Alma Mater Studiorum · University of Bologna

Department of Physics and Astronomy

Master Degree in Science of Climate

Assessing the Influence of Groundwater Storage Trends on Internal Migration and Conflicts

Supervisor:
Prof. Raya Muttarak

Co-supervisor:
Prof. Duccio Rocchini

Submitted by:
Filippo Galluzzi

Academic Year 2023/2024

A Sofia

Ai miei genitori

Abstract

Water resources are extremely important for the health of ecosystems, human life, and for a large part of human activities. Like many natural resources, their quantity and quality are imperiled by many threats, including pollution, anthropogenic overexploitation, and ongoing climate change, with clear consequences for society and nature. In particular groundwater, a resource that has not yet been extensively studied due to the difficulty in measuring it directly, is intimately linked to the health of ecosystems and provides many direct benefits to economic, industrial, and domestic human activities. The purpose of this thesis was to assess how quantity and variability of ground water resources, obtained from the simulation of ISIMIP3a model, are associated with human security as measured by conflict occurrence and internal migration. To this end, two different empirical approaches were used. The first approach consisted of a simple linear regression with fixed effects (administrative region and year) between some variables constructed from the amount of groundwater and, separately, the number of internal migrants and the number of conflicts for each region. As a second approach, two random forest models, a regression and a classification, were used in order to search for the possible presence of nonlinear relationships respectively for migrations and conflicts. In order to explore the heterogeneities, countries were classified into different categories by their geographical and socioeconomic characteristics. Based on the linear regression, it is found that the variations in ground water resources contribute to both an increase and a decrease in conflict and migration depending on a context. An additional finding from the random forest methods indicates that the relationships are nonlinear, although still within the subdivisions of the various classes.

Contents

1 Introduction

Climate change caused by the high concentration of greenhouse gases emitted into the atmosphere have already had considerable global impacts on human activities and ecosystems, particularly on water security, food production, and infrastructure. The risks that society is facing will largely depend on the socio-economic pathways that global policies manage to implement. Regardless of the chosen path and the adaptation measures adopted, some impacts due to current climate change are, however, inevitable. The global warming threshold of 1.5°C has been exceeded for twelve consecutive months [1], and even if the exceeding of the threshold is evaluated on a decadal time scale, this implies that achieving the Paris agreements is increasingly difficult. With very high confidence, as reported in the 'AR6 Climate Change 2022: Impacts, Adaptation and Vulnerability' report [2], exceeding global warming of 1.5°C will cause inevitable increases in damages to ecosystems and populations.

Of particular interest for this thesis work are the impacts that the climate change has on the variability of water resources, and how changes in water availability affect human security as measured by conflict and internal migration. In fact, from the report, it is noted that: "Climate change, including increases in frequency and intensity of extremes, have reduced food and water security, hindering efforts to meet Sustainable Development Goals (high confidence)." Therefore, not only are the ongoing climate changes having a significant impact on water and food security; they are even reducing the capabilities of policies to implement the SDGs and consequently limiting mitigation and adaptation measures. It is noteworthy that in 2022, about half of the world's population suffered from severe water scarcity (the lack of fresh water in relation to a region's needs) for at least part of the year. Much of the current literature focuses on surface water resources, largely neglecting groundwater resources. This is mainly due to the difficulty of correctly quantifying groundwater resources.

Groundwater provides not only direct benefits to economic or industrial activities or to domestic freshwater needs. It is intimately linked to the health of ecosystems and surface vegetation; consequently, high stress on this resource will also affect surface natural resources. One of the most significant impacts of climate change is that it alters the regularity of certain natural phenomena—regularity on which humanity has always relied for its activities, from planning agricultural endeavors to designing infrastructure. Quoting Pasini and Mastrojeni [3]: "With such uncertainty, society becomes insecure, chaotic, conflictual, and unstable. In practice, climate changes create instability in two different ways. First, because the climate of a more energetic atmosphere produces more violent phenomena that directly damage human physiology [...]. Second, because they render unpredictable the cyclical arrival of certain 'services' of nature on which we rely for an orderly organization of society and production." These changes are therefore also altering the distribution of groundwater, with consequences on agriculture, ecosystems, and domestic use. Natural resources will be distributed in regions different from the current ones. "Many will lose; some, even, will gain. But surely there will be a 'competition' to recover the lost resources" [3].

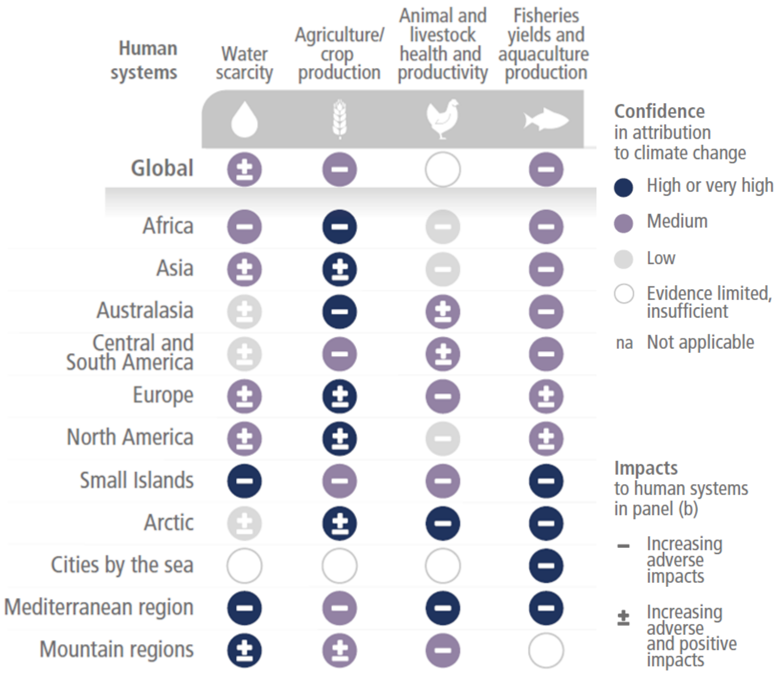

Figure 1.1: Climate change attributed impacts on water scarcity and food production, at global and regional level [2].

As can be observed from the figure 1.1, the effects of climate change on water scarcity do not have a univocal direction in any of the different regions, except for Africa and Mediterranean regions. This makes the evaluation of the impacts of water scarcity on society complex and ambiguous. Various studies have highlighted the correlation between natural resources and migrations and conflicts, although with often controversial and conflicting results. Although studies agree that climatic variables have a significant impact on these phenomena, they are less concordant on the intensities and different manifestations. In the short term, both violent conflicts and migrations will be driven by socioeconomic variables rather than climate ones. But, again from the report [2], with greater global warming, "impacts of weather and climate extremes, particularly drought, by increasing vulnerability will increasingly affect violent intrastate conflict (medium confidence)" [2].

Climate change is contributing substantially to humanitarian crises, especially in regions where climatic damages and high vulnerability meet. It is therefore necessary to integrate climate risk, and climate change, into future policy choices.

With this thesis work, an attempt was made to evaluate the impact of the quantity, variability, and temporal trends of groundwater on internal migrations and the number of conflicts in various regions of the world. To achieve this, two different approaches were used. The first consists of a simple linear regression with fixed effects between certain variables concerning groundwater and, separately, the rate of internal migration and the number of conflicts in these regions. Subsequently, two random forest algorithms were employed to assess the presence of non-linear relationships—which are not detectable by linear analysis—still focusing on the rate of internal migration and the presence or absence of conflicts. In this case, two different models were used: for internal migration, a regression algorithm using random forest was applied; for conflicts, since the problem involves evaluating the presence or absence of conflicts rather than their number, a random forest classification algorithm was utilized. Specifically, regarding the availability of groundwater values, an ISIMIP3a model (Inter-Sectoral Impact Model Intercomparison Project) was used, which provides projections on the future impacts of climate change on a global scale. Therefore, the values are not measured but simulated based on hydrological models and climatic inputs [4].

The chapters of the thesis are structured as follows. Chapter 2 briefly summarizes the results found in the literature regarding the relationships between natural resources, climate change, migrations, and conflicts. The first section is dedicated to groundwater, offering a brief overview of this resource. The second section describes the main results on the relationships between groundwater and migrations, their different manifestations, and the main variables. Finally, the last section is devoted to describing the relationships between natural resources, climate change, and violent conflicts. In Chapter 3, the preparation of the datasets used in the subsequent analyses is described, along with the operations performed and the names of the variables. Chapter 4 details the methods used in the linear and non-linear analyses, including tables and descriptions of the results. Chapter 5 presents the conclusions of the thesis. Appendix A gathers the names of the countries considered in the analyses, based on the data subdivision. Appendix B reports the codes used in the analyses conducted.

2 Literature Review

Climate change, caused by the high concentration of greenhouse gases produced by human activities, is having and will continue to have significant impacts on the dynamics of the Earth system as a whole and on human activities. Specifically, regarding the water resources present in the Earth's system, the latest IPCC report [5] states with high confidence that 'Continued global warming is projected to further intensify the global water cycle, including its variability, global monsoon precipitation, and very wet and very dry weather and climate events and seasons.' This intensification of the hydrological cycle will therefore lead to changes in local water availability, both positively and negatively, altering the fragile balances present in nature and human activities. Changes of this kind can exacerbate already stressed social situations, creating a vicious circle from which it is difficult to escape. To cite two recent examples, we can foresee what the future developments of the links between water resources and human activities will be. In 2020, the conflict in the Sudd region of South Sudan, in addition to the severe flooding, caused the forced displacement of about 2.6 million people [6]. These people have not yet been able to go back to their homes due to the persistence of extreme events in this region (for example, recently, the floods in September 2024). Considering the opposite phenomenon, namely the chronic scarcity of water resources due to changes in precipitation patterns and increasing drought, in Ecuador, to address the severe energy crisis caused by the lack of water resources in artificial reservoirs (hydroelectric plants provide 70% of the energy produced), the government had to implement scheduled blackouts, heavily impacting the region's economic activities [7]. From the IPCC report, it can be read that, with medium confidence, 'roughly half of the world's population currently experiences severe water scarcity for at least part of the year due to a combination of climatic and non-climatic drivers.' This can give an idea of the complexity of the water-society nexus, and in particular the water-migration-conflict nexus, where many variables act simultaneously, with different intensities and in different ways depending on the region studied.

2.1 Groundwater

The term groundwater refers to the water present underground, in the pores and fractures of rocks, soil, or sand. Groundwater is part of the hydrological cycle, which is the continuous movement of water through the various components of the Earth's system. It is described [8] as "all water found beneath the ground surface in the saturated zone." Groundwater is contained in aquifers, which are geological layers capable of storing and allowing the flow of water [9]. It can be found anywhere, at depths that vary depending on the geological characteristics of the region. The amount within an aquifer depends on recharge: when precipitation reaches the ground, a fraction of it evaporates directly into the atmosphere, another fraction flows on the surface until it reaches rivers or other water bodies, and the remaining part infiltrates the soil. Only the fraction that manages to reach the water table—i.e., the surface where the rocks or soil are permanently saturated with water—and that is not intercepted by the roots of vegetation, will, in the case of a surplus, increase the level of groundwater. Groundwater may also be recharged by leaks from water supply systems and through over-irrigation, when more water is supplied than is required by crops. About 70% of the water that reaches the soil through precipitation evaporates or is transpired by vegetation [10]. The importance of groundwater lies in the fact that it constitutes the largest liquid freshwater reserve on the planet, making it essential for the adaptation of vegetation and wildlife to climate variability. In terms of volume stored, groundwater is approximately one hundred times more abundant than surface water; despite that, in terms of renewal rate, the rate of groundwater renewal is approximately 30% of the rate of renewal of surface water [11]: this underlines the need for careful management of groundwater usage groundwater usage.

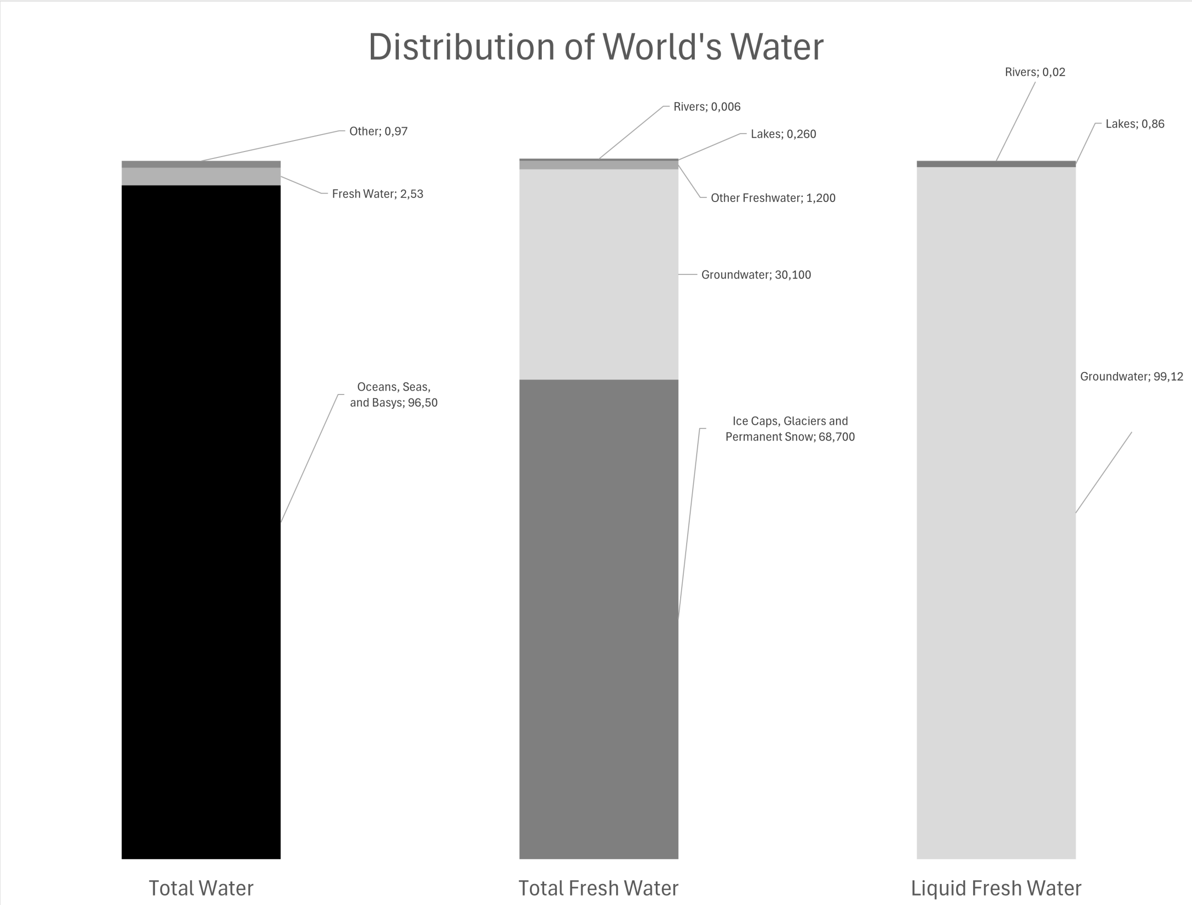

Figure 2.1: Distribution of Water on Earth (values are percentages) [12].

Groundwater is a vital resource for over two billion people [13][14], as it provides drinking water (36% of global supply), it is used for irrigating agriculture (42%), and supporting industry (24%) [15][16]. In addition, groundwater acts as a buffer and regulator of water flows, especially during periods of drought or low rainfall: groundwater helps maintain the baseflow in rivers and aquatic ecosystems. Annually, about half of the water discharged by rivers into the sea is groundwater. The availability of groundwater is changing primarily due to three factors: climate change, which affects precipitation patterns and thus aquifer recharge; natural variability; and the unsustainable extraction of groundwater for industrial, urban, or agricultural uses [17][13][18][16]. Due to the increase in precipitation variability, climate change impacts the stability of both surface and groundwater resources. Climate projections show significant variations in groundwater recharge across different regions of the world [19]. A distinction needs to be done between water shortage and water stress: the former refers to a lack of per capita water, while the latter refers to the scarcity of water based on available water resources. Water stress is thus a measure of the health of the aquifer, whereas water shortage also considers social and economic dynamics [20]. Per capita consumption did not increase significantly from 1900 to 2000, rising from around 209 m^3 per year to 230 m^3. Taking a look at the absolute terms however, global water consumption increased by 1,142 km^3 over the same period, from 359 km^3 to 1,500 km^3. With the increase in the global population, the demand for energy and food has also grown, impacting the global hydrological cycle, ecosystems, and the economies of countries in arid and semi-arid regions. The economic sectors responsible for the largest groundwater extractions are, respectively: irrigation, domestic use, electricity generation, livestock, mining, and manufacturing [21]. Irrigation accounts for about 70% of global freshwater withdrawals [18], and between 90-94% when considering blue water, which is surface and groundwater available for irrigation and human consumption, in regions such as South Asia and the Middle East [20]. Globally, groundwater provides drinking water directly or indirectly to half of the world's population: approximately 2.5 billion people rely exclusively on groundwater resources to meet their daily water needs [22]. Regions identified in the literature as critical due to groundwater overexploitation for agriculture [18] include the North China Plain, northwestern India, and California's Central Valley. Central Africa and the Amazon are also considered critical, where water availability is influenced by natural phenomena such as El Niño [17]. In addition, trends in groundwater storage are not the same everywhere: regions such as Northern Europe, Eastern Africa, and India show increases in recharge, while others like South America, the Mediterranean, and parts of the United States show significant reductions [19]. Regions with the highest water scarcity remain South Asia, North Africa, and the Middle East, while smaller proportions of the population affected by water scarcity are found in regions such as South America, Australia, and Southeast Asia [20].

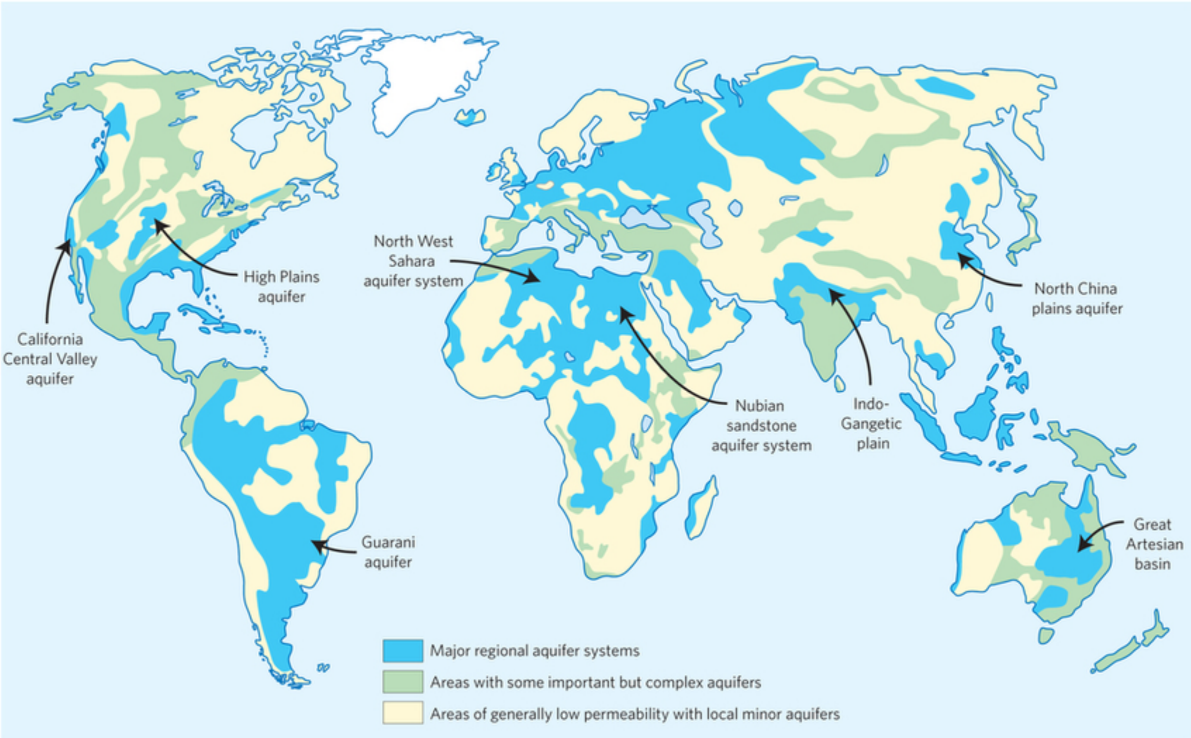

Figure 2.2: Simplified map of regional aquifer systems [18].

About half of the world's aquifers may have already reached or exceeded their tipping point, meaning the groundwater level below which natural recovery would take hundreds of years. Specifically, observations obtained through the GRACE mission show that 21 of the 37 major global aquifers are in exhaustion [16]. Of particular importance is the Nubian system, the largest non-renewable aquifer system on the planet (in this case, groundwater is referred to as fossil water), located in parts of Egypt, Libya, Sudan, and Chad. This system is considered to be under high stress due to the unsustainable withdrawal rates driven by population growth in these regions [9]. It is useful to consider the aridity level of different regions of the world, since it is a direct variable that influences the recharge and the health of water reserves. To this end, it is necessary to define the evapotranspiration (ET), as the combination of evaporation from small bodies of water from groundsurface and the transpiration of plants and vegetation in general, and the Potential Evapotranspiration (PET), as the amount of water that could be evapotranspired if there was an unlimited supply of water in the soil: it is higher in areas with higher temperatures, lower humidity and high wind speed, and it's closely related to aridity. Climate aridity is then defined as the ratio between precipitation and potential evapotranspiration. Arid regions are home to 2.5 billion people, grow roughly 44% of the world's food and raise about half of the world livestock [9].

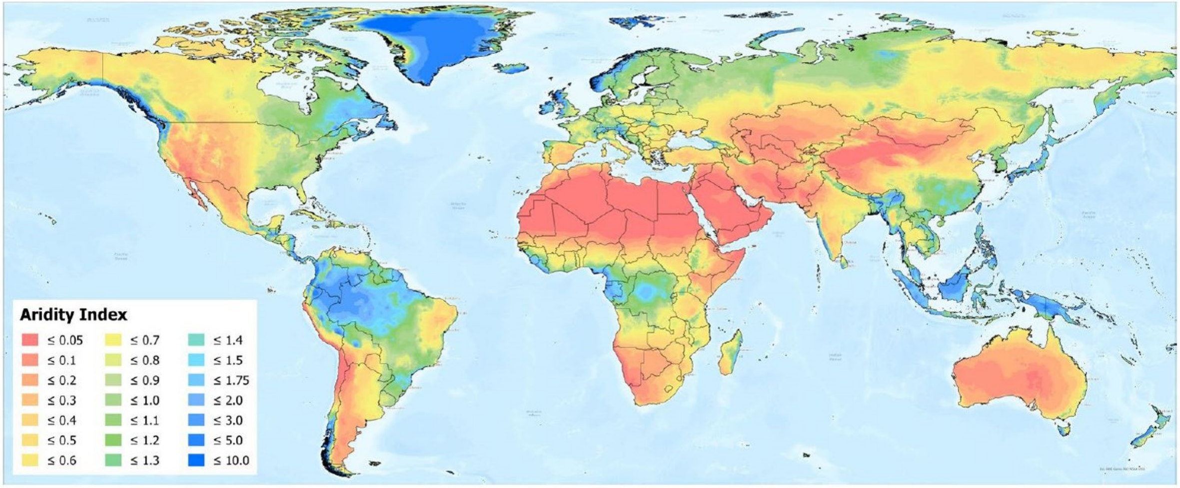

Figure 2.3: Global Aridity Index for the period 1970–2000 [23].

Groundwater for irrigation is a useful method to adapt to climate variations, but to do so, it is necessary to implement strategies that take into account the physical structure and mechanisms of both groundwater and surface water. Many countries have adequate regulations to sustainably manage water resources, but in many cases, what is lacking is the state's capacity to enforce them. One of the definitions of sustainability, with regard to groundwater use, is "the long-term maintenance of stable and high-quality reserves through inclusive and fair governance" [13]. It is therefore not just about maintaining physical stability of the system, but also about finding a balance in management among different stakeholders. Although groundwater is a global resource, its management must occur at the local level due to the differing water dynamics in various regions [15]. This is due to the different hydrological, social, and political characteristics present in each region. Nevertheless, in order to address sustainability issues and dynamics, it is necessary to implement a global perspective, given the structure of the hydrological cycle [13].

2.2 Migrations

The impacts of climate change on society directly or indirectly affects complex phenomena such as migration. Current literature indicates that climate change acts as a multiplier of existing threats rather than a direct trigger of migratory phenomena and violence [24]. Despite this, the IOM has proposed the following definition of environmental migrant: "Environmental migrants are persons or groups of persons who, predominantly for reasons of sudden or progressive change in the environment that adversely affects their lives or living conditions, are obliged to leave their habitual homes, or choose to do so, either temporarily or permanently, and who move either within their country or abroad" [25]. Migratory phenomena are typically influenced by many different factors that interact one with each other in complex ways [26][27]. While there is solid evidence regarding short-term shocks on migration, such as extreme precipitation, floods, and heatwaves, there are few studies in the literature on the long-term effects of climate change, for example chronic issues such as droughts and changing temperatures [28]. Between 0.5 and 3.9 billion people could be affected by these changes (water-related disasters such as floods, droughts, and glacier melt) by 2050 [29]. Migrations caused by sudden weather events, such as floods or droughts, are often temporary and internal, while gradual events can lead to permanent migrations, even across national borders. This is in contrast with other findings that state that migrations driven by environmental factors, in most cases, occurs within national borders rather than crossing them.

However, despite the significant impact of these events, it is rare for them to be the only determining factor: migration is a multifactorial process, where environmental factors interact with economic and social elements, such as the Human Development Index [30][27]. In regions with weak institutions and scarce resources (such as water), conflicts, violence, and political instability can be exacerbated by increased climate variability [31][32][33][34]. Approximately half of global migration occurs in areas characterized by high environmental vulnerability and low social adaptation capacity [27]. Globally, drought is one of the main factors driving the increase in conflicts and migrations, along with low levels of governance and pre-existing political tensions [32].

The same is true in East Africa, where a significant increase in the number of people crossing international borders as refugees has been recorded [35]. In this case, factors such as population growth and political stability also play a fundamental role, together with poor water resource management, mainly due to inefficient policies. For example, the reduction of the Aral Sea's surface, caused mainly by overexploitation area has led to large-scale migration: estimates indicate around 100,000 people [36]. Water shocks, particularly prolonged droughts, have proven to be a significant cause of migration: rainfall deficits can explain 10% of the increase in global migration during the period from 1970 to 2000 [37]. The reduction of Lake Chad, due in part to variability in precipitation patterns and in part to overexploitation, has led to mass migrations and conflicts between local communities over the remaining resources [36]. In particular, variability in water availability, water-related disasters, disruptions to water systems, and water pollution have been identified as the main channels through which the relationships between water and migration can be modeled [31]. These factors have varying effects on the type of migration, particularly in terms of the time frame (whether the migration is permanent or temporary). Sahel, Sub-Saharan Africa, the MENA region (Middle East and North

Africa), and various regions in Asia , are the most vulnerable regions to migrations related to water resources [31][32][33][38]. These areas are characterized by a strong dependence on agriculture and they heavily rely on natural resources, which are highly affected by water scarcity and rising temperatures. In particular, in agricultural regions, an increase in temperatures above 14°C (the optimal growth threshold for many crops) can decrease the productivity: in middle-income countries this phenomenon is linked to an increase of migrations [39]. On the other hand, the link between the worsening of productivity and an increase in migrations is not observed in low-income countries, since the former make it impossible to obtain the minimum financial resources necessary to migrate. For example, in countries dependent on agriculture, such as those in the Sahel and the MENA region, water scarcity increases the risk of poor crop production, therefore driving migration [31]. Rising temperatures and increasing evaporation reduce the water available for irrigation, creating a vicious cycle of decreasing production and increasing migratory flows, especially in the most vulnerable agricultural areas [40][34]. In Sub-Saharan Africa, for example, many economies rely predominantly on subsistence agriculture, making countries particularly vulnerable to climate variations. These anomalies effectively reduce rural wages, leading to increased internal migration flows from rural regions to urban areas, where economic conditions may be only slightly better [41]. Climatic variables, including total annual rainfall, average annual temperature, and exposure to temperatures above 30°C during the crop growing season, played an important role in explaining migration flows from the Sahel region of Africa to Italy between 1995 and 2009 [42]. These variables had a significant impact on crop yields, thus contributing to an increase in migration. Economic costs and legal restrictions often limit the possibility of international migration, especially in low-income countries [30]. In middle-income countries, extreme climate events can increase international migration, but in the poorest countries, paradoxically, migration tends to decrease, as climate shocks reduce the financial capacity of the population to migrate, leading to a sort of ”poverty trap”[26][37]. Income emerges as the most important factor in explaining observed migration, indicating that economic disparities are a powerful driver and that a certain minimum level of financial well-being is necessary to facilitate mobility [33][27]. Therefore, negative precipitation shocks tend to reduce migration in low-income countries.

Other phenomena that influences migrations are often overlooked. For instance: domestic water insecurity, meaning the difficulty in accessing safe and sufficient water resources at home. This phenomenon can have direct consequences on the psychological and physical well-being of people, pushing them to migrate, especially in areas with poor water infrastructure [43]. As already stated, the lack of sufficient water undermines economic opportunities, further exacerbating the vulnerability of populations already affected by other climate-related stresses. But populations affected by climate events are not passive actors. Many communities demonstrated resilience and high decision-making capacity in their adaptation strategies [26]. Even in the presence of severe and continuous environmental stress, many people choose to stay or migrate only as a last option. The decision to migrate, therefore, depends mainly on the socioeconomic context and the opportunities available to cope with climate change. Another fundamental element that plays a big role in explaining migratory events is the level of governance. Regions characterized by week institutions or internal conflicts are more exposed to the risks associated with water scarcity. In the MENA region, for example, water scarcity has exacerbated existing tensions and sharpened conflicts, as seen in Libya and Sudan. In particular, in Libya and Sudan, water scarcity is closely tied to decades of internal conflicts and the exploitation of natural resources, combined with weak governance [44]. In these contexts, water scarcity is described as a key factor in amplifying pre-existing tensions. Water has become a contested resource and is even used as a political pressure tool, as seen with the interruptions of the Great Man-Made River flow by armed groups in Libya.

Climate-related migration is a complex and multifactorial phenomenon. While climate change plays an important role in amplifying tensions, these phenomena must be analyzed within the broader context of the socioeconomic and political conditions of the affected regions. Resource scarcity, particularly water, acts as a powerful factor for migration, especially in agricultural areas. However, the level of governance and economic conditions remain critical factors in understanding the actual incidence of climate-related migration flows.

2.3 Conflicts

Regarding the dynamics linking the variability of natural resources to the increase in conflicts and violence in certain regions, the current literature distinguishes two possible pathways [45]: the first is the resource curse, which posits that the abundance of natural resources increases the likelihood of conflict due to competition among various actors. The second is resource scarcity, which argues that it is the lack of resources that primarily heightens the probability of conflict. This seemingly paradoxical description can be partly explained by distinguishing, on a case-by-case basis, the actors involved, the type of resource analyzed, the environmental context, and other specifics of the analysis method. In particular, the literature highlights that agricultural resource scarcity, influenced by climate variability and the resulting variability in water resources, plays a decisive role in amplifying the risk of conflict [45]. The relationship between climate events and armed conflicts is complex and depends on specific factors within the socio-political context. Recent studies show that the incidence of civil conflicts does not increase significantly following migrations induced by climate events, although in contexts with weak governance and pre-existing tensions, this link may be more evident [46]. For instance, in some African countries with high vulnerability, migration tends to increase internal conflict. Resource scarcity in these regions raises the likelihood of migration towards less vulnerable areas, increasing competition for resources and creating social tensions. Moreover, countries where investments in adaptation are implemented (such as improvements in governance or access to and distribution of natural resources) tend to be less vulnerable to conflicts related to climate change [47]. However, the relationship becomes

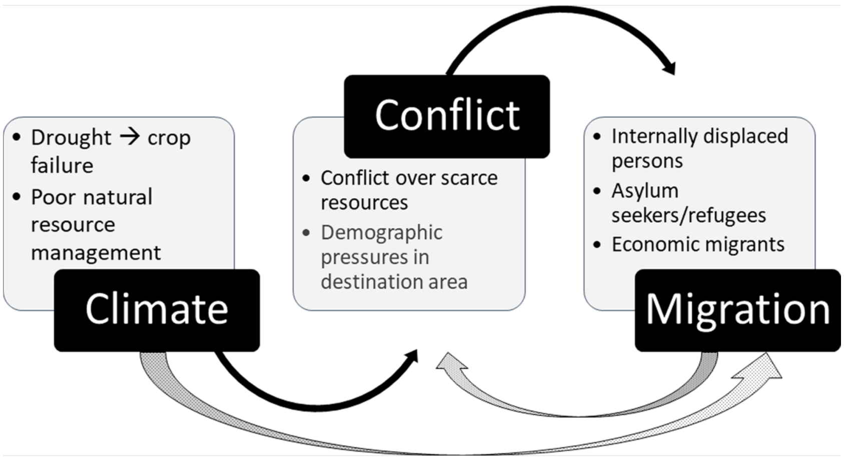

Figure 2.4: Interrelationships among Climate, Conflict and Migration [32].

less significant if the threshold for defining armed conflict is raised, suggesting that the effects of climate change on conflicts are limited to certain contexts. Therefore, the link between climate change and the risk of conflict remains confined to specific regions and countries with low levels of economic development and political marginalization [48]. For example, considering African countries, they recorded a significant increase in armed and non-armed conflicts during the period 1995-2017, with causes linked to resource scarcity and climate change—events that strongly affect African economies dependent on agro-pastoral activities [47].

Poor management of water resources, including inefficient agricultural policies and poor infrastructure, is the main cause of a region's water scarcity rather than physical scarcity itself [49][35]. The presence of strong institutions, and therefore a high level of governance capable of increasing the resilience of communities affected by environmental changes, limits the relationship between environmental resources and the probability of conflicts [45]. In fact, the countries most vulnerable to conflicts are partially democratic regimes because, unlike full democracies, they do not offer the same guarantees, and unlike autocracies, they are not sufficiently authoritarian to prevent riots and unrests [50]. The regions where the risk of conflict exacerbated by climate change overlaps with those identified for migratory phenomena are areas where water stress is already critical: the MENA region, Sub-Saharan Africa, and South America. In these cases, the literature has not reached unanimous conclusions on the causal relationships between variability in water resources and conflicts, but water stress is certainly an important factor to consider. For example, it has been observed that the risk of conflict in countries affected by ENSO doubles during El Niño events: 21% of the civil conflicts observed between 1950 and 2004 could be attributed to ENSO's effects on natural resources[51]. Many crises in the MENA region have been amplified by the inability of governments to meet the water needs of their populations. Terrorist groups like Boko Haram (active mainly in Nigeria, Cameroon, Niger, and Chad) and Al-Shabaab (active mainly in Somalia) exploit the control of water resources as a weapon to consolidate power [49]. In Yemen, the aquifer used to supply the capital Sana'a is running dry, and some water infrastructure has been deliberately damaged to exacerbate water scarcity and put pressure on the population. The same is observed in Syria [52]. It has been observed that a reduction in local water mass triples the likelihood of local social conflicts. Despite that, a correct management of the access to groundwater and surface waters can mitigate these effects, namely the impacts of water scarcity on conflicts [53]. As a negative example, in Africa the construction of dams or other water infrastructure (as irrigation systems) led to a changing in the distribution of water resources, that in some cases sharpened tensions among different groups of users over the control of the water resources [54]. The Mediterranean region in general is one of the areas most vulnerable to climate change, with projections estimating increases in temperature and decreases in precipitation—two factors that will be responsible for exacerbating issues like desertification, water scarcity, and food insecurity [55]. These are just some of the variables also identified in the report published by the European Commission on the impact of climate change and environmental degradation on peace, security, and defense [56]. In addition to these are extreme weather events, rising temperatures and sea levels, loss of biodiversity, and pollution. As indicated in the report, twelve of the twenty countries identified as the most vulnerable and least prepared for climate change were in conflict in 2020.

Water can serve as a diplomatic lever between states that share water basins; however, the worsening of crises related to water resources makes negotiations increasingly difficult and complex, and diplomatic efforts more challenging [49].

3 Datasets Description

In this paragraph I report the descriptions of the datasets that were used for the project. In brackets are the names of the datasets.

Shapefile (shp)

This dataset [57] is a shapefile that contains the geometries of the administrative regions where household data were recorded. To handle temporal changes in borders, this dataset has been originally harmonized to create temporally stable data. The dataset contains the following variables:

- CNTRY_NAME, the name of the country (renamed to 'country').
- ADMIN_NAME, the name of the administrative region (renamed to 'region').
- CNTRY_CODE, a digit that corresponds to the country name (removed).
- GEOLEVEL1, a digit that corresponds to the region name (renamed to 'orig').
- BPL_CODE, indicates the person's country of birth (removed).
- geometry, the geometry multi-polygon.

The dataset contained 20 empty geometries and 11 invalid geometries that were removed. Regarding invalid geometries, this means that they do not comply with the topological rules defined for this type of geometry. Specifically, two issues have been identified: incorrectly intersecting polygon geometries and duplicate vertices. The removed regions had no name. The names of the countries with empty geometries are: Argentina, Ecuador, Fiji (two regions), Honduras, Israel, Lesotho, Mali, Nicaragua, Nigeria, Paraguay, Slovenia, South Africa, Spain, Uganda, United States, Ethiopia, France, Morocco, and Uruguay. The names of the countries with invalid geometries are: Argentina (two geometries), Bangladesh, Belgium, Cambodia, Canada, Fiji, India, Côte d'Ivoire, and the Netherlands. Since some of the regions did not contain administrative regions, the NA values in this column were replaced simply by the name of the country. Finally, the CRS was manually set ("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"). This is the plot of the sub-national regions:

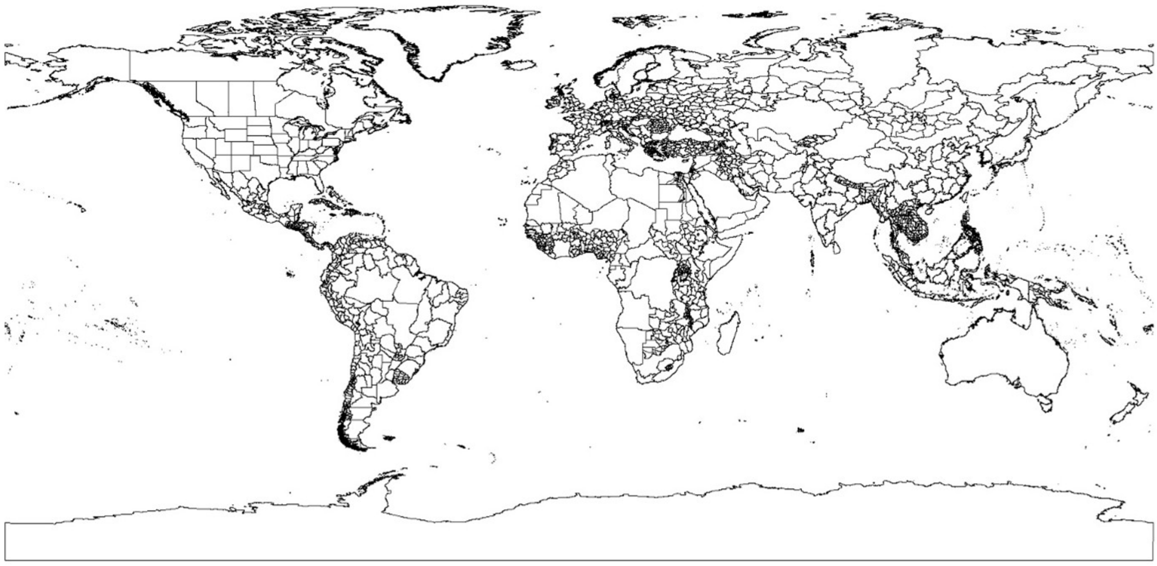

Figure 3.1: Administrative regions map.

Groundwater Storage (gws)

This dataset [58] is a raster data that contains monthly global values of groundwater storage (Kg/m^2) with a spatial resolution of 0.5°. These values are modeled data, obtained by ISIMIP3a using the CWatM impact model and the GSWP3-W5E5 climate forcing. The time period spans from 1901 to 2019. The CRS was set to be the same as the shapefile. Then, an annual average was calculated since the dataset contained monthly data. Afterward, the groundwater values were merged with the shapefile geometries to obtain an annual value of groundwater storage for each region in the shapefile. The variables 'country', 'region', and 'orig' were added, and this new dataset was reshaped into a long format. Finally, the names of the observation periods were changed, and only the period spanning from 1958 to 2019 was selected. Afterward, the groundwater values were merged with the shapefile geometries to obtain an annual value of groundwater storage for each region in the shapefile. The variables 'country', 'region', and 'orig' were added, and this new dataset was reshaped into a long format. Finally, the names of the observation periods were changed, and only the period spanning from 1958 to 2019 was selected. This is a section of the new dataset:

year	country	region	value	orig
1971	Brazil	São Paulo	1107.079	76035
1971	Canada	Nova Scotia	1104.930	124012
1971	Kenya	Western	1348.372	404008
1972	Brazil	São Paulo	1184.437	76035
1972	Canada	Nova Scotia	1145.043	124012
1972	Kenya	Western	1296.110	404008
1973	Brazil	São Paulo	1222.988	76035
1973	Canada	Nova Scotia	1191.834	124012
1973	Kenya	Western	1303.246	404008
1974	Brazil	São Paulo	1291.833	76035

Figure 3.2: Section of the Groundwater Storage dataset.

Population (pop)

This dataset [59] is a raster that contains the number of people per cell (with a spatial resolution of 30 arc seconds) derived from CIESIN GPWv4.11. The period estimates span from 1975 to 2020 with 5-year intervals and include two projections for 2025 and 2030. This raster has been merged with the administrative regions using addition instead of averaging. A new dataset has been created with the number of people per region and per year. Here is a section of the resulting dataset:

year	country	region	pop
2020	Argentina	City of Buenos Aires	2750869.00
1990	Armenia	Shirak	281005.72
2003	Botswana	Lobatse	62679.12
2012	Turkey	Çanakkale	525461.50
2016	Vietnam	Quang Binh	866001.25
2005	Zimbabwe	Mashonaland West	1256513.38
1993	Austria	Wien	1533169.75
1976	Chile	Petorca	54431.04
2020	Afghanistan	Afghanistan	38831220.00

Figure 3.3: Section of the Population dataset.

Potential Evapotranspiration (pet)

This dataset [60] is a raster that contains the values of PET for each cell (with a spatial resolution of 30 arc seconds) derived from the CGIAR Consortium for Spatial Information (CGIAR-CSI), for the year 2019. This year was selected because the dataset provides complete data with higher resolution, allowing for more detailed and accurate analysis. The procedure is the same as for the groundwater data: set the same CRS as the shapefile, then merge the raster with the administrative regions. After that, the values were averaged for each region of each country and sorted. This dataset has been used to divide the data into two subsets: high and low potential evapotranspiration countries. Figure 3.4 shows a section of the resulting dataset:

country	pet
Finland	594.2361
Denmark	731.1081
Montenegro	1054.3097
Macedonia	1192.4620
Tajikistan	1210.1937
Chile	1377.9034

Figure 3.4: PET Dataset.

Governance (gov)

This dataset [61] contains the governance effectiveness ranking for the year 2005. The year 2005 was chosen as it represents approximately a midpoint in the observation period, providing a balanced view of governance effectiveness trends. Government effectiveness "captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from -2.5 (the lowest level of governance) to +2.5 (the highest level of governance). This dataset has been used to divide the countries into three categories: low, medium, and high level of governance. Table 3.5 shows the list of the countries in the low governance category:

```
 [1] "Ukraine"              "Algeria"                  "Vanuatu"              "Bolivia"
 [5] "Uganda"               "Brazil"                   "Lao PDR"              "Paraguay"
 [9] "Pakistan"             "Niger"                    "Togo"                 "Zambia"
[13] "Gambia, The"          "Solomon Islands"          "Eswatini"             "Russian Federation"
[17] "Mauritania"           "Mozambique"               "Gabon"                "Ethiopia"
[21] "Tajikistan"           "Bangladesh"               "Timor-Leste"          "Djibouti"
[25] "Papua New Guinea"     "Burkina Faso"             "Malawi"               "Sao Tome and Principe"
[29] "Belarus"              "Honduras"                 "Cameroon"             "Iran, Islamic Rep."
[33] "Kyrgyz Republic"      "Lesotho"                  "West Bank and Gaza"   "Guatemala"
[37] "Nepal"                "Guinea"                   "Suriname"             "Madagascar"
[41] "Nigeria"              "Nicaragua"                "Angola"               "Bosnia and Herzegovina"
[45] "Turkmenistan"         "Sierra Leone"             "Mali"                 "Zimbabwe"
[49] "Burundi"              "Equatorial Guinea"        "Iraq"                 "Congo, Rep."
[53] "Chad"                 "Liberia"                  "Lebanon"              "Korea, Dem. People's Rep."
[57] "Guinea-Bissau"        "Comoros"                  "Myanmar"              "Venezuela, RB"
[61] "Sudan"                "Central African Republic" "Eritrea"              "Congo, Dem. Rep."
[65] "Syrian Arab Republic" "Libya"                    "Afghanistan"          "Somalia"
[69] "Haiti"                "Yemen, Rep."              "South Sudan"
```

Figure 3.5: List of Low Governance Countries.

Income

Regarding the income values, the package 'WDI' was used. This package contains data from the World Bank. The data used is "NY.GDP.MKTP.PP.KD," representing GDP at Purchasing Power Parity (PPP) in constant prices for the year 2005 (for the same reason as the previous dataset). The selected variable 'income' represents the country's income classification. This dataset was used to divide the data into four categories: low, low-middle, middle-high, and high-income countries. Here are the names of the countries in the lower-middle income dataset:

```
 [1] "Algeria"               "Angola"             "Bangladesh"       "Benin"            "Bhutan"
 [6] "Bolivia"               "Cabo Verde"         "Cambodia"         "Cameroon"         "Comoros"
[11] "Congo, Rep."           "Cote d'Ivoire"      "Djibouti"         "Egypt, Arab Rep." "El Salvador"
[16] "Eswatini"              "Ghana"              "Haiti"            "Honduras"         "India"
[21] "Indonesia"             "Iran, Islamic Rep." "Kenya"            "Kiribati"         "Kyrgyz Republic"
[26] "Lao PDR"               "Lebanon"            "Lesotho"          "Mauritania"       "Micronesia, Fed. Sts."
[31] "Mongolia"              "Morocco"            "Myanmar"          "Nepal"            "Nicaragua"
[36] "Nigeria"               "Pakistan"           "Papua New Guinea" "Philippines"      "Samoa"
[41] "Sao Tome and Principe" "Senegal"            "Solomon Islands"  "Sri Lanka"        "Tajikistan"
[46] "Tanzania"              "Timor-Leste"        "Tunisia"          "Ukraine"          "Uzbekistan"
[51] "Vanuatu"               "West Bank and Gaza" "Zimbabwe"
```

Figure 3.6: List of Lower Middle Income Countries.

Migration (gws_migr)

This dataset [33] contains information on internal migration within a country obtained from census micro data (kindly made available by the authors of this paper). The dataset contains the following variables:

- orig, the GEOLEVEL1 variable.
- year, the year of census data collection.
- year_cat10, a factor variable with decade information (removed).
- country_name, the name of country (renamed to ‘country’).
- population, total population size according to census.
- mig_interval, time interval over which the migration measure was estimates (1 or 5; renamed to ‘interval’).
- worldregion, South America, Northeastern Europe & Central Asia, Africa & Middle East, East Asia& Pacific, North America, Central America & Caribbean, Southern Europe, South Asia (removed).
- flow, the total number of internal migrants in a region and in the considered migration time interval (1 or 5).
- flow_annual, the number of migrants per year (flow/mig_interval; removed).
- outflow_rate_annual, the number of migrants per year/population (removed).

Then, this dataset was merged with the groundwater and population datasets. The dataset only included population values for each region for the year which a specific census was collected. Only countries for which migration data were recorded have been selected:

```
 [1] "Mexico"       "Panama"          "Ecuador"            "Costa Rica"          "Colombia"  "Guatemala"
 [7] "Argentina"    "Brazil"          "Thailand"           "United States"       "Greece"    "Haiti"
[13] "Indonesia"    "Nicaragua"       "Paraguay"           "Honduras"            "Uruguay"   "Bolivia"
[19] "Cameroon"     "Fiji"            "Romania"            "Benin"               "Kenya"     "Papua New Guinea"
[25] "Botswana"     "Canada"          "Dominican Republic" "Ireland"             "Portugal"  "Venezuela"
[31] "Chile"        "Jamaica"         "India"              "Israel"              "Senegal"   "Vietnam"
[37] "China"        "Mauritius"       "Philippines"        "Trinidad and Tobago" "Zambia"    "Malaysia"
[43] "Spain"        "United Kingdom"  "El Salvador"        "Egypt"               "Guinea"    "Mozambique"
[49] "Mali"         "Belarus"         "Kyrghzstan"         "Ghana"               "Mongolia"  "Armenia"
[55] "Nepal"        "South Africa"    "Poland"             "Slovenia"            "Tanzania"  "Cambodia"
[61] "Morocco"      "Peru"            "Malawi"             "South Sudan"         "Sudan"     "Russia"
[67] "Togo"         "Cuba"            "Suriname"           "Myanmar"             "Uganda"    "Sierra Leone"
```

Figure 3.7: List of countries in which migration data were collected.

Then, a new variable containing the migration data was created by dividing the flow (the number of internal migrants in a region) by the population. This variable was normalized using the logarithmic method: log(1+x). Finally, regions where migration flows exceeded the population were removed; this occurred in one region of Portugal:

	year	country	region	value	population	interval	flow	pop	migrants
1	1981	Portugal	Other Center	0.0001252538	579720	1	539440	553284.4	0.9749778
2	1991	Portugal	Other Center	0.0001657909	538620	1	527280	516799.2	1.0202801
3	2001	Portugal	Other Center	0.0002170192	521240	1	554340	507651.8	1.0919688
4	2011	Portugal	Other Center	0.0001554136	484280	1	728520	451695.1	1.6128580

Figure 3.8: List of regions removed from the migration dataset.

Conflict (gws_events)

This dataset (UCDP Georeferenced Event Dataset (GED) Global version 24.1) contains the coordinates of individual events of organized violence for the period 1989-2022. For the purpose of this thesis, only the following variables were selected:

- country, the country in which the event occurred.
- year, the year in which the event occurred.
- type_of_violence, 1 (state-based conflict); 2 (non-state conflict); 3 (one-sided violence).
- latitude and longitude, the coordinate of the event.
- best, the best estimate of the causalities.

The variable ‘type_of_violence’ was renamed to ‘type’. Then, the values of the variable ‘type’ were renamed as it follows:
1 : state
2 : Nstate
3: onesided
The CRS was set to be the same as the administrative regions. Then, an intersection between the coordinates of the events and the regions of the shapefile was performed to obtain a new variable representing the number of events (by type) for each region and year ('conflicts'). The data from 2020, 2021, and 2022 were removed due to the lack of groundwater data for the same period. Next, this dataset was merged with the groundwater and population datasets. For the regions and years in which no conflicts occurred, a value of 0 was assigned instead of NA. Since the conflict data starts from 1989, only the subset with years > 1978 was selected. Regions with NA values for groundwater storage were removed. Additionally, only regions with populations larger than 2,000 people were selected to maintain consistency. The list of the removed regions can be found in the Appendix A, Figure A.1.

Finally, a new variable was created to represent the number of events, irrespective of their type (‘count’).

New Variables

The following variables are the same for the two datasets gws_migr and gws_confl.

- Groundwater storage (gws) value per capita: $value = \frac{value_t}{pop}$
- Normalization of gws per capita: $n\_value = log(1 + value)$
- Averages for gws for 1-5-10 years (with y as the considered year): $gws\_avgk = \frac{value_y}{value_{y-k}}$, with $k = 1, 5, 10$
- gws logarithmic return: $logretk = log(\frac{value_t}{value_{t-k}})$, with $k = 1, 5, 10$
- Anomalies for 1-5-10 years: $gws\_anomk = \frac{n\_valuek - mean\_region}{std}$, with $mean\_region$ and std calculated for each region for the period (1980-2010).
- Coefficient of variations for 1-5-10 years: $\frac{gws\_stdk}{mean\_region} * 100\%$, with $k = 1, 5, 10$.

After that, the values for logret, anomalies, and coefficient of variation with NA values—those for which the values of GWS in the previous years were 0—were changed to 0.

4 Analysis

4.1 Methods

As a first approach, a generalized linear regression with fixed effects (specifically, region and year) was performed, applied independently for each variable related to ground water storage. These are the forumlae, respectively, for migrations and conflicts:

$$\log(1 + \text{migr}_{i,t}) = \alpha + \beta \cdot \mathrm{X}_{i,t} + \gamma_{\text{region}_i} + \delta_{\text{year}_t} + \epsilon_{i,t} \tag{4.1}$$

$$\log(E[\text{count}_{i,t}]) = \alpha + \beta \cdot \mathrm{X}_{i,t} + \gamma_{\text{region}_i} + \delta_{\text{year}_t} \tag{4.2}$$

where:
migr is the number of internal migrants in a region in a specific year, divided by the population of that region;
$E[\text{count}_{i,t}]$ is the expected value of the number of conflicts (since the distribution quasi-Poisson was used);
X is the considered independent variable;
α is the intercept;
β is the coefficient that measures the effect of X on migrations / conflicts;
γ and δ are the fixed effects respectively for region and year;
ϵ is the casual error, normally distributed.

The fixed effects allow controlling for variations that are specific to each region and year. Specifically, the independent variables are:

1. The value of groundwater storage for the given year, as well as the average over the last 5 and 10 years.
2. The anomalies for the given year and the averages of the anomalies of the last 5 and 10 years.
3. The coefficients of variation for that given year, as well as for the last 5 and 10 years.
4. The logarithmic return using the previous year, 5 years prior, and 10 years prior.

The target variables used, respectively for migrations and conflicts, were: the number of internal migrants in a certain region, divided by the total population of the region and normalized using the formula log(1+x); and the total number of conflicts (regardless of

the type) in a region in a given year.
To try to isolate key variables that may influence migrations and conflicts, different heterogeneity analyses were performed using subsets of data by the following categories:

- Continents and Geographic Macro-Regions: To investigate possible patterns related to geographical subdivisions, an analysis was conducted by dividing the data by continent. The continents considered are Africa, Asia, Europe, North America, South America, and Central America. Data for Oceania were too sparse to be used, so they were not considered. For the continental subdivision, the 'countrycode' function from the 'countrycode' package was used. Since current literature ([33],[31]) has identified four specific geographic regions as potential hotspots for migrations and conflicts influenced by climate change, the data were also divided into the following major geographic regions: Southeast Asia, South and Central America, MENA (Middle East and North Africa), and Sub-Saharan Africa.

- Income: The data were also divided based on the income level of the country. This division was made using World Bank data for 2005. The data are part of the indicator NY.GDP.MKTP.PP.KD, which represents GDP at purchasing power parity (PPP) in constant 2017 international dollars. This indicator accounts for inflation and allows for GDP comparisons between countries with different currencies. The data were divided into four groups: High, Upper Middle, Lower Middle, and Low Income.

- Government Effectiveness: The data were divided based on governance levels using the dataset from the Worldwide Governance Indicators. Three classes were created: High, Middle, and Low level of governance. Government Effectiveness measures six key dimensions used to capture the state capacity, that is a government's ability to achieve policy objectives.

- Potential Evapotranspiration (PET): The data were divided into two classes, High and Low Potential Evapotranspiration, to identify potential effects for possible effects of aridity on migrations and conflicts. The evapotranspiration is defined as "the evaporation from an extended surface of a short green crop which fully shades the ground, exerts little or negligible resistance to the flow of water, and is always well supplied with water. Potential evapotranspiration cannot exceed free water evaporation under the same weather conditions."

- Type of Conflicts: This subdivision was used only for the analysis on conflicts. Here I report the definitions of the three types of conflicts from UCDP:

i. State-based armed conflict: A contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state.

ii. Non-state conflict: The use of armed force between two organized armed groups, neither of which is the government of a state.

iii. One-sided violence: Is the use of armed force by the government of a state or by a formally organized group against civilians.

After the linear analysis carried out on the different subdivisions, in order to asses non-linear relationships between groundwater variables and migrations and conflicts, a Random Forest approach was performed. In this case, only the subdivisions for Governance and Income values were used.

Regarding the internal migration data, for each class and for each year of interval of migration, the data have been divided into three sets: train, test and validation, with the following proportions: 30 - 59.5 - 10.5. The train set has been used to search for the 6 best independent variables among the ones indicated at the beginning of the chapter, using a RandomForestRegressor algorithm. Then, using these 6 variables and using the RandomizedSearchCV, the tuning of the parameters was implemented on the test set, with a 5-fold cross validation and using as scoring 'R2'. Then the trained and tuned model has been applied to the validation data, to see if the model still holds for data that has never observed (data for the class "Low Income" and the year 5 were too little, therefore it wasn't possible to apply the algorithm).

Regarding the conflict dataset, in this case also, the subdivision in governance and income level was used. The data were divided into four classes: train, test, validation over the period 1989-2016 and another class with the data for the period 2017-2019. The latter has been done in order to see if the model still holds in time. Unlike for migrations, in this case the algorithm used was a classification one (RandomForestClassifier). This means that the aim of the model was to see if groundwater variables could be predictors either for the trigger of conflict or not. To do this, the variable 'count', that measures the number of conflict in a certain region for a definite year, was transformed into a binary one: 1 if there were conflicts, 0 otherwise. Regarding the classifier, since the dataset is unbalanced (there are many more zeros than ones), it was necessary to implement a weight. In this case, for each class, the ratio between the number of non-occurrences and occurrences was taken as weight. Then, the train set was used to determine the 6 best independent variables. With these variables, the model was tuned using a RandomizedSearchCV with a stratified 5-fold cross validation and using as scoring the average precision. In this case, two different test was done on the model: the first one on the validation set over the same period of the training data, and the second one on the data for the period

2017-2019, outside the training period. As metrics to evaluate the model's performance, precision, F1-score, and recall were used (TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative):

- Precision: measures the ratio between positive predictions and all the predictions made by the model
$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.3}$$
- Recall: measures the ability of the model to asses the correct positive values
$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$
- F1-Score: is the harmonic mean of precision and recall
$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.5}$$

4.2 Linear Analysis for Internal Migration

Since the migration data are measured over different time intervals (one and five years), the data were divided according to these intervals during the various analyses. Regarding the distribution family, the 'gaussian' family was used for the dependent variable in the generalized linear regression model, assuming that data are normally distributed.

Geographical Regions

In this section I report the results for the geographic subdivision of data. The first analysis has been done over all countries, irrespective of their geographical affiliation ('global'). This analysis returned statistically significant values (p-value < 0.01) only for positive anomalies over 5 years and for the logarithmic return over 10 years. Specifically, for an increase in one unit of positive anomalies over 5 years we observe a decrease in internal migration of approximately 0.39%, while for a percentage increase in groundwater storage over 10 years we observe a decrease of internal migrations of approximately 0.4%. No statistically significant relationships were found across all geographical subdivisions. Regarding Asian countries, for the interval of observed internal migration of 1 year, we can find significant relationships between the coefficient of variation for all three intervals over which it is calculated, and the internal migration. Specifically, we find that for each unit increase in the coefficient of variation, we measure an increase in the migration of 0.5% for the one-year averages and 0.6% for the averages over 5 and 10 years (p-values < 0.001). Also for European countries, statistically significant correlations are found between positive anomalies and the internal migration. Specifically, for each unit increase in positive anomalies, there is a decrease in internal migrations of 0.6% (p-value < 0.001), of 2.47% (p-value < 0.001), and 1.99% (p-value < 0.01), respectively, for the averages of anomalies over intervals of 1, 5, and 10 years. In North America, however, significant relationships are found for the averages of groundwater values. Specifically, for each percentage variation in the quantity of groundwater, there is a decrease in internal migration of 19.71% (p-value < 0.001) and 20.58% (p-value < 0.001), respectively, based on the averages over one year and five years. For the countries in Central America, only one variable is significantly correlated with internal migration — namely, positive anomalies over the 10-year interval. For each unit increase in positive anomalies, there is a decrease of about 0.61% in internal migration (p-value < 0.001). The case is different for countries belonging to North Africa and the Middle East (MENA). In this instance, the averages of groundwater quantity over intervals of 1, 5, and 10 years are statistically significantly correlated with internal migration. For each percentage increase in average water quantity there are decreases in the number of migrants equal to 93.92%, 60.70%, and 92.78% (p-value < 0.001), respectively for the averages over 1, 5 and 10 years. Similarly, for countries in Central and South America, the average groundwater quantity appears to be the most important variable. Indeed, for each percentage increase in the average groundwater value over intervals of one year, five years, and ten years, increases in the number of migrants are observed at 13.12% (p-value < 0.01), 11.93% (p-value < 0.001), and 10.61% (p-value < 0.001), respectively. Finally, the last subdivision that provides statistically significant results is that of countries belonging to Southeast Asia. Significant correlations are observed for the coefficients of variation over all three time intervals. For each unit increase in the coefficient of variation, increases in the number of migrants are found at 0.05%, 0.1%, and 0.06% (p-value < 0.001).

Table 4.1: Migrations 1-year interval: Geographical subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Variables	Global	Asia	Africa	Europe	N. America	S. America
GWS	1.337 (1.271)	0.8495 (0.9866)	0.7438 (0.4966)	-19.49* (7.998)	-19.71*** (4.743)	10.48 (2.842)
GWS 5-y	1.231 (1.090)	1.048 (0.9266)	0.3584 (0.3612)	-26.02* (13.21)	-20.58*** (5.866)	10.78 (3.253)
GWS 10-y	1.438 (1.007)	1.153 (0.9498)	0.1442 (0.3155)	-16.95* (8.457)	-18.42* (7.719)	7.240 (3.298)
GWS Anomalies 1-y	-0.0012 (0.0009)	0.0024* (0.0012)	0.0049 (0.0021)	-0.0066*** (0.0025)	-0.0016* (0.0006)	0.0081 (0.0035)
GWS Anomalies 5-y	-0.0039** (0.0013)	0.0010 (0.0013)	0.0030 (0.0025)	-0.0247*** (0.0052)	-0.0010 (0.0007)	0.0021 (0.0047)
GWS Anomalies 10-y	-0.0007 (0.0011)	0.0022 (0.0013)	0.0010 (0.0025)	-0.0199** (0.0057)	-0.0010 (0.0007)	-0.0002 (0.0063)
Coefficient of Variation 1-y	-2.65e-6 (2.29e-5)	0.0005*** (7.41e-5)	-1.19e-5 (1.05e-5)	-0.0001 (7.75e-5)	-1.16e-5 (0.0001)	0.0004 (8.65e-5)
Coefficient of Variation 5-y	2.45e-5 (1.91e-5)	0.0006*** (0.0001)	-1.28e-5 (9.75e-6)	-8.89e-6 (5.16e-5)	-0.0002 (0.0001)	0.0002 (0.0001)
Coefficient of Variation 10-y	-1.37e-5 (1.42e-5)	0.0006*** (8.67e-5)	-2.86e-6 (9.78e-6)	-3.61e-5 (6.13e-5)	-6.39e-5 (7.82e-5)	0.0003 (0.0001)
Logarithmic Return 1-y	0.0008 (0.0017)	0.0218* (0.0085)	0.0020 (0.0017)	0.0044* (0.0021)	-0.0078 (0.0069)	-0.0068 (0.0022)
Logarithmic Return 5-y	-0.0032 (0.0020)	-0.0006 (0.0025)	0.0011 (0.0013)	-0.0091** (0.0031)	-0.0035 (0.0021)	0.0066 (0.0101)
Logarithmic Return 10-y	-0.0040** (0.0019)	-0.0030 (0.0031)	0.0019 (0.0016)	-0.0065* (0.0030)	-0.0019 (0.0025)	0.0084 (0.0120)
Observations	1,689	379	444	520	170	97

Variables	C. America	MENA	C. and S. America	Sub-Saharan Africa	South-East Asia
GWS	-0.2294 (0.0974)	-93.92*** (10.04)	13.12** (3.913)	0.7442 (0.4975)	0.8505 (0.9848)
GWS 5-y	-0.1813 (0.0848)	-60.70*** (7.066)	11.93*** (3.462)	0.3557 (0.3618)	1.049 (0.9249)
GWS 10-y	-0.1494 (0.0837)	-92.78*** (10.22)	10.61** (3.119)	-0.1441 (0.3160)	1.153 (0.8482)
GWS Anomalies 1-y	0.0048 (0.0028)	-0.0008 (0.0009)	0.0121* (0.0046)	0.0050* (0.0022)	0.0025* (0.0012)
GWS Anomalies 5-y	-0.0077 (0.0016)	-0.0001 (0.0007)	0.0106* (0.0043)	0.0031 (0.0026)	0.0011 (0.0013)
GWS Anomalies 10-y	-0.0061*** (0.0005)	-0.0005 (0.0009)	0.0100* (0.0043)	0.0011 (0.0025)	0.0023. (0.0013)
Coefficient of Variation 1-y	0.0016 (0.0004)	-3.14e-6 (1.07e-5)	0.0010. (0.0006)	-1.2e-5 (1.07e-5)	0.0005*** (7.43e-5)
Coefficient of Variation 5-y	0.0008 (0.0004)	2.97e-6*** (3.77e-7)	0.0011 (0.0012)	-2.6e-5 (2e-5)	0.0010*** (0.0001)
Coefficient of Variation 10-y	0.0006 (0.0002)	4.14e-6*** (5.15e-7)	0.0011* (0.0005)	-3.69e-6 (1.1e-5)	0.0006*** (9.4e-5)
Logarithmic Return 1-y	0.0214 (0.0269)	-0.0017*** (0.0004)	-0.0549 (0.0454)	0.0023 (0.0019)	0.0249* (0.0096)
Logarithmic Return 5-y	0.0060 (0.0101)	-0.0006. (0.0003)	0.0407* (0.0179)	0.0017 (0.0018)	-0.0022 (0.0041)
Logarithmic Return 10-y	-0.0010 (0.0138)	-0.0002 (0.0003)	0.1735* (0.0670)	0.0037 (0.0030)	-0.0063 (0.0046)
Observations	54	39	87	405	357

Different results were found for the 5-year internal migration interval. For example, is observed that in Asia, positive anomalies are statistically correlated with an increase in migrations: for each unit increase in positive anomalies over 5 and 10 years, increases in the number of migrants of approximately 0.44% and 0.37% are found (p-value < 0.001). In Southeast Asia, it is found that a percentage increase in the average groundwater values over a 5-year interval is correlated with an increase in migrations of about 4.9% (p-value < 0.01). In this case, it is also noted that an increase in positive anomalies over the same time interval is correlated with an increase in the number of migrants of approximately 0.34% (p-value < 0.01). For the other subdivisions, no statistically significant values are found except with very small coefficients.

Table 4.2: Migrations 1-year interval: Geographical subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Variables	Global	Asia	Africa	Europe	N. America	S. America
GWS	3.976** (1.404)	1.577 (1.345)	-3.262 (7.660)	2.560 (3.812)	-1.331 (13.90)	-0.0565 (0.8650)
GWS 5-y	3.723* (1.505)	1.297 (1.758)	-56.46 (44.75)	2.283 (3.740)	7.426 (15.60)	-0.1919 (0.6803))
GWS 10-y	3.393* (1.414)	0.8913 (1.566)	-11.66 (18.53)	1.581 (3.514)	-6.806 (18.10)	-0.0872 (0.5834)
GWS Anomalies 1-y	0.0022** (0.0009)	0.0027* (0.0011)	0.0017 (0.0022)	-0.0190 (0.0106)	0.0013 (0.0008)	0.0008 (0.0028)
GWS Anomalies 5-y	0.0026 (0.0014)	0.0044*** (0.0011)	0.0023 (0.0034)	-0.0146 (0.0131)	0.0011 (0.0011)	0.0011 (0.0029)
GWS Anomalies 10-y	0.0014 (0.0015)	0.0037*** (0.0010)	0.0007 (0.0014)	-0.0128 (0.0069)	-0.0014 (0.0016)	0.0053 (0.0036)
Coefficient of Variation 1-y	7.18e-5 (6.47e-5)	0.0001 (8.26e-5)	-2.18e-5 (1.35e-5)	-4.63e-5 (0.0003)	-0.0002 (8.81e-5)	0.001 (0.0002)
Coefficient of Variation 5-y	0.0002** (8.7e-5)	6.95e-5 (0.0001)	-4.22e-5 (2.59e-5)	-0.0002 (0.0001)	-3.67e-5 (9.52e-5)	0.002 (0.0002)
Coefficient of Variation 10-y	2.91e-5 (4.47e-5)	6.65e-5 (4.73e-5)	-0.0002 (9.58e-5)	-0.0002* (8.92e-5)	-0.0003*** (9.39e-5)	0.002 (0.0003)
Logarithmic Return 1-y	0.0016 (0.0013)	-0.0019 (0.0039)	-0.0004 (0.0002)	-0.0396 (0.0203)	0.0040 (0.0027)	0.0052 (0.0106)
Logarithmic Return 5-y	0.0023* (0.0009)	0.0038** (0.0015)	0.0003 (0.0002)	-0.0084 (0.0076)	0.0026* (0.0012)	-0.0026 (0.0082)
Logarithmic Return 10-y	0.0005 (0.0011)	-0.0012 (0.0017)	0.0003 (0.0002)	0.0033 (0.0048)	0.0044* (0.0017)	-0.0167 (0.0104)
Observations	1,867	661	60	48	352	201

Variables	C. America	MENA	C. and S. America	Sub-Saharan Africa	South-East Asia
GWS	2.768 (1.542)	-343.2 (222.7)	1.604 (1.138)	-3.262 (7.707)	3.307* (1.459)
GWS 5-y	2.162 (1.784)	-376.1 (202.5)	1.050 (1.231)	-56.46 (45.02)	4.902** (1.876)
GWS 10-y	2.518 (1.430)	-293.7 (161.9)	1.308 (1.014)	-11.66 (18.64)	3.597. (1.862)
GWS Anomalies 1-y	0.0029 (0.0021)	-0.0737 (0.0152)	0.0025 (0.0020)	0.0017 (0.0022)	0.0026* (0.0010)
GWS Anomalies 5-y	0.0032 (0.0029)	-0.0893 (0.1015)	0.0025 (0.0027)	0.0023 (0.0034)	0.0034** (0.0011)
GWS Anomalies 10-y	0.0060 (0.0027)	-0.0709 (0.0974)	0.0053* (0.0026)	0.0007 (0.0014)	0.0018. (0.0010)
Coefficient of Variation 1-y	8.61e-5 (0.0002)	-0.0008 (0.0007)	-1.86e-6 (0.0001)	-2.18e-5 (1.36e-5)	-4.12e-5 (9.38e-5)
Coefficient of Variation 5-y	9.41e-5 (0.0001)	-0.0001 (0.0013)	7.42e-5 (0.0001)	-4.22e-5 (2.6e-5)	0.0005* (0.0002)
Coefficient of Variation 10-y	0.0001 (0.0001)	-0.0002 (0.0002)	0.0001 (0.0001)	-0.0002* (9.64e-5)	0.0002* (0.0001)
Logarithmic Return 1-y	0.0075 (0.0066)	0.0095 (0.0311)	0.0066 (0.0062)	-0.0004 (0.0002)	0.0094* (0.0038)
Logarithmic Return 5-y	0.0001 (0.0032)	-0.0657 (0.0748)	-0.0003 (0.0031)	-0.0003 (0.0002)	0.0128** (0.0046)
Logarithmic Return 10-y	-0.0014 (0.0029)	-0.0825* (0.0052)	-0.0024 (0.0027)	-0.0003 (0.0002)	0.0010 (0.0015)
Observations	516	18	663	49	462

Income

Regarding the subdivision by average per capita income, for the 1-year internal migration interval, it is observed that for high-income countries, for each unit increase in positive anomalies over one year, the number of migrants per unit of population decreases by about 0.55% (p-value < 0.001). The same applies to the intervals of average anomalies measured over 5 and 10 years, with respective decreases in the average internal migration of 1.67% and 1.37% (p-values < 0.001). No statistically significant results are found for the other income classes, except for upper-middle-income countries: for each unit increase in one-year positive anomalies, an increase in the number of migrants per unit of population of about 0.45% is observed (p-value < 0.001), thus in the opposite direction compared to high-income countries.

Variables	High Income	Upper Middle Income	Low Middle Income	Low Income
GWS	-14.82* (6.109)	1.557 (0.8339)	-0.9841 (0.9589)	3.441 (3.589)
GWS 5-y	-24.67* (12.02)	0.8211 (0.5689)	-0.2497 (1.010)	2.727 (2.457)
GWS 10-y	-17.26* (8.364)	0.4893 (0.5129)	0.0128 (1.214)	2.369 (2.783)
GWS Anomalies 1-y	-0.0055*** (0.0020)	0.0045*** (0.0010)	0.0084 (0.0051)	0.0033 (0.0020)
GWS Anomalies 5-y	-0.0167*** (0.0037)	0.0020 (0.0011)	0.0113 (0.0092)	0.0031 (0.0019)
GWS Anomalies 10-y	-0.0137*** (0.0038)	0.0018 (0.0014)	0.0161 (0.0112)	0.0023 (0.0018)
Coefficient of Variation 1-y	-0.0001 (7.4e-5)	1.24e-5 (1.39e-5)	0.0011 (0.0007)	0.0006 (0.0006)
Coefficient of Variation 5-y	1.46e-6 (5.29e-5)	2.43e-5 (2.55e-5)	9.02e-5 (8.03e-5)	0.0003 (0.0002)
Coefficient of Variation 10-y	-1.59e-5 (6.15e-5)	1.94e-6 (9.66e-6)	0.0002 (0.0001)	-2.42e-6 (0.0001)
Logarithmic Return 1-y	0.0035 (0.0021)	0.0040 (0.0028)	-0.0052 (0.0117)	-0.0337 (0.021)
Logarithmic Return 5-y	-0.0088** (0.0028)	0.0101* (0.0050)	0.0024 (0.0030)	0.0066 (0.0117)
Logarithmic Return 10-y	-0.0063* (0.0028)	0.0080* (0.0037)	0.0041 (0.0071)	0.0103 (0.0049)
Observations	616	412	320	208

Table 4.3: Migrations 1-year interval: Income subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Regarding the measurements of migrations over a 5-year interval, the subdivision into income brackets yielded statistically significant results only for upper-middle-income countries. Specifically, for each unit increase in positive anomalies over one year(p-value < 0.01) , five years (p-value < 0.01), and ten years (p-value < 0.001), respective decreases are observed in internal migration of 0.24%, 0.46%, and 0.66%.

Variables	High Income	Upper Middle Income	Low Middle Income	Low Income
GWS	0.4016 (1.409)	-1.212 (0.9148)	0.9975 (0.8146)	-0.4780 (0.7246)
GWS 5-y	-1.935 (2.799)	-0.7927 (0.7825)	0.8463 (0.7555)	-0.5168 (0.7804)
GWS 10-y	-1.287 (2.642)	-0.7488 (0.7562)	0.6982 (0.6662)	-0.4463 (0.6656)
GWS Anomalies 1-y	0.0024 (0.0017)	0.0024** (0.0010)	0.0010 (0.0012)	0.0097 (0.0122)
GWS Anomalies 5-y	0.0023 (0.0024)	0.0046** (0.0014)	-0.0004 (0.0011)	0.0031 (0.0112)
GWS Anomalies 10-y	0.0024 (0.0025)	0.0066*** (0.0018)	-0.0002 (0.0009)	0.0037 (0.0175)
Coefficient of Variation 1-y	3.03e-5 (0.0001)	7.88e-5 (9.39e-5)	6.46e-5 (8.84e-5)	-5.54e-7 (0.0143)
Coefficient of Variation 5-y	7.05e-5 (0.0001)	6.36e-5 (9.71e-5)	1.12e-5 (6.69e-5)	0.0033 (0.0010)
Coefficient of Variation 10-y	-2.01e-5 (8.92e-5)	-2.44e-5 (7.89e-5)	6.98e-5 (4.56e-5)	0.0098 (0.0034)
Logarithmic Return 1-y	0.0102 (0.0082)	0.0052* (0.0025)	-0.0019 (0.0017)	-0.4074 (0.1846)
Logarithmic Return 5-y	0.0006 (0.0022)	0.0002 (0.0016)	0.0010 (0.0011)	0.0315 (0.0157)
Logarithmic Return 10-y	-0.0009 (0.0026)	0.0020 (0.0013)	0.0017 (0.0014)	0.0739 (0.0288)
Observations	459	654	615	14

Table 4.4: Migrations 5-year interval: Income subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

<u>*Governance*</u>

Regarding the subdivision of countries by governance level, for the 1-year interval of internal migration, a greater number of statistically significant results is observed. Specifically, for countries with a high level of governance, it is found that for each unit increase in positive anomalies measured over 5 years, internal migration decreases by about 0.91% (p-value < 0.001). An opposite trend is observed for countries with a medium level of governance. Indeed, regarding positive anomalies measured over intervals of 1 and 10 years, it is found that for each unit increase, there is an increase in the internal migration of 0.28% in both cases (p-value < 0.01). Statistically significant results are also obtained for the coefficients of variation measured over intervals of 1 and 10 years. Specifically, still referring to countries with a medium level of governance, it is observed that for an increase of one unit in the coefficient of variation, there is an increase, in both cases, of about 0.04% (p-values < 0.001).

Regarding the migrations observed over the 5-year time interval, the class of countries with a medium level of governance is again the one providing statistically significant results. Indeed, it is found that for each percentage point increase in per capita groundwater quantity measured in the same year, the number of migrants per unit population increases by about 5.1% (p-value < 0.001). This value decreases slightly when considering the averages of per capita groundwater over 5 and 10 years, to 5.08% and 4.49%, respectively (p-values < 0.001). The anomalies also show statistically significant results for this class. Regarding the anomalies measured over the one-year interval, it is observed that for each unit increase in positive anomalies, the number of migrants per unit

Variables	High Governance	Medium Governance	Low Governance
GWS	1.471 (2.914)	3.385* (1.706)	-0.0836 (0.7966)
GWS 5-y	1.194 (2.114)	3.280 (1.796)	0.2506 (0.8086)
GWS 10-y	1.552 (1.912)	2.480 (1.296)	0.3779 (0.9451)
GWS Anomalies 1-y	-0.0026* (0.0013)	0.0028** (0.0010)	0.0044* (0.0021)
GWS Anomalies 5-y	-0.0091*** (0.0022)	0.0022* (0.0010)	0.0028 (0.0023)
GWS Anomalies 10-y	-0.0032 (0.0017)	0.0028** (0.0012)	0.0040 (0.0023)
Coefficient of Variation 1-y	-1.47e-5 (3.17e-5)	0.0004*** (6.05e-5)	0.0006 (0.0004)
Coefficient of Variation 5-y	1.486e-5 (3.10e-5)	0.0002** (7.9e-5)	0.0003* (0.0001)
Coefficient of Variation 10-y	-2.94e-5* (1.41e-5)	0.0004*** (8.21e-5)	0.0002 (0.0001)
Logarithmic Return 1-y	0.0008 (0.0019)	-0.0008 (0.0050)	0.0026 (0.0159)
Logarithmic Return 5-y	-0.0050* (0.0023)	0.0018 (0.0058)	0.0017 (0.0043)
Logarithmic Return 10-y	-0.0052* (0.0024)	0.0021 (0.0041)	0.0029 (0.0039)
Observations	617	651	280

Table 4.5: Migrations 1-year interval: Level of Governance subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

population increase by about 0.25% (p-value < 0.01). This value increases slightly when considering the positive anomalies averaged over intervals of 5 and 10 years, to 0.47% and 0.45%, respectively (p-values < 0.001).

Variables	High Governance	Medium Governance	Low Governance
GWS	0.4568 (1.421)	5.099*** (0.7487)	0.2536 (1.296)
GWS 5-y	-1.789 (2.798)	5.083*** (0.9048)	0.3592 (1.124)
GWS 10-y	-1.120 (2.650)	4.492*** (0.8978)	0.5283 (0.9853)
GWS Anomalies 1-y	0.0024 (0.0017)	0.0025** (0.0008)	-0.0028 (0.0021)
GWS Anomalies 5-y	0.0022 (0.0024)	0.0047*** (0.0010)	-0.0012 (0.0020)
GWS Anomalies 10-y	0.0024 (0.0025)	0.0045*** (0.0011)	0.0012 (0.0023)
Coefficient of Variation 1-y	2.25e-5 (0.0001)	7.5e-5 (6.62e-5)	-0.0002 (0.0002)
Coefficient of Variation 5-y	6.99e-5 (0.0001)	7.28e-5 (7.71e-5)	7.06e-5 (0.0002)
Coefficient of Variation 10-y	-2.69e-5 (0.0002)	3.96e-5 (4.2e-5)	0.0001 (0.0001)
Logarithmic Return 1-y	0.0107 (0.0082)	0.0004 (0.0013)	-0.0081 (0.0122)
Logarithmic Return 5-y	0.0009 (0.0022)	0.0010 (0.0007)	-0.0093 (0.0047)
Logarithmic Return 10-y	-0.0008 (0.0027)	-0.0005 (0.0010)	-0.0081 (0.0045)
Observations	489	1057	299

Table 4.6: Migrations 5-year interval: Level of Governance subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Potential Evapotranspiration

Regarding the subdivision of countries by aridity level (measured through potential evapotranspiration, PET), it is found that for both migrations measured over 1 year and migrations measured over 5 years, the only statistically significant results are in countries with low PET values (i.e., countries with low levels of aridity). Specifically, for the one-year migration interval, an increase of one unit in positive anomalies measured over a one-year interval is associated with a decrease in the number of migrants per unit population of about 0.64% (p-value < 0.001). This effect becomes more pronounced for positive anomalies measured over 5 and 10 years, with decreases of 1.71% and 1.47%, respectively (p-value < 0.001). An opposite trend is observed for the 5-year migration interval: increases of one unit in positive anomalies measured over 1 and 5 years are associated with increases in internal migration of approximately 0.30% and 0.33%, respectively (p-value < 0.01).

Variables	High PET	Low PET
GWS	1.067 (0.5915)	-19.03* (7.902)
GWS 5-y	0.8417 (0.4938)	-25.65 (13.12)
GWS 10-y	0.5806 (0.5074)	-16.11 (8.261)
GWS Anomalies 1-y	0.0032 (0.0010)	-0.0064*** (0.0020)
GWS Anomalies 5-y	0.0017 (0.0012)	-0.0171*** (0.0037)
GWS Anomalies 10-y	0.0016 (0.0013)	-0.0147*** (0.0037)
Coefficient of Variation 1-y	5.93e-6 (1.47e-5)	-0.0001 (7.85e-5)
Coefficient of Variation 5-y	1.23e-5 (1.48e-5)	-2.38e-5 (4.27e-5)
Coefficient of Variation 10-y	9.78e-6 (1.06e-5)	-4.17e-5 (5.64e-5)
Logarithmic Return 1-y	0.0018 (0.0032)	0.0039 (0.0020)
Logarithmic Return 5-y	0.0023 (0.0041)	-0.0087** (0.0029)
Logarithmic Return 10-y	0.0018 (0.0025)	-0.0069* (0.0030)
Observations	830	859

Table 4.7: Migrations 1-year interval: PET subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Variables	High PET	Low PET
GWS	-1.146 (0.8095)	0.2489 (0.9167)
GWS 5-y	-0.9295 (0.5991)	-0.2125 (1.093)
GWS 10-y	-0.5607 (0.5624)	-0.4392 (1.075)
GWS Anomalies 1-y	0.0009 (0.0011)	0.0030** (0.0011)
GWS Anomalies 5-y	0.0021 (0.0014)	0.0033** (0.0016)
GWS Anomalies 10-y	0.0037 (0.0017)	0.0016 (0.0018)
Coefficient of Variation 1-y	6.61e-6 (8.27e-5)	8.38e-5 (9.66e-5)
Coefficient of Variation 5-y	1.65e-6 (7.92e-5)	0.0003** (0.0001)
Coefficient of Variation 10-y	-2.62e-5 (4.61e-5)	6.02e-5 (9.45e-5)
Logarithmic Return 1-y	0.0004 (0.0014)	0.0059 (0.0039)
Logarithmic Return 5-y	0.0007 (0.0007)	0.0036 (0.0028)
Logarithmic Return 10-y	0.0009 (0.0013)	0.0013 (0.0018)
Observations	565	1302

Table 4.8: Migrations 5-year interval: PET subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

4.3 Linear Analysis for Conflicts

Geographical Regions

The geographic subdivision for conflict data didn't return many significant results. No significant results are found for global data, Asian and African countries. Regarding European countries, positive anomalies turn out to be statistically significant. An increase of one unit in positive anomalies over one year is linked to a decrease of 0.78% of conflicts (p-value < 0.001); an increase of one unit in positive anomalies over five years is correlated with a decrease of 1.92% in the number of conflicts (p-value < 0.001); and for the 10-year anomaly interval, a decrease of 2.46% is obtained (p-value < 0.01). The same trend for anomalies is observed for South American countries: an increase of one unit in positive anomalies is linked to a decrease in the number of conflicts. Over a one-year interval, we obtain a decrease of 0.44% (p-value < 0.01); over a five-year interval, a decrease of 0.89% (p-value < 0.001); and over a ten-year interval, a decrease of about 1.18% (p-value < 0.01). As for the geographical regions most vulnerable to climate change - namely: MENA, Central and South America, Sub-Saharan Africa, and Southeast Asia - we obtain less significant results. Regarding MENA, the only statistically significant result is the 5-year logarithmic return: for each percentage increase over five years in the quantity of groundwater, an increase in the number of conflicts of about 0.38% is observed (p-value < 0.001). Also in Central and South American countries, it is observed that an increase of one unit in positive anomalies over 5 and 10 years is linked respectively to decreases in the number of conflicts of about 0.77% (p-value < 0.001) and 1.1% (p-value < 0.01).

Table 4.9: Conflicts: Geographical subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Variables	Global	Asia	Africa	Europe	N. America	C. America
GWS	77.02 (115.2)	520.8 (398.1)	363.6 (228.2)	-39,687.6 (33,686.8)	12,179.8 (6,708.7)	580.4 (310.1)
GWS 5-y	72.26 (108.0)	508.3 (396.7)	414.4 (228.1)	-119,470.7 (85,552.1)	13,264.0 (6,831.4)	603.9 (350.6)
GWS 10-y	70.97 (93.93)	548.0 (394.4)	297.7* (126.6)	-91,192.2 (136,892.3)	8,596.3 (7,466.5)	869.1 (596.5)
GWS Anomalies 1-y	-0.0860 (0.2265)	0.2245 (0.1120)	0.1561 (0.1894)	-0.7793*** (0.2278)	0.7447** (0.2352)	0.6315 (0.2561)
GWS Anomalies 5-y	-0.3856 (0.2997)	0.1552 (0.2857)	0.2388 (0.2753)	-1.924*** (0.4885)	0.7332 (0.3972)	0.5524 (0.2844)
GWS Anomalies 10-y	-0.6320* (0.3032)	-0.2091 (0.3914)	0.0162 (0.2540)	-2.463** (0.8825)	0.5841 (0.4002)	0.7324 (0.2470)
Coefficient of Variation 1-y	0.0014 (0.0027)	0.0053 (0.0036)	-0.0133 (0.0125)	-0.0037 (0.0030)	0.0333 (0.0215)	0.0187 (0.0135)
Coefficient of Variation 5-y	0.0007 (0.0042)	0.0103 (0.0071)	-0.0186 (0.0178)	-0.0025 (0.0029)	-0.0061 (0.0175)	0.0126 (0.0348)
Coefficient of Variation 10-y	-0.0017 (0.0052)	0.0105 (0.0083)	-0.0130 (0.0141)	-0.0041 (0.0064)	0.0578 (0.0243)	-0.0470 (0.0345)
Logarithmic Return 1-y	-0.0039 (0.0045)	0.0010 (0.0060)	-0.0324 (0.2339)	0.2730 (0.1988)	-0.0222 (0.0428)	0.8303 (0.3778)
Logarithmic Return 5-y	0.2437 (0.1984)	0.1823 (0.1671)	0.3680 (0.3357)	0.0249 (0.2161)	0.2616 (0.0872)	0.4368 (0.4803)
Logarithmic Return 10-y	0.0267 (0.0295)	0.0225 (0.0208)	0.0881 (0.2466)	-0.7586* (0.3604)	0.3173* (0.1221)	0.8477 (0.6844)
Observations	28,272	10,416	8,773	2,449	1,160	1,530

Variables	S. America	MENA	C. and S. America	Sub-Saharan Africa	South East Asia
GWS	-40.68 (60.93)	3,153.1 (1,879.4)	-2.991 (8.708)	407.3 (209.3)	140.8** (42.54)
GWS 5-y	-71.59 (70.24)	5,281.7 (3,067.8)	-9.950 (14.15)	426.8 (222.6)	150.7 (55.20)
GWS 10-y	-74.02 (65.26)	5,706.8 (3,096.8)	-14.93 (24.48)	248.3 (124.5)	121.8 (69.94)
GWS Anomalies 1-y	-0.4353** (0.1583)	0.7047 (0.3620)	-0.2828 (0.1445)	0.4415 (0.2239)	0.0973 (0.0818)
GWS Anomalies 5-y	-0.8918*** (0.2231)	0.5388 (0.6516)	-0.7741*** (0.2081)	0.9019 (0.2959)	0.1960 (0.1046)
GWS Anomalies 10-y	-1.184** (0.3880)	-0.1199 (0.9114)	-1.103** (0.3803)	0.6019 (0.3032)	0.4847** (0.1834)
Coefficient of Variation 1-y	-0.0017 (0.0106)	0.0104 (0.0063)	-0.0011 (0.0096)	0.0023 (0.0117)	0.0071 (0.0043)
Coefficient of Variation 5-y	0.0064 (0.0117)	0.0240 (0.0151)	0.0073 (0.0108)	0.0001 (0.0113)	0.0074 (0.0055)
Coefficient of Variation 10-y	0.0061 (0.0158)	0.0070 (0.0103)	0.0040 (0.0155)	-0.0063 (0.0109)	0.0169* (0.0074)
Logarithmic Return 1-y	0.2433 (0.2730)	-0.0451 (0.0326)	0.3664 (0.2575)	-0.9030 (0.6298)	-0.2376 (0.1657)
Logarithmic Return 5-y	-0.2126 (0.2084)	0.3842*** (0.0612)	-0.0713 (0.2254)	0.3620 (0.5758)	-0.2479 (0.1579)
Logarithmic Return 10-y	-0.5229** (0.1648)	0.5397* (0.2530)	-0.4003* (0.1748)	0.3416 (0.3752)	-0.7330 (0.2509)
Observations	2,945	3,193	4,371	7,130	4,371

Income

Regarding the subdivision of countries based on average per capita income, statistically significant results are obtained for all classes except for lower-middle-income countries. As for high-income countries, it is found that an increase of one unit in positive anomalies is linked to decreases of 0.83%, 1.45%, and 1.87%, respectively, over anomaly intervals of 1, 5, and 10 years (p-values < 0.001). For upper-middle-income countries, however, it was found that an increase of one unit in the coefficient of variation over 5 and 10 years is linked to decreases in the number of conflicts of about 0.022% (p-value < 0.01) and 0.037% (p-value < 0.001), respectively. Finally, for low-income countries, it is observed that logarithmic growth rates are correlated with increases in the number of conflicts: for each percentage increase over one year in the quantity of groundwater, there is an increase in the number of conflicts of about 0.44% (p-value < 0.001); this value tends to grow for the 5 and 10-year intervals, where increases of 0.82% and 1.2% are obtained (p-values < 0.01).

Variables	High Income	Upper Middle Income	Lower Middle Income	Low Income
GWS	-917.3 (2,701.8)	153.3 (97.29)	77.29 (61.90)	1,096.2* (443.9)
GWS 5-y	-334.3 (1,582.5)	99.55 (84.09)	83.08 (59.87)	1,192.8* (536.1)
GWS 10-y	65.89 (537.2)	81.58 (73.25)	75.91 (42.64)	779.8* (332.7)
GWS Anomalies 1-y	-0.8288*** (0.1662)	-0.0791 (0.0740)	-0.0417 (0.0854)	0.7722** (0.2796)
GWS Anomalies 5-y	-1.496*** (0.2304)	-0.2070 (0.1999)	-0.0568 (0.1158)	1.003* (0.3994)
GWS Anomalies 10-y	-1.866** (0.4317)	-0.5071 (0.2701)	-0.1043 (0.1842)	0.7947* (0.3635)
Coefficient of Variation 1-y	-0.0155 (0.0146)	-0.0090 (0.0058)	0.0088 (0.0102)	0.0043 (0.0049)
Coefficient of Variation 5-y	-0.0227 (0.0150)	-0.0218** (0.0080)	0.0104 (0.0131)	-0.0049 (0.0162)
Coefficient of Variation 10-y	-0.0002 (0.0047)	-0.0368*** (0.0098)	-0.0057 (0.0089)	-0.0324 (0.0202)
Logarithmic Return 1-y	0.1489* (0.034)	-0.0723 (0.0970)	-0.1398 (0.1308)	0.4417*** (0.1108)
Logarithmic Return 5-y	-0.1420 (0.1262)	0.1480 (0.0877)	0.0044 (0.1099)	0.8220** (0.2562)
Logarithmic Return 10-y	-0.7121 (0.3755)	0.1099 (0.0776)	0.2177 (0.3624)	1.201** (0.4080)
Observations	2,139	5,983	9,982	4,619

Table 4.10: Conflicts: Income subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Governance

Regarding the division of countries based on governance level, for the class with a high level of governance, it was found that a unit increase in positive anomalies over 1-year and 10-year intervals is linked to decreases in the number of conflicts of 0.63% and 1.71%, respectively (p-value < 0.01). Greater statistical significance is found, still for the same class, for the logarithmic return: for each percentage increase over 5 and 10 years in the quantity of groundwater, a decrease in the number of conflicts of about 0.58% (p-value < 0.01) and 1.49% (p-value < 0.001) is obtained. As for the class with a low level of governance, statistically significant results were obtained for the 10-year coefficient of variation: an increase of one unit in positive anomalies over this interval is linked to decreases in the number of conflicts of about 0.033% (p-value < 0.001).

Variables	High Governance	Medium Governance	Low Governance
GWS	-18,920.2 (23,350.8)	188.3* (94.24)	720.6* (297.0)
GWS 5-y	-22,871.0 (33,545.1)	168.7 (89.99)	764.1* (331.4)
GWS 10-y	-6,797.0 (18,920.3)	142.9 (80.41)	531.6* (214.3)
GWS Anomalies 1-y	-0.6273** (0.2165)	0.0015 (0.0544)	0.3562 (0.2266)
GWS Anomalies 5-y	-0.9614* (0.4083)	-0.0877 (0.0981)	0.4945 (0.3428)
GWS Anomalies 10-y	-1.709** (0.5530)	-0.1335 (0.1730)	0.2325 (0.2552)
Coefficient of Variation 1-y	-0.0255** (0.0091)	-0.0041 (0.0073)	-0.0005 (0.0059)
Coefficient of Variation 5-y	-0.0187 (0.0148)	-0.0134 (0.0143)	-0.0168 (0.0089)
Coefficient of Variation 10-y	0.0434* (0.0208)	-0.0146 (0.0114)	-0.0329*** (0.0092)
Logarithmic Return 1-y	-0.4902 (0.2687)	-0.0456 (0.0998)	0.1295 (0.1023)
Logarithmic Return 5-y	-0.5755** (0.1695)	0.1386 (0.0843)	0.1517 (0.1309)
Logarithmic Return 10-y	-1.494*** (0.3620)	0.2389 (0.2928)	0.5200* (0.2495)
Observations	2,046	10,974	9,393

Table 4.11: Conflicts: Level of Governance subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

PET

The division into countries with high and low PET did not yield statistically significant results except for the one-year logarithmic return for low PET countries: a percentage increase in the quantity of groundwater over a one-year interval corresponds to an increase in the number of conflicts of about 0.25% (p-value < 0.01).

Variables	High PET	Low PET
GWS	177.4 (145.4)	232.8 (157.7)
GWS 5-y	217.4 (194.9)	160.4 (163.3)
GWS 10-y	259.5 (184.6)	57.78 (107.2)
GWS Anomalies 1-y	0.2314 (0.1456)	-0.1827 (0.2377)
GWS Anomalies 5-y	0.1575 (0.2373)	-0.3953 (0.3157)
GWS Anomalies 10-y	0.1803 (0.2949)	-0.5371 (0.3384)
Coefficient of Variation 1-y	0.0049 (0.0031)	-0.0003 (0.0018)
Coefficient of Variation 5-y	0.0093 (0.0062)	9.87e-5 (0.0035)
Coefficient of Variation 10-y	0.0090 (0.0072)	-0.0118 (0.0071)
Logarithmic Return 1-y	-0.0037 (0.0064)	0.2541** (0.0950)
Logarithmic Return 5-y	0.1969 (0.1786)	0.3419 (0.2176)
Logarithmic Return 10-y	0.0213 (0.0199)	0.0100 (0.2087)
Observations	15,500	12,400

Table 4.12: Conflicts: PET subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

Type

The same can be said for the subdivision by type of conflict (state, non-state, one-sided violence): no statistically significant results were found.

Variables	State	Non-State	Onesided
GWS	247.8 (153.2)	-39.24 (20.18)	238.5 (180.7)
GWS 5-y	221.2 (140.4)	-30.69 (22.61)	NaN
GWS 10-y	194.0 (118.2)	-40.45 (24.92)	NaN
GWS Anomalies 1-y	-0.2252 (0.2794)	0.0193 (0.1920)	0.0581 (0.1141)
GWS Anomalies 5-y	-0.5947 (0.3179)	0.2130 (0.4636)	-0.0179 (0.1671)
GWS Anomalies 10-y	-0.8119 (0.3309)	-0.2643 (0.3176)	0.2281 (0.1483)
Coefficient of Variation 1-y	0.0029 (0.0035)	0.0016 (0.0018)	0.0009 (0.0020)
Coefficient of Variation 5-y	-0.0006 (0.0043)	0.0040 (0.0036)	0.0004 (0.0038)
Coefficient of Variation 10-y	-0.0072 (0.0060)	-0.0056 (0.0087)	-0.0017 (0.0047)
Logarithmic Return 1-y	-0.0043 (0.0144)	-0.2554 (0.1478)	0.0262 (0.0279)
Logarithmic Return 5-y	0.4171 (0.1511)	-0.0754 (0.0886)	0.3690 (0.1861)
Logarithmic Return 10-y	0.0722 (0.0668)	0.2126 (0.1771)	0.0631 (0.0467)
Observations	21,328	11,935	21,824

Table 4.13: Conflicts: Type subdivision. P-values: * p<0.05; ** p<0.01; *** p<0.001.

4.4 Random Forest Analysis for Internal Migration

Regarding internal migrations measured over a one-year interval, using all the data without any distinction, it is observed that variables related to the quantity, variability, and temporal trends of groundwater can explain about 5.5% of the variability of observed migrations. Regarding the division of countries based on income, over the same time interval, it was found that the variables can explain about 23% of internal migrations for high-income countries and about 35.84% for lower-middle-income countries. As for the two remaining subdivisions, namely upper-middle-income and low-income countries, the results yield a negative R^2. This indicates that the random forest model for these two classes is performing worse than a model that simply predicts the average of migrations. Therefore, the model is not able to capture the relationships between these variables and the migration rate, and this may be due to the presence of other relevant variables not included in the model that predominantly influence the migration rate. For the subdivision of countries based on the level of governance, it was found that the model used is able to explain about 12.98% of the variability in the class of countries with a high level of governance and about 19.98% in the class of countries with a low level of governance. Also in this case, for the class of countries with a medium level of governance, a negative R^2 is obtained.

Variables	MSE	R2	Train/Test/Validation
Global, 1-year	0.0017	0.0552	506 / 1005 / 178
High Income, 1-year	0.0017	0.2308	184 / 367 / 65
Upper Middle Income, 1-year	0.0004	-0.0082	123 / 245 / 44
Middle Low Income, 1-year	0.0002	0.3584	96 / 190 / 34
Low Income, 1-year	0.0101	-0.1190	62 / 124 / 22
High Governance, 1-year	0.0017	0.1298	185 / 367 / 65
Medium Governance, 1-year	0.0029	-0.0011	195 / 387 / 69
Low Governance, 1-year	0.0018	0.1908	84 / 166 / 30
Global, 5-year	0.0011	0.0694	560 / 1110 / 197
High Income, 5-year	0.0007	0.0910	137 / 273 / 49
Upper Middle Income, 5-year	0.0018	0.1651	193 / 384 / 68
Middle Low Income, 5-year	0.0007	0.2819	184 / 366 / 65
Low Income, 5-year	0.0014	0.0079	89 / 178 / 32
High Governance, 5-year	0.0017	0.0549	146 / 291 / 52
Medium Governance, 5-year	0.0010	0.1685	317 / 629 / 111
Low Governance, 5-year	0.0014	0.0694	89 / 178 / 32

Table 4.14: Random Forest Results for 1-year and 5-year Migrations.

Regarding the observation for interval of migrations over five years, using all data (global), it is found that the random forest model is able to explain about 6.94% of the variability of internal migration, so with a slight increase with respect to the 1-year interval. As for the subdivision of countries by average per capita income, it is found that as income decreases, the model is able to explain greater variability in the data: for the high-income class, an R^2 of 9% is found; for the upper-middle-income class, 16.54%; and for the lower-middle-income class, 28.1%. As for the low-income class, an R^2 of 0.79% was found. Finally, regarding the subdivision based on the level of governance, it was found that the maximum variability explained by the model is for the medium governance class, about 16.85%, while for the other two classes, respectively high governance and low governance, values of about 5.49% and 6.94% are found.

4.5 Random Forest Analysis for Conflicts

As described in the methods' description, in this case, the results were tested over two different sets: the first is the validation set, that is a partition of the data on the same time interval as the data used for training and parameter optimization; the second is the temporal validation set, used to see how the model's performance changes over time, that is data of conflicts observed in an interval that spans from 2017 to 2019. At the global level, it is found that for predictions for the temporal validation set, the accuracy for the class related to the presence of conflicts is 22%, with an F1-score of 0.18.

Table 4.15: Random Forest Results for Conflicts.
P.: Precison; F1: F1-Score; R.: Recall. Classes (0) and (1) refers to predictions on occurrences of conflicts. Class Ratio refers to the ratio between occurrences and no-occurrences in the selected class.

Variables	P. (0)	P. (1)	F1 (0)	F1 (1)	R. (0)	R (1)	Class Ratio (%)	Train/Test/Valid./Prev.
Global, Valid	0.96	0.62	0.94	0.67	0.93	0.73	15.59	49694/83486/32468/17748
Global, 2017-2019	0.85	0.22	0.87	0.18	0.90	0.15	15.59	49694/83486/32468/17748
High Income, Valid	1.00	0.63	0.99	0.70	0.99	0.78	2.18	11592/19474/7574/4140
High Income, 2017-2019	0.99	0.15	0.99	0.17	0.98	0.18	2.18	11592/19474/7574/4140
Upper Middle Income, Valid	0.97	0.83	0.97	0.82	0.97	0.81	14.00	11541/19390/7541/4122
Upper Middle Income, 2017-2019	0.88	0.51	0.91	0.37	0.95	0.28	14.00	11541/19390/7541/4122
Middle Low Income, Valid	0.94	0.74	0.93	0.77	0.92	0.80	26.80	11919/20025/7788/4257
Middle Low Income, 2017-2019	0.80	0.43	0.82	0.39	0.84	0.36	26.80	11919/20025/7788/4257
Low Income, Valid	0.95	0.85	0.96	0.81	0.97	0.78	24.19	5241/8806/3425/1872
Low Income, 2017-2019	0.82	0.67	0.88	0.41	0.96	0.30	24.19	5241/8806/3425/1872
High Governance, Valid	1.00	0.51	0.99	0.64	0.98	0.86	2.50	11466/19262/7492/4095
High Governance, 2017-2019	0.98	0.06	0.99	0.05	0.99	0.04	2.50	11466/19262/7492/4095
Medium Governance, Valid	0.97	0.77	0.97	0.79	0.96	0.81	15.71	18975/31879/12398/6777
Medium Governance, 2017-2019	0.84	0.43	0.89	0.25	0.95	0.17	15.71	18975/31879/12398/6777
Low Governance, Valid	0.95	0.81	0.95	0.82	0.95	0.82	27.51	10508/17654/6866/3753
Low Governance, 2017-2019	0.84	0.59	0.87	0.48	0.91	0.41	27.51	10508/17654/6866/3753

Regarding the subdivision of data based on average per capita income, it is found that the accuracy for prediction of conflicts increases as income decreases: it is found a precision of 0.15 for the high income class, 0.51 for the upper-middle income class, 0.43 for the lower-middle income class, and 0.67 for the low income class. This suggests that for the low-income class, about 67% of conflict predictions over the 2017–2019 interval are correct. The same trend is observed for the F1-scores, indicating an increase in the model's ability to manage both precision and recall in predicting conflicts when average per capita income lowers. Regarding the subdivision by level of governance, differences are also observed based on the class of belonging. The accuracy for class 1 (presence of conflict) increases from 0.06 for high levels of governance, to 0.43 for medium levels, up to 0.59 for low levels of governance. Also in this case, the same trend is observed for the F1-score. As for the validation set — meaning the validation set on data related to the training interval — the same trend is found for both subdivisions but with higher values, suggesting that the model loses part of its predictive capacity for the prediction for the interval time 2017-2019.

5 Conclusions

Regarding internal migration, the results obtained from the fixed-effects linear regression showed some differences between short-term and medium-term observations (1 and 5 years). Specifically, for the one-year interval, more statistically significant results were observed compared to the five-year interval. Nevertheless, in both cases, the correlation coefficients are rather small, implying that groundwater variables play a minor role in internal migration phenomena, and that the greater difficulty in describing the relationships between natural resources and migration phenomena lies in isolating the former from socio-economic factors. The variables for which greater statistically significant results were found are the positive anomalies—that is, the difference between the groundwater value and the average value in the period 1980–2010, divided by the standard deviation of the reference period—and the average of groundwater values over intervals of 1, 5, and 10 years.

The subdivision by geographical regions showed, in the case of the one-year interval, some variations between countries. For example, the average groundwater value over the one-year interval showed a negative correlation both in North American countries, where a percentage decrease is linked to a decrease in internal migration of about 20%, and in MENA region countries, where there is a decrease up to 94%. Conversely, the inverse relationship was found in Central and South American countries, where a percentage decrease in groundwater quantity over a five-year interval is correlated with an increase in internal migration of 10%. However, similar relationships were not found for internal migration over the medium term. Regarding medium-term internal migrations, a correlation was found in Southeast Asian countries between the amount of groundwater averaged over five years and an increase in internal migration of about 4.9%. Geographically, therefore, the results indicate that variables related to groundwater can explain only a very small percentage of internal migration phenomena.

Similarly, in the case of subdivisions by average per capita income and level of governance, the variables that seem to be most important are the positive anomalies and the average groundwater level, still with rather small coefficients. The subdivision by average per capita income did not lead to further improvements in the results, except for a limited correlation between anomalies measured over intervals of 1 and 5 years, respectively with upper-middle-income and high-income countries, showing minimal percentage values (+0.45% and -1.67%). Again, a reduction in the number of statistically significant results is observed for medium-term internal migration; a weak correlation was found between anomalies and internal migration for the Upper-Middle Income class. Regarding the subdivision based on the level of governance, the class where the effects of groundwater variables are more evident is the one with a medium level of governance. Indeed, in the short term, a slight correlation is observed between anomalies and internal migration, albeit with small values. More concrete results were found for medium-term internal migration and the average groundwater level over 5 years, where a percentage increase in the latter is correlated with an increase in internal migration of about 5%.

Subdivisions based on the level of PET did not yield significant results, except for a weak correlation between the 5-year positive anomalies and a 1.7% decrease in the class of countries with low PET (low aridity).

In the case of conflicts, however, it was found that, in addition to positive anomalies, the most important variable turns out to be the logarithmic return. In this instance, the geographical subdivision did not show statistically significant results with appreciably large coefficients. Significant values were found for the 5-year positive anomalies in countries in Central and South America (a 0.77% decrease in the number of conflicts). A statistically significant correlation was also found in Europe with the 10-year anomalies, showing a 1.9% decrease in conflicts.

For the subdivisions by average per capita income level and governance level, weak correlations were found concerning the high and low-income classes, as well as the class with a high level of governance. Specifically, regarding the high-income class, values that deviate positively from the long-term average (1980–2010) are correlated with a decrease of up to 1.86% in the number of conflicts. An inverse result is observed for the low-income class, where the growth of groundwater (measured using the logarithmic return) is correlated with an increase of about 1.2% in the number of conflicts.

No statistically significant results were found for the subdivisions based on the degree of PET and the type of conflict.

From the results of the random forest analysis on internal migration, independent of the geographical affiliation of the countries, it was shown that the variables related to groundwater manage to capture approximately 5.5% and 6.95% of the variability in the data for short-term and medium-term migration, respectively. Regarding short-term migration, it can be noted that for the Upper-Middle Income, Low Income, and Medium Governance classes, the model does not perform better than a simple average, suggesting that other variables not considered in the model are fundamental. However, for the Lower-Middle Income class, the random forest model manages to explain about 35.84% of the variability, a percentage that decreases to 23.08% for the high-income countries class. This result is also confirmed in the case of medium-term migrations, where the model is able to explain about 28% of the variability in the data for the Lower-Middle Income class, while for the other classes the percentage decreases (9.1% for High Income, 16.51% for Upper-Middle Income, and 0.7% for Low Income). As for governance, for the one-year interval, it is observed that the model manages to explain about 19% of the variability in the data for the Low Governance class, which decreases to just under 7% for the five-year interval. This may indicate that, in countries with a low level of governance, variables related to groundwater have a greater impact on migration in the short term rather than in the long term. Conversely, the opposite trend is observed for the High Governance class, which decreases from about 13% to 5.5%. From the comparison between the two models, linear regression and random forest regression, it can be noted that the random forest algorithm was able to explain a greater proportion of the variability in the data. This may suggest that the variables related to groundwater and internal migrations are linked together by non-linear relationships rather than by linear ones.

The results from classification model show that the accuracy of the model to predict conflicts slightly increases as the average per capita income level of the countries decrease. From the results obtained using the time validation set (2017-2019), we observe that the F1-score values span from 0.17 for the High Income class, to 0.37 for the Upper-Middle Income class, to 0.39 for the Lower-Middle Income class, and finally to 0.41 for the Low Income class. The same trend, but more accentuated, is observed for the Level of Governance classification: F1-scores span from 0.05 for the High Governance class, to 0.25 for the Medium Governance class, to 0.48 for the Low Governance class. From these results, it is also possible to notice that these values are lower than the values obtained from the validation set, that consist of data from the same period as the training set. This suggests that the model loses part of its capability to predict conflicts in time.

The thesis work was based on groundwater values obtained through the ISIMIP3a model. This model relies on simulations of groundwater quantity in response to climate changes, human use, and evapotranspiration. The climate data are also obtained from simulations of climate models and emission scenarios. Consequently, the groundwater values do not faithfully reflect the actual conditions that can be found in different countries, due to the inevitable simplification of hydrological processes, spatial resolution limitations, and uncertainties arising from the climate data. The lack of high-resolution, global groundwater measurements is, however, a well-known gap in hydrology. This deficiency was addressed by utilizing the ISIMIP3a model, which, despite its limitations, offers a coherent estimate of global trends. However, this methodological choice may have introduced uncertainties in the results. In addition, the relationships that link migrations to socio-economic variables, identified in the literature as predominant for this kind of phenomena, are impossible to isolate exactly. In this study, the drivers were isolated by subdividing countries into various classes. This has certainly limited the ability to distinguish between the effects of socio-economic variables and those of natural variables related to groundwater. Integrating more advanced socio-economic models or using actually measured hydrological data could, therefore, improve the understanding of the phenomenon and the separation between socio-economic variables on one hand and natural variables on the other.

A Appendix A

Here, for completeness, I report the name of countries used in the analysis.

A.1 Internal Migration

Geographic Subdivision (Tables 4.1 and 4.2)

1-year internal migration interval.
Asia: Thailand, Philippines, Armenia, Cambodia, Myanmar.
Africa: Botswana, Cameroon, Kenya, Zambia, Benin, Egypt, Guinea, Mozambique, Mali, Tanzania, South Africa, Malawi, South Sudan, Sudan, Togo, Senegal, Uganda.
Europe: Ireland, Portugal, Greece, Spain, United Kingdom, Romania, Belarus, Poland, Slovenia, Russia.
North America: USA, Canada, Mexico.
Central America: Panama, Jamaica, Trinidad and Tobago, El Salvador, Cuba.
South America: Brazil, Venezuela, Suriname ('region' fixed effect was removed).
MENA: Egypt, Sudan. Israel, Morocco.('region' fixed effect was removed).
Central and South America: Brazil, Panama, Venezuela, El Salvador, Suriname.
Sub-Saharan Africa: Botswana, Cameroon, Kenya, Zambia, Benin, Guinea, Mozambique, Mali, Tanzania, South Africa, Malawi, South Sudan, Togo, Senegal, Uganda.
South-East Asia: Thailand, Philippines, Cambodia, Myanmar.

5-years internal migration interval.
Asia: Indonesia, India, Israel, Vietnam, China, Malaysia, Mongolia, Philippines, Nepal.
Africa: Senegal, Ghana, South Africa, Morocco, Cameroon, Sierra Leone.
Europe: Greece ('region' fixed effects was removed).
North America: USA, Canada, Mexico.
Central America: Dominican Republic, Guatemala, Haiti, Costa Rica, Honduras, Nicaragua.
South America: Argentina, Chile, Paraguay, Uruguay, Ecuador, Brazil, Bolivia, Venezuela, Peru.
MENA: Israel, Morocco ('region' fixed effect was removed).
Central and South America: Argentina, Guatemala, Chile, Paraguay, Costa Rica, Uruguay, Honduras, Ecuador, Brazil, Bolivia, Nicaragua, Venezuela, Peru
Sub-Saharan Africa: Senegal, Ghana, South Africa, Cameroon, Sierra Leone ('region' fixed effect was removed)
South-East Asia: Indonesia, Vietnam, Malaysia, Philippines.

Income Subdivision (Tables 4.3 and 4.4)

1-year internal migration interval.
High Income: Panama, Ireland, Portugal, Trinidad and Tobago, Canada, Greece, Spain, United Kingdom, Romania, Poland, Slovenia, United States.
Upper Middle Income: Brazil, Thailand, Botswana, Jamaica, Belarus, Armenia, South Africa, Cuba,

Suriname.
Lower Middle Income: Zambia, Guinea, Mozambique, Mali, Malawi, South Sudan, Sudan, Togo, Uganda.
Low Income: Cameroon, Kenya, Papua New Guinea, Philippines, Benin, El Salvador, Tanzania, Cambodia, Senegal, Myanmar.

5-years internal migration interval.
High Income: United States, Canada, Greece, Chile, Israel, Uruguay.
Upper Middle Income: Argentina, Dominican Republic, Guatemala, Paraguay, Costa Rica, Fiji, China, Ecuador, Mexico, Brazil, Malaysia, South Africa, Peru.
Lower Middle Income: Sierra Leone ('region' fixed effect was removed).
Low Income: Indonesia, Papua New Guinea, Haiti, India, Honduras, Senegal, Bolivia, Nicaragua.

Governance Subdivision (Tables 4.5 and 4.6)

1-year interval migration interval.
High Level of Governance: Botswana, Ireland, Portugal, Canada, Greece, Spain, United Kingdom, Poland, Slovenia, United States, South Africa.
Medium Level of Governance: Brazil, Panama, Thailand, Jamaica, Philippines, Trinidad and Tobago, Benin, El Salvador, Romania, Mozambique, Armenia, Tanzania, Cuba, Suriname, Senegal, Uganda.
Low Level of Governance: Cameroon, Kenya, Papua New Guinea, Zambia, Guinea, Mali, Belarus, Cambodia, Malawi, Sudan, Togo, Myanmar.

5-years interval migration interval.
High Level of Governance: United States, Canada, Greece, Chile, Israel, Uruguay, Malaysia, South Africa.
Medium Level of Governance: Argentina, Indonesia, Dominican Republic, India, Costa Rica, Fiji, Honduras, Senegal, Vietnam, China, Mexico, Brazil, Ghana, Mongolia, Philippines, Morocco.
Low Level of Governance: Papua New Guinea, Guatemala, Haiti, Paraguay, Ecuador, Bolivia, Nicaragua, Nepal, Cameroon, Peru, Sierra Leone.

PET Subdivision (Tables 4.7 and 4.8)

1-year interval migration interval.
High PET: Panama, Thailand, Botswana, Venezuela, Jamaica, Kenya, Trinidad and Tobago, Zambia, Benin, El Salvador, Egypt, Guinea, Mozambique, Mali, Tanzania, Cambodia, South Africa, Malawi, South Sudan, Sudan, Togo, Cuba, Suriname, Senegal, Uganda. Low PET: Brazil, Ireland, Portugal, Cameroon, Papua New Guinea, Philippines, Canada, Greece, Spain, United Kingdom, Romania, Belarus, Kyrgyzstan, Armenia, Poland, Slovenia, United States, Russia, Myanmar.

5-years interval migration interval.
High PET: Dominican Republic, Guatemala, Haiti, Paraguay, India, Israel, Honduras, Senegal, Mexico, Nicaragua, Ghana, South Africa, Venezuela, Morocco.
Low PET: Argentina, Indonesia, Papua New Guinea, United States, Canada, Greece, Chile, Costa

Rica, Uruguay, Fiji, Vietnam, China, Ecuador, Brazil, Malaysia, Bolivia, Mongolia, Philippines, Nepal, Cameroon, Peru, Sierra Leone.

A.2 Conflicts

Geographic Subdivision (Table 4.12)

Asia: Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Brunei, Cambodia, China, Cyprus, Georgia, Hong Kong (China), India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Korean islands under UN jurisdiction, Kuwait, Lao People's Democratic Republic, Lebanon, Macau (China), Malaysia, Maldives, Mongolia, Myanmar, Nepal, North Korea, Oman, Pakistan, Palestine, Palestine Territories: Gaza Strip, Philippines, Qatar, Saudi Arabia, Singapore, South Korea, Sri Lanka, Syria, Taiwan, Tajikistan, Thailand, Timor Leste, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, Yemen.
Africa: Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Congo DRC, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, In dispute South Sudan/Sudan, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Sao Tome & Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, St. Helena, Sudan, Swaziland, Tanzania, The Gambia, Togo, Tunisia, Uganda, Zambia, Zimbabwe.
Europe: Albania, Andorra, Austria, Belarus, Bosnia & Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Faroe Is., Finland, France, Germany, Gibraltar (UK), Greece, Guernsey, Hungary, Iceland, Ireland, Isle of Man, Italy, Jersey, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Monaco, Montenegro, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom, Vatican City.
North America: Canada, Mexico, United States.
Central America: Barbados, Belize, Costa Rica, Cuba, Dominica, Dominican Republic, El Salvador, Grenada, Guadeloupe, Guatemala, Haiti, Honduras, Jamaica, Martinique, Nicaragua, Panama, Puerto Rico, Saint Lucia, Trinidad and Tobago.
South America: Argentina, Aruba, Bolivia, Brazil, Chile, Colombia, Ecuador, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela.
MENA: Algeria, Bahrain, Chad, Djibouti, Egypt, Eritrea, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Yemen.
Central and South America: Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Honduras, Nicaragua, Panama, Paraguay, Peru, Suriname, Uruguay, Venezuela.
Sub-Saharan Africa: Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Comoros, Congo, Equatorial Guinea, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Seychelles, Sierra Leone, South Africa, South Sudan, Tanzania, Togo, Uganda, Zambia, Zimbabwe.

South-East Asia: Brunei, Cambodia, Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam.

Income Subdivision (Table 4.10)

High Income: Andorra, Aruba, Australia, Austria, Bahrain, Barbados, Bermuda, Canada, Chile, Croatia, Cyprus, Denmark, Estonia, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hungary, Iceland, Ireland, Isle of Man, Israel, Italy, Japan, Kuwait, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Nauru, New Caledonia, New Zealand, Norway, Oman, Panama, Poland, Portugal, Puerto Rico, Qatar, Romania, San Marino, Saudi Arabia, Seychelles, Singapore, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, United States, Uruguay.
Upper Middle Income: Albania, Argentina, Armenia, Azerbaijan, Belarus, Belize, Botswana, Brazil, Bulgaria, China, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Equatorial Guinea, Fiji, Gabon, Georgia, Grenada, Guatemala, Guyana, Iraq, Jamaica, Jordan, Kazakhstan, Kosovo, Libya, Malaysia, Maldives, Mauritius, Mexico, Moldova, Montenegro, Namibia, Palau, Paraguay, Peru, Serbia, South Africa, Suriname, Thailand, Tonga, Turkmenistan.
Lower Middle Income: Algeria, Angola, Bangladesh, Benin, Bhutan, Bolivia, Cambodia, Cameroon, Comoros, Djibouti, El Salvador, Ghana, Haiti, Honduras, India, Indonesia, Kenya, Kiribati, Lebanon, Lesotho, Mauritania, Mongolia, Morocco, Myanmar, Nepal, Nicaragua, Nigeria, Pakistan, Papua New Guinea, Philippines, Samoa, Senegal, Sri Lanka, Tajikistan, Tanzania, Tunisia, Ukraine, Uzbekistan, Vanuatu, Zimbabwe.
Low Income: Afghanistan, Burkina Faso, Burundi, Central African Republic, Chad, Eritrea, Ethiopia, Guinea, Guinea-Bissau, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Rwanda, Sierra Leone, Somalia, South Sudan, Sudan, Togo, Uganda, Zambia.

Governance Subdivision (Table 4.11)

High Level of Governance: Andorra, Anguilla, Aruba, Australia, Austria, Barbados, Bermuda, Botswana, Canada, Chile, Croatia, Cyprus, Denmark, Dominica, Estonia, Finland, France, French Guiana, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Liechtenstein, Lithuania, Luxembourg, Malaysia, Malta, Martinique, Mauritius, Netherlands Antilles, New Zealand, Norway, Poland, Portugal, Puerto Rico, Qatar, Reunion, Samoa, Seychelles, Singapore, Slovenia, South Africa, Spain, Sweden, Switzerland, United Arab Emirates, United Kingdom, United States, Uruguay.
Medium Level of Governance: Algeria, Argentina, Armenia, Bahrain, Belize, Benin, Bhutan, Brazil, Bulgaria, Burkina Faso, China, Colombia, Costa Rica, Cuba, Dominican Republic, El Salvador, Fiji, Georgia, Ghana, Grenada, Guam, Guyana, Honduras, India, Indonesia, Jamaica, Jordan, Kazakhstan, Kiribati, Kuwait, Lebanon, Lesotho, Madagascar, Maldives, Mauritania, Mexico, Mongolia, Montenegro, Morocco, Mozambique, Namibia, Oman, Pakistan, Panama, Philippines, Romania, Saudi Arabia, Senegal, Serbia, Sri Lanka, Suriname, Tanzania, Thailand, Tonga, Trinidad and Tobago, Tunisia, Uganda, Ukraine, Vanuatu, Vietnam.
Low Level of Governance: Afghanistan, Albania, Angola, Azerbaijan, Bangladesh, Belarus, Bolivia, Burundi, Cambodia, Cameroon, Central African Republic, Chad, Comoros, Djibouti, Ecuador, Equa-

torial Guinea, Eritrea, Ethiopia, Gabon, Guatemala, Guinea, Guinea-Bissau, Haiti, Iraq, Kenya, Liberia, Libya, Malawi, Mali, Moldova, Myanmar, Nepal, Nicaragua, Niger, Nigeria, Papua New Guinea, Paraguay, Peru, Rwanda, Sierra Leone, Somalia, Sudan, Tajikistan, Togo, Turkmenistan, Uzbekistan, Zambia, Zimbabwe.

PET Subdivision (Table 4.12)

High PET: Afghanistan, Akrotiri, Algeria, Angola, Anguilla, Antigua & Barbuda, Aruba, Australia, Bahrain, Barbados, Belize, Benin, Bermuda, Botswana, British Virgin Is., Burkina Faso, Cambodia, Cape Verde, Cayman Is., Central African Republic, Chad, Comoros, Cuba, Cyprus, Dhekelia, Djibouti, Dominica, Dominican Republic, Egypt, El Salvador, Eritrea, Ethiopia, Ghana, Grenada, Guadeloupe, Guatemala, Guinea, Guinea-Bissau, Haiti, Honduras, India, Iran, Iraq, Israel, Jamaica, Jordan, Kenya, Kiribati, Kuwait, Lebanon, Libya, Madagascar, Malawi, Maldives, Mali, Marshall Is., Martinique, Mauritania, Mayotte, Mexico, Micronesia, Montserrat, Morocco, Mozambique, Namibia, Nauru, Netherlands Antilles, New Caledonia, Nicaragua, Niger, Nigeria, Northern Mariana Is., Oman, Pakistan, Palestine, Panama, Paraguay, Puerto Rico, Qatar, Saint Lucia, Saint Martin, Saudi Arabia, Senegal, Seychelles, Sint Maarten, Somalia, South Africa, South Sudan, Sri Lanka, St. Kitts & Nevis, St. Vincent & the Grenadines, Sudan, Suriname, Swaziland, Syria, Tanzania, Thailand, The Bahamas, The Gambia, Timor Leste, Togo, Trinidad and Tobago, Tunisia, Turkmenistan, Turks & Caicos Is., Uganda, United Arab Emirates, Uzbekistan, Venezuela, Virgin Is., Yemen, Zambia, Zimbabwe.
Low PET: Albania, American Samoa, Andorra, Argentina, Armenia, Austria, Azerbaijan, Bangladesh, Belarus, Bhutan, Bolivia, Bosnia & Herzegovina, Brazil, Brunei, Bulgaria, Burundi, Cameroon, Canada, Chile, China, Colombia, Congo, Congo DRC, Cook Is., Costa Rica, Croatia, Czech Republic, Denmark, Ecuador, Equatorial Guinea, Estonia, Faroe Is., Fiji, Finland, France, French Guiana, French Polynesia, Gabon, Georgia, Germany, Gibraltar, Greece, Greenland, Guam, Guernsey, Guyana, Hong Kong, Hungary, Iceland, Indonesia, Ireland, Isle of Man, Italy, Japan, Jersey, Kazakhstan, Korean islands under UN jurisdiction, Kosovo, Kyrgyzstan, Lao People's Democratic Republic, Latvia, Lesotho, Liberia, Liechtenstein, Lithuania, Luxembourg, Macau, Macedonia, Malaysia, Malta, Mauritius, Moldova, Monaco, Mongolia, Montenegro, Myanmar, Nepal, New Zealand, North Korea, Norway, Palau, Papua New Guinea, Peru, Philippines, Poland, Portugal, Reunion, Romania, Russia, Rwanda, Samoa, San Marino, Sao Tome & Principe, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Is., South Korea, Spain, St. Helena, St. Pierre & Miquelon, Sweden, Switzerland, Taiwan, Tajikistan, Tonga, Turkey, Ukraine, United Kingdom, United States, Uruguay, Vanuatu, Vatican City, Vietnam, Wallis & Futuna.

Type Subdivision (Table 4.13)

Regarding the subdivision based on the different type of conflict, the countries are the same for each class:
Afghanistan, Akrotiri, Albania, Algeria, American Samoa, Andorra, Angola, Anguilla, Antigua & Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Barbados, Belarus, Belize, Benin, Bermuda, Bhutan, Bolivia, Bosnia & Herzegovina, Botswana, Brazil, British Virgin Is., Brunei, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Cape Verde, Cayman Is., Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo, Congo DRC, Cook Is., Costa

Rica, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Dhekelia, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Faroe Is., Fiji, Finland, France, French Guiana, French Polynesia, Gabon, Georgia, Germany, Ghana, Gibraltar, Greece, Greenland, Grenada, Guadeloupe, Guam, Guatemala, Guernsey, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Isle of Man, Israel, Italy, Jamaica, Japan, Jersey, Jordan, Kazakhstan, Kenya, Kiribati, Korean islands under UN jurisdiction, Kosovo, Kuwait, Kyrgyzstan, Lao People's Democratic Republic, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, Macau, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Is., Martinique, Mauritania, Mauritius, Mayotte, Mexico, Micronesia, Moldova, Monaco, Mongolia, Montenegro, Montserrat, Morocco, Mozambique, Myanmar, Namibia, Nauru, Nepal, Netherlands Antilles, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, North Korea, Northern Mariana Is., Norway, Oman, Pakistan, Palau, Palestine, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Reunion, Romania, Russia, Rwanda, Saint Lucia, Saint Martin, Samoa, San Marino, Sao Tome & Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Sint Maarten, Slovakia, Slovenia, Solomon Is., Somalia, South Africa, South Korea, South Sudan, Spain, Sri Lanka, St. Helena, St. Kitts & Nevis, St. Pierre & Miquelon, St. Vincent & the Grenadines, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, The Bahamas, The Gambia, Timor Leste, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Turks & Caicos Is., Uganda, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Uzbekistan, Vanuatu, Vatican City, Venezuela, Vietnam, Virgin Is., Wallis & Futuna, Yemen, Zambia, Zimbabwe.

A.3 Other Tables

```
 [1] "Falkland Islands"                        "South Georgia and South Sandwich Islands" "Lake Sevan"
 [4] "British Indian Ocean Territory"          "Christmas I."                             "Cocos Is."
 [7] "Coral Sea Islands (Australia)"           "Lake Enriquilla"                          "Falkland Is."
[10] "French Southern & Antarctic Lands"       "Heard I. & McDonald Is."                  "In dispute-Paracel Islands"
[13] "In dispute Belize/Honduras"              "In dispute Benin/Burkina Faso"            "In dispute Bhutan/China"
[16] "In dispute Brazil/Uruguay"               "In dispute Canada/Denmark"                "In dispute Croatia/Slovenia"
[19] "In dispute Djibouti/Eritrea"             "In dispute Egypt/Saudi Arabia"            "In dispute El Salvador/Honduras"
[22] "In dispute Equatorial Guinea/Gabon"      "In dispute India/Nepal"                   "In dispute India/Pakistan"
[25] "In dispute Morocco/Spain"                "Lake Urmia"                               "Jan Mayen"
[28] "Jan Mayen (Norway)"                      "Jarvis I."                                "Waterbody"
[31] "Pitcairn Is."                            "Sint Eustatius (Neth)"                    "South Georgia & the South Sandwich Is."
[34] "Tokelau"                                 "Tuvalu"                                   "Niue"
[37] "Norfolk Islands"
```

Figure A.1: List of regions removed from the conflict dataset.

B Codes

B.1 Dataset Reshaping

```
# These are the packages used for the dataset preparation
suppressPackageStartupMessages({library(sf);library(sp);library(plyr);library(raster);
library(ncdf4);library(exactextractr);library(dplyr);library(stringr);library(reshape2);
library(ggplot2);library(ggrepel);library(lubridate);library(zoo);library(foreign);
library(countrycode); library(fixest); library(broom);library(knitr);
library(stargazer);library(xtable)} )

############################################################################
#### INITIAL OPERATIONS FOR THE SHAPEFILE ####
############################################################################

# Open the datset
shp <- sf::read_sf("~Data/~Raw_Data/world_geolev1_2021/world_geolev1_2021.shp")

# Remove the undesired variables
shp$BPL_CODE=NULL; shp$CNTRY_CODE=NULL

# Remove regions with geometry error and invalid geometries
empty <- st_is_empty(shp); shp <- shp[!empty, ]
shp <- shp[st_is_valid(shp), ]

# Rename the variables country and region
shp <- shp %>%
  rename(country = CNTRY_NAME,
         region = ADMIN_NAME)
# Set the country name equal to the region if the country has no regions
shp$region <- ifelse(is.na(shp$region), shp$country, shp$region)

# Set the CRS
shp <- sf::st_transform(shp, sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"))


############################################################################
#### GWS DATASET ####
############################################################################

# Open the dataset
r <- raster::brick("~Data/~Raw_Data/ISIMIP3a/cwatm_gswp3-
w5e5_obsclim_histsoc_default_groundwstor_global_monthly_1901_2019.nc")

# Set the same CRS of the shapefile
proj4string(r) <- raster::crs(shp)

# Annual mean for all the rasters
media_annuale <- lapply(1:119, function(i) {
  anno_iniziale <- (i - 1) * 12 + 1
  anno_finale <- i * 12
  media <- mean(r[[anno_iniziale:anno_finale]])
  return(media)})

# New rasterbrick with annual averaged values
gws <- brick(media_annuale)

# Set the format and the name for the variable
years <- unique(format(as.Date(names(r), format = "X%Y.%m.%d"), "%Y"))
names(gws) <- paste0("gws", years)

# Merging data
gws_t <- gws
gws <- exactextractr::exact_extract(gws_t, shp, fun="mean")

# Add columns for regions and countries
gws$region <- shp$region ; gws$country <- shp$country; gws$orig<-shp$GEOLEVEL1

# Reshape the dataset into a long form
gws <- reshape2::melt(gws, id.vars=c("country", "region", "orig"))
```

```
# Rename the years
gws$variable <- gsub("mean.X", "", gws$variable)  ## Remove "mean.X"
gws$year <- as.integer(gsub("\\D", "", gws$variable)) + 1900
gws$variable=NULL
gws <- gws[, c("year","country", "region", "value","orig")]

# Since the migration data starts in 1960 (I need 10 years for the averages
and the anomalies)
gws <- gws %>%
  filter(year > 1958)


###############################################################################
####  POPULATION DATASET  ####
###############################################################################

file_info <- list(
  "1975" = "^Data/^Raw_Data/population/GHS_POP_E1975_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E1975_GLOBE_R2023A_54009_1000_V1_0.tif",
  "1980" = "^Data/^Raw_Data/population/GHS_POP_E1980_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E1980_GLOBE_R2023A_54009_1000_V1_0.tif",
  "1985" = "^Data/^Raw_Data/population/GHS_POP_E1985_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E1985_GLOBE_R2023A_54009_1000_V1_0.tif",
  "1990" = "^Data/^Raw_Data/population/GHS_POP_E1990_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E1990_GLOBE_R2023A_54009_1000_V1_0.tif",
  "1995" = "^Data/^Raw_Data/population/GHS_POP_E1995_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E1995_GLOBE_R2023A_54009_1000_V1_0.tif",
  "2000" = "^Data/^Raw_Data/population/GHS_POP_E2000_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E2000_GLOBE_R2023A_54009_1000_V1_0.tif",
  "2005" = "^Data/^Raw_Data/population/GHS_POP_E2005_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E2005_GLOBE_R2023A_54009_1000_V1_0.tif",
  "2010" = "^Data/^Raw_Data/population/GHS_POP_E2010_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E2010_GLOBE_R2023A_54009_1000_V1_0.tif",
  "2015" = "^Data/^Raw_Data/population/GHS_POP_E2015_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E2015_GLOBE_R2023A_54009_1000_V1_0.tif",
  "2020" = "^Data/^Raw_Data/population/GHS_POP_E2020_GLOBE_R2023A_54009_1000_V1_0/
--GHS_POP_E2020_GLOBE_R2023A_54009_1000_V1_0.tif")

for (year in names(file_info)) {
  file_tiff <- file_info[[year]]
  pop_t <- raster(file_tiff)
  pop <- exactextractr::exact_extract(pop_t, shp, fun="sum")

  # Crea un dataframe per l'anno corrente
  pop_df <- data.frame(pop = pop)
  pop_df$region <- shp$region
  pop_df$country <- shp$country
  pop_df$year <- as.integer(year)
  assign(paste0("pop", year), pop_df)}

datasets <- list(pop1975, pop1980, pop1985, pop1990, pop1995, pop2000,
pop2005, pop2010, pop2015, pop2020)
pop <- bind_rows(datasets)

# Funzione per copiare i valori di pop sugli anni mancanti
add_missing_years <- function(data) {
  years <- seq(1975, 2020, by = 5)
  new_data <- data.frame()
  for (year in years) {
    current_rows <- data %>% filter(year == year)
    for (offset in 1:4) {
      new_rows <- current_rows %>%
        mutate(year = year - offset)
      new_data <- bind_rows(new_data, new_rows)}}
  combined_data <- bind_rows(data, new_data) %>%
    distinct()
  return(combined_data)}

# Applica la funzione a ciascuna combinazione di country e region
pop <- pop %>%
  group_by(country, region) %>%
  do(add_missing_years(.)) %>%
  ungroup()

pop <- pop %>%
```

```r
  select(year, country, region, pop)


#############################################################################
####   PET DATASET   ####
#############################################################################

# Open the dataset and set the coordinate system
pet_t <- raster::brick("~Data/~Raw_Data/Global-AI_ET0_v3_annual/et0_v3_yr.tif")
proj4string(pet_t) <- raster::crs(shp)

# Reduce the resolution
factor <- 0.25 / res(pet_t)[1]
pet_t <- aggregate(pet_t, fact=factor, fun=mean, expand=TRUE)

# Merging data
pet <- exactextractr::exact_extract(pet_t, shp, fun="mean")

# Create a new dataset
country <- shp$country
pet <- data.frame(country=country, pet = pet)

# Media per ogni country
pet <- pet %>%
  group_by(country) %>%
  summarize(pet = mean(pet))

# Ordinare il dataset in base al valore di PET
pet <- pet %>%
  arrange(pet)


#############################################################################
####   MIGRATION DATASET   ####
#############################################################################

# Open the datasets
migr <-read.csv("~Data/~Raw_Data/Global_migr_raw.csv")

# Sort the order of the variables ofmigr dataset
migr <- migr[,c("year", "country_name", "worldregion",
"population","mig_interval","year_cat10","flow","flow_annual",
"outflow_rate_annual", "orig")]

# Rename variables
migr <- migr %>%
  rename(country = country_name,
         interval=mig_interval)

# Convert the values of 'orig' of gws into integers
gws$orig <- as.integer(gws$orig)

# Merge the datasets
gws_migr <- left_join(gws, migr, by=c("year", "orig"))

# Sort and rename the variables
gws_migr <- gws_migr %>%
  rename(country=country.x)
gws_migr$country.y=NULL
gws_migr <- gws_migr[,c("year", "country", "region", "worldregion",
"value", "population","interval", "flow","flow_annual",
"outflow_rate_annual","year_cat10", "orig")]

# Merge with population values
gws_migr <- merge(gws_migr, pop, by = c("year", "country", "region"), all.x = TRUE)

# Remove undesired variables
gws_migr$orig=NULL; gws_migr$flow_annual=NULL; gws_migr$worldregion=NULL
gws_migr$outflow_rate_annual=NULL; gws_migr$flow_annual=NULL; gws_migr$year_cat10=NULL

# Remove regions with NA values for 'value' variable (for regions such as Antarctica)
gws_migr <- gws_migr %>%
  filter(!is.na(value))

# Select only countries for which there are migration measures
data <- subset(gws_migr, flow>0)
```

```
data <- data %>%
  filter(!is.na(flow))
nomi <- unique(data$region)
gws_migr <- gws_migr %>%
  filter(region %in% nomi)

# Number of migrants divided by the population of that region
gws_migr <- gws_migr %>%
  mutate(migrants=(flow/pop))

# GWS per capita value
gws_migr <- gws_migr %>%
  mutate(value_t = value)
gws_migr <- gws_migr %>%
  mutate(value = value/pop)

# Normalization of gws per capita
gws_migr <- gws_migr %>%
  mutate(n_value = log(1+value))

# Normalization of migrants
gws_migr <- gws_migr %>%
  mutate(n_migr = log(1+migrants))

# 1-5-10 years averages for normalized value
gws_migr <- gws_migr %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(n_gws_avg1 = (lag(n_value) + n_value)/2,
         n_gws_avg5 = rollmean(n_value, k = 5, align = "right", fill = NA),
         n_gws_avg10 = rollmean(n_value, k = 10, align = "right", fill = NA))

# GWS logarithmic return for 1-5-10 years
gws_migr <- gws_migr %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(gws_logret=(log(value/(lag(value, n=1)))),
         gws_logret5=(log(value/(lag(value, n=4)))),
         gws_logret10=(log(value/(lag(value, n=9)))))

# GWS anomalies for1-5-10 years
# Create two new variables: mean and std over 1980-2010
medie <- gws_migr %>%
  select(year, country, region, n_value)
medie <- medie %>%
  filter(year >= 1980 & year <= 2010) %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(mean_region = mean(n_value))
medie$year <- NULL; medie$n_value <- NULL
medie <- medie %>%
  distinct(country, region, .keep_all = TRUE)
gws_migr <- left_join(gws_migr,medie,by=c("country","region"))
std_t <- gws_migr %>%
  select(year, country, region, n_value)
std_t <- std_t %>%
  filter(year >= 1980 & year <= 2010) %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(std = sd(n_value))
std_t$year <- NULL; std_t$n_value <- NULL
std_t <- std_t %>%
  distinct(country, region, .keep_all = TRUE)
gws_migr <- left_join(gws_migr,std_t,by=c("country","region"))
# Create anomalies for 1, 5, 10 years (averages)
gws_migr <- gws_migr %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(gws_anomalies = (n_value-mean_region)/std,
         gws_anomalies5 = (n_gws_avg5-mean_region)/std,
         gws_anomalies10 = (n_gws_avg10-mean_region)/std)

# Coefficiente di variazione (%)
# GWS standard deviation for 1-5-10 years
gws_migr <- gws_migr %>%
  arrange(year, country, region) %>%
```

```
  group_by(country, region) %>%
  mutate(gws_std1= rollapply(n_value, width = 2, FUN = sd, align = "right", fill = NA),
         gws_std5= rollapply(n_value, width = 5, FUN = sd, align = "right", fill = NA),
         gws_std10= rollapply(n_value, width = 10, FUN = sd, align = "right", fill = NA))
gws_migr <- gws_migr %>%
  arrange(year, country, region) %>%
  group_by(country, region) %>%
  mutate(CV1=(gws_std1/mean_region)*100,
         CV5=(gws_std5/mean_region)*100,
         CV10=(gws_std10/mean_region)*100)

# Remove useless values
gws_migr <- gws_migr %>%
  filter(!is.na(population))
gws_migr <- gws_migr %>%
  filter(!is.na(CV10))
# Remove regions ini which flow>pop
gws_migr <- subset(gws_migr, n_migr<0.6)

write.csv(gws_migr, paste0("~Data/", "gws_migr", ".csv"), row.names=FALSE)


################################################################################
####  CONFLICT DATASET  ####
################################################################################

# Open the conflict dataset
events <- read.csv("~Data/~Raw_Data/Conflict_Data/Global.csv")

# Select the variables of interest
events <- events[, c("country" ,"year", "type_of_violence","latitude" ,
"longitude", "best")]

# Rename the variables
events <- events %>%
  rename(type = type_of_violence,
         number_best = best)
events <- mutate(events,
                 type = case_when(
                   type == 1 ~ "state",
                   type == 2 ~ "Nstate",
                   type == 3 ~ "onesided"))

# Set the coordinate system
events <- st_as_sf(events, coords = c("longitude", "latitude"), crs = st_crs(shp))
events <- st_transform(events, st_crs(shp))

# Intersection shapefile-events and aggregate data
events_joined <- st_join(events, shp)
events_joined <- events_joined %>%
  rename(country = country.y)
events_joined$geometry=NULL
events_joined$country.x=NULL

# Create 2 variables: number of conflicts and best (per year)
events1 <- events_joined %>%
  group_by(year, country, region, type, GEOLEVEL1) %>%
  summarise(best = sum(number_best, na.rm = TRUE))
events2 <- events_joined %>%
  group_by(year, country, region, type, GEOLEVEL1) %>%
  summarise(conflicts = n())

events <- left_join(events1, events2, by=c("year", "country","region","type","GEOLEVEL1"))
events <- events[, c("year","country", "region","type","best", "conflicts","GEOLEVEL1")]

# Rename GEOLEVEL1 -> orig
events <- events %>%
  rename(orig = GEOLEVEL1)

# Sort datasets by year
events <- events[order(events$country),]
events <- events[order(events$year),]

events_data <- events %>%
  filter(year<2020)
```

```
vettore <- expand.grid(year=1944:2019, type=c("state","Nstate","onesided"))
gws_events <- left_join(gws, vettore, by=c("year"))

# Merge the datasets
gws_events <- left_join(gws_events,events_data,by=c("country","region","year","type","orig"))
gws_events$best[is.na(gws_events$best)] = 0  ## Assign a zero to each
month/province where no data is observed
gws_events$conflicts[is.na(gws_events$conflicts)] = 0  ## Assign a zero
to each month/province where no data is observed
gws_events <- gws_events[, c("year","country", "region","type","best",
"conflicts","value","orig")]

# Merge with population values
gws_events <- merge(gws_events, pop, by = c("year", "country", "region"), all.x = TRUE)

# Since the conflict datasets start from 1989 i just need data from 1979
gws_events <- gws_events %>%
  filter(year>1978)
gws_events$orig=NULL; gws_events$best=NULL
gws_events <- gws_events %>%
  filter(!is.na(value))  ## Regioni come Antartide in cui non ci sono valori di GWS

# Eliminare le regioni che hanno almeno un anno con pop<2000
data <- subset(gws_events, pop<2000)
nomi <- unique(data$region)
gws_events <- subset(gws_events, !(region %in% nomi))

# GWS PER CAPITA VALUE
gws_events <- gws_events %>%
  mutate(value_t = value)
gws_events <- gws_events %>%
  mutate(value = value/pop)

# TOTAL NUMBER OF CONFLICTS PER YEAR
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(year, country, region) %>%
  mutate(count = sum(conflicts))

# NORMALIZATION OF CONLFLITCS
gws_events <- gws_events %>%
  mutate(n_confl = log(1+conflicts))

# NORMALIZATION OF COUNT
gws_events <- gws_events %>%
  mutate(n_count = log(1+count))

# NORMALIZATION OF GWS VALUES (fortemente non simmetrica)
gws_events <- gws_events %>%
  mutate(n_value = log(1+value))

# AVERAGES FOR 1-5-10 YEARS (NORMALIZED)
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(n_gws_avg1 = (lag(n_value) + n_value)/2,
         n_gws_avg5 = rollmean(n_value, k = 5, align = "right", fill = NA),
         n_gws_avg10 = rollmean(n_value, k = 10, align = "right", fill = NA))

# GWS LOGARITHMIC RETURN 1-5-10 YEARS
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(gws_logret=(log(value/(lag(value, n=1)))),
         gws_logret5=(log(value/(lag(value, n=4)))),
         gws_logret10=(log(value/(lag(value, n=9)))))

# GWS STANDARD DEVIATION 1-5-10 YEARS
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(gws_std1= rollapply(n_value, width = 2, FUN = sd, align = "right", fill = NA),
         gws_std5= rollapply(n_value, width = 5, FUN = sd, align = "right", fill = NA),
         gws_std10= rollapply(n_value, width = 10, FUN = sd, align = "right", fill = NA))

# ANOMALIES (1980-2010)
```

```
medie <- gws_events %>%
  select(year, country, region, type, n_value)
medie <- medie %>%
  filter(year >= 1980 & year <= 2010) %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(mean_region = mean(n_value))
medie$year <- NULL; medie$type <- NULL; medie$n_value <- NULL
medie <- medie %>%
  distinct(country, region, .keep_all = TRUE)
gws_events <- left_join(gws_events,medie,by=c("country","region"))
std_t <- gws_events %>%
  select(year, country, region, type, n_value)
std_t <- std_t %>%
  filter(year >= 1980 & year <= 2010) %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(std = sd(n_value))
std_t$year <- NULL; std_t$type <- NULL; std_t$n_value <- NULL
std_t <- std_t %>%
  distinct(country, region, .keep_all = TRUE)
gws_events <- left_join(gws_events,std_t,by=c("country","region"))
# Create anomalies for 1, 5, 10 years (averages)
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(gws_anomalies = (n_value-mean_region)/std,
         gws_anomalies5 = (n_gws_avg5-mean_region)/std,
         gws_anomalies10 = (n_gws_avg10-mean_region)/std)

# Coefficiente di variazione (%)
gws_events <- gws_events %>%
  arrange(year, country, region, type) %>%
  group_by(country, region, type) %>%
  mutate(CV1=(gws_std1/mean_region)*100,
         CV5=(gws_std5/mean_region)*100,
         CV10=(gws_std10/mean_region)*100)

gws_events <- gws_events %>%
  filter(year>1988)
gws_events$gws_logret[is.nan(gws_events$gws_logret)] <- 0
gws_events$gws_logret5[is.nan(gws_events$gws_logret5)] <- 0
gws_events$gws_logret10[is.nan(gws_events$gws_logret10)] <- 0
gws_events$gws_anomalies[is.nan(gws_events$gws_anomalies)] <- 0
gws_events$gws_anomalies5[is.nan(gws_events$gws_anomalies5)] <- 0
gws_events$gws_anomalies10[is.nan(gws_events$gws_anomalies10)] <- 0
gws_events$CV1[is.nan(gws_events$CV1)] <- 0
gws_events$CV5[is.nan(gws_events$CV5)] <- 0
gws_events$CV10[is.nan(gws_events$CV10)] <- 0

write.csv(gws_events, paste0("~Data/", "gws_events", ".csv"), row.names=FALSE)
```

B.2 Linear Analysis

```
##########################################################################
#### CONFLICTS ANALYSIS ####
##########################################################################

# Upload of the groundwater-events dataset
ge <- read.csv("~Data/gws_events.csv")
# Create a subset of the dataset (because the variables are counted thrice (one for each type of conflict)
events_sum <- subset(ge, type=="state")

## Global data
model <- fixest::feglm(data=events_sum, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret,
gws_logret5, gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(model)

## GEOGRAPHIC SUBDIVISION
```

```
continent <- "Africa"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- events_sum %>%
  filter(get_continent(country) == continent)
Africa <- fixest::feglm(data = data_continent, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret,
gws_logret5, gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(Africa)
continent <- "Asia"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- events_sum %>%
  filter(get_continent(country) == continent)
Asia <- fixest::feglm(data = data_continent, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret,gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(Asia)
continent <- "Oceania"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- events_sum %>%
  filter(get_continent(country) == continent)
Oceania <- fixest::feglm(data = data_continent, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5, gws_logret10)|region + year, fam
tabella <- etable(Oceania)
continent <- "Europe"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- events_sum %>%
  filter(get_continent(country) == continent)
Europe <- fixest::feglm(data = data_continent, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(Europe)
North_America <- c("United States", "Canada", "Mexico")
N_A <- events_sum[events_sum$country %in% North_America, ]
Namerica <- fixest::feglm(data = N_A, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(Namerica)
South_America <- c("Argentina", "Bolivia", "Brazil", "Chile", "Colombia", "Ecuador", "Guyana", "Paraguay", "Peru", "Suri
S_A <- events_sum[events_sum$country %in% South_America, ]
Samerica <- fixest::feglm(data = S_A, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(Samerica)
Central_America <- c("Belize", "Costa Rica", "El Salvador", "Guatemala", "Honduras", "Nicaragua", "Panama", "Bahamas", "
C_A <- events_sum[events_sum$country %in% Central_America, ]
Camerica <- fixest::feglm(data = C_A, count~sw(n_value,n_gws_avg5,n_gws_avg10,gws_anomalies, gws_anomalies5, gws_anomali
tabella <- etable(Camerica)

## LIST OF NAMES OF COUNTRIES FOR: MENA, SUB-SAHARAN AFRICA, SOUTH-EAST ASIA, CENTRAL AND SOUTH AMERICA
lista_1 <- sahel_mena_countries <- c(
  "Algeria", "Bahrain", "Chad", "Djibouti", "Egypt", "Eritrea", "Iran", "Iraq",
  "Israel", "Jordan", "Kuwait", "Lebanon", "Libya", "Mauritania", "Morocco",
  "Oman", "Qatar", "Saudi Arabia", "Somalia", "Sudan", "Syria", "Tunisia",
  "United Arab Emirates", "Yemen", "Western Sahara")
data_1 <- events_sum[events_sum$country %in% lista_1, ]
lista_2 <- c("Angola", "Benin", "Botswana", "Burkina Faso", "Burundi", "Cabo Verde",
             "Cameroon", "Central African Republic", "Comoros", "Congo",
             "Democratic Republic of the Congo", "Equatorial Guinea", "Eswatini",
             "Ethiopia", "Gabon", "Gambia", "Ghana", "Guinea", "Guinea-Bissau",
             "Ivory Coast", "Kenya", "Lesotho", "Liberia", "Madagascar", "Malawi",
             "Mali", "Mozambique", "Namibia", "Niger", "Nigeria", "Rwanda",
             "Sao Tome and Principe", "Senegal", "Seychelles", "Sierra Leone",
             "South Africa", "South Sudan", "Tanzania", "Togo", "Uganda",
             "Zambia", "Zimbabwe")
data_2 <- events_sum[events_sum$country %in% lista_2, ]
lista_3 <- c("Brunei", "Cambodia", "East Timor", "Indonesia", "Laos",
             "Malaysia", "Myanmar", "Philippines", "Singapore", "Thailand",
             "Vietnam")
data_3 <- events_sum[events_sum$country %in% lista_3, ]
lista_4 <- c("Belize", "Costa Rica", "El Salvador", "Guatemala", "Honduras",
             "Nicaragua", "Panama", "Argentina", "Bolivia", "Brazil",
             "Chile", "Colombia", "Ecuador", "Guyana", "Paraguay",
```

```
             "Peru", "Suriname", "Uruguay", "Venezuela")
data_4 <- events_sum[events_sum$country %in% lista_4, ]


mena <- fixest::feglm(data = data_1, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(mena)
sub_sahara <- fixest::feglm(data = data_2, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(sub_sahara)
sud_est_asia <- fixest::feglm(data = data_3, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(sud_est_asia)
cs_america <- fixest::feglm(data = data_4, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(cs_america)


## INCOME
# Select the GDP list for one year (2005)
gdp_data <- WDI(indicator = "NY.GDP.MKTP.PP.KD", start = 2005, end = 2005, extra = TRUE)
gdp_data <- subset(gdp_data, year== 2005)
# Divide the countries into four categories
gdp_high <- subset(gdp_data, income == "High-income");
name_high <- unique(gdp_high$country)
gdp_low <- subset(gdp_data, income == "Low-income");
name_low <- unique(gdp_low$country)
gdp_lowmid <- subset(gdp_data, income == "Lower-middle-income");
name_lowmid <- unique(gdp_lowmid$country)
gdp_highmid <- subset(gdp_data, income == "Upper-middle-income");
name_highmid <- unique(gdp_highmid$country)

ge_high <- subset(events_sum, country %in% name_high)
high <- fixest::feglm(data=ge_high, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(high)
ge_low <- subset(events_sum, country %in% name_low)
low <- fixest::feglm(data=ge_low, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(low)
ge_highmid <- subset(events_sum, country %in% name_highmid)
midhigh <- fixest::feglm(data=ge_highmid, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(midhigh)
ge_lowmid <- subset(events_sum, country %in% name_lowmid)
lowmid <- fixest::feglm(data=ge_lowmid, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(lowmid)


## GOVERNANCE
gov <- read.csv("~Data/Govern.csv")
# Select the governance list
gov <- head(gov, -5)
gov <- gov[, -c(1, 2,4)]
names(gov) <- c("country", "govern")
gov$govern <- as.numeric(gov$govern)
gov <- gov %>%
  arrange(desc(govern))
gov <- na.omit(gov)

# Create 3 classes for high, medium and low governance
gov1 <- gov[1:71, ]; name_gov1 <- unique(gov1$country)  ## high
gov2 <- gov[72:142, ]; name_gov2 <- unique(gov2$country)  ## medium
gov3 <- gov[143:213, ]; name_gov3 <- unique(gov3$country)   ## low

ge_gov1 <- subset(events_sum, country %in% name_gov1)
gov1 <- fixest::feglm(data=ge_gov1, count~sw(n_value,n_gws_avg5,n_gws_avg10,
```

```
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(gov1)
ge_gov2 <- subset(events_sum, country %in% name_gov2)
gov2 <- fixest::feglm(data=ge_gov2, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(gov2)
ge_gov3 <- subset(events_sum, country %in% name_gov3)
gov3 <- fixest::feglm(data=ge_gov3, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(gov3)


## PET
pet <- read.csv("~Data/pet.csv")
# Create 2 classes for high and low PET
pet_l <- pet[1:140, ]; name_pet_l <- unique(pet_l$country)  ## low
pet_h <- pet[141:280, ]; name_pet_h <- unique(pet_h$country)  ## high

pet_H <- subset(events_sum, country %in% name_pet_h)
pet_high <- fixest::feglm(data=pet_H, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(pet_high)
pet_L <- subset(events_sum, country %in% name_pet_l)
pet_low <- fixest::feglm(data=pet_L, count~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(pet_low)


## TYPE OF CONFLICT
state <- subset(ge, type=="state")
model <- fixest::feglm(data=state, conflicts~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(model)
Nstate <- subset(ge, type=="Nstate")
model <- fixest::feglm(data=Nstate, conflicts~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(model)
onesided <- subset(ge, type=="onesided")
model <- fixest::feglm(data=onesided, conflicts~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella <- etable(model)


###########################################################################
#### INTERNAL MIGRATION ANALYSIS ####
###########################################################################

gm <- read.csv("~Data/gws_migr.csv")
data_1 <- subset(gm, interval==1); data_5 <- subset(gm, interval==5)

## GLOBAL
model1 <- fixest::feglm(data=data_1, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
model5 <- fixest::feglm(data=data_5, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(model1); tabella5 <- etable(model5)


## GEOGRAPHIC SUBDIVISION
continent <- "Africa"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- data_1 %>%
  filter(get_continent(country) == continent)
Africa1 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
```

```
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Africa1)
data_continent <- data_5 %>%
  filter(get_continent(country) == continent)
Africa5 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(Africa5)
continent <- "Asia"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- data_1 %>%
  filter(get_continent(country) == continent)
Asia1 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Asia1)
data_continent <- data_5 %>%
  filter(get_continent(country) == continent)
Asia5 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
continent <- "Europe"
get_continent <- function(countries) {
  countrycode(countries, "country.name", "continent")}
data_continent <- data_1 %>%
  filter(get_continent(country) == continent)
Europe1 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Europe1)
data_continent <- data_5 %>%
  filter(get_continent(country) == continent)
Europe5 <- fixest::feglm(data=data_continent, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(Europe5)
North_America <- c("United States", "Canada", "Mexico")
N_A <- data_1[data_1$country %in% North_America, ]
Namerica1 <- fixest::feglm(data=N_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Namerica1)
N_A <- data_5[data_5$country %in% North_America, ]
Namerica5 <- fixest::feglm(data=N_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(Namerica5)
South_America <- c("Argentina", "Bolivia", "Brazil", "Chile", "Colombia", "Ecuador",
"Guyana", "Paraguay", "Peru", "Suriname", "Uruguay", "Venezuela", "Aruba",
"Falkland Islands", "French Guiana", "South Georgia and the South Sandwich Islands")
S_A <- data_1[data_1$country %in% South_America, ]
Samerica1 <- fixest::feglm(data=S_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Samerica1)
S_A <- data_5[data_5$country %in% South_America, ]
Samerica5 <- fixest::feglm(data=S_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(Samerica5)
Central_America <- c("Belize", "Costa Rica", "El Salvador", "Guatemala", "Honduras",
"Nicaragua", "Panama", "Bahamas", "Barbados", "Cuba", "Dominica", "Jamaica", "Haiti",
"Trinidad and Tobago", "Sint Maarten", "Saint Vincent and the Grenadines", "Saint Lucia",
"Saint Kitts and Nevis", "Puerto Rico", "Dominican Republic", "Grenada", "Martinique",
"Saint Martin", "Virgin Islands", "Turks and Caicos Islands", "Cayman Islands",
"British Virgin Islands", "Guadeloupe", "Antigua and Barbuda", "Bonaire", "Curacao",
"Saint Barthelemy", "Saba", "Saint Eustatius", "Saint Pierre and Miquelon",
"British West Indies")
C_A <- data_1[data_1$country %in% Central_America, ]
Camerica1 <- fixest::feglm(data=C_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(Camerica1)
C_A <- data_5[data_5$country %in% Central_America, ]
Camerica5 <- fixest::feglm(data=C_A, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
```

```
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(Camerica5)

# SAME LIST OF COUNTIRES FOR: MENA, SUB-SAHARAN AFRICA, SOUTH-EAST ASIA, CENTRAL AND SOUTH AMERICA
data <- subset(data_1, interval==1)
mena <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella1 <- etable(mena)
data <- subset(data_1, interval==5)
mena <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella5 <- etable(mena)
data <- subset(data_2, interval==1)
sub_sahara <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella1 <- etable(sub_sahara)
data <- subset(data_2, interval==5)
sub_sahara <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella5 <- etable(sub_sahara)
data <- subset(data_3, interval==1)
sud_est_asia <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella1 <- etable(sud_est_asia)
data <- subset(data_3, interval==5)
sud_est_asia <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella5 <- etable(sud_est_asia)
data <- subset(data_4, interval==1)
cs_america <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella1 <- etable(cs_america)
data <- subset(data_4, interval==5)
cs_america <- fixest::feglm(data = data, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=quasipoisson)
tabella5 <- etable(cs_america)


## INCOME

gm_high <- subset(gm_1, country %in% name_high)
high1 <- fixest::feglm(data=gm_high, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(high1)
gm_high <- subset(gm_5, country %in% name_high)
high5 <- fixest::feglm(data=gm_high, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(high5)
gm_low <- subset(gm_1, country %in% name_low)
low1 <- fixest::feglm(data=gm_low, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(low1)
gm_low <- subset(gm_5, country %in% name_low)
low5 <- fixest::feglm(data=gm_low, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(low5)
gm_highmid <- subset(gm_1, country %in% name_highmid)
midhigh1 <- fixest::feglm(data=gm_highmid, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(midhigh1)
gm_highmid <- subset(gm_5, country %in% name_highmid)
midhigh5 <- fixest::feglm(data=gm_highmid, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
```

```
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(midhigh5)
gm_lowmid <- subset(gm_1, country %in% name_lowmid)
lowmid1 <- fixest::feglm(data=gm_lowmid, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(lowmid1)
gm_lowmid <- subset(gm_5, country %in% name_lowmid)
lowmid5 <- fixest::feglm(data=gm_lowmid, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(lowmid5)


## GOVERNANCE
gm_gov1 <- subset(gm_1, country %in% name_gov1)
gov11 <- fixest::feglm(data=gm_gov1, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(gov11)
gm_gov1 <- subset(gm_5, country %in% name_gov1)
gov15 <- fixest::feglm(data=gm_gov1, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(gov15)
gm_gov2 <- subset(gm_1, country %in% name_gov2)
gov21 <- fixest::feglm(data=gm_gov2, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(gov21)
gm_gov2 <- subset(gm_5, country %in% name_gov2)
gov25 <- fixest::feglm(data=gm_gov2, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(gov25)
gm_gov3 <- subset(gm_1, country %in% name_gov3)
gov31 <- fixest::feglm(data=gm_gov3, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(gov31)
gm_gov3 <- subset(gm_5, country %in% name_gov3)
gov35 <- fixest::feglm(data=gm_gov3, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(gov35)


## PET
pet_H <- subset(gm_1, country %in% name_pet_h)
pet_high1  <- fixest::feglm(data=pet_H, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(pet_high1)
pet_H <- subset(gm_5, country %in% name_pet_h)
pet_high5  <- fixest::feglm(data=pet_H, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(pet_high5)
pet_L <- subset(gm_1, country %in% name_pet_l)
pet_low1 <- fixest::feglm(data=pet_L, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella1 <- etable(pet_low1)
pet_L <- subset(gm_5, country %in% name_pet_l)
pet_low5 <- fixest::feglm(data=pet_L, n_migr~sw(n_value,n_gws_avg5,n_gws_avg10,
gws_anomalies, gws_anomalies5, gws_anomalies10,CV1, CV5, CV10,gws_logret, gws_logret5,
gws_logret10)|region + year, family=gaussian)
tabella5 <- etable(pet_low5)
```

B.3 Random Forest Analysis

```
# Import the libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor, RandomForestClassifier
from sklearn.model_selection import train_test_split, RandomizedSearchCV, KFold,
cross_val_score, StratifiedKFold
from sklearn.metrics import mean_squared_error, r2_score, accuracy_score, make_scorer,
mean_absolute_error, classification_report, roc_auc_score, f1_score, confusion_matrix
import xgboost
from xgboost import XGBClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.feature_selection import RFE
from imblearn.over_sampling import RandomOverSampler
from sklearn.utils.class_weight import compute_sample_weight
import seaborn as sns

# Mount GDrive
import os, sys
from google.colab import drive
drivedir='/content/drive'
drive.mount(drivedir)
os.chdir(drivedir)
datadir=drivedir+'/MyDrive/'


# Upload datasets

gm = pd.read_csv(datadir+'gws_migr.csv')  ## Migrations
ge = pd.read_csv(datadir+'gws_events.csv')  ## Conflicts

gov = pd.read_csv(datadir+'Govern.csv')  ## Level of governance
inc = pd.read_csv(datadir+'income.csv')  ## Income

# Operations on the governance dataset
gov = gov.iloc[:-5]
gov = gov.drop(gov.columns[[0, 1, 3]], axis=1)
gov.columns = ['country', 'govern']
gov['govern'] = pd.to_numeric(gov['govern'], errors='coerce')
gov = gov.sort_values(by='govern', ascending=False)
gov = gov.dropna()

# Select features
features = ['n_value', 'n_gws_avg5', 'n_gws_avg10', 'gws_anomalies', 'gws_anomalies5',
'gws_anomalies10',
            'CV1', 'CV5', 'CV10','gws_logret', 'gws_logret5', 'gws_logret10']
scaler = StandardScaler()

# Create 3 classes for high, medium, and low governance
gov1 = gov.iloc[:71]; name_gov1 = gov1['country'].unique()  ## High Governance
gov2 = gov.iloc[71:142] ; name_gov2 = gov2['country'].unique()  ## Medium Governance
gov3 = gov.iloc[142:213]; name_gov3 = gov3['country'].unique()  ## Low Governance

# Create 4 classes for high, upper-middle, middle-low, low income
inc1 = inc[inc['income']=='High income']
inc2 = inc[inc['income']=='Upper middle income']
inc3 = inc[inc['income']=='Lower middle income']
inc4 = inc[inc['income']=='Low income']
name_inc1 = inc1['country'].unique(); name_inc2 = inc2['country'].unique()
name_inc3 = inc3['country'].unique(); name_inc4 = inc4['country'].unique()




##############################################################################
####  MIGRATIONS  ####
##############################################################################


# Classi con nomi per ciclo for
```

```
anni = [1, 5]
name_global=gm['country'].unique()
class_names = {'name_global':name_global,'name_inc1': name_inc1,'name_inc2': name_inc2,
'name_inc3': name_inc3,'name_inc4': name_inc4,'name_gov1': name_gov1,
'name_gov2': name_gov2,'name_gov3': name_gov3}
classi = [(name_global, 'Global'),(name_inc1, 'High Income'),
(name_inc2, 'Upper Middle Income'), (name_inc3, 'Middle Low Income'),
(name_inc4, 'Low Income'), (name_gov1, 'High Governance'),
(name_gov2, 'Medium Governance'), (name_gov3, 'Low Governance')]
gm[features] = gm[features].apply(pd.to_numeric, errors='coerce')

#### RANDOM FOREST ####
results = []

for anno in anni:
    for classe, class_name in classi:
        gm_y = gm[gm['interval'] == anno]  ## Select the interval of migrations
        gm_y = gm_y[gm_y['country'].isin(classe)]  ## Select the class
        gm_y = gm_y.reset_index(drop=True)
        X = gm_y[features]; y = gm_y['n_migr']  ## Select the features and the target

        if X.shape[0] < 50:
            print(f"Pochi dati per {anno} e {class_name}: {X.shape[0]}")
            continue

        # Division into train-test-validation (30% for feature selection, 60% for
        hyperparameter tuning, 10% for validation).
        X_train, X_t, y_train, y_t = train_test_split(X, y, test_size=0.4, random_state=10)
        X_test, X_cv, y_test, y_cv = train_test_split(X_t, y_t, test_size=0.15, random_state=68)
        train = X_train.shape[0]; test = X_test.shape[0]; cv = X_cv.shape[0]
        train_test_valid = f'{train} / {test} / {cv}'  ## Ratio among data

        # Best 6 variables
        rf = RandomForestRegressor(random_state=4)
        rf.fit(X_train, y_train)
        importances = rf.feature_importances_
        indices = np.argsort(importances)[::-1]
        top_features = [features[i] for i in indices[:6]]

        # Tuning Hyperparameters
        parametri = {
            'n_estimators': [i for i in range(10, 400, 10)],
            'max_depth': [i for i in range(2, 10)],
            'min_samples_split': [i for i in range(3, 10)],
            'min_samples_leaf': [i for i in range(3, 10)],
            'max_features': ['sqrt'],
            'bootstrap': [True]}

        # Randomized SearchCV
        cv = KFold(n_splits=10, shuffle=True, random_state=43)
        random_search = RandomizedSearchCV(estimator=rf, param_distributions=parametri,
                                           n_iter=50, cv=cv, n_jobs=-1, scoring='r2',
                                           verbose=0, random_state=16)
        X_test = X_test[top_features]
        random_search.fit(X_test, y_test)
        rf_best = random_search.best_estimator_

        # Predictions on the validation set
        X_cv = X_cv[top_features]
        rf_best.fit(X_test, y_test)
        y_pred_cv = rf_best.predict(X_cv)
        mse_cv = mean_squared_error(y_cv, y_pred_cv)
        r2_cv = r2_score(y_cv, y_pred_cv)

        # Creation of a dictionary for the results table
        results.append({'Anno': anno,'Classe': class_name,'MSE Validation': mse_cv,
                        'R2 Validation': r2_cv,'Migliori Variabili': top_features,
                        'Train/Test/Validation':train_test_valid})

# Table of results
results_df = pd.DataFrame(results)
results_df[' '] = results_df['Classe'] + ', ' + results_df['Anno'].astype(str) + '-year'
df_summary = results_df[[' ', 'MSE Validation', 'R2 Validation',  'Train/Test/Validation']]
df_summary = df_summary.set_index(' ')
df_summary = df_summary.rename(columns={'MSE Validation': 'MSE','R2 Validation': 'R2'})
table = df_summary.style.set_properties(**{'text-align': 'center'})
```

```
table = table.format(precision=4)
table


################################################################################
#### CONFLICTS ####
#################################################################################
# Classes with names for cycle
name_global=ge['country'].unique()
class_names = {'name_global':name_global, 'name_inc1': name_inc1,
'name_inc2': name_inc2,'name_inc3': name_inc3,'name_inc4': name_inc4,
'name_gov1': name_gov1,'name_gov2': name_gov2,'name_gov3': name_gov3}
classi = [(name_global, 'Global'),(name_inc1, 'High-Income'),
(name_inc2, 'Upper-Middle-Income'), (name_inc3, 'Middle-Low-Income'),
(name_inc4, 'Low-Income'), (name_gov1, 'High-Governance'),
(name_gov2, 'Medium-Governance'),
          (name_gov3, 'Low-Governance')]

# Conversion of 'count' into a binary variable
ge['count'] = (ge['count'] > 0).astype(int)

# Select the last three years to evaluate the goodness of the model in time
ge_prima = ge[ge['year'] < 2017]
ge_dopo = ge[ge['year'] >= 2017]

#### RANDOM FOREST ####
results = []
for classe, class_name in classi:
    ge_y = ge_prima[ge_prima['country'].isin(classe)]  ## Select the class
    ge_y = ge_y.reset_index(drop=True)
    ge_w = ge_dopo[ge_dopo['country'].isin(classe)]  ## Select the class
    ge_w = ge_w.reset_index(drop=True)

    # Subdivision into train-test-validation
    X = ge_y[features]; y = ge_y['count']
    X_train, X_t, y_train, y_t = train_test_split(X, y, test_size=0.7, random_state=17)
    ## (30% for feature selection, 50% for tuning, and 20% validation)
    X_test, X_cv, y_test, y_cv = train_test_split(X_t, y_t, test_size=0.28, random_state=80)
    train_size = X_train.shape[0]; test_size = X_test.shape[0]; cv_size = X_cv.shape[0]; prev = ge_w.shape[0]
    data_size = f'{train_size} / {test_size} / {cv_size} / {prev}'  ## Ratio among sets

    # Weight calculation for different classes
    y_train_1 = (y_train == 1).sum(); y_test_1 = (y_test == 1).sum()
    y_train_0 = (y_train == 0).sum(); y_test_0 = (y_test == 0).sum()
    count_1 = y_train_1 + y_test_1; count_0 = y_train_0 + y_test_0
    ratio = (count_1 / count_0)
    rapp = (count_1 / count_0)*100  ## Percentage of occurrences in the selected class

    # Selection of the best 6 variables
    rf = RandomForestClassifier(class_weight={0: 1, 1: 1/ratio}, random_state=29)
    rf.fit(X_train, y_train)
    importances = rf.feature_importances_
    indices = np.argsort(importances)[::-1]
    top_features = [features[i] for i in indices[:6]]

    # Hyperparameter tuning
    parametri = {
        'n_estimators': [i for i in range(100, 1001, 100)],
        'max_depth': [i for i in range(6, 21, 1)],
        'min_samples_split': [i for i in range(2, 16, 1)],
        'min_samples_leaf': [i for i in range(2, 16, 1)],
        'bootstrap': [True]}

    # RandomSearchCV
    cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=36)
    random_search = RandomizedSearchCV(estimator=rf, param_distributions=parametri,
        cv=cv, scoring='average_precision', n_jobs=-1, verbose=0, n_iter=20,
        random_state=2)
    X_test = X_test[top_features]
    random_search.fit(X_test, y_test)
    best_rf = random_search.best_estimator_

    # Predictions on the validation set
    X_cv = X_cv[top_features]
    best_rf.fit(X_test,y_test)
    y_pred_cv = best_rf.predict(X_cv)
```

```
    report_cv = classification_report(y_cv, y_pred_cv, output_dict=True)

    # Predictions over 2017-2019 data
    X_w = ge_w[top_features]; y_w = ge_w['count']
    y_pred_w = best_rf.predict(X_w)
    report_w = classification_report(y_w, y_pred_w, output_dict=True)

    # Results
    results.append({'Classe': class_name, 'Validation': report_cv,'2017-2019': report_w,
                    'Migliori Variabili': top_features,'Migliori Parametri': random_search.best_params_,
                    'Rapp': rapp,'Data Size': data_size})


# Table of the resulst
def extract_metrics_combined(results):
    classi = []; dataset_type = []; precision_0 = []; recall_0 = []
    f1_score_0 = []; precision_1 = []; recall_1 = []; f1_score_1 = []
    weight_precision = []; weight_recall = []; weight_f1_score = []; accuracy = []
    rapp_list = []; data_size_list = []

    for result in results:
        for set_name in ['Validation', '2017-2019']:
            classi.append(result['Classe'])
            dataset_type.append(set_name)
            precision_0.append(result[set_name]['0']['precision'])
            precision_1.append(result[set_name]['1']['precision'])
            weight_precision.append(result[set_name]['weighted avg']['precision'])
            f1_score_0.append(result[set_name]['0']['f1-score'])
            f1_score_1.append(result[set_name]['1']['f1-score'])
            weight_f1_score.append(result[set_name]['weighted avg']['f1-score'])
            recall_0.append(result[set_name]['0']['recall'])
            recall_1.append(result[set_name]['1']['recall'])
            weight_recall.append(result[set_name]['weighted avg']['recall'])
            accuracy.append(result[set_name]['accuracy'])
            rapp_list.append(result['Rapp'])
            data_size_list.append(result['Data Size'])

    df = pd.DataFrame({'Classe': classi,'Dataset': dataset_type,'Precision No Conflicts': precision_0,
    'Precision Conflicts': precision_1, 'Weighted Precision':weight_precision,
    'F1-score No Conflicts': f1_score_0, 'F1-score Conflicts': f1_score_1,
    'Weighted F1-score':weight_f1_score, 'Recall No Conflicts': recall_0,
    'Recall Conflicts': recall_1, 'Weighted Recall':weight_recall,
    'Rapp': rapp_list,'Data Size': data_size_list})
    return df

df_combined = extract_metrics_combined(results)
df_combined[' '] = df_combined['Classe'] + ', ' + df_combined['Dataset']
df_combined = df_combined.drop(columns=['Classe', 'Dataset'])
df_combined.set_index(' ', inplace=True)
df_combined = df_combined.round(4)
df_combined = df_combined.rename(columns={
    'Precision No Conflicts': 'Precision (No Conflicts)',
    'Precision Conflicts': 'Precision (Conflicts)',
    'Weighted Precision': 'Weighted Precision',
    'F1-score No Conflicts': 'F1-score (No Conflicts)',
    'F1-score Conflicts': 'F1-score (Conflicts)',
    'Weighted F1-score': 'Weighted F1-score',
    'Recall No Conflicts': 'Recall (No Conflicts)',
    'Recall Conflicts': 'Recall (Conflicts)',
    'Weighted Recall': 'Weighted Recall',
    'Rapp': 'Class Ratio (%)',
    'Data Size': 'Data Size (Train/Test/CV/Prev)'})

table = df_combined.style.set_properties(**{'text-align': 'center'})
table = table.set_table_styles([{'selector': 'th', 'props': [('text-align', 'center')] }])
table = table.format(precision=2)
table
```

Bibliography

(1) Copernicus: June 2024 marks 12th month of global temperature reaching 1.5°C above pre-industrial `https://climate.copernicus.eu/copernicus-june-2024-marks-12th-month-global-temperature-reaching-15degc-above-pre-industrial`.

(2) *AR6 Climate Change 2022: Impacts, Adaptation and Vulnerability*; tech. rep.; IPCC, 2022.

(3) Pasini, A.; Mastrojeni, G., *Effetto serra, effetto guerra*; Chiarelettere: 2020.

(4) ISIMIP3 simulation protocol. Water (global) `https://protocol.isimip.org/#/ISIMIP3a/water_global/42-output-variables`.

(5) *AR6 Climate Change 2023: Synthesis Report. Summary for Policymakers*; tech. rep.; IPCC, 2023.

(6) The Conversation, 10/09/2024. South Sudan floods: the first example of a mass population permanently displaced by climate change? `https://theconversation.com/south-sudan-floods-the-first-example-of-a-mass-population-permanently-displaced-by-climate-change-238461`.

(7) BBC, 22/09/2024. South America drought brings wildfires and blackouts. `https://www.bbc.com/news/articles/cly7nxz48klo`.

(8) IGRAC. What is Groundwater? `https://www.un-igrac.org/what-groundwater`.

(9) Poeter, E.; et al., *Groundwater in Our Water Cycle*; Groundwater Project: 2020.

(10) Dai, A.; Trenberth K., E. *Journal of Hydrometeorology* **2002**.

(11) Vrba, J.; Van der Gun, J. *The world's groundwater resources. Contribution to Chapter 4 of WWDR-2*; tech. rep.; IGRAC, 2004.

(12) USGS. The World's Water `https://web.archive.org/web/20131214091601/http://ga.water.usgs.gov/edu/earthwherewater.html`.

(13) Gleeson, T.; et al. *Annual Review of Earth and Planetary Sciences* **2020**.

(14) Li, B.; et al. *Water Resources Research* **2019**.

(15) Foster, S.; et al. *Environmental Sustainability* **2013**.

(16) Lall, U.; et al. *Annual Review of Environment and Resources* **2020**.

(17) Rodell, M.; et al. **2018**.

(18) Taylor, R.; et al. *Nature Climate Change* **2012**.

(19) Reinecke, R.; et al. *Hydrology and Earth System Sciences* **2021**.

(20) Kummu, M.; et al. *Nature Scientific Reports* **2016**.

(21) Huang, Z.; et al. *European Geosciences Union* **2018**.

(22) UNWWAP *The United Nations world water development report 2015: water for a sustainable world*; tech. rep.; UNESCO, 2015.

(23) CMCC. The global map of aridity `https://www.cmcc.it/article/the-global-map-of-aridity`.

(24) Buhaug, H.; et al. *Implications of Climate Change for Armed Conflict*; tech. rep.; World Bank Group, 2008.

(25) IOM, Environmental Migration `https://environmentalmigration.iom.int/environmental-migration`.

(26) Piguet, E. *WIREs* **2021**.

(27) Niva, V.; et al. *Environmental Research Letters* **2021**.

(28) Daoust, G.; Selby, J. *WIREs* **2024**.

(29) Carretta M., A.; et al. *Climatic Change* **2023**.

(30) Flavell, A.; et al. *Migration, environment and climate change: Literature review*; tech. rep.; German Environment Agency, 2019.

(31) Xu, L.; Famiglietti S., J. *WIREs WATER* **2023**.

(32) Abel G., J.; et al. *Global Environmental Change* **2019**.

(33) Hoffmann R., e. a. *Climate Change, Aridity, and Internal Migration: Evidence from Census Microdata for 72 Countries; IIASA Working Paper. Laxenburg, Austria: WP-23-008* **2023**.

(34) Wrathall J., D.; et al. *Water stress and human migration: a global, georeferenced review of empirical research*; tech. rep.; FAO, 2018.

(35) Owain E., L.; Maslin M., A. *Humanities and social sciences communications* **2018**.

(36) Nagabhatla, N.; et al. *Water and Migration: A Global Overview*; tech. rep.; 2020.

(37) Zaveri, E.; et al. *Ebb and Flow*; tech. rep.; World Bank Group, 2021.

(38) Borderon, M.; et al *Demographic Research* **2018**.

(39) Cattaneo, C.; Peri, G. *Journal of Development Economics* **2016**.

(40) Berlemann, M.; Steinhardt M., F. *Oxford Academic* **2017**.

(41) Marchiori, L.; et al. *Journal of Environmental Economics and Management* **2012**.

(42) Pasini, A.; Amendola, S. *Environmental Research Communications* **2019**.

(43) Stoler, J.; et al. *Current Opinion in Environmental Sustainability* **2021**.

(44) *Climate Change, Water Scarcity and Migration*; tech. rep.; 2023.

(45) Vesco, P.; et al. *Ecological Economics* **2020**.

(46) Cattaneo, C.; Bosetti, V. *Climate-induced international migration and conflicts*; tech. rep.; CMCC, 2017.

(47) Tsomb, E.; et al. *Environment, Development and Sustainability* **2023**.

(48) Kuobi, V. *Annual Review of Political Science* **2019**.

(49) *The Role of Water Stress in Instability and Conflict*; tech. rep.; CNA, 2017.

(50) Nagano, T.; Sekiyama, T. *MDPI Climate* **2023**.

(51) Hsiang S., M.; et al. *Nature* **2011**.

(52) Robins, N.; Fergusson, J. *Earth perspectives* **2014**.

(53) Unifried, K.; et al. *Journal of Environmental Economics and Management* **2022**.

(54) Konhert, D. **2023**.

(55) Scheffran, J.; Brauch, H. G., *The Mediterranean Sea: Its history and present challenges*; Springer Link: 2014.

(56) *A new outlook on the climate and security nexus: Addressing the impact of climate change and environmental degradation on peace, security and defence*; tech. rep.; European Commission, 2023.

(57) IPUMS International: `https://international.ipums.org/international/gis_harmonized_1st.shtml`.

(58) ISIMIP Repository: `https://data.isimip.org/datasets/74c1496d-ec24-48c6-866a-02faad6a12e8`.

(59) Joint Research Centre Data Catalogue: `https://data.jrc.ec.europa.eu/dataset/2ff68a52-5b5b-4a22-8f40-c41da8332cfe`.

(60) Global Aridity and PET Database: `https://csidotinfo.wordpress.com/data/global-aridity-and-pet-database`.

(61) Worldwide Governance Indicators: `https://databank.worldbank.org/source/worldwide-governance-indicators`.

List of Figures

List of Tables

Acknowledgement

A special thanks goes to Professor Raya Muttarak, my thesis supervisor, for giving me the opportunity to carry out an internship at the International Institute for Applied Systems Analysis (IIASA) in Vienna. Her support, along with her valuable guidance, was essential in every phase of my work.
I would also like to thank Professor Duccio Rocchini, my co-supervisor, for taking the time to review my project and provide valuable feedbacks.
I am extremely grateful to IIASA for welcoming me and providing a stimulating and opportunity-rich work environment.
Last but not least, a special thanks to Gregor Zens and Roman Hoffmann, who supervised my internship and provided invaluable support in setting up my research work.