistics - List of survey questions

| FIES - GHS Questions |
|---|
| *In the past seven days, how many days you or someone in your household had to :* |
| (1) Rely on less preferred food? |
| (2) Limit the variety of food eaten? |
| (3) Limit the portion size at meal-times? |
| (4) Reduce number of meals eaten in a day? |
| (5) Restrict consumption by adults in order for small children to eat? |
| (6) Borrow food, or rely on help from a friend or relative? |
| (7) Have no food of any kind? |
| (8) Go at sleep hungry because there is not enough food? |
| (9) Go a whole day and night without eating? |

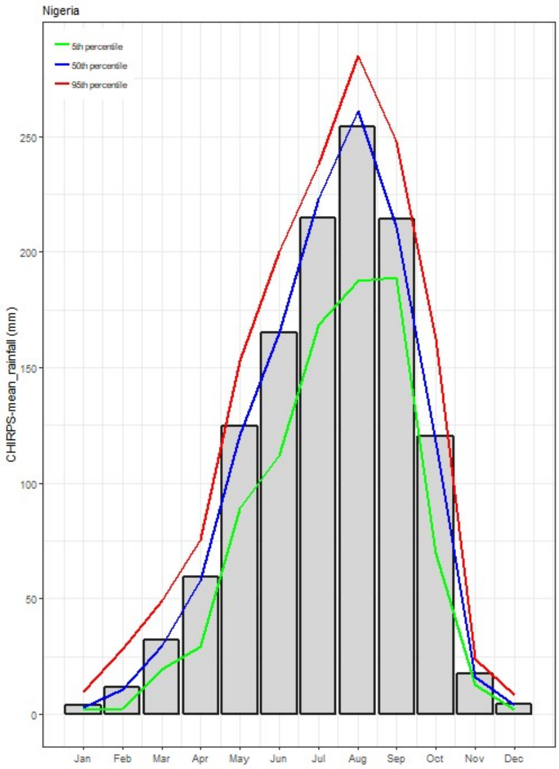
Table C.4: Descriptive Statistics of main variables

| | Mean (1) | SD (2) | Med (3) | Min (4) | Max (5) | Obs. (6) |
|---|---|---|---|---|---|---|
| **Household Characteristics** | | | | | | |
| Household size | | | | | | |
| Number of adults | 3.38 | 1.89 | 3 | 1 | 26 | 4041 |
| Head is female | 0.16 | 0.37 | 0 | 0 | 1 | 4041 |
| Age of head | 51.7 | 14.7 | 50 | 22 | 87 | 4041 |
| **Agricultural Production** | | | | | | |
| GPS measured total land holding (ha) | 1 | 1.52 | 0.53 | 0 | 26.5 | 1758 |
| SR total cultivated land holding (ha) | 8.4 | 219 | 0.98 | 0 | $10^4$ | 1758 |
| GPS measured plot area (ha) - average per hh | 0.56 | 0.83 | 0.35 | $10^{-4}$ | 26.5 | 1758 |
| SR cultivated plot area (ha)- average per hh | 4.44 | 103 | 0.58 | 0 | 4800 | 1758 |
| SR total crop production (kg) | 4435 | 25591 | 1751 | 0 | $10^6$ | 1758 |
| SR Yields (kg/ha) | 2369 | 2617 | 1436 | 202 | 14880 | 1758 |
| SR Yields main crops (kg/ha) | 2491 | 2788 | 1435 | 210 | 15104 | 1758 |
| SR Yields maize (kg/ha) | 1670 | 1805 | 1000 | 148 | 11752 | 777 |
| **Agricultural Practices** | | | | | | |
| Number of cultivated plots per hh | 2.08 | 1.18 | 2 | 1 | 11 | 1758 |
| Number of cultivated crops per plots - average per hh | 2.37 | 1.16 | 2 | 1 | 12 | 1758 |
| **Food Security** | | | | | | |
| FIES | -3.21 | 6.21 | 0.51 | -57 | 0.7 | 3907 |
| Food insufficiency | 0.20 | 0.40 | 0 | 0 | 1 | 4055 |
| HDDS | 8.22 | 2.06 | 8 | 1 | 12 | 4041 |

*Notes*: Descriptive statistics are displayed for the panel sample including the three first waves of the GHS. The mean and other mathematics are given for the three waves, the number of observations are the number of households (the real number of observation being three times the one given). Household characteristics outcomes are given for the whole sample. Agricultural production and practices are displayed for the panel of households that cultivated at least one plot in each waves (the one used in the main regression analysis), while the Food security outcomes for the panel sample for which we were able to build the indicator for each wave. Please note that SR Yields have been treated to correct the outliers, winsorized by the median.

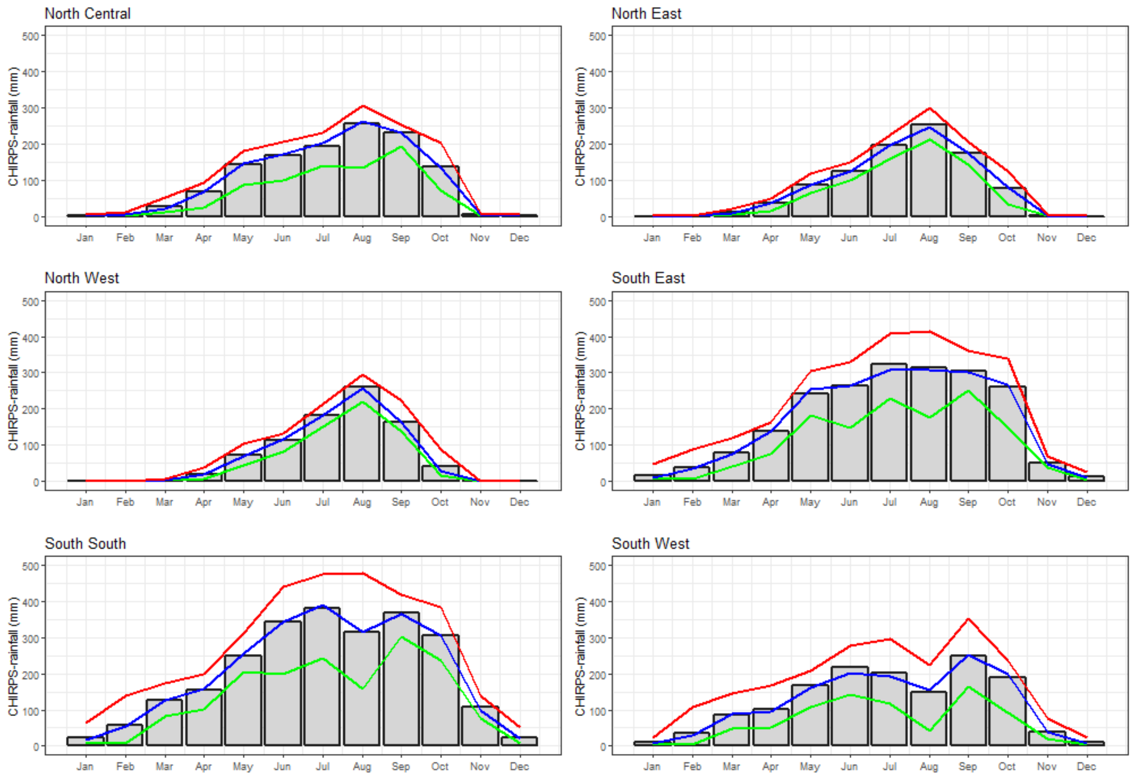
#### C.0.1.1 Figures

Figure C.1: Monthly precipitation of long-term average



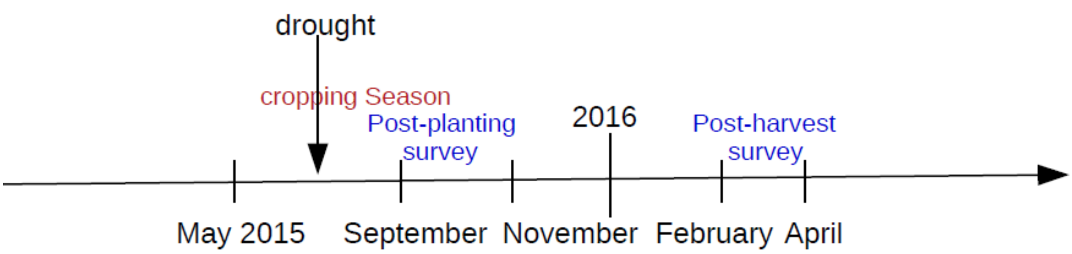

**Notes:** The Figure represent the long-term average (1981-2019) of the monthly precipitation over Nigeria. Red lines draw the 95th percentile of the long-term rainfall distribution, while the blue line the 50th percentile and the green line the 5th percentile.
*Sources:* Author's elaboration on CHIRPS

Figure C.2: Monthly precipitation of long-term average per geopolitical zones



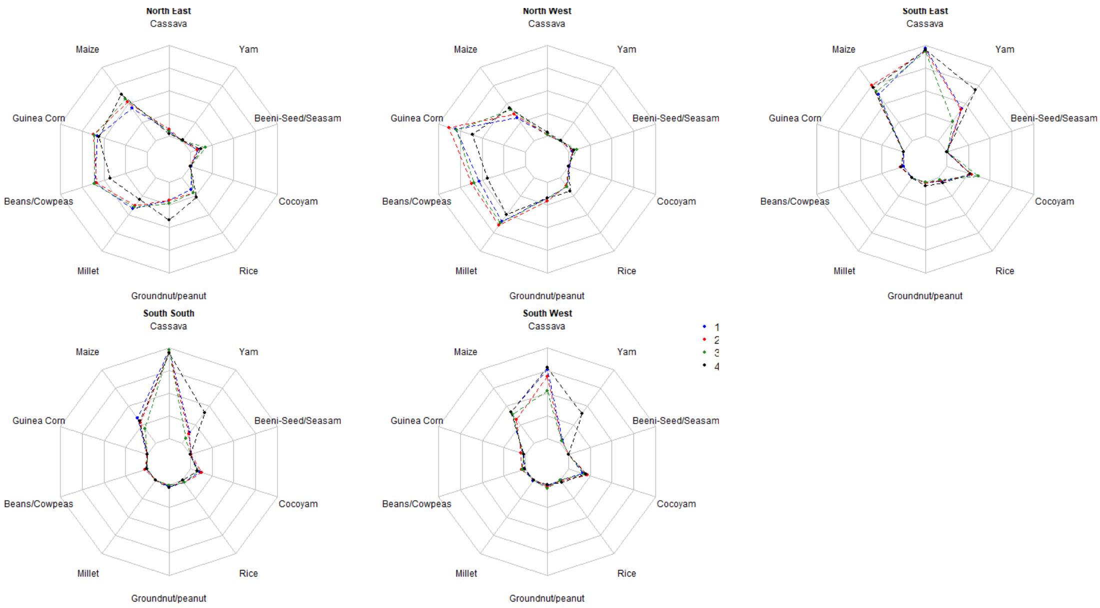

**Notes:** The Figure represent the long-term average (1981-2019) of the monthly precipitation over the six geopolitical zones of Nigeria. Red lines draw the 95th percentile of the long-term rainfall distribution, while the blue line the 50th percentile and the green line the 5th percentile.
*Sources:* Author's elaboration on CHIRPS

Figure C.3: Timeline of cropping season and survey rounds




**Notes:** The Figure gives the timeline of the cropping season and the post-planting and post-harvest survey rounds.
*Sources:* Author's elaboration on GHS data

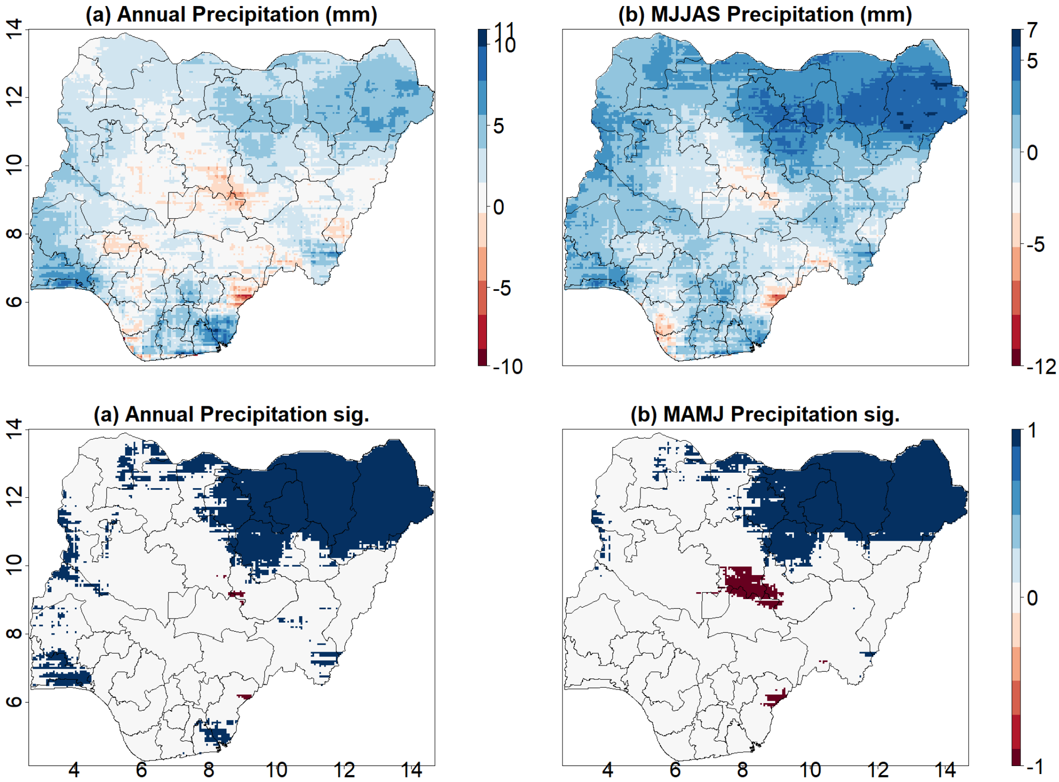
Figure C.4: Main cultivated crops per geopolitical zones per GHS waves




**Notes:** The Figure plots the distribution of main cultivated crops per geopolitical zones of Nigeria, for each GHS waves (wave 1 in blue, wave 2 in red, wave 3 in green and wave 4 in dark).
*Sources:* Author's elaboration on GHS data

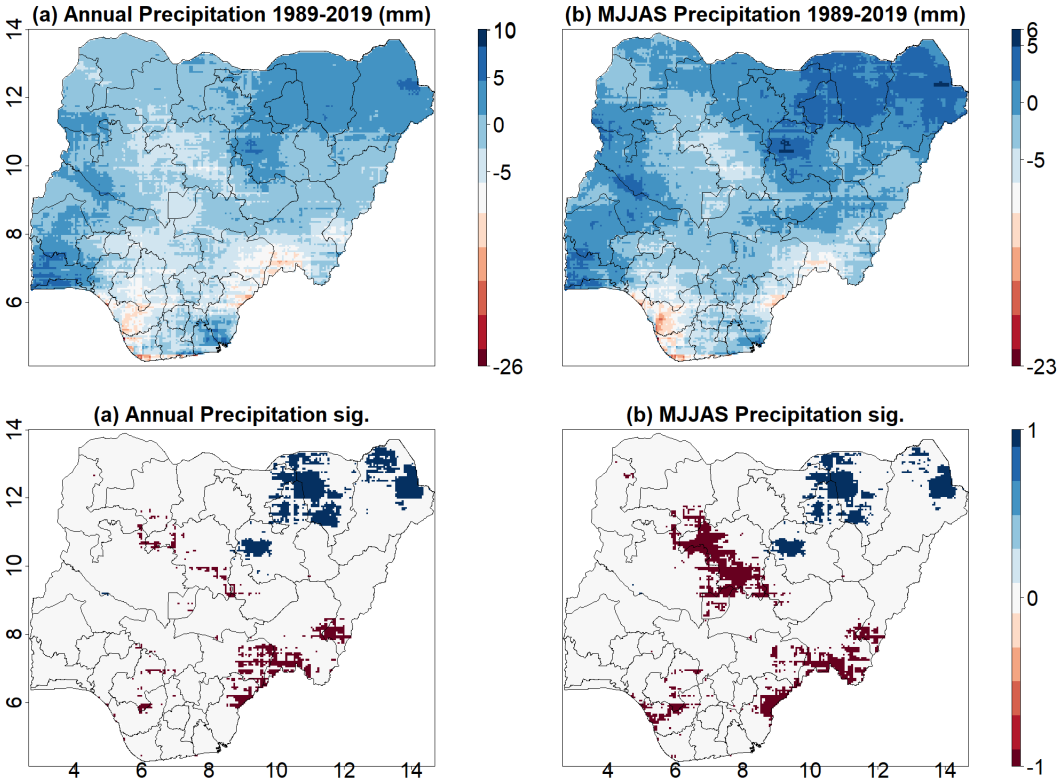
### C.0.2 Long-term changes in rainfall patterns and characteristics

Figure C.5: Long-term trends of climate indicators - 1981-2019 long-term period



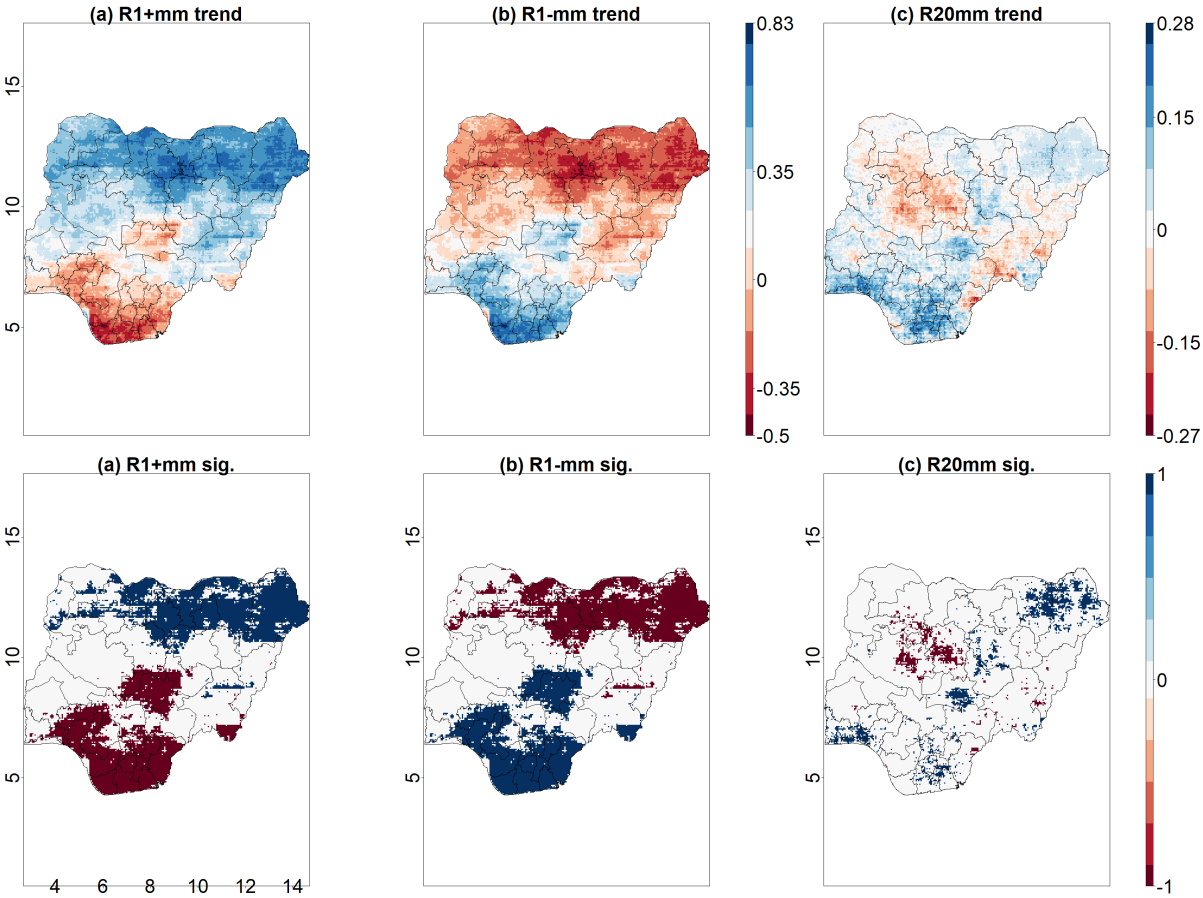

**Notes:** The Figure plots the annual (a) and long-rainy season trends (b) of precipitation amounts (mm) during the **long-term period 1981-2019**, based on CHIRPS data. Bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays a significant increasing trends, while red (-1) a significant decreasing one and 0 non significant changes.
*Sources:* Author's elaboration on CHIRPS data.

Figure C.6: Long-term trends of climate indicators - 1989-2019 long-term period



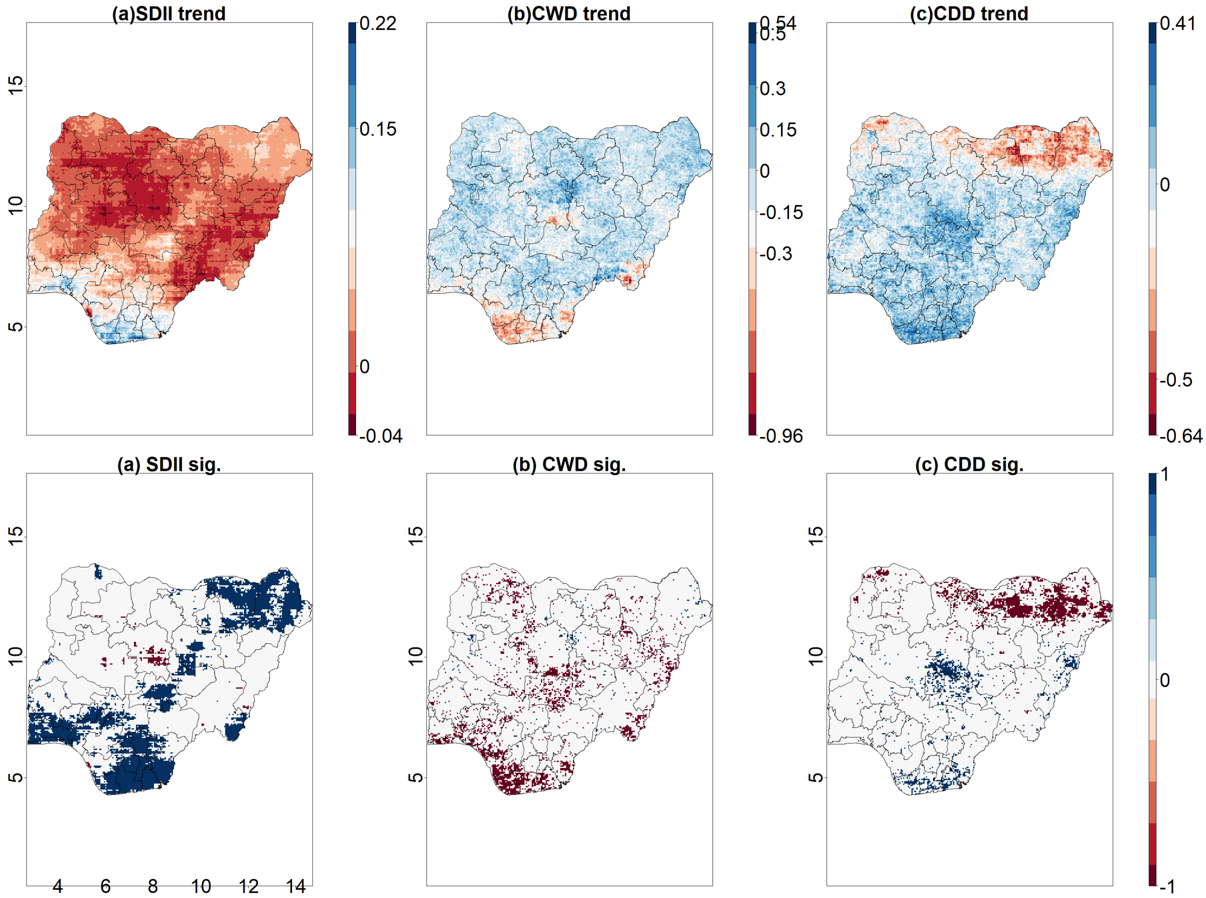

**Notes:** The Figure plots the annual (a) and long-rainy season trends (b) of precipitation amounts (mm) during the **long-term period 1989-2019**, based on CHIRPS data. Bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays a significant increasing trends, while red (-1) a significant decreasing one and 0 non significant changes.
*Sources:* Author's elaboration on CHIRPS data.

Figure C.7: Long-term trends of climate indicators (2) -1981-2019 long-term period




**Notes:** The Figure plots (a)R1+mm(b)R1+mm and (c) R20mm trends over the long-rainy season (days), during the **long-term period 1981-2019**, based on CHIRPS data. Bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays a significant increasing trends, while red (-1) a significant decreasing one and 0 non significant changes. R1+mm indicator is the number of wet days (i.e the rains are strictly positive), R1-mm is the number of dry days (i.e when the rains equal zero), and R20mm the number of heavy rains (i.e when rains aver over 20mm). By construction, R1+mm and R1-mm account for the total period and are symmetric.
*Sources:* Author's elaboration on CHIRPS data.

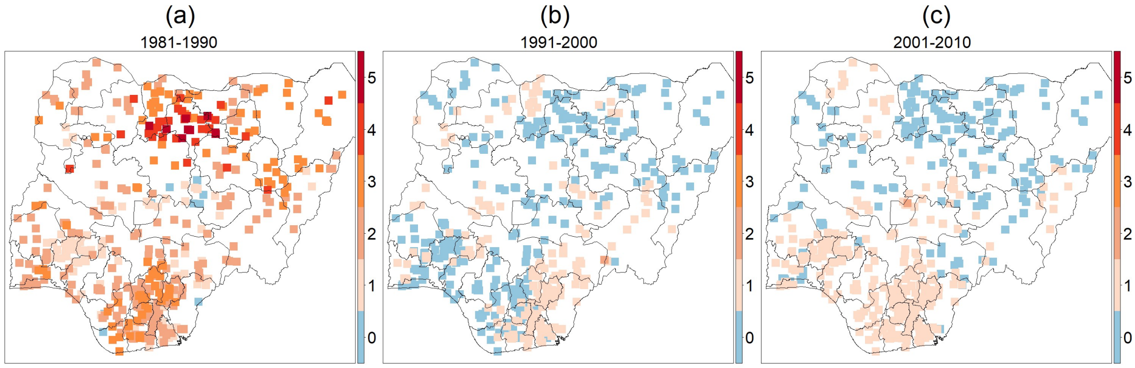
Figure C.8: Long-term trends of climate indicators (3) - 1981-2019 long-term period




**Notes:** The Figure plots (a)SDII ($mm.day^{-1}$)(b)CWD and (c) CDD(days) trends over the long-rainy season (days), during the **long-term period 1981-2019**, based on CHIRPS data. Bottom panels show the significance of the trends at $p < 0.05$. Blue (+1) displays a significant increasing trends, while red (-1) a significant decreasing one and 0 non significant changes. SDII is the Simple Daily Intesity Index, CWD the Consecutive Wet day Index and CDD the Consecutive Dry Day Index.
*Sources:* Author's elaboration on CHIRPS data.

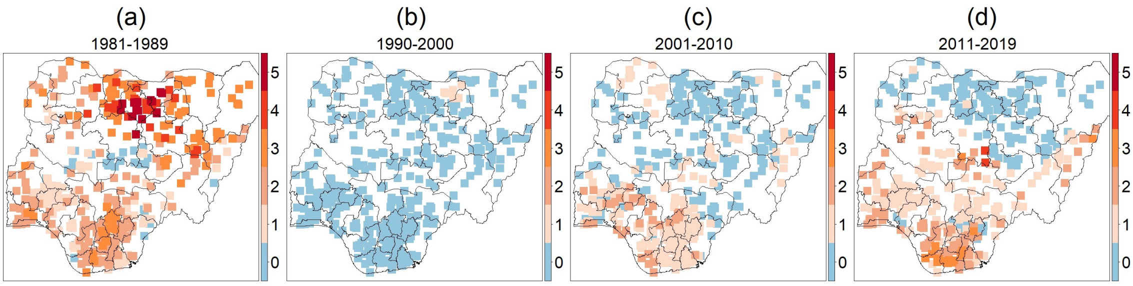
### C.0.3 Main Shock of interest

Figure C.9: Spatial distribution of the number of dry rainy seasons - past years




**Notes:** The Figures plot the number of dry rainy season for each GHS village from the panel, during the three decades over the thirty years period 1981-2010. Figure (a) plots the number of dry rainy season from 1981 to 1990 (including), Figure (b) from 1991 to 2000, while Figure (c) from 2001 to 2010. Dry years are defined according to the 1981-2010 long-term average. Please note that the number of dry rainy season reach 5 for some GHS village, as the dummy is constructed using the 10th percentile of the normal distribution of the rains (but this is scarce).
*Sources:* Author's elaboration on CHIRPS and GHS data.

Figure C.10: Spatial distribution of the number of dry rainy seasons - past years




**Notes:** The Figures plot the number of dry rainy season for each GHS village from the panel, comparing 4 time periods over the full long-term period from 1981 to 2019. This Figure makes it possible to compare the intensity of dry spells, comparing the dry period of the 80s to the 2001 drought and the more recent years, used as contemporaneous short shocks in the first stage analysis. Figure (a) plots the number of dry rainy season from 1981 to 1989 (including), Figure (b) from 1990 to 2000, Figure (c) from 2001 to 2010, and Figure (d) from 2011 to 2019. Dry years are defined according to the 1981-2019 long-term average. Please note that the length of the 4 length periods vary in order to have the same number of years between Figure (a) and Figure (d), which are the main shocks that we intend to compare in this Figure.
*Sources:* Author's elaboration on CHIRPS and GHS data.

# List of Tables

# List of Figures