Droughts, Migration and Population in Kenya

Mélanie Gittard*

January 15, 2024

Abstract

Since 2000, Kenya has experienced an increase in the frequency of droughts, significantly affecting agriculture and driving labor force migration. This paper investigates strategic migration patterns among farmers and pastoralists in response to repetitive droughts. I use fine-grained data that enables the capture of short-distance migration and heterogeneity, combining satellite-based data on daily rainfalls (CHIRPS) with exhaustive censuses from 1989,1999, and 2009. I use a two-way fixed-effect model to exploit the spatial variation in drought frequency across 2,518 sub-locations, comparing their demographic growth according to the number of dry-rainy seasons over each decade. First, I show that increased drought frequency triggers out-migration, as one additional drought decreases demographic growth by 1.7 p.p, equivalent to a 1% population decline. This result is consistent within the [15; 65] age group, excluding other demographic effects and confirming migration as the driving factor. The main contribution of this paper is the identification of different migration strategies across livelihoods. Rural areas dominated by pastoral activities experience significant out-migration, leading to a rural-rural shift from pastoral to agriculture-oriented regions. Herders' migration displays little heterogeneity, suggesting the migration of entire households and consistent with migration as a last resort. Agricultural rural areas are less vulnerable to drought and display significant heterogeneity. The results show the migration of the most educated individuals in the working age, while uneducated individuals are trapped in affected areas. This paper highlights the importance of using detailed data to understand diverse migration strategies, thereby facilitating the implementation of effective policies.

Keywords: Kenya, Droughts, Migration, Population, Census data.
JEL codes: Q51, Q54,R23,O12,O13,O15.

---

*Paris School of Economics (PSE), Centre International de Recherche sur l'Environnement et le Développement (CIRED) and Ecole Nationale des Ponts et Chaussées (ENPC), France. Contact author's email : melanie.gittard@psemail.eu . I am grateful to my advisors Denis Cogneau and Philippe Quirion for their support and extensive feedbacks on this project. I also would like to thank Eric Strobl, Christelle Dumas, Liam Wren-Lewis, Raja Chakir Katrin Millock, Serge Janicot, Tamara Ben-Ari, Tristan Le Cotty, Wolfram Schlenker, my many colleagues in Paris School of Economics (PSE) and the Centre International de Recherche sur l'Environnement et le Développement (CIRED) for helpful comments and discussion, the participants at the Casual Development Seminar at PSE, the participants of the CIRED seminar. I also thank seminar participants at the NOVAFRICA, CISEA, SISC, NDCE, IDCE, JMA, FAERE, EAERE and EEA. The usual disclaimer applies. Declaration of conflicts of interest: none.

1 Introduction

Over the past decades, East Africa has been facing an increase in the frequency of droughts, associated with a decrease in the agricultural season length and an increase in dry conditions [Gebrechorkos et al., 2019], significantly affecting agricultural activities and making adaptation necessary. Migration is a possible strategy for adaptation to climate shocks, which can act as 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 [Cattaneo et al., 2019a]. Migration might also occur when local strategies have failed and be a last resort strategy. 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]. Understanding the characteristics of climate-induced migration in rural settings at the local scale is critical to reducing vulnerability to climate events.

This paper investigates internal strategic migration patterns among farmers and pastoralists in response to the increase in drought frequency. It contributes to prior research using fine-grained data that enables the capture of short-distance migration and heterogeneity. To analyze rainfall patterns over the 1983-2013 period, I use high-resolution satellite-based data on daily rainfall and temperature, 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. 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. 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.

First, 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 (or a 1% population decline) compared to normal rainfall conditions. This result is consistent within the [15; 65] age group, excluding other demographic effects and confirming migration as the driving factor. 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. I find no significant results of the effects of an additional flood during the main agricultural season, which shows the determinant impact of droughts.

The main contribution of this paper is the heterogeneity, which identifies different migration strategies across livelihoods. Overall, the heterogeneity analysis shows that migrants are mainly young individuals of working age and with minimum education, as non-educated 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, consistent with migration being a strategy of last resort. On the other hand, agriculture-oriented areas display high heterogeneity, as the out-migration is driven by young and skilled individuals and non-educated individuals are trapped in affected rural areas, which is in line with individual migration as a response to climate shocks and with an income diversification strategy. Finally, I also find a structural change in the labor market in treated urban areas, as business owners seem to fall into unemployment.

The intensive margin analysis shows the critical effects of drought repetition and suggests a rural-rural shift from pastoral to agriculture-oriented regions. The out-migration increases when the number of droughts over a decade increases. The results from changing the threshold intensity suggest that herders relocate into rural agriculture-oriented areas with favorable rainfall conditions. I use the land-cover ESA Glob-Cover data to distinguish areas where croplands are dominant, a proxy for the intensity of agricultural activity, and those where pastures prevail. The results show that the overall effects is driven by the out-migration of mixed areas, containing both pastoral areas and croplands, showing that the results are driven by pastoralist systems with diversified livelihoods. Overall, these results suggest that the response to repetitive droughts in Kenya results in rural-rural migration.

The findings of this paper are supported by a battery of robustness checks. The main result is robust in controlling for average temperature, 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 climate-induced migration 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 across livelihoods. 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 2 reviews the literature and presents the contributions. Section 3 details the data used in the paper, and Section 4 the context as well as spatial and temporal changes in precipitation in Kenya. Section 5 presents the main empirical strategy. Section 6 lays out the main results, while Section 7 investigates the heterogeneity. Section 8 looks at the intensive margin of the results, and Section 9 proposes a list of robustness checks and tests. Section 10 concludes.

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 et al., 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.

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 et al. [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].

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.

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 et al., 2011]. It displays mixed evidence of climate-induced migration according to the dependence on the agricultural sector [Cai et al., 2016; Feng et al., 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 et al., 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 et al. [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 et al. [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 et al. [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.

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 et al., 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; Thiede and Gray, 2017; Albert et al., 2021], historical archive records [Hornbeck, 2012; Lynham et al., 2017], or panel survey [Dillon et al., 2011; 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 et al., 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 et al. [2022] use administrative censuses as well and finds evidence of net outflows in response to dry shocks driven from the SPEI Vicente-Serrano et al. [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.

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 et al., 2019b]. 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.

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].

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].

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 14 in Section B.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 B.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.

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 in-

[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.

dividual 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 B.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 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

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

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.

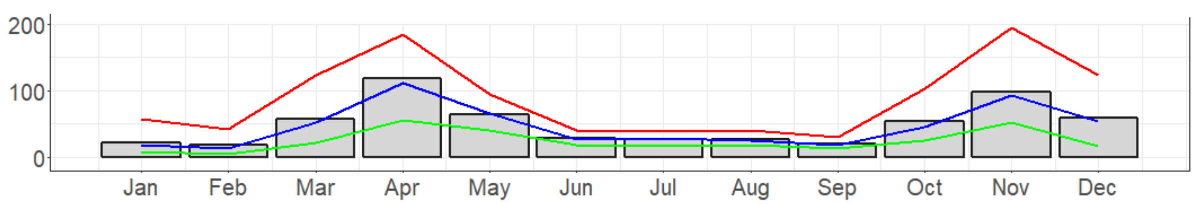
4 Context and descriptive statistics

4.1 Setting

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.

Figure 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.

[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.

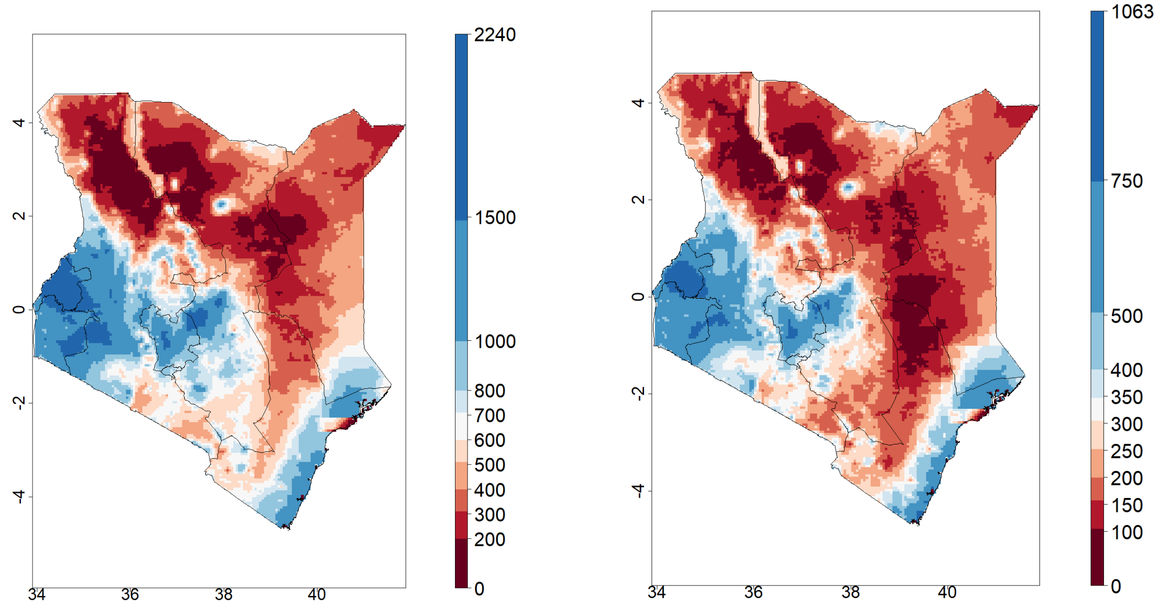
Lying across the equator, Kenya experiences a bimodal seasonal pattern, as shown in Figure 1. The two wet seasons are the long rains and the short rains and are separated by two other seasons with little rainfall. The long rainy season is the primary agricultural season, and extends from March to June (MAMJ), with a peak centered around March/-May, and is modulated by local factors rather than global scale ones [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 21 in Section C.1 displays the seasonal patterns across provinces. If some heterogeneity is observed, all provinces show a bimodal 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 2 plots the spatial variation of the long-term average of annual rains (Figure 2a) and the long-rainy season (MAMJ) cumulative precipitations (Figure 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 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 2. Figure 22 in Section C.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 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.

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

Figure 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 .

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 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].

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

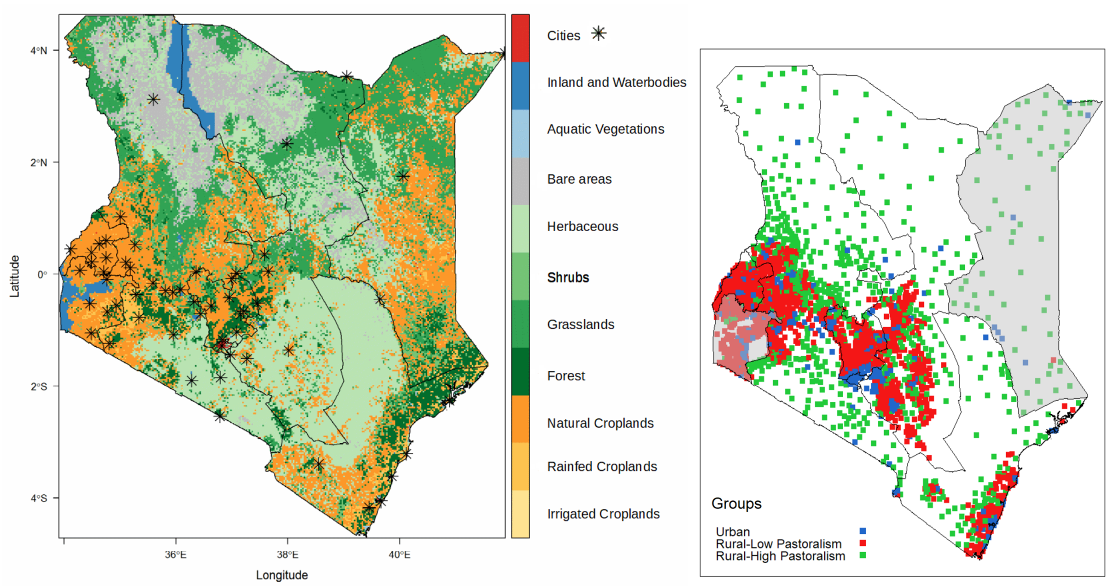
[9]Percentages are computed using the ESA GlobCover data

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 2 and 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 21 22 in Section C.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 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

Figure 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.

events [Hesse and MacGregor, 2006].

Thus, the majority of the labor force is involved in agricultural activity including both farming and husbandry. Figure 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 3 shows the correlation between the dummy indicating pastoralist activity and the presence of grasslands and pastures [10].

4.1.3 Long-term rainfall trends

Section C.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

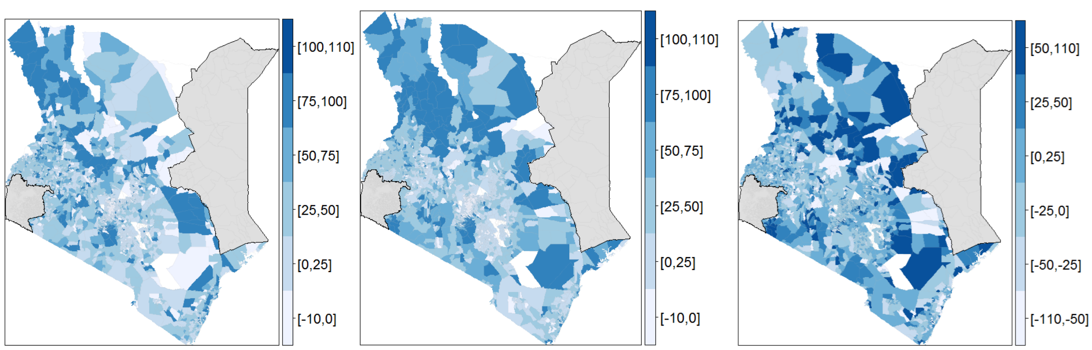
[10] The dummy for high pastoralism correlates at 46^{***}% with the presence of grasslands

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.

4.2 Temporal and spatial variation

4.2.1 Population and migration

Figure 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 G). Kenya faces high population growth, and Central and Western provinces are the most populated and dense areas (aside from Nairobi) (Tables 19 20).

Figure 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. Figure4 shows spatially clustered population growth trends, which will be discussed in Section 9.3. Figure 27 and 26 in Section C.3 plot spatial patterns of population size and density for each sublocation.

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 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 28 from Section C.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 28 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.

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, 2017]. Figure 29 plots times series of seasonal anomalies, and displays high intraseasonal variability. It shows that the decline

[11] Arid and semi-arid region

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

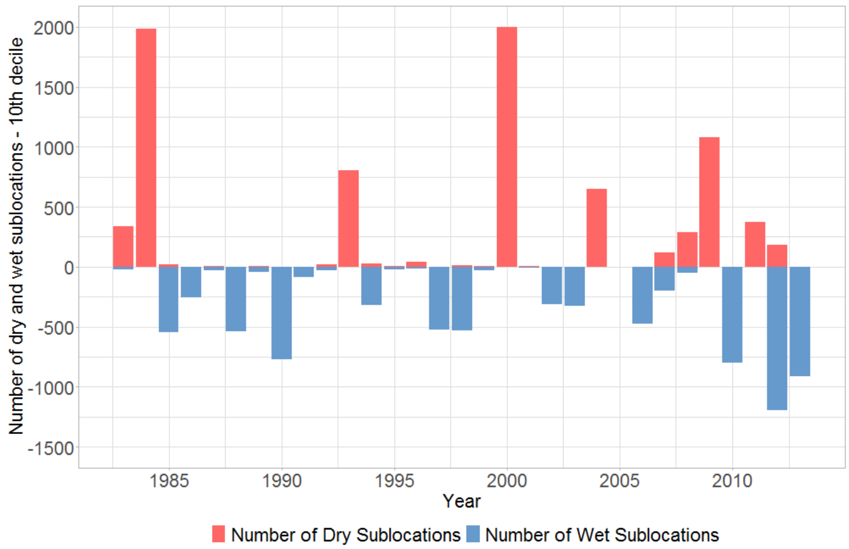
in precipitation in Kenya is mainly borne by the decrease during the rainy season MAMJ, while the increase in excessive rains is mainly borne by an increase in wet conditions during the short rainy season OND. As this paper focuses on the effect of drought repetition, the main analysis focuses on the occurrence of rainfall over the long-rainy season MAMJ.

Figure 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 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 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 6 displays as well the occurrence of excessive rains over MAMJ, which are more distributed over the years and less intense than droughts. Figure 30 in Section C.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 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: $z_{i,t} = \sum_{j=1}^{10} 1_{i,t-j}$, such as:

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

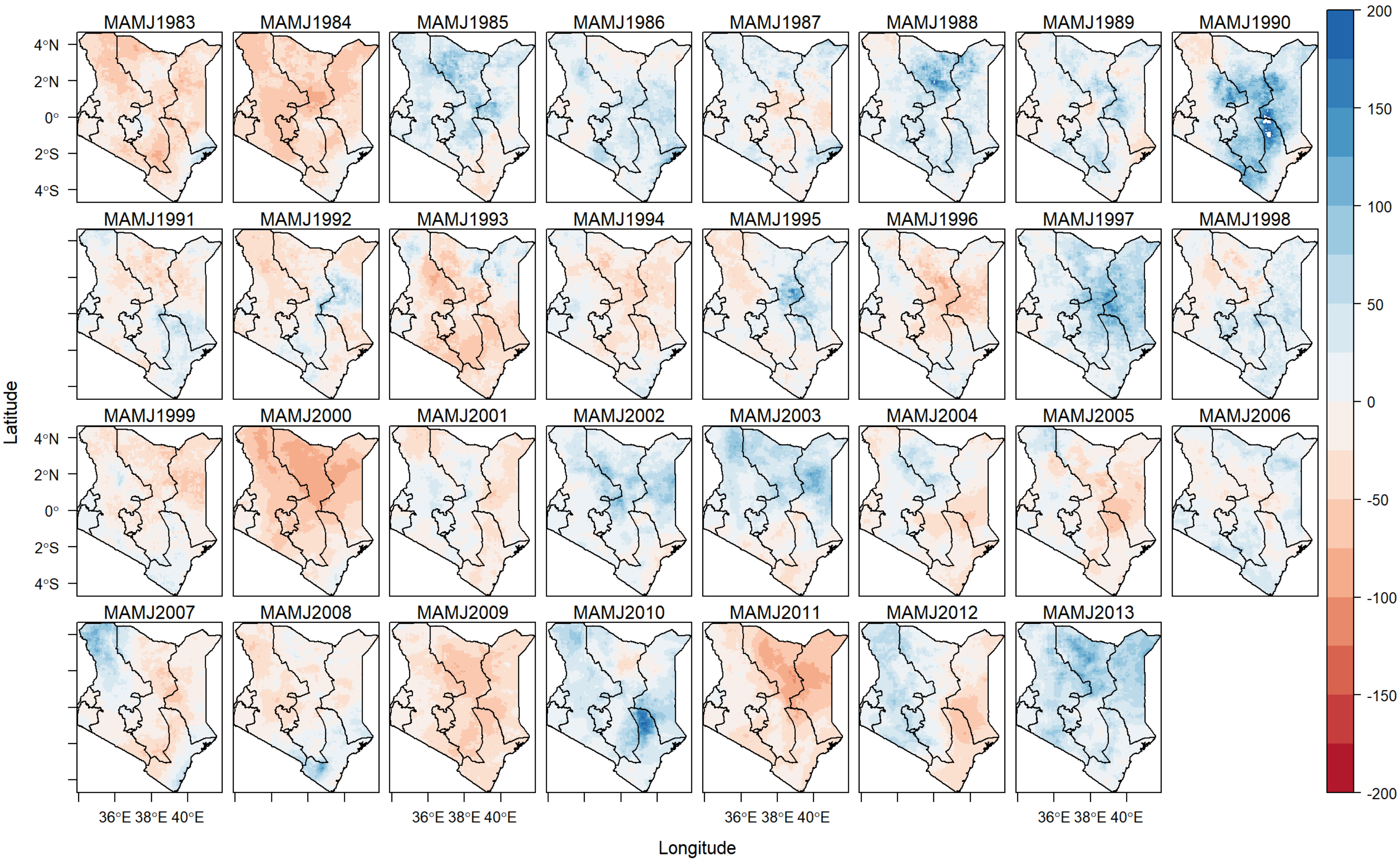
Figure 5: Number of dry and wet sublocations across years - 10th decile



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.

Figure 5 plots for each year the total number of sublocations for which the dummy $1_{i,t}$ 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 5 shows that 2009 was particularly intense, its effects on Kenyan demography will not be analyzed due to the availability of population data. Figure 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 29, Figure 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 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.

5 Empirical strategy

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 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 9.1 for further discussion). More formally, the empirical strategy can be written as follows :

$$DPGR_{i,t} = \alpha_0 + \alpha_1 dens_{i,t_0} + \alpha_2 z_{i,t} + \alpha_3 dens_{i,t_0} \times z_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t} \qquad (1)$$

With $z_{i,t} = \sum_{j=1}^{10} 1_{i,t-j}$, the number of years considered as dry over the decade (definition based on the long-rainy season, cf Section 4.2.2). γ_i and γ_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 26 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 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, α_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.

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 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 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 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 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.

Eventually, a threat to the heterogeneity analysis would be that economic status and

[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

education are endogenous to rainfall shocks. Section 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 7.3, I can not rule out the fact that the results suggest changes in the labor market rather than pure migration effects.

6 Main results

6.1 Drought intensity

This section gives the main result of the paper estimated from equation 1. Table 1 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 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 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.

[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

[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%.

Table 1: Effects of the number of dry rainy season 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].

Table 2 gives the same results as Table 1, 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

[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*.

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

Table 2: Effects of the number of dry rainy season 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.

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. 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 7 investigates the heterogeneity of the migration across sublocation characteristics to better understand these different forms of migration.

6.2 Flood intensity

Table 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 26 in Section F.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 D.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 3: Effects of the number of wet rainy season on the DPGR

	All Kenya	Urban	Rural	Low Pastoralism	High Pastoralism
	(1)	(2)	(3)	(4)	(5)
(6)					
Number of wet years	0.841	0.454	0.811	1.423	1.564
	[0.520]	[1.201]	[0.606]	[0.912]	[1.259]
Number of wet years × density	0.000182	0.000149	0.00153	-0.000956	0.00186
	[0.000240]	[0.000244]	[0.00220]	[0.00254]	[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.

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 7.1 gives the heterogeneity of the migration across gender and age brackets, Section 7.2 according to the educational level (defined according to past schooling attendance of adults), and Section 7.3 according to the economic activity.

7.1 Age and Gender

7.1.1 Gender

Table 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 1 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 22 in Section E.1 reproduces the same heterogeneity analysis across gender following the cohort of individuals aged from 15 to 65 years old as in Table 2. Again,

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

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 4. As in Table 2, the effect is significant in rural areas with low pastoralism, in line with the hypothesis of the working population.

Table 4: Effects of the number of dry rainy season 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

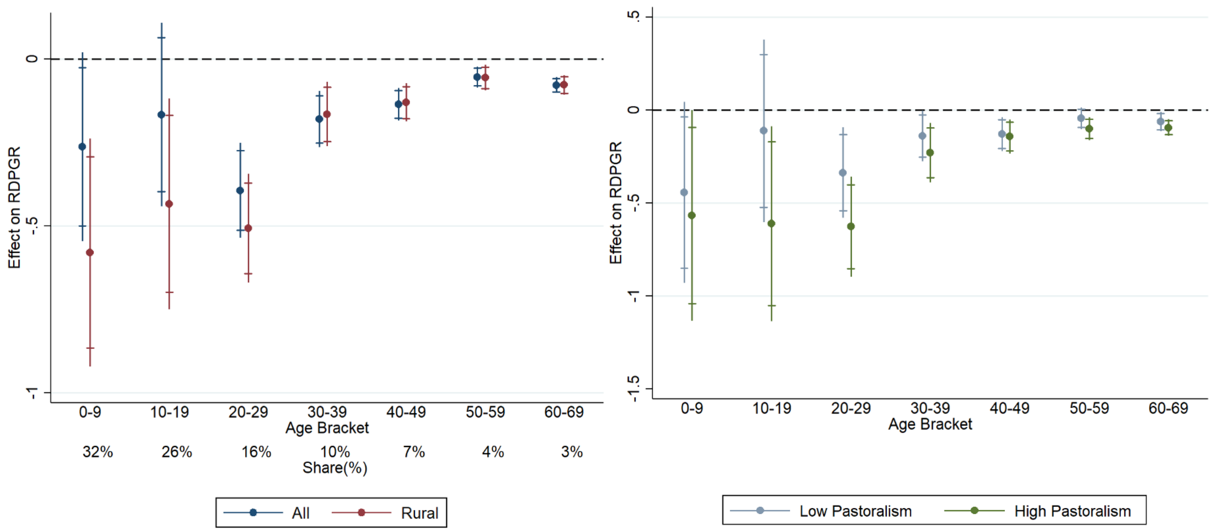
Notes: Standard errors clustered at the sublocation level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

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 7 breaks down the effect on the RDPGR of each 10-year bracket within [0-69], each bracket following cohorts. Table 23 in section E.1 gives the effect on all the [0-69] cohort, which is the sum of each regression dot per location type. Figure 7a plots the effect for all the sublocations and rural sublocations, and Figure 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 7a.

Overall Kenya, Figure 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 interpreted as

Figure 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.

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

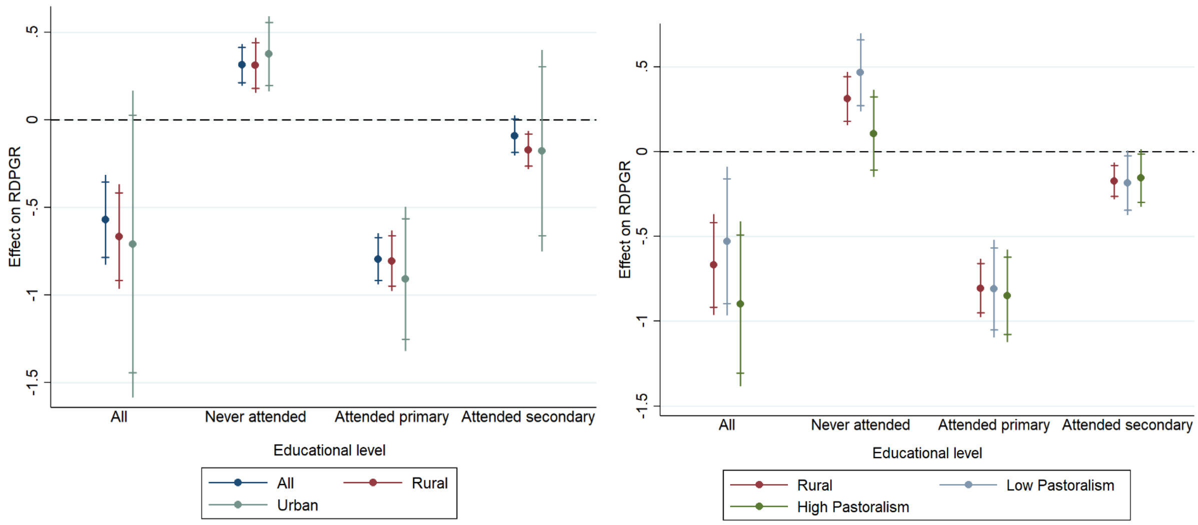
Figure 7 shows that the effect is mainly driven by the migration of young individuals in the age of working ([10-19] and [20-29] brackets), across all types of sublocation types. This is especially the case in low pastoralist areas, as the effect is only significant for individuals aged between [20-29], while it is more homogeneous across age brackets in high pastoralist areas (Figure 7b). This result is in line with an out-migration of the young working population within agricultural areas.

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 8 distinguishes individuals that never attended schooling, that at least have attended primary school [20],

[20]attended or completed primary education

Figure 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.

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. Figure8 shows a reverse effect of individuals that never went to school, which can be interpreted as a proxy for the illiterate population. Figure 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 proxy for richness, this result can be explained by the fact that individuals are too credit-constrained to fi-

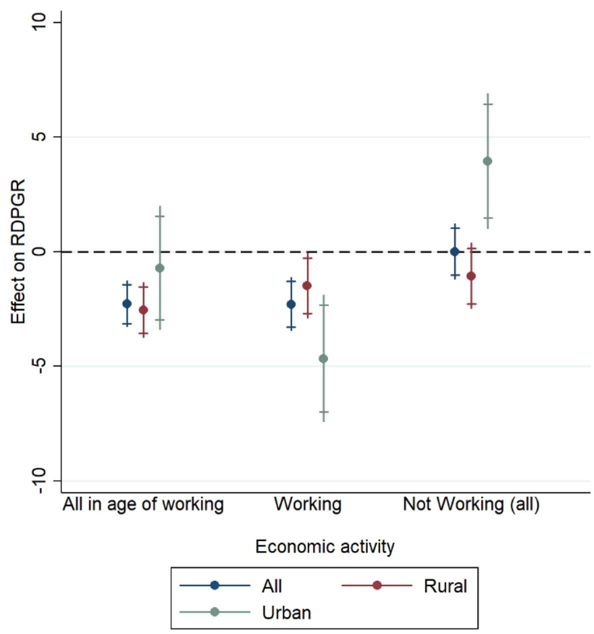
[21] attended or completed secondary education. This includes individuals that went to university. As they only represent 0.72% of the population, this subgroup could not be distinguished

nance 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.

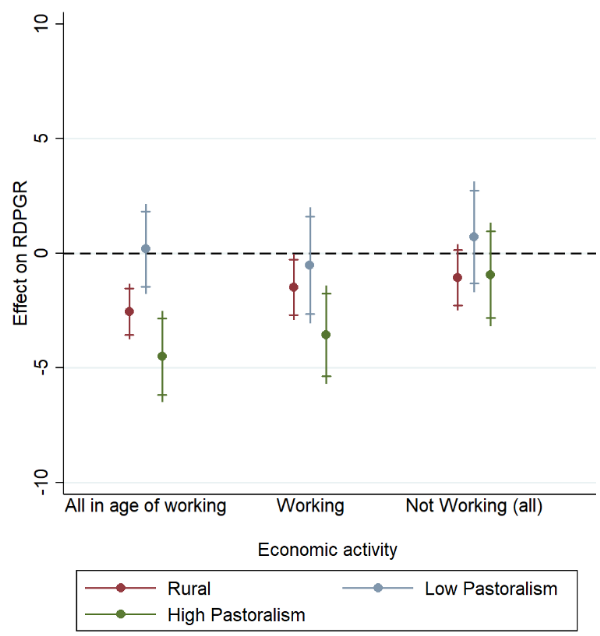
Figure 31 in section E.2 displays the results on education according to age brackets, to verify that these results are not driven by any age effects.

7.3 Economic Activity

Figure 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 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 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 9 translates an effect of climate variability on labor allocation rather than migration. An important result from Figure 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 32 from section E.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

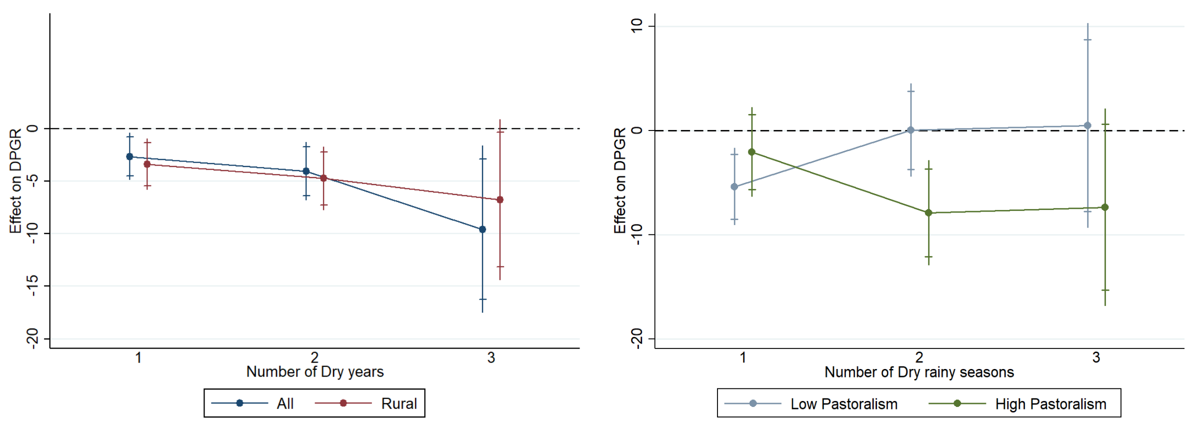
8 Intensive Margin

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 14a and 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 5.

Figure 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 10b shows that the effect on rural areas where pastoralism prevails is mainly driven by sublocations that have been hit by 2 droughts.

Figure 10: Intensive margin - Effect of the number of dry years on the DPGR



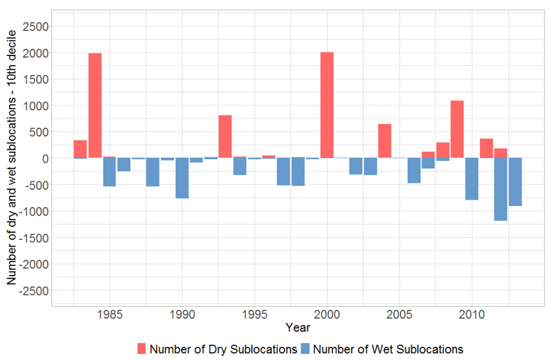
(a) All Kenya (b) Rural areas

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.

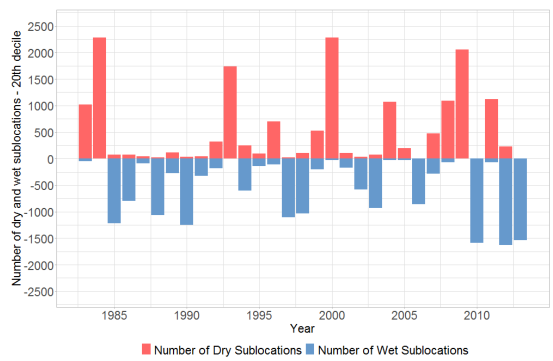
8.2 Threshold intensity

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 11 replicates Figure 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.

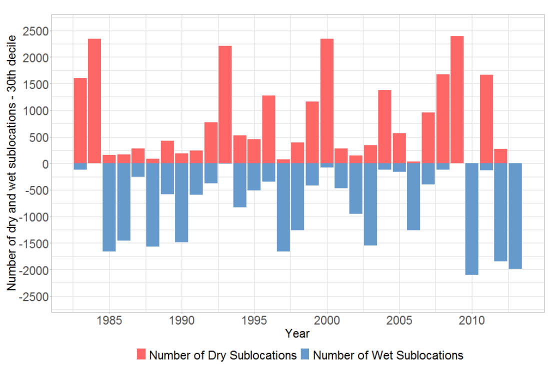
Figure 11: Number of dry and wet sublocations across years and deciles



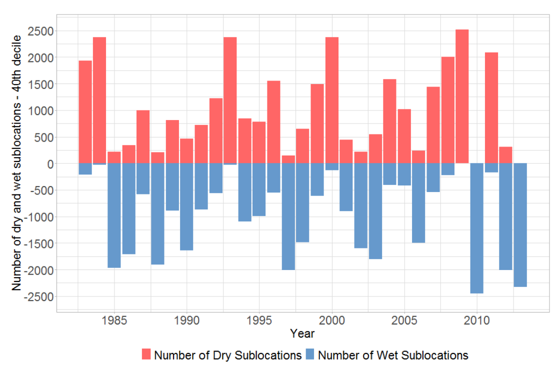
(a) 10th decile



(b) 20th decile



(c) 30th decile



(d) 40th decile

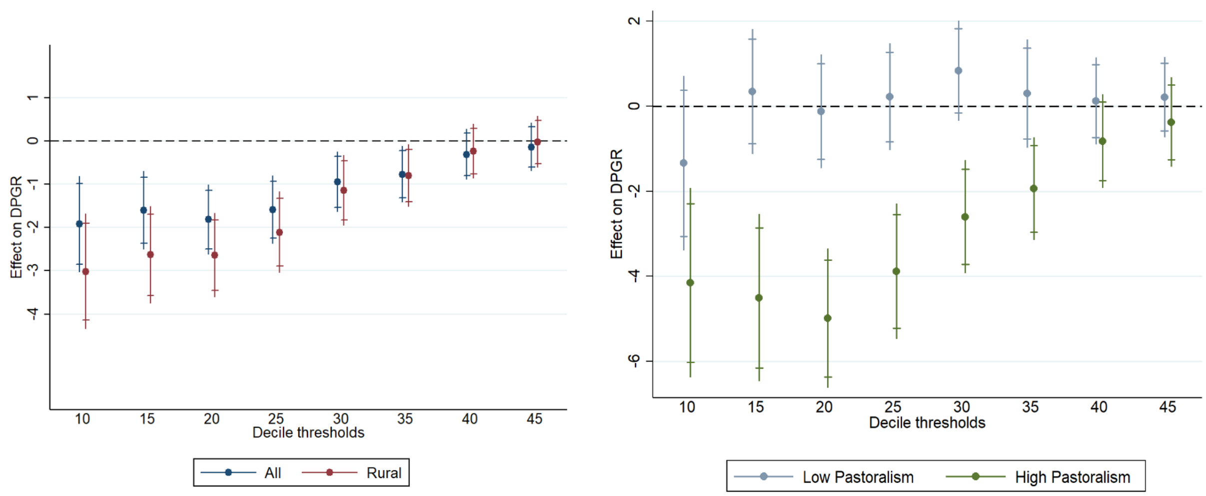
Notes: These Figures in red are the number of sublocations for which the rainy season is dry and wet across years, for four different decile thresholds.

Figure 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 1. Figure 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 12b), the decrease of the DPGR is still high up to the 20th decile and then reduces as the threshold increases.

Figure 13 replicates the same analysis for excessive rains. Effects on the 90th decile correspond to the main results from Table 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 13b shows that this attraction effect is mainly borne by sublocations where agriculture is the main livelihood strategy, and pastoralism is not predominant.

Figure 12: Intensive margin - Effect of the number of dry years on the DPGR

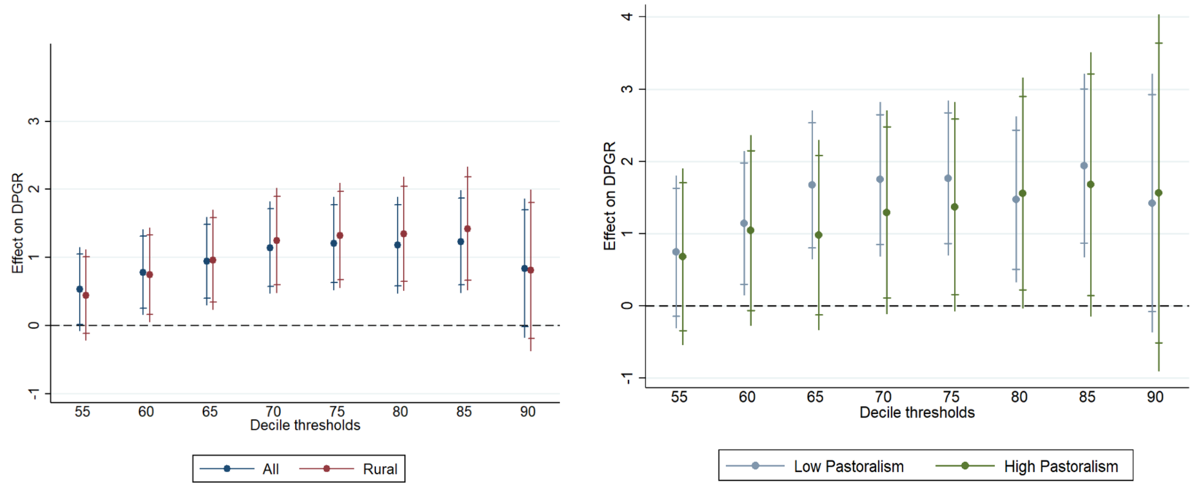


(a) All Kenya

(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 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.

8.3 Land Cover and Agricultural activity

Table 5 replicates the main analysis according to a land-use classification of sublocations, based on the ESA GlobCover data, as illustrated in Figure 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 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 is mainly pastoral areas. It shows no effect within cropland regions, and that the overall effects are

[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

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 5: Effects of the number of dry rainy season on the DPGR across land cover categories

	Cropland areas		Pastoral areas	
	(1)	(2)	(3)	(4)
Number of dry years	-1.148	-0.106	-2.165**	3.251
	[0.714]	[0.721]	[0.954]	[2.484]
Number of dry years × Pasture presence		-2.054***		
		[0.762]		
Number of dry years × Cropland presence				-5.567**
				[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 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 migrating towards agriculture-oriented rural areas, and maybe changing their main economic activity.

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

9 Robustness checks

9.1 Binary treatment and de Chaisemartin and d'Haultfœuille [2020]

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 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 increase in the occurrence of dry rainy seasons since 2000. As Figures 5 and 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]:

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

[28] I exclude from the analysis sublocations that have been impacted both in 1993 and 2000. As the

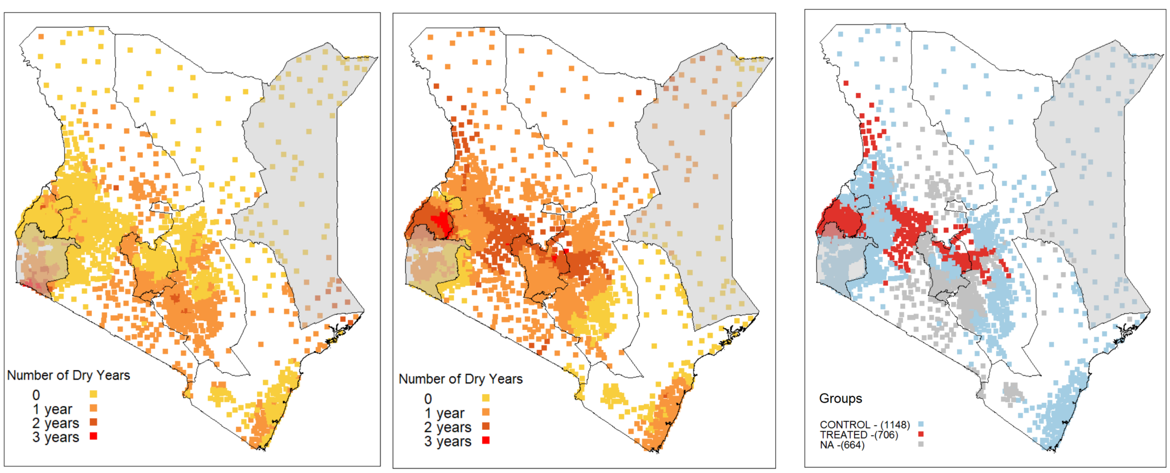
$$D_i = \begin{cases} \mathbf{1} & \text{if sublocation i was hit by at least 2 droughts in period 2} \\ & \text{(2000 and 2004 and/or 2007/2008) and no more than one in period 1 (in 1993)} \\ \mathbf{0} & \text{ if sublocation i has either known a drought in 1993, either 2000, or none} \\ . & \text{if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

The empirical strategy can be formally written as follows :

$$DPGR_{i,t} = \frac{\Delta pop_{i,[t-10,t]}}{pop_{t-10}} = \alpha_0 + \alpha_1 D_i + \alpha_2 P_t + \alpha_3 D_i \times P_t + \gamma_i + \epsilon_{i,t} \tag{2}$$

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

Figure 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 14 maps the number of dry rainy seasons over the two periods and identifies 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 4). This is not the case for the Central and Rift Valley,

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 with hard geographic conditions, a high share of pastoralism (3b) and decreasing DPGR trends. As treated sublocation are clustered, Section 9.3.1 adjusts standard errors for both spatial and serial auto-correlation.

Table 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 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

[29] This is equivalent to the DiD coefficient, α_3 from 2

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 7 displays the results of the binary treatment from equation 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 24 in Section F.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 9.1.2.

Table 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.

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 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 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 8: Effects of the number of dry rainy season 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 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 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 et al. [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 Group 1 and Group 3, for which the stable assumption holds [32]. Column (1) shows that, when excluding Group 2 from the analysis,

[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.

[32]between the two periods, there are sublocations for which the treatment is stable, i.e the number of

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 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 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.

9.2 Common trend assumption

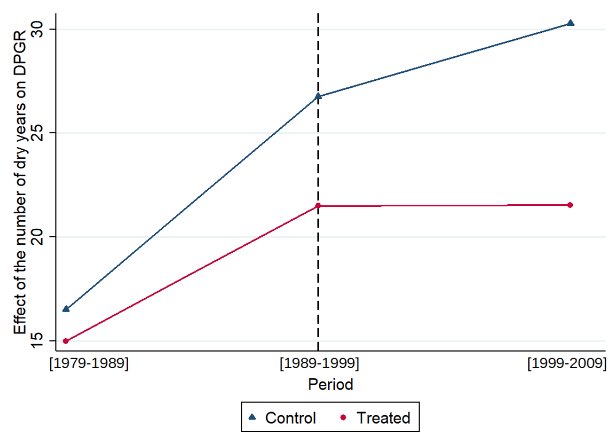
The key assumption of the DiD strategy is that the dependent variable would follow the same time trends in the absence of droughts both in treated and control groups. To test for the common trends assumption, one can observe the pre-treatment data and the evolution of the DPGR before the two periods.

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 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

droughts does not change, Assumption 10 from de Chaisemartin and d'Haultfœuille [2020] : it is Group 3 the$DPGR_{[1979,1989]}$ by 100 to look at averages across groups in Figure 15[33]. Table 25 in Appendix F.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 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 9.1.1.

Figure 15 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 15 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 7, Figure 15 shows that the DiD estimators are driven by a deceleration of the demographic growth for treated sublocations, in comparison to control sublocations.

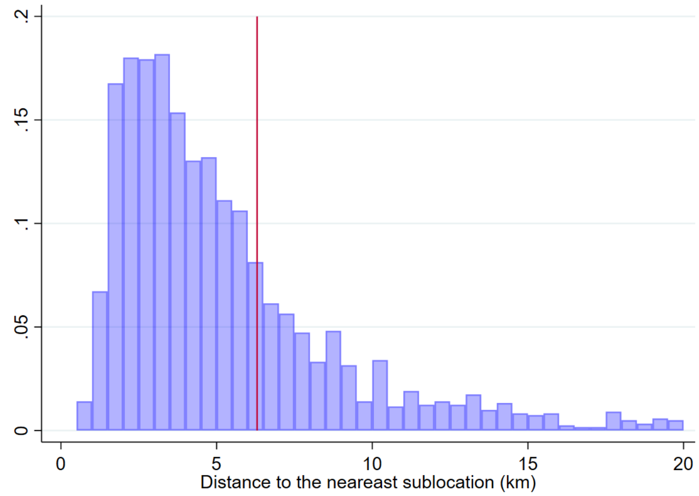
[33]I divided by 100 to have coherent numbers. Besides, following the main analysis, I winsorize the $DPGR_{[1979,1989]}$ at the 5% level

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

9.3 Spurious correlation

9.3.1 Spatial correlation

Figure 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 14 and 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 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 9 shows the stability of the significance of the main result (Column (2) Table 1) 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 9: Effect of the number of dry rainy season 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.

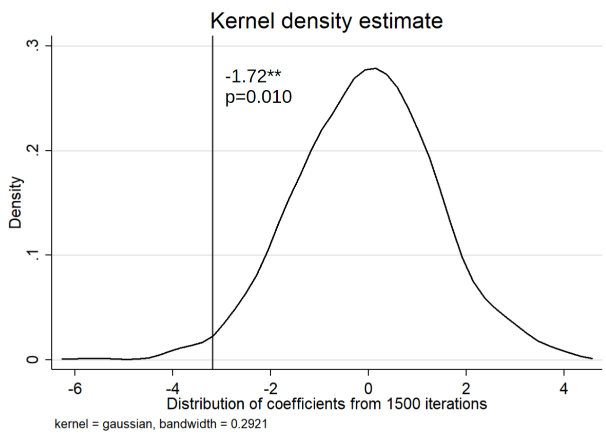
9.3.2 Placebo tests

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 17a 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 17b randomly changes the timing (and thus the total number) of the rainfall shocks for each sublocation.

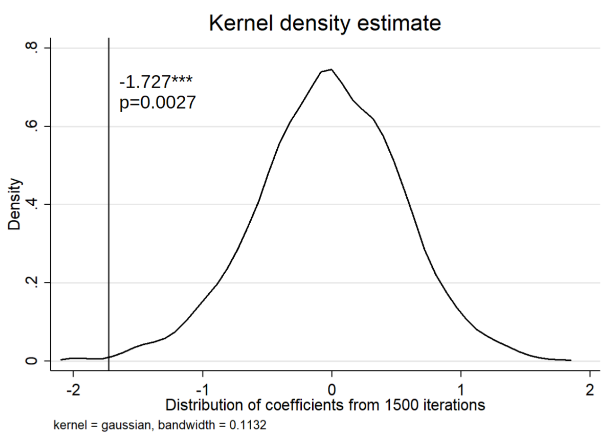
Both simulations show that the distribution of the treatment effects are shifted around zero, and are almost perfect replication of the standard normal distribution. The vertical lines indicate the location of the estimates under the implemented treatment assignment (Table 1 Column 2), indicating the rejection regions, and gives the new estimated p-value. I am sure at the 1% level (Figure 17b) and 5% level (Figure 17a) that the model is not misspecified. Figure 33 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 1% level.

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

Figure 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 1 Column 2), and gives the new estimated p-value. *Sources:* Author's elaboration on CHIRPS and KNBS data.

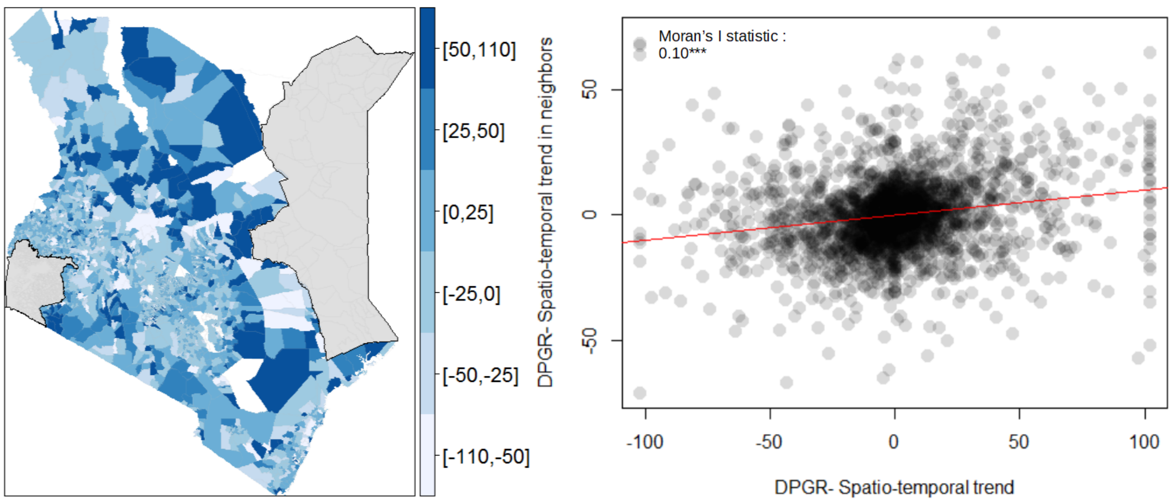
9.3.3 Spatially dependent trends

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 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 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 18 illustrates the spatial dependency of the dependent variable. Figure 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 at-

Figure 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.

tenuated, decreasing trends in the west and the center. Figure 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 (2.2×10^{-16}), and shows the spatial correlation of the migration outcome. Figure 14 shows the spatial clustering of the trends of rainfall shocks as well.

As spatial pattern is found, Table 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 1×6 terms to control for spatiotemporal trends as proposed in Lind [2019], and show that my estimation is robust [38].

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

[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 $i \geq 2$, $P_i(x) = [(2i-1)xP_{i-1}(x) - (i-1)P_{i-1}(x)]/i$

Table 10: Effects of the number of dry rainy season 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

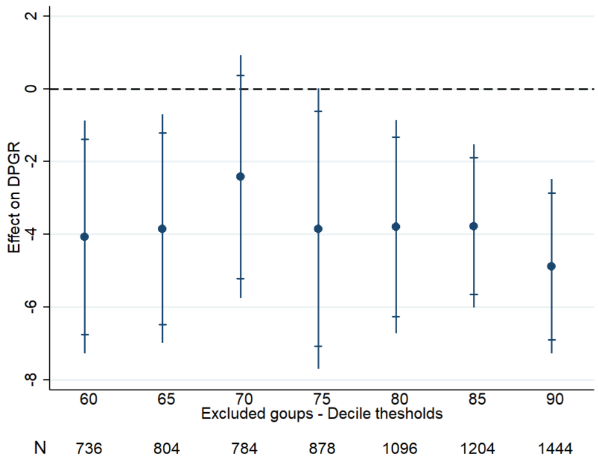
Notes: Standard errors clustered at the sublocation level, $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$. Nyanza and North Eastern provinces are excluded. Each demographic variable is winsorized at the 5% threshold.

9.4 Contamination of the Control group

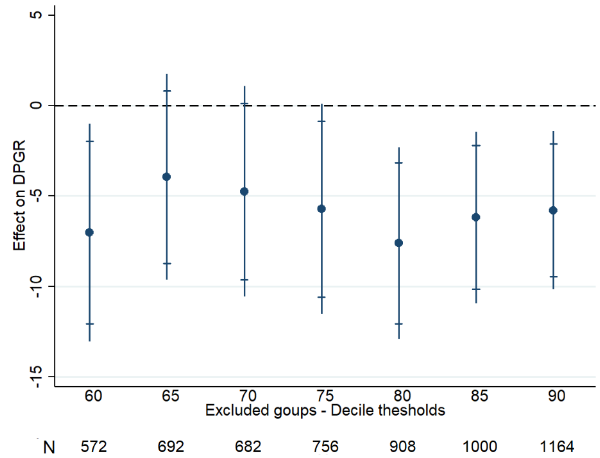
One concern in the main analysis is the overestimation of the effect due to the contamination of the control group. The main result shows that being hit by an additional dry year increases the likelihood of out-migration. As the results suggest a rural-rural migration mainly born by herders, there is 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 8.2 suggests that the individuals migrate within sublocations with humid conditions. Figure 13 shows that being under an additional moderate wet year 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 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 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 19b replicates the same exercise for the binary treatment as defined in Section 9.1.1.

9.5 Other climate indicators

9.5.1 The role of temperature and evapotranspiration

Table 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 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 11: Effects of the number of dry rainy season 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 et al., 2010], which controls for both temperature and PET (Section F.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 34b in Section F.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.

9.5.2 Other indicators

Table 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 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 [0.530]	-0.0120 [0.409]	-0.546 [0.694]	-1.261** [0.493]	-0.210 [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 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 12 shows that it is mainly the drop in cumulative rains over MAMJ that play a key role in migration.

Table 26 in Section F.4 shows no effects of droughts nor floods occurring during the short rainy season OND.

9.6 Replication at the district level

Table 13: Effects of the number of dry rainy season on district DPGR

	All Kenya		Urban	Rural	Low Pastoralism	High Pastoralism
	(1)	(2)	(3)	(4)	(5)	(6)
Number of dry years	-5.035*	-5.526*	1.691	-5.858*	2.722	-7.594**
	[2.844]	[3.124]	[2.905]	[3.304]	[3.094]	[3.472]
Number of dry years × density		0.00748	-0.00348	0.00828	0.00204	0.0114
		[0.00968]	[0.00727]	[0.0105]	[0.00714]	[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 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 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 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 G proposes another estimation, looking at the effects of yearly droughts on the inter-district bilateral migration flows using a PPML estimation. Again, as the results

[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.

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.

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 across livelihoods. 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.

References

Albert, C., Bustos, P., and Ponticelli, J. (2021). The effects of climate change on labor and capital reallocation. Working Paper 28995, National Bureau of Economic Research.

Auffhammer, M., Hsiang, S., Schlenker, W., and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, 7(2):181–198.

Barrios, S., Bertinelli, L., and Strobl, E. (2006). Climatic change and rural–urban migration: The case of sub-saharan africa. *Journal of Urban Economics*, 60(3):357 – 371.

Beine, M. and Parsons, C. (2015). Climatic factors as determinants of international migration. *The Scandinavian Journal of Economics*, 117(2):723–767.

Bertoli, S., Docquier, F., Rapoport, H., and Ruyssen, I. (2021). Weather shocks and migration intentions in western africa: insights from a multilevel analysis. *Journal of Economic Geography*, 22(2):289–323.

Bryan, E., Ringler, C., Okoba, B., Roncoli, C., Silvestri, S., and Herrero, M. (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, R., Feng, S., Oppenheimer, M., and Pytlikova, M. (2016). Climate variability and international migration: The importance of the agricultural linkage. *Journal of Environmental Economics and Management*, 79:135–151.

Campbell, D. J. and Axinn, G. H. (1980). Pastoralism in kenya. Report, American Universities Field Staff.

Cattaneo, C., Beine, M., Fröhlich, C. J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., Millock, K., Piguet, E., and Schraven, B. (2019a). Human migration in the era of climate change. *Review of Environmental Economics and Policy*, 13(2):189–206.

Cattaneo, C., Beine, M., Frohlich, C. J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., Millock, K., Piguet, E., and Schraven, B. (2019b). Human migration in the era of climate change. *Review of Environmental Economics and Policy*, 13(2):189–206.

Cattaneo, C. and Peri, G. (2016). The migration response to increasing temperatures. *Journal of Development Economics*, 122(Supplement C):127 – 146.

Colella, F., Lalive, R., Sakalli, S. O., and Thoenig, M. (2019). Inference with arbitrary clustering. IZA Discussion Paper n. 12584.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Dallmann, I. and Millock, K. (2017). Climate variability and inter-state migration in india. *CESifo Economic Studies*, 63(4):560–594.

de Chaisemartin, C. and d'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

de Chaisemartin, C., d'Haultfœuille, X., and Guyonvarch, Y. (2019). Fuzzy differences-in-differences with stata. *The Stata Journal*, 19(2):435–458.

Defrance, D., Delesalle, E., and Gubert, F. (2022). Migration response to drought in mali. an analysis using panel data on malian localities over the 1987-2009 period. *Environment and Development Economics*, page 1–20.

Dell, M., Jones, B., and Olken, B. (2014). What do we learn from the weather? the new climate–economy literature. *Journal of Economic Literature*, 52(3):740–798.

Dillon, A., Mueller, V., and Salau, S. (2011). Migratory responses to agricultural risk in northern nigeria. *American Journal of Agricultural Economics*, 93(4):1048–1061.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., and Ceccato, P. (2018). Validation of the chirps satellite rainfall estimates over eastern africa. *Quarterly Journal of the Royal Meteorological Society*, 144(S1):292–312.

Endris, H. S., Omondi, P., Jain, S., Lennard, C., Hewitson, B., Chang'a, L., Awange, J. L., Dosio, A., Ketiem, P., Nikulin, G., Panitz, H.-J., Büchner, M., Stordal, F., and Tazalika, L. (2013). Assessment of the performance of cordex regional climate models in simulating east african rainfall. *Journal of Climate*, 26(21):8453–8475.

Feng, S., Krueger, A. B., and Oppenheimer, M. (2010). Linkages among climate change, crop yields and mexico–us cross-border migration. *Proceedings of the National Academy of Sciences*, 107(32):14257–14262.

Findley, S. E. (1994). Does drought increase migration? a study of migration from rural mali during the 1983-1985 drought. *International Migration Review*, 28(3):539–553.

Funk, C., Peterson, P., Peterson, S., Shukla, S., Davenport, F., Michaelsen, J., Knapp, K. R., Landsfeld, M., Husak, G., Harrison, L., Rowland, J., Budde, M., Meiburg, A., Dinku, T., Pedreros, D., and Mata, N. (2019). A high-resolution 1983–2016 tmax climate data record based on infrared temperatures and stations by the climate hazard center. *Journal of Climate*, 32(17):5639–5658.

Gebrechorkos, S., Hülsmann, S., and Bernhofer, C. (2019). Changes in temperature and precipitation extremes in ethiopia, kenya, and tanzania. *International Journal of Climatology*, 39(1):18–30.

Gray, C. L. and Mueller, V. (2012). Natural disasters and population mobility in bangladesh. *Proceedings of the National Academy of Sciences*, 109(16):6000–6005.

Gráda, C. and O'Rourke, K. H. (1997). Migration as disaster relief: Lessons from the great irish famine. *European Review of Economic History*, 1(1):3–25.

Henderson, J. V., Storeygard, A., and Deichmann, U. (2017). Has climate change driven urbanization in africa? *Journal of Development Economics*, 124(C):60–82.

Herrero, M., Ringer, C., Van de Steeg, J., Thornton, P., Zhu, T., Bryan, E., Omolo, A., Koo, J., and Notenbaert, A. (2010). Climate variability and climate change and their impacts on kenya's agricultural sector. *International Livestock Research Institute ILRI Report*.

Hesse, C. and MacGregor, J. (2006). Pastoralism: drylands' invisible asset? pastoral civil society in east africa. Technical Report 142, IIED Issue Pape.

Hoerling, M., Hurrell, J., Eischeid, J., and Phillips, A. (2006). Detection and attribution of twentieth-century northern and southern african rainfall change. *Journal of Climate*, 19(16):3989–4008.

Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short- and long-run adjustments to environmental catastrophe. *American Economic Review*, 102.

Joseph, G. and Wodon, Q. (2013). Is internal migration in yemen driven by climate or socio-economic factors? *Review of International Economics*, 21(2):295–310.

Liebmann, B., Hoerling, M. P., Funk, C., Bladé, I., Dole, R. M., Allured, D., Quan, X., Pegion, P., and Eischeid, J. K. (2014). Understanding recent eastern horn of africa rainfall variability and change. *Journal of Climate*, 27(23):8630–8645.

Lind, J. T. (2019). Spurious weather effects. *Journal of Regional Science*, 59(2):322–354.

Long, J. and Siu, H. (2018). Refugees from dust and shrinking land: Tracking the dust bowl migrants. *The Journal of Economic History*, 78(4):1001–1033.

Lynham, J., Noy, I., and Page, J. (2017). The 1960 tsunami in hawaii: Long-term consequences of a coastal disaster. *World Development*, 94:106 – 118.

Lyon, B. and DeWitt, D. (2012). A recent and abrupt decline in the east africa long rains. *Geophysical Research Letters*, 39:2702–.

Marchiori, L., Maystadt, J., and Schumacher, I. (2012). The impact of weather anomalies on migration in sub-saharan africa. *Journal of Environmental Economics and Management*, 63(3):355 – 374.

McGuirk, E. F. and Nunn, N. (2020). Transhumant pastoralism, climate change, and conflict in africa. NBER Working Papers 28243, National Bureau of Economic Research, Inc.

Missirian, A. and Schlenker, W. (2017). Asylum applications respond to temperature fluctuations. *Science*, 358(6370).

Mogaka, H., Gichere, S., Davis, R., and Hirji, R. (2006). Climate variability and water resources degradation in kenya: Improving water resources development and management. *World Bank Working Paper*, pages 1–105.

Mueller, V., Gray, C., and Hopping, D. (2020). Climate-induced migration and unemployment in middle-income africa. *Global Environmental Change*, 65:102183.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the u. s. labor market. *The Quarterly Journal of Economics*, 118(2):549–599.

Nicholson, S. (2015). An analysis of recent rainfall conditions in eastern africa. *International Journal of Climatology*, 36.

Nicholson, S. (2017). Climate and climatic variability of rainfall over eastern africa. *Reviews of Geophysics*, 55(3):590–635.

Nyariki, D. and Amwata, D. (2019). The value of pastoralism in kenya: Application of total economic value approach. *Pastoralism 9*, 9.

Omondi, P., Ogallo, L. A., Anyah, R., Muthama, J. M., and Ininda, J. (2013). Linkages between global sea surface temperatures and decadal rainfall variability over eastern africa region. *International Journal of Climatology*, 33(8):2082–2104.

Piguet, E., Pécoud, A., and de Guchteneire, P. (2011). Migration and climate change: An overview. *Refugee Survey Quarterly*, 30(3):1–23.

Randell, H. and Gray, C. (2016). Climate variability and educational attainment: Evidence from rural ethiopia. *Global Environmental Change*, 41:111–123.

Reuveny, R. and Moore, W. H. (2009). Does environmental degradation influence migration? emigration to developed countries in the late 1980s and 1990s. *Social Science Quarterly*, 90(3):461–479.

Sedova, B. and Kalkuhl, M. (2020). Who are the climate migrants and where do they go? evidence from rural india. *World Development*, 129(C).

Silvestri, S., Bryan, E., Ringler, C., Herrero, M., and Okoba, B. (2012). Climate change perception and adaptation of agro-pastoral communities in kenya. *Regional Environmental Change*, 12.

Strobl, E. and Valfort, M.-A. (2015). The effect of weather-induced internal migration on local labor markets. evidence from uganda. *World Bank Economic Review*, 29(2):385–412.

Thiede, B. C. and Gray, C. L. (2017). Heterogeneous climate effects on human migration in indonesia. *Population and Environment*, 39(2):147–195.

Tobin, L. (2017). Essais sur l'urbanisation en afrique subsaharienne. Technical report.

Vicente-Serrano, S., Beguería, S., and López-Moreno, J. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23:1696–1718.

WFP (2018). Pastoral and agro-pastoral production system in the arid and semi-ari areas. Field practitioners guide no. 3.

Özden, C., Parsons, C. R., and Schiff, M. (2011). Where on earth is everybody? *World Bank Economic Review*, 25(1):12–56.

A Appendix

B Data

B.1 Population censuses

Table 14 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 15 and 16 show in more detail that it is necessary to exclude the Nyanza and North Eastern provinces from the analysis.

Table 15 (resp. 16) shows descriptive statistics of the population distribution within Nyanza (resp. North Eastern), at the scale of districts [40]. 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 16 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 19 which shows that 16% of the total observations are missing in the censuses. Such as Nyanza province, I can not

[40] districts matched over the censuses to the districts in 1989

conclude that these missing data are driven by some districts or sublocations and exclude North Eastern province from the analysis. Besides, Table 16 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 14: 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.
	(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 15: 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 16: Returns from the micro data in North Eastern province, 1989, 1999 and 2009

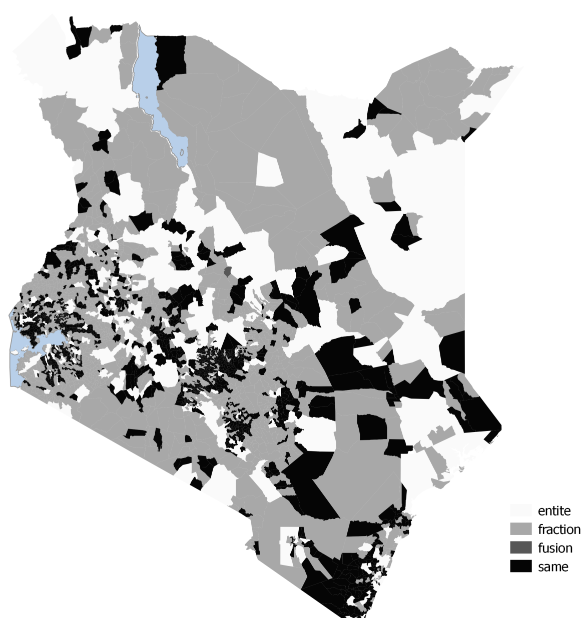
	North Eastern 1989			North Eastern 1999			North Eastern 2009		
	N(indiv.)	Mean(SD)	Min(Max)	N(indiv.)	Mean(SD)	Min(Max)	N(indiv.)	Mean(SD)	Min(Max)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
North Eastern	364923	3119 (2509.88)	236 (14869)	807198	2603.86 (2378.21)	112 (19718)	2301746	7238.19 (7477.14)	373 (65386)
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)

B.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 20 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 20: 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 17 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 17: 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

C Setting

C.1 Climatogy - descriptive statistics

Figure 21 plots the long-term average of monthly precipitations across provinces. It shows that all Kenyan provinces follow a bimodal seasonal pattern. Figure 22 plots the long-term mean of rainfall characteristics over the MAMJ season across several indicators. Table 18 gives the definition of each indicator.

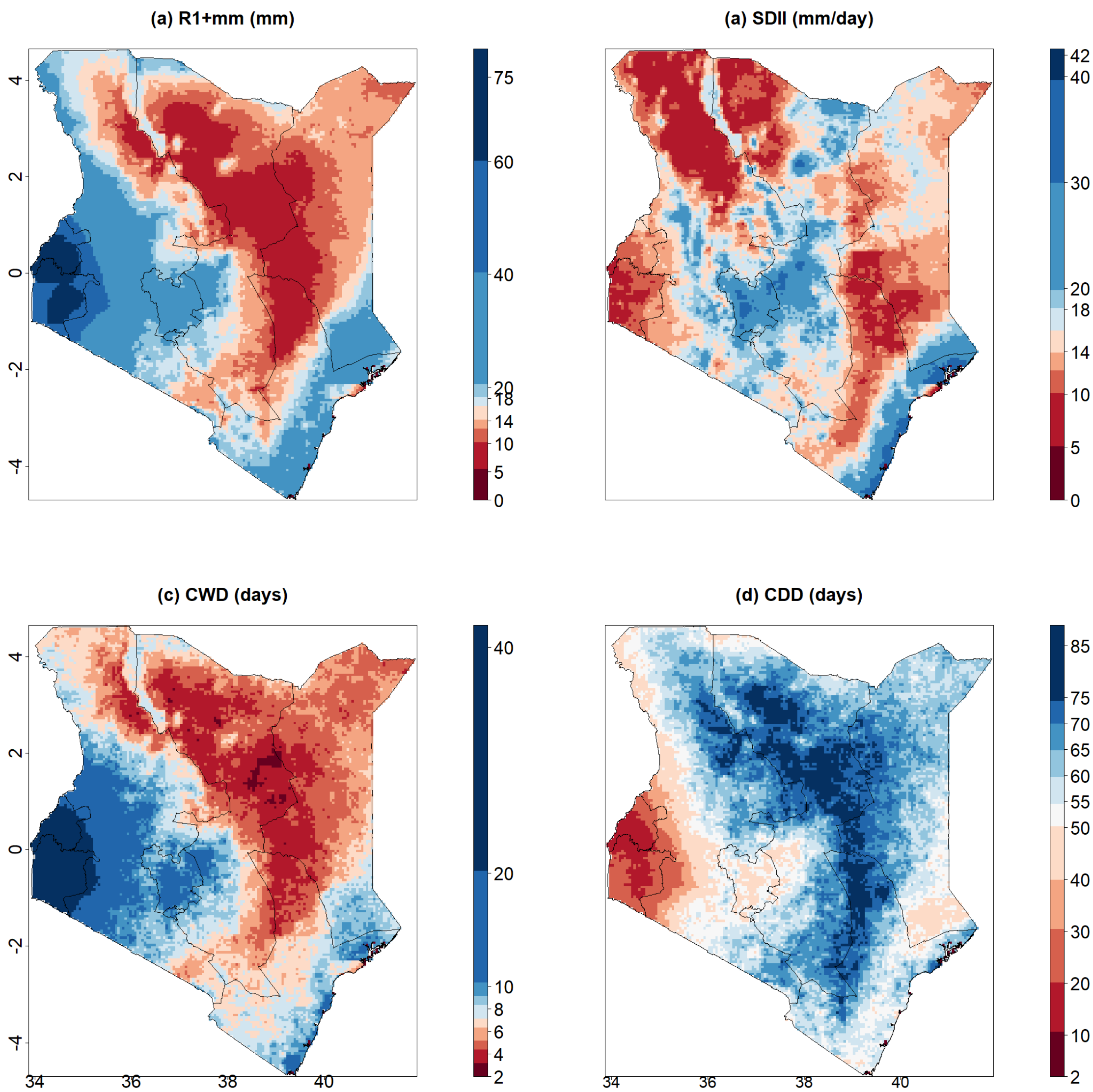
Table 18: Extreme precipitation indices and their ETCCDI and ECA definitions (from Gebrechorkos et al. [2019])

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 21: 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.

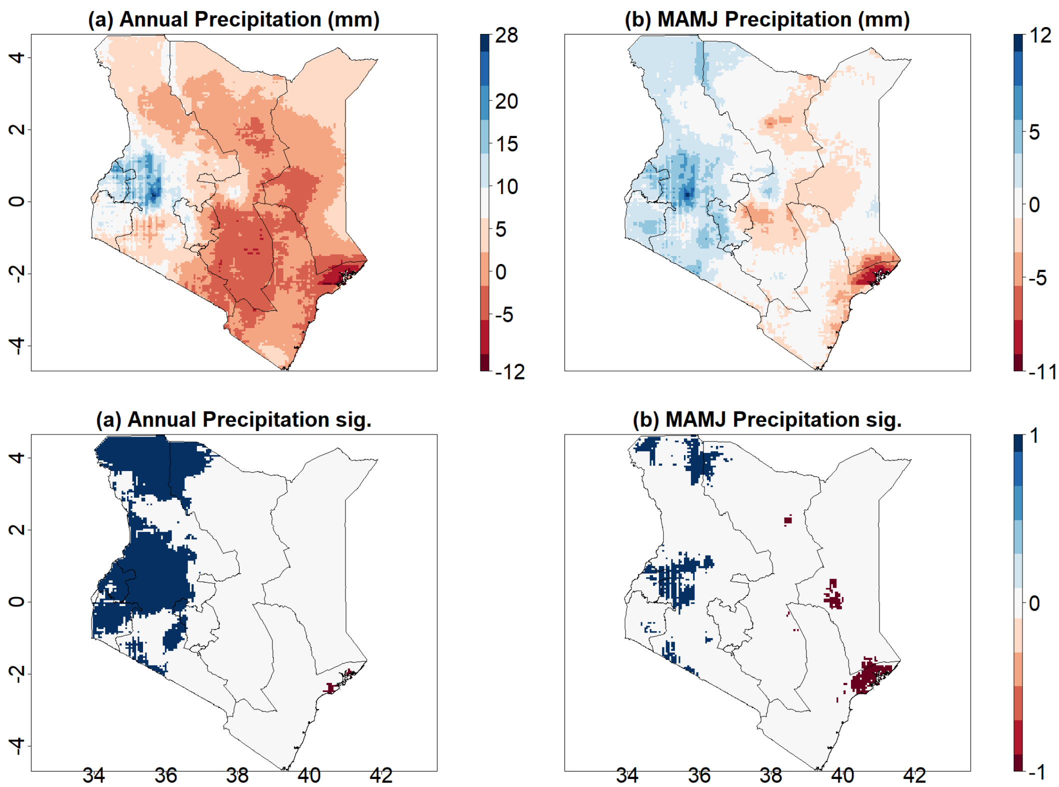
Figure 22: Spatial distribution of the long-term average of climate indicators



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.

C.2 Long-term evolution of rainfall

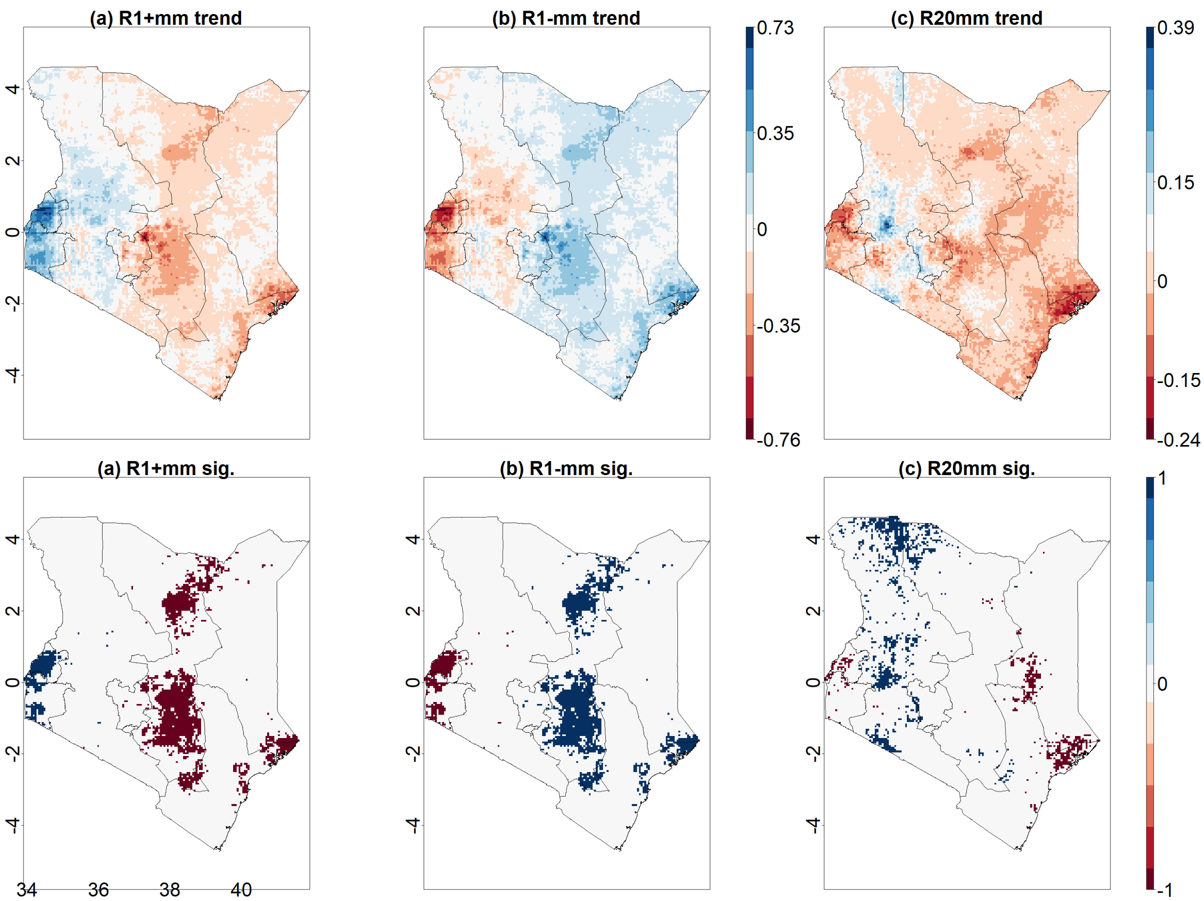
Figure 23: Long-term trends of climate indicators



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 23 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 24: 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 24 and Figure 25 show a long-term modification in precipitation characteristics, each indicator being described in Table 18. Figure 24 (a) and (b) are symmetric *de facto* [41]. 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

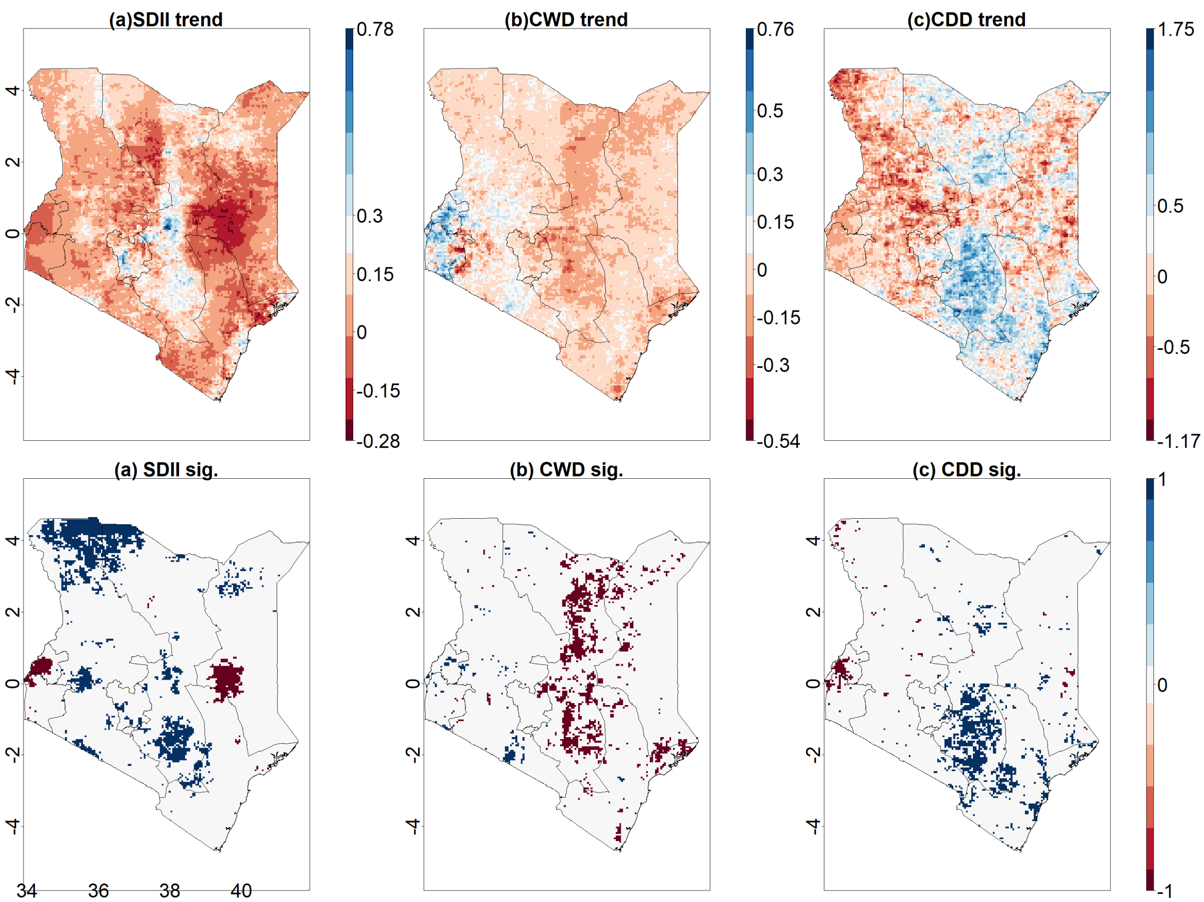
[41] By construction, R1+mm + R1-mm =122 days, i.e equals the total number of days during MAMJ

Intensity Index (SDII) (Figure 25 (a), between +0.3 and +0.78 mm per day), and significant decreases (resp. increases) in the Consecutive 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 23 (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]).

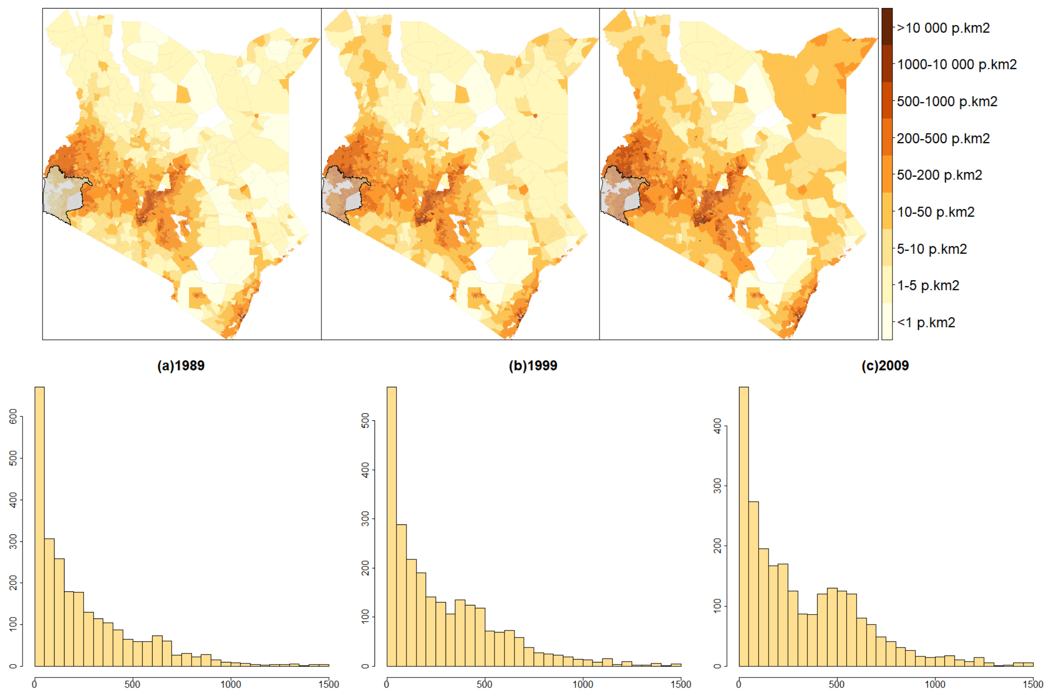
Figure 25: Long-term trends of climate indicators (3)



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.

C.3 Temporal and spatial variation of population and migration

Figure 26: 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.

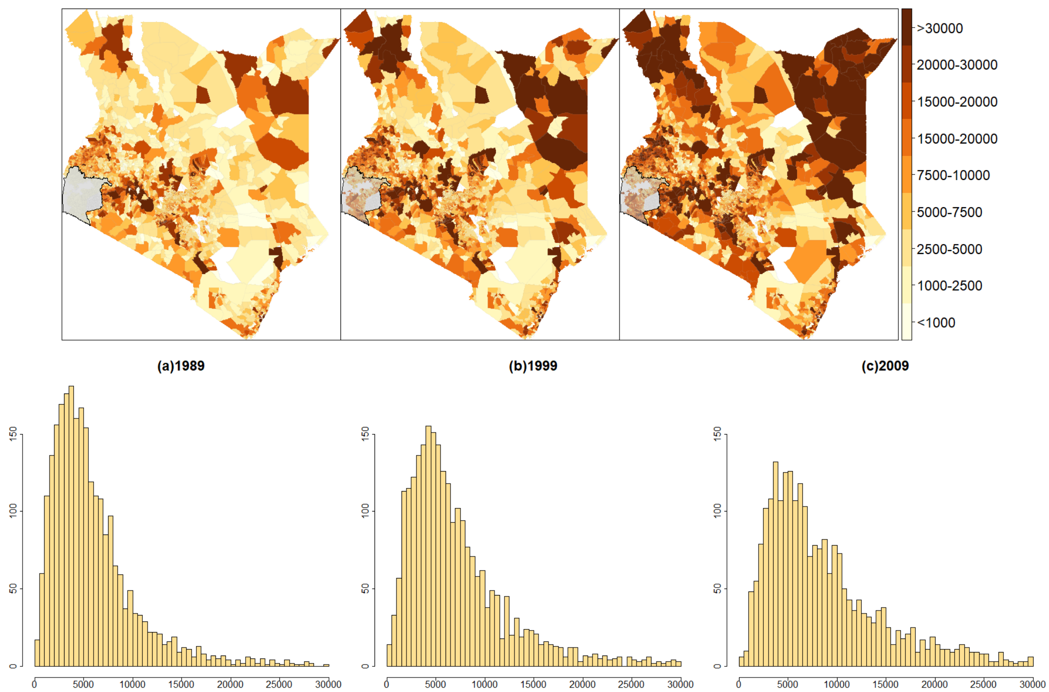
Table 19 and Table 20 display descriptive statistics of population, density, and population growth variables across period for each province [42]. Apart from Nyanza and North Eastern, The yearly (p.a) and decadal (DPGR) growth of the population are stable when calculated from different censuses [43]. 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 6 (Column (9)) is not erratic and is in line with the descriptive statistics of Table 20.

[42] The per annum population growth rate is defined as follows: $p.a = exp(\frac{log(\frac{pop(t_2)}{pop(t_1)})}{10}) - 1$

[43] As the Table 20 gives information about Nyanza and North Eastern provinces, the national average includes both provinces with discrepancies

Figure 27 plots the spatial distribution of the population size for each census wave, while Figure 26 plots the spatial distribution of the population density for each census wave.

Figure 27: 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 19: Descriptive Statistics of Province Population from microdata 1989, 1999 and 2009

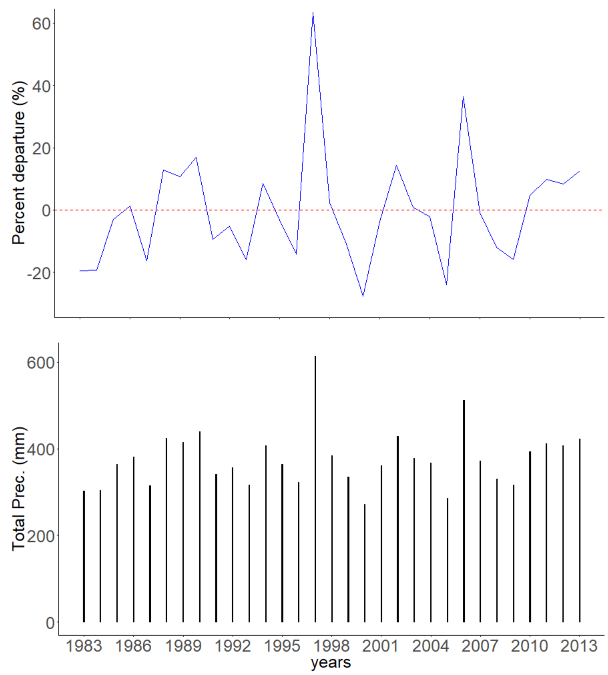
	1989			1999			2009			[1999-1989]	[2009-1999]
	Tot	Mean\(SD)	Min\(Max)	Tot	Mean\(SD)	Min\(Max)	Tot	Mean\(SD)	Min\(Max)	DPGR-1\(p.value)	DPGR-1\(p.value)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Kenya											
Population	17168480	5720.92	10	25953524	8648.29	82	36394457	12127.44	31	2.59	-0.61
		(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		
		(385.22)	(4632.43)		(348.88)	(2557.51)		(407.97)	(3533.53)		

Table 20: 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)		

C.4 Temporal and spatial variation of rainfall

Figure 28: Long-term of yearly rainfall departures

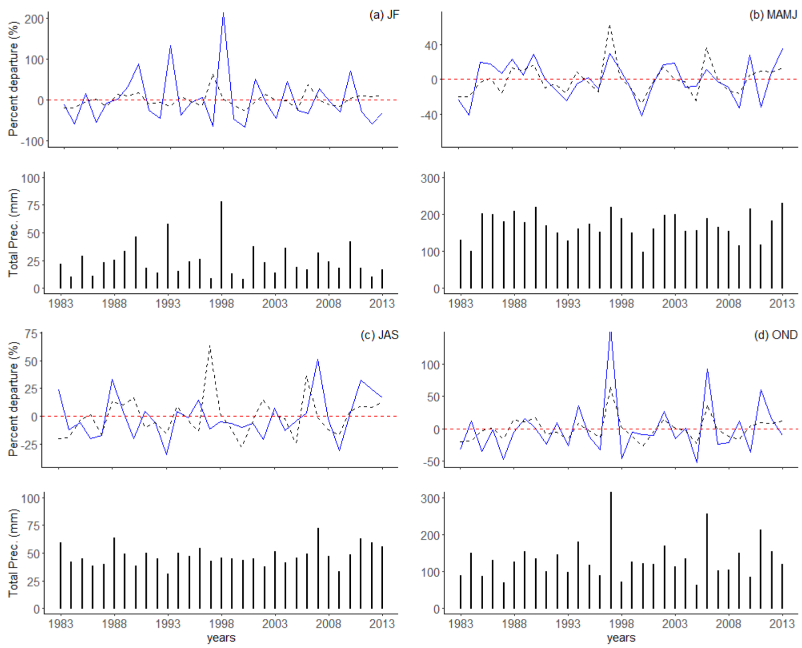


Notes: The Figures represent the time series of yearly rainfall departures (% above or below 1983-2013 mean) and time series of annual precipitations (mm)
Sources : author's elaboration on CHIRPS data.

Figure 28 plots the time series of annual precipitation departures aggregated over the country and interannual rainfall variability. Figure 29 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 29 (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 29 (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).

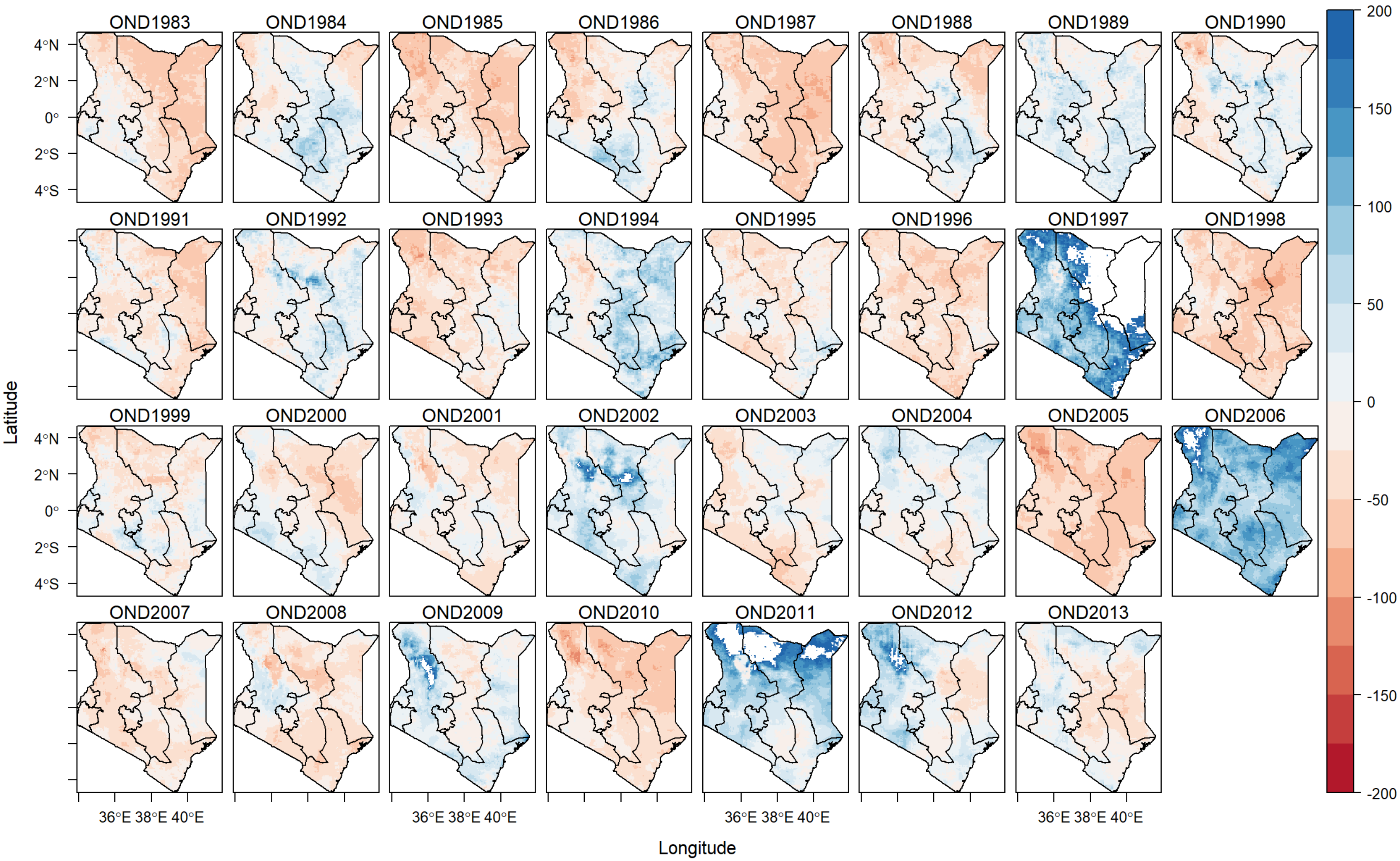
Figure 30 plots the spatial pattern of the departures of the rains over the short-rainy season from October to December (OND). Severe floods over the short-rainy season can be identified in 1997 and 2006.

Figure 29: 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 30: Rainfall percent departures of the short-rainy season (OND) from the 1983-2013 mean



D Main result

D.1 Climate variability

Table 21 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 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 [44], as suggested in Figure 5, the results might be biased due to multicollinearity.

[44] coefficient correlation: 0.0053

Table 21: Effects of the number of dry and wet rainy season 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.

E Heterogeneity

E.1 Gender and age brackets

Table 22 gives the results across gender and sublocation types restricting to the [15;65] years old cohort. Table 23 gives the results on the whole [0;69] cohort for comparison purposes.

Table 22: Effects of the number of dry rainy season 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 23: Effects of the number of dry rainy season on RDPGR

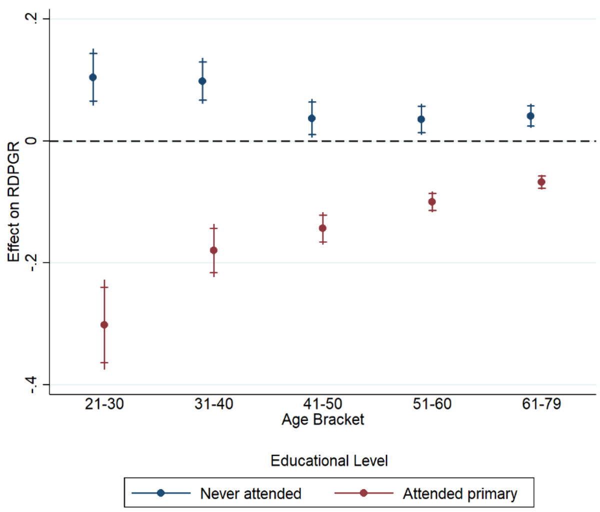
	All Kenya	Urban	Rural	Low Pastoralism	High Pastoralism
[0,69]	(1)	(2)	(3)	(4)	(5)
Number of dry years	-1.196***	-0.0900	-1.875***	-1.205*	-2.278***
	[0.382]	[1.065]	[0.459]	[0.669]	[0.761]
Number of dry years × density	0.000758	0.0000579	0.00494***	0.00369***	0.00895***
	[0.000491]	[0.000383]	[0.00112]	[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.

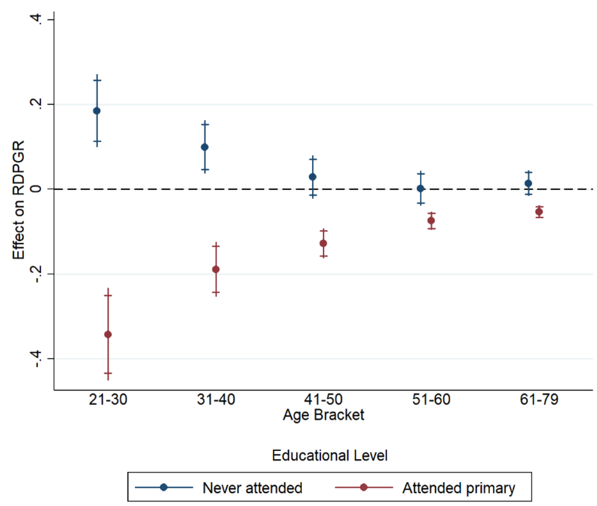
E.2 Education

Figure 31 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 31 shows that it is not the case, and the result across the skill distribution is not driven by any age effects.

Figure 31: Effect of the number of dry rainy seasons across educational level and age brackets



(a) All Kenya



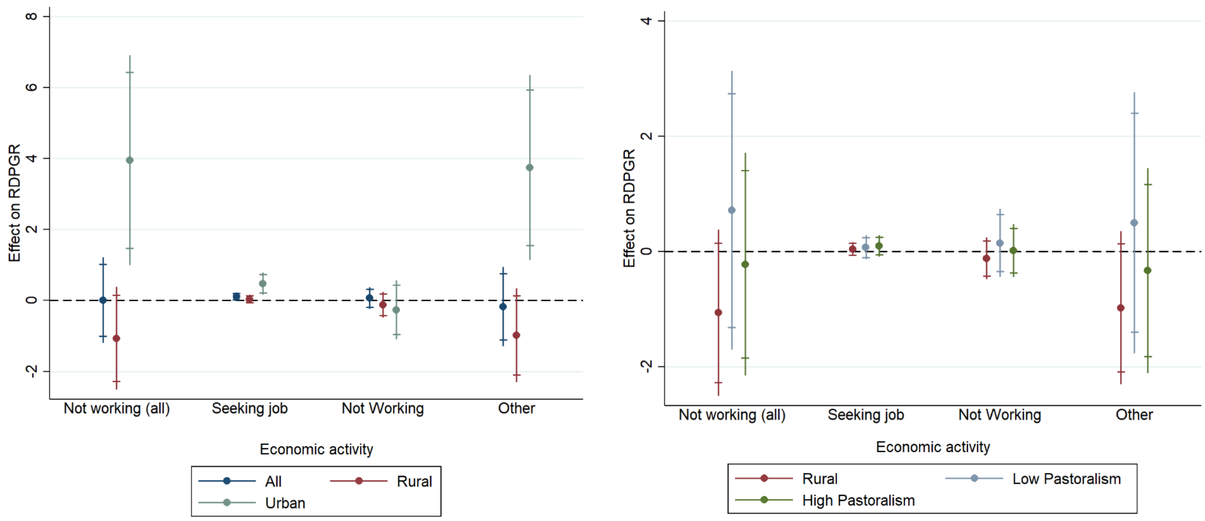
(b) Rural areas - Low-Pastoralism

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.

E.3 Economic Activity

Figure 32 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 32: 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.

F Robustness

F.1 Binary treatment

Table 24 gives the results of the binary treatment for the $DPGR_{[15,65]}$.

F.2 Common trend assumption

Table 25 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 9.2.

Table 24: 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.

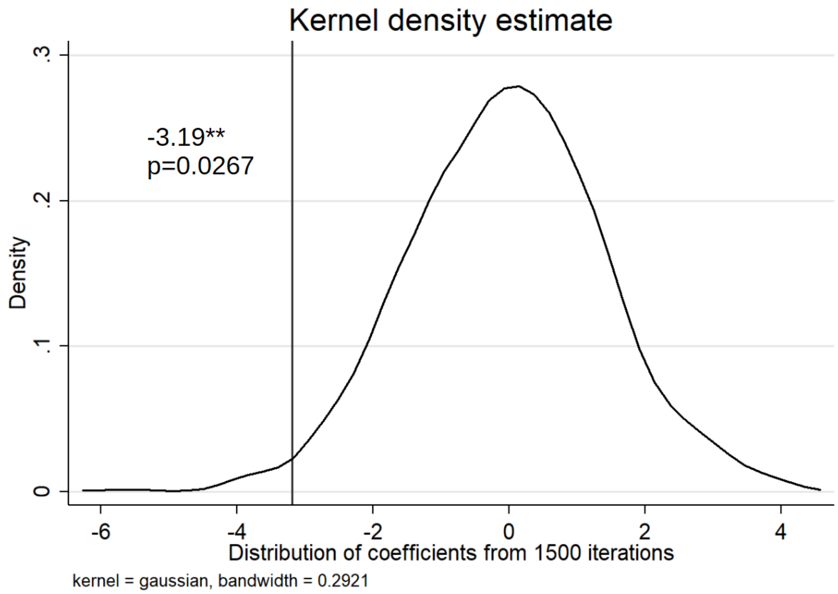
Table 25: Effects of the number of dry rainy season on the DPGR - Continuous and Binary treatment - Restricted Sample

	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.

F.3 Spurious correlation

Figure 33: 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 7 Column 1) and gives the new estimated p-value. *Sources:* Author's elaboration on CHIRPS and KNBS data.

F.4 Other climate indicators

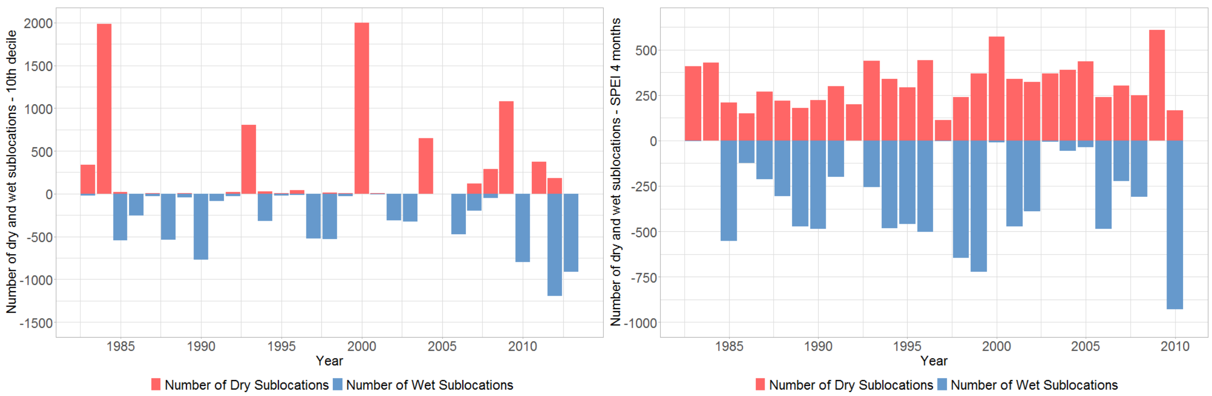
Figure 34 compares the distribution of the number of sublocations hit by droughts across years, according to the 10th decile and SPEI definitions. Table 26 displays the effects of the number of dry and wet years, based on cumulative rains over the short rainy season OND being under the 10th decile and 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 et al., 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 [45], which

[45] The SPEI has been computed using the R library SPEI developed by Vicente-Serrano et al. [2010]

gives the standard 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 34: 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 26: Effects of the number of dry short-rainy season (OND) on the DPGR

	All Kenya		Rural areas	
	(1)	(2)	(3)	(4)
Number of dry years	-0.222		0.313	
	[0.778]		[0.815]	
Number of wet years		-0.253		-0.442
		[0.755]		[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.

G 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.

G.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.

G.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 [46]. 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:

$$M_{od,t} = \frac{m_{od,t}}{pop_{oo,t}} = \alpha_0 + \alpha_1 D_{o,t} + \gamma_o + \delta_{d,t} + \beta_{od} + \epsilon_{od,t} \tag{3}$$

Where $M_{od,t}$ measures migration rates from the district of origin to the district of destination d. γ_o,$\delta_{d,t}$ and β_{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, β_{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.

[46] 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)

G.3 Results

Table 27 compares an OLS to a PPML estimator of equation 3. 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 27: 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 [0.00185]	0.000308 [0.00410]	0.0789 [0.133]	0.0826 [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, $^{*}p<0.1, ^{**}p<0.05, ^{***}p<0.01$.