Associating refugee populations with climate exposure

A data analytics approach

by

Thomas Rous

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday September 24, 2024 at 09:00 AM.

Student number:	4963946	
Project duration:	March 1, 2024 – September 24, 2024	
Thesis committee:	Chair - Prof. Dr. T. Comes,	TU Delft
	First supervisor - Dr. N.Y. Aydin,	TU Delft
	External supervisor - Dr. S. Fransen,	Erasmus University

This thesis is confidential and cannot be made public until September 30, 2024.

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Jupyter Lab (data analytics) files are available at https://github.com/Thomas-4444/Master_Thesis_TR.

TU Delft

Preface

This report marks the completion of the Master's program Engineering and Policy Analysis (EPA) from the faculty Technology, Policy and Management (TPM) at TU Delft. With this, my period of studying and my time as a student have come to an end. It was a great experience and I developed both personally and academically.

What I find most promising -besides the education aspect- is the variety of opportunities you get while studying. In my case, I engaged in committee work with the study association of TPM; I was the chair of the FSC of TPM; I chaired the student union (VSSD); I am the chair of the NGO Energy for Refugees, and I am still participating in the Impact Studio of TU Delft. For these chances, I am very grateful, and I found it impressive to have these types of opportunities due to TU Delft, besides developing technical and theoretical knowledge while studying.

Regarding this thesis, I would like to thank my thesis committee. First of all, thanks to Nazli Aydin, for being my weekly discussion partner. Thanks to Tina Comes for the conceptual thoughts and dataset choices. I also thank Trivik Verma for not only being a great teacher of one of the courses in the EPA program, but also for his constructive feedback during my thesis progress meetings. I thank Sonja Fransen for her theoretical input. In addition, I want to thank Maarten Kroesen and Enno Schröder for having chats with me to discuss statistical tests.

Over the past six months, I have learned a lot about refugee movements, as well as climate change, both chancing over years. Hopefully, this report will also provide you with new insights.

Thomas Rous
Delft, September 2024

Summary

Currently, the number of refugees is increasing, while the effects of climate events are worsening. Given this context, it is becoming increasingly important to understand the potential relationships between refugee movements and climate exposure, particularly at the origin and destination locations of refugees. This understanding is crucial for predicting future movements of climate-related refugees and identifying effective climate adaptation measures at both origin and destination locations.

Research on the global relationships between refugee movements and climate exposure is lacking in the current literature, especially in terms of data-driven, analytical approaches. This research addresses this gap.

The main research question of this thesis is: How is climate exposure associated with global refugee movements? To answer this, a literature review and data analytics are employed. The literature review identifies climate-related hazards that serve as inputs for the data analysis. The hazards considered in this study include droughts, sea-level rise, coastal flooding, riverine flooding, and cyclones. Various types of (possible) associations are examined using correlation analyses, regression analyses, spatial cluster identifications, and bivariate choropleth maps. All analyses consider changes over the years, specifically from 2003 to 2022.

As one of the results, this research presents an equation for relatively accurate predictions of refugees fleeing from countries in recent years, based on climate indicators while also incorporating social factors.

Moreover, the results of the study indicate that sea-level rise and riverine flooding are significantly associated with refugees fleeing from a country. Future research should explore whether this relationship could potentially be causal.

Additionally, the study identifies hotspots for different climatic conditions as well as for refugee movements (both fleeing to and from countries). The hotspots for climatic indicators and refugee movements align relatively well. Generally, climate exposure and large refugee movements are concentrated in Central-East Africa, the Middle East, and South-East Asia. These hotspots have not changed significantly over the last 20 years.

Lastly, the thesis compares top origin-destination locations of refugees at both generic (statistical) and individual (visual) levels. While statistical differences between top origin and destination locations are sometimes inconclusive, individual comparisons of countries within the same regions provide key insights into the differences in climatic exposure for refugees.

More effective climate adaptation measures can be implemented at refugee destination locations by tailoring them to specific climatic exposures, taking into account the conditions at their origin locations. Specific investments regarding adaptation measures have been identified for particular regions and countries, and these are recommended to policymakers.

The procedures and tests developed in this research can be applied to similar studies using different or newer climate and refugee data.

Contents

1 Introduction

1.1. Motivation

The number of refugees worldwide increased from 27.1 million in 2021 to 35.3 million at the end of 2022, the largest yearly increase ever recorded, according to UNHCR's statistics on forced displacement (UNHCR, 2023). Also when considering a longer time period, the number of refugees increases. At the same time, climate change is worsening (Jotzo & Howden, 2023). For example, at some places in Bangladesh, a country with also a lot of refugee movements, temperatures are rising and rainfall is more erratic and intense already (Rojas, 2021). This means that over the years, refugees will encounter even more livability problems.

Given all of this, yet limited research has been conducted on potential relations between refugee movements and climate exposure. There is, for instance, no research that examines differences in climate exposure between refugees' origin and destination locations on a global scale. Furthermore, in the literature review later in this Chapter, no worldwide, quantitative study has been focusing on climate impacts being a potential driver of refugee movements.

When looking at the international grand challenges or sustainable development goals (United Nations, n.d.), having a systematic data analytics approach to determine the refugee movement and climate impact connection would contribute to reduced inequalities (10) because of more accurate climate adaptation measures that can be arranged. Also, future refugee influx/outflow due to climate exposure can be determined more accurate if the connection between the two is more clear. Therefore, it would be highly relevant to map-out possible connections between refugee movements and climate conditions, also at origin-destination locations, and give advice to policy makers at United Nations High Commissioner for Refugees (UNHCR) and governments. This research calls for integrating the difference in climate exposure at origin and destination locations of refugees in the climate adaptation decision making process, as well as planning future climate-refugee influx/outflow more accurate. The latter means that governments and non-governmental organizations (NGOs) can better estimate how many refugees will leave a country and flee to a country in the future.

The context of finding relations between refugee movements and climate events is a complex and sociotechnical one. The technical aspects, such as combining data sources to identify origin and destination locations of refugees and climate-analysis on those data, as well as the social aspect of discovering which policy makers offer room for different climate adaptation measures make up this complex sociotechnical system. In research of Raadschelders, Larrison, and Thapar (2019), refugee movement is even categorized as a wicked problem.

This research takes a problem-driven approach. The argument for this is that an understanding of relations between climate exposure and refugee movements can help policy makers at governments and NGOs in planning for refugee movements (related to climate), and design of climate adaptation measures at origin and destination locations of refugees. Indeed, one can argue that this research is also data-driven, as in order to do this research, (climate and refugee) data needs to be available.

In scientific literature, there is a lack of understanding of the climatic conditions refugees come from and go to, as well as a global measure of climate events being a potential driver of refugee flows. With the latter, a global estimation of climate refugee streams can be developed. Having these insights would contribute to the Oxford Journal of Refugee Studies, and therein, especially to the review of Nicholson (2012), who criticized the claimed multi-disciplinary researched relation between climate and displacement. In terms of modeling refugee movements, this thesis contributes in the Journal of Refugee Studies to Edwards (2008) by including specific climate events and regression analysis. However, this thesis adds most insights to the work of Farbotko and Lazrus (2012) in the Global Environmental Change Journal, by identifying situations in which climate-vulnerable refugees are, and use that for adaptation planning, which is what their paper argued for. Also, with the identification of climate adaptation measures at origin-destination locations of refugees, there is contribution to sustainability research, especially for vulnerable people (refugees), building on research of Turner et al. (2003a) as well as Turner et al. (2003b). Specifically, theoretical relevance is in this case about identifying relative importance of climate indicators for refugees fleeing, to base adaptation measures upon. In general, contributions to the researches above represents the complete theoretical relevance of this thesis. Translated to scientific methodologies, a worldwide data analytics approach to inspect 'climate-refugee' relations is missing. Researching the global climate-refugee association with data analytics represents the methodological relevance of this thesis.

1.2. Existing theories and gaps

In this section, findings of the relevant literature are synthesized and knowledge gaps are presented. In Appendix A, the search plan is presented.

1.2.1. Findings of literature study

In this section, the funnel structure is used to report the findings from the literature study. This means that starting from a -to some extent- broader article context, the articles are narrowed down to being more specific for the research topic.

In general, the historical development in this field is low, as nearly all publications are written between 2015 and 2024 (which is noticeable in the paragraphs below). Before 2015, the link between climate and migration was not researched extensively. This means that this is an upcoming research area.

1.2.1.1. Link between migration and climate exposure

In research of Luetz (2018) is stated that climate change cannot be isolated as the definitive cause of this movement, but that it is impossible to dismiss it as a contributing factor. Similar findings are reported in a quantitative and generic study on the link between climate change and migration regarding immigrants of Ningxia, China. The authors find an interaction between climate change and migration, although the effect is not direct. There is an environmental push reaction from the origin and a pull effect of the place of destination (Han, Kumar, & Kumar, 2024). The latter is confirmed by Thorn et al. (2023), because they showed that in Namibia, there is an increase in climate-related push factors. In addition, the study of Byravan and Rajan (2022) admits that the increased risks of environmental hazards, including climate change, have intensified the push to migrate. Nevertheless, the relation between climate change and forced displacement is not direct. Moreover, in a qualitative New Zealand case study, the actual climate-migration effect seems limited. However, the issue of climate change sticks in the mind of migrants and it is still a driver of migration (Shen & Gemenne, 2011). A somewhat different, clearer, claim is made by Wesselbaum and Aburn (2019), who even state that 'climate change is a more important driver than income and political freedom at origin together'. Conversely, a study in Yemen finds that 'climate variables do affect migration, but in a limited way, with socio-economic and cost factors playing a much more prominent role' (Joseph & Wodon, 2013). Finally, in Africa is shown that the effect of climate change on migration differs per country (Mueller, Gray, & Hopping, 2020). This is interesting, because sometimes higher temperatures/more precipitation leads to decline in migration, while in other countries it leads to an increase. Therefore, it is an advantage that this thesis investigates globally per country origin and destination effects of climate change for refugees. These country-specific findings highlight the contextually-specific nature of climate and refugee movement relationships. Also, Rikani, Frieler, and Schewe (2022) write about effects of climate change on international migration, which they try to quantify and explain through economic development. This again suggests that climate effects may only have an indirect effect on migration movements. This would imply that, for example, the correlation between the number of refugees moving from a country and the climate exposure in that country is low. That is, among other things, what this thesis will examine. The authors also emphasize that the relation differs per movement flow between countries. The latter will also be investigated in this thesis. In general, the thesis will examine the potential form and magnitude of (in)direct relationships between climate exposure indicators and refugees fleeing countries worldwide.

1.2.1.2. Origin and destination location exposure

Many dimensions of migration are poorly understood. In particular, factors of migration in destination areas and areas of origin are still lacking comprehensive analysis (Pirani, Marino, & Petrucci, 2019; Schürmann, Kleemann, Teucher, Fürst, & Conrad, 2022). However, the development and implementation of environmental safety assessments and mitigation strategies for countries of origin and destination of environmental migration is earlier conducted by Reznikova and Danilina (2021). They claim that environmental migration is one of the main threats to the security of regions and countries and is linked to their sustainable development (or lack thereof).

Migration may be increasingly used as adaptation strategy to reduce populations' exposure and vulnerability to climate change impacts. Conversely, due to a lack of information at destinations, people might move to locations where they are more exposed to climate exposure than at their origin locations (Benveniste, Oppenheimer, & Fleurbaey, 2020). Also, climate exposure and vulnerability tend to be higher in developing regions. It is shown that most migrants from developing regions tend to move to areas where they are less affected by climate conditions than where they came from (Benveniste et al., 2020). The reason for this is not reflected upon. It is important to verify globally whether refugees can end up in more exposed areas. This will be examined in this thesis.

The qualitative interview study of Jennath and Paul (2024) indicates that those fleeing from environmental stress often move to nearby areas. They also found that climate conditions of the potential destination and its geographic and social proximity to area or origin, are influential factors in destination selection. For the this thesis, it is interesting to (de)validate findings on a worldwide scale regarding refugee origin-destination movements, such that climate adaptation policies can be specified by location characteristics.

1.2.1.3. Specific climate indicators regarding migration

There is a need for more research on the relation between movements and climate exposure. This is illustrated by a case study in Germany (Nowak-Lehmann, Cardozo, & Martínez-Zarzoso, 2021). They find that increasing average temperatures is mostly negatively correlated with total migration flows. However, and completely contrary to the former finding, the authors see that increasing average temperatures trigger emigration among asylum seekers. In addition, they find that an increase in average precipitation has a migration-increasing effect. Conversely, in another study that focuses on migration in Less Developed Countries, is found that there is a negative relation between rainfall and migration, but it is unknown why this would be (Lilleør & Van den Broeck, 2011).

In India, changing precipitation patterns pose pressure on rural livelihoods through the increasing frequency and severity of droughts, contributing to rural-to-urban mobility. At destination, however, insufficient information is available on the complex mobility backgrounds of the new arrivals. Combining a household survey with in-depth interviews and monthly precipitation data on district level, they show a significant relationship between drought at origin and mobility to Pune. Paradoxically, migrants affected by droughts at origin face increased flood risk at destination (Karutz & Kabisch, 2023). The latter is related to what is going to be examined with in this thesis: do refugees end up in destination locations that are heavier exposed to climate hazards than their origin destination?

In the quantitative US case study of Paglino (2024) is found that excess migration after tropical storms is rare and generally fails to reduce the number of people at risk of experiencing future natural disasters. Only the most destructive tropical storms are associated with significant excess migration. It is interesting for this thesis to research if the absence of these movement patterns is also there when looking at worldwide refugee movement data.

Sivisaca, Robalino, Cascante, Imbach, and Sandoval (2021) show the effects of weather events on migration in Guatemala. On average, the presence of a drought in a municipality of origin significantly reduces migration, although, on average, extreme precipitation events increase migration significantly. For the thesis, it is valuable to examine if the same conclusions can be drawn on a global level, regarding refugee movements. These findings show the importance of considering the type of weather events when thinking about their impact on movements.

In a study on the driving forces of international migration out of Africa, the focus is on the direct effect of climate variables: temperature, rainfall, and weather-related disaster (Wesselbaum, 2021). The findings show that climate can be a driver and an inhibitor of migration depending on the size of the temperature shock. The temperature has non-linear effects on migration. This gives already an indication of what climate factors to focus on in this research. Can these associations also be found in the results of this study?

A Brazilian Amazon study on the out-migration includes temperature, precipitation and extreme weather events as environmental change indicators (Gori Maia & Schons, 2020). It thereby confirms, like other studies reviewed above, to include these climate factors also in the global study of this research. However, in the end, the exact relation between climate change and migration is unclear; there are many possibilities (Obokata, Veronis, & McLeman, 2014). It is difficult to make general conclusions as to which types of environmental factors may be most likely to influence migration. No definite conclusions can yet be drawn with respect to how migration trends or patterns respond to specific environmental factors. Nevertheless, first instances of understanding conceptual relationships are being investigated in this study. This thesis aims to explain these relations more, specifically for refugees.

1.2.1.4. Adaptation policies

In a qualitative case study in Kenya and Ethiopia is claimed that climate change is a trigger for migration (Leal Filho et al., 2023). The study highlights the need for governments, international organizations, and other stakeholders better to understand the complex linkages between climate change and migration. Also, the paper concludes by stating the urgent need for policies and programs that support climate change-induced migrants. A call for good adaptation strategies is also made by Hauer, Jacobs, and Kulp (2024) in their quantitative US case study.

The study of Nawrotzki, Riosmena, Hunter, and Runfola (2015) in Mexico shows that social networks could facilitate climate change adaptation in place. It is argued that networks may not always have multiplier effects on migration. The possibility of a "suppression mechanism" whereby social networks facilitate climate change adaptation at home and reduce the influence of environmental factors should be considered within policy.

This study will investigate climate conditions at origin and destination locations, and from that, climate adaptation strategies can be developed to enhance livability for refugees at their new place.

1.2.2. Knowledge gaps

Based on the previous section, the following knowledge gaps are identified.

- No study focuses on associations between climate conditions and refugee movements Only migrants are considered, and not refugees specifically.
- Existing studies research the link between climate exposure and movements only in specific countries or regions. No global (worldwide) analyses have been conducted.
- Little research has been conducted on the form and magnitude of relations between climatic conditions and refugee movements. For instance, there is no current investigation into the numeric differences in climate exposure between origin and destination locations of refugees. Namely, many studies are qualitative in nature. Therefore, this quantitative thesis will produce additional numeric information besides interview -and other qualitative- insights.

1.3. Research questions

In this section, the research questions are presented. A procedural approach is adopted in constructing the sub-questions. This means that when combining the answers to the three sub-questions together, the main question can be answered. However, one can argue that sub-question 2, 3, 4 and 5 below also have issue-based characteristics, as they make individual contributions to the literature.

How is climate exposure associated with global refugee movements?

To answer the main question, five sub-questions (SQs) are defined.

SQ 1. Which events reflect climate-related hazards?

SQ 2. How strongly could climate exposure drive people to flee, over years?

SQ 3. How do countries cluster spatially over time, based on climate exposure?

SQ 4. How do countries cluster spatially over time, based on refugee movements?

SQ 5. What is the difference in climate exposure between origin and destination locations of refugees worldwide, over years?

First, relevant climate events have to be identified in order to conduct the analyses (SQ 1). After identifying these events, climate exposure as possible driver for refugee movements must be examined (SQ 2). In other words, sub-question 2 dives into the climatic reasons of people fleeing. The answer to the second sub-question explains the role of climate intensity exposure scores on refugee fleeing from/to a country as well as predictive power for refugee flows over years. In addition, geographical clusters of (dis)similar countries are identified across the globe based on climatic conditions (SQ 3) and based on refugee movements (refugees fleeing to/from countries) (SQ 4). The answer to sub-question 3 helps to map where in the world many climate alike countries are over years and the answer to sub-question 4 helps to map where in the world many/few refugee movements between countries take place over years. Together, overlap between climate exposure and refugee movement is identified. The identified cluster form the basis for sub-question 5. This last sub-question compares the climatic situations of refugee origin and destination locations over the years, explaining the differences with a specific focus on the key regions identified in sub-questions 3 and 4.

With the answers to SQ 2 - SQ 5 (for which the answer to sub-question 1 is required), one can conclude about climate exposure and refugee movements associations over years in general, and thereby answer the main question. This is because both the climatic contribution prior to fleeing (SQ 2), as well as the climatic condition differences after people have fled are examined (SQ 5). Also, clusters of (dis)similar climate exposure (SQ 3) and refugee movements (SQ 4) situations are identified across the globe. Multiple type of associations are studied with the sub-questions. The answers to these sub-questions are integrated to conclude on overall association. From this, climate adaptation policy advice for (local) governments and NGOs in different countries can be derived, along with improved planning for refugee movements.
Moreover, with answering these questions, the knowledge gaps are addressed. Firstly, the analyses in SQ 2 - SQ 5 are carried out with refugee data and not migrant data. Secondly, these analyses are carried out globally, with worldwide country data, and not regionally or locally. Lastly, all these analyses are quantitative, on which will be reflected in Chapter 2.

1.4. Concepts

This section clarifies the key terms and concepts used throughout this thesis regarding refugees as well as climate and its impacts, to ensure consistency and clarity in this report.

1.4.1. Refugees

In this thesis, specifically refugee movements are considered and not general migrant in- or outflows. The knowledge gaps indicate that many research has focused on migrant flows, and not refugees specifically. That draws attention to the difference between the two concepts.

In 1951, the Refugee Convention defined a refugee as someone who is forced to flee their own country. Refugees cannot return to their origin country due to (feared) persecution for reasons like race, politics, religion and membership of a social group, outside of their country of origin and are not protected by their origin country (UNHCR, n.d.-c).
This definition largely focuses on reasons other than climate. That also shows the importance of SQ 5 (and SQ 3 and 4 before 5): although refugees flee for other reasons than climate, how does their

climate conditions differ in their destination location, such that this already vulnerable group can receive specific help, tailored to their needs.
However, the definition above is only limited. That is why in 1969 and 1984, extended definitions were initiated specifically on the continents Africa and Latin America (UNHCR, n.d.-b). Those extensions take a more general approach in defining refugees by including several statements. In Africa, the refugee definition now includes those fleeing due to "events seriously disturbing public order in either part or the whole of the country of origin". In Latin America, also "other circumstances which have seriously disturbed public order" are taken into account. This broadens the refugee scope.
These new definitions offer more room for climate being also a potential driver of refugee movements. The UNHCR provides insights into this.

> "The key is that climate refugees is a phrase often used in the media to describe people who are forced to move from their homes due to climate-related events, but it is not a term officially recognized in international law. The 1951 Refugee Convention offers protection only to those fleeing war, violence, conflict or persecution who have crossed an international border to find safety. Although displacement solely in the context of climate change or disasters is not covered by the 1951 Convention, it can apply when an individual's risk of facing persecution or violence is increased by climate change. For example, in northern Cameroon in 2021, hundreds of people were killed and tens of thousands fled to neighboring Chad following violence between herders and fishermen that was sparked by dwindling water resources linked to climate change. Regional refugee laws may also provide protection. Refugee definitions in the Organization of African Unity Convention and Latin America's Cartagena Declaration both include those seeking refuge due to events seriously disturbing public order, which could include climate-related events" (Siegfried, 2023).

This thesis follows the definition of the Convention of 1951, with the extensions (in some regions), and given the fact that climate displacement and violent conflicts can go hand in hand, as one can read above. The data used in this thesis (see Chapter 2) also uses this definition, as the main source of the data is UNHCR.

The term 'migrant' is also not easy to define, because it is not defined under international law (UNHCR, n.d.-a). The word migrant has been used to refer to people who move by choice rather than to escape conflict or persecution, usually across an international border, for instance to join family members already abroad, to search for a livelihood, or for a range of other purposes. The term is increasingly used as an umbrella term to refer to any person who moves away from their usual place of residence, regardless of whether the movement is forced or not.
The migrant term in literature either refers to people not being forced to flee (totally distinct group from refugees) or is used to broadly refer to people leaving their place of origin without a specific reason (refugees are part of this definition). In this thesis, only refugee movements are studied, which is in latest definitions a subpart of migrants, but not the largest groups, because most migrants move voluntarily.

1.4.2. Climate

The three most prominent terms regarding climate are: *hazards*, *exposure* and *vulnerability*. In the framework of the European Commission and European Environment Agency (n.d.), a *climate-related hazard* refers to "current climate conditions and (future changes), which determine the likelihood of an area being affected by extreme events (e.g. heatwaves) or slow-onset events (e.g. sea-level rise)." An example in this thesis context is a heatwave striking a country. Then, *exposure* is defined as the "presence of people, infrastructure or ecosystems in areas that could be adversely affected". Continuing with the heatwave example, exposure would be the people living in the country who are affected by the heatwave. In addition, *vulnerability* is defined as "the susceptibility of exposed systems and components to be adversely affected." This susceptibility is influenced by two factors: *sensitivity* ("the degree to which a system or species is affected, either adversely or positively by climate variability or change") and *adaptive capacity* ("the ability of people, sectors or systems to adapt to potential damage, seize opportunities or respond to consequences"). The example here is that refugees are more sensitive to heatwaves than other inhabitants. At last, *adaptation response* is about the means that are available to act in different situations. Policies in this context are about actions to reduce exposures to hazards of

people, and/or reduce vulnerability through enhancing adaptive capacity (European Commission and European Environment Agency, n.d.).
Given this, Ferrazzi, Kalantzis, and Zwart (2021) state in their report that the climate crisis is a global phenomenon, but the ways in which countries are exposed indeed differ greatly. In this thesis, the climate data presented in Chapter 2 measures climate-related hazards (e.g. droughts and riverine floodings) in terms of exposure to intensity of the phenomena.

This research considers that refugees usually are more vulnerable than others due to their displacement, lack of resources, low adaptive capacity, and uncertain living conditions. However, this study does not aim to quantify the exact vulnerability of refugees, if possible at all. In this study, exposure to climate-related hazards in countries (with high refugee inflows and outflows) is examined. This country-level exposure affects the refugees who flee to or from them. Therefore, if refugees flee to countries with worse climatic conditions, for example in terms of coastal floodings, their exposure to climate-related hazards increases. This does not necessarily imply a corresponding increase in climate vulnerability, as their sensitivity and adaptive capacity might increase, dependent on the destination country. The government can ensure a part of the vulnerability is reduced/mitigated, as one could read earlier in the adaptation definition. The latter is taken into account in this thesis. For example, in the Netherlands, the coastal flooding risk intensity is high. However, the Netherlands as a country has measures in place to adapt to this, with e.g. dykes and sand banks. Therefore, the focus in this report is not on 'rich', Western countries with rather strong governments as refugee destination locations: although those countries can be exposed to climate-related hazards, the countries (and therefore also the inhabitants) are relatively invulnerable.

There are many relevant remaining climate terms. For the purposes of this study, *climate change* is defined as the change in climate-related hazards, in line with the framework provided by the European Commission and European Environment Agency (n.d. As a result of climate change, exposure and vulnerability can potentially also change.

Several other terms related to climate are also used throughout the thesis. *Climate events* is a synonym for climate hazard, while *climate conditions* and *climate situations* are similar, but tend slightly more towards exposure. *Climate indicators* refer to climate events/hazards, like *climate factors* and *climate variables* do. *Climate effects* also refers to climate events in this thesis. Lastly, *climate impact* tends to refer more to climate exposure than climate hazards. *Climate risk* refers to climate exposure in this thesis. Note that *climate* can be replaced by *climatic* in all the concepts above.

To conclude, given the nature of the climate data used in this study (see also Chapter 2) and the definitions presented above, exposure is the central term around which the research questions revolve. Consequently, climate exposure is the primary focus of this thesis, while based on that, policy recommendations regarding climate adaptation measures will be provided, also incorporating the vulnerability mitigation, i.e. adaptation measures, (Western) countries have in place already.

1.5. Structure of the report

The remainder of this thesis report consists of the following chapters and subjects. At first, the methodology is presented in Chapter 2, where is explained how the research questions are answered. Then, the literature review to answer sub-question 1 is conducted in Chapter 3. What follows are the chapters to answer the sub-questions centered around the data analytic parts: answering sub-question 2 with Chapter 4. Sub-questions 3 and 4 are answered with Chapter 5. Sub-question 5 is answered with Chapter 6. At last, the discussion (Chapter 8) and the conclusion (Chapter 7) are presented.

2 Methodology

2.1. Research design

In this section, the research design is shown. The introduction in Chapter 1 and the research questions outlined in that same Chapter indicate that quantitative data analytics is the most central method for this research. That is because data on climate events at different locations needs to be gathered, processed, analyzed and compared. Also, refugee movements need to be mapped and analyzed. To compare, research of Fransen, Werntges, Hunns, Sirenko, and Comes (2023) and Owen, Kruczkiewicz, and Van Den Hoek (2023) also use quantitative data analytics to identify climate exposure among refugees at different locations.

This study is descriptive and explanatory in nature: it looks at climate events as possible driver of refugee movements and from which origin countries refugees flee to which destination countries, along with the difference in climate exposure between origin and destination locations.

This chapter will further outline the research process, it will present the mode of inquiry, the literature review approach, the data sets and the data analytics approach.

2.1.1. Research process

In this section, the different research stages are visually presented along with the research questions and the methods (see Figure 2.1). Many elements in Figure 2.1 are self-explanatory. After the approval of the earlier submitted research proposal, answering sub-question 1 could start. This sub-question is part of the qualitative part of this study. A literature study is conducted to identify climate events that are relevant to examine associations with refugee movements. This approach is preferred due to the extensive body of scientific articles available about climate events, eliminating the necessity for domain-specific knowledge among policy makers to obtain the latest updates in the form of interviews. The climate theory is already well-established in literature. Thereafter, sub-question 2 and 3 + 4 can be answered in parallel. The identified climate events serve as input for the data analysis part in which, at first, data is collected for the identified climate events at country levels. After the collection, the data is explored and cleaned, and after that analyzed. From the analysis, interpretative figures will be made, such that conclusions can be drawn and recommendations can be given. This, of course, also holds for sub-question 5. To a certain extent, sub-questions 3 and 4 serve as input for sub-question 5. The spatial clusters identified in sub-questions 3 and 4 form the center of attention for the comparison of certain regions/countries in answering sub-question 5.

2.1.2. Mode of inquiry

This section dives into the 'how' of answering the research questions. Below, for every sub-questions the details of the research approach are specified (see Table 2.1 and Table 2.2). More specifically, the following questions are answered in that order in the table rows: 1) What data is required/used? 2) How will the necessary data be collected? 3) How will the data be cleaned/processed? 4) How will the data be analyzed? 5) What tools will be used for these steps when answering the sub-question?

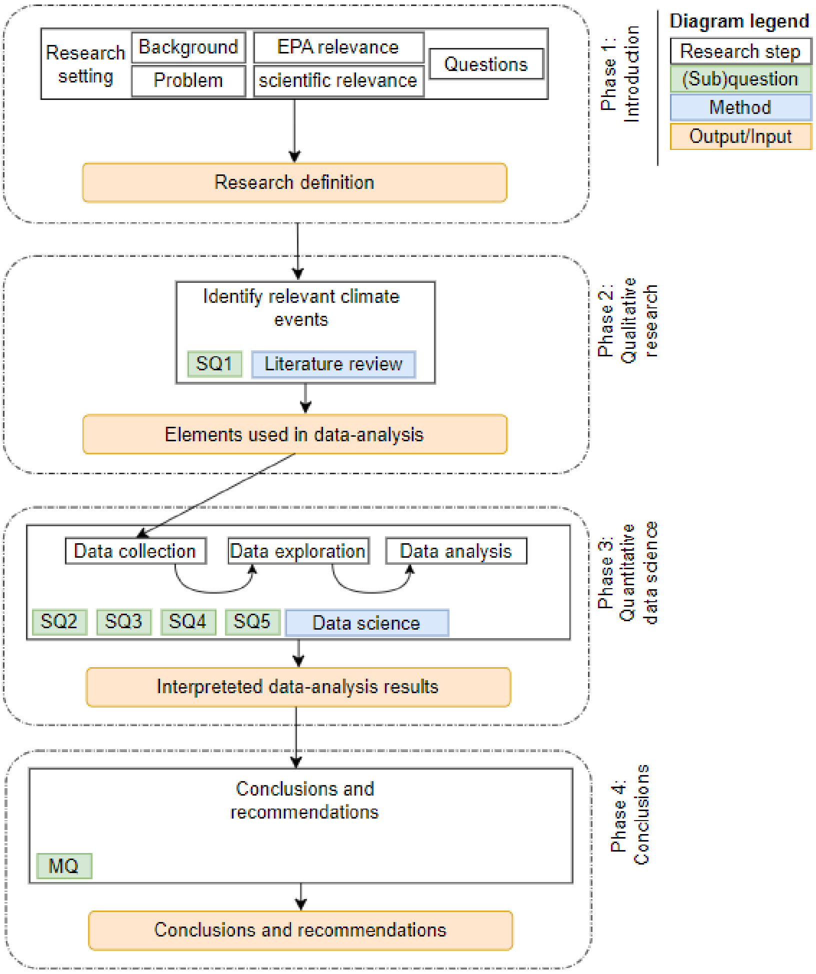


Figure 2.1: Research flow diagram

Table 2.1: Specification of answering research questions 1, 2 and 5

	SQ1: Which events reflect climate-related hazards?	*SQ2: How strongly could climate exposure drive people to flee, over years?*	*SQ5: What is the difference in climate exposure between origin and destination locations of refugees worldwide, over years?*
Data required	Climate hazards (qualitative)	Climate hazards (quantitative) & Refugee movement numbers (quantitative)	Climate change factors (quantitative) & Number of refugees that are hosted in/leaved specific countries (quantitative)

Continued on next page

	SQ1: Which events reflect climate-related hazards?	*SQ2: How strongly could climate exposure drive people to flee, over years?*	*SQ5: What is the difference in climate exposure between origin and destination locations of refugees worldwide, over years?*
Data gathering	Literature study	Refugee data (UN High Commissioner for Refugees (via World Bank), n.d.-a, n.d.-b); Geo-data + Climate hazards (Institute for International Law of Peace and Armed Conflict (IFHV), n.d.)	Refugee data (UN High Commissioner for Refugees (via World Bank), n.d.-a, n.d.-b); Geo-data + Climate hazards (Institute for International Law of Peace and Armed Conflict (IFHV), n.d.)
Data cleaning	Identify inclusion factors	See GitHub data_prep file provided by this thesis	See GitHub data_prep file provided by this thesis
Data analysis	Reading, synthesizing	Correlation matrices + Regressions results and performance (see further in Figure 2.2)	Bivariate choropleths (see further in Figure 2.2)
Tool usage	Scopus	Jupyter Lab (Python)	Jupyter Lab (Python)

Table 2.2: Specification of answering research questions 3 and 4

	SQ3: How do countries cluster spatially over time, based on climate exposure?	*SQ4:How do countries cluster spatially over time, based on refugee movements?*
Data required	Climate hazards (quantitative)	Refugee movement numbers (quantitative)
Data gathering	Climate hazards (Institute for International Law of Peace and Armed Conflict (IFHV), n.d.)	Refugee data (UN High Commissioner for Refugees (via World Bank), n.d.-a, n.d.-b)
Data cleaning	See GitHub data_prep file provided by this thesis	See GitHub data_prep file provided by this thesis
Data analysis	Local Indicators of Spatial Association (see further in Figure 2.2)	Local Indicators of Spatial Association (see further in Figure 2.2)
Tool usage	Jupyter Lab (Python)	Jupyter Lab (Python)

2.1.3. Literature review approach

This section is centered around the first sub-question: Which events reflect climate related hazards? This question will be answered through a literature study. One has to be very specific in choosing alternative terms in the search strategy, as not being concise and specific in this research area leads to tens or even hundred of thousands of hits. One cannot include a term like climate change, without begin specific. That being mentioned, the procedure is outlined below.

For this review, pre-determined steps are followed to create a systematic approach for selecting articles, and to make the literature review reproducible. The search plan is build based on the Designing policy relevant research document on Brightspace for the course Preparation Master Thesis (EPA2934)[1] and TULib (TU Delft, n.d.).

1. Analyze research topic

[1]This source is not publicly accessible.

The key concepts for this review are ‘events’ and ‘climate related hazards’. For each concept, synonyms and alternative terms are identified (see Table 2.3).

Table 2.3: Concepts and alternative search terms

Combine the concepts with AND		
Combine the synonyms with OR	**climate related hazards**	**events**
	climate related extremes	impacts
	climate related threats	outcomes
	environmental related hazards	effects
	environmental related extremes	occurrences
	environmental related threats	situations
		phenomena

2. Formulating a search query

   Based on Table 2.3, per column, all the elements are combined with OR (vertically) and then the combined column elements are again combined with AND (horizontally). This results in the following search term:

   ((”climate related threats” OR ”climate related hazards” OR ”climate related extremes” OR ”environmental related hazards” OR ”environmental related extremes” OR ”environmental related threats”) AND (impacts OR effects OR occurrences OR outcomes OR phenomena OR situations OR events))

3. Choosing a database

   Keyword-search-based platforms are used in this study, because they are most common, most developed and most extensive. Springer and Elsevier, for example, are within this category marked as publisher database. Aggregators are for example Google Scholar and Scopus. Aggregators contain also information from publishing databases, so an aggregator is used in this thesis.

   Scopus is chosen, because it is an internationally well-known; mostly peer-reviewed database, and because it is centered around public policy. Google Scholar is not used, as also (many) non peer reviewed scientific articles are in that database.

4. Screening

   This search leads to 453 results. When excluding the subject areas ‘Agricultural and Biological Sciences’, ‘Medicine’, ‘Energy’ ‘Business, Management and Accounting’, ‘Biochemistry, Genetics and Molecular Biology’, ‘Nursing’, ‘ Immunology and Microbiology’, ‘ Economics, Econometrics and Finance’, ’‘Physics and Astronomy’, ‘Materials Science’, ‘Arts and Humanities’ ‘Chemical Engineering’, ‘Chemistry’, ‘Neuroscience’, ‘Health Professions’, and ‘Pharmacology, Toxicology and Pharmaceutics’ (topics totally unrelated with the topic of this research) in Scopus, 279 results remain.

   In further selection, only articles that are in their final Publication Stage are selected. As a result, there are 276 articles remaining.

   Further, only English publications are selected. This results in the selection of 272 publications for the analysis.

   In addition, only Articles are included as document type, and not, for example, Books. This, together with selecting only Journals as source type, instead of Book series, for example. This leads to 199 articles in the final inclusion. The final search query in Scopus is as follows:

   TITLE-ABS-KEY ((”climate related threats” OR ”climate related hazards” OR ”climate related extremes” OR ”environmental related hazards” OR ”environmental related extremes” OR ”environmental related threats”) AND (impacts OR effects OR occurrences OR outcomes OR phenomena OR situations OR events)) AND (EXCLUDE (SUBJAREA , ”AGRI”) OR EXCLUDE (SUBJAREA , ”MEDI”) OR EXCLUDE (SUBJAREA , ”ENER”) OR EXCLUDE (SUBJAREA , ”BUSI”) OR EXCLUDE (SUBJAREA , ”ECON”) OR EXCLUDE (SUBJAREA , ”BIOC”) OR

EXCLUDE (SUBJAREA , ”PHYS”) OR EXCLUDE (SUBJAREA , ”MATE”) OR EXCLUDE (SUBJAREA , ”NURS”) OR EXCLUDE (SUBJAREA , ”ARTS”) OR EXCLUDE (SUBJAREA , ”PHAR”) OR EXCLUDE (SUBJAREA , ”CHEM”) OR EXCLUDE (SUBJAREA , ”CENG”) OR EXCLUDE (SUBJAREA , ”HEAL”) OR EXCLUDE (SUBJAREA , ”IMMU”) OR EXCLUDE (SUBJAREA , ”NEUR”)) AND (LIMIT-TO (PUBSTAGE , ”final”)) AND (LIMIT-TO (LANGUAGE , ”English”)) AND (LIMIT-TO (DOCTYPE , ”ar”)) AND (LIMIT-TO (SRCTYPE , ”j”)).

When analyzing those 199 results, 110 do not even mention a specific instance of a climate hazard in the abstract, title or keywords. In the end, 89 are relevant, based on title, keywords and abstract. Those articles identify, select, include or measure specific climatic indicators or events in their studies. Inclusion of one or more specific climate events is key. But also from the remaining 89 results, in which are specific climatic events mentioned, the climatic events are not always central nor threatening in the analysis. Namely, the focus is often on adaptation and some climatic events are taken as ‘given’ or as examples, and not researched/identified in the study themselves. Of the 89 relevant publications, 57 articles are also excluded after reading the abstract with the main reason being a ‘wrong study design’. Some publication turned out to be off-scope, researching a specific topic, for example publications that focus on crop yields and the relation with climatic events. No ‘such as’ examples of climatic events are valid. The selection of these 35 events is purely on researching a climate event, and not a risk, adaptation, planning or resilience strategy. Those are very important, but this literate search is purely for the identification of relevant threatening climatic events for the research. Many articles do not have the right study design, researching consequences of climate (changing) events; adaptation, vulnerability and/or resilience techniques as a result of climate change. In these articles, the climate (changing) hazards, the events, are not specified and are not central. It is about consequences of climate events in general. Those topics are out of scope for this research.

5. Full-text eligibility

   From these 32 specific articles, 4 have an irrelevant context. This means it is not about identifying climate hazards. One article dives into a comparison of different models for ‘estimating’ rainfall. The other on is about the relation between two pre-specified climatic indicators in Italy. Another study focuses on making flooding maps and early warning systems. The last study focuses on making a model/assessment on flood risk in general for informal settlements. These articles are not about identification. Also, 3 publications are not fully accessible (only the title, abstract and key-words are accessible).

6. Inclusion

   The 25 remaining publication describe threatening climatic events. Those are synthesized in Chapter 3.

2.1.4. Data sets

In Table 2.1 and 2.2, the data sources are already presented. In this section is explained why these data sets are chosen for this research. In the Discussion (see Chapter 8) is reflected on the limitations of the use of these data sets.

The World Risk Index includes global country level data on various climatic and social indicators. Examples of climatic events and societal variables that are included are: cyclones, droughts, but also average population killed in conflicts and control of corruption scores. The World Risk Index is not the first tool to assess risks and exposures on a global scale. Two prominent other approaches are the Disaster Risk Index and the Natural Disaster Hotspot. However, the World Risk Index focuses not only on mortality and economic effects of climate-related hazards, but also exposure of societies (Welle & Birkmann, 2015). This is an important reason to choose for the World Risk Index, as country level data is needed for the analyses in this study, but also social factors (in broad sense) are needed to include in the analyses to study interactions between social and climate events, and also to control for social events, to correctly estimate effects of climate exposure (see Chapter 4). This is necessary as many articles outline the importance of especially conflict and violence as main driver behind refugee movements (Hatton, 2009; Moore & Shellman, 2004; Schmeidl, 1997). In the World Risk Index, violence and conflict are captured explicitly in the variable lack of coping capacities. Economic and intervening

variables have far less impact, but still play a role (Moore & Shellman, 2004; Schmeidl, 1997). In the World Risk Index, lack of socio-economic development, societal disparities, socio-economic deprivation and lack of adaptive capacities capture these aspects.
Moreover, the theoretical fundament in the World Risk Index is considered reliable and is for example in line with definitions provided by IPCC (Welle & Birkmann, 2015).
In addition, the World Risk Index is used as framework for other indexes, including the Disaster Risk Indicators in Brazil (DRIB) index, because of the modular structure of the World Risk Index (de Almeida, Welle, & Birkmann, 2016). This shows the popularity of the index. However, the World Risk Index also induces limitations on which will be reflected in the Discussion (see Chapter 8).

The source of the refugee data is UNHCR, and the data is processed by the World Bank (see Table 2.1). Refugee data from UNHCR has been widely used in many applications (Marbach, 2018). There are a lot of examples of scientific studies using UNHCR data for refugee locations analyses (Fransen et al., 2023; Owen et al., 2023). Also, Fransen and De Haas (2022) mention that the UNHCR Statistics Database in general is the most comprehensive one available, although it has limitations. This boosts confidence in using UNHCR data for refugee movement analyses. Also, in combination with the above, the format of the data (yearly at country level) limits the data set choice to UNHCR data. The UNHCR data consists of refugee stocks, meaning the number of refugees at a certain point in time fleeing from/to a country. No flows are documented, which will be reflected upon in the Discussion (see Chapter 8).

2.1.5. Data analytics approach

A data analytics approach is required to answer sub-questions 2, 3, 4 and 5. This section specifies the types of analyses that will be used to answer these sub-questions.

First, before the start of all the individual analyses, the data sets listed in Table 2.1 and Table 2.2 (Data gathering) are merged in Python based on country codes. That is possible, because both the refugee and climate data are available for each country. For example, the data includes the number of refugees hosted in each country and a relative score for riverine flooding intensity. As can be derived from Table 2.1 and Table 2.2, geospatial data is also available alongside the climate data from Institute for International Law of Peace and Armed Conflict (IFHV) (n.d.). This means that country data can be plotted.

After the data is merged and cleaned, the analyses start. The analyses differ between the sub-questions. That is, in short, visually presented in Figure 2.2 and thereafter explained in detail in text below.

- **SQ 2.** To establish whether climate conditions could be a reason for people fleeing, first a correlation analysis will be conducted in the period 2003-2022. Researching correlation over time is in line with research of Casali, Aydin, and Comes (2024). Also, Owen et al. (2023) use a correlation analysis to track the relationship between climatic exposure and the establishment year of refugee camps, allowing for temporal analysis of different exposure levels.
  Correlation between two variables is (usually) required to determine whether there could be a (causal) relation. In order to examine different types of correlation possibilities, Kendall and Distance correlation will be investigated (see Chapter 4 for the reasoning behind this). There are several steps to take in this correlation analysis. 1) Correlation matrices will be made for every climate indicator with the number of refugees fleeing from/coming to a country, over all the years. In other words, different climatic indicators will be correlated with number of refugees fleeing from a country and also number of refugees coming to a country, over multiple years. For example, sea level rise will be correlated with number of refugees fleeing from a country, over multiple years. Then, per year, one can look at the size of the correlation coefficient in order to make a first estimate of the magnitude of the relations per year. 2) To statistically determine whether the correlation coefficients change over year with a certain pattern, the Mann-Kendall Test will be used. This test verifies whether there is a trend in time series data (Alashan, 2020). If it is verified that there is a statistical trend with the Mann-Kendall Test, the Theil-Sen Slope will be calculated. This test gives an estimation of the slope of the trend. For example, the 20 correlation coefficients (from 2003 up to and including 2022) between droughts score in countries and the number of refugees that are fleeing from countries will be gathered. This will be done for every climate indicator. Then, with the p-values of the Man-Kendall Test and Theil-Sen Slope, one can determine for every relation if there is a statistically significant trend, and, if there is, the

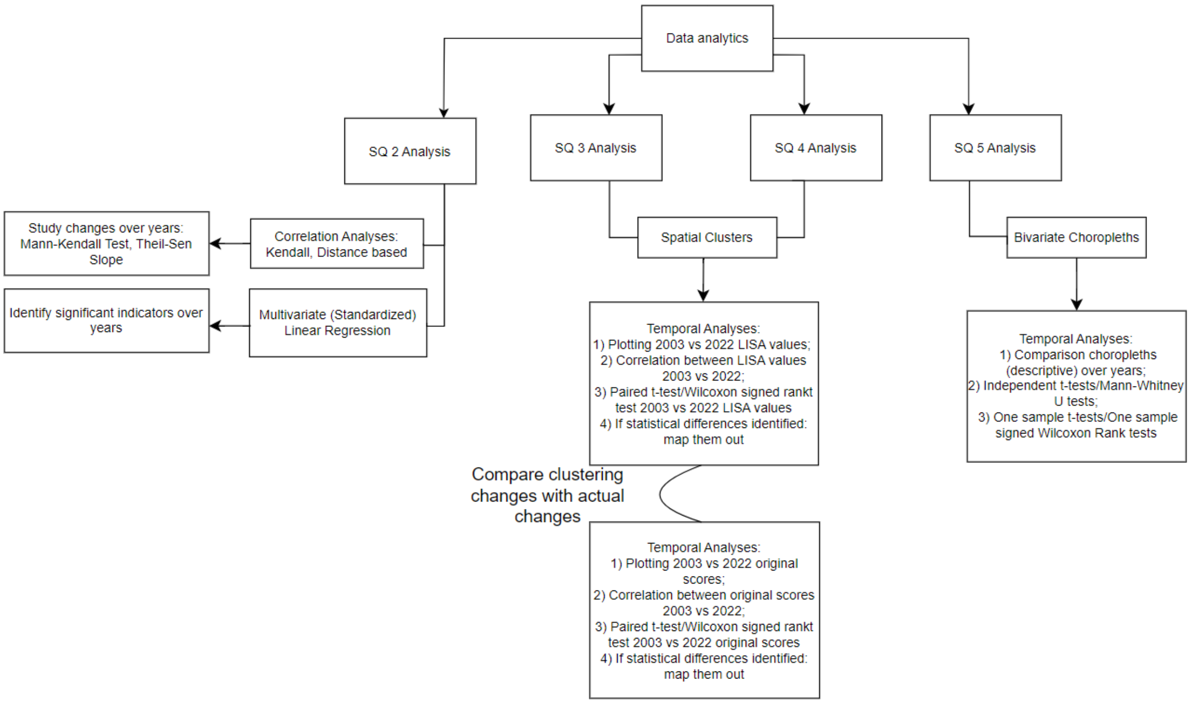


Figure 2.2: Data analytics approach diagram

magnitude of the trend for the correlations over years.

Because regression can tell more about a relationship than correlation and can be used for prediction (Shi & Conrad, 2009), also a multiple linear regression model will be fitted for every year, focusing on estimating the coefficients of the climate indicators. In data from the World Risk Index (Institute for International Law of Peace and Armed Conflict (IFHV), n.d.) also conflict and economic data is made available, besides the climate data. These factors, such as control of corruption, political stability and absence of violence and terror, as well as gross national income per capita, will be included in the regression formula, so there can be 'controlled' for these factors when necessary. With this principle, the effect of a climatic situation on refugees fleeing can be estimated more accurate. Again, this analysis will be conducted for every year. There will be one regression formula, with the number of refugees fleeing being the dependent variable and the climate indicators and the economic and conflict variables being the independent variables. The gathered coefficients for the climate indicators over the years, will be compared based on their significance over the years.

With all this, a valid estimate can be made of the effect of (changing) climatic conditions on refugees fleeing to or from a country.

- **SQ 3 & 4.** To identify worldwide (dis)similar refugee and climate situations among countries, spatial clusters will be identified, following the line of Casali et al. (2024). This means for SQ 3 that areas with similar climatic exposure and dissimilar climatic exposure will be identified, and this will be also be done, for SQ 4, with refugee movement data (e.g. areas with high number of refugees fleeing from it). This is based on statistical analysis. Local Indicators of Spatial Association (LISAs) are designed for this type of analyses. The core idea is to identify cases in which the comparison between the value of a country and the average of its neighbors is either more similar (High-High, Low-Low) or dissimilar (High-Low, Low-High) than one would expect from pure chance.

  The alternative of LISA is the Hotspot analysis, which is also often used in spatial studies, such as in work of Aydin, Yigitbasi, Casali, and van Wee (2023) and Casali et al. (2024). In this Hotspot analysis, the main statistic used is the Getis-Ord Gi*, which only identifies hotspots and coldspots with high or low value concentrations. Hotspots (coldspots) should then have high (low) values and be surrounded by other high (low) value countries. That would mean for this study that only countries which are surrounded by high value countries (low) and have high (low) values themselves, are identified as significant. On purpose is therefore chosen for LISA with four categories, as in the context of refugees fleeing to/from countries, it is extremely important to know what countries have high (low) values and are surrounded by countries with low (high) values, because neighboring countries could be safer (unsafer) in terms of climatic exposure. Refugees often flee to neighboring countries, in which the climatic situation can be different. That is what this study wants to identify. When countries are not identified as significant in a cluster in which high (low) values are surrounded with low (high) values, a lot of information is missed.

No global spatial analysis will be used, because it does not specifically identify the locations of clusters within the spatial data, it only gives an indication for the overall clustering worldwide in one single score.

For every year -in separate analyses for every climate indicator, refugee origin data and refugee receiving data- each country gets a LISA score. The LISA scores are compared visually by colored maps and conclusions are drawn based on these insights, following the research of Casali et al. (2024).

Temporal relationships are examined by comparing the LISA scores over the different years for every indicator, based on plots, correlations and statistical tests (see Figure 2.2 and Chapter 5). This way, one can verify differences in spatial clusters over years, for every indicator (for climate as well as refugee data).

- **SQ 5.** To show the difference in climate exposure at origin and destination locations of refugees, at first bivariate choropleths will be made for the period 2003-2022. This is a descriptive visualization, in which climatic exposure and refugee movement data at country level will be presented. This will be done for every individual climate indicator and refugee movement situation (fleeing from or to a country) over years. (See Chapter 6 for more explanation.) A color map shows the climate exposure for a certain indicator and on this maps, different dot sizes per country give an indication of the number of refugees fleeing to or from a country per year. From this, the top 25 countries with the highest number of refugees fleeing from and the top 25 countries receiving refugees will be identified per year. The climate scores of these countries will then be compared and reflected upon (descriptive analysis) over years. Dependent on the characteristics of the specific data, a statistical test is chosen to make this comparison over time (see Chapter 6).

3
Results SQ1: climatic events

This chapter is centered around answering the first sub-question: Which events reflect climate related hazards? This question will be answered through synthesizing results of a literature study. The procedure followed in advance of the literature findings is outlined in Chapter 2.

3.1. Findings of literature study

In this section, the findings from the literature study are presented per world continent. That way, it can easily be assessed which climate events are researched in more than one continent in the world.

In general, the historical development in this field is low, as nearly all publications are written between 2015 and 2024 (which is noticeable in the paragraphs below). Before 2015, the link between climate events/hazards were not researched extensively. This means that this is an upcoming research area.

3.1.1. America

In a United States case study, Prein, Holland, Rasmussen, Clark, and Tye (2016) note that changes in precipitation have profound impacts on society and ecological systems, as highlighted by recent severe droughts in California and Oklahoma. Droughts, alongside tropical cyclones, are among the costliest climate-related extreme events in the U.S. There weather type (WT) analysis of reanalysis data from 1979–2014 shows that precipitation trends from 1980–2010 are influenced by shifts in WT frequencies and precipitation intensities. In the North Atlantic and Midwest, increased precipitation intensity drives rising precipitation trends. Conversely, in the U.S. Southwest, changes in WT frequencies result in significant precipitation decreases of up to 25% due to increased anticyclonic conditions in the North Pacific, though this trend is partially mitigated by rising precipitation intensities.

In the Upper Midwestern United States, the study of Khan, Bhattarai, and Chen (2024) identifies other climatic events and indicators than precipitation. The study examined trends in extreme heat and cold events from 1979 to 2021, revealing significant regional variations and their impacts. Extreme heat stress increased in parts of Michigan, Wisconsin, Ohio, and Indiana, while it decreased in western regions like Minnesota and Iowa. Extreme cold events generally declined across the western regions, though some areas, including Iowa and northern Minnesota, Michigan, and Wisconsin, experienced increases.

A specific New York case study, investigates the variability of climate-related extreme events affecting the city over the past 140 years (Depietri & McPhearson, 2018). Although New York was once considered less prone to hazards, this view has shifted, especially following the severe impacts of Hurricane Sandy in 2012. The authors developed a comprehensive database of significant climatic events and evaluated their multi-sector impacts. The findings reveal that New York has been systematically affected by hazards, with heat waves being the deadliest and hurricanes the most costly. They specifically analyzed trends in heat waves and flooding, as data for these events span the entire study period. Using data from The New York Times, the analysis shows that both flooding and heat waves have frequently impacted the city, with a trend toward an increasing average number of such events per decade. Another, related study in New York, finds flooding as the major risk for the city (Depietri, Dahal, & McPhearson, 2018). Their analysis indicates that coastal areas, particularly midtown and downtown Manhattan, Harlem, and parts of Brooklyn, face the highest risk due to coastal flooding.

As opposed to the United States studies above, the study of Pauli et al. (2021) research climate related hazards in Cambodia, South America. In the Mekong Basin, communities have historically managed variations in temperature, rainfall, and flooding. However, rapid environmental changes now bring new challenges, including increased drought, altered rainfall patterns, more frequent floods, and rising water demands for hydropower and irrigation. This study uses a multi-method approach to combine local knowledge of floods, droughts, and rainfall with scientific data, including satellite imagery and historical rainfall trends. By mapping local indicators and comparing them with 35 years of rainfall data, the research identifies key spatial pressure points. The findings of a study in Cambodia are similar compared to the results of a study in Bolivia (Seiler, Hutjes, & Kabat, 2013). Climate variability in Bolivia is likely to exacerbate extreme weather events such as droughts and floods. Droughts typically affect the southern lowlands and Altiplano from June to August, while floods occur mainly in the northern lowlands' savannas and the catchment areas of Lake Titicaca and Poopó from January to March.

3.1.2. Asia

The study of H. Chen and Sun (2020) examines how population exposure to extreme precipitation events across China may change using high-resolution simulations under future mid- and high-emission scenarios. As the climate warms rapidly in the future, extreme precipitation events, are expected to increase significantly across China. The Tibetan Plateau is projected to experience the largest rise of about 13.5%, while Southwest China will see the smallest increase of around 4.9% by the end of the century compared to current conditions. Overall, the study highlights precipitation as the most severe climate threat facing the country. In China, another study confirms the findings of rainfall being an important climatic factor (Tang, Gan, Zhao, & Gao, 2006). This study analyzes persistent heavy rainfall (PHR) events in China from 1951 to 2004 using daily rainfall data. PHR events are defined by their intensity, duration, extent, and persistence. The study identified 197 such events and classified them by intensity, circulation patterns, and geographic location. It then examined the most severe, severe, and moderate events to understand their frequency, intensity, and distribution.

More extreme rainfall patterns are also identified by research in northern Thailand (Manandhar, Pratoomchai, Ono, Kazama, & Komori, 2015). They add that nearly 45% of households perceive climate change, with 47% informed through various sources. Rainfall changes, particularly in timing. While most households correctly perceive changes in rainfall patterns, such as increased rainfall, fewer rainy days, and late rainfall, 82% do not notice temperature changes. Droughts, floods, and landslides are major concerns, with many misunderstanding the causes of droughts. Flooding is linked to increased rainfall, but landslides also contribute. Increasing rain leads to land slides and floods.

Heavy rainfall is also identified in Bangladesh, along with increase in temperature, resulting in more hot days and nights (Shahid et al., 2016). Bangladesh's urban population faces significant vulnerability to climate change due to high density, inadequate infrastructure, and low adaptive capacity. Analyzing 55 years of climate data (1958–2012) from five major cities reveals notable increases in both daily maximum and minimum temperatures. There is also a rise in extreme weather events, including heavy rainfall (>20 mm), hot days (>32 °C), and hot nights (>25 °C). These trends are projected to continue into the twenty-first century. Similar things hold for a study in Sri Lanka, in which extreme rainfall events are examined in the dry zone using rainfall data from 1971 to 2017 across 19 meteorological stations (Abeysingha, Kularathna, Bandara, & Ray, 2023). This trend reflects broader climate change impacts, with notable shifts in extreme rainfall occurrences observed month-to-month and across different stations. The results underscore the influence of climate change on increased annual and seasonal rainfall variability.

The study of Paliwal and Patwardhan (2013) puts emphasis on cyclones, by stating that tropical cyclones pose a major climate risk in South Asia. To assess this risk effectively, it is crucial to understand historical cyclone patterns. This study examines cyclone tracks in the North Indian Ocean and identifies five distinct clusters based on their spatial characteristics. Each cluster represents a specific region with unique patterns in cyclone formation, likelihood of landfall, duration, and intensity. Some clusters are particularly significant because they have higher chances of making landfall and produce more intense cyclones.

3.1.3. Africa

The study of Orimoloye et al. (2019) explores a spatially synergistic methodology to evaluate drought occurrences in the Cape Town area of South Africa from 2014 to 2018, leveraging remotely sensed data. This research underscores the crucial role of remote sensing (RS) and geographic information systems (GIS) in assessing drought severity. Advancements in modern RS technologies facilitated comparative analyses of this natural hazard and other potential environmental disasters. The study incorporated five land use features—vegetation, built-up areas, water bodies, bare surfaces, and sparse vegetation—derived from a spectral mixture classification as indicators for comparison with drought indices. Central is their focus on drought severity. Drought as prominent climate factor is also identified by Hagenlocher, Lang, Hölbling, Tiede, and Kienberger (2013). They focused on identifying hotspots in North-West Africa where extreme climatic conditions, such as flooding and droughts, are most intense. They developed a spatial composite indicator to highlight areas experiencing significant changes in temperature and precipitation. Key indicators in this analysis included seasonal temperature and precipitation trends, drought occurrences, and major flood events.

In addition, in South Africa, the Table Mountain National Park has key climate stressors drought, warming temperatures and declining rainfall (Chikodzi, Nhamo, Dube, & Chapungu, 2022). But in addition to these droughts, in other national parks, other stressors are identified. The other sites assessed were Mapungubwe World Heritage Site and Thulamela Ruins in Kruger National Park. Climate exposure for Mapungubwe was shown to be high, with high temperatures, flooding and intense rainfall being the main climatic stressors affecting the site. The same climatic stressors were observed at Thulamela (Chikodzi et al., 2022).

In Southern Ethiopia, a study evaluates trends in temperature and precipitation extremes in Gurage Zone, revealing significant findings (Dendir & Birhanu, 2022). Temperature extremes are increasing, with more warm days and nights and fewer cold days and nights. Positive trends in temperature indices and longer warm spells were observed, while cold spell duration trends were inconsistent. Precipitation extremes show spatial inconsistency, with drier conditions in both lowland and highland areas and variable intensity of extreme precipitation. Also, flood and drought risks are identified.

At last, a Nigerian study purely focuses on intense heat events (Taiwo, Olaniran, & Osayomi, 2012). Local perceptions of climate hazards, specifically seasonal heat events, in Ibadan, Nigeria are examined. Residents attribute intense heat to factors such as strong sunlight, climate change, and prolonged periods without rain, with notable variations across residential density areas. Key elements influencing heat exposure include electricity availability, proximity to water sources, and the presence of neighborhood trees.

3.1.4. Europe

In European coastal cities, research of Laino and Iglesias (2024a) identifies the most important and significant climate related hazards to be storm, coastal and land flooding, and coastal erosion, potentially caused by sea level rise as the overall important factor. Similar events are confirmed by Hawchar, Naughton, Nolan, Stewart, and Ryan (2020) in Ireland. The framework in that study identifies fluvial flooding, coastal flooding, erosion, sea level rise, and extreme wind speeds as significant climate change impacts on critical infrastructure. Another, but related, study in Europe (Laino & Iglesias, 2024b), finds also that coastal cities face a range of climate-related hazards with significant variability. Sea-level rise rates differ widely, from 2.4 mm/year in Sligo to 5.4 mm/year in Vilanova i la Geltrú, necessitating city-specific adaptation strategies. Extreme sea levels are largely influenced by wave action, with Sligo, Oarsoaldea, and Oeiras experiencing the highest levels. Coastal flooding extents are often smaller than low-elevation coastal zones (LECZ), with Gdansk showing the highest LECZ proportion (46.8%). Coastal erosion is severe in cities like Massa, while artificial seafronts in other cities reduce erosion susceptibility. River flooding risks are notable in Gdansk, Samsun, and Dublin. Heavy rainfall varies, with Massa and Piran seeing the most frequent events, whereas drought conditions are most severe in Oeiras. Temperature extremes vary, with Oeiras recording the highest maximum temperatures and Gdansk the lowest minimums. Strong wind events are most frequent in Sligo and Dublin, while landslide susceptibility is highest in Massa. These findings emphasize the diverse climate risks.

A study in Germany analyzed drought frequency, persistence, and severity in the Ruhr River basin from 1961 to 2007 using the Standardized Precipitation Index (SPI) (Khadr, Morgenschweis, & Schlenkhoff, 2009). Calculated over various timescales (1 to 24 months), the SPI revealed multiple drought events, ranging from mild to severe, with both positive and negative trends. Droughts affected the Ruhr basin in a random pattern, occurring in both summer and winter, with the most severe event in winter. Despite an overall increase in winter precipitation, the study found no significant long-term change in drought conditions.

In a generic European study, that is, not for/in a specific country, Forzieri et al. (2016) introduce a multi-hazard framework for assessing exposure to various climate extremes throughout the 21st century. By analyzing an ensemble of climate projections, the framework evaluates changes in the frequency of heatwaves, cold waves, river and coastal flooding, streamflow droughts, wildfires, and windstorms. The results reveal significant variations in hazard exposure, influenced by rising temperatures and local climate conditions. Europe is expected to experience a gradual increase in overall climate hazards, with the most severe impacts in south-western regions, driven primarily by heatwaves, droughts, and wildfires. Coastal areas and floodplains, which are often densely populated and economically significant, are identified as critical hotspots where floods and windstorms could compound other climate hazards.

A study in Serbia evaluates the impact of intense heat waves in Novi Sad during the exceptionally hot summer of 2015 (Savić et al., 2018). By analyzing nocturnal Urban Heat Island intensity, it finds that densely built areas face the highest heat-related risks. The results highlight the uneven distribution of heat wave impacts across the city.

3.1.5. Global

The study of Lange et al. (2020) measures how climate change affects exposure to six extreme weather events: river floods, tropical cyclones, crop failures, wildfires, droughts, and heatwaves. A 1°C increase in global temperatures has already raised the global land area and population exposed to these events by about 140% and 130%, respectively, with droughts and heatwaves being the largest contributors. Especially heatwaves are also confirmed as important climate related hazard by Jones, Tebaldi, O'Neill, Oleson, and Gao (2018). The study projects future global, regional, and subnational exposure to heat extremes, finding that heat waves will increase significantly under all scenarios. Again is found that urban areas experience more heat waves due to the urban heat island effect, suggesting that models should account for urban-rural differences. The study highlights the need for multi-model approaches and further exploration of vulnerability factors to better understand future heat impacts.

3.2. Selection of climatic events

The selection of which variables will be used in the analyses in this thesis is, of course, dependent on the results of the literature study above. In addition, the selection is also dependent on data (format) availability. The most important instance of a climatic event that is very relevant, but not available in the correct format and the correct time frame for this research, is extreme precipitation (or heavy rainfall). Precipitation is an important climatic factor. However, data is only available per year till 2020, and in another format it is not available per country. For the analyses in this thesis, per year country availability and latest data (up to and including 2022) is required. Therefore, precipitation data is not included.

Extreme heat (or intense heat events, and to a lesser extent cold waves) is also researched very frequently. However, unfortunately, the same holds here as for the precipitation data: the data format (per country data) combined with the required time frame (up to and including 2022) does not fit. In addition, these factors are difficult to assess, i.e. there is ambiguity about when something is an extreme heat event, and that definition differs around the world. That makes it even more difficult to examine this indicator in a global and quantitative academic study.

Other (important) climatic events that are frequently researched in the literature are included in this study. Flooding, coastal as well as riverine flooding, is researched often in different continents in the world. Also, droughts, cyclones and sea level rise are examined very often in different parts of the world and identified as important climate related hazards.

Some events that one may expect, like tsunamis and earthquakes, are not selected, because not one source identified these kind of events in the search above. This may be due to the fact that these events are not (directly) caused by climate change.

To conclude, the following indicators are selected for the remainder of this thesis: Droughts, Cyclones, Coastal Flooding, Riverine Flooding and Sea Level Rise. These are important climate related hazards identified in the literature study. Also, this data is available in the required format. With this selection, sub-question 1 is answered.

4
Results SQ2: Climate as driver for fleeing

This chapter is centered around the second sub-question, exploring the drivers behind refugees fleeing. To establish causation (in this refugee movement context) is very difficult. Nine requirements that should be met are identified by Hill as a road map to reflect on causality (Rothman & Greenland, 2005). The following aspects of an association should be considered in attempting to distinguish causal from non-causal associations: strength, consistency, specificity, temporality, biologic gradient, plausibility, coherence, experimental evidence, and analogy. These criteria are 'the best' to use for causal inference (Swaen & van Amelsvoort, 2009).

Not on all those nine requirements will be extensively reflected in this thesis. However, on strength (= strong associations are more likely to be causal than weak associations) is reflected in the correlation and regression analysis. Also, Biologic Gradient (=the presence of a monotone (unidirectional) relation) is taken into account with using the Kendall correlation coefficients, which is about monotonicity; regression and plots are also considered. In addition, plausibility (=biologic plausibility of the hypothesis) is met, because social and climatic factors could cause refugees to flee. It is not unrealistic to assume that these factors can cause refugees to flee. Moreover, as an example, temporal precedence or temporality, is difficult to determine: is there first a climatic event/social conflict and do refugees then flee, or did refugees already flee before the event took place? That cannot be derived from the data used in this study and requires another analysis. This will not be addressed in this thesis.

Although regression, which will be conducted in this study, can suggest causality if the model is well-specified and assumptions are met (Schomaker, 2020), it does not prove causation on its own, because it is often nearly impossible to know if a model is truly well-specified, and then still some other assumptions need to be met. However, estimated coefficients represent more than simple correlations. The regression coefficient represents the partial/conditional relationship between the dependent variable and a specific independent variable, controlling for the effects of other variables in the model. Unlike a correlation, which just measures the strength of a relationship, the regression coefficient accounts for the influence of other variables in the model. This means it can isolate the unique contribution of each predictor to the dependent variable, assuming a 'correct' (i.e. well-specified) model. The sign and magnitude of the coefficient indicate the direction (positive or negative) and strength of the association between the predictor and the outcome variable, after accounting for the other predictors. In terms of Hills criteria, it helps in determining true Strength and Biologic Gradient, when one can control for third variables (confounding factors). Ruling out confounding is extremely important when reflecting on causality, and is indeed also related to Hill (Cox Jr, 2018).

In the remainder of this chapter, at first, the (insights regarding the) findings of the correlation analysis are presented. Thereafter, different statistical tests are used to indicate whether the correlations differ over time. Then, the multivariate regression analysis is presented.

4.1. Significance

In the correlation matrices presented in the section below, and the trend and regression analysis later in this chapter, only statistically significant correlation coefficients are taken into account. An attentive reader may think this is remarkable, because this study works with population values and not sample data, and significance only plays a role in inference from sample to population (hypothesis testing). However, there are four reasons why only statistical significant coefficients are taken into account. 1) At first, the population in this study are all countries in the world over multiple years, and also in the future. There is no data available for all the countries in the world on climate, social and refugee data, and only approximately 85% of all the (independent) countries are included in this study. So, one may argue that this study just has a very large sample. 2) Also, forecasts into the future are made, which represent a new, unknown population (unknown values). 3) In the data analysis for this study, also all the correlation coefficients were compared with only the statistical significant ones in Python[1]. It turns out that all of the insignificant statistical coefficients, also are practical insignificant (often (far) below or around 0.1 correlation coefficient size). From a practical point of view, this is the most important reason: the statistical significance aligns with the practical significance. 4) The last argument is a combination of the small population size and the years (particular dates) the data is analyzed in. In this study, the population is small, because in the world there exist around 200 countries. The data in this study is analyzed for a specific period from 2003-2022 and some forecasts are made about the future. So, in this study is aimed to infer conclusions on similar small groups over time and into the future, that one can treat every relatively small group per period (year) as sample to compare.

4.2. Correlation analysis

The correlation coefficients selection with extensive explanation is presented in Appendix B. The selected climatic events in Chapter 3 are included in the analyses as well as many of the available generic social factors included in the World Risk Index: societal disparities, lack of socio-economic development, lack of coping capacities, socio-economic deprivation and lack of adaptive capacities. In Chapter 2 is reflected on the reasons for this.

4.2.1. Results refugees origin

In this section is reflected on the Kendall correlation coefficients as well as the Distance correlation coefficients for the refugee origin data, which means that the focus in this section is on correlating different climatic/social indicators with refugees fleeing from a country.

4.2.1.1. Kendall correlation

In the correlation matrices below, the Kendall Tau correlation coefficients are shown for all the years between all the indicators. Only the significant coefficients are shown, that is why some squares are white/empty. Also, the upper part of the square is masked due to the symmetry of the correlation matrix.

In this study, the most important column in the matrices is the first column. The first column describes the correlation between refugees fleeing from a country and all the indicators, climate indicators as well as social indicators.

In 2003 (see Figure 4.1), one can see moderate correlation between riverine floodings and refugees fleeing (0.33). The same holds for the lack of socio-economic development (0.3) and lack of adaptive capacities (0.3). Lack of coping capacities is the only correlation coefficient that is -to some extent- high, with a coefficient of 0.45. Other correlations between indicators and refugees origin are quite low. Only coastal flooding correlation coefficient with refugees fleeing is not significant, which means that there is no correlation between those variables, thus there is no association and there cannot be causation in 2003, according to this analysis. Further, one can see that socio-economic deprivation could cause multicollinearity later because of the high correlation with other social factors, and that cyclones as well as riverine floodings do nearly not have any association with other indicators in this study. Also, what is important to notice is that the variables cyclones and sea level rise have a (counter intuitive) negative sign. This means that more cyclones and more sea level rise is associated with fewer people fleeing, which of course does not represent a rational mechanism that connects the variables like this in reality. From this can be concluded that although these factors measure association, there cannot be a causal

[1] If one wants to verify this: see the by this thesis provided GitHub Correlation analysis file cells 6-26.

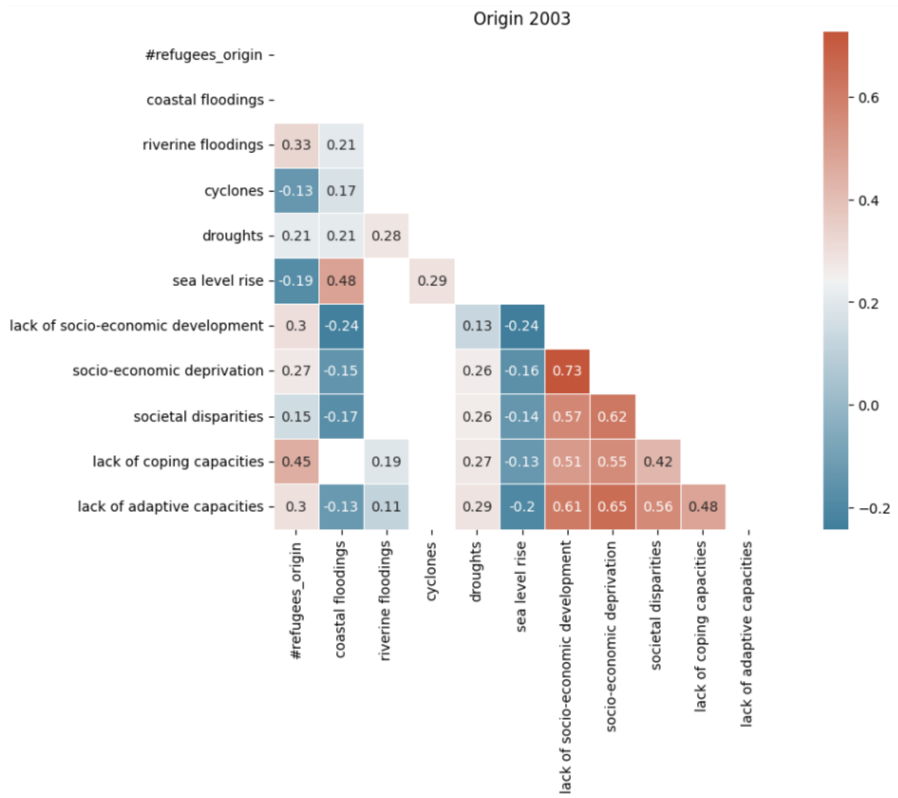


Figure 4.1: Refugee origin location 2003 Kendall correlation coefficients

connection between the factors, according to the Kendall correlation coefficients. The other indicators have an intuitive positive sign, which indicates that, for example, more riverine flooding or more lack of coping capacities, is associated with more refugees fleeing from that country. Note that signs could change when conducting regression and controlling for other variables. The same overall structure holds for 2004 and 2005.

From 2006 onwards (see Figure 4.2), one can see that also the Kendall correlation coefficients between cyclones and refugees fleeing are not (statistically) significant, and therefore there is then no association between those indicators and also no causal relation. Approximately the same structure remains present up to and including 2022. In 2021 (see Figure 4.3), one can see that the correlation between droughts and refugees fleeing has gone up from 0.21 in 2006 to 0.27. Also correlations between refugees fleeing and lack of coping capacities/societal disparities are higher. So, the structure is the same, but the correlations between all social variables and refugees fleeing have slightly increased (approximately with 0.04).

4.2.1.2. Distance based correlation

The distance based correlation (see Appendix B for explanation) fully aligns with the Kendall correlation regarding cyclones and coastal flooding not having a significant correlation association with refugees fleeing, which means that these variables are independent from each other. Other coefficients seem to align also, with an exception for the higher correlations estimated between the social factors (0.85 between deprivation and lack of development), indicating multicollinearity again. The biggest difference is in the correlation with sea level rise and refugees fleeing from a country, which in this Distance correlation analysis does have the expected positive sign, in contrast with the negative sign found with the Kendall correlation coefficient. This finding is explainable, because the range of the Distance based correlation is between [0,1] and not [-1,1]. This type of correlation is purely meant to conclude on (in)dependence between variables. One cannot conclude on the sign of the association. Up to 2007, the same conclusions can be drawn as based on Kendall correlation.

Then, from 2007 onwards (see Figure 4.4), what differs with the Kendall correlation coefficients, is that also sea level rise is not significant anymore in the correlation with refugees fleeing from a country.

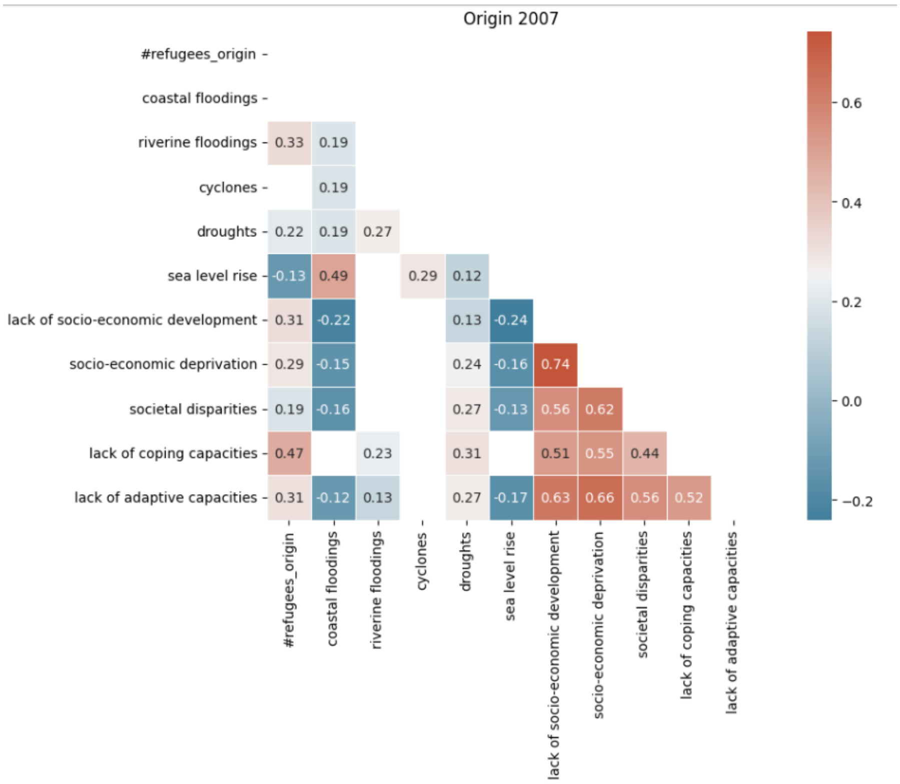


Figure 4.2: Refugee origin location 2007 Kendall correlation coefficients

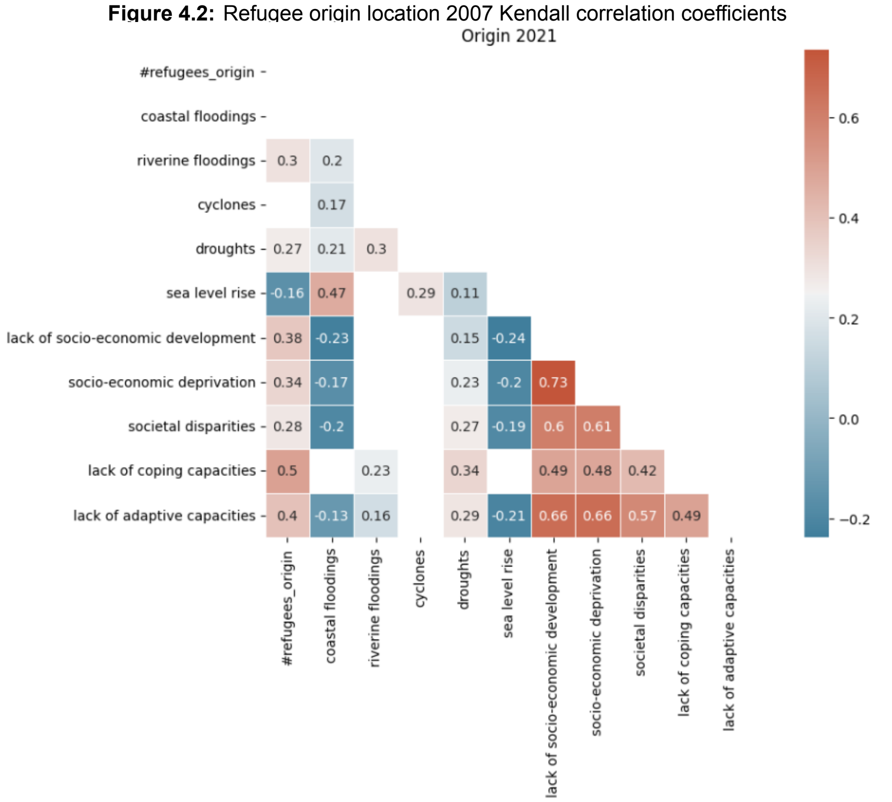


Figure 4.3: Refugee origin location 2021 Kendall correlation coefficients

This means that these variables are independent, according to this analysis. Then from 2014 onwards, droughts also are not significant anymore in the correlation with refugees fleeing from a country. This

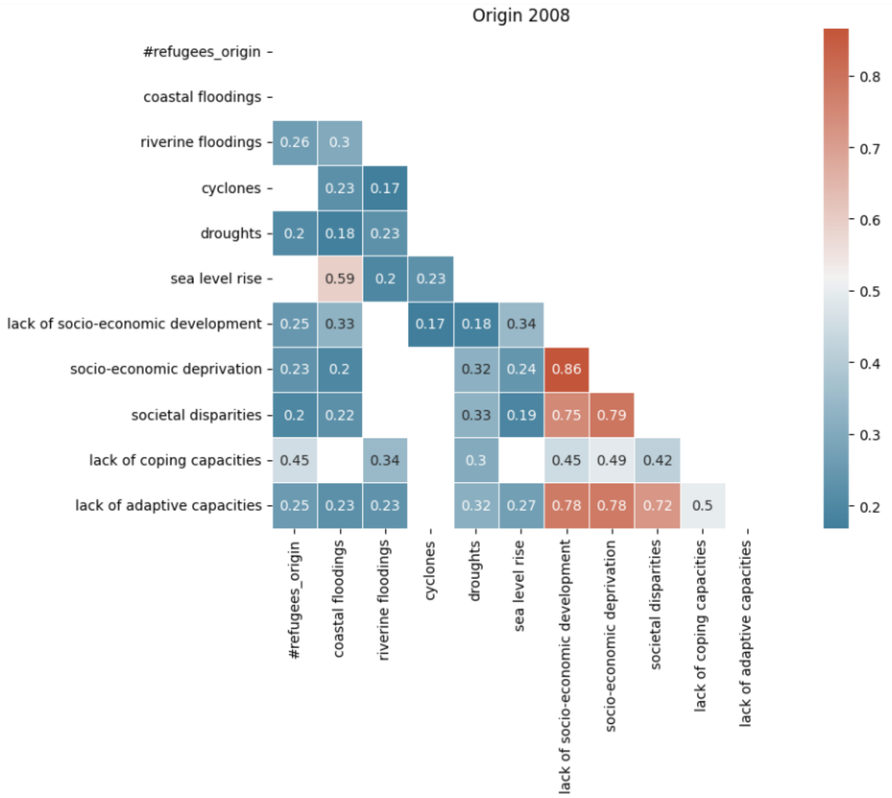


Figure 4.4: Refugee origin location 2008 Distance correlation coefficients

means that in the period 2014-2022 (see Figure 4.5), only the climatic indicator riverine flooding does not seem independent from refugees fleeing a country. Especially the drought correlation coefficient with refugees fleeing is a significant difference compared to the Kendall correlation coefficient results. The other correlation coefficients align quite well.

4.2.2. Results refugees destination

Again, here will be reflected on the Kendall and Distance based correlation, but now for the refugee destination data, which means that the focus in this section is on correlating different climatic/social indicators with refugees fleeing to a country.

4.2.2.1. Kendall correlation

Compared to correlations between indicators and refugees fleeing from a country, there are less significant correlations coefficients between indicators and refugees fleeing to a country. The logical explanation for this could be that in determining where to flee to, climate and social factors play a limited role: refugees just have to flee from the country they come from, without having to much eye for the country they go to. Also, nearly all correlation coefficients have an 'unexpected' positive sign, which can be seen below. Therefore, it cannot be that the identified associations below are the result of an underlying, potential causal mechanism. The explanation for these associations could be that refugees often flee from climatic and/or social poor countries to 'neighboring' countries, which also have poor climate/social conditions. However, it is not the case that one can potentially interpret the association in the way that an increase in lack of coping capacities or riverine floodings leads to more refugees fleeing to a country. It is just a descriptive finding of associations. Note that these signs could change when conducting regression and controlling for other variables.

In 2003 (see Figure 4.6), only the correlation coefficients for refugees fleeing to a country with riverine floodings, droughts and lack of coping capacities are (statistically) significant, and those are quite low (below 0.2) and have the positive sign, which is reflected on above.

Indeed, the correlation between climate indicators and social indicators, and within both categories themselves, remain the same, compared to the origin case, as the same climate and social scores still

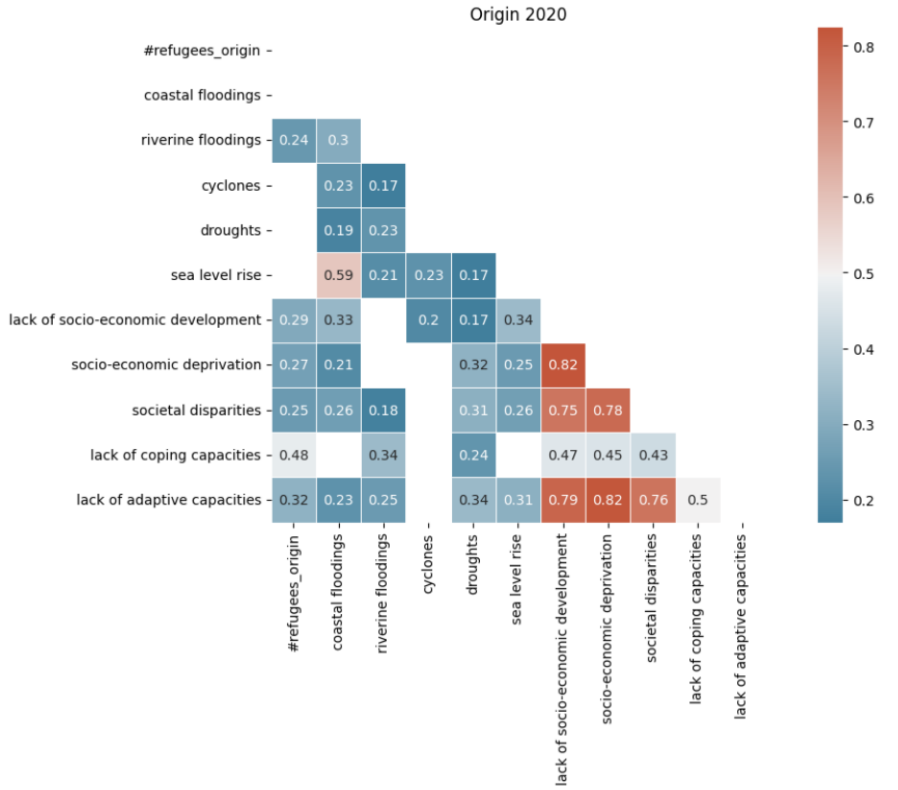


Figure 4.5: Refugee origin location 2020 Distance correlation coefficients

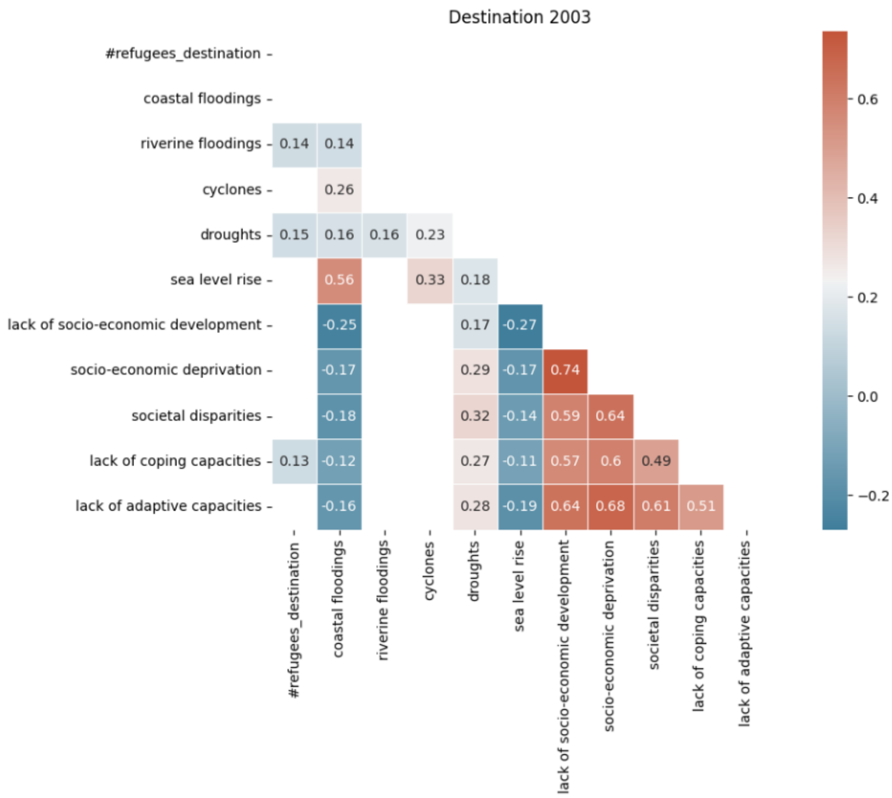


Figure 4.6: Refugee destination location 2003 Kendall correlation coefficients

belong to a country: only the refugee situation and numbers change. (There are some minor differences because not every country has data for both refugees fleeing to an from a country.) So, one can for example still see high correlations between the social factors, again indicating multicollinearity.

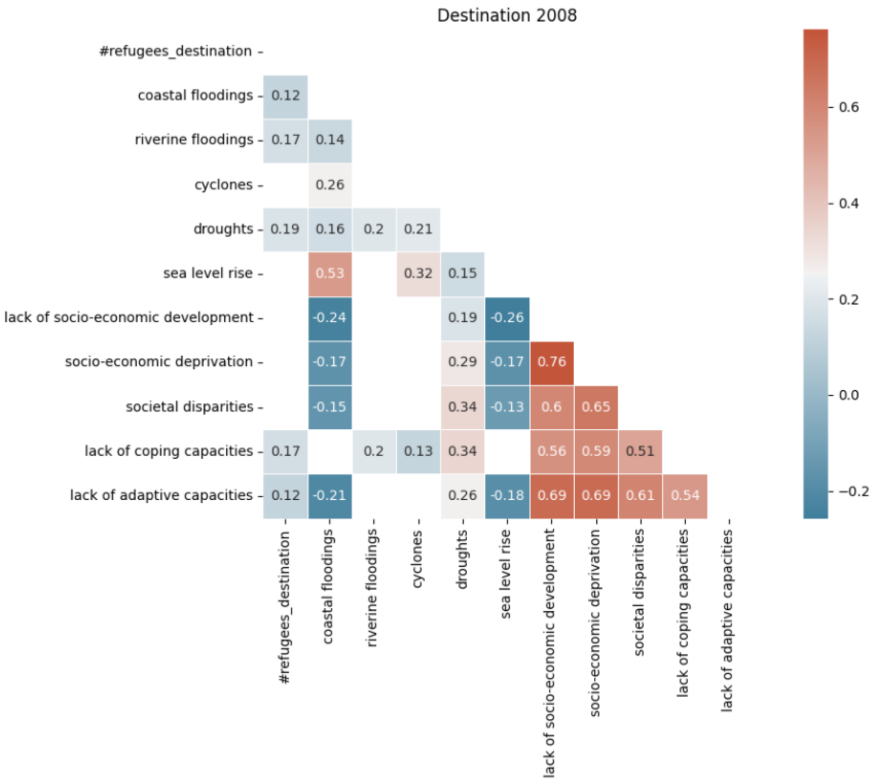


Figure 4.7: Refugee destination location 2008 Kendall correlation coefficients

In the years that follow till 2007, the correlation matrix structure remains the same, with varied significant and insignificant coefficients for lack of coping capacities and lack of adaptive capacities with refugees fleeing to a country (size of approximately. 0.1). Also those have 'unexpected' positive signs. All other correlations between climate/social factors and refugees fleeing to a country are still lower than 0.2. All those coefficients (still) have 'unexpected' positive sign.

In 2007 and 2008 (see Figure 4.7), also the correlation between coastal floodings and refugees fleeing to a country becomes significant (0.13), but this lasts only for two years. Also, sea level rise is significant in 2007 (0.12). But all associations cannot potential be causal due to the absence of a theory: the numbers would then potentially indicate that having more social and climatic problems leads to more refugees fleeing to a country, which is unexplainable. To make this concrete, it is not that the Netherlands, for example, will have more refugees fleeing to it if it has more droughts or less coping capacities. There is of course not an underlying mechanism, there is no theory that describes this, and that is potentially also why the correlations are low. This also indicates that the claim about not thinking about situations in hosting countries could be true: refugees just want to flee from the country they come from.

The above structure (with positive correlations between refugees coming to a country and more social/-climatic problems) remains till 2011. Then, one thing changes: the correlation between cyclones and refugees coming to a country is negative (-0.15) and significant. This could reflect a causal mechanism, potentially concluding that more cyclones could lead to less refugees coming to that country. This structure remains till 2019, thereafter only positive significant correlations remain again. There should not be putted much attention on these, because they have no potential causal interpretation: one can only note that refugees flee to countries where not always the climate and social situating is considered safe. In 2021 (see Figure 4.8) and 2022, cyclones is significant again and slightly negative (around -0.15). All these signs could change when conducting regression and controlling for other variables.

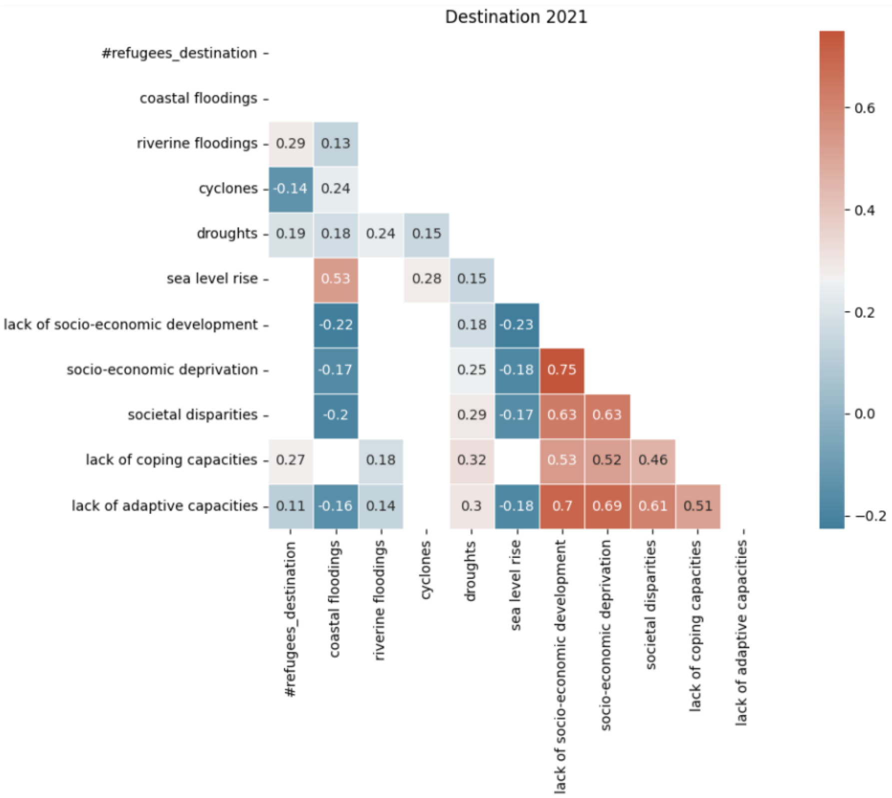


Figure 4.8: Refugee destination location 2021 Kendall correlation coefficients

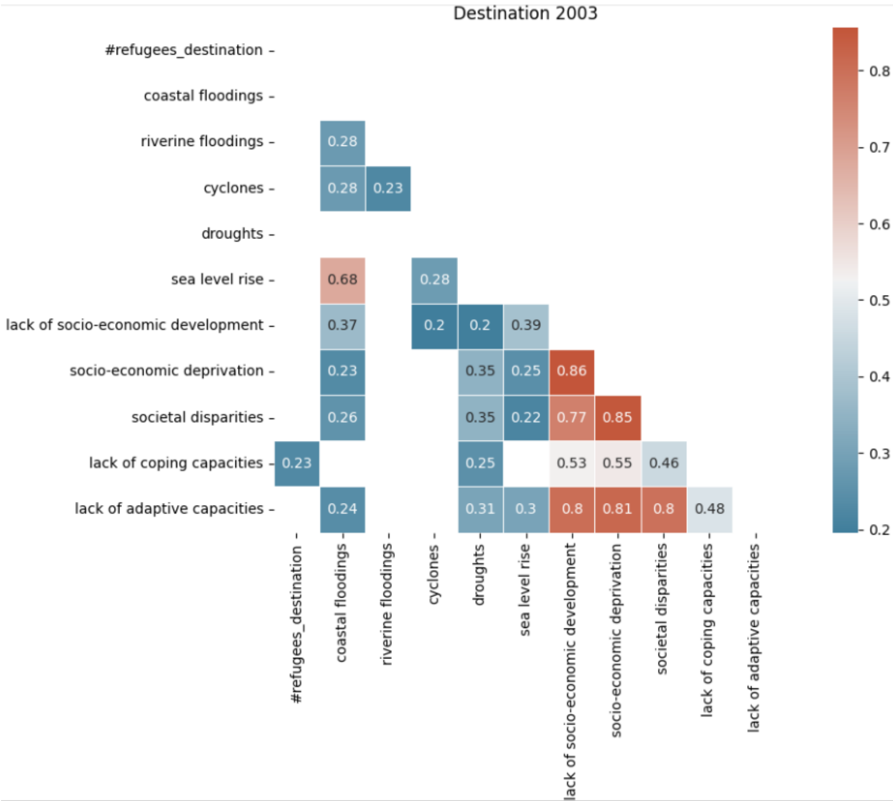


Figure 4.9: Refugee destination location 2003 Distance correlation coefficients

4.2.2.2. Distance based correlation

In 2003 (see Figure 4.9), only the correlation coefficient for refugees fleeing to a country with lack of coping capacities is (statistically) significant (0.23). No other correlations with refugees fleeing to a country are statistically significant. This structure remains till 2009, with one exception in 2006: then non of the correlation coefficients with refugees fleeing to a country are significant, indicating independence according to this Distance based correlation. Again, as indicated in the Kendall analysis, the significant coefficients have 'unexpected signs', indicating no causal mechanism for explanation. However, this can change when conducting multivariate regression.

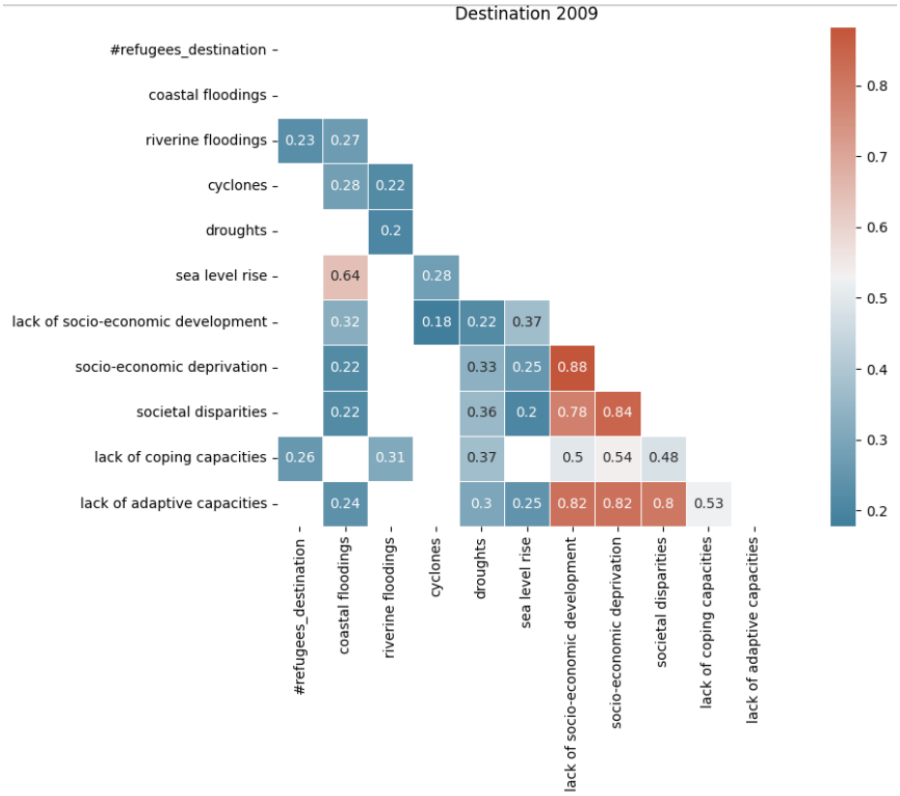


Figure 4.10: Refugee destination location 2009 Distance correlation coefficients

Indeed, the correlations between climate indicators and social indicators, and within both categories themselves, remain the same, compared to the origin case, as the same climate and social scores still belong to a country: only the refugee situation and numbers change. (There are some minor differences because not every country has data for both refugees fleeing to an from a country.) So, one can for example still see high correlations between the social factors, again indicating multicollinearity.

In 2009 (see Figure 4.10), besides lack of coping capacities, also riverine floodings becomes significant in the association with refugees fleeing to a country (0.23). This structure remains till 2012. In 2012, 2013, 2014 and 2015, lack of adaptive capacities also is statistically significant in the correlation with refugees fleeing to a country (around 0.2). Thereafter, this coefficient is again insignificant. In 2013, coastal floodings is one time significant (0.18). After 2015 (see Figure 4.11), again only riverine floodings (0.25) and lack of coping capacities (0.45) are (statistically) significant in their correlation with refugees fleeing to a country.

Overall, according to this Distance based correlation analysis, one can say that over multiple years, refugees fleeing to a country only depends on riverine floodings and lack of coping capacities. All other indicators, apart from some exceptions in one or a couple of years, are overall independent from refugee destination numbers. There is quite some difference here when comparing with the conclusions of Kendall correlation: namely, in most years was concluded that droughts, cyclones and lack of adaptive capacities, besides lack of coping capacities and riverine floodings, also have significant coefficients in their correlation with refugees fleeing a country. Note that again, as indicated in the Kendall analysis, the significant coefficients have 'unexpected' signs, indication no causal theory for explanation. However, this can change when conducting multivariate regression.

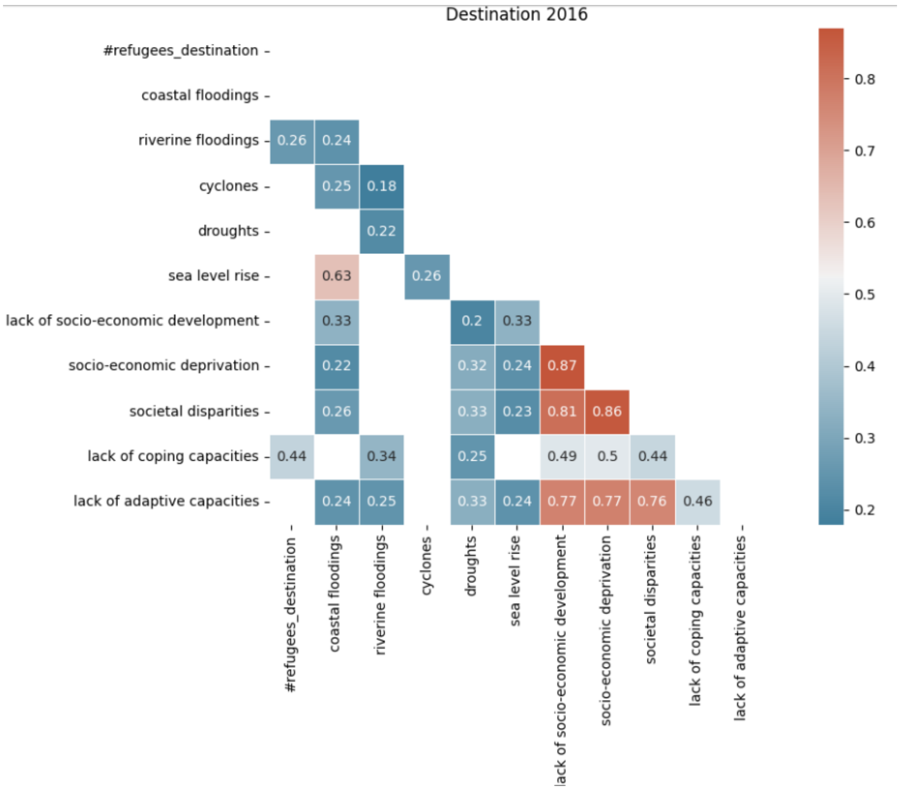


Figure 4.11: Refugee destination location 2016 Distance correlation coefficients

4.2.3. Key observations

4.2.3.1. For refugee origin data based on Kendall correlation

1. Only the climatic indicators riverine floodings and droughts are significant or have the 'right', expected sign, in their correlation with refugees fleeing from a country being around 0.3. Again, other climatic correlation coefficients are not significant, or are low and have the 'unexpected' sign (around -0.15 for sea level rise, cyclones and coastal floodings not significant). Note that these signs could change when conducting regression and controlling for other variables.
2. Social factors correlate relatively high with refugees fleeing, especially lack of coping capacity (around 0.5) and lack of adaptive capacity (0.4). All correlations are significant and have the expected sign. Note again that these signs could change when conducting regression and controlling for other variables.
3. Social factors, taken all together, could cause multicollinearity in later stages of regression analysis.
4. Cyclones and coastal floodings do nearly not have any significant correlation with other indicators.
5. Obviously, correlation between sea level rise and coastal floodings is relatively high (around 0.5 over years).

4.2.3.2. For refugee origin data based on Distance correlation

1. Most results align with results from Kendall correlation, taken into account the different range of possible correlation values. Most important is the alignment on cyclones and coastal floodings not being significant in their correlation with refugees fleeing from a country.
2. Sea level rise is in this analysis also not significant in later years (after 2007), which could be logical, because in Kendall correlation, the correlation did not have the expected sign.
3. The only big difference with Kendall correlation is that the drought correlation coefficient is not significant in the period 2014-2022, which would suggest independence between these variables (droughts and refugees fleeing from a country).

4. Only riverine flooding is an important climatic indicator in association with refugees fleeing away, according to this analysis.
5. Social factor correlations with refugees fleeing from a country, and correlations between social and climate factors themselves, align with Kendall correlation findings, again taking into account the different ranges of values the different correlation techniques have.

4.2.3.3. For refugee origin data based on Kendall and Distance correlation

- Riverine floodings stand out as the most 'important' climate events in association with refugees fleeing (correlation coefficient above 0.3).
- The correlation association between refugees fleeing from a country and the drought variable is inconclusive, because results of Kendall and Distance correlation differ (0.3 significant Kendall correlation vs insignificant Distance correlation).
- Other climate factors are not dependent/significant, or have an 'unexpected sign', in their association with refugees fleeing. Note again that these signs and sizes could change when conducting regression and controlling for other variables.
- Most social factors are associated with refugees fleeing, especially lack of coping capacity (correlation coefficient of approximately 0.5).
- High correlations are observed between different social factors among themselves.
- Relatively high correlation coefficients between sea level rise and coastal flooding is observed.

4.2.3.4. For refugee destination data based on Kendall correlation

1. The only constant significant correlation coefficients over multiple years are those between refugees fleeing to a country and riverine floodings, droughts, lack of coping capacities and lack of adaptive capacities.
2. All these coefficients are rather low, around 0.2 or lower. Only from 2014 onwards, the correlations coefficients for riverine floodings and lack of coping capacities become higher (around 0.3 for riverine floodings and around 0.4 for lack of coping capacities), with an exception for lack of coping capacities in 2022 (0.16). The other correlation coefficients remain below/around (0.2).
3. Those coefficients have 'unexpected' positive signs, indicating that in this analysis there is no causal mechanism to explain this association. However, signs and sizes could change when conducting multivariate regression.
4. Later, after 2011, the correlation between cyclones and refugees fleeing to a country becomes statistically significant (-0.15) and has the expected sign.
5. Other indicators do not play a role, because they have no consistent association with refugees fleeing to a country over years.
6. Less coefficients are statistically significant compared to the origin case, which could possibly be explained by the fact that refugees have to flee their country of origin, without looking at where they will arrive.

4.2.3.5. For refugee destination data based on Distance correlation

1. Only lack of coping capacities coefficient is constantly significant (0.23 in 2003, increases to 0.45 after 2015) in the association with refugees fleeing to a country until 2009. After 2009, also riverine floodings becomes constantly statistically significant in the association with refugees fleeing to a country (0.23).
2. All other indicators are independent from refugees fleeing to a country, according to this analysis.
3. There are differences in comparison with the conclusions of Kendall correlation for the destination data: in most years in the Kendall correlation analysis was concluded that droughts, cyclones and lack of adaptive capacities, besides lack of coping capacities and riverine floodings, also have significant coefficients in their correlation with refugees fleeing a country.

4.2.3.6. For refugee destination data based on Kendall and Distance correlation

- The only constantly significant, important correlation coefficients identified by both analysis are those between refugees fleeing to a country and riverine floodings / lack of coping capacities and those start around size 0.2 and end around 0.3 (riverine floodings) and 0.4 (lack of coping capacities)
- The significant coefficients have 'unexpected positive' signs, indicating that in this analysis there is no causal mechanism to explain this association. However, signs and sizes could change when conducting multivariate regression. Till now, the reason could be that refugees often flee from climatic and/or social poor countries to 'neighboring' countries, which also have poor climate/social conditions. However, it is not the case that one can potentially interpret the association in the way that an increase in lack of coping capacities or riverine floodings leads to more refugees fleeing to a country. It is just a descriptive finding of certain associations.
- Less coefficients are statistically significant compared to the origin case, which could possibly be explained by the fact that refugees have to flee their country of origin, without looking at where they will arrive.

4.2.3.7. All correlation analyses combined

- All correlation coefficients between social/climate factors and refugee movements are rather low (all lower/around 0.3), except for lack of coping capacities with refugees fleeing from a country (around 0.5), lack of adaptive capacities with refugees fleeing from a country (around 0.4), and riverine floodings with refugees fleeing from a country (above 0.3).
- Especially in correlations between different indicators and refugees fleeing to a country, many coefficients are insignificant or have very low values and/or have in this analysis the 'unexpected' positive sign, indicating that increasingly worse conditions are associated with more refugees fleeing to a country. The latter is of course not explained by a causal mechanism. Note however that signs and sizes of coefficients can change in multivariate regression analysis, which is conducted later in this chapter.

4.2.4. Testing correlation changes over years

All procedures and statistical test selections are presented in Appendix B.

4.2.4.1. Testing correlation changes over years for origin refugee data

	Indicator	H value	Mann-Kendall Statistic	Mann-Kendall p-value	Kendall Tau	Theil-Sen Slope	S value	Conclusion
0	coastal floodings	False	-0.746219	4.555354e-01	-0.126316	-0.001128	-24.0	There is no significant trend.
1	cyclones	True	2.952430	3.152837e-03	0.484211	0.002976	92.0	There is a significant Increasing trend.
2	droughts	True	3.017318	2.550218e-03	0.494737	0.002762	94.0	There is a significant Increasing trend.
3	lack of adaptive capacities	True	3.860870	1.129841e-04	0.631579	0.003867	120.0	There is a significant Increasing trend.
4	lack of coping capacities	True	3.276873	1.049637e-03	0.536842	0.003763	102.0	There is a significant Increasing trend.
5	lack of socio-economic development	True	4.315090	1.595380e-05	0.705263	0.006585	134.0	There is a significant Increasing trend.
6	riverine floodings	True	-3.082207	2.054719e-03	-0.505263	-0.002415	-96.0	There is a significant Decreasing trend.
7	sea level rise	False	1.070661	2.843217e-01	0.178947	0.001176	34.0	There is no significant trend.
8	societal disparities	True	5.158641	2.487485e-07	0.842105	0.008121	160.0	There is a significant Increasing trend.
9	socio-economic deprivation	True	3.990647	6.589330e-05	0.652632	0.003664	124.0	There is a significant Increasing trend.

Figure 4.12: Mann-Kendall Test and Theil-Sen Slope for Correlation with Origin data

The table above (see Table 4.12) shows for every indicator whether the correlation coefficient between the indicator and refugees fleeing from a country changes over time. In other words, here is indicated whether or not the correlation increases or decreases in the years 2003-2022, and the magnitude of the change is specified. This represents a trend analysis.

One can immediately conclude that for the indicators coastal floodings and sea level rise, the correlation with refugees fleeing from a country, does not change significantly over time. There is no increase or decrease of these associations.

The eight other indicators do have a significant trend, of which only one indicator has a decreasing correlation coefficient with refugees fleeing from a country over time: riverine floodings. However, the decrease is very minimal: the slope is -0.002. This means that every year the correlation coefficients decreases with 0.002. So in 20 years, this means a decrease of 0.040. This is very minimal.

The other correlation coefficients increase over time, but five of them only minimally. Around 0.003 for cyclones increase per year; the same for droughts; 0.004 for lack of adaptive and lack of coping capacities; and also 0.004 for socio-economic deprivation. This means that in 20 years, the correlation increases with 0.06 or 0.08, for example from 0.2 to 0.28. This is of course significant, but not a drastic increase.

A much higher increase in correlation holds for societal disparities and lack of socio-economic development with refugees fleeing from a country, with respectively a slope of 0.008 and 0.007. This means in 20 years, an increase in correlation coefficients of 0.16 and 0.14.

These results mean that, if this trend continues, social-economic factors have a stronger association with refugees fleeing from countries with a bad social-economic situation. At least, in the last 20 years this was the case. To a somehow lesser extent this also holds for some of the climate factors: droughts and cyclones. For the other climatic indicators there is no clear trend or a very slightly negative one. Note that this does not say anything about the (current) size of the correlation, only about increase or decrease.

One final point worth noting concerns the results in the Kendall Tau column. This column indicates the correlation between the year and the earlier identified correlation coefficients between the indicator and refugees fleeing from a country. The correlation can be quite high, for example 0.63 for lack of adaptive capacities. Then one may think that over the years, the correlation coefficients also rise heavily. This is, however, not the case, because the trend value slope is quite low (0.004 per year). It is only that the two (years and the coefficients) increase together, but the increase is minimal. So, in this example one can better look at the real increase or decrease of the trend.

4.2.4.2. Testing correlation changes over years for destination refugee data

One can immediately conclude that for the indicators coastal floodings, droughts, lack of adaptive capacities, sea level rise, societal disparities and lack of socio-economic development, the correlation with refugees fleeing to a country does not change significantly over time (see Table 4.13). There is no increase or decrease of these associations. There are a lot more insignificant trends compared to the origin case described in the previous section.

The four other indicators do have a significant trend, of which two indicators have a decreasing correlation coefficient with refugees fleeing to a country over time: cyclones and socio-economic deprivation. The decreases are minimal: the slopes are respectively -0.003 and -0.002. This means that every year the correlation coefficients decrease with 0.003 and 0.002 respectively. So, in 20 years, this means a decrease of 0.060 or 0.040. This is very little. It is interesting, because in the case of cyclones, this actually means that the correlation becomes stronger, but more negative. Namely, the correlations coefficient was slightly negative and becomes more negative over years, which could intuitively be plausible because refugees would then not flee to countries which have more cyclones, they rather do not experience more cyclone intensity. In the case of socio-economic deprivation, the coefficient also decreases but was not significantly negative already, so first the correlation becomes weaker (closer to zero), before it can potentially become negative and follow the potential causal explanation.

The other correlation coefficients increase over time (the other two indicators with refugees fleeing to a country). Around 0.008 for riverine floodings and around 0.01 increase per year for lack of coping

	Indicator	H value	Mann-Kendall Statistic	Mann-Kendall p-value	Kendall Tau	Theil-Sen Slope	S value	Conclusion
0	coastal floodings	False	1.524881	1.272887e-01	0.252632	0.002605	48.0	There is no significant trend.
1	cyclones	True	-3.276873	1.049637e-03	-0.536842	-0.003370	-102.0	There is a significant Decreasing trend.
2	droughts	False	1.135550	2.561450e-01	0.189474	0.001058	36.0	There is no significant trend.
3	lack of adaptive capacities	False	-0.291999	7.702877e-01	-0.052632	-0.000602	-10.0	There is no significant trend.
4	lack of coping capacities	True	3.990647	6.589330e-05	0.652632	0.010101	124.0	There is a significant Increasing trend.
5	lack of socio-economic development	False	-0.291999	7.702877e-01	-0.052632	-0.000234	-10.0	There is no significant trend.
6	riverine floodings	True	5.288418	1.233786e-07	0.863158	0.008489	164.0	There is a significant Increasing trend.
7	sea level rise	False	0.000000	1.000000e+00	0.000000	-0.000073	0.0	There is no significant trend.
8	societal disparities	False	-0.551553	5.812548e-01	-0.094737	-0.000351	-18.0	There is no significant trend.
9	socio-economic deprivation	True	-3.276873	1.049637e-03	-0.536842	-0.002455	-102.0	There is a significant Decreasing trend.

Figure 4.13: Mann-Kendall Test and Theil-Sen Slope for Correlation with Destination data

capacities. This is remarkable. This means that in 20 years, the correlation increases with 0.16 or 0.2 respectively. This is of course significant, and a relatively large increase. It is in this analysis unexplainable (from a causal mechanism point of view), because these coefficients were positive already and the trend is that the correlation becomes more positive. The association that becomes stronger is: more riverine floodings/more lack of coping capacities increases together with more refugees fleeing to those countries. This is, as mentioned in the explanation around the correlation destination matrices, not explainable with a causal mechanism, but signs and sizes of coefficients and effects can change when doing multivariate regression. Note that (increasing) associating could just reflect the fact that people flee to 'neighboring' countries that are also societal/climatic in a poor situation. And apparently this effect is increasing over time (over the last years). But it is definitely not causal that more negatives effects make it more attractive for refugees to flee to those countries.

Note that the results in the table above (see Table 4.13) do not say anything about the (current) size of the correlation, only about trend. One final point worth noting concerns the results in the Kendall Tau column. This column indicates the correlation between the year and the earlier identified correlation coefficients between the indicator and refugees fleeing from a country. The correlation can be quite high, for example 0.86 for riverine floodings. Then one may think that over the years, the correlation coefficients will also rise heavily. This is not the case, because the trend value slope is much lower (0.008 per year). It is only that the two variables (years and the coefficients) increase together, but the increase is not very high. So, in this example one can better look at the real increase or decrease of trend.

4.2.4.3. Verification of changing correlation results

Also, Hamed Rao Modified Mann-Kendall tests are conducted, to correct for possible serial autocorrelation in the data (see Figures B.4 and B.5 in Appendix B). The conclusions remain the same with these tests compared to the original, non-modified, Mann-Kendall tests.

4.3. Regression analysis

The previous analyses in this chapter serve as the foundation for this analysis. In this analysis is shown what the signs, sizes and effects are of indicators, when interaction, higher order effects and other variables are taken into account at the same time, in the relation with refugee movements.

The entire regression preparation, manipulation and choices made are presented in Appendix B. In this analysis, the logarithm of refugee movements are the dependent variables (y) and the social/climatic indicators are the independent variables (X). Obviously, for refugee destination as well as refugee origin, different models are estimated, for every year. So, every year, there is a formula that 'predicts' the number of refugees fleeing to/from a country, based on all indicators and interactions plus higher order effects.
The assumed 'theoretical model' or conceptual model in this multiple regression study is the following. Cyclone intensity, sea level rise risk, riverine flooding intensity, coastal flooding intensity, drought intensity, lack of socio-economic development and lack of coping capacities are included as independent variables. Between these independent variables, all interactions are included to control for possible enhancers of social-climate risk and climate exposures/social factors enhancing each other. In addition, higher order effects (squared and cubic) are included of every factor to address the non-linear relations between independent and dependent variables. Again, see Appendix B for more details.

4.3.1. Results

Below, per section, two example equations for two years will be provided. On the GitHub provided by this report, in the file Correlation analysis, coefficients for all the years be be found between lines 113 and 124.

4.3.1.1. Importance

In this section, the data is standardized, such that the model produces comparable coefficients. Those coefficients all have the same interpretation, and are measured in terms of standard deviations. This means that the indicators can be compared, and because of that, the importance/weights of different indicators can be concluded upon. However, the interpretation of a coefficient is now not in original units of indicators and dependent variable, but in standard deviations.

Origin
Below, two equations are shown and explained.

2018

$$\begin{aligned}\text{z-score log(refugees fleeing away)} &= 1.33\cdot\text{z-score sea level rise}+1.59\cdot\text{z-score lack of coping capac}\\ &- 0.63\cdot\text{z-score lack of socio-economic devel}\cdot\text{z-score sea level rise}\\ &+ 4.72\cdot\text{z-score lack of socio-economic devel}^2 - 3.19\cdot\text{z-score lack of socio-economic devel}^3 + e\end{aligned}$$

2020

$$\begin{aligned}\text{z-score log(refugees fleeing away)} &= 1.13\cdot\text{z-score sea level rise}+2.09\cdot\text{z-score lack of coping capac}\\ &- 1.89\cdot\text{z-score riverine flood}^2 + 3.94\cdot\text{z-score lack of socio-economic devel}^2\\ &- 2.81\cdot\text{z-score lack of socio-economic devel}^3 + e\end{aligned}$$

What can be concluded from these standardized coefficients, is that lack of coping capacities has the biggest linear influence on refugees fleeing their country. That is logical, as conflicts and instability of a country are taken into account in this indicator. At the linear second place, however, sea level rise has the biggest influence on refugees fleeing their country. Sea level rise of course can have many effects, of which some are (coastal) erosion, flooding and salinization. Also, lack of socio-economic development and riverine flooding play a role. With the for last definitely having more influence than the last one.

However, per year, there are also quite some differences sometimes. What overall, mostly can be seen, in the equations above, but also in the Table in GitHub line 120, is that the two social factors nearly always play a role in every year. And from the climatic indicators, sea level rise (also almost always plays a role) is definitely the most prominent significant one, followed by riverine floodings and drought sometimes, mostly as interaction terms with riverine floodings. Coastal flooding and cyclones play no (consistent) significant role, and droughts only a small role. It is mostly about sea level rise and the second most important climatic indicator is riverine floodings.

Destination
Below, two equations are shown and explained.

2018

$$\begin{aligned}\text{z-score log(refugees fleeing to)} = & -8.87 \cdot \text{z-score cyclones} \\ & - 1.11 \cdot \text{z-score riverine floodings} \cdot \text{z-score lack of socio-economic devel} \\ & + 18.58 \cdot \text{z-score cyclones}^2 - 10.91 \cdot \text{z-score cyclones}^3 + e\end{aligned}$$

2020

$$\begin{aligned}\text{z-score log(refugees fleeing to)} = & -8.23 \cdot \text{z-score cyclones} \\ & - 1.14 \cdot \text{z-score riverine floodings} \cdot \text{z-score lack of socio-economic devel} + 15.59 \cdot \text{z-score cyclones}^2 \\ & - 5.94 \cdot \text{z-score droughts}^2 - 9.04 \cdot \text{z-score cyclones}^3 + 3.44 \cdot \text{z-score droughts}^3 + e\end{aligned}$$

What can be concluded from these standardized coefficients, is that cyclone intensity has the biggest influence on refugees fleeing to a country. That is remarkable, as it is very different compared to the important origin indicators. There, the social indicators are the most prominent, and cyclone intensity plays no role. Also, droughts and lack of socio-economic development have some influence. Riverine flooding only has effect in interactions with lack of socio-economic development. Sea level rise and coastal flooding play no constant significant role over years. Also lack of coping capacities has a small influence only.

However, per year, there are also quite some differences sometimes, but less than in the origin case. What overall, mostly can be seen, in the equations above, but also in the Table on GitHub line 123, is that cyclone intensity always plays a significant role in every year. This climatic indicator is definitely the most prominent significant one, followed by riverine floodings and droughts. Coastal flooding and especially sea level rise play no (consistent) significant role. The social factors have a relatively small influence, and are only significant in interactions in the latest years.

4.3.1.2. Prediction

In this section, original indicator values are used, to predict real refugee movement numbers. All the indicators have their own unit of measurement and cannot be compared in terms of importance. (That is because all indicators have different unit measurements.) These coefficients can be interpreted as follows: one unit change in the indicator variable causes the value of the coefficients*100 percentage change in the dependent variable. Obviously, the equations include the same indicators as in the standardized case. Only the coefficients change. Again, with the equations below, refugee numbers can be predicted in a certain year.

Origin

2018

$$\begin{aligned}\text{log(refugees fleeing away)} = & \ 0.205 \cdot \text{sea level rise} + 0.254 \cdot \text{lack of coping capac} \\ & - 0.002 \cdot \text{lack of socio-economic devel} \cdot \text{sea level rise} \\ & + 0.011 \cdot \text{lack of socio-economic devel}^2 - 0.000097 \cdot \text{lack of socio-economic devel}^3 + e\end{aligned}$$

2020

$$\begin{aligned}\text{log(refugees fleeing away)} = 0.177 \cdot \text{sea level rise} + 0.322 \cdot \text{lack of coping capac} \\ - 0.004 \cdot \text{riverine flood}^2 + 0.009 \cdot \text{lack of socio-economic devel}^2 \\ - 0.00009 \cdot \text{lack of socio-economic devel}^3 + e\end{aligned}$$

The coefficient of sea level rise in 2020 can be interpreted as: one unit increase in sea level rise intensity score, results in 17.7% more refugees, controlling for all other indicators. In 2018, this is nearly the same, but there cannot be controlled for all other significant effects, due to the interaction between lack of socio-economic development and sea level rise. The total effect on refugees fleeing away is therefore dependent on the score on lack of socio-economic development. The contribution of sea level rise is in 2018: 0.205 - 0.002 * lack of socio-economic development. If one includes the mean score of all the countries for this indicator, which is approximately 48, the contribution of sea level rise to refugees fleeing, under the mean score of lack of socio-economic development and keeping other indicators constant, is 11%.

The coefficient of lack of coping capacities can be interpreted directly in 2018 as well as 2020. When the score of lack of coping capacities increases with one, the number of refugees increases with around 30% (25% in 2018 and 32% in 2020).

The other significant indicators are lack of socio-economic development and riverine floodings. But those have no linear effect, but only higher order and/or interaction. The contribution of lack of socio-economic development to refugees fleeing in 2018 is: -0.002 * sea level rise + 0.022 * lack of socio-economic development - 0.000291 * lack of socio-economic development squared. So, obviously, the effect of lack of socio-economic development is dependent on itself and sea level rise. Say, one takes the mean value of both indicators (28 for sea level rise and 48 for lack of socio-economic development), then the contribution of lack of socio-economic development when it increases one unit on refugees fleeing is 33%, holding other factors constant. In 2020, this contribution is: -0.018 - 0.00027 * lack of socio-economic development squared. Under the mean value, this is 24%.

At last, the contribution of riverine flooding also depends on its own value and is, when assuming the mean value of riverine floodings (33): -0.008 * 33 is -26%. This is remarkable, and does not make sense from a potential causal point of view. It cannot be that an increase in riverine flooding leads to less people fleeing from a country. That would not be logical that an increase in floodings, means less refugees fleeing away. Again, that cannot be causal, but it is empirical. Apparently, a lot of refugees are/stay in riverine flooding areas.

In the end, in terms of potential causal contribution to refugees fleeing from a country, only sea level rise, lack of coping capacities and lack of socio-economic development play a consistent significant role. This is also the generic conclusion over all the years. Thereafter, riverine floodings plays a role, but sometimes with the counter intuitive negative sign for contribution, as can be seen above. The drought indicator only has a small role and is not often significant over the (latest) years. Coastal floodings and cyclones play no role at all in contributing to refugees fleeing.

Destination

2018

$$\begin{aligned}\text{log(refugees fleeing to)} = -1.6813 \cdot \text{cyclones} \\ - 0.0031 \cdot \text{riverine floodings} \cdot \text{lack of socio-economic devel} \\ + 0.0460 \cdot \text{cyclones}^2 - 0.0003 \cdot \text{cyclones}^3 + e\end{aligned}$$

2020

$$
\begin{aligned}
\text{log(refugees fleeing to)} = -1.4711 \cdot \text{cyclones} \\
- 0.0031 \cdot \text{riverine floodings} \cdot \text{lack of socio-economic devel} + 0.0364 \cdot \text{cyclones}^2 \\
- 0.0149 \cdot \text{droughts}^2 - 0.0003 \cdot \text{cyclones}^3 + 0.0002 \cdot \text{droughts}^3 + e
\end{aligned}
$$

In 2018, the contribution of one unit increase in cyclone intensity has the following influence on refugees fleeing to a country: -1.6813 + 0.092 * cyclones - 0.0009 * cyclones squared, controlled for the other indicators. With the mean value of cyclone intensity, approximately 5, the contribution is -124%. Approximately the same holds in 2020. This means that in this situation, 124% less refugees flee to countries with higher cyclone intensity values. This could be a logical finding. It is at least a potential causal explainable, empirical result. However, it does not necessarily mean that refugees look at cyclone intensity to determine where to flee, but it is at least an empirical finding.

Also riverine flooding has approximately the same contribution in 2018 as in 2020. Only the interaction effect with lack of socio-economic development is significant. This means that the influence of riverine floodings on refugees fleeing to a country is dependent on lack of socio economic development: -0.0031 * lack of socio-economic development. Assuming again the mean value for this indicator (48), the result is -15%, controlling for the other factors. So, countries with more riverine flooding intensity are paired with less refugees fleeing to that country.

The contribution for lack of socio-economic development is also the same in 2018 and 2020, and dependent on riverine floodings, due to the interaction one can see in the equation above. One unit increase in lack of socio-economic development, assuming mean value of riverine floodings (33), contributes -10% to the prediction of refugees fleeing to a country, controlling for the other indicators.

At last, the influence of drought intensity on refugees fleeing to a country is examined. This indicator has only significant contribution in 2020 and not in 2018. Only the higher order drought terms are significant, no interaction or linear effect is. The contribution of a one unit increase in drought intensity score to the prediction of the number of refugees coming to a country, is, assuming mean drought value (14): -0.0298 * 14 + 0.0006 * 196, controlling for other indicators. This is approximately -30%. This is very significant, meaning that in an above average drought situation, countries will receive more than 30% less refugees.

In terms of potential causal contribution to refugees fleeing to a country, cyclone intensity is by far the biggest significant indicator, consistent over years. Thereafter, far below consistently significance numbers of cyclone intensity, riverine flooding is the second most consistent significant indicator over the years, a bit more significant than the lack of socio-economic development, over years. The social factor with the biggest influence is then lack of socio-economic development. Throughout the years, the effect of this indicator is nearly always significant, but always in an interaction with another factor. In the equations above, one can see examples of that. In nearly all the cases where lack of socio-economic development is significant, the interaction is between riverine floodings and lack of socio-economic developments. These two indicators enhance each other. What is interesting, is that in latest years, often, the droughts indicator is significant with sometimes a big influence, as one can see above in the 2020 case. Also, one could read earlier that one of the most important indicators in refugees fleeing from a country, is sea level rise. Here, in the destination case (= fleeing to a country) sea level rise plays no role; is not significant over the years. The same holds (to a somewhat lesser extent) for lack of coping capacities. This indicator was the most consistently significant in the earlier origin analysis, but in this destination analysis, the contribution of this indicator is after 2015 not significant anymore. Before that year, the contribution is also limited. Coastal flooding also does not have a consistent significant influence over the years.

4.3.2. Performance

Obviously, the standardized, as well as the original, model have the same performance. Therefore, the origin as well as the destination performance has to be examined only one time.

	Year	R_squared	Adjusted_R_squared	MAE	MSE	RMSE
0	2003	0.686729	0.587801	1.457445	3.454409	1.858604
1	2004	0.643681	0.527670	1.474645	3.705929	1.925079
2	2005	0.685981	0.585303	1.348311	3.135123	1.770628
3	2006	0.670294	0.568473	1.361876	3.162676	1.778392
4	2007	0.703982	0.611200	1.256632	2.716373	1.648142
5	2008	0.677694	0.577421	1.355248	3.020099	1.737843
6	2009	0.666303	0.562486	1.401574	3.178459	1.782823
7	2010	0.659133	0.553865	1.418567	3.254862	1.804124
8	2011	0.673696	0.573662	1.391639	3.096122	1.759580
9	2012	0.679753	0.582286	1.378211	3.024342	1.739063
10	2013	0.692776	0.599946	1.303259	2.931711	1.712224
11	2014	0.685561	0.589862	1.370134	3.062575	1.750021
12	2015	0.680271	0.580800	1.348368	3.072986	1.752993
13	2016	0.663872	0.556922	1.395390	3.313621	1.820335
14	2017	0.679320	0.578808	1.377390	3.263074	1.806398
15	2018	0.703219	0.611566	1.351320	3.087773	1.757206
16	2019	0.721091	0.634958	1.337980	2.928495	1.711285
17	2020	0.717734	0.630564	1.377421	3.046770	1.745500
18	2021	0.717951	0.632727	1.417305	3.152434	1.775510
19	2022	0.704124	0.614074	1.429585	3.304195	1.817745

Figure 4.14: Performance of the linear regression model in all years (y = refugees fleeing from a country)

4.3.2.1. Origin

From Table 4.14, the (adjusted) R-squared is the easiest to interpret and compare. The adjusted R-squared corrects for the predictors that are used in the model. In this case, that are many, due to all the interactions and higher order effects. The range of this performance indicator is between 0 and 1, with 0 being the ability of the model with independent variables to explain nothing (=no variance) in the dependent variable, and 1 being the ability to explain every outcome. One can see, that especially in latest years (R-squared of around 0.7 and adjusted R-squared above 0.6), the model is 'quite good', i.e. above moderate in explaining variance in refugees fleeing from a country. This gives confidence in the prediction capabilities of the models, also for future years with similar indicators.

The results of a relatively good performance also boosts two other things: 1) confidence that not (many) important factors are omitted that should have been included in the analysis; 2) that the coefficients approach the true values because then the prediction error is low. However, note for the latter point that accurate and significant coefficients do tend to improve performance metrics, but a model can still have significant coefficients without a high R-squared if it does not capture enough of the total variance in the outcome.

One also can see that the models become better over time. Over the years, all the performance metrics, from R-squared to Root Mean Squared Error (RMSE), indicate a better fit, i.e. smaller error.

4.3.2.2. Destination

	Year	R_squared	Adjusted_R_squared	MAE	MSE	RMSE
0	2003	0.394717	0.143015	1.655167	4.896560	2.212817
1	2004	0.418886	0.179604	1.686928	4.695085	2.166814
2	2005	0.433484	0.197903	1.612932	4.330550	2.080997
3	2006	0.430922	0.198870	1.645940	4.423322	2.103170
4	2007	0.412562	0.179803	1.725293	4.935515	2.221602
5	2008	0.418768	0.192734	1.728723	4.998173	2.235659
6	2009	0.451691	0.242337	1.701313	4.795688	2.189906
7	2010	0.420191	0.202762	1.744050	4.990077	2.233848
8	2011	0.455736	0.256962	1.777361	4.862101	2.205017
9	2012	0.462511	0.262736	1.763132	4.720009	2.172558
10	2013	0.514253	0.335294	1.676226	4.454216	2.110501
11	2014	0.554110	0.394047	1.687821	4.368862	2.090182
12	2015	0.578308	0.428215	1.620441	4.218383	2.053870
13	2016	0.601784	0.460047	1.599507	4.050499	2.012585
14	2017	0.567900	0.414102	1.635055	4.359555	2.087955
15	2018	0.602398	0.460879	1.570228	3.974702	1.993665
16	2019	0.609873	0.467392	1.501164	3.569664	1.889355
17	2020	0.609852	0.467363	1.520627	3.532392	1.879466
18	2021	0.631947	0.499826	1.472879	3.332538	1.825524
19	2022	0.580274	0.428305	1.626803	4.239288	2.058953

Figure 4.15: Performance of the linear regression model in all years (y = refugees fleeing to a country)

In the destination case, the performance results are noticeably worse than in the origin case. Positive is that the performance, in general, also increases over years. However, the variance in refugees fleeing to a country that can be explained by the model with the independent variables stays below 50% with the adjusted R-squared in latest years. Also, the difference between R-squared and adjusted R-squared is quite high, indicating that lots of indicators cannot explain significant variance in the dependent variable. The difference between the R-squared and the adjusted R-squared is not a big problem, because the sample of countries in this study is very large, approaching the entire population. Therefore potential overfitting plays a small role.

So, in this destination case, the confidence in the predictions is less compared than in the origin case. It suggests that possibly some important other factors are omitted that should be included in the model and that the prediction error is relatively high. However, the significant coefficients still can be interpreted as such, because although the overall variance explanation in the dependent variable may be not very high, that does not mean that all the individual coefficients are not trustworthy.

It is on the other hand not the case that the model is very bad. Especially in 2021 (and 2018, 2019, 2020) to model is moderate, explaining about 50% (adjusted R-squared) or 60% (R-squared) of the variance in the dependent variable.

4.4. Summary

In this chapter, a correlation analysis as well as a regression analysis, both over years, are conducted to identify climatic drivers for refugees fleeing to/from countries. Also, explicitly, changes over years are examined. Early in this chapter is reflected on the fact that causality between climate exposure and refugee movements can not be established in this research. However, significant associations found in the analyses can still point towards possible causality, and are drivers, according to this research.

4.4.1. Correlation

Two correlation coefficients are studied: Kendall and Distance based correlation between climate/societal factors and refugees fleeing from (origin) and to (destination) a country.
Key findings are in the origin case that riverine floodings stand out as the most 'important' climate events in association with refugees fleeing from a country (correlation coefficient above 0.3). Droughts are identified as statistically significant by the Kendall correlation (correlation coefficients a bit below 0.3), but not by Distance based correlation. Other climate factors do not have a (consistent/structural) significant correlation coefficient, or have an 'unexpected sign', in their association with refugees fleeing from a country. (Note that these signs and sizes could change when conducting regression and controlling for other variables.) All social factors are statistically significantly associated with refugees fleeing from a country, especially lack of coping capacity (correlation coefficient of approximately 0.5).
Trends over time in correlation coefficients are examined. A high increase in correlation over years holds for societal disparities and lack of socio-economic development with refugees fleeing from a country, with respectively a slope of 0.008 and 0.007 per year. This means in 20 years, an increase in correlation coefficients of 0.16 and 0.14. If this trend continues, social-economic factors have a stronger association with refugees fleeing from countries with a bad social-economic situation. To a far lesser extent, this also holds for some of the climate factors: droughts and cyclones (increase of around 0.07 in 20 years). For the other climatic indicators there is no clear trend or a very slightly negative one.

Key findings in the destination case are that the only constantly significant correlation coefficients identified by both analysis are those between refugees fleeing to a country and riverine floodings/lack of coping capacities. Those coefficients are around 0.2 in early 2000s and increase to approximately 0.3 (riverine floodings) and 0.4 (lack of coping capacities) by the 2020s. Droughts and lack of adaptive capacity coefficients are partially significant, but have low coefficients. More importantly, all those significant coefficients have 'unexpected positive' signs, indicating that in this analysis there is no causal mechanism to explain this association. (However, signs and sizes could change when conducting multivariate regression.) The reason for these findings could be that refugees often flee from climatic and/or social poor countries to 'neighboring' countries, which also have poor climate/social conditions. Regardless, it is not the case that one can potentially interpret the association in the way that an increase in lack of coping capacities or riverine floodings leads to more refugees fleeing to a country. It is just a descriptive finding of associations. Moreover, in general, less coefficients are (statistically) significant compared to the origin case, which could possibly be explained by the fact that refugees have to flee their country of origin, without looking at where they will arrive. Later, after 2011, the correlation between cyclones and refugees fleeing to a country becomes statistically significant (-0.15) and has the 'expected' negative sign, meaning more cyclone intensity is associated with less refugees fleeing to a country.
Trends over time in correlation coefficients are also examined. The rather low correlation coefficients between coastal flooding, droughts, lack of adaptive capacities, sea level rise, societal disparities, and lack of socio-economic development with refugees fleeing to a country remain relatively stable over time, with many insignificant trends. However, two indicators —cyclones and socio-economic deprivation— show minimal decreases in correlation over time, with slopes of -0.003 and -0.002, respectively. This suggests a stronger negative correlation for cyclones and a weakening correlation for socio-economic deprivation. In contrast, the correlation for riverine floodings and lack of coping capacities increases

by 0.008 and 0.01 per year, respectively, leading to relatively large increases over 20 years. Again, although these increases are 'unexpected', they may reflect that refugees flee to neighboring countries with similar socio-climatic issues, rather than suggesting that negative conditions attract more refugees.

4.4.2. Regression

In the regression analysis, the dependent variables (either refugees fleeing from or to a country) are log transformed. In this chapter is reflected on importance of climatic indicators, placed in perspective with strength of associations with social factors. Further, predictions and performances are examined.

The results indicate for the origin case that the standardized coefficient of lack of coping capacities has the greatest direct linear influence on refugees fleeing their country, as it reflects conflicts and instability. Sea level rise also has a relatively big direct influence. Lack of socio-economic development and riverine flooding also play a role. Social factors are consistently more influential each year than climate variables. Sea level rise is the most prominent climatic factor in all years, followed by riverine floodings and, at times, droughts. Coastal flooding and cyclones have no consistent significant association, and drought only has a minor role.
The results indicate for the destination case that cyclone intensity has the largest influence on refugees fleeing -not- to a country, which contrasts with origin results where social factors are more prominent, and cyclone intensity plays no role. Droughts and lack of socio-economic development also have some impact, while riverine flooding only matters when interacting with socio-economic development. Sea level rise and coastal flooding are not consistently significant, and lack of coping capacities only has a small influence. Cyclone intensity is the most consistently significant climatic factor, followed by riverine flooding and droughts, while social factors have a minor influence, only becoming significant in recent years.

Also, equations for predictions are presented in this chapter, along with percentages increases of refugee movements for one-unit increases in the dependent climate and social variables.

Lastly, the performance of the models is reflected upon. For the origin case, in recent years, the R-squared is around 0.7 and the adjusted R-squared is above 0.6, indicating the model is 'quite good' or above moderate at explaining the variance in refugees fleeing from a country with the independent social and climatic indicators. This suggests the model has reasonable predictive power for future years with similar indicators and equations.
In the destination case, the model performs worse than in the origin case. While performance still improves over time, the adjusted R-squared remains below 50% in recent years, indicating that less than half of the variance in refugees fleeing to a country is explained by the model. The large gap between the R-squared and adjusted R-squared suggests many indicators do not explain significant variance.

5 Results SQ3 and SQ4: Spatial distributions

In this chapter, spatial clusters of climate and refugee variables will be identified, presented and analyzed. Also, a comparison will be made with original, regular indicator values in addition to the clustering values. Moreover, the (statistical) changes over years of both the clustering and regular values are examined.

Spatial distributions are examined to gain a deeper understanding of the variables individually; to identify potential changes over years better, and to gain spatial structure. In these analyses, ‘spatial weight’ is a key notion. Spatial weight matrices formalize geographical relationships between the observations in a dataset. Each cell in the matrix contains a value that represents the degree of spatial interaction between observations. An important concept in this context is that of neighbor. In this case, a neighbor is a country that shares a border with the country of interest.

Also, ‘spatial lag’ is a key notion. Spatial lag, when weights are row-standardized, can be interpreted as the average climatic (e.g. drought) score in the surrounding countries of a given country. The way to interpret the spatial lag for a country is as follows, e.g.: Afghanistan, where the drought intensity score is 2.44 (which is quite low) is surrounded by neighboring countries where, on average, the drought score is above 10 (way higher).

An appropriate tool is still needed that indicates where statistical significant clusters are. Therefore, one needs a local measure of spatial autocorrelation. That is what Local Indicators of Spatial Association (LISAs) do. The core idea is to identify cases in which the comparison between the value of an observation and the average of its neighbors is either more similar (High-High, Low-Low) or dissimilar (High-Low, Low-High) than one would expect from pure chance. LISAs are widely used in many fields to identify clusters of values in space. They are a very useful tool that can quickly return areas in which values are concentrated and provide suggestive evidence about the processes that might be at work. What is typically done is to create a map, a cluster map, that extracts the significant observations (those that are highly unlikely to have come from pure chance) and plots them with a specific color depending on their quadrant category. These maps will be visualized in these analyses for every year, for every indicator.

The alternative of LISA is the Hotspot analysis, which is also often used in spatial studies, such as in work of Casali et al. (2024) and Aydin et al. (2023). In this Hotspot analysis, the main statistic used is the Getis-Ord Gi*, which only identifies two categories: High-Highs (HH) and Low-Lows (LL), i.e. hotspots and coldspots. This would mean for this study that only countries which are surrounded by high value countries (low) and have high (low) values themselves, are identified as significant. In this study, on purpose is chosen for LISA with four categories, as in the context of refugees fleeing to/from countries, it is extremely important to know what countries have high (low) values and are surrounded by countries with low (high) values, because neighboring countries could be safer (unsafer) then in terms of climatic exposure. Refugees often flee to neighboring countries, in which the climatic situation can be very different. This is what this study wants to identify. If then countries are not identified in a cluster in which high (low) values are surrounded with low (high) values, a lot of information is missed.

5.1. Droughts

5.1.1. Spatial clustering over years

In the beginning of this chapter, and foremost in the GitHub Drought file provided by this thesis, the spatial weights, lags and autocorrelations are explained carefully. Therefore, the analysis below will focus on the results.

5.1.1.1. Visual spatial clusters

Regarding the local spatial autocorrelation (LISA scores), the same continents/countries are clustered together over time (see Figure 5.1). The clustering nearly does not change in 20 years. One can see that in Central and mainly (South) East Europe, and Turkey, drought scores are low. Countries there have low drought intensity scores and are surrounded by other countries with a similar low score.

In those 20 years, the continent South America is always a cluster of high drought scores, surrounded by high drought scores. The same pattern in Central Africa can be identified on a smaller scale. Further, one sees over the years 2000-2022 that Russia often can be classified as a country with high drought scores, surrounded by countries with low drought scores. The opposite holds for the United States: that country often is classified as a country with low drought scores, surrounded by countries with high drought scores. What is interesting is that Nepal always is classified as HH, meaning that drought scores are high, surrounded by other high drought scores. This means that India, China and Bangladesh, for example, have high drought scores. Other continents mostly have non-significant data.

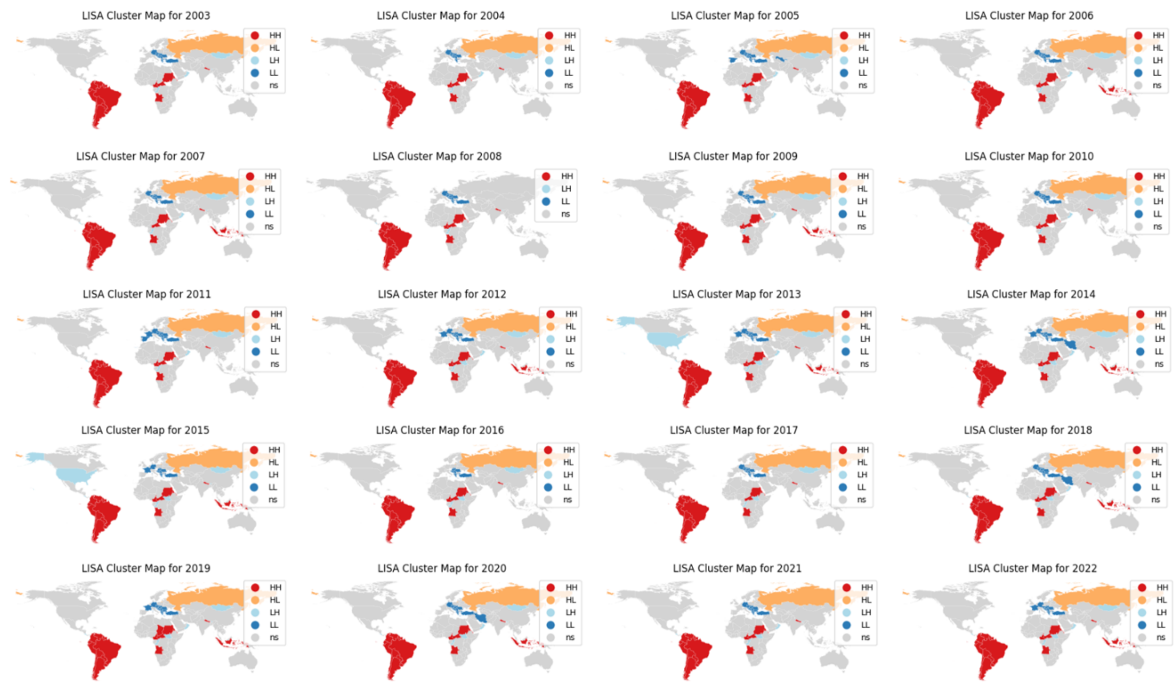


Figure 5.1: Droughts LISA Cluster Maps 2003-2022

So, especially refugee movements in Africa and a bit in Asia are interesting to study from a drought intensity perspective, because there are also significant drought differences between neighbors. In South-America, every country has high drought intensity, so from a drought perspective it is less interesting to which specific country people flee. In Europe, nearly all the countries have low drought intensity, so it does not matter to which country in Europe refugees flee, from a drought perspective. Other regions do not have significant local spatial autocorrelation, which means no specific, but random, clustering. This still can mean that neighboring countries differ heavily in terms of drought scores, but not in a structured way across all neighbors, which could be very interesting when refugees flee to neighboring countries. Therefore, no significance gives also a lot of information and can also be interesting to investigate from a 'fleeing to neighboring countries' perspective. To conclude, HH and LL indicate that refugees will flee to similar drought exposure countries. LH or HL indicates difference in drought exposure for neighboring countries.

5.1.1.2. Statistical cluster changes

In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare changes.

The Kendall correlation result of 0.93 (see Appendix C) between 2003 and 2022 LISA cluster values confirms the line of thought that drought intensity scores did not really change over the last 20 years. Namely, combining this correlation information with the scatter plot (see Figure C.1, in which there are not many changes, one can conclude that the local spatial clustering remains approximately the same over years. A correlation of 0.93 is extremely high, approaching 1. With 1, the drought numbers would be exactly the same in this context.

Apart from a simple correlation coefficient and bivariate plot, a statistical test is conducted to test the differences (see Appendix C) between LISA cluster values of 2003 and 2022. The Wilcoxon Signed-Rank test (W statistic = 3339.0, p-value = 0.0122) indicates a significant difference between the LISA values in 2003 and 2022, only at the 5% level, not at the one percent level. The outlined differences can be seen in Figure C.3.

Although the test being significant, in general, the changes are minor. In addition, more changes are negative than positive, so drought clustering of countries with similar values seems to be decreasing over time in general. Although the biggest clustering changes are positive, the most cluster changes are slightly negative. Small clustering changes do not hold for three countries: Cambodia, United Arab Emirates and Yemen. This means that for those three countries, the local spatial clustering changed heavily (more clustering with countries with similar values HH or LL for Cambodia and United Arab Emirates) and more towards clustering of dissimilar values HL or LH with other countries for this drought indicator for Yemen. But other than this, rather small changes are observed.

5.1.2. Drought changes over years

In this section, thee original, regular drought values are compared over years, and not the changing clustering values (LISA values), as was done in the previous section.

Before comparing the change in 2003 and 2022 drought values, it needs to be examined which correlation coefficient and statistical test can be used. Details on this procedure can be found in Appendix C.

The Kendall correlation between 2003 and 2022 original drought values is 0.98. The Wilcoxon Signed-Rank indicates that the difference in drought indicator values between 2003 and 2022 is not statistically significant (Test Statistic: 2163.5, p-value: 0.090). It is clear from the results above (also see Figure C.5) that the actual drought numbers are rather similar in 2003 and in 2022. This is because the correlation is approaching 1 and the test results indicate no significant change in median between the two year groups. Also, nearly all the points in the scatter plot are on a straight line. So, according to this analysis, drought intensity numbers have not changed over the last 20 years.

5.2. Riverine floodings

5.2.1. Spatial clustering over years

5.2.1.1. Visual spatial clusters

In the beginning of this chapter, and foremost in the riverine flooding indicator file on GitHub, the extensive explanation of spatial weight, lag and autocorrelations are presented. Therefore, in the analysis below, the focus is on the results.

Firstly, what can be seen in Figure 5.2 are many insignificant values/clusters. That is, 'random' spatial clustering, and it means: no structure. Only in Asia, Africa and the Middle East, there are significant clusters.

Figure 5.2: Riverine floodings LISA Cluster Maps 2003-2022

In Asia, nearly all the countries have high riverine flooding intensity and are surrounded by countries with this similar exposure. There are three exceptions: Nepal, Mongolia and Kirghistan. Those countries themselves have low riverine flooding intensity, but are all surrounded with high riverine flooding exposure countries. This is the situation for Asia in the last 20 years. Overall, it should be clear that Asia it the hotspot of riverine flooding danger.

In the Middle East, there is a much smaller cluster of low riverine flooding intensity countries surrounded by countries with also little riverine flooding intensity around Saudi Arabia and Oman. This means that neighboring countries of those countries also have low riverine flooding intensity.

In Africa, the Central Africa Republic forms the core of a cluster with high riverine flooding intensity numbers. For the rest, there are no stable clusters over many years. However, sometimes in the far North and far South in Africa, there are small clusters of countries with low riverine flooding intensities surrounded by similar countries. In the South East, Mozambique forms the core of a cluster with high riverine flooding exposure surrounded by low riverine flooding exposure. The same as for Mozambique holds for Somalia at the East Coast.

5.2.1.2. Statistical cluster changes

In Appendix C is reflected on the choice for the correlation coefficient and statistical test.

The Kendall correlation coefficient between LISA scores of 2003 and 2022 is 0.83. Together with the scatter plot (see Figure C.6), this confirms that the sea level rise risk clusters in the world, are very similar over the years. Namely, combining this information one can conclude that the local spatial clustering remains approximately the same over years.

The the Wilcoxon Signed-Rank test is conducted (W statistic = 4432.0, p-value = 0.9579) to formally test the statistical differences. It is indeed concluded that there is no significant difference between 2003 and 2022 LISA riverine flooding values.

5.2.2. Riverine flooding changes over years

In this section, the original, regular sea level rise risk values are compared from 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The Pearson correlation coefficient between 2003 and 2022 riverine flooding values is 0.99. The paired t-test indicates no significant differences between the 2003 and 2022 values (T-Statistic: -0.13, p-value: 0.90). Together with the scatter plot of the actual values (see Figure C.8) can be concluded that sea level rise did not change heavily in those 20 years. Actually, changes are very minimal.

It is clear from the results above that the actual riverine flooding intensity numbers are relatively similar in 2003 and 2022. This is because the correlation is very high and the test results indicate no significant change in means between the two year groups. Also, nearly all the points in the scatter plot are on a straight line with the same values on the x and y-axes. So, according to this analysis, riverine flooding intensity has not changed over the last 20 years.

5.3. Coastal floodings

5.3.1. Spatial clustering over years

5.3.1.1. Visual spatial clusters

In the beginning of this chapter, and foremost in the coastal flooding file on GitHub, the extensive explanation of the spatial weight, lag and autocorrelations are carefully documented. Therefore, in the analysis below, the focus is on the results.

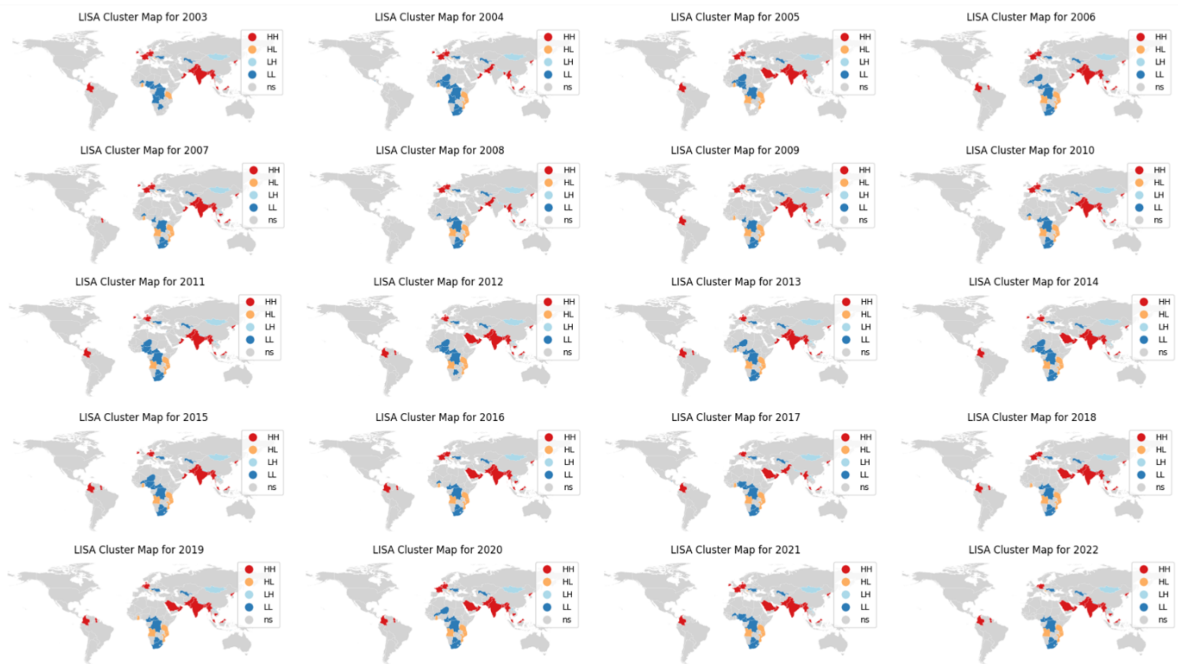


Figure 5.3: Coastal floodings LISA Cluster Maps 2003-2022

The statistical clusters do not seem to change much over time (see Figure 5.3). On the continent America, only one cluster is statistically significant. In Northern South America, high coastal flooding intensity is surrounded by other high coastal flooding value countries. This is the case around Colombia and Guyana.

Western Europe also forms a big hotspot of high exposed coastal flooding countries surrounded with other high coastal flooding intensity countries. Only Switzerland is a county with low coastal flooding intensity, surrounded by high flooding intensity. Then in Eastern Europe, there is a cluster of low coastal flooding exposure surrounded by low coastal exposure. Except for Croatia, that is the only country with high coastal flooding intensity.

In Africa, around the country South-Africa there is a cluster of low coastal flooding intensity surrounded by low coastal flooding intensity, However, in the South on the West and East coasts, there are clusters of high exposure to coastal floodings, centered around Angola (West) and Mozambique and Tanzania (East). In the Center, Congo forms the core of the cluster of countries with low coastal flooding exposure surrounded by low coastal flooding risk. This is logical, as no coast is near these countries. This cluster extends in later years to the North West, reaching through Nigeria and Niger, all with low coastal flooding values. However, Ghana ends this cluster, being a country with high coastal flooding intensity, surrounded by the low valued countries. This picture varies somewhat over the years, but overall, this remains the general situation.

In the Middle East, there is a cluster of high coastal flooding intensity around Oman and Saudi Arabia, but mostly insignificant values in other countries in the Middle East. Here, in general, coastal floodings are not a big problem.

In (East) Asia, Pakistan, India, Bhutan, Bangladesh, Myanmar, Cambodia and Malaysia form clusters of high coastal flooding intensity. Clearly, the most exposure is in (East) Asia. The same is true for North Korea, meaning China and South Korea also have high exposure to coastal flooding risks. In this region, only Uzbekistan forms the core of a cluster of low coastal flooding intensity values surrounded by other low scores. This is more towards the (North) West of Asia. At last, Nepal, Laos and Mongolia are low coastal flooding exposed countries, surrounded by the other high value countries in this region. Again, this is logical, as those countries are not near a coast. Those countries are the safe haven in East Asia, in terms of exposure to coastal floodings.

The description above is more or less constant over the years. So, in terms of interesting areas for refugee movements: Central and East Africa, East Asia and the Middle East, Africa and Asia are the most interesting from a refugee movement point of view, because high as well as low (cluster) coastal flooding intensity countries are here.

5.3.1.2. Statistical cluster changes

To examine statistically that coastal flooding clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used in this analysis. Therefore, assumptions are examined. This discussion and selection can be found in Appendix C.

The Kendall correlation coefficient between LISA scores of 2003 and 2022 is 0.91. Together with the scatter plot (see Figure C.10), this confirms that the coastal flooding clusters in the world, are very similar over the years. Namely, combining this information, one can conclude that the local spatial clustering remains approximately the same over years.

The Wilcoxon Signed-Rank test is conducted (W statistic = 4366.0, p-value = 0.84) to formally test the statistical differences. It is indeed concluded that there is no significant difference between 2003 and 2022 LISA coastal flooding values.

5.3.2. Coastal flooding changes over years

In this section, the original, regular coastal flooding intensity values are compared in 2003 and 2022, and not the changing clustering values (LISA values) are compared, as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The Kendall correlation coefficient between 2003 and 2022 coastal flooding values is 0.98. Together with the scatter plot of the actual values (see Figure C.13), it can be concluded that coastal flooding intensity did not change heavily in those 20 years. Actually, changes are minimal.

Also the findings of the Wilcoxon Signed-Rank test confirm this: the difference between the values is not statistically significant (Test Statistic: 1140.0, p-value: 0.13).

It is clear from the results above that the actual coastal flooding numbers are pretty much similar in 2003 and in 2022. This is because the correlation is very high and the test results indicate no significant change in median between the two year groups. Also, nearly all the points in the scatter plot are on a straight line with the same values on the x and y-axes. So, according to this analysis, coastal flooding intensity values have not changed significantly over the last 20 years.

5.4. Sea level rise

5.4.1. Spatial clustering over years

5.4.1.1. Visual spatial clusters

In the beginning of this chapter, and foremost in the sea level rise file on GitHub, the extensive explanation of the spatial weight, lag and autocorrelations are presented. Therefore, in the analysis below, the focus will be on the results.

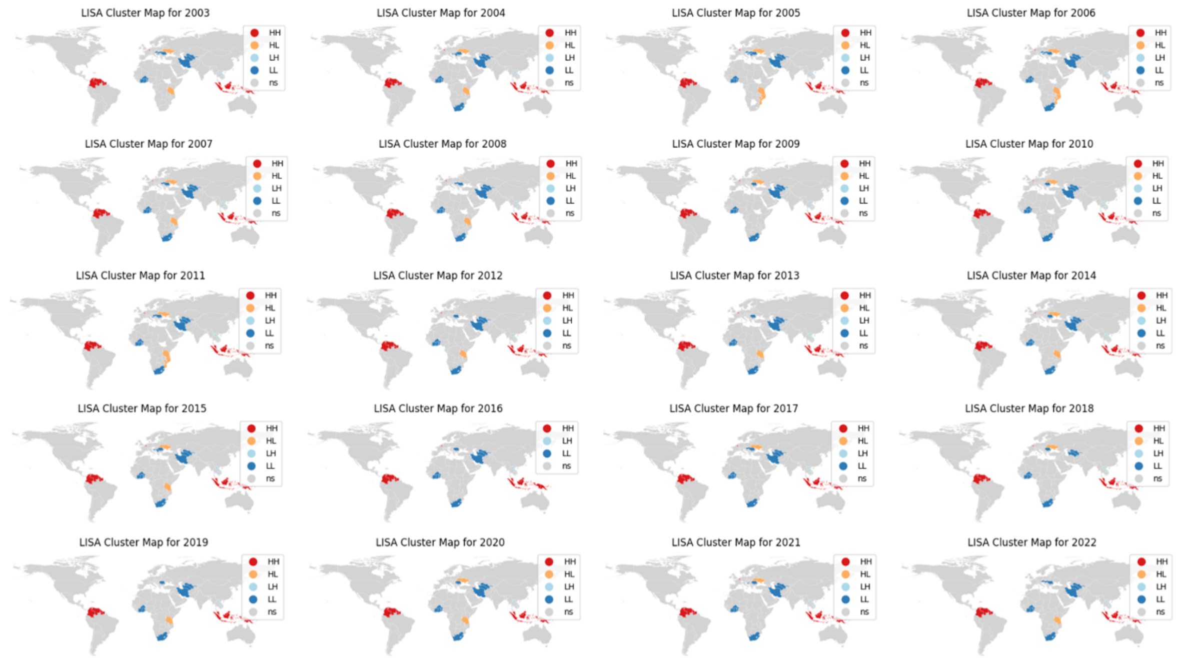


Figure 5.4: Sea level rise LISA Cluster Maps 2003-2022

Firstly, one can see many insignificant values/clusters (see Figure 5.4). That is, random spatial clustering, and means: no geographical structure.

Over the years, high sea level risk values are present in Northern South-America. And around Indonesia, over the years, there is a persistent and stable cluster of high sea level risk areas.

Remarkable is the fact that for the most interesting regions in the world in terms of refugee movements: Africa, Middle East and (East) Asia, not many clusters are formed. In Africa, there are some small clusters of low sea level rise risk in South-Africa and in (North) West-Africa. In addition, what is very interesting, is that Tanzania (and Malawi) are the core of a cluster in which Tanzania has high sea level rise risk, but is surrounded by countries with little risk. So, in terms of sea level rise risk, Tanzania (and Malawi) are unsafe, according to this analysis (see Figure 5.4).

Laos is interesting in Asia, because it forms the core of a cluster in which Laos has little sea level rise risk, but is surrounded by high sea level rise risk areas. So, in terms of sea level rise risk, Laos is a safe haven in high risk surrounded countries.

In the Middle East, there is in general very little sea level rise risk. Iran, Turkmenistan and Uzbekistan form the core of a big cluster in which all neighboring countries and the core itself have little sea level rise risk. Sea level rise risk forms no problem in this region.

Europe as region is not that interesting for this study. This study wants to map out climate exposure/changes in areas where people, refugees and the government cannot cope with it themselves. In Europe, often, can be coped with climate events. Besides, there are also no refugees fleeing from Europe (apart from 2022 Ukraine), but only to Europe on this continent. However, there are also no big consistent significant clusters in Europe, regarding sea level rise risk. East Europe forms usually a cluster of low sea level rise risk values, except for Ukraine, that has high sea level rise risk. In West Europe, there is a small cluster of high sea level rise risk values, centered around Belgium.

5.4.1.2. Statistical cluster changes

To examine statistically that sea level rise clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used. This discussion and selection can be found in Appendix C.

The Kendall correlation coefficient between LISA scores of 2003 and 2022 is 0.89. Together, with the scatter plot (see Figure C.14), this confirms that the sea level rise risk clusters in the world, are very similar over the years. Namely, combining this information one can conclude that the local spatial clustering remains approximately the same over years.

The Wilcoxon Signed-Rank test is conducted (W statistic = 3796.0, p-value = 0.14) to formally test the statistical differences. It is indeed concluded that there is no significant difference between 2003 and 2022 LISA sea level rise values.

5.4.2. Sea level rise changes over years

In this section, the original, regular sea level rise risk values are compared in 2003 and 2022, and not the changing clustering values (LISA values) are studied, as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The Kendall correlation coefficient between 2003 and 2022 sea level rise values is 0.96. Together with the scatter plot of the actual values (see Figure C.17) can be concluded that sea level rise did not change heavily in those 20 years. Actually, changes are very minimal. Also, the findings of the Wilcoxon Signed-Rank test confirm that: the difference between the values is not statistically significant (Test Statistic: 1768.0, p-value: 0.54).

It is clear from the results above that the actual sea level rise numbers are pretty much similar in 2003 and in 2022. This is because the correlation is very high and the test results indicate no significant change in median between the two year groups. Also, nearly all the points in the scatter plot are on a straight line, with the same values on the x and y-axes. So, according to this analysis, sea level rise risk values have not changed over the last 20 years.

5.5. Cyclones

5.5.1. Spatial clustering over years

5.5.1.1. Visual spatial clusters

One can immediately see in Figure 5.5 that significant clusters are always in North America, Africa, Europe and East Asia. However, in Europe and Africa, the clusters seem to change over years, i.e. cyclone exposure clusters are different over years. North America and East Asia clusters are rather constant over the last twenty years.

In North America, one can see that the United States (and Mexico) form(s) the cluster of high cyclone intensity exposure surrounded with high cyclone intensity exposure countries. Canada has low exposure to cyclone intensity. So, North America, apart from Canada has relatively high exposure to cyclone risks.

In East Asia, there are mostly clusters of high cyclone risk countries. Around China, Korea, Vietnam, Bangladesh, India, countries with high cyclone exposure are centered. However, some countries have low cyclone intensity exposure. These countries are Nepal, Pakistan, Mongolia, Bhutan and Kazakhstan. These countries are surrounded by countries with high exposure to cyclone risks.

In Europe, there sometimes is a big cluster of mostly West European countries with low cyclone risk exposure surrounded by similar countries. However, sometimes this cluster (of for example France, Germany and neighboring countries) are not significant, and then there is no Western European cluster. However, sometimes also South-East Europe is a significant cluster of low cyclone risk countries. There does not seem to be a structure in this over years.

In Africa, sometimes only central Africa forms the cluster of low cyclone exposure risk countries, but sometimes also/only the whole of North Africa belong to the same category of low cyclone intensity exposure. There does not seem to be a structure in this over years.

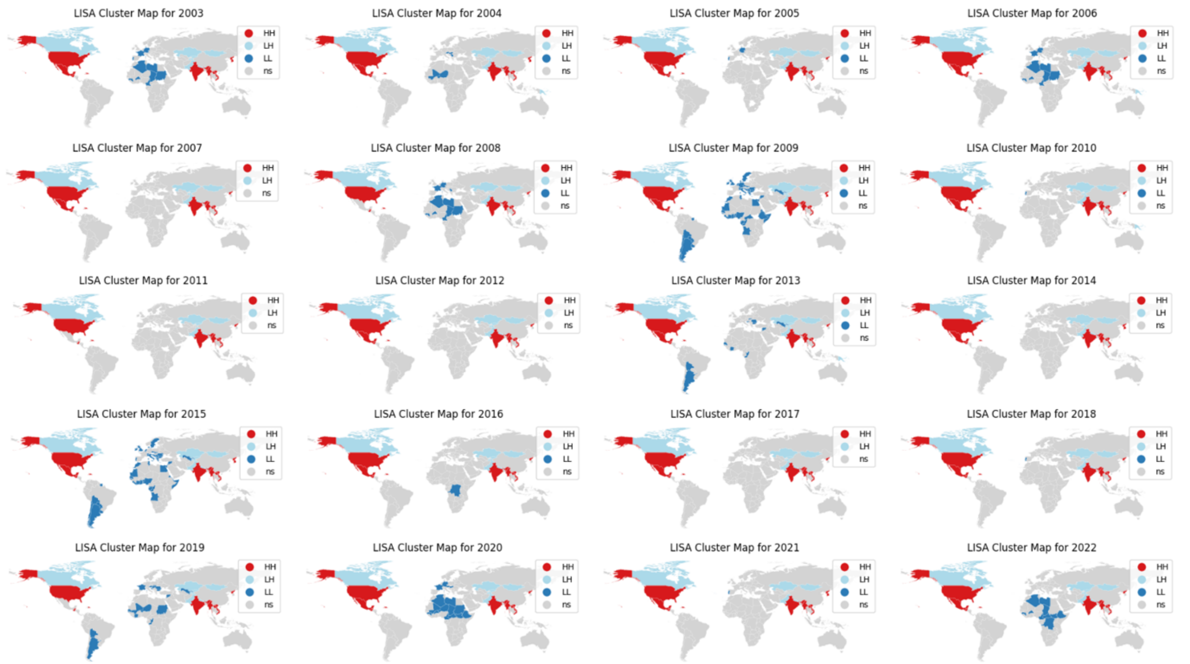


Figure 5.5: Cyclones LISA Cluster Maps 2003-2022

Sometimes, in South America, there is also a cluster of low cyclone exposure countries surrounded by low concentration countries, especially surrounded around Argentina and Bolivia.

To conclude, on the continent America, it seems clear that the United States is highly exposed to intense cyclones, surrounded by similar countries, apart from Canada. So the focus is on the United States. Further, in South America, Africa and Europe, only low cyclone intensity clusters or insignificant values are visible. This means low cyclone risk values on these continents.

Asia is the most difficult continent to assess, because mostly in the East there seem to be very intense cyclone countries surrounded by similar countries, but with quite some exceptions, especially more in the West and North. Asia definitely is the most interesting area to study more regarding cyclone intensities (and refugee movements).

5.5.1.2. Statistical cluster changes

To examine statistically that cyclone intensity clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used. This discussion and selection can be found in Appendix C.

The Kendall correlation coefficient between LISA scores of 2003 and 2022 is 0.92. Together, with the scatter plot (see Figure C.18), this confirms that the cyclone intensity clusters in the world, are very similar over the years. Namely, combining this information, one may conclude that the local spatial clustering remains approximately the same over years.

The Wilcoxon Signed-Rank test is conducted (W statistic = 1456.0, p-value = 1.3897e-11) to formally test the statistical differences. It is concluded that there is a significant difference between 2003 and 2022 LISA cyclone intensity values. That was not expected from the plots (see Appendix C) and correlation coefficient, but it could be expected from the global map presented in Figure 5.5.

The differences are pointed out in Figure C.20. Most differences are rather small, apart from Mexico and the United States. Those two countries experience a huge increase in positive spatial autocorrelation, which means that there neighboring countries (each other, for instance) get more aligned values in the period 2003-2022.

5.5.2. Cyclone changes over years

In this section, the original, regular cyclone intensity risk values are compared in 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The Kendall correlation coefficient between 2003 and 2022 cyclone intensity values is 1.00. Together with the scatter plot of the actual values (see Figure C.22) can be concluded that cyclone intensity did not change heavily in those 20 years. Actually, changes are very minimal. Also, the findings of the Wilcoxon Signed-Rank test confirms that: the difference between the values is not statistically significant (Test Statistic: 33.5, p-value: 0.23).

It is clear from the results above that the actual cyclone intensity numbers are similar in 2003 and in 2022. This is because the correlation is 1, and the test results indicate no significant change in median between the two year groups. Also, nearly all the points in the scatter plot are on a straight line with the same values on the x and y-axes. So, according to this analysis, cyclone risk values have not changed over the last 20 years.

5.6. Countries refugees flee from

5.6.1. Spatial clustering over years

In this section, the spatial analysis for the refugee origin data is conducted.

5.6.1.1. Visual spatial clusters

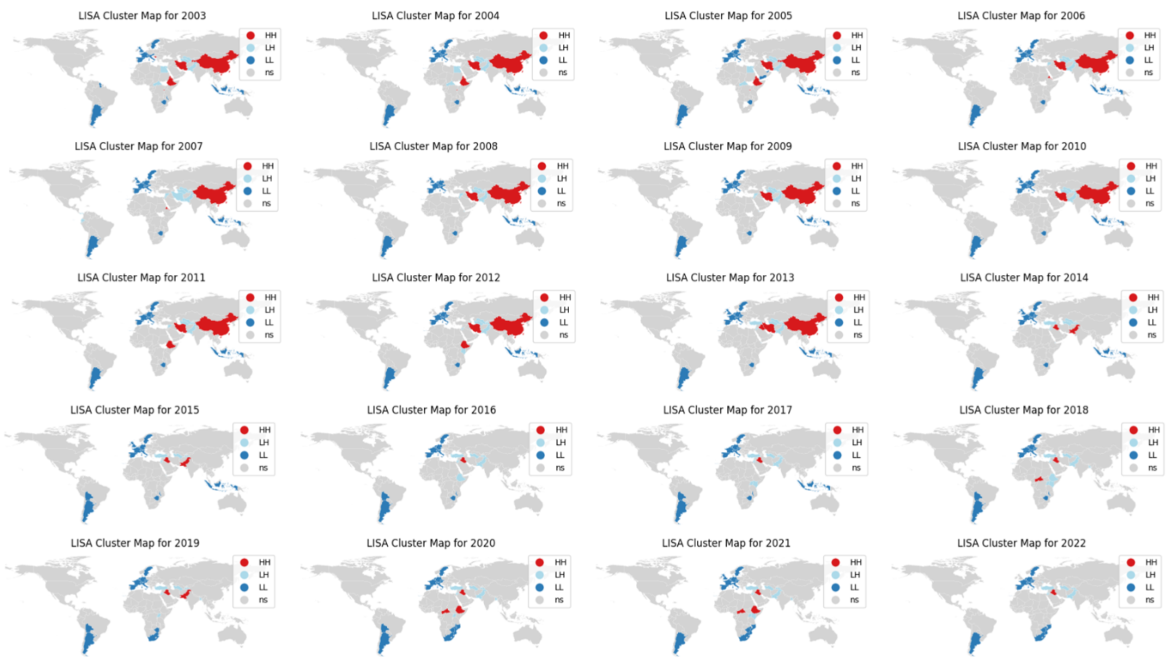


Figure 5.6: Refugees origin LISA Cluster Maps 2003-2022

What can be seen in the last 20 years from Figure 5.6, is that in Europe and the lower part of South America, as well as in Indonesia and parts of Southern Africa, clusters of countries are centered with low refugees fleeing from it.

In the Middle East, Central-East Africa, and in and around China (East/South Asia), clusters of countries are present with high refugees numbers fleeing from them. What potentially is the most interesting for this study, is that in some part of the Middle East and Central Africa, there are countries with relatively few refugees fleeing from it, while the cluster around it (i.e. neighbors around those countries) are countries with high number of refugees fleeing from it. So, in the map, red and light blue countries are the most interesting to study in the origin-destination comparison in the next chapter.

This situation stays the same over years, but in some years, also the North-West of South America has a situation in which one country (Ecuador) with few refugees fleeing from it, is surrounded by countries with high outflow of refugees. This also holds for Panama.

This map identifies regions of interest for this study regarding refugees fleeing from countries: Middle East, Central-East Asia, Central-East Africa and the Northern South-America. On purpose, ‘regions’ is written, because individual countries can still turn out to be interesting from a climate or refugee perspective, while the cluster is not significant. The gray countries are not in a cluster surrounded by significant high or low values, but maybe surrounded with sometimes high and sometimes low value neighbors. That could still be interesting. But this map only identifies entire regions of interest and shows refugee fleeing hotspots, coldspots and mixed spots (HL-LH) in the world.

5.6.1.2. Statistical cluster changes

To statistically establish whether the LISA values change over years, first is examined which correlation coefficient and statistical tests should be used. This discussion and selection can be found in Appendix C.

The Kendall correlation coefficient between 2003 and 2022 LISA cluster values is 0.42. The Wilcoxon Signed-Rank test indicates there is a significant different between the LISA 2003 and 2022 cluster data (W statistic = 3318.0, p-value = 6.33e-05).

The scatter plot (see Figure C.24) indicates that most values are aligned, but that there are also quite some shifts in the cluster data, especially around cluster values of approximately zero in 2003. This is shown further in Figure C.25. The countries that have a difference in LISA score of more than 0.5 or where the difference is greater than -0.5 are reflected upon. Interesting to see is that there are no differences (value LISA 2022 - value LISA 2003) greater than 0.5. This indicates that the relatively high changes are all more negative than -0.5, which means that the cluster score in 2003 was higher than in 2022. The latter holds for the following countries: Afghanistan (-0.894), Bosnia and Herzegovina (-1.168), Burundi (-1.329), Iran (-0.642), Lebanon (-0.799), Moldova (-0.669), Syria (-0.908), Ukraine (-1.262). A high positive value signals a local spatial clustering of similar values, whilst a high negative value signals a local spatial clustering of dissimilar values (Haining, 2001). This means that over the years, for these countries, they became part of a cluster (i.g. their neighbors) with more dissimilar values. So these countries move significantly from a more HH or LL cluster to HL or LH. So this means that in some part in the world, neighboring countries have more dissimilar values in those regions in the world than in 2003.

5.6.2. Changes in refugee origin numbers over years

In this section, the original, regular refugee numbers fleeing from a country are compared in 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The refugee fleeing away numbers in 2003 and 2022 have a Kendall correlation coefficient of 0.62. It is interesting that the difference in actual numbers of refugees at origin locations between 2003 and 2022 is not statistically significant based on the Wilcoxon Signed-Rank test (Test Statistic: 4335.5, p-value: 0.059). However, the p-value is 0.059, so at the 6% significance level, the difference would be significant.

Most countries have similar refugee fleeing values in both years, but of course some countries have very different values (see Figure C.27), due to conflicts, for example Ukraine. But overall, again, generic differences are only significant at the 6% level. Individual country differences are shown in Figure C.28).

There are countries with more than 100.000 extra refugees fleeing from it in 2022, compared to 2003. Those are: Afghanistan, Cameroon, Central African Republic, Democratic Republic of Congo, Eritrea, Mali, Myanmar, Nigeria, Rwanda, Somalia, Sudan, Syria, Ukraine and Venezuela. This are nearly all African countries.

There are also countries with over 100,000 fewer refugees fleeing from them in 2022 compared to 2003. These countries include: Angola, Azerbaijan, Bosnia and Herzegovina, Burundi, Croatia, Liberia, Serbia, Vietnam.

Based on the lists above, it again seems important to focus on Africa, Middle East and Asia as most interesting refugee movement regions.

5.7. Countries refugees flee to

5.7.1. Spatial clustering over years

5.7.1.1. Visual spatial clusters

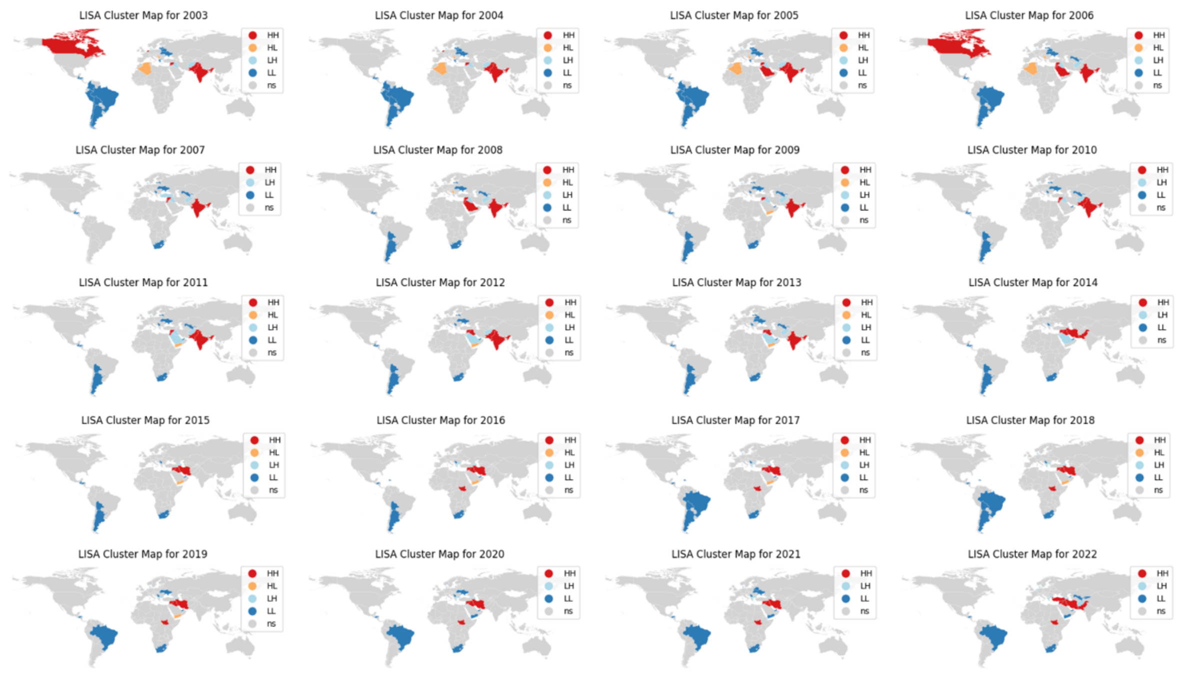


Figure 5.7: Refugees destination LISA Cluster Maps 2003-2022

In the clustering of countries where refugees flee to (host countries), interesting patterns can be observed (see Figure 5.7). Firstly, North America often contains countries that host large numbers of refugees. In contrast, almost all of South America has countries that receive very few refugees across the entire region. Additionally, in the far east and slightly north of Europe, there are clusters of countries that also receive few refugees. Additionally, there is a cluster of low values around Panama and Mexico, indicating that these countries receive very few refugees.

Clusters of high numbers of refugees fleeing to countries can be observed in the Middle East and Asia. This pattern is due to the high number of refugees that are also fleeing from countries within these regions. Specifically, LH (light blue) clusters are present in those regions, indicating countries with low refugee arrivals surrounded by countries with high refugee inflows. This pattern supports the previous observation.

In Africa, an interesting pattern emerges. Occasionally, a cluster forms where one country has a high number of refugees fleeing to it, while being surrounded by countries with low refugee inflows. This suggests that, in this region, refugees are fleeing from neighboring countries to this particular country.

Also in Northern Europe, there is a small LH cluster, which means that Denmark hosts few refugees, but countries around have high refugee receiving numbers, like Germany.

5.7.1.2. Statistical cluster changes

To statistically establish whether the LISA values change over years, first is examined which correlation coefficient and statistical tests should be used. This discussion and selection can be found in Appendix C.

The Kendall correlation coefficients between clusters refugees flee to is 0.11. This is quite low. However, the Wilcoxon Signed-Rank rest indicates no significant difference between 2003 and 2022 values (W statistic = 4280.0, p-value = 0.69). And indeed, overall there is not a structural pattern identified in the differences, most values align (see Figure C.30). Most differences are centered around zero, some are much higher, some much lower. The following countries are much more spatially aligned (HH or LL) with their neighbors in 2022 than in 2003: Czech Republic, Germany, Iran, Iraq, Poland and Turkey. The following countries are spatially more dissimilar than their neighbors (HL or LH), compared to 20 years ago: Armenia, Azerbaijan, Denmark, Georgia, Jordan, Pakistan, Russia, Saudi-Arabia and Syria.

5.7.2. Changes in refugee destination numbers over years

In this section, the original, regular refugee numbers fleeing to a country are compared in 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section. In Appendix C is reflected on the choice for the correlation coefficient and statistical test to compare the change.

The Kendall correlation coefficient between 2003 and 2022 refugee destination data is 0.34 and the Wilcoxon Signed-Rank test indicates a difference in destination location numbers between 2003 and 2022 (Test Statistic: 2251.5, p-value: 7.43e-07). A scatter plot between 2003 and 2022 values is shown in Figure C.32. Many countries have more refugees fleeing to them in 2022 than in 2003. The following countries are identified that have more than 100.000 refugees fleeing to it in 2022 than there were in 2003: Austria, Bangladesh, Belgium, Bulgaria, Cameroon, Chad, Czech Republic, Democratic Republic of Congo, Egypt, Ethiopia, France, Germany, Greece, Iran (Islamic Republic of), Iraq, Italy, Jordan, Kenya, Lebanon, Malaysia, Mauritania, Republic of Moldova, Niger, Pakistan, Poland, Romania, Russian Federation, Spain, Sudan, Sweden, Switzerland, Syrian Arab Republic, Turkey and Uganda. These countries have a huge influx of refugees. Far fewer countries have less than 100.000 refugees fleeing to it, compared between 2022 and 2003: Armenia, China, Guinea, Nepal, Saudi Arabia, Serbia, Tanzania and Zambia. So these countries receive fewer refugees. This information about differences can overall be seen in Figure C.33.

From these lists, it again seems important to focus on Africa, Middle East and Asia as most interesting refugee movement regions, apart from Europe. Europe is less interesting, because those countries can more easily cope with climatic circumstances/changes and also have very formal procedures to help refugees.

5.8. Summary

In this chapter, geographical clusters of (dis)similar countries regarding climate and refugee variables are identified across the globe. Also, changes over years of these clusters are examined. A comparison of cluster values with the original, non-cluster data is made. For every variable, a summary of the findings is presented below.

5.8.1. Climate clusters

In Central and (South) East Europe, drought intensity scores are low. Turkey also has low drought scores and is surrounded by other low-scoring countries. South America consistently shows high drought scores, as does Central Africa on a smaller scale. Russia has high drought scores, and is surrounded by low-scoring neighbors, while the U.S. has low drought scores, and is surrounded by high-scoring countries. Nepal forms the core of a high drought score cluster, with India, China, and Bangladesh therefore having a similar high drought exposure score. Africa and Asia show significant drought differences between neighboring countries, making refugee movements in these regions interesting. In South America and Europe, uniform drought scores make refugee destinations less relevant from a climate movement perspective: areas are then similarly exposed. Statistically significant changes in drought clusters occurred between 2003 and 2022, but practically, most changes are minor, with more negative than positive shifts. This means drought clustering of countries with similar values is generally slightly decreasing over time. Large clustering changes are only observed in Cambodia, the UAE, and Yemen. Cambodia and the UAE show increased clustering with neighboring countries of similar values, while Yemen shows more clustering with dissimilar drought scores.
However, no global statistical changes in original drought intensity scores were observed in the period 2003-2022, showing stability over the past 20 years.

Most regions show insignificant or random spatial riverine flooding clustering, with no structure, except for Asia, Africa, and the Middle East. In Asia, nearly all countries have high riverine flooding intensity,

except Nepal, Mongolia, and Kirghistan, which have low intensity but are surrounded by high-intensity countries. Asia is the hotspot for riverine flooding danger. In the Middle East, a small cluster of low riverine flooding intensity exists around Saudi Arabia and Oman, with neighboring countries also experiencing low intensity. In Africa, the Central African Republic forms the core of a high riverine flooding intensity cluster, while Mozambique and Somalia experience high exposure but are surrounded by low-intensity countries.
No statistically significant differences between 2003 and 2022 local riverine flooding clusters are identified. The same holds for the original, regular riverine flooding intensity values of countries, which have shown stability over the past 20 years.

In the Americas, only one significant cluster is in Northern South America, where high coastal flooding intensity is found. Western Europe primarily also forms a large hotspot of high coastal flooding exposure. Eastern Europe mostly has low exposure. In Africa, South Africa has low coastal flooding intensity, but clusters of high exposure are found around Angola, Mozambique, and Tanzania. Central Africa shows low exposure, extending to Nigeria and Niger, with Ghana being an exception with high intensity. In the Middle East, Oman and Saudi Arabia form the core of a high coastal flooding intensity cluster, while other countries have insignificant values. (South East) Asia is highly exposed, with only Uzbekistan as the core of a cluster with low intensity. Nepal, Laos, and Mongolia are low-exposure countries, surrounded by high-exposure neighbors. To conclude, refugee movements are most relevant in Central and East Africa, East Asia, and the Middle East due to varying coastal flooding intensities.
No statistically significant differences between 2003 and 2022 local coastal flooding clusters are identified. The same holds for the original, regular coastal flooding intensity values of countries, which have shown stability over the past 20 years.

Regarding sea level rise, there are many insignificant values, indicating random spatial clustering with no clear structure. Notable clusters include high sea level risk in northern South America and a persistent cluster around Indonesia. In Africa, small clusters of low sea level rise risk are in South Africa and (North) West Africa, with Tanzania and Malawi as a high-risk core surrounded by low-risk countries. In Asia, Laos is a safe haven with low sea level rise risk surrounded by high-risk areas. The Middle East generally shows low sea level rise risk, with Iran, Turkmenistan, and Uzbekistan forming a low-risk cluster. Europe has few significant clusters; Eastern Europe usually has low risk except for high-risk Ukraine, and Western Europe has a small high-risk cluster around Belgium.
No statistically significant differences between 2003 and 2022 local sea level rise clusters are identified. The same holds for the original, regular sea level rise intensity values of countries, which have shown stability over the past 20 years.

In North America, the United States and Mexico, consistently show high cyclone intensity, while Canada has low exposure. In (South) East Asia, most countries experience high cyclone intensity exposure, except for Nepal, Pakistan and Bhutan. More North and North West, Mongolia, Tajikistan, Kyrgyzstan and Kazakhstan experience low exposure but these countries are surrounded by high-intensity neighbors. In Europe, clusters of low cyclone intensity are found in West and South-East Europe. Africa often shows clusters of low cyclone intensity exposure in central and North Africa. South America occasionally has clusters of low cyclone exposure around Argentina and Bolivia. Overall, North America, particularly the United States, has high cyclone exposure. Asia is the most complex, with often intense cyclone clusters in the East and various exceptions elsewhere, making it the most interesting region for studying cyclone intensities and refugee movements.
Statistically significant changes in cluster values were observed between 2003 and 2022. The United States and Mexico experienced a substantial increase in alignment with their neighboring countries, while other changes were small. However, the original, actual cyclone intensity numbers did not change significantly over this period.

5.8.2. Refugee clusters

Over the last 20 years, clusters of countries with low refugee outflows are seen in Europe, lower South America, Indonesia, and parts of Southern Africa. In contrast, clusters with high refugee numbers fleeing away are found in the Middle East, Central-East Africa, and around East/South Asia. Notably, some countries in the Middle East and Central Africa have few refugees fleeing from it while their neighbors

have high outflows.
This pattern has remained consistent over the years, with occasional exceptions in North-West South America, such as Ecuador and Panama, which have low refugee outflows surrounded by countries with high refugee outflow numbers. Key regions of interest for studying refugees fleeing from countries include the Middle East, Central-East and South-East Asia, Central-East Africa, and Northern South America.
Analysis of cluster data from 2003 and 2022 shows statistical significant changes, with 2022 clusters often being more dissimilar than in 2003. The countries Afghanistan, Bosnia and Herzegovina, Burundi, Iran, Lebanon, Moldova, Syria, and Ukraine have shifted from similar (High surrounded by High) to dissimilar (High-Low or Low-High) refugee fleeing clusters, indicating greater dissimilarity over years in refugee outflow values compared to their neighbors.
However, the change in the actual number of refugees fleeing between 2003 and 2022 is only statistically significant at the 6% level. Many countries have similar refugee fleeing values in both years, but some show notable increases or decreases. Countries with over 100,000 additional refugees fleeing away in 2022 compared to 2003 include Afghanistan, Cameroon, Central African Republic, Democratic Republic of Congo, Eritrea, Mali, Myanmar, Nigeria, Rwanda, Somalia, Sudan, Syria, Ukraine, and Venezuela. Conversely, countries with over 100,000 fewer refugees fleeing in 2022 compared to 2003 include Angola, Azerbaijan, Bosnia and Herzegovina, Burundi, Croatia, Liberia, Serbia, and Vietnam.

In analyzing refugee host country clusters, North America often hosts large numbers of refugees, while almost all South America countries and certain areas in Eastern and Northern Europe receive very few. Also, a cluster of low refugee inflow numbers is centered around Panama and Mexico.
High refugee inflows are clustered in the Middle East and Asia, with also clusters of countries with low arrivals surrounded by those with high inflows. Northern Europe also has a small Low-High cluster with Denmark as core receiving few refugees while neighboring countries, like Germany, have high inflows. In Africa, an opposite pattern appears in which a country forms the core of a cluster with high receive numbers, but is surrounded by low refugee inflow countries. There is also one similar type of cluster in the Middle East around Yemen.
There is no statistical significant difference in clustering patterns between 2003 and 2022. However, there is a statistically significant difference in actual refugee destination location numbers. In 2022, many countries had more than 100,000 additional refugees fleeing to them compared to 2003. These include Austria, Bangladesh, Belgium, Bulgaria, Cameroon, Chad, Czech Republic, Democratic Republic of Congo, Egypt, Ethiopia, France, Germany, Greece, Iran, Iraq, Italy, Jordan, Kenya, Lebanon, Malaysia, Mauritania, Moldova, Niger, Pakistan, Poland, Romania, Russia, Spain, Sudan, Sweden, Switzerland, Syria, Turkey, and Uganda. Conversely, countries with fewer than 100,000 refugees fleeing to them in 2022 compared to 2003 only include Armenia, China, Guinea, Nepal, Saudi Arabia, Serbia, Tanzania, and Zambia.

6
Results SQ5: Comparing origin and destination locations of refugees

In this chapter, geographical data is utilized to illustrate different climatic conditions for refugees in their origin and destination locations. Each section focuses on a specific climatic indicator, providing a detailed analysis of how different climatic events influence refugee origin and destination locations. Instead of relying on an overall score that combines various climatic exposures -an approach that can obscure the underlying reasons for differences and potentially lead to incorrect conclusions— individual indicators are analyzed separately. For instance, a general score might suggest a country faces low climatic exposure, which could mask significant individual threats like e.g. severe droughts, for which is ‘compensated’ by other, low climatic exposures. By comparing individual indicators (visually), it becomes possible to accurately identify which countries are exposed to specific climatic exposures, thereby gaining crucial information that might be lost in aggregated data. This detailed analysis of individual indicators allows for a more precise assessment of total climate exposure for each location, by combine insights from different individual climatic exposures to mark the total climate exposure.

The subsequent sections first present world maps displaying climatic indicators alongside refugee numbers over time, showing either those fleeing to or from a country. Thereafter, the focus shifts to the 25 countries with the highest numbers of refugees, comparing their specific climatic exposures.

In the worldwide bivariate maps, a high number of refugees is defined as at least 100,000 fleeing to (destination) or from (origin) a country, though these ‘high numbers’ often exceeds 500,000 or even reaches into the millions. A low number of refugees indicates fewer than 100,000 refugees, or sometimes even fewer than 10,000.

In the top 25 plots, the countries with the highest numbers of refugees fleeing from it, have a minimum of approximately 60,000 people fleeing and a maximum of around 6.5 million, depending on the year. For refugee destination countries, the top 25 countries experience a minimum of roughly 100,000 refugees arriving and a maximum of around 3.5 million, also depending on the year. The significant differences among these top 25 countries are immediately apparent. Of course, the ranges per individual year are smaller. There is a huge increase of the number of people fleeing over years. For example, 6.5 million people fleeing from one country only happened in 2022.

6.1. Droughts

Drought scores refer to drought intensity values, where higher scores indicate greater intensity. The scores originate from the World Risk Index (see Chapter 2), range from 0.01 to 70.58 and are relative measures. A score of 10 is low, ‘good’, and means little drought intensity exposure, while 60 is ‘bad’ and means high drought intensity exposure. Below, in the plots, more red colors indicate (very) high drought intensity, whereas green means (very) little drought intensity.

6.1.1. Droughts: Refugees origin

6.1.1.1. Drought intensity: Refugees fleeing from countries worldwide

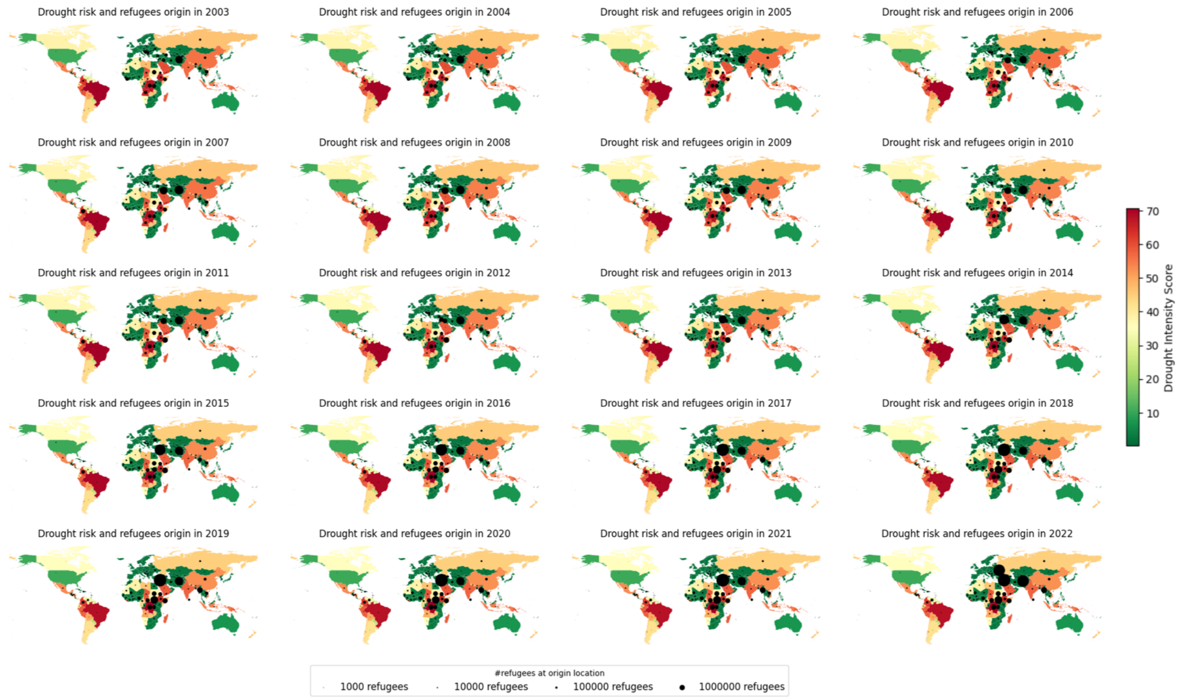


Figure 6.1: Bivariate plots of global drought scores and refugees fleeing from a country over years

It is evident that drought scores have remained relatively stable over the past 20 years (see Figure 6.1), indicating that drought intensity exposure has not significantly changed during this period. This stability suggests that there has not been a substantial increase in drought intensity due to climate change over the last two decades.

High drought exposure is evident in parts of Central and South America, with nearly every country scoring above a moderate threshold of 40 and some even reaching high scores around 60. This pattern is also observed in Central and East Africa, as well as in the southern part of the Middle East, and Southeast Asia. In Africa, Asia and the Middle East, there are also countries with (very) low drought exposure scores. This variability is particularly significant as it suggests that refugees fleeing from these areas might experience very different levels of drought exposure in their destination countries. Other regions in the world, generally report moderate or even very low drought exposure, with scores around 30 or less. Exceptions to this are Russia and New Zealand, which also display relatively high drought scores of around 50.

In contrast to drought intensity, the number of refugees fleeing from certain countries has fluctuated over the years, but consistently remains high in regions such as Central-East and West Africa, as well as the Middle East.

In the early 2000s, Balkan countries and Turkey also exhibited high numbers of refugees fleeing away. However, the number of refugees fleeing from the Balkans began to decline from 2006 onward and became minimal by 2015.

It is notable that very few refugees originate from South or Central America, as well as from Indonesia. An exception is Colombia, where refugee numbers increased significantly between 2003 and 2010, with a particularly sharp rise in 2007, followed by a decline starting in 2017. Additionally, from 2019 onwards, there has been a substantial increase in the number of refugees fleeing from Venezuela.

Throughout the period from 2003 to 2020, refugee numbers in Africa have remained consistently high. However, there has been a noticeable increase over time of refugees fleeing from African countries, with a significant rise occurring especially from 2016 onwards.

The Middle East consistently shows the highest numbers of refugees per country over time, with Afghanistan and Iraq particularly affected due to ongoing conflicts. Iran has also seen relatively high numbers of refugees fleeing from it over the years, as has Turkey. This pattern remains stable until 2012, after which refugee numbers from Syria surge dramatically. Starting in 2013, the entire Middle East region experiences a significant increase in refugees, with large numbers fleeing from Afghanistan, Syria, Iraq, and, to a lesser extent, Iran and Turkey.

For China, India, and Russia, the number of refugees fleeing from those countries has remained relatively constant over the years. In 2022, of course, Ukraine saw a significant spike in refugees fleeing away. Notably, this increase did not occur simultaneously in Russia, which did not experience a comparable rise in refugees fleeing away during the same period. It is possible that changes in Russia's refugee numbers might be reflected in data from 2023 and 2024.

This analysis highlights that particular attention should be directed toward the Middle East and Central, East, and West Africa, where the number of refugees fleeing away is significantly higher than in other regions. In these areas, various climatic exposures can severely impact already vulnerable people. The high concentration of refugees in these regions, coupled with their financial inability to cope with changing environments, underscores the critical importance of mapping out climatic exposure differences in the countries to which they flee. Many of the destination countries in those regions are also poor countries and have their own climatic exposure. By understanding the environmental changes that refugees encounter, NGOs and local governments can provide more targeted and effective assistance.

6.1.1.2. Drought intensity: Top 25 countries refugees flee from

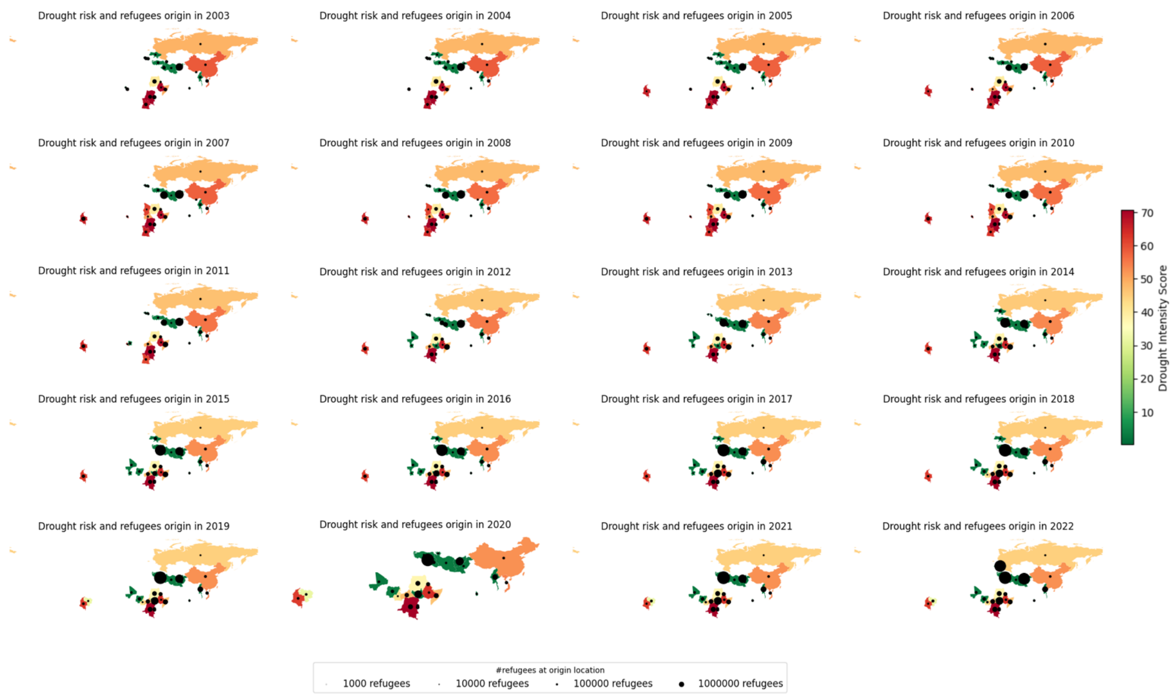


Figure 6.2: Bivariate plots of top hosting countries: drought scores and refugees fleeing from a country over years

Indeed, this more scoped map (see Figure 6.2) confirms two previous findings: 1) no visible changes in terms of drought intensity exposure for these selected countries. 2) (Central) Africa and the Middle East are prominent areas in terms of number of refugees fleeing away. Other insights are that China and Russia are always in the top 25 countries refugees flee from, while their numbers seem to be rather constant over time.

Among the smaller countries in this top 25, some notable examples extend beyond Africa and the Middle East. In addition to Colombia and Venezuela, which were previously highlighted, Sri Lanka, Myanmar, and Vietnam also stand out for their high numbers of refugees fleeing from these countries. Bhutan was also notable regarding this aspect from 2003 to 2011, although it no longer appears on the top 25 list after 2011. In West Africa, Sierra Leone and Liberia showed very high refugee numbers in 2003, with Liberia continuing to be a significant source of refugees fleeing away thereafter. This is especially remarkable when considering the relatively small population sizes of these countries.

In Eastern Europe, the Balkans experienced high refugee numbers primarily in the early 2000s. Croatia and Serbia also showed significant refugee fleeing figures until 2014. Ukraine has been a notable source of refugees in the early 2000s, and shows particularly high refugee fleeing numbers around 2015.

Drought intensity scores are notably high in Colombia, Central-East Africa, as well as in China and Russia to a lesser extent. However, in the top 25 countries, the Middle Eastern countries included do not exhibit high drought intensity scores. Among the top 25 countries included from East Asia, only some experience high levels of drought exposure.

The overall conclusion regarding the origin of refugees is that Africa and the Middle East are the primary regions of interest for study, with East Asia also being particularly significant for this research. Refugees from these regions often have limited resources to adapt to climatic events. This situation is also true for Colombia and Venezuela.

6.1.2. Droughts: Refugees destination

6.1.2.1. Drought intensity: Refugees fleeing to countries worldwide

The drought intensity scores obviously remain the same as in the origin analysis, because only the refugee situation changes in this analysis: refugee destination countries are examined (see Figure 6.3). In the global map, the drought scores remain the same.

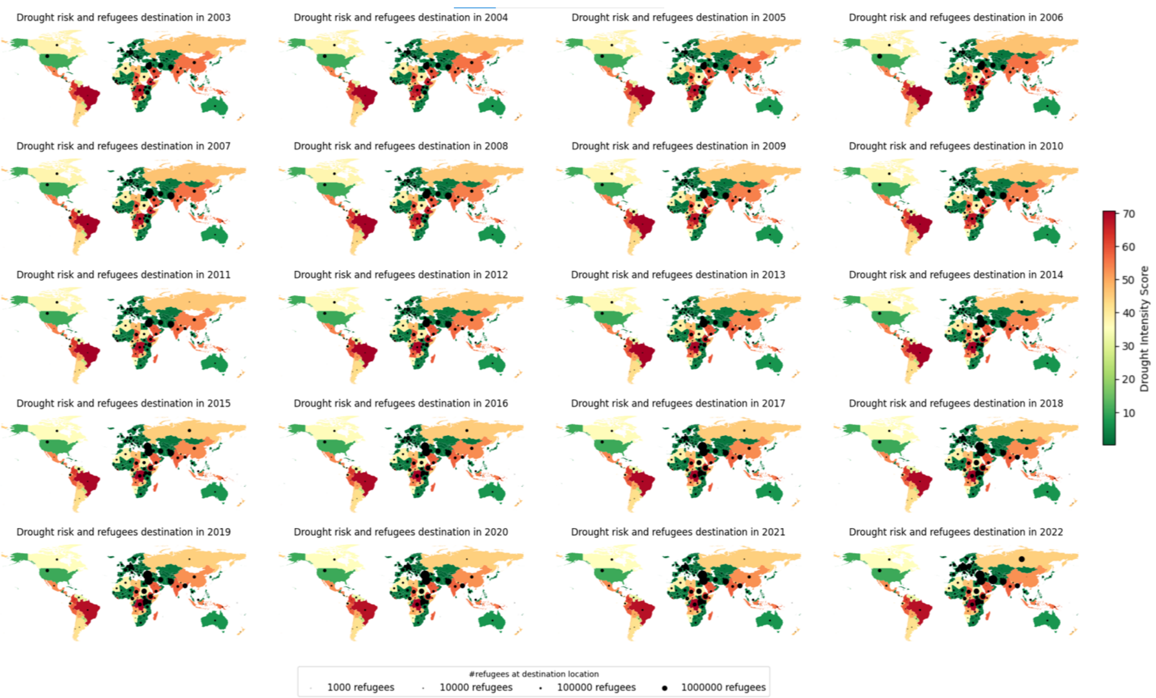


Figure 6.3: Bivariate plots of global drought scores and refugees fleeing to a country over years

There are relatively high numbers of refugees arriving in North America and Europe, as these regions are popular destinations for refugees. Australia is also a moderately popular destination. However, high refugee destination numbers are also found in neighboring countries to those from which refugees are fleeing. In some cases, countries are both major sources of refugees fleeing and significant destination locations for other refugees. In Africa and, especially, the Middle East, there are countries that also receive large numbers of refugees. This latter group is particularly relevant to this study, as refugees often move to neighboring countries within the same region, which are also frequently impoverished and face different climatic conditions. These countries typically offer less governmental support compared to Western Europe and North America.

In addition, Venezuela and Ecuador experienced a relatively high influx of refugees from 2007 onwards, although this increased flow began to diminish after 2015. Nepal has seen a consistently large refugee influx throughout the entire period (2003-2022).

Countries such as India and China also receive high numbers of refugees. Russia, on the other hand, has seen a modest refugee influx overall, with relatively higher numbers in 2014, 2015, and 2016, which then decreased afterward.

The refugee influx into India remains constant over the years. In Bangladesh, the influx increase significantly after 2009, with minimal figures before that year. The inflow continues to rise even more sharply after 2016. In Nepal, the refugee numbers are relatively high until 2009, but then decrease and remain stable after 2011.

6.1.2.2. Drought intensity: Top 25 countries refugees flee to

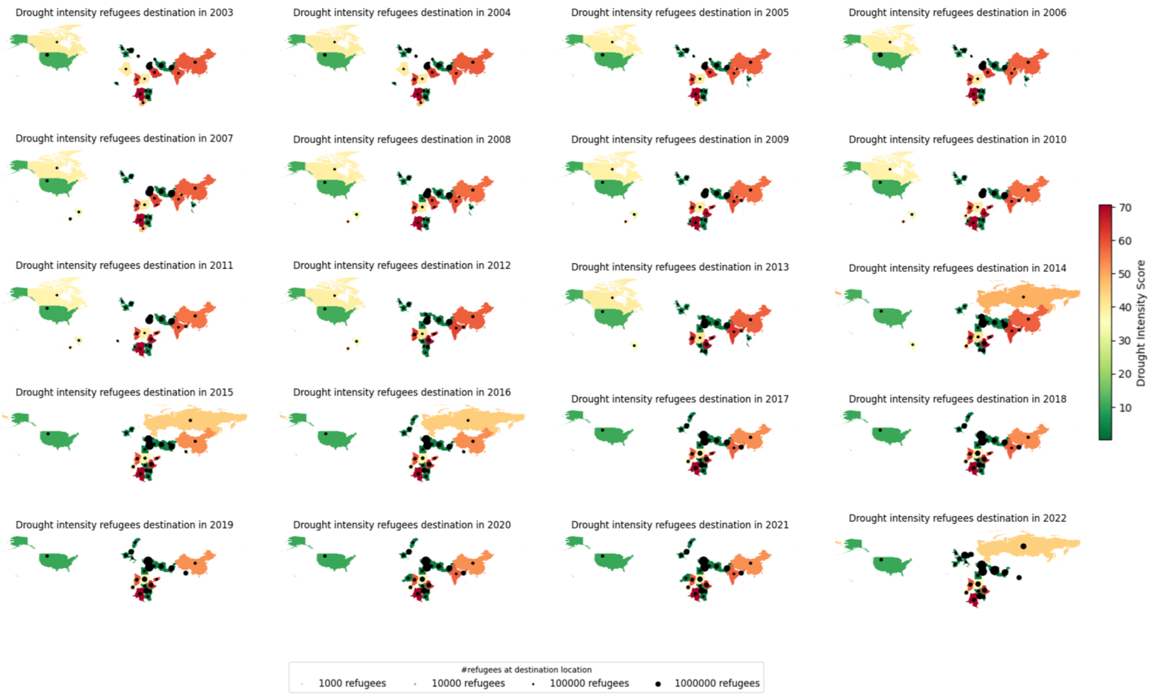


Figure 6.4: Bivariate plots of top hosting countries: drought scores and refugees fleeing to a country over years

From these plots (see Figure 6.4), it can be concluded that North America, particularly the USA, is consistently among the top 25 refugee-receiving countries in the world. Western Europe, especially France, Germany, and Great Britain, also consistently ranks among the top. Additionally, India and Russia occasionally appear in the top 25. China has consistently been among the top 25 refugee-receiving countries over the past 20 years.

Regarding Africa and the Middle East, many countries in these regions have consistently appeared on the list of top refugee-receiving countries over the last 20 years. However, in the Middle East, drought intensity is relatively uniform across countries, with scores generally indicating low drought intensity (around 10). Thus, from a climatic perspective, the Middle East is less critical in this context. In contrast, Africa exhibits significant variation in drought intensity scores between countries. Some African countries experience severe exposure to droughts, while others do not. This variability can occur over relatively short distances, particularly within Central Africa. This discrepancy makes it especially relevant to examine in the next subsection, where the focus will be on comparing the drought conditions of origin and destination countries, to understand how drought exposure differs for refugees fleeing to neighboring African countries.

Asia also hosts a large number of refugees. In West Asia, drought exposure is generally very low, whereas in East Asia, drought exposure is generally high.

Overall, the most prominent finding is that a significant amount of in-region fleeing occurs, particularly within Africa, the Middle East, and West Asia, and East Asia. These are relatively impoverished regions where refugees can be particularly more vulnerable to variations in drought conditions than in western countries. Therefore, the focus should be on these areas. To a lesser extent, the same consideration applies to Ecuador and Venezuela in South America.

6.1.3. Droughts: Comparing origin-destination of refugees

The previous sections and identifications lead to this section.

6.1.3.1. Droughts: Statistical generic difference tests

Due to the fact that some countries are both top origin and destination countries (for example, country X is both a country where many refugees flee from as well as a country where many refugees flee to), their drought scores cannot be compared directly to each other. That is because every statistical test to compare groups (parametric as well as non-parametric), for example an independent t-test or Mann-Whitney U test, requires that the 'observations' do not overlap between the groups. That means that one country either belongs to group A or group B, with A en B being the top 25 countries refugees flee from and flee to in this study.

Due to the overlap between the top origin and destination locations, it becomes difficult to conduct statistical tests to map out generic differences. However, two options remain possible (which themselves also consist of two options). These options are explained and statistical assumptions are tested in Appendix D. Here, only the results are presented.

First category of options

1) Testing differences of means

The median drought score of the means over years is 23.39 for origin locations, and 19.66 is the median score of the means over years for destinations. The results of the Mann-Whitney U test indicate differences over years between the groups, with significance at a high level degree of confidence (U-statistic: 344.0, p-value: 0.0001). So, the differences of the means in de results above are significant. Origin countries are in general, on average, more exposed to drought intensity than destination locations.

2) Testing differences of medians

The median drought score of all the medians over the years for origin locations is 3.20. The median score of all the medians over years for destination locations is 3.06. The result also indicates differences between the groups with high degree of confidence (U-statistic: 308.0, p-value: 0.0036). So, the difference in medians over years is significant.

The partly, generic conclusion is that the top 25 destination countries are in general less exposed to drought intensity than the top 25 origin countries, over years, comparing means as well as medians.

Second category of options

This section concludes per year on the differences between groups.

3) Testing differences of origin values with median destination values (see Figure 6.5)

4) Testing differences of destination values with median origin values (see Figure 6.6)

The results are inconclusive. When the origin drought values are tested against the median destination drought scores, 50% of the times (10 years), the null hypothesis is rejected. For the other ten years, the null hypothesis is not rejected. This means that in some years, the origin values differ from the median destination value. That is, in some years, the origin values do not come from the same 'population' as the destination median value. In the latest year, the median drought score from the top origin destination is higher than the median of the destination values.

However, when the origin medians are taken as constant over the years, the destination values in nearly all years seem to come from the same distribution, i.e. the null hypothesis is not rejected in almost every year. In this case, no differences can be detected.

	Year	Median Origin	Median Destination	Test-statistic	p-value	Conclusion
0	2003	3.03	2.98	102.0	0.107315	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
1	2004	3.03	2.98	109.0	0.156338	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
2	2005	3.03	2.86	90.0	0.051576	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
3	2006	3.25	2.86	76.0	0.018745	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	36.93	10.28	78.0	0.021908	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	36.71	3.06	74.0	0.015973	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	36.93	3.28	78.0	0.021908	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	37.33	3.27	78.0	0.021908	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	3.36	3.27	89.0	0.048262	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	3.06	3.07	107.5	0.148480	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
10	2013	3.09	3.07	107.0	0.140915	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
11	2014	3.14	3.09	93.5	0.106430	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
12	2015	3.17	3.07	101.0	0.101397	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
13	2016	3.17	3.06	100.0	0.095733	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
14	2017	3.17	3.06	101.0	0.101397	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
15	2018	3.18	3.07	95.0	0.070984	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
16	2019	3.21	3.00	64.0	0.014000	Reject the null hypothesis. The median is significantly different from the specified value.
17	2020	3.21	3.00	77.0	0.036985	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	3.21	3.00	64.0	0.013995	Reject the null hypothesis. The median is significantly different from the specified value.
19	2022	3.21	2.85	69.0	0.010511	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.5: Wilcoxon signed rank test: Differences of origin values

	Year	Median Origin	Median Destination	Test Statistic	p-value	Conclusion
0	2003	3.03	2.98	113.0	0.190814	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
1	2004	3.03	2.98	113.0	0.190814	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
2	2005	3.03	2.86	120.0	0.263476	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
3	2006	3.25	2.86	120.0	0.263476	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
4	2007	36.93	10.28	74.0	0.015973	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	36.71	3.06	62.0	0.005579	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	36.93	3.28	62.0	0.005579	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	37.33	3.27	52.0	0.005108	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	3.36	3.27	91.0	0.055070	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
9	2012	3.06	3.07	110.0	0.164496	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
10	2013	3.09	3.07	111.0	0.265132	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
11	2014	3.14	3.09	114.0	0.200216	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
12	2015	3.17	3.07	132.0	0.426142	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
13	2016	3.17	3.06	132.0	0.426142	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
14	2017	3.17	3.06	132.0	0.426142	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
15	2018	3.18	3.07	132.0	0.426142	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
16	2019	3.21	3.00	135.0	0.668235	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
17	2020	3.21	3.00	135.0	0.668235	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
18	2021	3.21	3.00	135.0	0.668235	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
19	2022	3.21	2.85	148.0	0.954431	Fail to reject the null hypothesis. The median is not significantly different from the specified value.

Figure 6.6: Wilcoxon signed rank test: Differences of destination values

In general, the statistical tests give inconclusive information. However, a part of the statistical information indicates that origin locations in general are more exposed to droughts than destination locations. The focus will be on the visual comparisons below to map out individual differences between countries, which in the end is what refugees experience.

6.1.3.2. Droughts: Visual individual comparisons

In this section, individual country level differences between top 25 origin and destination locations are compared side by side (see Figure 6.7). This is presented such that individual differences between top hosting and fleeing countries can be pointed out.

This comparison is presented for four years instead of twenty years, because: 1) that is easier to present and interpret; 2) over the years, in origin as well as destination locations, the top 25 refugee receiving/fleeing countries and drought scores do not change heavily. That are the reasons to summarize and skip years in between.

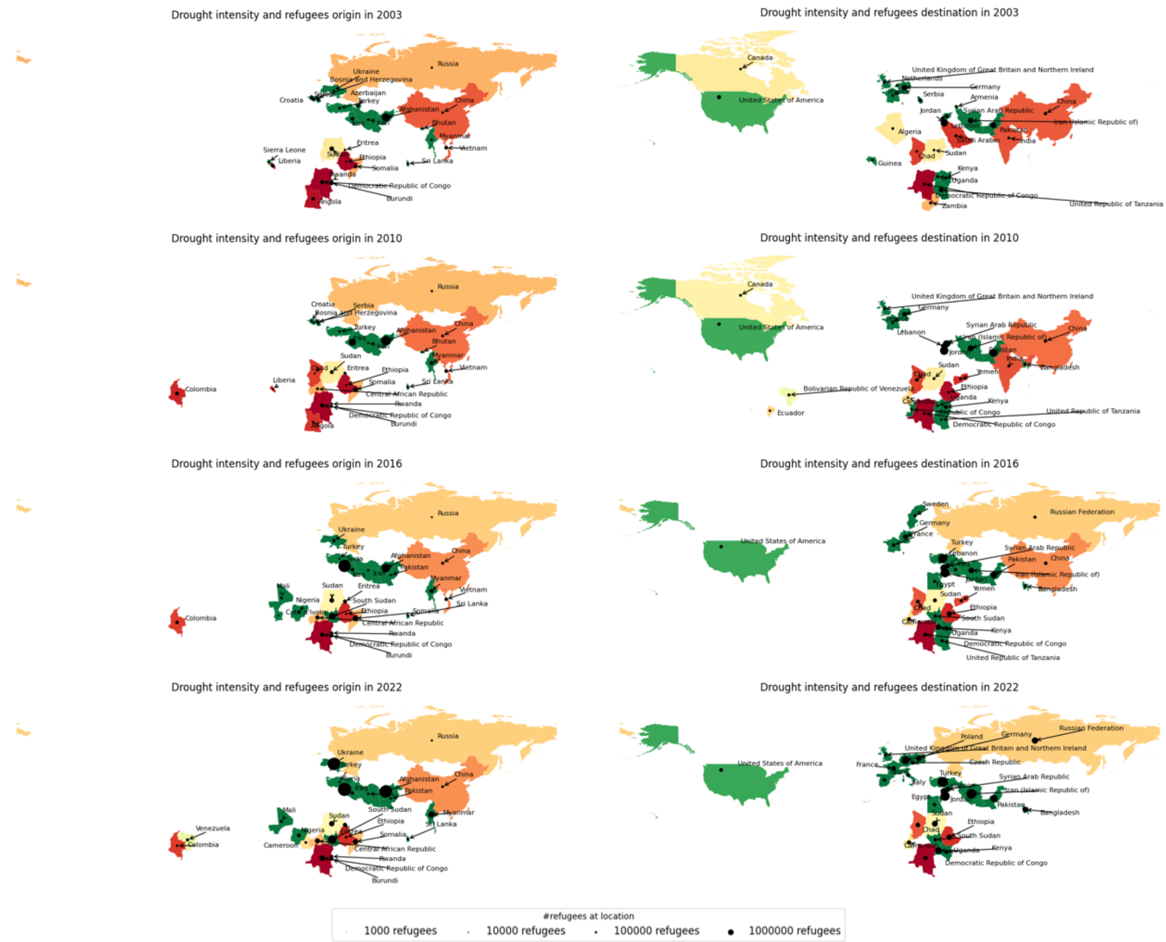


Figure 6.7: Comparing top origin and destination locations by drought scores

From earlier analyses it became apparent to specifically focus on Africa, the Middle East and Asia. To start with the latest years, what stands out in 2022, is that refugees who flee to top destination countries Congo, Ethiopia, Cameroon, Sudan and Chad arrive in high intense drought situations. It is interesting that at no other place, apart from Russia, drought intensity scores are high. The other way around is even true: in other receiving countries, drought intensity scores are low. So, if people flee to these countries from Mali, Nigeria or South-Sudan (which are the top origin countries), in which the drought intensity scores are low (good), then they should receive support regarding this change in drought exposure. They will namely end up in a more exposed place, from a drought perspective.

In 2016, there was a very similar situation. From the 2016 results, one can still learn that the observations above also holds for fleeing from Côte d'Ivoire: that country has a low drought intensity score and some neighboring countries have a high drought score. And refugees who flee to Yemen also need help to cope with the drought intensity, as it is a high refugee receiving country in 2016 and the drought score is very high.

In the Middle East, origin as well as destination locations score similar in terms of drought intensity. So, from a drought point of view, it does not matter where a refugee lives (same drought intensity). That also holds for Afghanistan, Bhutan, Sri Lanka, Myanmar and Bangladesh in Asia (all low drought intensity scores). However, Vietnam, India and China experience in this region all high drought intensity scores. From these countries, only Bangladesh, India and China are top destination countries. So, refugees fleeing to India and China (both in the top 25 destination list), end up in an at least evenly exposed, or worse, country in terms of drought intensity.

Earlier years more or less confirm this pattern, with some differences, but the emphasis in policy making should be on the countries described above. In this conclusion is also taken into account the very poor status of these continents in the world (e.g. Africa is the poorest continent in the world, which immediately also makes it vulnerable, because people there, generally, cannot adapt to climate change effects).

To conclude, Congo, Ethiopia, Cameroon, Sudan and Chad are countries with consistently high drought intensity scores and they are also top 25 destination countries. Refugees fleeing from other countries to these countries should be assisted in terms of drought management, but especially those who come from Mali, Nigeria or South-Sudan, as those will come from a far less intense drought situation, in general.

6.2. Sea level rise

In the bivariate plots presented below, the refugee numbers remain consistent with those shown earlier for the drought indicator. For each climatic indicator, the refugee numbers are identical, while the colors in the bivariate maps change to reflect the different indicators. The focus in this section is specifically on the sea level rise risk indicator. The dots representing refugee numbers are the same in every bivariate plot for each different indicator, with variations occurring only between origin and destination locations.

Sea level rise risk is scored between 0 and 80, with 10 indicating ‘good’, little sea level rise risk, and 70 representing ‘bad’, high sea level rise risk. Below, more red colors indicate (very) high sea level rise risk, whereas green means (very) little sea level rise risk.

6.2.1. Sea level rise: Refugees origin

6.2.1.1. Sea level rise: Refugees fleeing from countries worldwide

For refugee numbers, attention is focused on the dots, not the colors in the maps (see Figure 6.8). As observed earlier, the primary refugee hotspots are typically located in Central and East Africa, the Middle East, and to a somewhat lesser extent, in East Asia.

In South America, high refugee fleeing numbers are found primarily in Colombia and Venezuela, with Colombia being particularly prominent. In Asia, Sri Lanka, Vietnam, and Burma stand out, each with a million or more refugees fleeing from these countries. In contrast, Europe (apart from Ukraine in 2022), North America, and Oceania have very low refugee fleeing numbers, which is to be expected. Indonesia also has very few refugees fleeing from it (fewer than 100,000), which is remarkable given the country’s large size.

In South and North Africa, significantly fewer refugees are fleeing away, compared to Central and East Africa. West Africa also has a considerable number of refugees fleeing away, with numbers ranging between 100,000 and one million, but these numbers are still much lower than those in East and Central Africa.

In Eastern Europe, particularly in Croatia, Serbia, and Bosnia and Herzegovina, a significant number of refugees flee from these countries between 2003 and 2013. After that, these numbers declined, while refugee numbers in Ukraine began to rise slightly due to the Crimea conflict, and then more substantially from 2022 onwards due to the broader national war.

There are relatively small but significant numbers of refugees fleeing from Honduras, Haiti, Guatemala, and El Salvador, with figures around 50,000 or slightly less (approximately 30,000) for Haiti and Guatemala.

India has relatively low refugee fleeing numbers, around 20,000 or fewer over the years. China has higher numbers, approximately 200,000 in recent years. Sri Lanka, despite its small size, has around

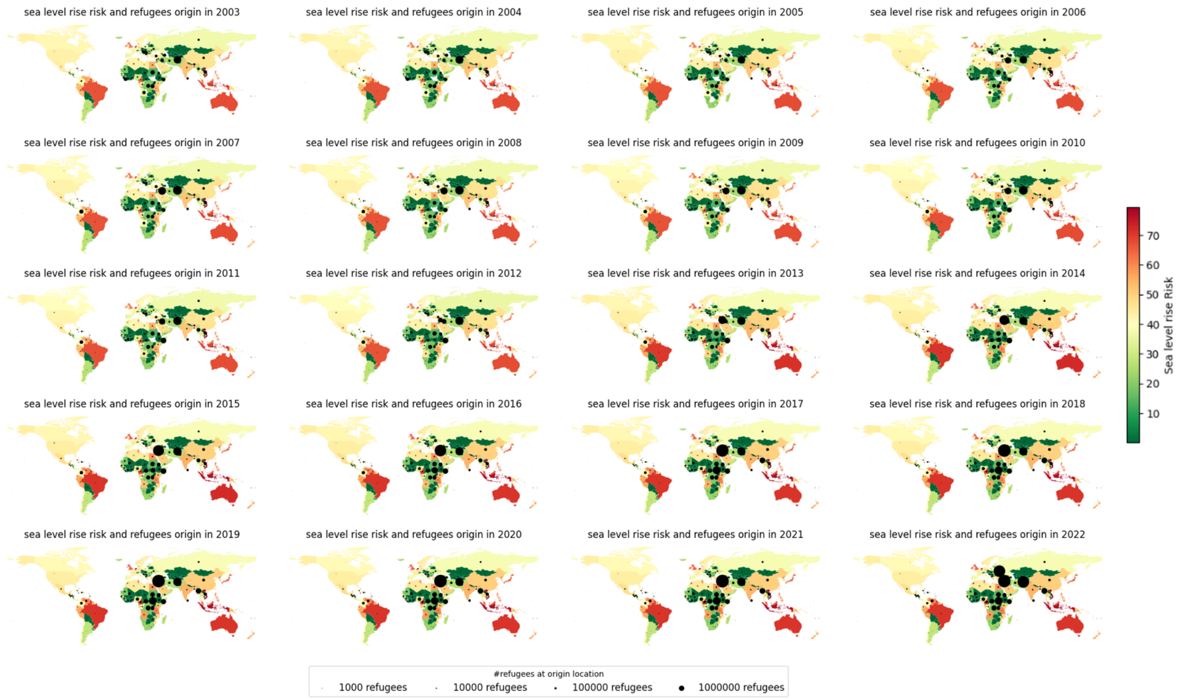


Figure 6.8: Bivariate plots of global sea level rise scores and refugees fleeing from a country over years

150,000 refugees fleeing from it, which is significant for such a small country. Russia had around 100,000 or more refugees in the early 2000s, but in the last nine years, the numbers have dropped to around 60,000 to 70,000. Colombia and especially Venezuela have seen high refugee numbers in recent years, with both countries reaching around 200,000.

Vietnam consistently had over 300,000 refugees fleeing from it in latest years. Myanmar (Burma) also consistently had very high numbers, with over a million refugees in recent years.

As indicated earlier, East and Central Africa (e.g. Sudan and South Sudan); the Middle East (e.g. Syria) and Asia (e.g. Afghanistan) experience significantly higher refugee fleeing numbers than other continents, reaching into the millions or at least hundreds of thousands. In the Middle East, particularly, the numbers can be around 5 million refugees in recent years.

In terms of sea level rise exposure, the focus is on the colors rather than the dots. The exposure indicator value ranges from 0 to around 70. Values below 10 represent very low risk, indicating minimal sea level rise, and are depicted in green on the map. Values above 55 signify high risk and are shown in red. Generally, more green indicates better conditions, while more red indicates worse conditions.

There are relatively few areas with high sea level rise exposure, with notable exceptions including Brazil, Australia, Indonesia, Vietnam, Cameroon, Japan, Great Britain, and the Netherlands. The sea level rise risk in Venezuela increases over the years and becomes significant after 2010. Additionally, risk levels are relatively high for Bangladesh, Malaysia, and the Philippines.

Above average sea level rise risk is observed in the following countries: Peru, Ecuador, Tunisia, Egypt, Angola, and Tanzania. To a somewhat lesser extent: India, Kenya, Mozambique, North Korea, and New Zealand also face moderate risk. Other countries with average sea level rise risk include the United States, Mexico, Colombia, some North African countries, as well as Turkey, Saudi Arabia, and China.

All other countries have little to very little sea level rise risk. Overall, while the risk numbers do seem to show a slight increase over the years, the intensity of this increase looks not very significant. This will be further analyzed and tested with statistical methods later in this section.

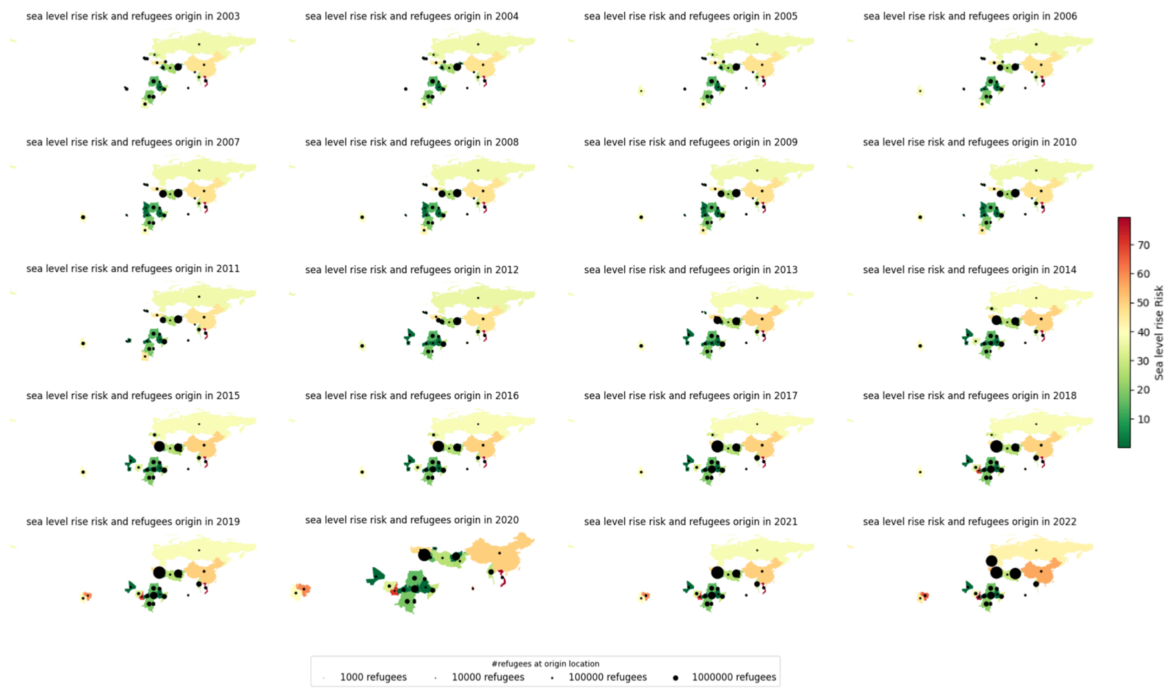


Figure 6.9: Bivariate plots of top refugee hosting countries: sea level rise scores and refugees fleeing from a country over years

6.2.1.2. Sea level rise: Top 25 countries refugees flee from

Again, refugee numbers (represented by dots) remain consistent across all indicators; only the colors change to reflect different aspects of climate exposure. High refugee fleeing numbers are observed in Central and East Africa, the Middle East, and (East) Asia, and later in Colombia and Venezuela (northern of South America). This pattern has already been reflected upon by the previous drought indicator and is discussed above in the worldwide analysis.

What is particularly interesting (see Figure 6.9) is that nearly all countries with high refugee fleeing numbers have low sea level rise risk, with notable exceptions being Vietnam, Sri Lanka, Cameroon, and Venezuela. China and Turkey have moderate sea level rise risk. The remaining of the top 25 countries refugees flee from, including those in Africa and other parts of the Middle East, generally have very low sea level rise risk.

6.2.2. Sea level rise: Refugees destination

6.2.2.1. Sea level rise: Refugees fleeing to countries worldwide

Compared to the origin case above, the drought numbers remain unchanged, so detailed discussion on these scores is not necessary here. The sea level rise risk is also consistent with the previous global plot in the origin case, however, the refugee numbers (dots) differ. These refugee numbers align then again with those seen in the drought destination case. Therefore, the focus is primarily on the combination of the two indicators, as that information provides new insights (see Figure 6.10).

In North America, there are moderate to high numbers of refugees arriving in the USA and Canada. Canada receives approximately 150,000 refugees over the years, while the USA receives around 400,000.

In 2006, Venezuela had 716 refugees arriving, and from 2007 to 2016, the number of refugees are around 200,000, an enormous increase. However, this number decreased significantly to just 30,000 by 2022. In contrast, Ecuador had around 10,000 refugees arriving until 2006. In 2007, this number surged to nearly 265,000 and remained high, exceeding 100,000, until 2021.

The three biggest countries in the world: China, Russia, India, have hundreds of thousands refugees receiving every year. However, Bangladesh stands out with nearly a million refugees a year fleeing to

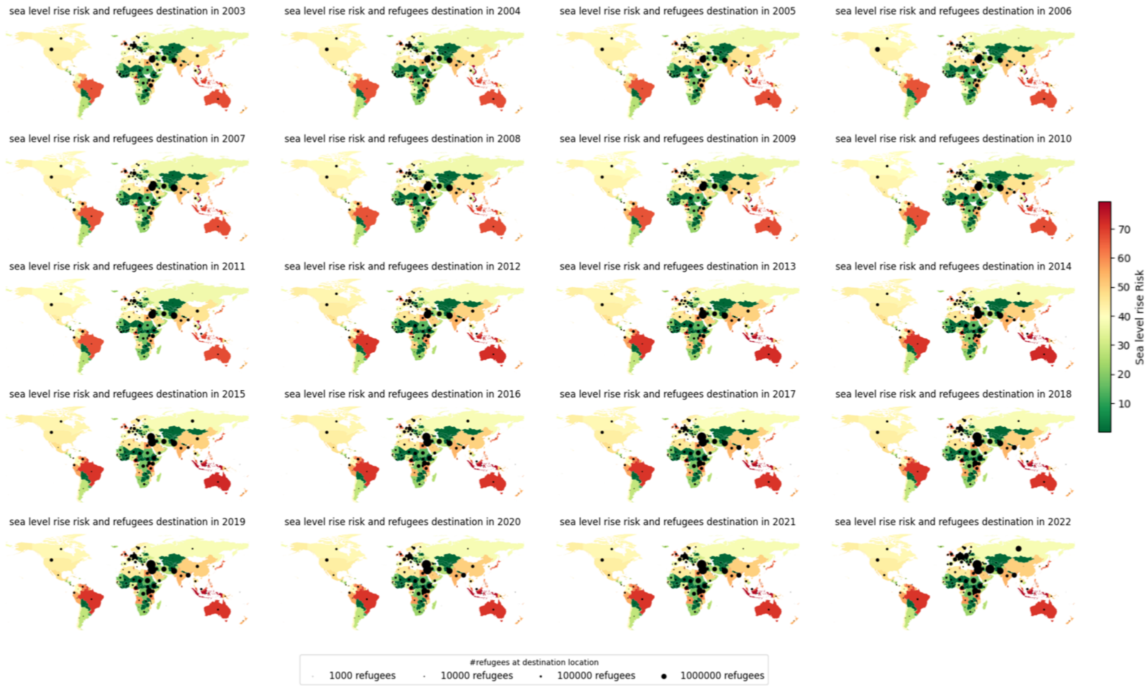


Figure 6.10: Bivariate plots of global sea level rise scores and refugees fleeing to a country over years

that country in the last couple of years. Russia receives over 1 million refugees only in 2022 due to war with Ukraine. Thailand has around 100,000 refugees fleeing to it yearly.

In the Middle East, refugee receiving numbers are extremely high. For example, Jordan has received around one million refugees annually in recent years. Iran also hosts a similar number of refugees, and Turkey receives approximately 3.6 million refugees in latest years. Lebanon similarly sees over one million refugees arriving annually in latest years. In the Middle East, refugee receiving numbers are the highest in the world. In Africa, while the numbers are also extremely high, they are somewhat lower compared to the Middle East. For instance, Sudan received approximately one million refugees in 2021, and Uganda has seen nearly 1.5 million refugees in recent years. Other countries in Central-East Africa also host hundreds of thousands of refugees. The numbers in these regions have increased over the years.

In Europe, Great Britain, France and Germany stand out with receiving a couple of hundreds thousands refugees every year, with Germany receiving the most: hitting even a million after the Syrian war. But also the Netherlands, Greece and Austria receive around 100,000 refugees. Sweden even receives a bit above 200,000 refugees. This gives an indication of what European countries do. Other regions host less refugees, around tens of thousands or even less.

From this destination map, it is clear that many refugees stay in the same region/on the same continent as where they came from. Also, some refugees flee to the 'rich' west (America, Canada or Europe) but definitely not the majority in terms of numbers. Many refugees stay in Africa, the Middle East or Asia, depending on where they flee from. And in those regions where refugees flee to, usually the sea level risk is not high.

6.2.2.2. Sea level rise: Top 25 countries refugees flee to

The refugee receiving numbers are of course the same as in the global maps above in the destination case (and destination maps of other indicators). The climatic situation is also the same as in the section above. The new information is the combination of the top 25 refugee receiving country selection and the sea level rise exposure data (see Figure 6.11).

As discussed in the section above, most refugees do not flee to areas with high sea level rise exposure, which is a positive finding. However, highly exposed destination countries are: Tanzania, Cameroon,

Figure 6.11: Bivariate plots of top refugee hosting countries: sea level rise scores and refugees fleeing to a country over years

and Venezuela. And to a lesser extent: India, Ecuador, Kenya, Saudi Arabia, Thailand, China, Congo, Egypt, Bangladesh, and Turkey also have notable sea level rise risks.

6.2.3. Sea level rise: Comparing origin-destination of refugees

The previous sections and identifications lead to this analysis.

6.2.3.1. Sea level rise: Statistical generic difference tests

Due to the fact that some countries are both top origin and destination countries (i.e., one country is a country where many refugees flee from as well as a country where many refugees flee to), the sea level rise scores cannot be compared directly to each other to test differences. That is because every test to compare groups (parametric as well as non-parametric), for example an independent t-test or Mann-Whitney U test, requires that the observations do not overlap (Leon, 1998). That means that one country should either belong to the group of the top 25 countries refugees flee from or flee to.

Therefore, it becomes difficult to do statistical tests to test generic differences between top 25 origin and destination countries. However, there are two options that are still available (which themselves also consist of two options). These options are explained and statistical assumptions are tested in Appendix D. Here, only the results are presented.

First category of options

1) Testing differences of means

The median of origin sea level rise score means over years is 21.91, while the median of the destinations means over years is 31.17. The conclusion of the Mann-Whitney U test indicate significant differences between the groups at a high level degree of confidence (U-statistic: 0.0, p-value: 6.79e-08). So, the difference of the means in de results above are significant. Destination means are higher than origin values, so one can conclude that overall, destination countries are on average more exposed to sea level risk, over years.

2) Testing differences of medians

The median of the origin medians over years is 18.52, while the median of the destination medians over years is 35.52. The conclusion of the Mann-Whitney U test is as follows. The results conclude differences also, and again with high degree of confidence (U-statistic: 0.0, p-value: 6.56e-08). So, those median differences are significant. Again, the same conclusion can be drawn: destination locations are in general, overall more exposed to sea level rise risk than origin locations.

The partly, generic conclusion is that the top 25 destination countries are in general more exposed to sea level rise risk than the top 25 origin countries, over years, comparing means as well as medians.

Second category of options
This section concludes per year on the differences between groups.

3) Testing differences of origin values with median destination values (see Figure 6.12).

	Year	Median Origin	Median Destination	Test-statistic	p-value	Conclusion
0	2003	18.52	34.25	63.0	0.006129	Reject the null hypothesis. The median is significantly different from the specified value.
1	2004	17.66	34.42	59.0	0.004175	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	18.52	34.09	67.0	0.008822	Reject the null hypothesis. The median is significantly different from the specified value.
3	2006	17.84	33.92	63.0	0.006129	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	17.66	40.62	35.0	0.000250	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	17.66	41.72	33.0	0.000188	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	17.65	40.02	37.0	0.000329	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	17.66	41.30	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	17.47	39.84	38.0	0.000376	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	17.47	41.29	31.0	0.000140	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	17.97	39.66	34.0	0.000217	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	23.47	34.60	64.0	0.006726	Reject the null hypothesis. The median is significantly different from the specified value.
12	2015	23.35	32.06	68.0	0.018624	Reject the null hypothesis. The median is significantly different from the specified value.
13	2016	23.47	34.77	65.0	0.007371	Reject the null hypothesis. The median is significantly different from the specified value.
14	2017	23.35	35.10	64.0	0.006726	Reject the null hypothesis. The median is significantly different from the specified value.
15	2018	25.04	35.27	80.0	0.025505	Reject the null hypothesis. The median is significantly different from the specified value.
16	2019	25.29	35.43	91.0	0.055070	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
17	2020	23.35	35.60	84.0	0.034174	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	25.29	35.60	91.0	0.055070	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
19	2022	25.29	35.60	71.0	0.012466	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.12: Wilcoxon signed rank test: Differences of origin values

4) Testing differences of destination values with median origin values (see Figure 6.13).

The results are inconclusive. When the median of the origin is taken as 'population value' and is compared with the destination values, the results differ heavily compared to when the situation is the other way around. Namely, in the 'origin as median' case, nearly all the hypotheses are rejected, indicating differences between the values. However, in the 'destination as median' case, all the hypotheses from 2014 onwards, cannot be rejected, indicating that the medians do not differ.

In this case, the statistical tests produce inconclusive information in general. Therefore, the focus is on the visual comparisons below. However, the majority of the statistical information indicates that destination locations in general are more exposed to sea level rise than origin locations.

6.2.3.2. Sea level rise: Visual individual comparisons

In this section, individual country level differences for top 25 origin and destination locations are analyzed side by side (see Figure 6.14). We do this so that we can map out individual differences between top hosting and fleeing countries. We only do this for four years instead of twenty years, because 1) that is easier to present and interpret and most importantly 2) over the years, in origin as well as destination locations, the top 25 refugee receiving/fleeing countries and sea level rise scores do not change heavily, so we can summarize and skip some years in between. (All results can be found in the provided GitHub link by this thesis in the file for the specific indicator, in this case sea level rise.)

	Year	Median Origin	Median Destination	Test Statistic	p-value	Conclusion
0	2003	18.52	34.25	56.0	0.007178	Reject the null hypothesis. The median is significantly different from the specified value.
1	2004	17.66	34.42	63.0	0.006129	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	18.52	34.09	60.0	0.009994	Reject the null hypothesis. The median is significantly different from the specified value.
3	2006	17.84	33.92	67.0	0.008822	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	17.66	40.62	46.0	0.001027	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	17.66	41.72	34.0	0.000217	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	17.65	40.02	46.0	0.001027	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	17.66	41.30	34.0	0.000217	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	17.47	39.84	41.0	0.000556	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	17.47	41.29	40.0	0.000489	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	17.97	39.66	48.0	0.001296	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	23.47	34.60	94.0	0.109236	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
12	2015	23.35	32.06	105.0	0.198084	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
13	2016	23.47	34.77	89.0	0.081203	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
14	2017	23.35	35.10	85.0	0.063156	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
15	2018	25.04	35.27	98.0	0.137154	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
16	2019	25.29	35.43	106.0	0.208470	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
17	2020	23.35	35.60	95.0	0.115711	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
18	2021	25.29	35.60	111.0	0.264669	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
19	2022	25.29	35.60	111.0	0.264669	Fail to reject the null hypothesis. The median is not significantly different from the specified value.

Figure 6.13: Wilcoxon signed rank test: Differences of destination values

What stands out in the years till 2020, is that nearly all origin locations do not have sea level rise risk, apart from Vietnam and Sri Lanka. All the other top 25 countries refugees flee from score relatively low on sea level rise risk. In recent years, Turkey, Colombia, Venezuela and China, as well as Cameroon, also have relatively high risk scores among the top 25 refugees origin locations, besides Vietnam and Sri Lanka. Also, in latest years, sea level risks in China, Turkey and Venezuela seem to increase.

In the top destination locations, in the early 2000s, Congo and India have relatively high scores for sea level rise risk. Algeria and Saudi Arabia have moderate risk. On purpose, the Netherlands and the United Kingdom are not mentioned, because these countries have great coping mechanisms to deal with sea level rise, and have formal procedures to help refugees. (The same holds in later years for Canada and the United States.) Around 2010, Ecuador, Congo and Kenya have high sea level rise risk scores, while Cameroon, Venezuela and Tanzania have very extremely high sea level rise risks. For India and China, the risks are also relatively high. In 2016, there is a similar situation, but Egypt enters the list of top refugee destination country, and Egypt also has a remarkable high sea level rise risk. The same holds for Uganda. Bangladesh has moderate sea level rise risk. In latest years, nothing changes in terms of top refugee destination countries with high sea level rise risk.

To conclude, the focus for this refugee and climatic situation should be on Africa. In this continent, relatively few destination countries have very high risk of sea level rise: Egypt, Tanzania, Kenya, Cameroon and Uganda. Refugees who flee to these countries from Mali, Sudan, South Sudan, Eritrea, Ethiopia, Central African Republic, Rwanda, Somalia, Nigeria and Burundi, which are all top countries refugees flee from, should be assisted, at least informed, regarding sea level rise risk. They come from another African country with low sea level rise risk, and end up in a country with very high sea level rise risk. Also refugees who flee from, for example Myanmar (top origin country) to Bangladesh (top destination country) should be informed and assisted, because they experience a similar situation. They should be prioritized as top refugee destination countries in risk prevention. In the Middle East, origin as well as destination locations do not have high risks of sea level rise. Only Turkey has slightly higher risk than the other countries in that region.

The emphasis in policy making should be on the countries described above. In this conclusion is also taken into account the very poor status of continents in the world (Africa, for example, is the poorest

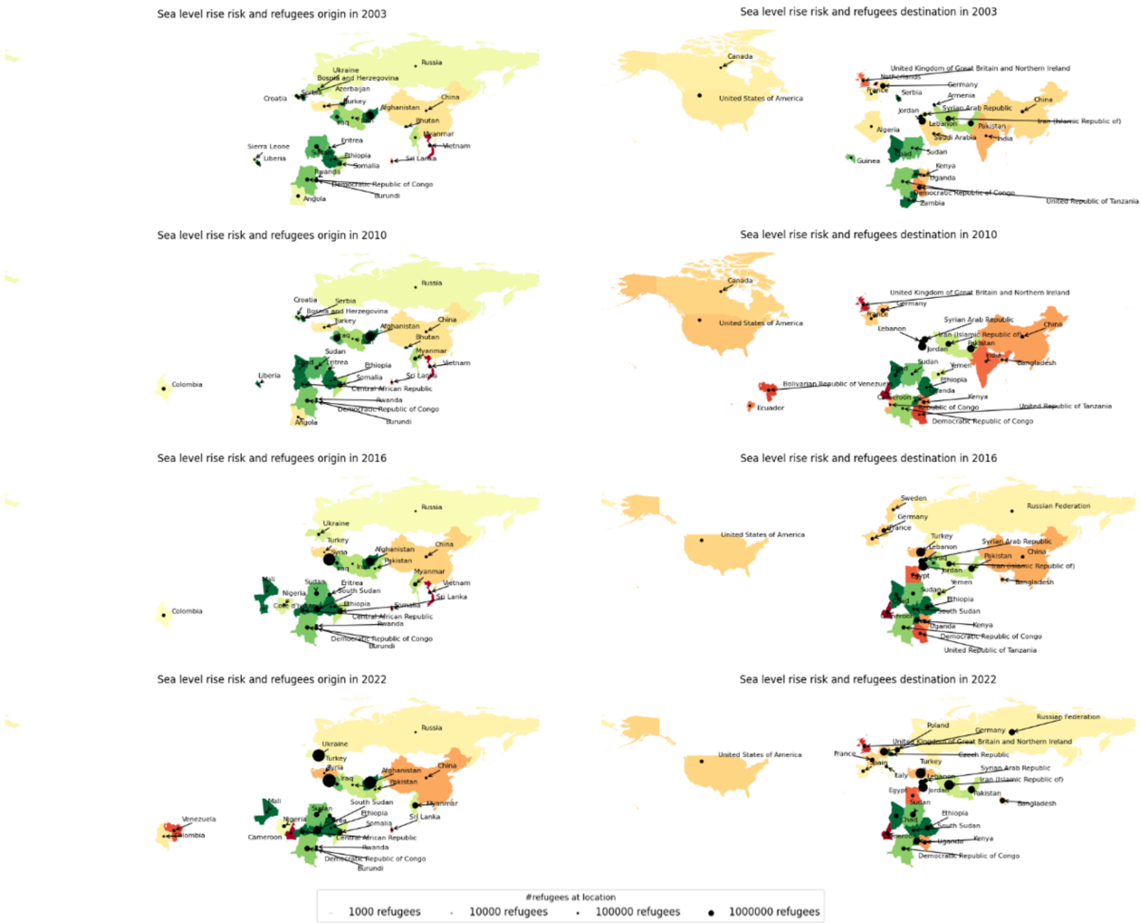


Figure 6.14: Comparing top origin and destination locations by sea level rise risk

continent in the world, which immediately also makes the refugees vulnerable, because people there, generally, cannot adapt to climate change effects).

6.3. Riverine floodings

In this section, riverine flooding intensities are examined together with refugee movements over years. Riverine flooding intensity is scored on a scale between 0 and 90, with 10 being low, ‘good’, very little riverine flooding intensity, and 80 being, ‘bad’, very high riverine flooding intensity. Below, more red colors indicate (very) high riverine flooding intensity, whereas green means (very) little riverine flooding intensity.

6.3.1. Riverine floodings: Refugees origin

6.3.1.1. Riverine floodings: Refugees fleeing from countries worldwide

As indicated earlier, the refugee numbers do not change over different climate analyses: in the bivariate plots (see Figure 6.15), only the colors are new, not the dots in the countries (indicating the number of refugees fleeing from a country in this case). The link between the ‘new’ climate indicator and the already ‘known’ refugee numbers provides new information. What again stands out, is that overall, the climatic situation (in this case riverine floodings) does not seems to change (heavily) over years. The latter is concluded due to the fact that the legend as well as the colors do not change heavily over the years (see Figure 6.15).

The riverine flooding risk in America is, in general, over the years, moderate. Also, in Europe, only the intensities in the Netherlands, Germany, France and Romania are relatively high. The rest of the

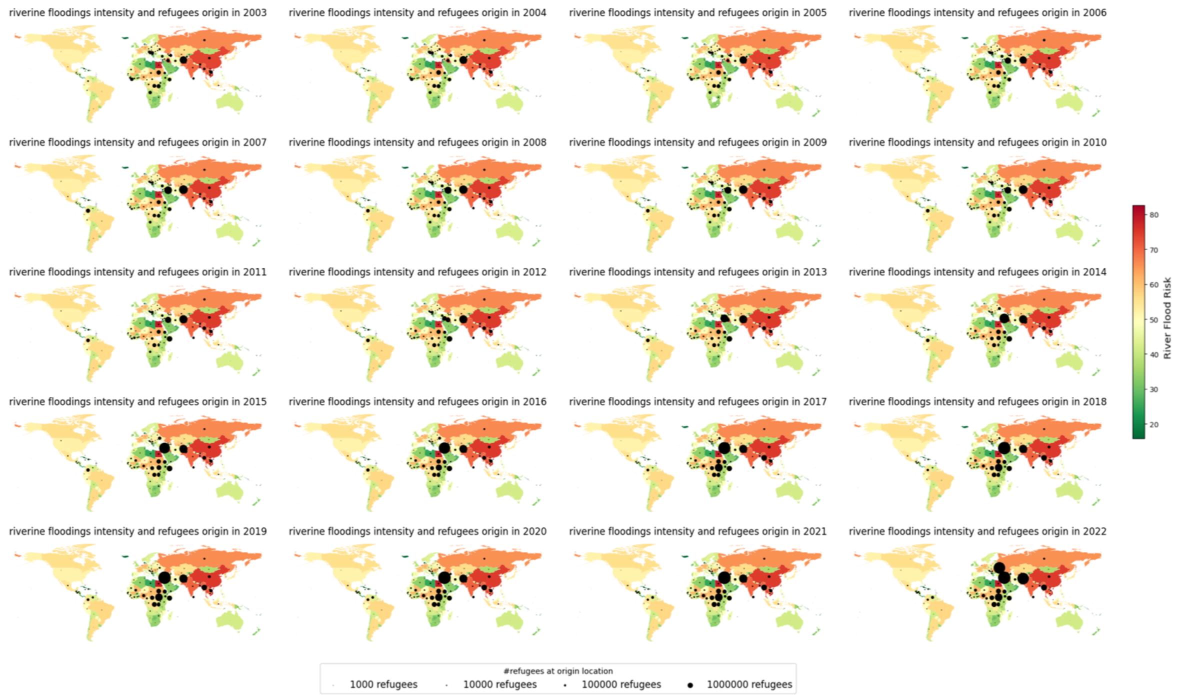


Figure 6.15: Bivariate plots of global riverine flooding intensity scores and refugees fleeing from a country over years

European countries have below moderate or little riverine flooding exposure. In Russia, the riverine flooding intensity is high.

In Africa, there is an interesting situation with countries outside the center of the continent (Central Africa). Those countries score below moderate to little on riverine flooding intensity, while the center countries have moderate to high exposure. Only Egypt is the exception to this. That country, by far, has the highest riverine flooding intensity, even taking into account the whole world. Also, Mali is an exception, as it is not in the 'center' of Africa, but the riverine flooding intensity is above moderate. In the Middle East, riverine flooding intensity is relatively low, except for Iran, there the intensity is high. In Oceania, there is little riverine flooding intensity.

In Asia, in general, the highest riverine floodings intensities can be found. Especially India, China, Pakistan, Bangladesh, Myanmar (Burma), Thailand, Vietnam have very high intensity scores (above 70). Other Asian countries -Kazakhstan, Turkmenistan and Uzbekistan- have moderate risk. Some countries have low riverine flooding intensity on this continent: Sri Lanka, Mongolia, Kyrgyzstan and Bhutan. In Indonesia, the riverine flooding intensity is always moderate.

For the refugee numbers, we look at the dots, and not the colors in Figure 6.15. Again, as seen earlier, the refugee hotspots usually are in Central and East Africa, the Middle East and to a somewhat lesser extent in (East) Asia. In South America, the only high refugee fleeing numbers are in Columbia and Venezuela. Apart from these countries, relatively little refugees flee from South or Mid America. Again, the only country in South America that increases heavily regarding refugee numbers between 2003 and 2010 is Colombia. In 2007, this country has a huge increase in refugee outflow, which decreases again from 2017 onwards. From 2019 onwards, the number of refugees fleeing from Venezuela increases heavily.

In Asia, especially Sri Lanka, Vietnam and Burma stand out, with a million or more refugees fleeing from those countries. In Europe (apart from Ukraine 2022), North-America and Oceania, we see of course very little refugee fleeing numbers. Indonesia also has relatively few refugees fleeing from it, less than 100,000. That is interesting, because it is an enormous big -and relatively poor- country. Especially in South, but also in North Africa, far fewer refugees are fleeing away compared to Central and East Africa. West Africa also has a relatively high number of refugees fleeing that region, between 100,000 and a million, but far fewer than in East and Central Africa.

In Eastern Europe (specifically Croatia, Serbia, Bosnia and Herzegovina), a significant number of refugees flee from these countries until 2013. Thereafter, these numbers shrink and the numbers in Ukraine increase, due to the Crimea conflict, and later, the more extensive national war in 2022. Moreover, we see some relatively small, but significant number of refugees fleeing from Honduras, Haiti, Guatemala and El Salvador with around 50,000 refugees or slightly fewer (Haiti and Guatemala around 30,000).

Also India has relatively few refugees fleeing from it, around 20,000 per year. China has higher refugee fleeing numbers, approximately 200,000 in the latest years. Sri Lanka has around 150,000 refugees fleeing from it per year, which is enormous for such a small country. Russia has numbers around and over 100,000 in the early 2000s, but in the latest 9 years around 60,000 or 70,000. Columbia and especially Venezuela have high refugee fleeing numbers in more latest years: both around 200,000. Vietnam always has over 300,000 refugees fleeing from it. And Myanmar (Burma) also always shows high numbers, with over a million refugees fleeing from it in the latest years. In early 2000s, Balkan countries and Turkey also have high refugees fleeing numbers. Refugees fleeing from the Balkan diminishes from 2006 onwards and is minimal from 2015 onwards.

As indicated earlier, in East/Central Africa (for example Sudan and South-Sudan) and the Middle East (for example Afghanistan and Syria), refugee numbers are much higher compared to the rest of the world, with over millions or hundreds of thousands fleeing from these regions. But especially in the Middle East, numbers can be around 5 million in latest years. In Africa, in this whole period 2003-2022, refugee fleeing numbers are constantly high. But also numbers in Africa increase over time and especially from 2016 onwards. However, the Middle East is definitely the area with the biggest numbers per country, over time. Especially Iraq and Afghanistan (Asia) have huge numbers due to the conflicts there. And Iran also has a relatively high number of refugees fleeing from it over the years. This latter also holds for Turkey. This pattern remains similar 2012. Then also, of course, the Syrian fleeing numbers increase heavily. And from 2013 onwards, this Middle East region explodes, with a lot of refugees fleeing from Afghanistan, Syria, Iraq and also to a lesser extent Iran and Turkey.

In combination (refugee fleeing numbers and riverine flooding risk), one can conclude that refugee fleeing numbers seem to increase over time. However, riverine flooding intensity seems to be relatively constant over the last 20 years. Additionally, only in East Asia, riverine flooding intensity is very high and in that region also relatively high number of refugees are fleeing away (from China, Thailand, Vietnam etc.). If these refugees flee to Nepal, West-Asia or Middle East, in terms of riverine flooding intensity, they will be in a safer place.

Iraq has high refugee numbers fleeing from it and also experiences high intensity of riverine floodings risk. Also, above moderate riverine intensity is present in Central Africa, with also high numbers of refugees fleeing from this region. In the Middle East, there is relatively low riverine flooding intensity, apart from Iraq.

So, from the three refugee fleeing hotspots: Central Africa, South East Asia and the Middle East, Asia is the region with high riverine flooding intensity. And Asian countries usually have difficulty to cope with changing environments. It is therefore extremely important to map out climatic exposure differences in the countries they flee to. This way, NGOs can better assist refugees in specific countries.

6.3.1.2. Riverine floodings: Top 25 countries refugees flee from

In addition to the previous global section, here stands out that countries in East Asia and Iraq experience high riverine flooding intensity from the top refugee fleeing countries (see Figure 6.16). Moderate risk is there in Sudan. For the rest of these top origin refugee countries, there is relatively little riverine flooding intensity. This holds for the time range 2003-2011. Thereafter, Mali enters the top 25 refugees fleeing list, which also has moderate riverine flooding intensity. Then in 2014, also Nigeria enters the top 25 list with moderate exposure. From 2015, also Pakistan is in the top refugees fleeing countries with high riverine flooding intensity. This pattern remains similar until 2022. The other top origin countries experience relatively low riverine flooding intensity.

Smaller countries in this top 25 are interesting, because many refugees flee from them, and some are outside of Africa and the Middle East. Columbia was already identified earlier, together with Venezuela. Additionally, Sri Lanka, Myanmar and Vietnam also stand out as countries with high refugee numbers

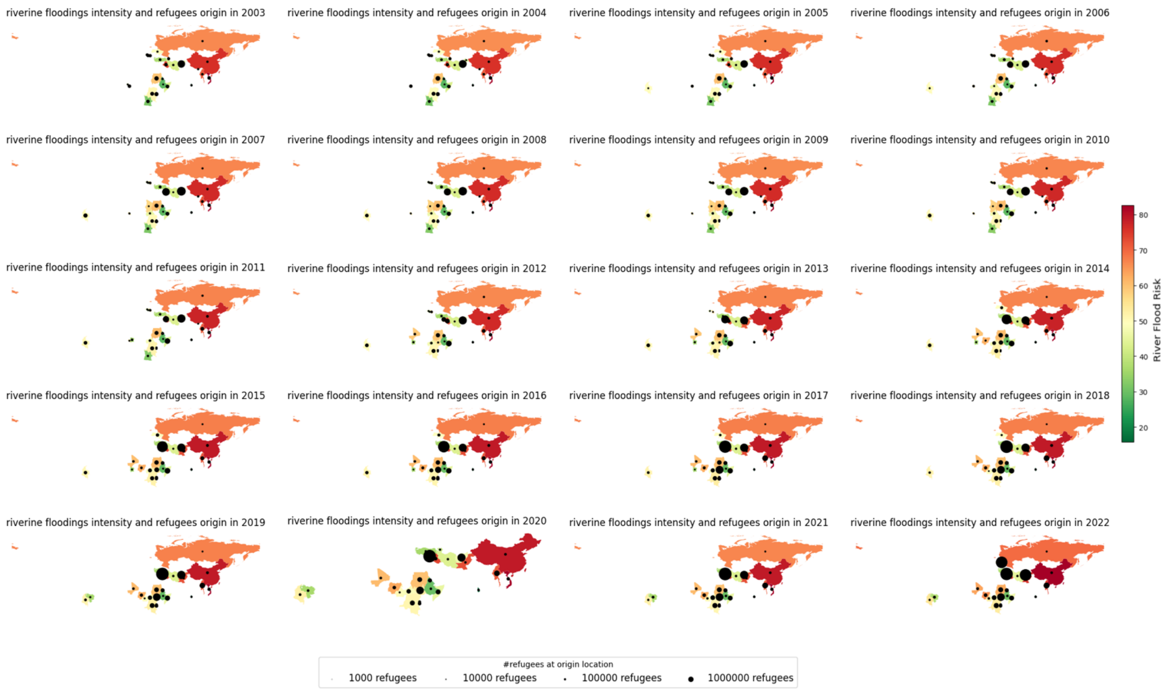


Figure 6.16: Bivariate plots of top refugee hosting countries: riverine flooding intensity scores and refugees fleeing from a country over years

fleeing from them. Also, in West-Africa, with in 2003 Sierra Leone and Liberia and thereafter only Liberia, there a very high refugees outflow numbers, especially if you take into account the population size of these countries. Other insights are that China and Russia consistently appear in the top 25 countries refugees flee from. Lastly, as noted earlier, the riverine floodings intensity scores do not seem to change heavily over years.

6.3.2. Riverine floodings: Refugees destination

Indeed, the climatic situation remains the same in this section, compared to the origin analysis above. Only the refugee numbers change, from origin to destination. The dots now reflect the number of refugees fleeing to a country. (However, this are again the same numbers as in the destination case for another climatic indicator.)

6.3.2.1. Riverine floodings: Refugees fleeing to countries worldwide

There is a moderate to high number of refugees coming to the USA and Canada (see Figure 6.17), with approximately 150,000 arriving in Canada and 400,000 in the USA, annually. In 2006, Venezuela only has 716 refugees fleeing to it, and from 2007 till 2016, this are around 200,000 refugees. Thereafter, the number shrinks again till 30,000 in 2022. Ecuador has around 10,000 refugees fleeing to it till 2006 and then in 2007, there are nearly 265.000 refugees and this number remain high (still above 100,000 in 2021).

The three biggest countries in the world: China, Russia, India, generally receive hundreds of thousands refugees every year (only Russia sometimes receives less than a hundred thousand refugees) and have a rather constant influx of refugees over the years. However, Bangladesh stands out, with nearly a million refugees a year fleeing to that country the last couple of years. The refugee receiving numbers of Bangladesh after 2009 increase heavily, before they are relatively small. After 2016, the numbers increase even more. Russia has over 1 million refugees fleeing to it in 2022 due to war with Ukraine. Thailand has around 100,000 refugees fleeing to it yearly. However, in the Middle East, we see the highest numbers. For example, around a million refugees fleeing to Jordan per year, in recent years. This also holds for Iran, and Turkey even hosts approximately 3.6 million refugees yearly. Lebanon also has over a million refugees fleeing to that country. In Africa, the numbers are also insanely high, but less compared to the Middle East. Sudan receives about a million refugees in 2021, while also Uganda in

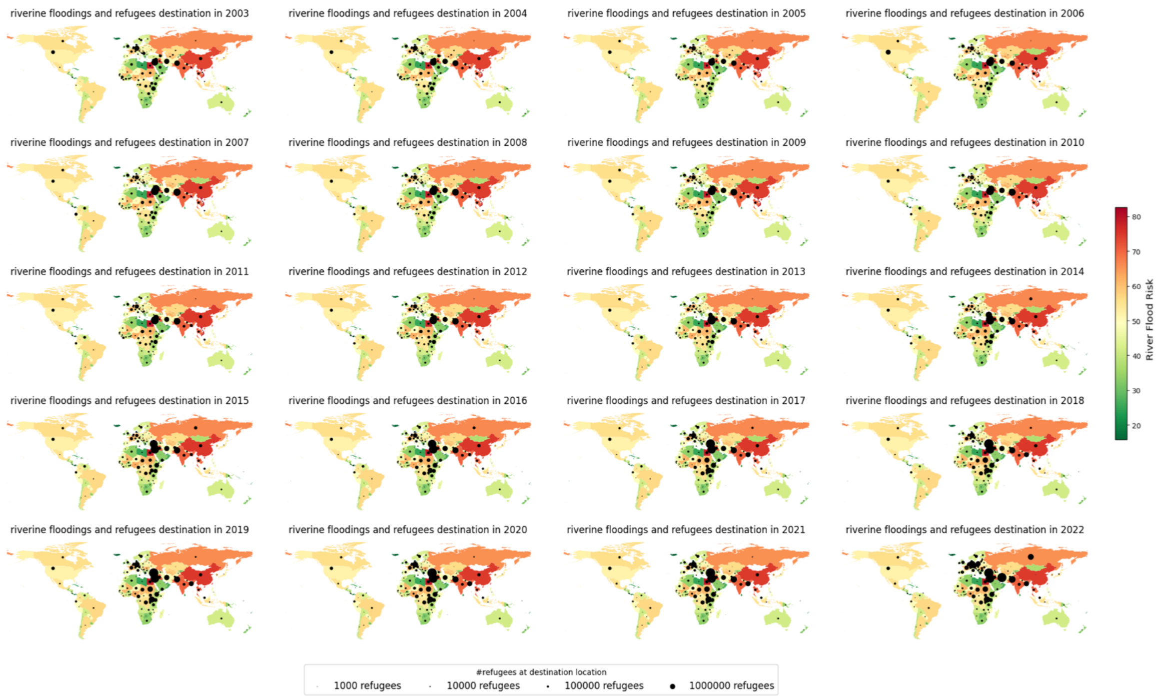


Figure 6.17: Bivariate plots of global riverine flooding intensity scores and refugees fleeing to a country over years

latest years receives nearly 1.5 million refugees. Other, Central East African countries receive hundreds of thousands refugees per year. But also in these regions, the numbers increase over the years. In Europe, Great Britain, France and Germany stand out with receiving a couple of hundred thousands refugees every year, with Germany hosting the most: hitting even a million after the Syrian war. But also the Netherlands, Greece and Austria receive around 100,000 refugees. Sweden receives more than 200,000 refugees. The refugee influx of Nepal is relatively high till 2009, but decreases after 2009 and remains constant after 2011. Other countries host fewer refugees, around tens of thousands or even less.

From this destination map (see Figure 6.17), it is clear that many refugees stay in the same region as where they came from. Also, some refugees flee to the ‘rich’ west (America, Canada or Europe) but definitely not the majority in terms of numbers. Many refugees stay in Africa, the Middle East or Asia, depending on where they flee from. And in the Middle East and Africa, riverine flooding intensity is usually not high, apart from Iraq and to some extent Sudan. Only in South East Asia, riverine flooding intensity is very high. However, relative little refugees (compared to other countries in the region) seem to flee to Iraq, which is good from a riverine flooding intensity point of view. In latest years, also many refugees flee to Sudan, which is not good from a riverine flooding point of view. The same holds for Bangladesh and Pakistan. But also China, India, Egypt and Thailand are relative popular destination locations of refugees with moderate to high riverine flooding risk. The latter countries are the ones to watch and support for refugees who arrive there should be prioritized. However, in most countries refugees flee to, the riverine flooding intensity usually is relatively low, apart from the exceptions above. Lastly, in general, refugee numbers increase over years. Riverine flooding intensities stay relatively similar over the years for countries.

6.3.2.2. Riverine floodings: Top 25 countries refugees flee to

What becomes clear from Figure 6.18, is that indeed China, India, Pakistan and Bangladesh are the only top receiving refugee countries with high riverine flooding intensity. Refugees fleeing to these countries should be prepared for riverine floodings. And to a lesser extent this also holds for Sudan and Chad. From 2012 onwards, Egypt is added to the list of very high intensity and huge refugee influx countries. From 2013 onwards, Iraq enters the list of top 25 refugee receiving country with also high riverine flooding intensity. Especially Pakistan and Bangladesh are important from the above selection, as these countries host the most refugees in latest years. Other top refugee receiving countries in less

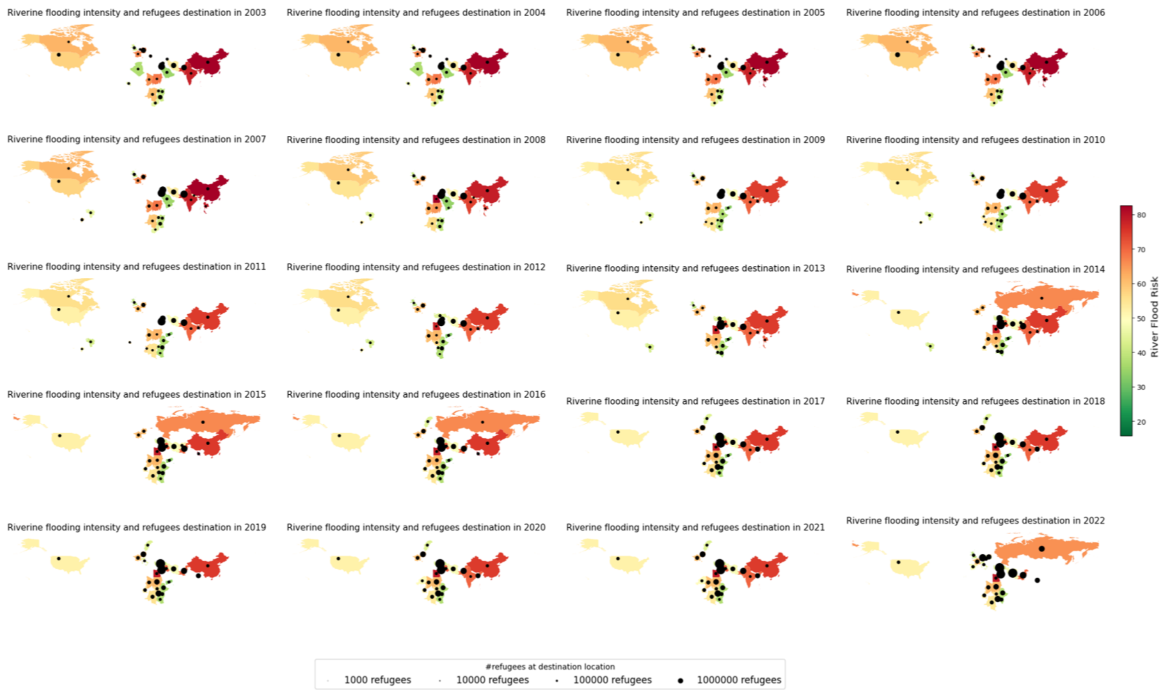


Figure 6.18: Bivariate plots of top refugee hosting countries: riverine flooding intensity scores and refugees fleeing from a country over years

economically developed areas (Africa, Asia, Middle East) have relatively low riverine flooding intensity.

6.3.3. Riverine floodings: Comparing origin-destination of refugees

The previous sections and identifications lead to this analysis.

6.3.3.1. Riverine floodings: Statistical generic difference tests

As one can see from the earlier top 25 countries refugees flee from/to analysis, there is overlap in top origin and destination locations. This prevents the possibility to directly conduct a statistical test to compare the top origin and destination countries. Namely, every statistical test to compare difference between two different groups, requires no overlap between those groups (for parametric as well as non-parametric test) (Leon, 1998). This is about an (alternative to the) independent or unpaired test, for which holds that one country should either belong to the group of the top 25 countries refugees flee from or flee to. But a paired test is also not a possibility, because the two groups are definitely also not totally the same (and it is not the case that you want to test differences of the same group at different time indices, you want to compare different groups, being top origin and destination countries).

Due to the above, it is difficult to do statistical tests to compare generic/overall differences between partly overlapping groups (top 25 origin and destination countries). However, there are still two options (which themselves also consist of two options). These options are explained and statistical assumptions are tested in Appendix D. Here, only the results are presented.

First category of options

1) Testing differences of means

The median of the origin means over years is 48.35, while the median of the destination means over years is 46.95. The conclusion of the Mann-Whitney U test is as follows: no significant difference between the groups (U-statistic: 250.0, p-value: 0.18). So, the difference of the means in de results above are not significant. Destination mean values are not statistically significant different than origin values. So, one can conclude that overall, destination countries are in equal exposed riverine flooding intensity areas.

2) Testing differences of medians

The median of the origin medians over years is 47.85, while the median of the destination medians over years is 48.75. The conclusion of the Mann-Whitney U test is as follows: no significant difference between the groups (U-statistic: 172.0, p-value: 0.46). So, those median differences are not significant. Again, the same conclusion can be drawn: origin and destination locations are in general, overall equally exposed to riverine flooding intensity.

The partly, generic conclusion is that top 25 origin and destination countries are in general equally exposed to riverine flooding intensity over years, comparing means and well as medians.

Second category of options
This section concludes per year on the differences between groups.

3) Testing differences of origin values with mean destination values (see Figure 6.19).

	Year	Mean Origin	Mean Destination	T-statistic	p-value	Conclusion
0	2003	46.2368	43.9108	0.751801	0.459480	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
1	2004	46.1448	43.8964	0.726051	0.474829	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
2	2005	46.4632	46.3024	0.813151	0.424125	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
3	2006	46.2888	46.2360	0.770151	0.448725	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
4	2007	47.0484	44.3812	0.966593	0.343388	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
5	2008	46.9892	46.4028	0.949590	0.351786	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
6	2009	47.1216	46.2068	0.992065	0.331063	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
7	2010	47.1136	46.4504	0.989250	0.332410	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
8	2011	46.1396	46.7028	0.733509	0.470353	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
9	2012	48.0476	47.1936	1.285754	0.210797	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
10	2013	49.6112	50.0600	1.687034	0.104554	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
11	2014	50.1464	50.2964	1.829158	0.079829	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
12	2015	50.0980	49.2588	1.821828	0.080968	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
13	2016	50.0688	48.7788	1.815899	0.081900	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
14	2017	50.1920	48.9996	1.851425	0.076452	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
15	2018	50.6780	48.9204	2.019397	0.054748	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
16	2019	50.1272	47.6328	1.845196	0.077384	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
17	2020	48.9112	49.7448	1.541744	0.136219	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
18	2021	50.0088	49.7448	1.827124	0.080144	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
19	2022	48.6612	46.6112	1.567217	0.130156	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.

Figure 6.19: T-test: Differences of origin values

4) Testing differences of destination values with mean origin values (see Figure 6.20).

The results are conclusive. Over all the years, whether the mean of the destination or origin group is taken as given and the other groups is tested against this mean, no generic statistical differences are found. This means that in general, top origin and destination locations are equally exposed to riverine flooding intensity. The focus will now be on the visual comparisons below to map out individual differences, which in the end is what a refugee experiences.

6.3.3.2. Riverine floodings: Visual individual comparisons

In this section, individual country level differences for top 25 origin and destination locations are compared side by side. This way, individual differences between top hosting and fleeing countries are mapped.

We only do this for four years instead of twenty years, because 1) that is easier to present and interpret, and mostly 2) over the years, in origin as well as destination locations, the top 25 refugee receiving/fleeing countries and riverine floodings scores do not change heavily, so we can summarize and skip some years in between.

	Year	Mean Origin	Mean Destination	Test Statistic	p-value	Conclusion
0	2003	46.2368	43.9108	-0.591138	0.559955	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
1	2004	46.1448	43.8964	-0.573127	0.571890	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
2	2005	46.4632	46.3024	-0.040224	0.968247	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
3	2006	46.2888	46.2360	-0.013182	0.989591	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
4	2007	47.0484	44.3812	-0.673116	0.507304	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
5	2008	46.9892	46.4028	-0.138708	0.890838	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
6	2009	47.1216	46.2068	-0.221020	0.826946	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
7	2010	47.1136	46.4504	-0.160243	0.874031	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
8	2011	46.1396	46.7028	0.135576	0.893287	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
9	2012	48.0476	47.1936	-0.193607	0.848113	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
10	2013	49.6112	50.0600	0.098168	0.922614	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
11	2014	50.1464	50.2964	0.033683	0.973409	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
12	2015	50.0980	49.2588	-0.190387	0.850607	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
13	2016	50.0688	48.7788	-0.290959	0.773582	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
14	2017	50.1920	48.9996	-0.266715	0.791967	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
15	2018	50.6780	48.9204	-0.393543	0.697393	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
16	2019	50.1272	47.6328	-0.569977	0.573991	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
17	2020	48.9112	49.7448	0.191381	0.849837	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
18	2021	50.0088	49.7448	-0.060610	0.952172	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.
19	2022	48.6612	46.6112	-0.511645	0.613575	Fail to reject the null hypothesis. The sample mean is not significantly different from the specified value.

Figure 6.20: T-test: Differences of destination values

Areas of interest are those where top origin and destination locations are within the same region and where the countries are economically less developed, in which coping with the impacts of climate change is difficult. This is the case in the regions Africa, Middle East and Asia.

Top refugee origin countries of interest in Africa are Angola, Rwanda, Burundi, Somalia, Ethiopia, Central African Republic, Sierra Leaone, Liberia, Cote d'Ivoire, Cameroon and Eritrea. Those are locations of interest as a result of two things: 1) the riverine flooding intensity is low, so refugees can become more exposed in their destination location, and 2) these countries are relatively close to areas where the riverine flooding intensity is worse, and also many refugees flee to. This are Egypt, Chad, Mali, Nigeria and Sudan. Congo, South Sudan and Cameroon have moderate exposure over the years and are top destination countries. So, refugees fleeing to those countries should be supported regarding riverine flooding intensity, especially if they come from the little exposed African countries mentioned before. Other countries in Africa where many refugees flee to (Zambia, Tanzania, Uganda, Kenya, Guinea, Yemen and Algeria) are also popular destination locations, but have relatively low riverine flooding intensity.

In Asia, top refugee fleeing countries are Afghanistan, China, Russia, Bhutan, Myanmar, Pakistan, Vietnam and Sri Lanka. All those countries, except for Afghanistan, Bhutan and Sri Lanka, are countries with very high riverine flooding intensity. However, the top destination countries in Asia: India, Bangladesh, China and Pakistan also have major riverine flooding intensity. This means that Asian refugees who flee in the region, are always in an equally bad or even worse situation, in terms of riverine flooding intensity. If Asian refugees flee to the Middle East, however, in nearly all the cases they are less exposed to riverine floodings.

The only country in the Middle East that has a lot of exposure to riverine floodings is Iraq. All the other countries in the Middle East have very low exposure to riverine flooding intensity. In this region, popular refugee destination locations are Turkey, Syria, Iran, Lebanon and Jordan. All those countries have low to moderate riverine flooding intensity. To conclude, refugees who flee from Iraq to other countries in this region, will be in a better place, in terms of riverine flooding intensity. Other refugee movements in this regions will not experience major differences in riverine flooding intensity. Iran and Afghanistan are top origin countries with low riverine flooding intensity, but are surrounded by top refugee receiving countries, Iraq and Pakistan, with very high riverine flooding intensity risks. This makes movement

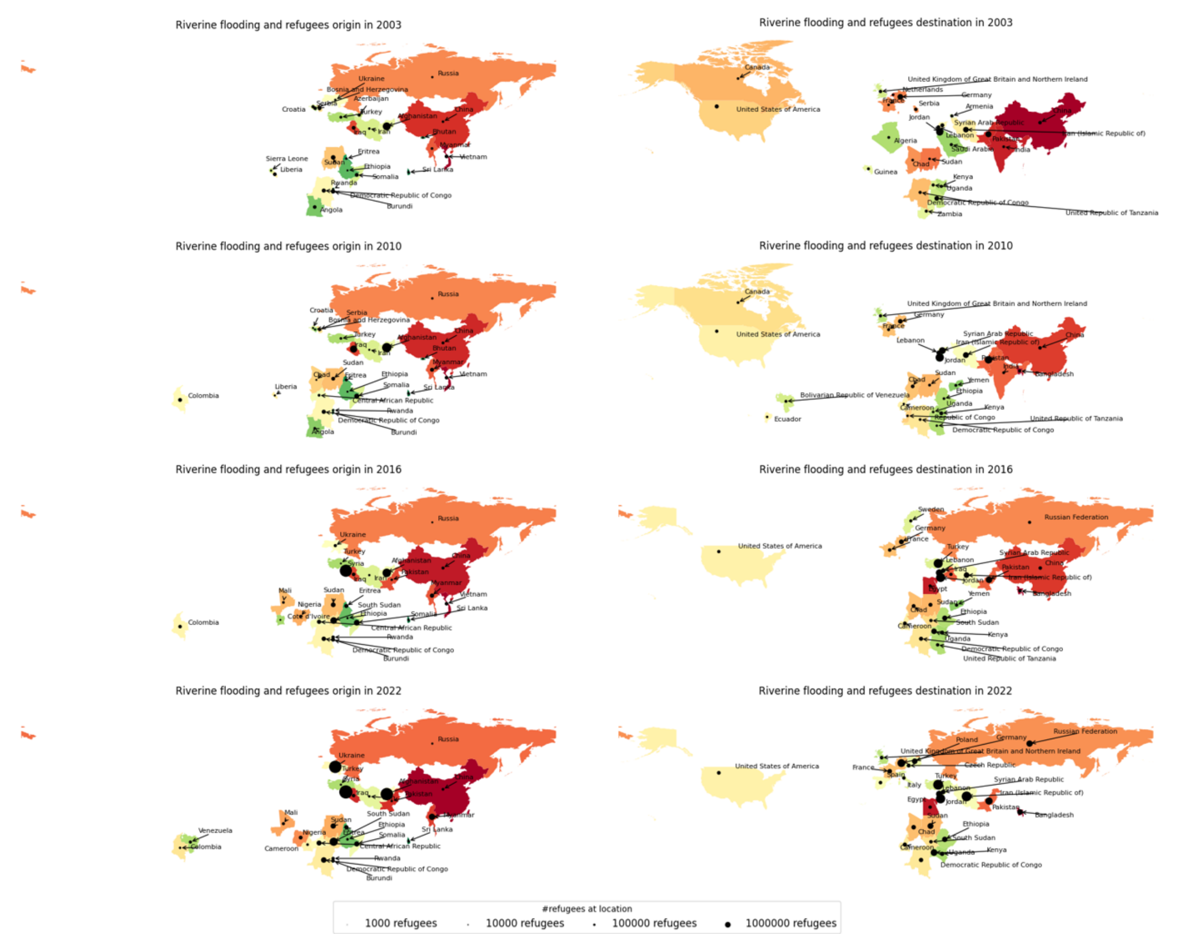


Figure 6.21: Comparing top origin and destination locations by riverine flooding scores

to these countries dangerous from a riverine flooding perspective, especially when fleeing from low intensity areas. Egypt and Jordan are two top destination countries located close to each other, with Egypt experiencing extremely high riverine flooding intensity and Jordan experiencing very low intensity. In terms of riverine flooding safety, refugees are better off in Jordan than in Egypt.

In northern South America, Colombia (and later Venezuela) is also on the list of countries from which many refugees flee. However, the top destination countries in this region include Ecuador and Venezuela, which have equal or less exposure to riverine flooding intensity. From this point of view, the movement does not have a huge impact on the live of refugees regarding riverine flooding intensity.

6.4. Coastal floodings

In this section, coastal floodings are examined in combination with refugee movements over years. Coastal flooding intensity is scored between 0 and 90, with 10 being low, ‘good’, very little coastal flooding intensity, and 80 being ‘bad’, very high coastal flooding intensity. Below, more red colors indicate (very) high coastal flooding intensity, whereas green means (very) little coastal flooding intensity.

6.4.1. Coastal floodings Refugees origin

6.4.1.1. Coastal floodings: Refugees fleeing from countries worldwide

There only are a couple of countries in the world where coastal flooding intensity is high/above moderate (see Figure 6.22). On the continent America, this are only Peru, Suriname and Guyana. The United States and Venezuela have moderate coastal flooding intensity. All the other countries on this continent are ‘green’, i.e. have low coastal flooding intensity numbers.

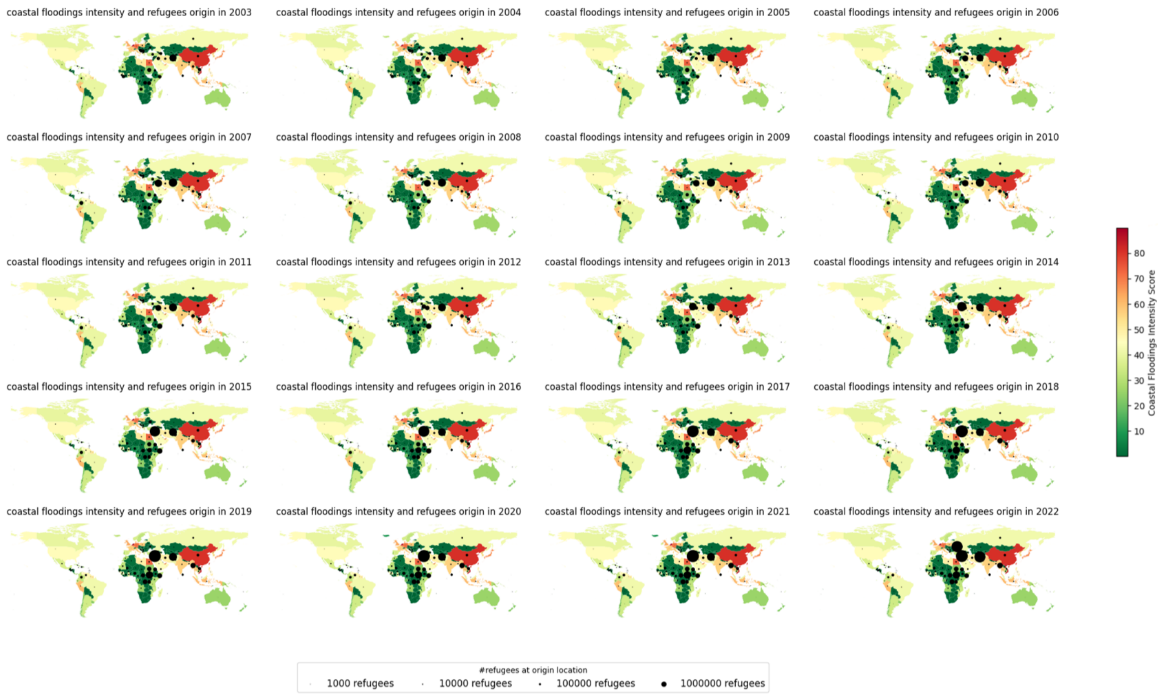


Figure 6.22: Bivariate plots of global coastal flooding intensity scores and refugees fleeing from a country over years

In Europe, the Netherlands has the highest risk, followed by Germany, United Kingdom, France, Italy, Denmark, Albania and Poland. All have above moderate coastal flooding intensity exposure. Other Europeans countries have low coastal flooding intensity numbers. Moreover, in Oceania, solely low coastal flooding intensity numbers can be found.

In Africa, only Egypt has high coastal flooding intensity. Further, Libya, Guinea and Ghana have moderate risk. The other countries on this continent all have low coastal flooding intensity.

In the Middle East, only the United Arab Emirates experiences high coastal flooding intensity. Saudi Arabia has below moderate coastal flooding risk, and Iran and Iraq have moderate risk. All other countries, like Turkey and Syria, have very low coastal flooding intensity.

In Asia, China and Vietnam are the only countries that have very high coastal flooding intensity exposure. Additionally, India, Japan, Indonesia, Thailand, Burma, Bangladesh, Korea and Philippines still have above moderate to high exposure. So, nearly all countries in South East Asia have high coastal flooding intensity. However, West-Asia (Kazakhstan, Pakistan, Afghanistan etc.) have very low coastal flooding risk. And in East-Asia, there are some countries of particular interest: Laos, Sri Lanka, Cambodia, Nepal, Malaysia and Bhutan. These are the only countries in South East Asia with low coastal flooding intensity scores. Therefore, refugee movements in Asia are of particular interest regarding this climatic indicator.

The coastal flooding intensities do not seem to change heavily over the years. The refugee movement situation is, of course, the same as in other bivariate maps for other indicators. So, again, the areas of particular interest in terms of refugees fleeing are Central and East Africa, the Middle East and South East Africa.

6.4.1.2. Coastal floodings: Top 25 countries refugees flee from

The top 25 origin countries for refugees (top countries refugees flee from) are the same across all climatic indicators. No major changes in the overall structure of these top 25 countries over the years can be detected. The core always is Central Africa, Middle East and South East Asia. In latest years, also some small parts of West-Africa occur in this top 25 list. From South-America, only Columbia appears over multiple years in this list. Venezuela is part of this list only in the latest years.

In terms of coastal flooding intensity, the only countries from the top refugee origin countries with high

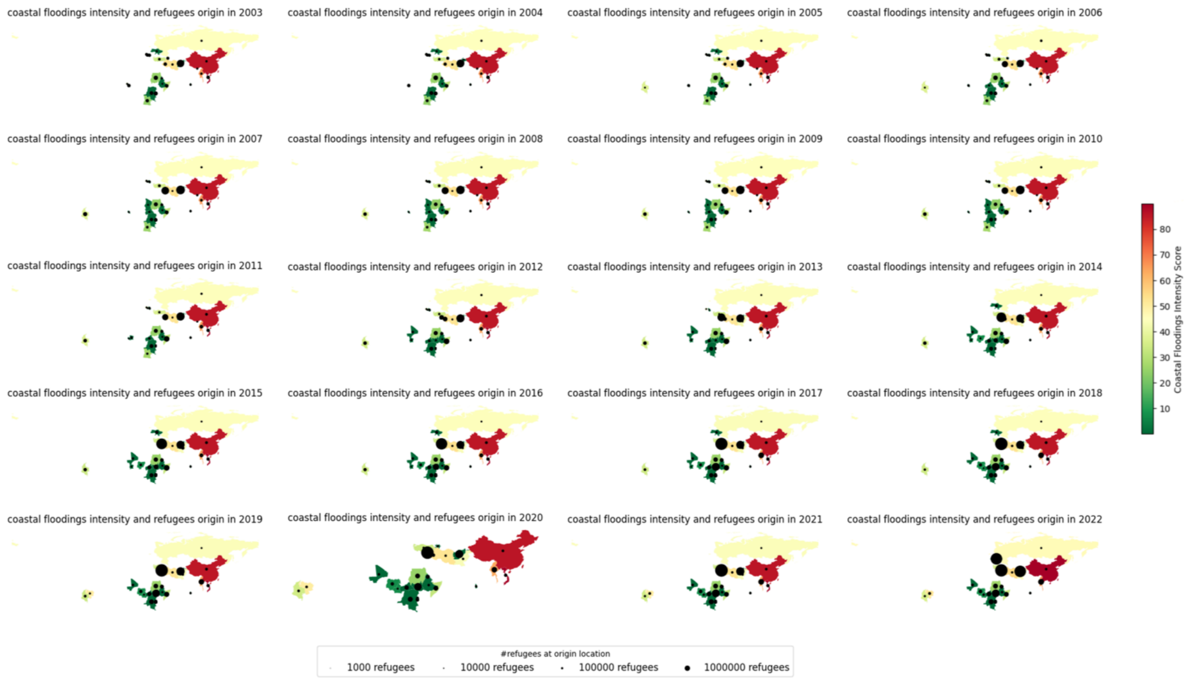


Figure 6.23: Bivariate plots of top hosting refugee countries: coastal flooding intensity scores and refugees fleeing from a country over years

exposure are: China, Vietnam and Burma. Further, Iran and Venezuela have moderate risk. The rest of the top countries have below moderate or low exposure to coastal floodings. Indeed, only refugees who flee from the mentioned countries here can end up in an even worse place regarding coastal flooding intensity. So, on these countries should be focused later in the origin-destination comparison below.

6.4.2. Coastal floodings: Refugees destination

6.4.2.1. Coastal floodings: Refugees fleeing to countries worldwide

Obviously, the coastal flooding situation is the same as in origin case (the world map is the same) and the refugee destination hotspots are also described in earlier analyses with different climate indicators.

In general (see Figure 6.24), not many countries in the world have high exposure to coastal flooding risks. From the previous section is clear that in South East Asia, there is a hotspots of countries with high coastal flooding intensity. Further, only Egypt, Libya, Peru, and United Arab Emirates are of high interest regarding coastal floodings. Moderate exposure is observed in Iraq and Iran, and Suriname. Western Europe and Poland also have relatively high coastal flooding exposure, but those countries are less interesting to study, because this are economically well developed countries and these can in general cope easier with climatic exposures. Also, they have a formal refugee influx procedure, making the refugees there less vulnerable.

Refugee influx hotspots are in Western Europe and North America, but most refugees stay in regions where they flee from. Therefore, there are high refugee receiving numbers in (Central and East) Africa, Middle East and Asia. Also this is described already in more detail by other climatic indicators.

6.4.2.2. Coastal floodings: Top 25 countries refugees flee to

The top refugee destination locations are rather constant over the years (see Figure 6.25). Apart from Europe and North America, numbers are also always high in Central, North and East Africa, the Middle East and Asia. In South America, sometimes Ecuador and Venezuela are among the top refugee destination countries. However, the core countries of interest are: China, Bangladesh, India, Pakistan, Iran, Turkey, Syria, Yemen, Egypt, Sudan, Ethiopia, Kenya, Tanzania, Congo, Uganda, Chad, Niger, South-Sudan and Cameroon. These are consistent top refugee destination countries in the world, especially in the latest years.

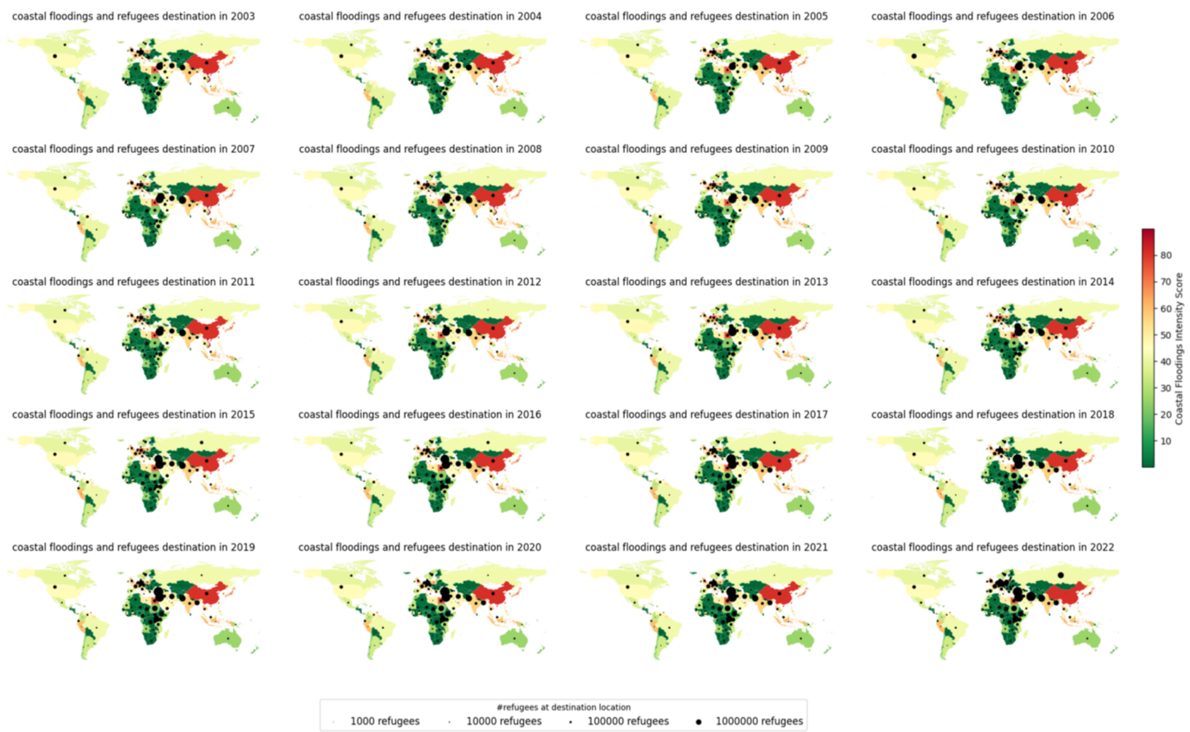


Figure 6.24: Bivariate plots of global coastal flooding intensity scores and refugees fleeing to a country over years

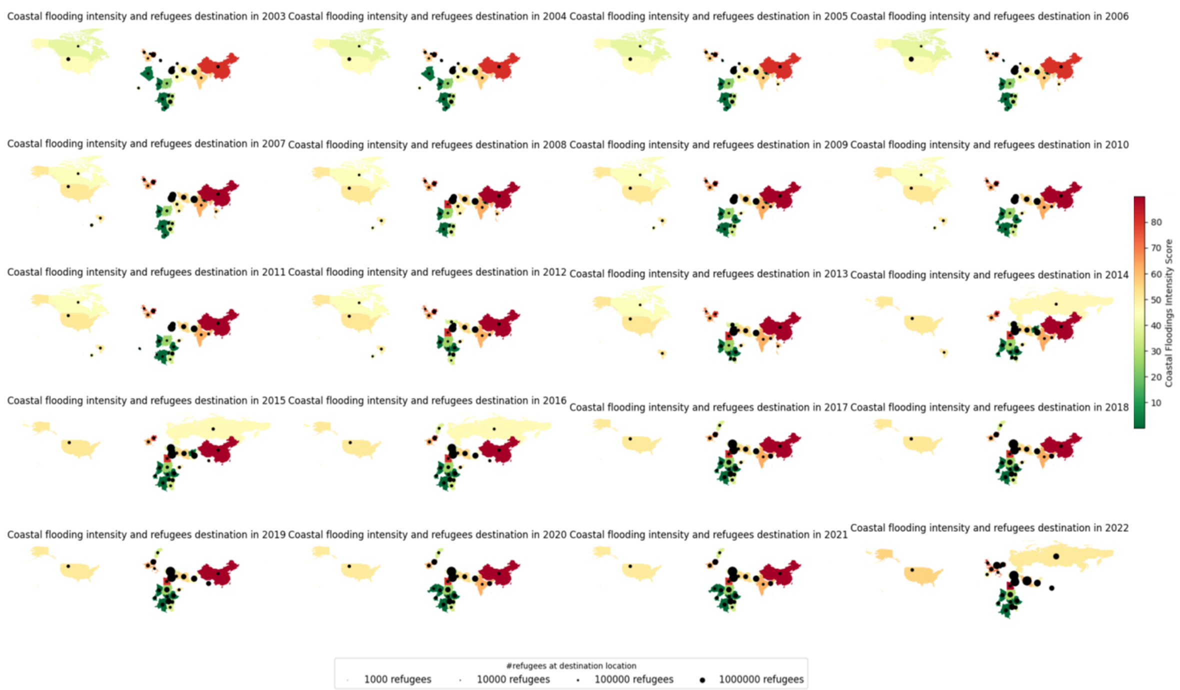


Figure 6.25: Bivariate plots of top refugee hosting countries: coastal flooding intensity scores and refugees fleeing to a country over years

From this top refugee destination list, most refugees are in Turkey and Syria, especially in the latest years. Also, a very high number of refugees is present in Pakistan, and Sudan and its neighboring countries. Those countries are on the top of the refugee receiving list, with Bangladesh also hosting many refugees compared to other countries in that region.

From the countries of interest described above, only China, Bangladesh, India and Egypt are highly exposed to coastal flooding intensities. Pakistan, Iran and Yemen have moderate coastal flooding intensity. In other interesting destination countries, coastal flooding risk is very low. This means that all African countries in this list, apart from Egypt, have very low coastal flooding intensity. In the Middle East, a low coastal flooding intensity score applies only to Turkey and Syria among the top destination countries.

6.4.3. Coastal floodings: Comparing origin-destination of refugees

The previous sections and identifications lead to this section.

6.4.3.1. Coastal floodings: Statistical generic difference tests

It can be seen, when comparing top 25 origin and destination maps in this report, that refugee origin and destination locations partly overlap. This prevents the possibility to directly conduct a statistical test to compare the top origin and destination countries. Namely, every statistical test to compare difference between two different groups, requires no overlap between those groups (for parametric as well as non-parametric test) (Leon, 1998). Due to the above, it becomes difficult to conduct suitable statistical tests to compare generic/overall differences between partly overlapping groups (top 25 origin and destination countries). However, there are two options that are still available (which themselves also consist of two options). These options are explained and statistical assumptions are tested in Appendix D. Here, only the results are presented.

First category of options

1) Testing differences of means

The median of the origin means over years is 23.00, while the median of the destinations means over years is 31.90. The conclusion of the Mann-Whitney U test is as follows: significant difference between the groups (U-statistic: 0.0, p-value: 6.79e-08). So, the difference of the means in de results above are significant. Destination mean values are statistically significant different from origin values. So, one can conclude that overall, destination countries are in more exposed coastal flooding intensity areas.

2) Testing differences of medians

The median of the top 25 origin medians gathered over 20 years is 2.0, while for the destination case it is 33.12. The conclusion of the Mann-Whitney U test is the following: significant differences between the groups (U-statistic: 0.0, p-value: 6.53e-08). So, those median differences are significant. Again, the same conclusion can be drawn: destination locations are in general, overall more exposed to coastal flooding intensity than origin locations.

The partly, generic conclusion is that top 25 origin and destination countries are overall not equally exposed to coastal flooding intensity over years, comparing means and well as medians. So, according to these tests, the destination locations are more exposed to coastal flooding than the origin locations.

Second category of options

This section concludes per year on the differences between groups.

3) Testing differences of origin values with median destination values (see Figure 6.26).

4) Testing differences of destination values with median origin values (see Figure 6.27).

In nearly all the cases every year, the conclusion is that the medians differ. The majority of the statistical information indicates that destination locations in general are more exposed to coastal flooding intensity than origin locations. The results are relatively conclusive. Over all the years, whether the mean of the destination or origin group is taken as given and the other group is tested against this mean, often, generic statistical differences are found. This means that overall, top origin and destination locations are differently exposed to coastal flooding intensity.

However, the results are also partly inconclusive. When the median of the origin is taken as 'population value' and is compared with the destination values, the results differ sometimes compared to when the situation is the other way around. Namely, in the 'origin as median' case, all the hypotheses are rejected, indicating differences between the values. However, in the 'destination as median' case, not all the hypotheses are rejected, indicating the medians do not differ. But, in most cases, the hypothesis

	Year	Median Origin	Median Destination	Test-statistic	p-value	Conclusion
0	2003	2.00	32.71	94.0	0.066702	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
1	2004	1.85	32.46	87.0	0.042150	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	2.00	32.69	95.0	0.070984	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
3	2006	2.00	32.85	94.0	0.066702	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
4	2007	2.00	40.89	75.0	0.017312	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	2.01	41.30	73.0	0.014722	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	1.97	34.92	88.0	0.045123	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	2.00	35.08	86.0	0.039339	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	1.97	34.56	89.0	0.048262	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	0.12	40.89	73.0	0.014722	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	1.90	46.27	39.0	0.000430	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	1.92	41.92	53.0	0.005189	Reject the null hypothesis. The median is significantly different from the specified value.
12	2015	1.95	32.78	87.0	0.042150	Reject the null hypothesis. The median is significantly different from the specified value.
13	2016	1.95	33.22	87.0	0.042150	Reject the null hypothesis. The median is significantly different from the specified value.
14	2017	1.94	33.18	88.0	0.045123	Reject the null hypothesis. The median is significantly different from the specified value.
15	2018	6.41	33.05	89.0	0.048262	Reject the null hypothesis. The median is significantly different from the specified value.
16	2019	6.56	32.89	98.0	0.085139	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
17	2020	6.42	32.53	80.0	0.044182	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	6.53	32.53	89.0	0.080054	Fail to reject the null hypothesis. The median is not significantly different from the specified value.
19	2022	6.42	32.53	66.0	0.015743	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.26: Wilcoxon signed rank test: Differences of origin values

	Year	Median Origin	Median Destination	Test Statistic	p-value	Conclusion
0	2003	2.00	32.71	55.0	0.002785	Reject the null hypothesis. The median is significantly different from the specified value.
1	2004	1.85	32.46	55.0	0.002785	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	2.00	32.69	55.0	0.002785	Reject the null hypothesis. The median is significantly different from the specified value.
3	2006	2.00	32.85	55.0	0.002785	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	2.00	40.89	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	2.01	41.30	28.0	0.000088	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	1.97	34.92	45.0	0.000912	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	2.00	35.08	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	1.97	34.56	36.0	0.001089	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	0.12	40.89	28.0	0.000088	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	1.90	46.27	28.0	0.000088	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	1.92	41.92	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
12	2015	1.95	32.78	45.0	0.000912	Reject the null hypothesis. The median is significantly different from the specified value.
13	2016	1.95	33.22	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
14	2017	1.94	33.18	36.0	0.000287	Reject the null hypothesis. The median is significantly different from the specified value.
15	2018	6.41	33.05	36.0	0.001089	Reject the null hypothesis. The median is significantly different from the specified value.
16	2019	6.56	32.89	45.0	0.000912	Reject the null hypothesis. The median is significantly different from the specified value.
17	2020	6.42	32.53	45.0	0.002587	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	6.53	32.53	55.0	0.002785	Reject the null hypothesis. The median is significantly different from the specified value.
19	2022	6.42	32.53	45.0	0.002587	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.27: Wilcoxon signed rank test: Differences of destination values

that the medians are equal, is rejected. The focus will be on the visual comparisons below to map out individual differences between countries, which in the end is what refugees experience.

6.4.3.2. Coastal floodings: Visual individual comparisons

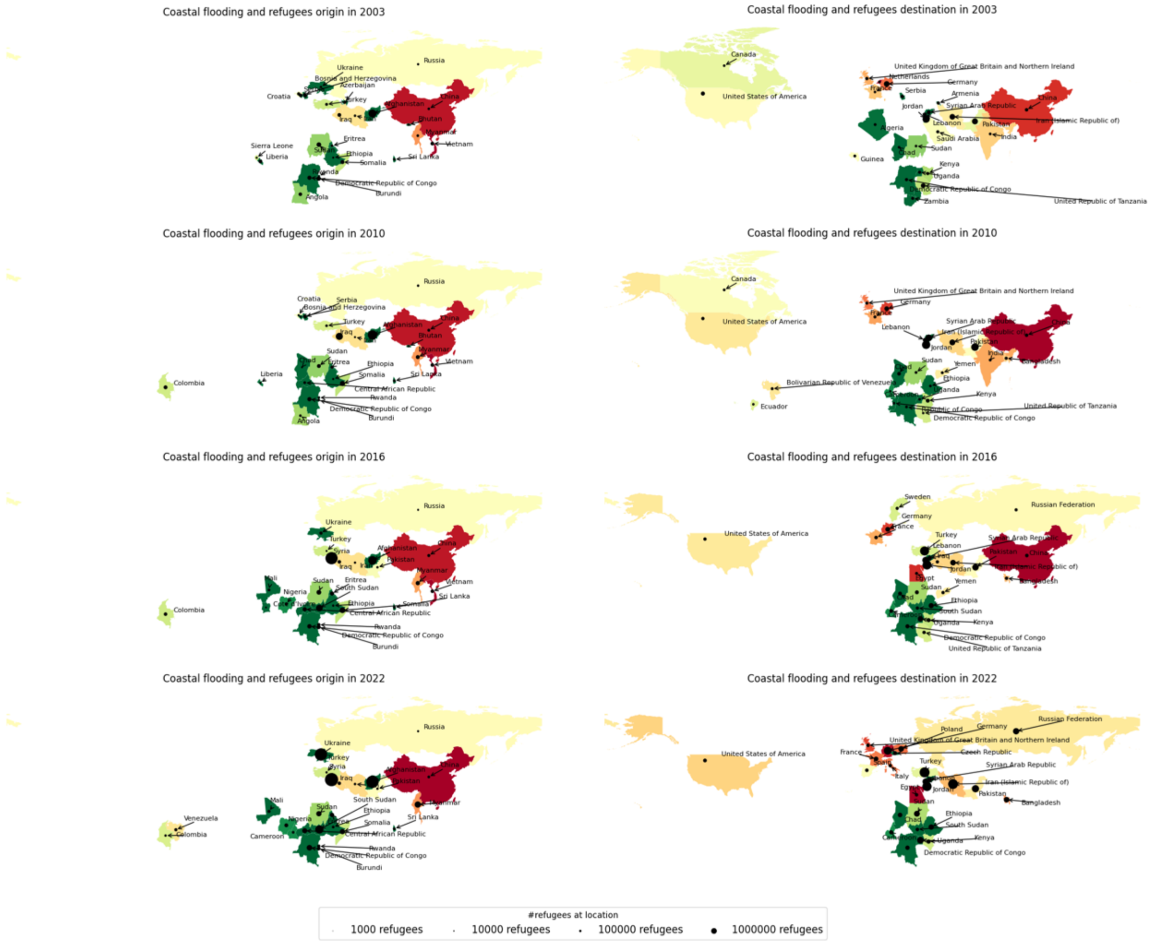


Figure 6.28: Comparing top origin and destination locations by coastal flooding scores

The top origin and destination locations are rather constant over the years (see Figure 6.28). In Africa, all the top origin countries have low coastal flooding intensity. This also holds for the top destination countries, except for Egypt. Egypt is the only country in Africa in the top destination list, that is highly exposed to coastal flooding intensity. So, from that point of view, fleeing to Egypt is the worst thing to do. However, when refugees from other countries in Africa, flee to Egypt, NGOs that support them should be aware of this different risk, and help them accordingly.

In South America, refugees sometimes flee from Colombia to Venezuela or Ecuador. Venezuela is way more exposed to coastal flooding risk than Ecuador and Colombia. Also, those refugees should be supported by making these differences clear to them and helping them avoid living near coastal areas in Venezuela, for example.

Refugees fleeing to Europe or North America are also exposed to moderate or high coastal flooding intensity. However, those countries usually can cope with this and have prevention measure into play. Also, they have a formal procedure for refugees coming in. Therefore, those refugee movements are of less interest in this study to help refugees. Note further that Oceania (and Indonesia) never are in any top origin or top destination refugee lists.

In the Middle East, top origin and destination locations overlap. That is partly also the case in Africa, but in this region it stands out more, because the overlap is larger. Namely, consistently top origin locations are: Turkey, Iraq, Iran, Syria (from 2014 onwards). Consistently top destination locations are: Turkey (after 2010 onwards), Jordan, Syria, Lebanon, Iran. So, Turkey, Syria and Iran are countries that are consistently top origin and top destination locations. Iraq usually only is a top origin country and not a top destination location. Jordan and Lebanon are top destination locations and not top origin. But, to be clear, from all top origin location in the list in the Middle East, only Iraq is not also a top destination location. From the top origin countries in this region, only Iran and Iraq are exposed to above moderate coastal flooding risk, the other countries all have very low exposure to coastal flooding. In the top destination locations, only Iran has above moderate exposure to coastal flooding. So, from a coastal flooding point of view, fleeing from other top origin Middle East countries to Iran is a problem, because then refugees are more exposed to coastal floodings compared to where they came from.

In Asia, there is less overlap in origin and destination countries, apart from China and Pakistan. China always is a top origin and destination country, and Pakistan always is a top destination and in latest years also a top origin country. Also Russia is always a top origin country and as well sometimes a top destination country. Further, consistently top origin countries over the years are: Afghanistan, China, Sri Lanka, Myanmar, Pakistan and Vietnam. Top destination countries in this region consistently are: Pakistan, India, China, Bangladesh. Afghanistan and Sri Lanka are top origin countries with very low exposure to coastal flooding. So, from this perspective, fleeing from those countries to top destination countries in this region, will always make refugees more exposed. Other top origin countries, China, Myanmar and Vietnam, score extremely bad in terms of coastal flooding intensity. So, fleeing from these countries will not place refugees in an even worse location, in terms of coastal floodings. Pakistan has below moderate coastal flooding exposure, like Russia. However, all (other) top destination countries, apart from Pakistan, have very high coastal flooding exposure. So, fleeing to Bangladesh, India or China, will put Pakistani, Russian, Afghan and Sri Lankan refugees in a worse coastal flooding place. In general, apart from Afghanistan, Sri Lanka and Pakistan, both top origin and destination countries in this region are highly exposed to coastal floodings.

6.5. Cyclones

The cyclone intensity scores below range between 0 (no exposure to cyclone intensity) and 80 (extremely high exposure to cyclone intensity). In other words, scores range from low to high, and from green to red in the plots below.

6.5.1. Cyclones: Refugees origin

6.5.1.1. Cyclones: Refugees fleeing from countries worldwide

Figure 6.29: Bivariate plots of global cyclone intensity scores and refugees fleeing from a country over years

One can immediately conclude from Figure 6.29 that most countries, over years, are not exposed to high cyclone intensity. Many countries in the world have the green color of low risk. Only in the United States, Mexico, China, Japan, South Korea's, Madagascar and Philippines, the cyclone intensity risk is above moderate to high. Australia is the only country with moderate cyclone intensity exposure. But, again, further, every country has this green color, indicating 'good', relative safe scores for cyclone risks. This patterns is the same over all these years, no changes are detected in the last 20 years.

Especially in the regions of interest regarding large refugee movements: Central, East Africa, Middle East and East Asia, all the countries have very little risk of intense cyclones, except for Madagascar, China, South Korea, Philippines and Japan. However, in terms of refugee movements, only China plays a role regarding cyclone intensity risk with lots of refugees fleeing to and from this country. The other countries have relatively small refugee numbers (influx as well as outflow). From China, approximately 200,000 refugees flee per year.

The refugee numbers (dots) confirm the previous statement: every relatively big dot is in a 'green' country, apart from China. This means that only refugees who flee from China, can arrive in a better place regarding cyclone intensity. Refugees who flee from other hotspots, can arrive in even worse places if they flee to the Asian countries listed above.

6.5.1.2. Cyclones: Top 25 countries refugees flee from

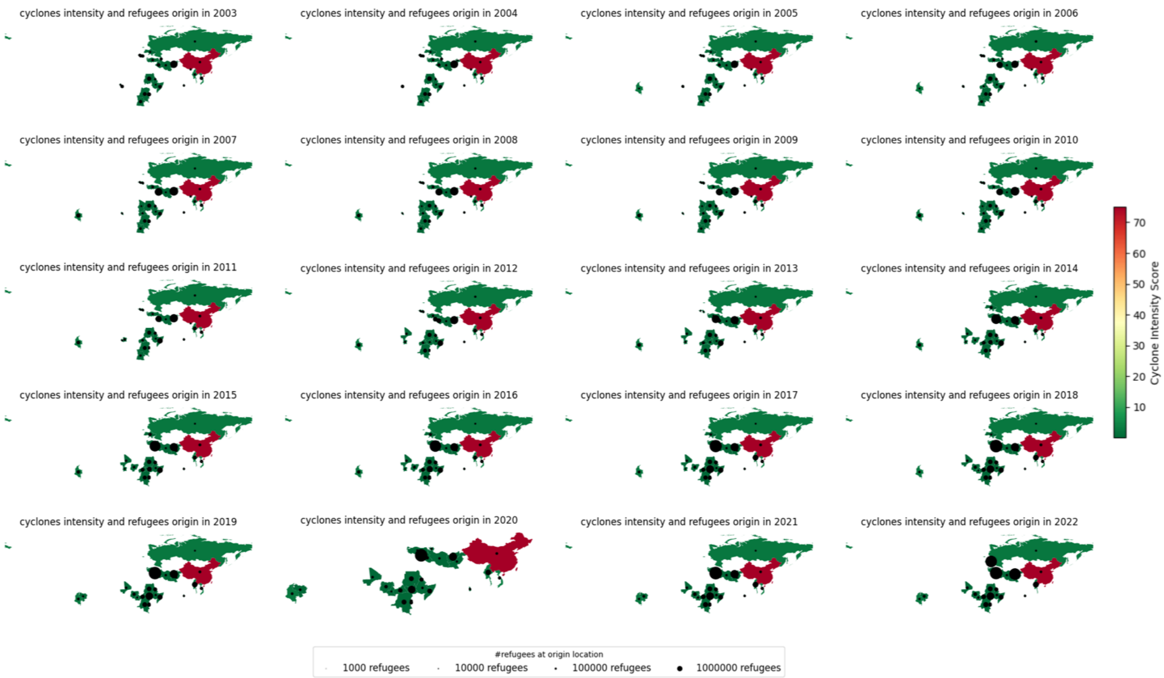


Figure 6.30: Bivariate plots of top refugee hosting countries: cyclone intensity scores and refugees fleeing from a country over years

The findings in the previous section are confirmed by this plot (see Figure 6.30). Of the top 25 countries refugees flee from over the last 20 years, only China has high scores for cyclones intensity. There is no exception, no other top origin country that has another color than green. The other top countries refugees flee from, do not have high or moderate cyclone intensity exposure, only (very) low. This means that, from a cyclone risk point of view, it is only 'good' to flee from China.

6.5.2. Cyclones: Refugees destination

Of course, the cyclone intensity exposure stays the same in this destination analysis, only the refugee numbers change, because here is examined to which countries refugees flee instead of from which countries they flee. However, also this refugee data is already examined for other climatic events. What is new, is the combination of the cyclone exposure and refugee destination locations.

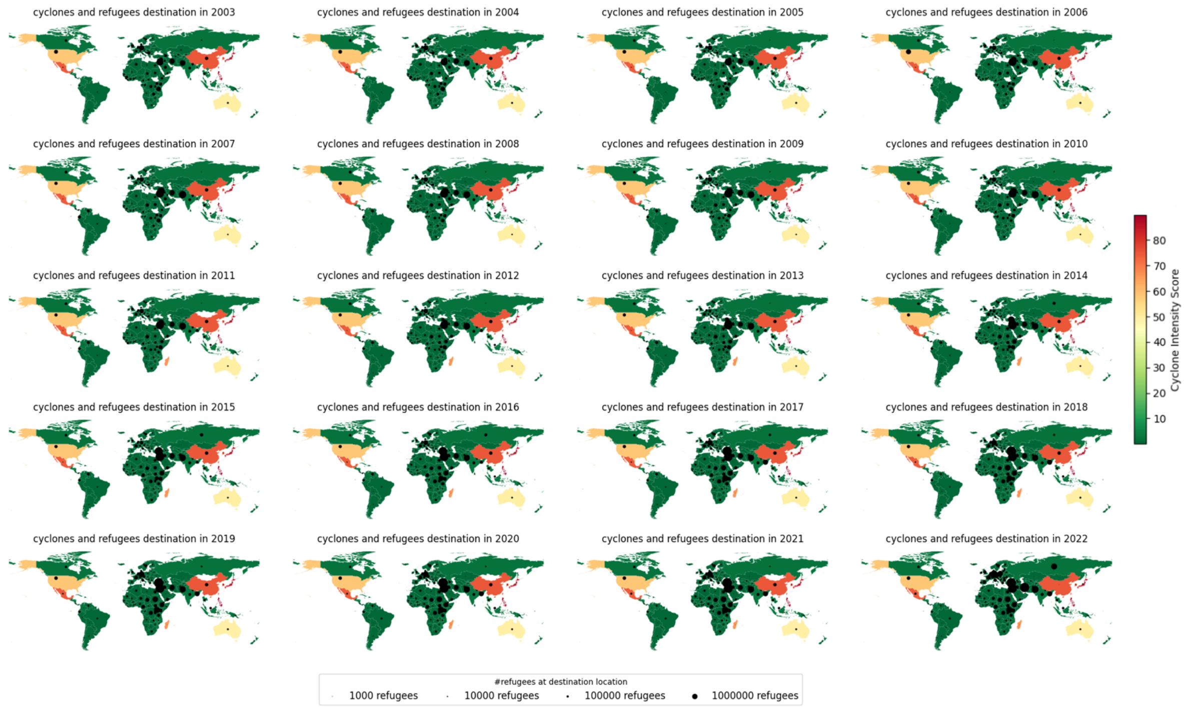


Figure 6.31: Bivariate plots of global cyclone intensity scores and refugees fleeing to a country over years

6.5.2.1. Cyclones: Refugees fleeing to countries worldwide

One can immediately conclude from Figure 6.31 that the only prominent destination countries of refugees with high cyclone intensity exposure are the United States and especially China. They both host around 300,000 refugees every year. Australia has moderate cyclone intensity risk and hosts not that many refugees, nearly always below/around 50,000 per year. All other countries in the world with significant refugee receiving numbers (say hosting above 30,000 refugees a year), have very low exposure to high cyclone intensity. Only in the latest three years, Mexico receives above 50,000 refugees, but it is still not a large refugee hosting country. However, this country also has high cyclone intensity risk.

The United States is not so interesting to focus on, as this country is able to cope with changing climate due to its economic development. Also, the Unites States has a formal support procedure for refugees who arrive. More interesting is to look at less economically developed places in which refugees are more vulnerable, due to the lack of help from institutions. However, these countries (in for example Middle East, Asia and Africa) usually have very low exposure to cyclone intensity.

So, from an interesting refugee destination point of view in combination with cyclone intensity exposure and coping mechanisms, only China is interesting to investigate further in the comparison between origin and destination countries below.

6.5.2.2. Cyclones: Top 25 countries refugees flee to

Previous findings above are confirmed in this plot (see Figure 6.32). Only the United States and China are high refugee receiving countries with high exposure to cyclone intensity. All other top refugee receiving countries, with especially in the Middle East countries taking in over millions of refugees, for example Turkey) are all not exposed to (high) cyclone risks. This also holds for Africa.

South American countries in this top list are minimal, only Venezuela and Peru are in this top 25 list sometimes, both with low cyclone risks. This also holds for countries in Europe that receive many refugees, especially Germany.

To conclude, refugees fleeing to China, from for example the Middle East or West Asian countries, should be prepared for the difference in cyclone exposure. At least, NGOs can help support them regarding awareness and location picking, as well as providing tips, tools and tricks to cope with the different cyclone intensity exposure.

Figure 6.32: Bivariate plots of top refugee hosting countries: cyclone intensity scores and refugees fleeing to a country over years

6.5.3. Cyclones: Comparing origin-destination of refugees

The previous sections and identifications lead to this section.

6.5.3.1. Cyclones: Statistical generic difference tests

One can conclude, when comparing top 25 origin and destination maps in this report, that refugee origin and destination locations partly overlap. This prevents the possibility to directly conduct a statistical test to compare the top origin and destination countries. Namely, every statistical test to compare difference between two different groups, requires no overlap between those groups (for parametric as well as non-parametric test) (Leon, 1998). Due to the above, it becomes difficult to do statistical tests to compare generic/overall differences between partly overlapping groups (top 25 origin and destination countries). However, there are two options that are still available (which themselves also consist of two options). These options are explained and statistical assumptions are tested in Appendix D. Here, only the results are presented.

First category of options

1) Testing differences of means

The median of the top 25 origin means over years is 3.46, while the median of the top 25 destination means over years is 5.63. The conclusion of the Mann-Whitney U test is as follows. The tests results indicate significant differences between the groups (U-statistic: 20.0, p-value: 1.19e-06). So, the difference of the means in de results above are significant. Destination mean values are statistically significant different from origin values. So, one can conclude that overall, destination countries are in more exposed cyclone intensity areas.

2) Testing differences of medians

Here one can see something interesting: all the medians, for origin as well as destination, for every year, always have the value 0.01 (see Figure D.1 in Appendix D). There is no spread, there is no distribution. 0.01 is the median each and every year, for all the top countries. So, both median of medians are also 0.01. And of course, there is also no difference between the groups (U-statistic: 200.0, p-value: 1.0).

This means that average/mean values of cyclone intensity are higher in destination locations compared to origin, in general. However, median values are the same in origin as well as destination locations.

This indicates that for all years, there is in general low exposure (0.01 is always the middle value of all the data, the median) in origin as well as destination. However, if there is exposure, that is only in a few countries, than the destination is more exposed than origin, in general, due to the higher mean of mean values in this group.

The partly, generic conclusion is that top 25 origin and destination countries are in general, overall mostly equally exposed to cyclone intensity over years, comparing means and medians. However, if there is exposure, in general, the cyclone intensity is higher in destination locations than in origin locations.

Second category of options
This section concludes per year on the differences between groups.

3) Testing differences of origin values with median destination values (see Figure 6.33).

	Year	Median Origin	Median Destination	Test-statistic	p-value	Conclusion
0	2003	0.01	0.01	0.0	0.027708	Reject the null hypothesis. The median is significantly different from the specified value.
1	2004	0.01	0.01	0.0	0.027708	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
3	2006	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	0.01	0.01	0.0	0.011719	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	0.01	0.01	0.0	0.011719	Reject the null hypothesis. The median is significantly different from the specified value.
12	2015	0.01	0.01	0.0	0.011719	Reject the null hypothesis. The median is significantly different from the specified value.
13	2016	0.01	0.01	0.0	0.011616	Reject the null hypothesis. The median is significantly different from the specified value.
14	2017	0.01	0.01	0.0	0.011616	Reject the null hypothesis. The median is significantly different from the specified value.
15	2018	0.01	0.01	0.0	0.011616	Reject the null hypothesis. The median is significantly different from the specified value.
16	2019	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
17	2020	0.01	0.01	0.0	0.011616	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
19	2022	0.01	0.01	0.0	0.011616	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.33: Wilcoxon signed rank test: Differences of origin values

4) Testing differences of destination values with median origin values (see Figure 6.34).

What one can see in both tables is very interesting: all null hypotheses, for every year, whether the origin or destination median is treated as given and the other group is tested against, is rejected, while the medians of the groups are the same. This raises many questions. However, once you know that the (one) signed Wilcoxon rank test only test differences from the specified median, it becomes intuitive. There are many equal ranks in the data, indicating a difference of zero, namely, the median is 0.01 (as we saw in the previous analysis) and one of the other group values is also 0.01. This difference is then not taken into account in the analysis. The tests only focuses on difference from the median, not equal values to the median. Therefore, these conclusions are logical, because if a country value is not 0.01 (which is the minimum score) it is higher than 0.01, and that value is then tested against the 0.01 median from the other group, then, of course, the difference is always positive. That is why the test statistic is 0.0 in all the cases. However, the practical value of this is minimal. One cannot conclude anything from this, apart from the fact that overall medians of origin and destination groups are the same, but that some individual countries differ heavily from this median 0.01 value and have more

	Year	Median Origin	Median Destination	Test Statistic	p-value	Conclusion
0	2003	0.01	0.01	0.0	0.011311	Reject the null hypothesis. The median is significantly different from the specified value.
1	2004	0.01	0.01	0.0	0.011311	Reject the null hypothesis. The median is significantly different from the specified value.
2	2005	0.01	0.01	0.0	0.007474	Reject the null hypothesis. The median is significantly different from the specified value.
3	2006	0.01	0.01	0.0	0.007474	Reject the null hypothesis. The median is significantly different from the specified value.
4	2007	0.01	0.01	0.0	0.004948	Reject the null hypothesis. The median is significantly different from the specified value.
5	2008	0.01	0.01	0.0	0.004948	Reject the null hypothesis. The median is significantly different from the specified value.
6	2009	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
7	2010	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
8	2011	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
9	2012	0.01	0.01	0.0	0.007632	Reject the null hypothesis. The median is significantly different from the specified value.
10	2013	0.01	0.01	0.0	0.007686	Reject the null hypothesis. The median is significantly different from the specified value.
11	2014	0.01	0.01	0.0	0.011719	Reject the null hypothesis. The median is significantly different from the specified value.
12	2015	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
13	2016	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
14	2017	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
15	2018	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
16	2019	0.01	0.01	0.0	0.027708	Reject the null hypothesis. The median is significantly different from the specified value.
17	2020	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
18	2021	0.01	0.01	0.0	0.017960	Reject the null hypothesis. The median is significantly different from the specified value.
19	2022	0.01	0.01	0.0	0.043114	Reject the null hypothesis. The median is significantly different from the specified value.

Figure 6.34: Wilcoxon signed rank test: Differences of destination values

cyclone exposure. In short, many countries have equal, 0.01, cyclone intensity exposure. However, some countries in origin as well as destination locations have (much) higher cyclone risk. The focus will be on the visual comparisons below to map out individual differences between countries, which in the end is what refugees experience.

6.5.3.2. Cyclones: Visual individual comparisons

It is interesting to see that only China has high exposure to cyclone intensity and is a top origin country as well as a top destination country for refugees. So, the refugees who flee from China to other top destination countries will be in a better place regarding cyclone intensity exposure.

Refugees fleeing from other top origin countries are likely to arrive in a similar, namely low, risk profile regarding cyclone intensity exposure. However, if they flee to China or the United States, they arrive in a higher exposed cyclone intensity country. All other top destination countries score low cyclone intensity exposure. This pattern does not really change over years: all other top origin and destination locations are 'green', i.e. have low cyclone intensity exposure.

To conclude, refugees fleeing from top origin countries to China, should be supported by NGOs in their cyclone intensity exposure transition. They should be aware of this and should have tools, tips and tricks to cope with these potential danger cyclone situation. Because (apart from China itself) refugees fleeing from other top origin countries will arrive in a worse cyclone intensity environment. To a lesser extent, this also holds for fleeing to the United States from top origin countries. The cyclone intensity risk is lower than in China, and coping mechanism and economically development are better in the United States, so the focus for this climatic indicator is primarily on China.

6.6. Summary

In this chapter, bivariate plots of climatic indicators (droughts, riverine floodings, sea level rise, coastal floodings and cyclones) with refugee numbers (fleeing to and from a country) are constructed in the period 2003-2022. These bivariate plots are presented in two ways: 1) for all the countries in the world (global) and 2) for only the top 25 origin/destination countries. Statistical differences between origin

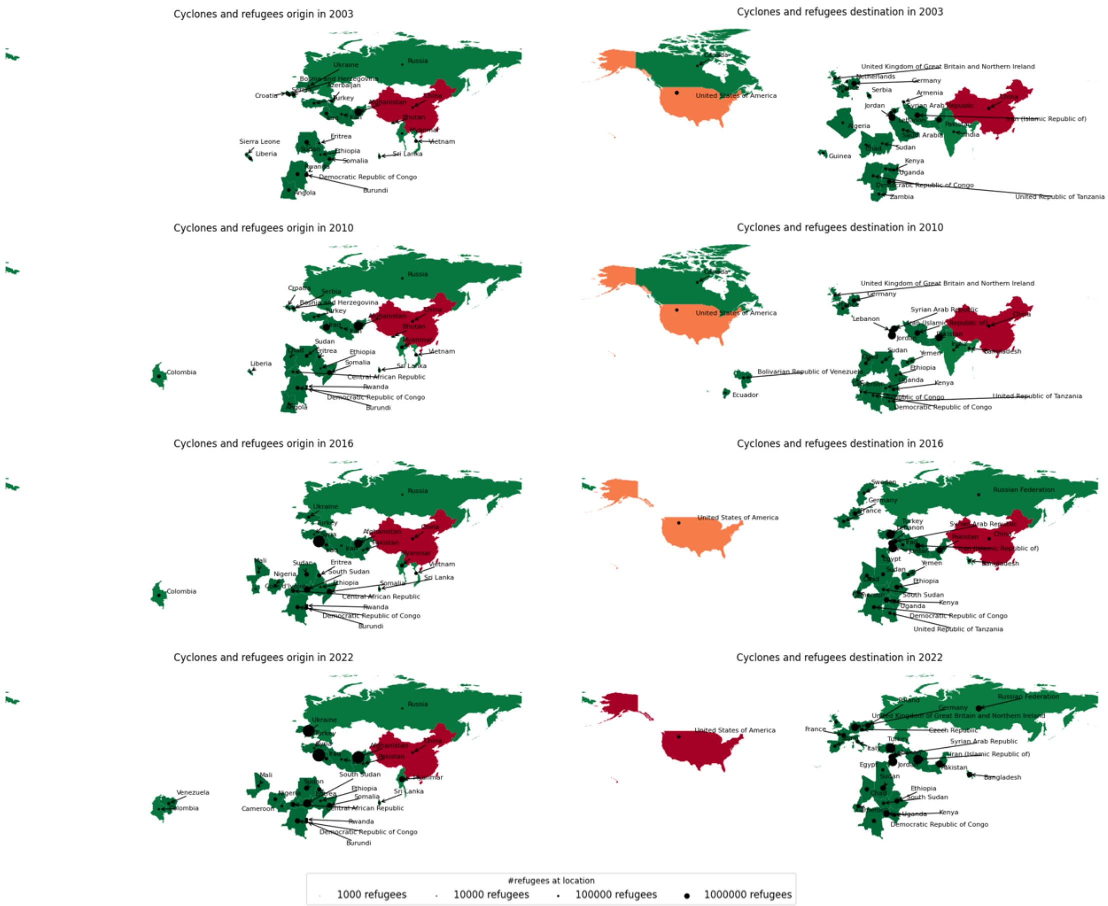


Figure 6.35: Comparing top origin and destination locations by cyclone scores

and destination locations are examined.

High drought exposure significantly affects Central and South America, Central-East Africa, parts of the Middle East, and Southeast Asia, while other regions generally experience lower drought intensity. Noteworthy regarding this is that Russia is consistently among the top 25 refugee origin countries, while it is often also a top 25 destination location, with an above moderate drought intensity score.
Specifically in Africa, Congo, Ethiopia, Cameroon, Sudan, Yemen and Chad are countries with consistently high drought intensity scores and they are also in the top 25 destination countries. Refugees fleeing from other countries to these countries should be assist in terms of drought management, but especially those who come from top 25 origin countries Mali, Côte d'Ivoire, Nigeria, Burundi, Rwanda and South Sudan, as those will come from a far less intense drought situation, in general.
In Asia, the countries Afghanistan, Bhutan, Sri Lanka, Myanmar, and Bangladesh have low drought intensity, whereas Vietnam, India, and China have high drought intensity. From these countries, only Bangladesh, India and China are top destination countries. Top origin countries are Myanmar, Vietnam, China and Sri Lanka. Refugees fleeing to India and China face at least equal or worse drought conditions.
In the Middle East, drought intensity is similar in both top origin and destination locations. However, in northern South America, Columbia and in latest year Venezuela are top countries refugees flee from, with Columbia experiencing more drought intensity exposure than Venezuela. In South America there are no top destination countries.
In general, the statistical difference tests between all top origin and destination countries over years, give inconclusive information. However, a part of the statistical information indicates that origin locations, on average, are more exposed to droughts than destination locations.

Outside the top origin and destination countries of refugees, sea level rise risk is extremely high in

Brazil, Australia, Indonesia and Japan. However, most countries face moderate to low risk.
Until 2020, most of the top origin countries had a low risk of sea level rise, with the exceptions of Vietnam and Sri Lanka. In more recent years, Turkey, Colombia, Venezuela, China, and Cameroon have been identified as having relatively high risk scores among the top 25 origin locations. Furthermore, China, Turkey, and Venezuela are experiencing increasing sea level rise risks.
In the early 2000s, among the top destination countries for refugees, Congo and India faced high sea level rise risks, while Algeria and Saudi Arabia were at moderate risk. Countries like the Netherlands, the UK, Canada, and the US are not highlighted due to their effective coping and adaptation mechanisms, as well as formal procedures for assisting refugees. By 2010, Ecuador, Congo, Kenya, Cameroon, Venezuela, and Tanzania were facing high or extreme risks from sea level rise. India and China also continued to have high risks. In 2016, Egypt and Uganda were added to the list of top destination countries with significantly high sea level rise risks, while Bangladesh had an above moderate risk. Refugees fleeing from the above low-risk African nations to these high-risk destinations, and from Myanmar to Bangladesh, India or China, need support and information regarding sea level rise risks. In recent years, there has been no change in the list of high risk top refugee destination countries.
The results of statistical difference tests between all top origin and destination locations are partly inconclusive: comparing origin medians with destination values shows significant differences, while comparing destination medians with origin values shows no significant differences from 2014 onwards. Overall, statistical tests suggest that top destination locations are generally more exposed to sea level rise than top origin locations.

Riverine flooding intensity exposure remains relatively stable over the years, but varies by region. In general, the entire continent America experiences above moderate risk, but South East Asia, as a major refugee hotspot, experiences very high riverine flooding intensity. Many refugees remain in their region of origin.
In Africa, top origin countries with low riverine flooding intensity are Angola, Rwanda, Burundi, Somalia, Ethiopia, the Central African Republic, Sierra Leone, Liberia, Côte d'Ivoire, Cameroon, and Eritrea. Refugees fleeing from these countries face higher exposure when relocating to top destination countries in the region, which are Egypt, Chad, Mali, Nigeria and Sudan, all of which have worse riverine flooding intensity. Congo, South Sudan and Cameroon, which have experienced moderate exposure over the years, are also top destination locations. Other popular destination countries in Africa with low riverine flooding intensity are Zambia, Tanzania, Uganda, Kenya, Guinea, Yemen and Algeria.
In Asia, top origin countries are Afghanistan, China, Russia, Bhutan, Myanmar, Pakistan, Vietnam and Sri Lanka. All of these countries face high riverine flooding intensity, with the exceptions of Afghanistan, Bhutan, and Sri Lanka. Top destination countries in Asia —India, Bangladesh, China, and Pakistan— also have high flooding intensity. This means that refugees fleeing within Asia typically encounter similar or even worse riverine flooding conditions when moving from top origin to top destination locations. However, refugees moving from Asia to the Middle East generally experience lower riverine flooding intensity.
In the Middle East, Iraq is the only country with high exposure to riverine flooding. Top destination countries in this region -Turkey, Syria, Iran, Lebanon, and Jordan— have low to moderate flooding intensity, meaning that refugees fleeing Iraq would encounter less severe riverine flooding conditions in these areas. On the other hand, Iran and Afghanistan are top origin countries with low riverine flooding intensity, but they are surrounded by top destination countries with higher exposure, namely Iraq and Pakistan. Among nearby countries, Jordan and Egypt are both top destination countries, but while Egypt experiences one of the highest riverine flooding intensity in the world, Jordan experiences low intensity exposure.
In South America, Colombia and Venezuela are countries where many people flee from (top origin countries). Top destination countries like Ecuador and Venezuela have equal or lower exposure to riverine flooding, minimizing the negative impact on moving refugees between those countries.
Overall, no significant statistical differences are found in riverine flooding intensity between all top origin and destination countries.

Apart from some top origin and destination locations (discussed below), only a few other countries worldwide experience high coastal flooding intensity. These include Peru, Suriname, and Guyana in America; the UAE in the Middle East; and Japan, Thailand and Indonesia in Asia. Further, many countries in Europe experience high coastal flooding exposure. In general, coastal flooding intensity

remains stable over time for countries.
In Africa, all top origin countries have low coastal flooding intensity, which is also true for most top destination countries, with the exception of Egypt, which faces a high risk of coastal flooding. Refugees fleeing to Egypt will face significantly higher coastal flooding risks. There is also some overlap between top origin and destination countries in this region.
In South America, refugees fleeing from Colombia to Venezuela or Ecuador encounter higher coastal flooding risks in Venezuela compared to Ecuador and Colombia. Support efforts should prioritize helping refugees understand and avoid high exposure coastal regions in Venezuela.
In Europe and North America, both top destination regions, coastal flooding intensity ranges from moderate to high. However, these countries have effective coping mechanisms and formal procedures for refugees, making them less of a focus in this study. Additionally, regions like Oceania and Indonesia do never appear in the top origin or destination lists.
In the Middle East, there is considerable overlap between top origin and destination countries. Top origin countries are Turkey, Iraq, Iran, and Syria, while top destination countries are Turkey, Jordan, Syria, Lebanon, and Iran. Notably, Turkey, Syria, and Iran consistently serve as both top origin and destination countries. Iraq and Iran are the only nations in this region facing above moderate to high coastal flooding risk, while other countries have low exposure. Refugees moving from other top origin countries in the Middle East to Iran will face increased coastal flooding risks.
In Asia, there is less overlap between top origin and destination countries, except for China and Pakistan, which are consistently both. Russia is also a top origin country and occasionally a top destination. Only consistently top origin countries are Afghanistan, Sri Lanka, Myanmar, Pakistan, and Vietnam. Additionally, top destination countries are India and Bangladesh. Refugees fleeing from top origin locations Afghanistan and Sri Lanka, which have very low coastal flooding exposure, will face increased exposure in top destination countries. In contrast, refugees fleeing from other top origin countries China, Myanmar, and Vietnam —countries with extremely high coastal flooding intensity— will not encounter worsened coastal flooding conditions when relocating. Pakistan and Russia have below moderate coastal flooding exposure, but most other top destination countries face very high exposure. Therefore, refugees fleeing from Pakistan, Russia, Afghanistan, and Sri Lanka will encounter higher coastal flooding exposure when moving to Bangladesh, India, or China. In general, apart from Afghanistan, Sri Lanka, and Pakistan, top origin as well as top destination countries are in regions highly exposed to coastal flooding.
Overall, statistical analyses mostly show that destination locations have higher coastal flooding intensity than origin locations, with significant differences observed over the years. However, results are mixed: while top origin and destination medians often differ, some comparisons show no significant difference.

Most countries have low cyclone intensity risk, with only the U.S., Mexico, China, Japan, South Korea, Madagascar, and the Philippines facing high exposure, and Australia experiencing moderate exposure. Nearly all top refugee destinations, including those in the Middle East, Africa, South America, and Europe, face low cyclone risks. Also, among the top 25 refugee origin countries, only China faces high cyclone intensity. All other top origin countries have low cyclone risk. Over the years, this patterns remains constant. Most refugee movements occur in low risk regions.
China stands out as the only country that is both a top origin and destination location with high cyclone intensity exposure. In general, refugees fleeing from China to other top destination countries will encounter lower cyclone intensity. Refugees fleeing from other top origin countries generally move to destinations with similar low cyclone intensity, with the exception of those fleeing to the top destination United States, which has high cyclone intensity exposure.
Overall, statistical differences between all top origin and destination countries are inconclusive when it comes to cyclone intensity exposure, meaning that they are in general equally exposed.

7

Conclusion

7.1. Drivers of refugee movements

This conclusion section addresses SQ2: How strongly could climate exposure drive people to flee, over years? To answer this question, two overarching analyses were conducted: a correlation analysis and multiple linear regression modeling. Changes in correlation coefficients over the years were statistically tested, and significant coefficients in the multiple linear regression models over the years were examined.

7.1.1. Correlation

7.1.1.1. Fleeing from a country (origin)

Based on this analysis, only riverine floodings -and to a lesser extent droughts- could potentially be climatic drivers of refugees fleeing away from a country. However, these climatic events can only partly act as drivers, and their potential driving principle is weaker compared to some social variables, especially the lack of coping capacities of a country (a term that includes aspects like conflicts and democratic principles). This social indicator shows a higher correlation association with refugees fleeing from a country compared to riverine floodings, with coefficients of 0.5 and 0.3 respectively. Note however that these signs, sizes and conclusions can change when conducting regression analysis.

Over the last 20 years, the correlation between coastal flooding/sea level rise, and refugees fleeing from a country has remained stable. Riverine flooding shows a minimal decrease, while cyclone intensity, droughts, lack of adaptive/coping capacities, and socio-economic deprivation have seen slight increases in their correlations with refugees fleeing. Notably, societal disparities and lack of socio-economic development saw a significant increasing trend of about 0.15 between 2003 and 2022. If these trends continue, socio-economic factors will have a stronger association with refugees fleeing from countries in the future.

7.1.1.2. Fleeing to a country (destination)

Based on this analysis, only cyclone intensity could potentially be a small, but real driver of refugees -not- fleeing to a country. Namely, this is the only indicator with a negative sign, meaning more cyclone intensity is associated with less refugees fleeing to a country (-0.15).
All other variables, even the social ones, have positive signs (and are also generally lower than compared to the origin case), indicating an association that, for instance, more riverine flooding intensity or less coping capacity (only those have constant significant correlation coefficients) is associated with more refugees fleeing to a country. This is an explainable association, as many refugees flee to neighboring countries where potentially there are also, for instance, conflicts and riverine floodings. However, those indicators cannot be a (causal) driver for determining whether or not to flee to a country, according to this correlation analysis. One cannot potentially interpret this as more droughts leads to more people fleeing to a country or lower coping capacities (e.g. more conflicts) leads to people fleeing to a country. It is just a descriptive finding of certain associations. However, signs and sizes could change when conducting multivariate regression.

The above, and also the fact that in general less correlation coefficients are (statistically) significant compared to the origin case, could potentially be explained by the fact that refugees have to flee their country of origin, without looking at where they will arrive, i.e. that refugees often flee from climatic and/or social poor countries to (neighboring) countries, which (also) have poor climate/social conditions. This statement is also supported by the fact no similar conclusions can be drawn, based on origin and destination correlations. This means that for none of the indicators, the origin correlation is highly positive (association with fleeing from) and in the destination case is highly negative (fleeing -not- to).

Over the last 20 years, the rather low correlations between refugees fleeing to a country and the indicators coastal floodings, droughts, lack of adaptive capacities, sea level rise, societal disparities and lack of socio-economic development has not changed significantly. Cyclones and socio-economic deprivation show a minimal decrease in their correlation with refugees fleeing to a country.
The correlation between riverine flooding/lack of coping capacities and refugees fleeing to a country increases by respectively 0.16 and 0.2 over 20 years. This significant rise suggests a stronger positive association between these factors, but, again, it should not be interpreted as a potential causal driving relationship that more flooding intensity/lack of coping capacities leads to more refugees fleeing to a country. Instead, it may indicate that refugees increasingly are fleeing to (neighboring) countries with societal challenges and/or climatic exposure.

7.1.2. Regression

7.1.2.1. Fleeing from a country

The performance/predictive capacity of this model is quite strong, especially in recent years, with an R-squared value of approximately 0.7, indicating that it can be effectively used for predictions. The relatively good performance boosts confidence that no important factors were omitted from the analysis that should have been included. In the model, the most influential and significant climate event is unequivocally sea level rise, followed at a considerable distance by riverine flooding, and even further behind by droughts, which play only a minor role in the relation with refugees fleeing from a country.
To contextualize these findings, it is important to note that social factors, for example the lack of coping capacity consistently has far greater explanatory power (approximately twice), more influence, than the most influential and significant climatic indicator, which is sea level rise. Nonetheless, sea level rise remains a potential causal driver according to this analysis. Cyclones and coastal flooding do not appear to play any significant role in the regression association with refugees fleeing from a country.
Further research should focus on the causal criteria presented in the study by Cox Jr (2018) to determine if the climatic events listed above (most importantly sea level rise) are indeed causal drivers. According to this research, the causal drivers are expected to be -at least- ‘the lack of coping capacities’, ‘the lack of socio-economic development’ and ‘sea level rise’.

7.1.2.2. Fleeing to a country

The performance (or predictive power) of this model is lower compared to the origin case above, with an R-squared value below 0.6 in recent years, indicating a poorer fit. However, this value has been improving over time. The lower predictive power may be due to refugees fleeing their country of origin out of necessity without much consideration for where they end up, rather than choosing their destinations based on specific criteria, resulting in less structure in their destination decisions.
Cyclone intensity is the most important factor in relation with refugees fleeing to a country, with fewer refugees fleeing to countries with intense cyclones. To a lesser extent, droughts and riverine flooding also have impact, while other climate indicators are insignificant.
The finding on cyclone intensity does not necessarily mean that refugees take cyclone intensity into account when fleeing to a country. This is an empirical finding. It is possible that the destinations they reach are jut not in cyclone-prone areas, but this is likely not due to refugees deliberately avoiding such zones. Further research should investigate if this association is a real causal driving mechanism.
Droughts, riverine flooding, and lack of socio-economic development also influence the prediction of number of refugees fleeing to a country, with flooding and development mostly only have an effect when interacting with each other. All results suggests that fewer refugees flee to less developed or more climate exposed countries.
Sea level rise and coastal flooding show no consistent impact on predicting refugee destinations over the years, and the lack of coping capacities has only a minor influence. While sea level rise is a key

climatic indicator in predicting refugees fleeing from a country, it does not significantly affect their destination location.

In short, cyclone intensity is by far the most influential indicator associated with refugees fleeing to a country, consistently significant over the years. This could potentially be causal. Far behind cyclone intensity, riverine flooding has been the most consistently significant indicator over the years, slightly more significant than the lack of socio-economic development. Among social factors, the lack of socio-economic development has the most substantial influence. In nearly all cases where the lack of socio-economic development is significant, the interaction is between riverine flooding and socio-economic development, with these two indicators enhancing each other. In the last three years, droughts is also significantly associated with refugee destination locations.

7.1.3. Answer SQ2

Sea level rise, far followed by riverine flooding exposure and, to an even smaller extent, drought intensity exposure, could be significant causal drivers of people fleeing their countries, also in the future. These climatic events have significant associations with refugees fleeing from a country. However, to put this in perspective, social factors, such as the lack of coping capacities, are always bigger potential drivers. Further research is needed to determine whether these associations are causal. Meanwhile, the predictive capacity of the model estimating refugees fleeing from a country based on the indicators is quite strong.

For fleeing to a specific country, fewer strong driving possibilities occur; predictions are worse than in the 'fleeing from a country' case (less explanatory power). This suggests the need of refugees to leave the country of origin, without deliberately choosing to which country to flee to, in general. However, the 'unexpected' correlation signs are turned around in the regression analysis, and cyclone intensity, and to a lesser extent, riverine flooding and droughts, are (at least) associated with destination choices of refugees. No social factors have stronger associative links with refugee destinations than cyclone intensity.

7.2. Spatial clusters of climate exposure

This conclusion section addresses SQ3: How do countries cluster spatially over time, based on climate exposure? To answer this question, worldwide spatial clusters were identified, and changes in these clusters over the years were tested and compared with the original value changes.

7.2.1. Droughts

High drought intensity clusters are found in South America and, on a smaller scale, in Central Africa, while low scores are clustered in Central and (South) East Europe, including Turkey. Africa and Asia are key regions due to significant drought cluster differences between neighboring countries (high score countries surrounded by low score countries and vice versa). Statistically significant changes in drought clusters occurred in 2003 and 2022, but practically the changes are minor, except around Cambodia (more similar values), United Arab Emirates (more similar values), and Yemen (more dissimilar drought scores of neighbors).

However, no global statistical changes in original drought intensity scores were observed between 2003 and 2022, indicating stability over the last 20 years.

7.2.2. Riverine floodings

In Asia, most countries have high riverine flooding intensity, except for Nepal, Mongolia, and Kirghistan, which have low intensity but are surrounded by high-intensity neighbors, making Asia a hotspot for riverine flooding danger. In the Middle East, a smaller cluster of low-intensity countries is found around Saudi Arabia and Oman. In Africa, the Central African Republic forms the core of a high-intensity cluster, while Mozambique and Somalia have high exposure but are surrounded by low-intensity countries.

There were no significant changes in local cluster scores or original riverine flooding intensity scores between 2003 and 2022, indicating stability over the last 20 years.

7.2.3. Coastal floodings

Northern South America and Western Europe have high coastal flooding intensity countries, surrounded by similarly high-exposure countries. Eastern Europe features a low exposure cluster. In Africa, high-intensity clusters are on the southern West and East coasts, while low value clusters are in Central Africa and the South. The Middle East has a high-intensity cluster around Oman and Saudi Arabia. (South East) Asia is the most exposed, with only Uzbekistan as the core of a low-intensity cluster (North West part of Asia), and Nepal, Laos, and Mongolia as low-exposure countries surrounded by high-intensity neighbors.
No significant changes were found in cluster values or original coastal flooding intensity values between 2003 and 2022, indicating stability over the last 20 years.

7.2.4. Sea level rise

Over the years, Northern South America and Indonesia form clusters of consistently high sea level rise risk. In Africa, small low-risk clusters exist in South Africa and (North) West Africa, while Tanzania and Malawi are the core of high-risk areas surrounded by low-risk countries, making those two countries particularly unsafe. In Asia, Laos stands out as a low-risk core surrounded by high-risk areas, serving as a safe haven. The Middle East generally has low sea level rise risk, with Iran, Turkmenistan, and Uzbekistan forming a large low-risk cluster. Eastern Europe mostly has low risk, except for Ukraine, which is high risk, while Western Europe has a small high-risk cluster centered around Belgium.
No significant changes in cluster values or sea level rise numbers were observed between 2003 and 2022. Also, actual sea level rise numbers are similar between those years, indicating no sea level rise changes.

7.2.5. Cyclones

In North America, the United States and Mexico form a high cyclone intensity exposure cluster, while Canada has low exposure. In (South) East Asia, most countries have high cyclone risk, except for Nepal, Pakistan and Bhutan. More North and North West, Mongolia, Tajikistan, Kyrgyzstan and Kazakhstan, have low exposure but are surrounded by high-intensity neighbors. In Europe, clusters of low cyclone exposure are found in Western and Southeast Europe. In Africa, North and Central Africa often form clusters of low cyclone exposure. Southern and Central South America occasionally form a cluster of low exposure countries over the years.
Significant differences in cluster values between 2003 and 2022 were observed, with the United States and Mexico showing a large increase in more aligned values with their neighbors, while other changes were minor. Original, actual cyclone intensity numbers remained statistically similar between 2003 and 2022.

7.2.6. Answer SQ3

Asia, particularly South East Asia, is the region most exposed to nearly all the climate events, with numerous high-value clusters, mixed with little exposed countries. Africa follows with a lot of mixed high and low exposure clusters, while the Middle East has less (severe) climatic exposure clusters compared to Asia and Africa. Northern South America often shows highly exposed clusters, particularly in droughts, sea level rise and coastal floodings. Other regions either lack multiple significant high, low or mixed climate exposure clusters, or when such clusters are present, these countries are able to cope with the climate exposure.
To emphasis, South East Asia and Africa are the most significant regions for climate exposure, with mixed clusters of similar and dissimilar values. This variation is interesting for refugee movements, as refugees may encounter different levels of exposure even within neighboring countries. In contrast, other continents have less statistical significant clusters, or only have clusters of the same values, which is for example the case in Northern of South America. That is less interesting from a refugee movement perspective, because fleeing to neighboring countries then leads to the same climatic exposure.
Notably, climatic scores and clusters have remained largely constant over the years, with minimal changes identified.

7.3. Spatial clusters of refugee movements

This conclusion section addresses SQ4: How do countries cluster spatially over time, based on refugee movements? To answer this question, worldwide spatial clusters were identified, and changes in these clusters over the years were tested and compared with original value changes.

7.3.1. Origin

Over the past 20 years, Europe, the lower part of South America, Indonesia, and parts of Southern Africa have clusters of countries with low numbers of refugees fleeing from them. In contrast, the Middle East, Central-East Africa, and East/South Asia, have clusters with high refugee outflow. Interestingly, some countries in the Middle East and Central Africa with relatively few refugees fleeing from it are surrounded by high-outflow countries, making these regions crucial for studying refugee dynamics.
This pattern has remained consistent over the years, with in some years the North-West of South America showing low refugee outflows in Ecuador, surrounded by high-outflow countries. This also holds for Panama. Key regions of interest for studying refugees fleeing from countries are the Middle East, Central-East and South-East Asia, Central-East Africa, and Northern South America.

Analysis of cluster data from 2003 and 2022 reveals significant changes, with clusters in 2022 often showing more dissimilarity compared to 2003. The countries Afghanistan, Bosnia and Herzegovina, Burundi, Iran, Lebanon, Moldova, Syria, and Ukraine have shifted from similar (High surrounded by High) to more dissimilar (High-Low or Low-High) refugee fleeing clusters. This means that over the years, these countries have more dissimilar refugee fleeing values compared to their neighbors.
However, the overall difference in the actual number of refugees fleeing from origin locations between 2003 and 2022 is only statistically significant at the 6% level. Many countries have similar refugee fleeing values in both years, but some have seen significant increases or decreases. Countries with over 100,000 additional refugees fleeing away in 2022 compared to 2003 are Afghanistan, Cameroon, Central African Republic, Democratic Republic of Congo, Eritrea, Mali, Myanmar, Nigeria, Rwanda, Somalia, Sudan, Syria, Ukraine, and Venezuela. Nearly all African countries. Countries with over 100,000 fewer refugees fleeing from them in 2022 compared to 2003 are Angola, Azerbaijan, Bosnia and Herzegovina, Burundi, Croatia, Liberia, Serbia and Vietnam.

7.3.2. Destination

Clusters of countries refugees flee to, reveal notable trends. North America frequently hosts large numbers of refugees, while almost entire South America show clusters of countries with very few refugees receiving numbers. Notably, Panama and Mexico also form the core of a cluster indicating low refugee inflows. Similarly, in Europe, Western countries, especially Germany, host many refugees, while Eastern Europe has low refugee-hosting number clusters.
High refugee inflows are prominent in the Middle East and Asia, where also Low-High clusters are observed. These clusters indicate countries with low refugee arrivals surrounded by those with high inflows, which is due to the also high number of refugees fleeing from countries within these regions. Similarly, Northern Europe has a small Low-High cluster with as core Denmark, which hosts few refugees but is surrounded by countries with high refugee receiving numbers like Germany.
However, in the Middle East, there is also a High-Low cluster. In Africa, a similar pattern appears where one country may receive many refugees while being surrounded by countries with low inflows, suggesting regional (neighboring country) refugee movement dynamics.

There is no significant difference in clustering patterns between 2003 and 2022. However, there is a notable difference in destination location numbers, with many countries seeing more than 100,000 additional refugees fleeing to them in 2022 compared to 2003. These countries include Austria, Bangladesh, Belgium, Bulgaria, Cameroon, Chad, Czech Republic, Democratic Republic of Congo, Egypt, Ethiopia, France, Germany, Greece, Iran, Iraq, Italy, Jordan, Kenya, Lebanon, Malaysia, Mauritania, Moldova, Niger, Pakistan, Poland, Romania, Russia, Spain, Sudan, Sweden, Switzerland, Syria, Turkey, and Uganda. Conversely, countries with fewer than 100,000 refugees fleeing to them in 2022 compared to 2003 only include Armenia, China, Guinea, Nepal, Saudi Arabia, Serbia, Tanzania, and Zambia.

7.3.3. Answer SQ4

In short, Central and East Africa, the Middle East, and South East Asia are the most significant regions for studying refugee movements, as they exhibit both high refugee outflows and inflows. To a far lesser extent, this also applies to Northern South America. West Europe and North America have relatively high refugee receiving numbers. However, those continents are less critical to study, because there, countries can more easily cope with, and adapt to, climatic exposure. In addition, these countries usually have formal procedures for assisting refugees.

7.4. Climate exposure in origin and destination locations of refugees

This conclusion section addresses SQ5: What is the difference in climate exposure between origin and destination locations of refugees worldwide, over years? To answer this question, global bivariate choropleths were constructed, and the top 25 refugee origin and destination countries were visualized using these plots. Statistical differences between the top countries were tested, and individual visual comparisons were made.

7.4.1. Droughts

In general, top origin locations have statistically higher drought exposure than top destination locations, on average.
In 2022, refugees fleeing to the top destination countries Congo, Ethiopia, Cameroon, Sudan, and Chad face high drought intensity, while other top receiving countries have low drought scores, except Russia. Refugees fleeing from top origin countries Mali, Nigeria, Burundi, Rwanda and South Sudan, which have lower drought intensity, should receive support for the increased drought exposure in their destination countries. A similar pattern was observed in 2016, with as addition that refugees fleeing from top origin country Côte d'Ivoire to the above listed top destination countries, end up in far more drought exposed countries. And Yemen adds to the list of top receiving country with high drought exposure.
In the Middle East, drought intensity is similar in both origin and destination locations. In Asia, countries like Afghanistan, Bhutan, Sri Lanka, Myanmar, and Bangladesh have low drought intensity, whereas Vietnam, India, and China have high drought intensity. From these countries, only Bangladesh, India and China are top destination countries. Refugees fleeing to India and China face at least equal or worse drought conditions.

7.4.2. Sea level rise

Top destination countries generally face higher sea level rise risk than top origin locations, on average. Until 2020, most top origin locations have low sea level rise risk, except Vietnam and Sri Lanka. In latest years, Turkey, Colombia, Venezuela, and China, alongside Cameroon, also have relatively high risk scores from the countries in the top 25 origin locations. Also from this list, China, Turkey, and Venezuela showing increasing sea level rise risks.
In the early 2000s, for the top destination locations, Congo and India had high sea level rise risk, with Algeria and Saudi Arabia experiencing moderate risk. The Netherlands and the UK, along with Canada and the US later on, are not mentioned due to their effective coping/adapting mechanisms and formal procedures to help refugees. By 2010, Ecuador, Congo, Kenya, Cameroon, Venezuela, and Tanzania had high or extreme sea level rise risks. India and China also had high risks. In 2016, Egypt and Uganda joined the list of top destination locations, and have remarkable high sea level rise risk. Bangladesh had an above moderate risk. Recent years show no change in high-risk top refugee destination countries.

7.4.3. Riverine floodings

Overall, on average, top destination countries are equally exposed to riverine flooding intensity as top origin countries.
In Africa, top origin countries with low riverine flooding intensity include Angola, Rwanda, Burundi, Somalia, Ethiopia, Central African Republic, Sierra Leone, Liberia, Côte d'Ivoire, Cameroon, and Eritrea. Refugees from these countries face higher exposure in top destination locations in the region: Egypt, Chad, Mali, Nigeria, and Sudan, which have worse flooding intensity. Congo, South Sudan, and Cameroon, with moderate exposure over the years, are also top destinations. Other popular des-

tination location in Africa are Zambia, Tanzania, Uganda, Kenya, Guinea, Yemen, and Algeria, which experience low riverine flooding intensity.
In Asia, top origin countries are Afghanistan, China, Russia, Bhutan, Myanmar, Pakistan, Vietnam, and Sri Lanka, with all having high riverine flooding intensity, except for Afghanistan, Bhutan and Sri Lanka. Top destinations in Asia are India, Bangladesh, China, and Pakistan, also have high flooding intensity. This means Asian refugees fleeing within the region from top origin to top destination locations, always face worse or similar riverine flooding intensity. However, refugees fleeing from Asia to the Middle East generally encounter lower riverine flooding intensity.
In the Middle East, only Iraq has high riverine flooding exposure. Popular top destination countries in this region are Turkey, Syria, Iran, Lebanon and Jordan, which have low to moderate flooding intensity. Refugees fleeing from Iraq to these countries will experience less riverine flooding intensity. Iran and Afghanistan are top origin countries with low riverine flooding intensity, surrounded by top destination countries with high exposure: Iraq and Pakistan. Jordan and Egypt are nearly neighbors and both top destination countries, with Egypt having the highest riverine flooding intensity in the world and Jordan having very low intensity.
In South America, Colombia and Venezuela are top origin countries (where many refugees flee from). Top destination countries like Ecuador and Venezuela have equal or less exposure to riverine flooding intensity, so the negative impact on refugees is minimal.

7.4.4. Coastal floodings

Overall, on average, destination countries are more exposed to coastal flooding than origin countries.
In Africa, all top origin countries have low coastal flooding intensity, which is also true for most top destination countries, except Egypt, which has high coastal flooding risk. Refugees fleeing to Egypt face significant coastal flooding risk. (In this region, top origin and destination countries also partly overlap.)
In South America, refugees fleeing from Colombia to Venezuela or Ecuador face higher coastal flooding risk in Venezuela compared to Ecuador and Colombia. Support should focus on helping refugees understand and avoid high exposure coastal regions in Venezuela.
In Europe and North America (top destination regions), coastal flooding intensity is moderate to high, but these countries have effective coping mechanisms and formal procedures for refugees, making them less of a focus for this study. Note further that regions in Oceania, and Indonesia, never are in the top origin or top destination list.
In the Middle East, there is considerable overlap between top origin and destination countries. Top origin locations are Turkey, Iraq, Iran, Syria. Top destination locations are Turkey, Jordan, Syria, Lebanon and Iran. So, Turkey, Syria and Iran are consistently both top origin and destination countries. Only Iraq and Iran have above moderate coastal flooding risk, other countries face low exposure. Refugees from other top origin countries in the Middle East who move to Iran face increased coastal flooding risk.
In Asia, there is less overlap between top origin and destination countries, except for China and Pakistan, which are consistently both. Russia is also a top origin and sometimes top destination country. Further, consistently top origin countries are Afghanistan, Sri Lanka, Myanmar, Pakistan, and Vietnam. Additional top destination countries are India and Bangladesh. Refugees fleeing from Afghanistan and Sri Lanka, with very low coastal flooding exposure, face increased exposure in top destination countries. Refugees fleeing from other top origin countries China, Myanmar, and Vietnam, which have extremely high coastal flooding intensity, do not face worsened conditions. Pakistan and Russia have below moderate coastal flooding exposure, while most other top destination countries have very high exposure. Hence, refugees fleeing from Pakistan, Russia, Afghanistan, and Sri Lanka will face increased coastal flooding exposure when moving to Bangladesh, India, or China. In general, apart from Afghanistan, Sri Lanka and Pakistan, both top origin and destination countries are in highly coastal flooding exposed places.

7.4.5. Cyclones

Overall, average statistical differences in origin and destination countries are inconclusive regarding this form of climate exposure.
Notably, China is the only top origin and destination country with high cyclone intensity exposure. Refugees fleeing from China to other top destination countries will experience lower cyclone intensity. Refugees fleeing from other top origin countries are likely to arrive in destinations with similar or lower cyclone intensity exposure, except for fleeing to China and the United States, which have higher

cyclone intensity. Other top destination countries have low cyclone intensity exposure. This pattern remains consistent over the years: all other top origin and destination locations have low cyclone intensity exposure.

7.4.6. Answer SQ5

This answer addresses top origin-destination country-specific climate exposure differences within the three regions with the most refugee movements and (varied) climate exposures (as identified by answering SQ3 and SQ4): Central and East Africa, the Middle East, and (South East) Asia. Table 7.1 enables the identification of specific differences in exposure per region and for individual top origin and destination countries within those regions. These findings are used to base regional specific policy recommendations upon in the discussion (see Chapter 8).
In general, climate exposure is not changing significantly in countries over the last 20 years, according to this analysis.

Table 7.1: Combined exposure (L=Low, M=Moderate, H=High) overview of top origin-destination countries

Top Origin	Droughts	Sea level rise	Riverine floodings	Coastal floodings	Cyclones	**Top Destination**	Droughts	Sea level rise	Riverine floodings	Coastal floodings	Cyclones
Africa											
Sierra Leone	L	L	L	L	L	Algeria	M	M	L	L	L
Liberia	H	L	M	L	L	Guinea	L	L	L	M	L
Angola	H	M	L	L	L	Egypt	L	H	H	H	L
Dem. Rep. Congo	H	L	M	L	L	Dem. Rep. Congo	H	L	M	L	L
Burundi	L	L	L	L	L	Zambia	H	L	L	L	L
Rwanda	L	L	L	L	L	Tanzania	L	H	L	L	L
Somalia	H	L	L	L	L	Uganda	L	L	L	L	L
Ethiopia	H	L	L	L	L	Ethiopia	H	L	L	L	L
Eritrea	M	L	L	L	L	Rep. Congo	L	H	H	L	L
Sudan	M	L	H	L	L	Sudan	M	L	H	L	L
Central African Rep.	H	L	M	L	L	Kenya	L	H	L	L	L
Chad	H	L	H	L	L	Chad	H	L	H	L	L
Mali	L	L	H	L	L	Yemen	H	L	L	M	L
Côte d’Ivoire	L	L	L	L	L						
Nigeria	L	M	H	L	L						
South Sudan	L	L	M	L	L	South Sudan	L	L	M	L	L
Cameroon	M	H	M	L	L	Cameroon	M	H	M	L	L
(Greater) Middle East											
Turkey	L	H	L	L	L	Turkey	L	H	L	L	L
Iraq	L	L	H	H	L	Iraq	L	L	H	H	L
Iran	L	L	M	H	L	Iran	L	L	M	H	L
Afghanistan	L	L	L	L	L	Saudi Arabia	H	H	L	M	L
Azerbaijan	L	L	L	L	L	Jordan	L	L	L	L	L
Syria	L	M	L	L	L	Syria	L	M	L	L	L
						Lebanon	L	M	L	L	L
						Armenia	L	L	L	L	L
(South East) Asia											
China	H	H	H	H	H	China	H	H	H	H	H
Bhutan	L	L	L	L	L	India	H	H	H	H	L

Continued on next page

Top Origin	Droughts	Sea level rise	Riverine floodings	Coastal floodings	Cyclones	**Top Destination**	Droughts	Sea level rise	Riverine floodings	Coastal floodings	Cyclones
Myanmar	L	L	H	H	L	Bangladesh	L	H	H	H	L
Vietnam	H	H	H	H	L						
Sri Lanka	L	H	L	L	L						
Pakistan	L	L	H	M	L	Pakistan	L	L	H	M	L

7.5. Associations between climate exposure and refugee movements

With the answers to the sub-questions, the main question can be answered: How is climate exposure associated with global refugee movements?

Firstly, sea level rise, riverine floodings and droughts, in that order of importance, are significant in their association with refugees fleeing from a country. This does not necessarily mean the effect is causal. Secondly, cyclone intensity, riverine floodings and droughts, in that order of importance, are associated with refugees fleeing to a country. This does not necessarily mean the effect is causal.

In addition, the most severe climate exposure is experienced in (South East) Asia, followed by Central and East Africa. Also, refugee movements are very large in those regions. Highest refugee movements are in the Middle East and Central East Africa, and thereafter high refugee in- and outflow is seen in South East Asia. This means that climate clusters of high and low exposed countries, and large refugee movement clusters in the world, align relatively closely.

Finally, the specific differences in climate exposure between top origin and destination countries are identified per relevant region in Table 7.1.

8 Discussion

8.1. Contribution to literature

8.1.1. Knowledge gaps

In this section is reflected on the introduction, and therein specifically the literature study for the identification of knowledge gaps. This reflection identifies where this thesis aligns with other scientific studies and where it diverges.

Findings in this thesis align with those of Luetz (2018) in that climate conditions cannot be isolated (yet) as definitive cause of movements, but that it is impossible to dismiss them as a contributing factor. This is because of the significant associations found in this thesis, for example for sea level rise and riverine floodings with refugees fleeing from a country. Related to this, Han et al. (2024) find an interaction between climate and migration in their local study in China. Results of this thesis also find significant interactions, between, for instance, riverine floodings and lack of socio-economic development, especially in earlier years, having a positive effect on refugees fleeing from a country. In addition, for refugees fleeing to a country, the interaction between riverine flooding and lack of socio-economic development is even more frequently significant over the years, also in latest years, suggesting that less people flee to a country with higher combined scores on riverine floodings and lack of socio-economic development. What this thesis adds, is that also interactions between climate indicators themselves are sometimes significant. For example, the interaction between droughts and riverine floodings was found to be statistically significant in associating more refugees fleeing away when combined exposure to these indicators is high, enhancing each other. Even more, this thesis adds higher order effects as most significant variables in associations with refugee movements. Frequently, higher order effects of climate indicators as well as social indicators were found to be statistically significant in the association with refugees fleeing from/to a country. This is to some extent contradicting the claim by Byravan and Rajan (2022) that there is no direct relation between displacement and climate exposure. According to this thesis, the possibility for a direct (causal) relation is not ruled out. However, effects will be smaller than the effects of social effects (including, for instance, conflicts), confirming thereby research of Joseph and Wodon (2013). Therefore, this thesis does not confirm that climate events “are a more important driver than income and political freedom at origin together” (Wesselbaum & Aburn, 2019). With all of the above, the knowledge gap regarding the lack of (quantitative) information on the form and magnitude of relations between climate conditions and refugee movements is shrinking.

In Table 7.1 is presented that climate exposure differs heavily per country, even in the same region. That is thereby a support for the research of Mueller et al. (2020) who indeed claimed that the effect of climate on migration differs per country. With Table 7.1 can also be identified that people sometimes move to locations where they are more exposed to climate exposure than at their origin locations (Benveniste et al., 2020). This thesis does not confirm that people who move from developing regions move to areas where they are less affected by climate conditions compared to where they came from, as indicated by Benveniste et al. (2020). Namely, this thesis identified relatively large overlap in top origin and destination location of refugees, because of which refugees not necessarily end up in less exposed areas (see Table 7.1). With the arguments in this paragraph, the knowledge gap on the lack of global (worldwide) analysis between climate exposure and refugee movements is filled.

In this thesis, there is no association found between cyclone intensity and refugees fleeing from a country. (There only is association between refugees not fleeing to cyclone intensity exposure destinations.) Paglino 2024 found that excess migration after tropical storms is rare and generally fails to reduce the number of people at risk of experiencing future events. This finding is thereby confirmed by this thesis: cyclone intensity is in general not associated with refugees fleeing from a place. With this reflection, the knowledge gap on the lack of findings for refugees compared to migrants is narrowed. Also, cyclone intensity knowledge is added to the gap on the lack of quantitative relationships between climate conditions and refugee movements.

Two studies researched the relation between droughts and movements. One study identified a significant relationship between drought at origin and moving (Karutz & Kabisch, 2023). Paradoxically, according to their study, migrants affected by droughts at origin face increased flood risk at destination. The study of Sivisaca et al. (2021) find that the presence of a drought in a municipality of origin significantly reduces migration. According to that study, those people cannot migrate due to the lack of financial means, as a result of agriculture loss due to the droughts.
This thesis finds droughts only having a minor association with refugees fleeing from a country (only significant in the relation with riverine floodings). In latest years, droughts influence is higher in the association with not fleeing to a country. Whether refugees who flee from drought areas end up in increased flood risk locations, depends on the specific destination as can be seen in Table 7.1. The reflection and findings in this paragraph contribute to filling two gaps: 1) the lack of refugee studies regarding climate and movement, compared to migrant studies; 2) the lack of quantitative insights on relations between climate conditions and refugee movements.

Wesselbaum (2021) showed that temperature shocks have a direct non-linear effect on migration. Again, this finding is related to general migration and is not specific to refugees.This thesis closes that gap. While temperature shocks are not specifically studied in this thesis, many non-linear association effects are identified for other climate indicators in their relation with refugee movements. These indicators include sea level rise and riverine floodings for refugees fleeing from a country, and cyclone intensity and droughts for refugees fleeing to a country. The latter also contributes to closing the gap regarding the lack of qualitative insights into relations between climate conditions and refugee movements, by identifying the non-linear relationships described above.

Finally, the scientific study of Leal Filho et al. (2023) highlights the need for governments and international organizations to better understand the complex linkages between climate and movements, which is what this thesis aims to make clearer. Even more, their paper points out the urgent needs for policies that support climate migrants. Such a call for adaptation strategies is also made by Hauer et al. (2024). This thesis addresses this in the ‘policy recommendations’ section below.

8.1.2. Additional reflection

Two studies warrant specific attention in this discussion. These numerical studies examine climate exposure of refugees specifically, rather than migrants, and evaluate multiple refugee locations. A comparison between those studies and this thesis will be made across different themes below.

8.1.2.1. Geographical scope

This thesis expands the scope of analysis, compared to the two other studies. Namely, unlike research of Owen et al. (2023) and Fransen et al. (2023), which primarily examines refugee populations at a regional scale and focuses mainly on destination locations, this study considers global refugee movements and emphasizes the differences of both origin and destination locations. In relation to this, this thesis incorporates global climatic exposure maps for refugees alongside the regional border exposure map in East Africa of Owen et al. (2023). This broader perspective allows for a more comprehensive understanding of global displacement patterns, but neglects regional details.

In addition, the research of Owen et al. (2023) notes that the variable selection for using their framework/model must be adapted to different geographies, where distinct drivers of exposure or hazards may exist. This thesis proposes the use of Table 7.1 as a guide for determining the appropriate inclusion of climatic exposure variables in a given study, providing a structured approach for adapting their framework to different country contexts.

Lastly, Owen et al. (2023) and Fransen et al. (2023) find that refugee camps are more exposed to climate hazards compared to other regional/national locations. This thesis identifies that, on a global scale, refugees do not necessarily end up in more or less exposed areas than those from which they fled. It depends on the specific location differences (see Table 7.1). This highlights the need for strategic climate adaptation investments tailored to specific refugee locations.

8.1.2.2. Comparison of methods

The methods of both studies differ substantially from each other and from this thesis, yet they share important points of reflection. The paper of Fransen et al. (2023) primarily employs descriptive statistics, such as means, standard deviations, and the signal-to-noise ratio, to compare refugee camp conditions with broader national averages and other camps. The analysis in that study tracks changes in exposure over time visually in a line graph. This thesis examines country differences over years also visually using scatter plots and global maps, but in addition uses correlation coefficients and statistical difference tests, thereby integrating both visual and statistical methods to analyze differences in refugee exposure over years. In short, richer statistical tools are used in this thesis.

The paper of Owen et al. (2023) constructs exposure hotspots, a technique also adopted in this thesis, although they do not use LISA cluster values. So, there is alignment with the type of spatial analysis done in this thesis, though the technique differs. They use a ‘self-constructed’ exposure formula from which hotspots are derived. It is not explained how their formula is supported by existing literature, nor is there an explanation of how it differs from existing literature, which makes it difficult to reflect on it.

8.1.2.3. Climatic events

A key distinction between the studies lies in the selection of climatic-related hazards and indicators. Owen et al. (2023) identify droughts and floods as most central documented events, while this thesis broadens that scope to include also cyclones and sea level rise. However, drivers of exposure in refugee camps are also precipitation, humidity, surface soil moisture, and temperature differences, and again drought severity is mentioned (Owen et al., 2023). This thesis did not include any of these drivers besides droughts. Fransen et al. (2023) examine a range of slow-onset (temperature, rainfall) and rapid-onset events (heatwaves, coldwaves, extreme rainfall), with no explicit overlap with the climatic events considered in this thesis. However, in their paper extreme rainfall is argued to be a proxy of floodings. This diversity in hazard selection among the studies highlights the need for richer, more comprehensive climate data in future research, a point that all three studies implicitly underscore.

Furthermore, Owen et al. (2023) integrate different normalized climatic exposures into a single exposure score, a significant methodological departure from this thesis and the paper of Fransen et al. (2023), where exposure is measured separately for each climatic indicator. This difference allows this thesis to distinguish the individual impacts of various climate events more clearly, whereas the aggregated exposure score presents a more generalized view. Further, in this thesis, different categorized exposure scores are combined using the geometric mean for each separate climatic indicator, unlike the non-categorized real value data per indicator of Fransen et al. (2023). The latter reflects better on actual, observable events per indicator, while the former allows easier comparison of exposure scores over different indicators.

8.1.2.4. Societal factors and interactions

This thesis not only examines climatic factors, but also incorporates societal components to better understand the drivers of refugee movements. These factors are integrated into a predictive model and are used to assess the relative importance of climate compared to social indicators, offering a nuanced approach to understanding and forecasting refugee flows.

Additionally, the article of Fransen et al. (2023) calls for further exploration of the interaction effects between climate-related hazards and societal factors. This thesis integrates such interactions into a regression model. Notably, this thesis finds thatthe interactions between climate and social indicators in the relation with refugees fleeing from a country are not consistently statistically significant over time. However, the interaction between riverine flooding and a lack of socio-economic development is significant for refugees fleeing to countries, creating a 'turbo effect' where higher values in both factors are associated with a stronger decrease in refugee inflows. This insight enriches the discussion on how climatic and social drivers can mutually reinforce each other.

8.1.2.5. Call for action

This thesis argues for targeted financial investments in specific top refugee movements countries where exposure to sea level rise, riverine flooding and drought, in that order of importance, is high (see therefore Table 7.1). This adds specific indicators to the call for more climate adaptation measures in the research of Fransen et al. (2023). The next section will elaborate on this matter.

8.2. Policy recommendations

Many refugee organizations and (local) Non Governmental Organizations (NGOs), as well as the UNHCR, are actively working to strengthen climate adaptation among refugees and come up with solutions; advance changes in policies, and forge innovative protection pathways (Climate Refugees, n.d.; Refugees International, n.d.; United Nations High Commissioner for Refugees, n.d.). UNHCR has even created a platform in which governments, NGOs and other stakeholders participate in "sharing best practices and resources in order to enhance access to climate actions resources and funding for refugee and host communities". Regarding this, for example, they work on a project in South Sudan to enhance climate resilience.
This thesis helps those organizations to focus on specific climate exposures that are associated most with refugee moments, particularly in regions with the highest climate exposure and refugee activity, as well as highlighting country specific differences in top origin and destination locations within climate exposed regions. This research helps to tailor the already limited available (financial) means of those organizations.

The answer to SQ2 helps determine which specific climate exposures should be prioritized for tailored resources. Namely, some exposures were more associated with refugees fleeing from/to countries than others. Sea level rise is the most driving climatic factor that warrants further research, as it could be a potential causal factor for refugees fleeing from a country. This holds to a lesser extent for riverine flooding and droughts. Coastal flooding and cyclones do not have a potential causal effect, although cyclone intensity is associated with less refugees fleeing to a country. The latter is recommended to research further. The focus in top origin countries on reducing sea level rise risk may help in reducing refugee fleeing from countries. This also holds for riverine floodings and to a lesser extent for droughts. Therefore, the focus should be on sea level rise and riverine flooding in terms of money and future research. Those climate indicators are promising to be potential causal and with addressing those, refugee movement numbers can potentially be decreased.

The answers to SQ3 and SQ4 help to address the specific regions in the world for which support should be priority, based on climate exposure and refugee movements. Overall, for policy makers in international public organizations and especially NGOs, with both having limited budgets, it is recommended to focus on climate adaptation measures for refugees only in the regions South-East Asia, Central-East Africa, the Middle East and to a lesser extent Northern South America. In those regions, refugee movements are the most intensive, and climate exposure is present and sometimes severe. Also, a lot of differences in climate exposure and refugees fleeing to/from countries are observed in those regions, making the movements particularly interesting. Specifically, different climate exposures are mapped to the regions where the focus should be. Cyclones primarily affect South East Asia. Coastal flooding is a significant issue in East Africa, South East Asia, Northern South America, and the Middle East. Riverine flooding impacts South East Asia, Central Africa (to a lesser extent), and the Middle East (to a lesser extent). Sea level rise is a concern in South East Asia, Northern South America, and (South) East Africa. Droughts are prevalent in Central Africa, Northern South America, and, to a lesser extent, the Middle East and East Asia.
To narrow this list even further, funds should be allocated to mitigate climate exposures for refugees through climate adaptation measures in the three primary clusters of refugee movement areas and their highest -within region- exposure differences: droughts in Central Africa; sea level rise in East Africa; cyclones and sea level rise in East Asia, as well as all types of flooding in Asia, and coastal flooding in the Middle East. The emphasis should be on adapting to sea level rise and riverine flooding risk due to their significant associations with refugees fleeing from countries, as explained earlier. Thereafter,

resources should be directed towards addressing droughts. Although the Middle East and especially Africa face considerable climate exposure, South-East Asia is the region with the highest combined climate exposure. It should be clear that NGOs and other international organizations should prioritize their climate adaptation efforts in the regions presented above.

The answer to SQ5 (see Table 7.1) helps in selecting priority countries within the relevant regions identified above. In Africa, Egypt is the only top destination country with high sea level rise, high riverine flooding intensity and high coastal flooding intensity. Therefore, Egypt should be a priority country for refugees arriving there from top origin countries. Besides, regarding riverine floodings, should be focused on the following top destination countries: Republic Congo (also for sea level rise risk in this country), Sudan and Chad. For sea level rise only, Cameroon, Kenya and Tanzania are priority countries. In terms of droughts, the priority countries are Democratic Republic Congo, Zambia, Ethiopia, Chad en Yemen. In conclusion, with limited resources, the top priority destination countries are: Egypt (for sea level rise, riverine and coastal floodings); Republic Congo (for sea level rise and riverine floodings) and Chad (for droughts and riverine flooding). Other top destination countries only experience high exposure to one of the climate events.
In the Middle East, there are only two top destination countries with high exposure to more than one indicator: Iraq (for riverine and coastal floodings) and Saudi Arabia (for droughts and sea level rise). Therefore, those countries are top priority in this region. Iran, as top destination country, only has high climate exposure to coastal floodings. Turkey only has high exposure to sea level rise. The other top destination countries in this region have low to moderate exposure scores and are therefore no priority.
In South-East Asia, all top destination countries are priority, except for Pakistan. Pakistan experiences only high exposure to riverine floodings and further has low to moderate exposure to the other climatic indicators. However, China, India and Bangladesh have high exposure scores to at least three climatic indicators. China has even high exposure to all of the climate conditions: cyclone intensity, drought intensity, sea level rise intensity, riverine and coastal floodings. India has the same score, except for not having high exposure to cyclone intensity, but rather low. Bangladesh has high exposure to sea level rise, riverine and coastal floodings. To conclude, China, Bangladesh and India are top priority countries in this region to allocate (financial) resources to for climate adaptation regarding the identified exposures.
Those climate exposure situations influence how refugees can produce food; where they can live; which clothes are suitable in which circumstances. The above identifications help to prioritize and determine a specific way to help refugees to adapt to climate exposure.

Lastly, policy makers in governments and NGOs can use the latest year regression equation fitted in this report to predict the numbers of refugees fleeing from a country, based on social situation and climate factors. This helps them in planning refugee moments for the countries mentioned above, enabling predictions or at least estimates of the number of future refugees and the corresponding level of assistance required.

8.3. Limitations

In the discussion of limitations, several important areas warrant attention. First, the UNHCR refugee movement data used in this study is limited to ‘stock data’, meaning it only reports the number of refugees leaving or arriving in a country each year, without detailing the specific flows between countries. This limitation prevents understanding where refugees in a particular country came from or where they subsequently moved. Unfortunately, no richer global data is publicly available that is also complete.
Also, the study should be extended by incorporating additional refugee data, particularly data on transit refugees—those who flee from a country other than their birth country. The current study assumes that refugees flee directly from their country of birth to their destination, but in reality, refugees often transit through other countries. Future research should address this gap by including data that captures these more complex movement patterns.
Further, a major limitation of UNHCR refugee data is that it is incomplete for earlier years, especially before 1951, but also before 1994 (Fransen & De Haas, 2022). This research addresses this limitation by using only data from the period 2003-2022. However, since 2007, UNHCR’s refugee statistics include ‘refugee-like’ situation numbers (Fransen & De Haas, 2022). This biases the results regarding increasing trends in refugee numbers between early 2000s and 2022.

Moreover, refugee definitions and administrative practices vary across countries. This was noted in the Introduction (see Chapter 1, where multiple descriptions of the refugee definition, ranging from narrow to broad, are presented. This means that someone may be considered a refugee in some countries but not in others. Related to this is the argument made by Marbach (2018), who state that UNHCR refugee numbers are not entirely accurate and that the UNHCR provides little information on how it collects data.

In addition, this thesis uses multiple regression to identify relationships between climate exposure and refugee movements. It thereby assumes cross-sectional data, on which a different equation is estimated in every year during the period 2003-2022. This approach estimates general -and not country specific- effects. This is valuable when one wants to measure overall global impacts of climate on refugee movements. One could also argue to better exploit the panel structure (observations on multiple countries over multiple years) of the available data by conducting fixed panel regression. This can be in the form of time fixed effect and individual specific effects. With the latter, individual country differences can be identified regarding the climate-refugee relation. With the former option, time fixed effects, in which time differences in relations can be identified, it would be interesting to research whether the same conclusions can be drawn about significance over years as in this study with a different equation with different coefficients for each year. In general, panel regression offers the advantage of reducing omitted variable bias, because the model can better control for individual year country characteristics. In other words, one can control for unobserved heterogeneity, and potentially make causal inference easier (Ba, Berrett, & Coupet, 2021; Liu, Liu, & Shi, 2020). However, this also comes with a couple of downsides (which are advantages of 'normal' multiple regression). Panel regression is more complex to specify and interpret, especially the regression equations are usually very large, with many factors to control for country and time specifics in one formula. This makes it less intuitive and transparent.
Another option, specifically to better identify potential causal relationships (inference), is incorporating Directed Acyclic Graphs (DAG) and Structural Equation Modeling (SEM) in further studies. This would enhance the robustness of the findings regarding the link between climate exposure and refugee movements.

This study is limited to correlation and regression analysis for determining the importance of climate indicators in predicting refugee movements. Further research should explore the use of Machine Learning models for predicting refugee movements, combined with Explainable AI (XAI) to assess the importance of various factors included in the analysis. This approach could verify whether the conclusions drawn align with those from the regression models used in this study. Additionally, it would be valuable to examine whether the predictive performance improves when using methods such as neural networks compared to regression models.

Furthermore, this study uses limited climate data. Future research should consider incorporating additional climate indicators, namely temperature and precipitation, to expand the scope of the analysis and to see how that influences the results.
Besides, the climate exposure scores used in this study solely originate from the World Risk Index. In general, the variability of the World Risk Index country scores over year is very limited, particularly for the climate indicators as shown in the results of Chapter 5 and 6. This raises concerns about the validity of the data, as noticeable climatic changes have occurred over the past 20 years (Aboye, Kinsella, & Leza, 2022; Y.-n. Chen, Li, Xu, & Hao, 2007; George, 2010).
Even more, the World Risk Index employs its own calculations to assign weighted scores to climatic and social indicators, each derived from multiple sub-indicators. For instance, the 'state and government' score is based on the geometric mean of scores on 'democratic principles' and 'governmental responsibilities'. Democratic principles, in turn, has a score equal to the geometric mean of 'control of corruption' and 'rule of law'. The two latter scores are gathered through another source (in this case the World Bank). This process reflects the extensive data manipulating through different levels, potentially leading to biased results. The article of Welle and Birkmann (2015) acknowledges that the calculation of exposures based on different databases with different thresholds leads to uncertainties in the World Risk Index. While the creators of the World Risk Index are transparent about their calculations and data sources, they rely heavily on the geometric mean to derive indicator scores. However, the geometric mean has notable drawbacks, including its sensitivity to both probability distributions and skewness, as well as its bias compared to most other central tendency measures (Vogel, 2022). In simpler terms, the geometric mean lacks a clear and concise physical interpretation compared to other measures, such

as the median, which could affect the results observed in this study. An alternative approach could involve using an Empirical Risk Index (alongside the World Risk Index) (Feldmeyer, Nowak, Jamshed, & Birkmann, 2021) or the INFORM index (Birkmann et al., 2022) to evaluate how different exposure scoring mechanisms influence the results.
Further research should outline differences when conducting the analyses in this study with: 1) incorporating additional climatic indicators; 2) utilizing alternative climate data sources beyond (only) the World Risk Index and 3) employing calculation metrics other than the geometric mean.

Lastly, this research solely takes a global approach. Focusing on local studies for key top origin-destination countries identified in this research (see Table 7.1) helps to detect climate exposure differences within countries, highlighting specific local areas of interest. This way, refugees can be better assisted on the ground. Birkmann et al. (2022) confirm that global approaches, like the World Risk Index with aggregated information, need to be complemented with context specific data at sub-national and local scales.
Research of Welle and Birkmann (2015) and de Almeida et al.(2016) suggest that the modular structure of the World Risk Index makes it adaptable for use at a more localized levels, such as regional scales, compared to its general global application. However, although many local studies employ a similar concept to the World Risk Index, accurate local data is often not available, which limits proper application (Ramli, Alias, Yusop, & Taib, 2020). Therefore, an alternative exposure calculation method may be required for local studies, depending on data availability in the specific countries of interest.

References

Abeysingha, N., Kularathna, K., Bandara, A., & Ray, R. L. (2023). Assessment of extreme rainfall through statistical process control-i chart. *Journal of Earth System Science*, *132*(1), 15.

Aboye, A. B., Kinsella, J., & Leza, T. (2022). Major climatic changes experienced by farm households: Evidence from the lowlands of southern ethiopia. *Advancements in Agricultural Development*, *3*(1), 87–102. doi:10.37433/aad.v3i1.163

Alashan, S. (2020). Combination of modified mann-kendall method and şen innovative trend analysis. *Engineering Reports*, *2*(3), e12131.

Alexopoulos, E. C. (2010). Introduction to multivariate regression analysis. *Hippokratia*, *14*(Suppl 1), 23.

Armstrong, R. A. (2019). Should pearson's correlation coefficient be avoided? *Ophthalmic and Physiological Optics*, *39*(5), 316–327.

Aydin, N. Y., Yigitbasi, E., Casali, Y., & van Wee, B. (2023). Assessing the evolution of educational accessibility with self-avoiding random walk: Insights from helsinki. *Applied Network Science*, *8*(1), 55.

Ba, Y., Berrett, J., & Coupet, J. (2021). Panel data analysis: A guide for nonprofit studies. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 1–16.

Benveniste, H., Oppenheimer, M., & Fleurbaey, M. (2020). Effect of border policy on exposure and vulnerability to climate change. *Proceedings of the National Academy of Sciences*, *117*(43), 26692–26702.

Birkmann, J., Jamshed, A., McMillan, J. M., Feldmeyer, D., Totin, E., Solecki, W., … Alegría, A. (2022). Understanding human vulnerability to climate change: A global perspective on index validation for adaptation planning. *Science of The Total Environment*, *803*, 150065. doi:https://doi.org/10.1016/j.scitotenv.2021.150065

Bonita, J., Ambolode, L., Rosenberg, B., Cellucci, C., Watanabe, T., Rapp, P., & Albano, A. (2014). Time domain measures of inter-channel eeg correlations: A comparison of linear, nonparametric and nonlinear measures. *Cognitive neurodynamics*, *8*, 1–15.

Byravan, S., & Rajan, S. C. (2022). Cross-border migration on a warming planet: A policy framework. *Wiley Interdisciplinary Reviews: Climate Change*, *13*(2), e763.

Casali, Y., Aydin, N. Y., & Comes, T. (2024). A data-driven approach to analyse the co-evolution of urban systems through a resilience lens: A helsinki case study. *Environment and Planning B: Urban Analytics and City Science*, *0*(0), 23998083241235246. doi:10.1177/23998083241235246. eprint: https://doi.org/10.1177/23998083241235246

Chen, H., & Sun, J. (2020). Increased population exposure to precipitation extremes in china under global warming scenarios. *Atmospheric and Oceanic Science Letters*, *13*(1), 63–70.

Chen, Y.-n., Li, W.-h., Xu, C.-c., & Hao, X.-m. (2007). Effects of climate change on water resources in tarim river basin, northwest china. *Journal of Environmental Sciences*, *19*(4), 488–493. doi:https://doi.org/10.1016/S1001-0742(07)60082-5

Chikodzi, D., Nhamo, G., Dube, K., & Chapungu, L. (2022). Climate change risk assessment of heritage tourism sites within south african national parks. *International Journal of Geoheritage and Parks*, *10*(3), 417–434.

Climate Refugees. (n.d.). What We Do. Retrieved from https://www.climate-refugees.org/foundingvision

Cox Jr, L. A. (2018). Modernizing the bradford hill criteria for assessing causal relationships in observational data. *Critical reviews in toxicology*, *48*(8), 682–712.

Croux, C., & Dehon, C. (2010). Influence functions of the spearman and kendall correlation measures. *Statistical methods & applications*, *19*, 497–515.

de Almeida, L. Q., Welle, T., & Birkmann, J. (2016). Disaster risk indicators in brazil: A proposal based on the world risk index. *International Journal of Disaster Risk Reduction*, *17*, 251–272. doi:https://doi.org/10.1016/j.ijdrr.2016.04.007

Dendir, Z., & Birhanu, B. S. (2022). Analysis of observed trends in daily temperature and precipitation extremes in different agroecologies of gurage zone, southern ethiopia. *Advances in Meteorology*, *2022*(1), 4745123.

Depietri, Y., Dahal, K., & McPhearson, T. (2018). Multi-hazard risks in new york city. *Natural Hazards and Earth System Sciences*, *18*(12), 3363–3381.

Depietri, Y., & McPhearson, T. (2018). Changing urban risk: 140 years of climatic hazards in new york city. *Climatic Change*, *148*(1), 95–108.

Edelmann, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the pearson and the distance correlation coefficients. *Statistics & probability letters*, *169*, 108960.

Edwards, S. (2008). Computational Tools in Predicting and Assessing Forced Migration. *Journal of Refugee Studies*, *21*(3), 347–359. doi:10.1093/jrs/fen024. eprint: https://academic.oup.com/jrs/article-pdf/21/3/347/4409266/fen024.pdf

European Commission and European Environment Agency. (n.d.). Step 2 Assessing climate risks and vulnerabilities. Retrieved from https://climate-adapt.eea.europa.eu/en/mission/knowledge-and-data/regional-adaptation-support-tool/step-2-assessing-climate-change-risks-and-vulnerabilities

Farbotko, C., & Lazrus, H. (2012). The first climate refugees? contesting global narratives of climate change in tuvalu. *Global Environmental Change*, *22*(2), 382–390. Adding Insult to Injury: Climate Change, Social Stratification, and the Inequities of Intervention. doi:https://doi.org/10.1016/j.gloenvcha.2011.11.014

Feldmeyer, D., Nowak, W., Jamshed, A., & Birkmann, J. (2021). An open resilience index: Crowd-sourced indicators empirically developed from natural hazard and climatic event data. *Science of The Total Environment*, *774*, 145734. doi:https://doi.org/10.1016/j.scitotenv.2021.145734

Ferrazzi, M., Kalantzis, F., & Zwart, S. (2021, May). *Assessing climate change risks at the country level: the EIB scoring model*. European Investment Bank. doi:10.2867/854649

Forzieri, G., Feyen, L., Russo, S., Vousdoukas, M., Alfieri, L., Outten, S., … Cid, A. (2016). Multi-hazard assessment in europe under climate change. *Climatic Change*, *137*, 105–119.

Fransen, S., & De Haas, H. (2022). Trends and patterns of global refugee migration. *Population and Development Review*, *48*(1), 97–128.

Fransen, S., Werntges, A., Hunns, A., Sirenko, M., & Comes, T. (2023). Refugee settlements are highly exposed to extreme weather conditions. *Proceedings of the National Academy of Sciences*, *120*(0), e2206189120.

George, G. (2010). *The impact of climate change on european lakes*. Springer.

Gori Maia, A., & Schons, S. Z. (2020). The effect of environmental change on out-migration in the brazilian amazon rainforest. *Population and Environment*, *42*(2), 183–218.

Hagenlocher, M., Lang, S., Hölbling, D., Tiede, D., & Kienberger, S. (2013). Modeling hotspots of climate change in the sahel using object-based regionalization of multidimensional gridded datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(1), 229–234.

Haining, R. (2001). Spatial autocorrelation. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 14763–14768). doi:https://doi.org/10.1016/B0-08-043076-7/02511-0

Han, Q., Kumar, R., & Kumar, A. (2024). Climate change and human migration: Perspectives for environmentally sustainable societies. *Journal of Geochemical Exploration*, *256*, 107352.

Hatton, T. J. (2009). The rise and fall of asylum: What happened and why? *The Economic Journal*, *119*(535), F183–F213.

Hauer, M. E., Jacobs, S. A., & Kulp, S. A. (2024). Climate migration amplifies demographic change and population aging. *Proceedings of the National Academy of Sciences*, *121*(3), e2206192119.

Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, *30*(2), 87–93.

Hawchar, L., Naughton, O., Nolan, P., Stewart, M. G., & Ryan, P. C. (2020). A gis-based framework for high-level climate change risk assessment of critical infrastructure. *Climate Risk Management*, *29*, 100235.

Hussain, M., & Mahmud, I. (2019). Pymannkendall: A python package for non parametric mann kendall family of trend tests. *Journal of Open Source Software*, *4*(39), 1556. doi:10.21105/joss.01556

Institute for International Law of Peace and Armed Conflict (IFHV). (n.d.). WorldRiskIndex - Humanitarian Data Exchange. Retrieved from https://data.humdata.org/dataset/worldriskindex

Jennath, A., & Paul, S. (2024). Trajectories of environmental migrants: Understanding migration destinations of the coastal population in kerala, india. *Human Ecology*, 1–15.

Jones, B., Tebaldi, C., O'Neill, B. C., Oleson, K., & Gao, J. (2018). Avoiding population exposure to heat-related extremes: Demographic change vs climate change. *Climatic change*, *146*, 423–437.

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

Jotzo, F., & Howden, M. (2023, March). Analysis: Latest IPCC report confirms climate change is worsening, but we have the tools to combat it. Retrieved from https://www.pbs.org/newshour/science/analysis-latest-ipcc-report-confirms-climate-change-is-worsening-but-we-have-the-tools-to-combat-it

Karutz, R., & Kabisch, S. (2023). Exploring the relationship between droughts and rural-to-urban mobility—a mixed methods approach for pune, india. *Frontiers in Climate*, *5*, 1168584.

Khadr, M., Morgenschweis, G., & Schlenkhoff, A. (2009). Analysis of meteorological drought in the ruhr basin by using the standardized precipitation index. *World Academy of Science, Engineering and Technology*, *33*, 607–616.

Khan, M., Bhattarai, R., & Chen, L. (2024). Assessment of deadly heat stress and extreme cold events in the upper midwestern united states. *Atmosphere*, *15*(5), 614.

King, A. P., & Eckersley, R. J. (2019). Inferential statistics iii: Nonparametric hypothesis testing. *Statistics for Biomedical Engineers and Scientists*, 119–145.

Laino, E., & Iglesias, G. (2024a). High-level characterisation and mapping of key climate-change hazards in european coastal cities. *Natural Hazards*, *120*(4), 3623–3659.

Laino, E., & Iglesias, G. (2024b). Multi-hazard assessment of climate-related hazards for european coastal cities. *Journal of Environmental Management*, *357*, 120787.

Lange, S., Volkholz, J., Geiger, T., Zhao, F., Vega, I., Veldkamp, T., … Jägermeyr, J., et al. (2020). Projecting exposure to extreme climate impact events across six event categories and three spatial scales. *Earth's Future*, *8*(12), e2020EF001616.

Leal Filho, W., Ayal, D. Y., Chamma, D. D., Kovaleva, M., Alverio, G. N., Nzengya, D. M., … Nagy, G. J. (2023). Assessing causes and implications of climate-induced migration in kenya and ethiopia. *Environmental Science & Policy*, *150*, 103577.

Leon, A. C. (1998). 3.12 - descriptive and inferential statistics. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (pp. 243–285). doi:https://doi.org/10.1016/B0080-4270(73)00264-9

Lilleør, H. B., & Van den Broeck, K. (2011). Economic drivers of migration and climate change in ldcs. *Global environmental change*, *21*, S70–S81.

Liu, T., Liu, S., & Shi, L. (2020). Regression analysis of panel data. In *Time series analysis using sas enterprise guide* (pp. 101–123). doi:10.1007/978-981-15-0321-4_6

Luetz, J. (2018). Climate change and migration in bangladesh: Empirically derived lessons and opportunities for policy makers and practitioners. *Limits to climate change adaptation*, 59–105.

MacFarland, T. W., Yates, J. M., MacFarland, T. W., & Yates, J. M. (2016). Mann–whitney u test. *Introduction to nonparametric statistics for the biological sciences using R*, 103–132.

Manandhar, S., Pratoomchai, W., Ono, K., Kazama, S., & Komori, D. (2015). Local people's perceptions of climate change and related hazards in mountainous areas of northern thailand. *International Journal of Disaster Risk Reduction*, *11*, 47–59.

Marbach, M. (2018). On imputing unhcr data. *Research & Politics*, *5*(4), 2053168018803239. doi:10.1177/2053168018803239

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67–72.

Moore, W. H., & Shellman, S. M. (2004). Fear of persecution: Forced migration, 1952-1995. *Journal of Conflict Resolution*, *48*(5), 723–745.

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

Nations, U. (2023, February). Climate Change-induced Sea-Level Rise Direct Threat to Millions around World, Secretary-General Tells Security Council. Retrieved from https://press.un.org/en/2023/sc15199.doc.htm

Nawrotzki, R. J., Riosmena, F., Hunter, L. M., & Runfola, D. M. (2015). Amplification or suppression: Social networks and the climate change—migration association in rural mexico. *Global Environmental Change*, *35*, 463–474.

Nicholson, C. T. M. (2012). Climate Change and Displacement: Multidisciplinary Perspectives. Edited by Jane McAdam. *Journal of Refugee Studies*, *25*(4), 584–585. doi:10.1093/jrs/fes043. eprint: https://academic.oup.com/jrs/article-pdf/25/4/584/4366129/fes043.pdf

Nowak-Lehmann, F., Cardozo, A., & Martínez-Zarzoso, I. (2021). Migration and asylum flows to germany: New insights into the motives. *Politics and Governance*, *9*(4), 210–223.

Obokata, R., Veronis, L., & McLeman, R. (2014). Empirical research on international environmental migration: A systematic review. *Population and environment*, *36*, 111–135.

Okoye, K., & Hosseini, S. (2024). Correlation tests in r: Pearson cor, kendall's tau, and spearman's rho. In *R programming: Statistical data analysis in research* (pp. 247–277). Springer.

Orimoloye, I. R., Ololade, O. O., Mazinyo, S. P., Kalumba, A. M., Ekundayo, O. Y., Busayo, E. T., … Nel, W. (2019). Spatial assessment of drought severity in cape town area, south africa. *Heliyon*, *5*(7).

Owen, M., Kruczkiewicz, A., & Van Den Hoek, J. (2023). Indexing climatic and environmental exposure of refugee camps with a case study in east africa. *Scientific Reports*, *13*(1), 1–14.

Paglino, E. (2024). Estimating excess migration associated with tropical storms in the usa 1990–2010. *Population and Environment*, *46*(2), 1–28.

Paliwal, M., & Patwardhan, A. (2013). Identification of clusters in tropical cyclone tracks of north indian ocean. *Natural hazards*, *68*, 645–656.

Pauli, N., Williams, M., Henningsen, S., Davies, K., Chhom, C., van Ogtrop, F., … Neef, A. (2021). "listening to the sounds of the water": Bringing together local knowledge and biophysical data to understand climate-related hazard dynamics. *International Journal of Disaster Risk Science*, *12*, 326–340.

Pirani, E., Marino, M. F., & Petrucci, A. (2019). Correlates of inter-districts migrations in tanzania. a gravity-type modeling approach. *Statistica*, *79*(2), 201–221.

Prein, A. F., Holland, G. J., Rasmussen, R. M., Clark, M. P., & Tye, M. R. (2016). Running dry: The us southwest's drift into a drier climate state. *Geophysical Research Letters*, *43*(3), 1272–1279.

Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2015). Effective use of spearman's and kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, *102*, 77–84.

Raadschelders, J. C., Larrison, J., & Thapar, A. V. (2019). Refugee migration as a "wicked problem": American controlling, palliative, and governance policies in global context. *World Affairs*, *182*(3), 228–255.

Ramachandran, K. M., & Tsokos, C. P. (2021). Chapter 12 - nonparametric statistics. In K. M. Ramachandran & C. P. Tsokos (Eds.), *Mathematical statistics with applications in r (third edition)* (Third Edition, pp. 491–530). doi:https://doi.org/10.1016/B978-0-12-817815-7.00012-9

Ramli, M. W. A., Alias, N., Yusop, Z., & Taib, S. (2020). Disaster risk index: A review of local scale concept and methodologies. In *Iop conference series: Earth and environmental science* (Vol. 479, p. 012023). IOP Publishing.

Refugees International. (n.d.). Climate displacement. Retrieved from https://www.refugeesinternational.org/climate-displacement/

Reznikova, I., & Danilina, N. (2021). Environmental migration and sustainable development of regions and countries. In *E3s web of conferences* (Vol. 250, p. 05003). EDP Sciences.

Riffenburgh, R. (2012). Chapter 11 - tests on ranked data. In R. Riffenburgh (Ed.), *Statistics in medicine (third edition)* (Third Edition, pp. 221–248). doi:https://doi.org/10.1016/B978-0-12-384864-2.00011-1

Rikani, A., Frieler, K., & Schewe, J. (2022). Climate change and international migration: Exploring the macroeconomic channel. *PLoS One*, *17*(11), e0276764.

Rojas, D. (2021, December). How the climate crisis is impacting Bangladesh. Retrieved from https://www.climaterealityproject.org/blog/how-climate-crisis-impacting-bangladesh

Rothman, K. J., & Greenland, S. (2005). H ill's criteria for causality. *Encyclopedia of biostatistics*, *4*.

Savić, S., Marković, V., Šećerov, I., Pavić, D., Arsenović, D., Milošević, D., … Pantelić, M. (2018). Heat wave risk assessment and mapping in urban areas: Case study for a midsized central european city, novi sad (serbia). *Natural hazards*, *91*, 891–911.

Schmeidl, S. (1997). Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, 284–308.
Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & analgesia*, *126*(5), 1763–1768.
Schomaker, M. (2020). Regression and causality. *arXiv preprint arXiv:2006.11754*.
Schürmann, A., Kleemann, J., Teucher, M., Fürst, C., & Conrad, C. (2022). Migration in west africa: A visual analysis of motivation, causes, and routes. *Ecology and Society*, *27*(3).
Seiler, C., Hutjes, R. W., & Kabat, P. (2013). Likely ranges of climate change in bolivia. *Journal of Applied Meteorology and Climatology*, *52*(6), 1303–1317.
Shahid, S., Wang, X.-J., Harun, S. B., Shamsudin, S. B., Ismail, T., & Minhans, A. (2016). Climate variability and changes in the major cities of bangladesh: Observations, possible impacts and adaptation. *Regional Environmental Change*, *16*, 459–471.
Shen, S., & Gemenne, F. (2011). Contrasted views on environmental change and migration: The case of tuvaluan migration to new zealand. *International Migration*, *49*, e224–e242.
Shi, R., & Conrad, S. A. (2009). Correlation and regression analysis. *Ann Allergy Asthma Immunol*, *103*(4), S34–S41.
Siegfried, K. (2023, November). Climate change and displacement: the myths and the facts. Retrieved from https://www.unhcr.org/news/stories/climate-change-and-displacement-myths-and-facts
Sivisaca, D. L., Robalino, J., Cascante, A. C., Imbach, P., & Sandoval, C. (2021). Effects of extreme weather events on internal migration in rural guatemala. *Extreme Events and Climate Change: A Multidisciplinary Approach*, 135–146.
Swaen, G., & van Amelsvoort, L. (2009). A weight of evidence approach to causal inference. *Journal of Clinical Epidemiology*, *62*(3), 270–277. doi:https://doi.org/10.1016/j.jclinepi.2008.06.013
Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. doi:10.1214/009053607000000505
Taiwo, O. J., Olaniran, H. D., & Osayomi, T. (2012). Perceived causes, exposures and adjustments to seasonal heat in different residential areas in ibadan, nigeria. *The Environmentalist*, *32*, 405–414.
Tang, Y., Gan, J., Zhao, L., & Gao, K. (2006). On the climatology of persistent heavy rainfall events in china. *Advances in Atmospheric Sciences*, *23*, 678–692.
Thorn, J. P., Nangolo, P., Biancardi, R. A., Shackleton, S., Marchant, R. A., Ajala, O., … Hejnowicz, A. P. (2023). Exploring the benefits and dis-benefits of climate migration as an adaptive strategy along the rural-peri-urban continuum in namibia. *Regional Environmental Change*, *23*(1), 10.
TU Delft. (n.d.). Making a search plan. Retrieved from https://www.tudelft.nl/tulib/searching-resources/making-a-search-plan
Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., … Schiller, A. (2003a). A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences*, *100*(14), 8074–8079. doi:10.1073/pnas.1231335100
Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., … Schiller, A. (2003b). A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences*, *100*(14), 8074–8079. doi:10.1073/pnas.1231335100. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1231335100
UN High Commissioner for Refugees (via World Bank). (n.d.-a). Refugee population by country or territory of asylum, 1960 to 2021. Retrieved from https://ourworldindata.org/grapher/refugee-population-by-country-or-territory-of-asylum?tab=table
UN High Commissioner for Refugees (via World Bank). (n.d.-b). Refugee population by country or territory of origin, 2002 to 2021. Retrieved from https://ourworldindata.org/grapher/refugee-population-by-country-or-territory-of-origin?tab=table&time=2002..latest
UNHCR. (n.d.-a). Migrant definition. Retrieved from https://emergency.unhcr.org/protection/legal-framework/migrant-definition
UNHCR. (n.d.-b). Refugee definition. Retrieved from https://emergency.unhcr.org/protection/legal-framework/refugee-definition
UNHCR. (n.d.-c). Refugees. Retrieved from https://www.unhcr.org/refugees
UNHCR. (2023, June). Global Trends. Retrieved from https://www.unhcr.org/global-trends
United Nations. (n.d.). THE 17 GOALS. Retrieved from https://sdgs.un.org/goals

United Nations High Commissioner for Refugees. (n.d.). Strengthening climate adaptation and resilience. Retrieved from https://www.unhcr.org/what-we-do/build-better-futures/climate-change-and-displacement/strengthening-climate-adaptation

University, R. (n.d.). Linear regression. Retrieved from https://www.ru.nl/socialsciences/stip/facilities-support/facilities/smap/frequently-asked-questions-faq/linear-regression/#checkinfluentialcases

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. 4th International Conference on New Horizons in Education. doi:https://doi.org/10.1016/j.sbspro.2013.12.027

Vogel, R. M. (2022). The geometric mean? *Communications in Statistics - Theory and Methods*, *51*(1), 82–94. doi:10.1080/03610926.2020.1743313

Welle, T., & Birkmann, J. (2015). The World Risk Index – An Approach to Assess Risk and Vulnerability on a Global Scale. *Journal of Extreme Events*, *02*(01). Retrieved from https://doi.org/10.1142/s2345737615500037

Wesselbaum, D. (2021). Revisiting the climate driver and inhibitor mechanisms of international migration. *Climate and Development*, *13*(1), 10–20.

Wesselbaum, D., & Aburn, A. (2019). Gone with the wind: International migration. *Global and Planetary Change*, *178*, 96–109.

Xu, W., Hou, Y., Hung, Y., & Zou, Y. (2013). A comparative analysis of spearman's rho and kendall's tau in normal and contaminated normal models. *Signal Processing*, *93*(1), 261–276.

A

Search Plan for Literature Review: knowledge gaps identification

In this Appendix, the search plan is presented. In the Introduction (Chapter 1) in the main text, the findings of the relevant literature are synthesized and knowledge gaps are presented.

In this literature review, pre-determined steps are followed to create a systematic approach for selecting articles, and to make the literature review reproducible.

A search plan is build based on the Designing policy relevant research document on Brightspace for the course Preparation Master Thesis (EPA2934)[1] and TULib (TU Delft, n.d.).

1. Analyze research topic

   The key concepts for this review are refugee populations, origin location, destination location, and climate change. For each concept, synonyms and alternative terms are identified (see Table A.1).

Table A.1: Concepts and alternative search terms for them in a search table

Combine the concepts with AND				
	refugee population	**origin location**	**destination location**	**climate change**
Combine the synonyms with OR	refugee*	origin	destination	climat*
	displaced	initial place*	last place*	environment*
	stateless	first location*	final countr*	milieu*
	migrant	birthplace*	last countr*	global warming
	*migration	birth countr*	last stop	global heating
	exile*	home countr*	arrival* site*	weather changes
	fleeing	native countr*	receiving countr*	ecological*
	fled	place of birth	host* countr*	temperature*
	escapee*	place of departure*	settlement* location*	ecosystem*
	asylum seeker*	start* point*	end* point*	greenhouse
	asylum-seeker*	initial destination*	last location*	
	evacuee*	departure site*	final settlement*	
	flee	source countr*	last settlement*	
	displacement*	first place*	final stop*	
		initial location*	rest* place*	
		motherland		

Continued on next page

[1]This source is not publicly accessible.

Combine the concepts with AND				
Combine the syn-onyms with OR	**refugee population**	**origin location**	**destination location**	**climate change**
		fatherland		
		homeland		

2. Formulating a search query

   Based on Table A.1, per column, all the elements are combined with OR (vertically) and then the combined column elements are again combined with AND (horizontally). This results in the following search term:

   (refugee* OR displaced OR stateless OR *migrant* OR *migration OR exile* OR fleeing OR fled OR escapee* OR asylum seeker* OR asylum-seeker* OR evacuee* OR flee OR displacement*) AND (origin OR initial place* OR first location* OR birthplace* OR birth countr* OR home countr* OR native countr* OR place of birth OR place of de-parture* OR start* point* OR initial destination* OR departure site* OR source countr* OR first place* OR initial location* OR motherland OR fatherland OR homeland) AND (destination OR last place* OR final countr* OR last countr* OR last stop OR arrival* site* OR receiving countr* OR host* countr* OR settlement* location* OR end* point* OR last location* OR final settlement* OR last settlement* OR final stop* OR rest* place*) AND (climat* OR environment* OR milieu* OR global warming OR global heating OR weather changes OR ecological* OR temperature* OR ecosystem* OR greenhouse)

3. Choosing a database

   Keyword-search-based platforms are used in this study, because they are most common, most developed and most extensive[2]. Springer and Elsevier, for example, are within this category marked as publisher database. Aggregators are for example Google Scholar and Scopus. Aggregators contain also information from publishing databases, so an aggregator is used in this thesis.

   Scopus is chosen, because it is an internationally well-known; mostly peer-reviewed database, and because it is centered around public policy according to the course document Designing policy relevant research (EPA2934). Google Scholar is not used, as also (many) non peer reviewed scientific articles are in that database.

   The search can be done in two ways: with quotation marks and without. In this thesis, only the first option is presented below, as that search is also recommended by TULib (TU Delft, n.d.).

   TITLE-ABS-KEY ((”refugee*” OR ”displaced” OR ”stateless” OR ”*migrant*” OR ”*migration” OR ”exile*” OR ”fleeing” OR ”fled” OR ”escapee*” OR ”asylum seeker*” OR ”asylum-seeker*” OR ”evacuee*” OR ”flee” OR ”displacement*”) AND (”origin” OR ”initial place*” OR ”first location*” OR ”birthplace*” OR ”birth countr*” OR ”home countr*” OR ”native countr*” OR ”place of birth” OR ”place of departure*” OR ”start* point*” OR ”initial destination*” OR ”departure site*” OR ”source countr*” OR ”first place*” OR ”initial location*” OR ”motherland” OR ”fatherland” OR ”homeland”) AND (”destination” OR ”last place*” OR ”final countr*” OR ”last countr*” OR ”last stop” OR ”arrival* site*” OR ”receiving countr*” OR ”host* countr*” OR ”settlement* location*” OR ”end* point*” OR ”last location*” OR ”final settlement*” OR ”last settlement*” OR ”final stop*” OR ”rest* place*”) AND (”climat*” OR ”environment*” OR ”milieu*” OR ”global warming” OR ”global heating” OR ”weather changes” OR ”ecological*” OR ”temperature*” OR ”ecosystem*” OR ”greenhouse”)).

4. Screening

   This search leads to 674 results. When excluding the subject areas ‘Medicine’, ‘Business, Management and Accounting’, ‘Biochemistry, Genetics and Molecular Biology’, ‘Nursing’, ‘Psychology’, ‘Immunology and Microbiology’, ‘Economics, Econometrics and Finance’, ’‘Physics and Astronomy’, ‘Materials Science’, ‘Chemical Engineering’ ,‘Chemistry’, ‘Neuroscience’, ‘Health Pro-

[2]Cited from the not-publicly accessible course document (EPA2934)

fessions', 'Veterinary' and 'Pharmacology, Toxicology and Pharmaceutics' (topics totally unrelated with the topic of this research) in Scopus, 390 results remain.

In further selection, only articles that are in their final Publication Stage are selected. As a result, there are 383 articles remaining.

Finally, only English publications are selected. This results in the final selection of 346 publications for the analysis.

This leads to the following final query in Scopus:

TITLE-ABS-KEY ((”refugee*” OR ”displaced” OR ”stateless” OR ”*migrant*” OR ”*migration” OR ”exile*” OR ”fleeing” OR ”fled” OR ”escapee*” OR ”asylum seeker*” OR ”asylum-seeker*” OR ”evacuee*” OR ”flee” OR ”displacement*”) AND (”origin” OR ”initial place*” OR ”first location*” OR ”birthplace*” OR ”birth countr*” OR ”home countr*” OR ”native countr*” OR ”place of birth” OR ”place of departure*” OR ”start* point*” OR ”initial destination*” OR ”departure site*” OR ”source countr*” OR ”first place*” OR ”initial location*” OR ”motherland” OR ”fatherland” OR ”homeland”) AND (”destination” OR ”last place*” OR ”final countr*” OR ”last countr*” OR ”last stop” OR ”arrival* site*” OR ”receiving countr*” OR ”host* countr*” OR ”settlement* location*” OR ”end* point*” OR ”last location*” OR ”final settlement*” OR ”last settlement*” OR ”final stop*” OR ”rest* place*”) AND (”climat*” OR ”environment*” OR ”milieu*” OR ”global warming” OR ”global heating” OR ”weather changes” OR ”ecological*” OR ”temperature*” OR ”ecosystem*” OR ”greenhouse”)) AND (EXCLUDE (SUBJAREA,”MEDI”) OR EXCLUDE (SUBJAREA,”BUSI”) OR EXCLUDE (SUBJAREA,”PSYC”) OR EXCLUDE (SUBJAREA,”ECON”) OR EXCLUDE (SUBJAREA,”BIOC”) OR EXCLUDE (SUBJAREA,”NURS”) OR EXCLUDE (SUBJAREA,”PHYS”) OR EXCLUDE (SUBJAREA,”NEUR”) OR EXCLUDE (SUBJAREA,”IMMU”) OR EXCLUDE (SUBJAREA,”CHEM”) OR EXCLUDE (SUBJAREA,”MATE”) OR EXCLUDE (SUBJAREA,”HEAL”) OR EXCLUDE (SUBJAREA,”CENG”) OR EXCLUDE (SUBJAREA,”PHAR”) OR EXCLUDE (SUBJAREA,”VETE”)) AND (LIMIT-TO (PUBSTAGE,”final”)) AND (LIMIT-TO (LANGUAGE, ”English”)).

When analyzing those 346 results, only 35 are relevant, based on title, keywords and abstract. Those publications are exactly about the research topic of this thesis. The others are not relevant, researching completely different topics. Some articles are, for instance, on animal migration or migrant payments. Those topics are out of scope for this research. Finally, there are some duplicates.

5. Full-text eligibility

   Of the 35 publications, 9 articles are excluded after reading the full text with the main reason being an irrelevant context. Some publication turned out to be off-scope, for example a publication that focused on generic implications of the Paris Agreements. Also, some publications are not fully accessible (only the title, abstract and key-words are accessible).

6. Inclusion

   The 26 remaining publication describe an aspect of the link between the effects of climate change and migration and are synthesized in the main text in Chapter 1. These articles are relevant for identifying the knowledge gaps and the historical studies that have been conducted till now in this research context.

B

Supportive results Chapter 4: Climate as driver for fleeing

B.1. Correlation analysis

The selection of correlation coefficients and identification of suitable trend tests are central in this section.

B.1.1. Visualizing correlation coefficients over years

As a first step to determine whether there is any relation/association between climate exposure and fleeing, the correlation coefficient between the two concepts is researched. This is because, to put simply, if there is no correlation, there cannot (in general) be a relation, let alone causation. So, if correlation coefficients are close to zero, this gives a lot of information about the relation between climate and fleeing. There are many types of correlation coefficients, with the most familiar ones being: Pearson, Spearman and Kendall.

To interpret the estimated correlation coefficients in a valid way, the data one uses to calculate the correlation coefficients should fulfill several assumptions. If the data does not fulfill those requirements, and the coefficients are interpreted anyway, the conclusions can be wrong or inaccurate.

The most important assumption the data should fulfill to use Pearson correlation is that the variables one measures the correlation between, should be linearly related (Hauke & Kossowski, 2011), while the article of Schober, Boer, and Schwarte (2018) mentions the absence of outliers and the fact that the variables need to be normally distributed (strictly bivariate normal distribution) as most important. All these assumptions are also identified as important by (Armstrong, 2019). The data, variables used in this study, fulfill none of the above assumptions. That is illustrated in Figure B.1. For different years and different refugee numbers (origin and destination), as well as different indicators as examples is shown that the overall structure is the same and that all the data/variables do not fulfill the assumptions. (Note that values are not zero in the bivariate plot, but close to zero.) From these plots it is clear that the Pearson correlation coefficients are not correct to use.

While Pearson correlation is used as a measure of linear association among variables (Okoye & Hosseini, 2024; Schober et al., 2018), Spearman and Kendall correlation coefficients are both used to measure monotonicity in relationships (Bonita et al., 2014). Spearman as well as Kendall are non-parametric alternatives of the Pearson correlation coefficient (Puth, Neuhäuser, & Ruxton, 2015).

In general, multiple sources indicate that a researcher can use one of the two coefficients as alternative to Pearson, because they both produce similar results (Hauke & Kossowski, 2011; Okoye & Hosseini, 2024). However, in Croux and Dehon (2010), the comparison between Spearman and Kendall is investigated in more detail and this study call for Kendall Tau correlation over Spearman correlation: “The Kendall correlation measure is more robust and slightly more efficient than Spearman’s rank correlation, making it the preferable estimator from both perspectives.” Moreover, Hauke and Kossowski

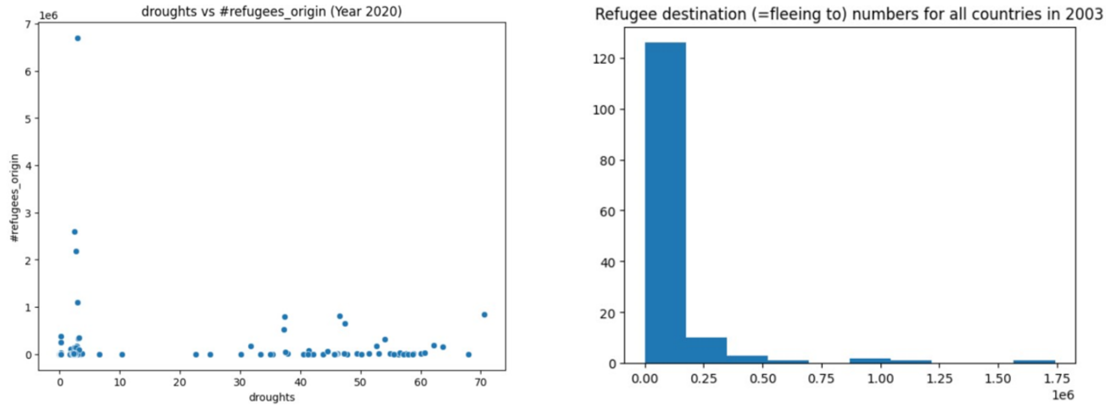


Figure B.1: Examples of not satisfying assumptions for Pearson correlation coefficients

(2011) calls for interpreting the one that estimates the lowest coefficients, which is Kendall correlation in this study case, as one will see below. Also Puth et al. (2015) confirms, in the use case for this study (without having ties in the data), to use Kendall correlation. In this refugee research, using Kendall instead of Spearman is also confirmed by Xu, Hou, Hung, and Zou (2013), because there is no really small sample size, but rather large scale, per year, per variable there are over 130 data points. Therefore, this study proceeds with Kendall correlation and uses it as the main method in the analysis.

In addition, this study identifies the results of the rather new Distance Based correlation coefficient. This correlation is also an alternative to Pearson correlation, but with the unique characteristic that it measures (in)dependence and has range [0,1] instead of [-1,1] (Edelmann, Móri, & Székely, 2021; Székely, Rizzo, & Bakirov, 2007). The lack of Distance correlation defines independence, that is, dCor(X, Y) = 0 if and only if X and Y are independent. Their empirical results suggest that their test may be more powerful than the parametric test when the dependence structure is nonlinear. Their proposed statistics are sensitive to all types of departures from independence, including nonlinear or nonmonotone dependence structure (Székely et al., 2007). Therefore, this coefficients can provide additional information, because not all relationships between indicators are monotonic, let alone linear. Also, independence can immediately be identified.

B.1.2. Testing correlation changes over years

In this section, statistical analyses are used to interpret correlation changes over years that can be seen in the matrices in the previous section. This way can be established whether correlation association between social and climatic factors with refugee movements increase, decrease or stay stable over time. For example, identifying an increasing trend between droughts and refugees fleeing from a country in recent years, could potentially indicate that droughts play a larger role in fleeing in the future. At least it indicates that more people fled away from more to drought exposed countries, over the years. In this section, the Kendall correlation coefficients are used to base the tests upon, as argued for in the previous section[1].

There are different test one can do to identify trends/changes over years. To investigate whether there is standard trend in Time Series data, one can do a Mann-Kendall test, Theil-Sen Slope test, simple linear regression between year and the correlation, or again a correlation between the correlation coefficients of the indicators and the years they were found in. The test one chooses depends on certain assumptions/characteristics of the data.

First, below, some bivariate plots are presented to make the factors visual between which should be determined whether there is a trend or not.

One can see that in the origin as well as the destination case, not all relations are strictly monotonic

[1]Kendall is used over Distance correlation, as Kendall correlation is well established and robust. Distance based correlation is rather new and relatively unknown, and under-researched, compared to Kendall correlation.

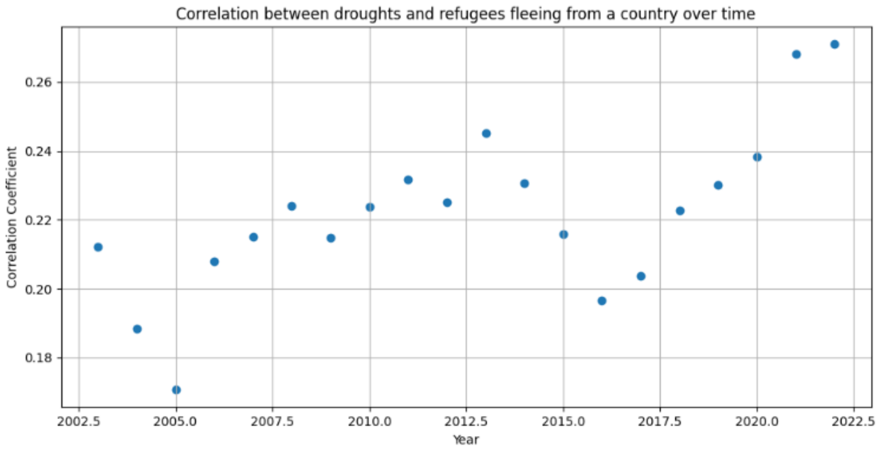


Figure B.2: Correlation between drought scores and refugee destination numbers

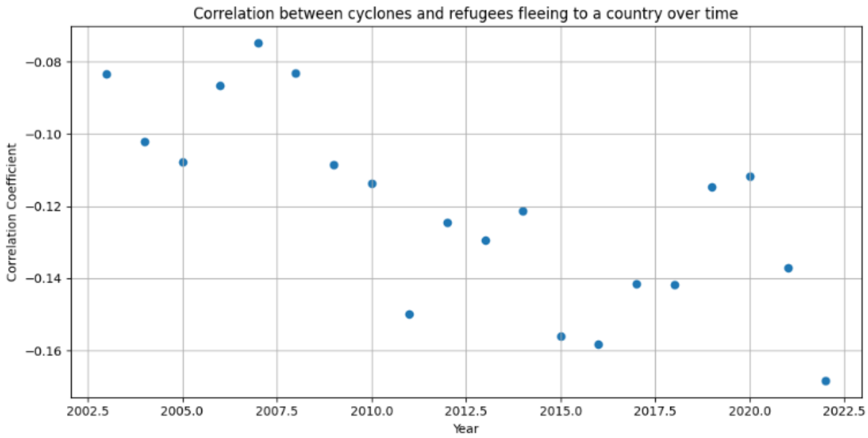


Figure B.3: Correlation between cyclone scores and refugee destination numbers

and linear, but deviations from monotonic behavior are not very drastic. Therefore, non-parametric test are used, of which the Mann-Kendall Test forms the core. The Mann-Kendall Trend Test (often referred to as the MK test) is employed to examine trends in time series data (Alashan, 2020). As a non-parametric test, it is applicable to all types of distributions, meaning it does not require the data to follow a normal distribution (Hussain & Mahmud, 2019). The MK test is a robust tool for trend analysis (Alashan, 2020; Hussain & Mahmud, 2019). However, the original test assumes that the data is free from serial correlation. If there is serial correlation in the data, it may affect the significance level (p-value) and potentially lead to incorrect interpretations (Hussain & Mahmud, 2019). To address this issue, researchers have developed several modified versions of the Mann-Kendall test, with the Hamed and Rao Modified MK Test being the most appropriate choice for this study to handle potential autocorrelation problems.

Below, the non-parametric results are presented for both the original MK and the Hamed and Rao modified test. Also, the Theil-Sen Slope is presented, with the Kendall Tau correlation coefficient between the years and the correlation coefficients, to identify the trend in another non-parametric way. In this case, the correlation is fitted between the year variable (increasing from 2003-2020) and the earlier identified correlation coefficients for every indicator with the refugee data.

B.1.2.1. Verification of changing correlation results

The Tables below verify the results presented in the main text.

B.2. Regression

This section focuses on the data selection and manipulation to prepare for the regression analysis.

	Indicator	H value	Mann-Kendall Statistic	Mann-Kendall p-value	Kendall Tau	Theil-Sen Slope	S value	Conclusion
0	coastal floodings	False	-0.619898	5.353247e-01	-0.126316	-0.001128	-24.0	There is no significant trend.
1	cyclones	True	2.261477	2.372971e-02	0.484211	0.002976	92.0	There is a significant Increasing trend.
2	droughts	True	2.255700	2.408944e-02	0.494737	0.002762	94.0	There is a significant Increasing trend.
3	lack of adaptive capacities	True	3.860870	1.129841e-04	0.631579	0.003867	120.0	There is a significant Increasing trend.
4	lack of coping capacities	True	4.078826	4.526364e-05	0.536842	0.003763	102.0	There is a significant Increasing trend.
5	lack of socio-economic development	True	3.633054	2.800866e-04	0.705263	0.006585	134.0	There is a significant Increasing trend.
6	riverine floodings	True	-3.082207	2.054719e-03	-0.505263	-0.002415	-96.0	There is a significant Decreasing trend.
7	sea level rise	False	0.772751	4.396695e-01	0.178947	0.001176	34.0	There is no significant trend.
8	societal disparities	True	5.158641	2.487485e-07	0.842105	0.008121	160.0	There is a significant Increasing trend.
9	socio-economic deprivation	True	3.990647	6.589330e-05	0.652632	0.003664	124.0	There is a significant Increasing trend.

Figure B.4: Hamed Rao Mann-Kendall Test and Theil-Sen Slope for Correlation with Destination data

	Indicator	H value	Mann-Kendall Statistic	Mann-Kendall p-value	Kendall Tau	Theil-Sen Slope	S value	Conclusion
0	coastal floodings	False	1.070366	2.844546e-01	0.252632	0.002605	48.0	There is no significant trend.
1	cyclones	True	-3.276873	1.049637e-03	-0.536842	-0.003370	-102.0	There is a significant Decreasing trend.
2	droughts	False	0.708987	4.783324e-01	0.189474	0.001058	36.0	There is no significant trend.
3	lack of adaptive capacities	False	-0.206861	8.361181e-01	-0.052632	-0.000602	-10.0	There is no significant trend.
4	lack of coping capacities	True	6.060660	1.355642e-09	0.652632	0.010101	124.0	There is a significant Increasing trend.
5	lack of socio-economic development	False	-0.203180	8.389943e-01	-0.052632	-0.000234	-10.0	There is no significant trend.
6	riverine floodings	True	5.288418	1.233786e-07	0.863158	0.008489	164.0	There is a significant Increasing trend.
7	sea level rise	False	0.000000	1.000000e+00	0.000000	-0.000073	0.0	There is no significant trend.
8	societal disparities	False	-0.408412	6.829710e-01	-0.094737	-0.000351	-18.0	There is no significant trend.
9	socio-economic deprivation	True	-2.374593	1.756830e-02	-0.536842	-0.002455	-102.0	There is a significant Decreasing trend.

Figure B.5: Hamed Rao Mann-Kendall Test and Theil-Sen Slope for Correlation with Destination data

B.2.1. Preparation

This regression analysis also comes with a lot of data manipulation to fulfill the regression assumptions and be able to conduct nice, neat and correct interpretations.

In this analysis, the refugee movements are the dependent variables (y) and the social/climatic indicators are the independent variables (X). Obviously, for refugee destination as well as refugee origin, different models are estimated, for every year. So, every year, there is a formula that 'predicts' the number of refugees fleeing to/from a country, based on all indicators and interactions plus higher order effects

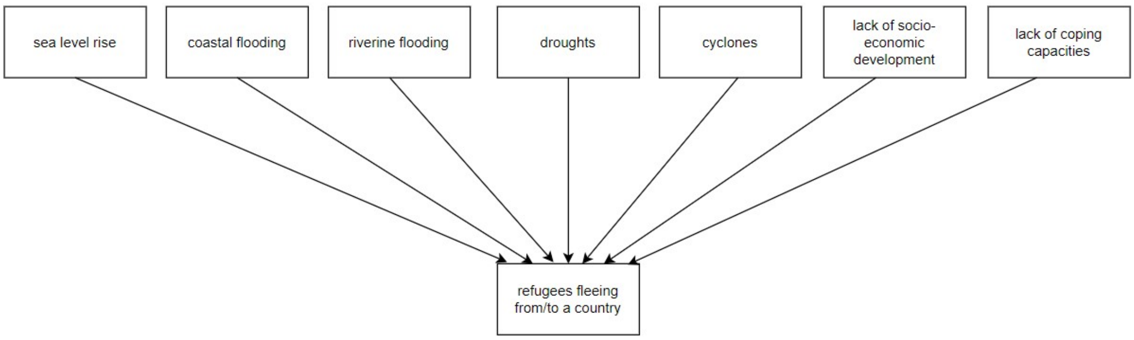


Figure B.6: Conceptual regression model

This is the assumed 'theoretical model' or conceptual model in this multiple regression study. The indicators are controlled for the effects of each other, and also interactions between climatic indicators and social factors are studied, as maybe the climatic situation could influence the relation between social factors and refugees fleeing, for example. That is what this interaction effects are measuring. Also, the interactions between other climatic indicators which each other, are measured, because multiple climatic effects could influence refugee movements more than just the additive separate direct effects of the individual climate events. At last, in the final model, also higher order effects of every variable are taken into account, to capture other non-linear relations.

Further, the separate inclusion of sea level rise as well as coastal flooding may seem remarkable for an attentive reader. One could think that the theoretical model assumes that sea level rise has no direct meaningful effect on refugee movements, only indirect through coastal flooding. And that therefore, this indicator is not incorporated in the model, because regular regression analysis would only fit a coefficient for a direct relation between sea level rise and refugee movements. And that indirect effects should be measured by Structural Equation Modeling (SEM), which is out of the scope for this report. However, there are several reasons for including sea level rise in the analysis. 1) Sea level rise is not only associated with (coastal) flooding, but also with erosion and/or salinization, for example (there are many more effects) (Nations, 2023). 2) When including sea level rise in the regression analysis, the adjusted R-squared becomes higher than without this indicator[2]. Also, the sea level rise indicator is significant in the origin case almost all years, while coastal flooding is almost in no year a significant indicator. This is even the case when sea level rise is excluded from the regression model. Then still coastal floodings cannot explain significant variance in refugee movement numbers. Therefore, sea level rise has added value that is identified by the model. That is also a big reason to include this indicator. Apparently, sea level rise can explain significant variance in refugees fleeing from a country.

In addition, one can see that only two social variables are incorporated in the model, while in the Correlation analysis more social variables are used. This has to do with multicollinearity, as will be reflected upon below.

What not comes as a surprise, is that, without manipulating the data, the data does not fulfill the assumptions to conduct multiple linear regression upon. One could see earlier in the Correlation analysis section that the data did not fulfill the assumption to estimate Pearson (parametric) correlation.

Note that linear regression does refer to linear in parameters, and not estimating only a linear function. For linear regression analysis, the following assumptions should hold, according to Alexopoulos (2010):
-For each value of the independent variable, the distribution of the dependent variable must be normal.
-The variance of the distribution of the dependent variable should be constant for all values of the independent variable. -The relationship between the dependent variable and the independent variables should be linear. -All observations should be independent. So the assumptions are: independence; linearity; normality; homoscedasticity.

According to University (n.d.), the following conditions should hold:
-Data should be independent.
-Residuals are normally distributed.
-Residuals are homoscedastic. (This is sometimes also called homogeneity of variance.)
-The relationship(s) between IV(s) and DV is linear.
-Avoid multicollinearity (i.e. highly correlated IVs).
-No influential cases.

The following assumptions are confirmed by Uyanık and Güler (2013) to conduct multiple regression analysis: normality, linearity, no extreme values.

So, one can see high overlap between listed assumptions of different academic studies. Some academic sources list more assumptions than others. These assumptions can be checked with tests and plots, which is shown below.

When including all the social (economic) factors, multicollinearity comes into play. This means that the coefficients cannot be estimated correctly. This is a problem. After several runs, including and excluding several social factors, it is found that including lack of coping capacities and lack of socio-economic developments makes sure multicollinearity is prevented. If more than two social factors are included, multicollinearity arises[3].

[2] One can verify this in the GitHub correlation analysis file provided by this report, by manipulating cell 106 (excluding sea level rise variables) and looking at decreased performance in cell 115 and 117 outcomes.

[3] In the Python correlation analysis file on GitHub provided by this thesis, one can then see then that the VIF in cell score 98 (origin) or 99 (destination) exceeds 10, if more than two social factors are included in cell 97.

One can see below, that without manipulating the data further, it makes no sense to run a linear regression analysis:

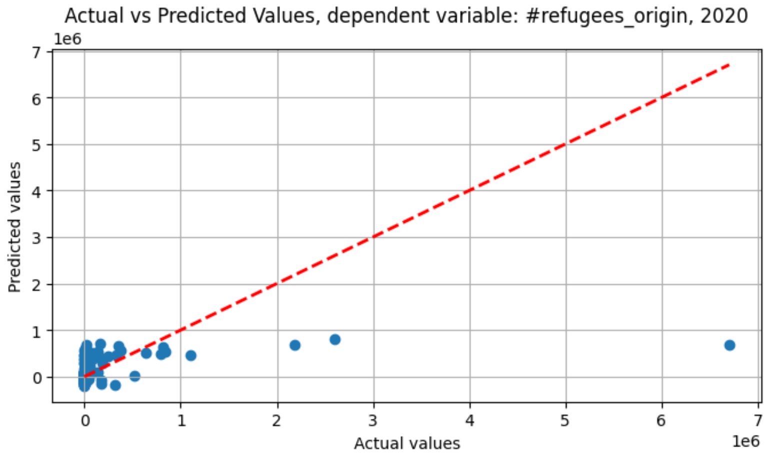


Figure B.7: Actual vs predicted refugees fleeing from a country, without scaling, higher order effects and interactions

This plot represents a standard example of an influential case. Also, the data does not fulfill all the tests as one can see in Figure B.8 for the origin case. Linearity and normality of residuals tests are failed, and sometimes there is positive autocorrelation as well as homoscedasticity. The destination case is similar.

When the dependent variable is log transformed, all the assumptions for conducting linear regression hold, especially in latest years in the origin case (as one can verify in line 94 of the GitHub correlation analysis file with all the plots and table). But without including interactions and higher order effects in the model, there is also still a low actual vs predicted fit, especially for earlier years and in the destination case (as one can verify in results line 95 of the GitHub correlation analysis file with all the plots). With the inclusion of higher order effects and interaction effects in the model, actual vs predicted plots are more aligned, for origin as well as destination, see plots below with different year examples:

The plots suggest already a 'rather good' fit, way better than without manipulating the data and without including higher order terms and interactions. And it is even better, for sure, in latest years, as one can verify in the plot here. The model improves over time, i.e. model fit is better when the years increase. Also, the residuals vs fitted values look 'good', as one can see in the example below.

There is no structure/pattern deviation from the zero line in the residual plot, that is key. This means that the linear model fits the data well. The same holds (to a little bit lesser extent) for the destination case. Both origin and destination models, with the log scaled y variable and the interactions and higher order effects, fulfill every test/assumption (as one can see in the lines 103 and 108 in the GitHub correlation analysis file for all the plots and tables).

So, with these plots (residuals vs fitted and actual vs predicted) and with the tests (tables) can be concluded that the assumptions for multivariate regression are met. This log scaled dependent variable model with interactions and higher order effects will be used in the analysis.

	Year	Linearity	Normality of Residuals	Homoscedasticity	Autocorrelation	Multicollinearity
0	2003	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
1	2004	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
2	2005	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
3	2006	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
4	2007	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
5	2008	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
6	2009	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
7	2010	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
8	2011	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
9	2012	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	Positive autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
10	2013	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
11	2014	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
12	2015	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
13	2016	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
14	2017	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
15	2018	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
16	2019	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
17	2020	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
18	2021	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Satisfies homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]
19	2022	Violates linearity assumption (Rainbow test)	Residuals are not normally distributed (Shapiro-Wilk)	Violates homoscedasticity assumption	No significant autocorrelation detected (Durbin-Watson test)	[coastal floodings has VIF <= 10, riverine floodings has VIF <= 10, cyclones has VIF <= 10, droughts has VIF <= 10, sea level rise has VIF <= 10, lack of socio-economic development has VIF <= 10, lack of coping capacities has VIF <= 10]

Figure B.8: Not fulfilling linear regressions assumptions for origin data

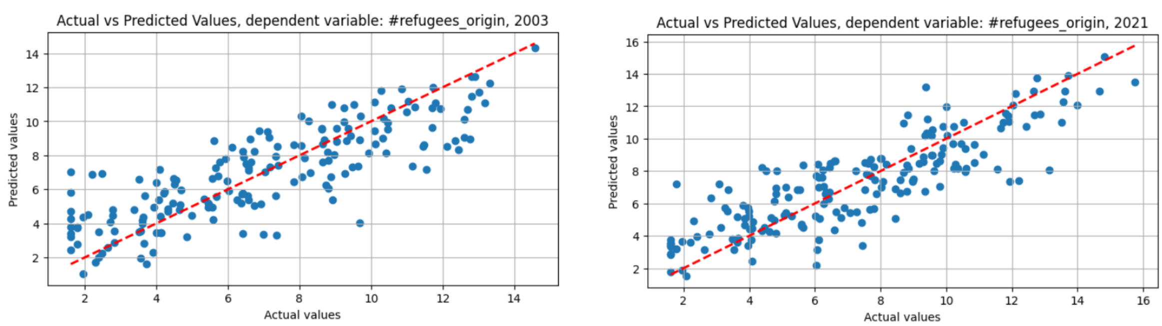


Figure B.9: Actual vs predicted refugees fleeing to a country, , with scaling, higher order effects and interactions

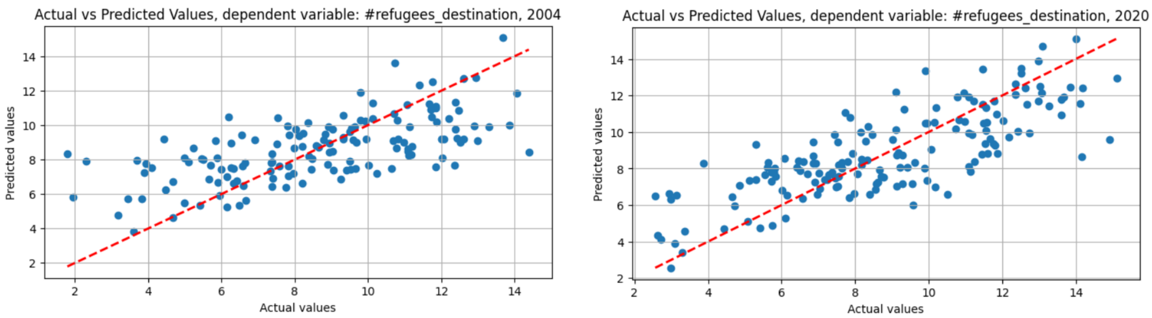


Figure B.10: Actual vs predicted refugees fleeing from a country, with scaling, higher order effects and interactions

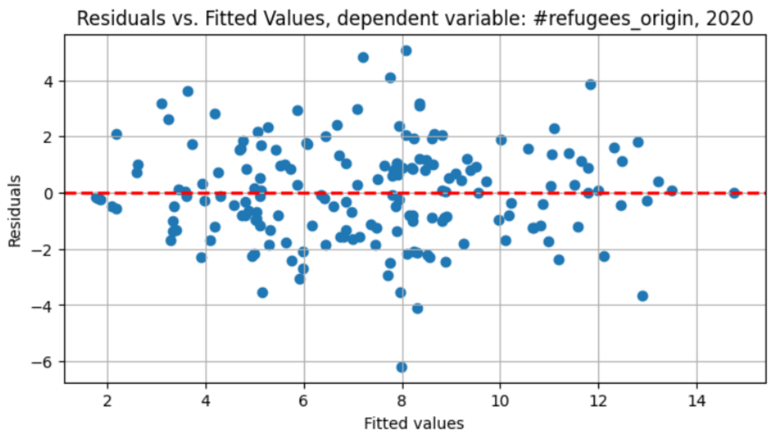


Figure B.11: Residual plot example 2020 refugees fleeing from a country

C

Supportive results Chapter 5: Spatial distributions

In this appendix, supportive evidence is presented for all the analyses conducted in Chapter 5. The comparison of scores over cluster as well as original scores is central in this appendix. Regarding these comparisons, (non-)parametric paired tests are used, because the same variables for the same countries are measured in 2003 and later in 2022. Details for every indicator are found below.

C.1. Droughts

This section shows supportive evidence for the drought analysis.

C.1.0.1. Statistical cluster changes

To show statistically that drought clusters do (not) change over time, plots and statistical tests are conducted.

From Figure C.1, one can conclude that the relation is approximately linear and that the distributions of the variables are approximately normal, but with heavier tails. Also, there is one 'outlier' visible in the scatter plot. However, the results between Pearson and Kendall correlation do not change heavily. The results of a Kendall correlation estimation between LISA values 2003 and 2022 is 0.93. For Pearson, this is 0.96. The statistical tests indicate that the data is not normal distributed. The Kolmogorov-Smirnov (KS) test is used to determine that, because the sample size is bigger than 50. Otherwise, the Shapiro-Wilk test should have been used (Mishra et al., 2019). However, both the KS and the Shapiro-Wilk test conclude the same: the data is not normally distributed (2003: KS statistic = 0.2469, p-value = 1.2588e-07. Shapiro-Wilk test: W statistic = 0.8679, p-value = 1.5761e-09. 2022: KS statistic = 0.2445, p-value = 1.7345e-07. Shapiro-Wilk test: W statistic = 0.8751, p-value = 3.3463e-09). This can be also be seen in Figure C.2. This holds for 2003 LISA data as well as 2022 LISA data.

So, one should use the Kendall Tau correlation coefficient, as not all assumption of Pearson correlation are met (there is for example an outlier and the data is not normally distributed). Kendall correlation is in this case also better than Spearman as there are equal ranks (approximately).

Below, statistical differences are tested, apart from a simple correlation coefficient. As indicated above, the LISA drought scores are not normally distributed. Therefore, in this, case an alternative of the paired t-test is chosen: The paired samples Wilcoxon test (also known as Wilcoxon signed-rank test). This is a non-parametric alternative to paired t-test used to compare paired data. That is earlier explained in Chapter 4. The Wilcoxon Signed-Rank test (W statistic = 3339.0, p-value = 0.0122) indicates a significant difference between the LISA values in 2003 and 2022, only at the 5% level, not at the one percent level. The outlined differences can be seen in Figure C.3. Main results can be found in Chapter 5.

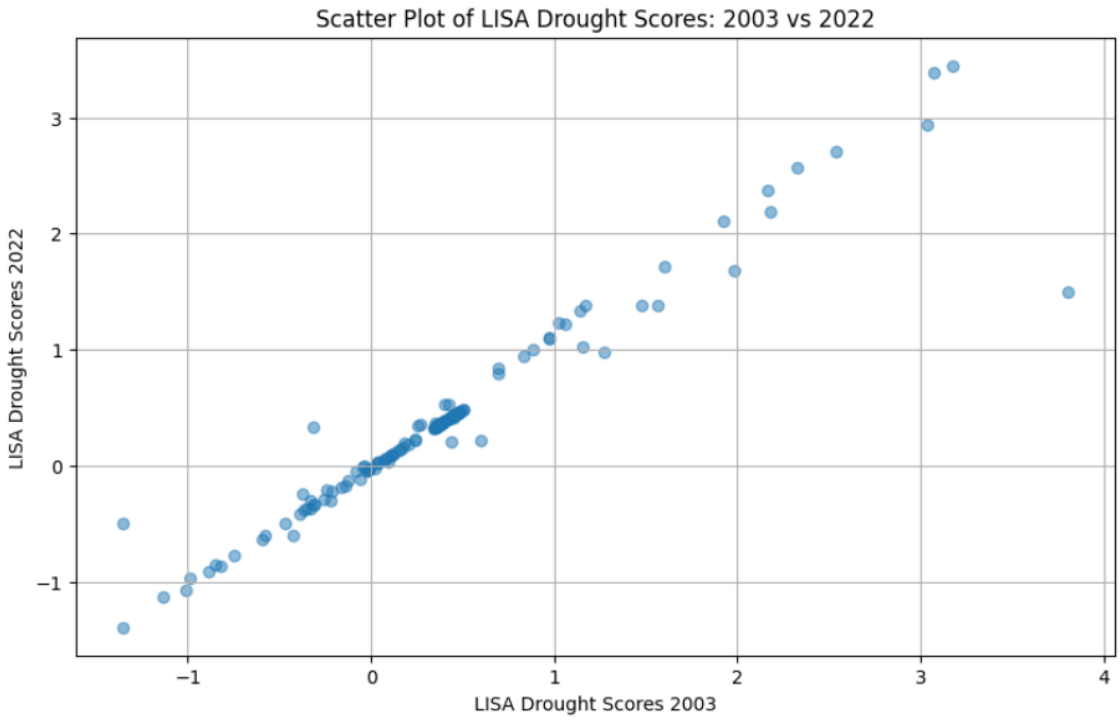


Figure C.1: Comparison of LISA Drought Scores

C.1.1. Drought changes over years

In this section, the original, regular drought values are compared over years, and not the changing clustering values (LISA values), as was done in the previous section.

Before comparing the change in 2003 and 2022 drought values, it needs to be examined which correlation coefficient and statistical test can be used. The data is not normally distributed (see Figure C.4. The Kolmogorov-Smirnov test also indicates that (2003: KS statistic = 0.3919, p-value = 3.4298e-19; 2022: KS statistic = 0.3965, p-value = 1.1959e-20).

C.2. Riverine floodings

This section shows supportive evidence for the riverine flooding analysis.

C.2.0.1. Statistical cluster changes

To examine statistically that riverine flooding clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used for the analysis. Therefore, statistical assumptions are examined. The relation between the 2003 and 2022 clustering values look linear, see the plot in Figure C.6. However, the LISA 2003 and 2022 values are not normally distributed (see Figure C.7). The latter is also indicated by the KS test (2003: KS statistic = 0.2319, p-value = 9.0631e-07; 2022: KS statistic = 0.2187, p-value = 4.5875e-06). Therefore, the Kendall correlation coefficient is used instead of the Pearson correlation coefficient. Earlier is reflected on why Spearman correlation is not chosen.

Also, assumptions for the paired t-test are not met, as one can see above. Therefore, the Wilcoxon Signed-Rank test is conducted to formally test differences. The results can be found in Chapter 5.

C.2.1. Riverine flooding changes over years

In this section, the original, regular sea level rise risk values are compared from 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section. Before comparing the change in 2003 and 2022 original sea level rise values, it needs to be examined which correlation coefficient and statistical test can be used. The relation between the originals values look linear, see

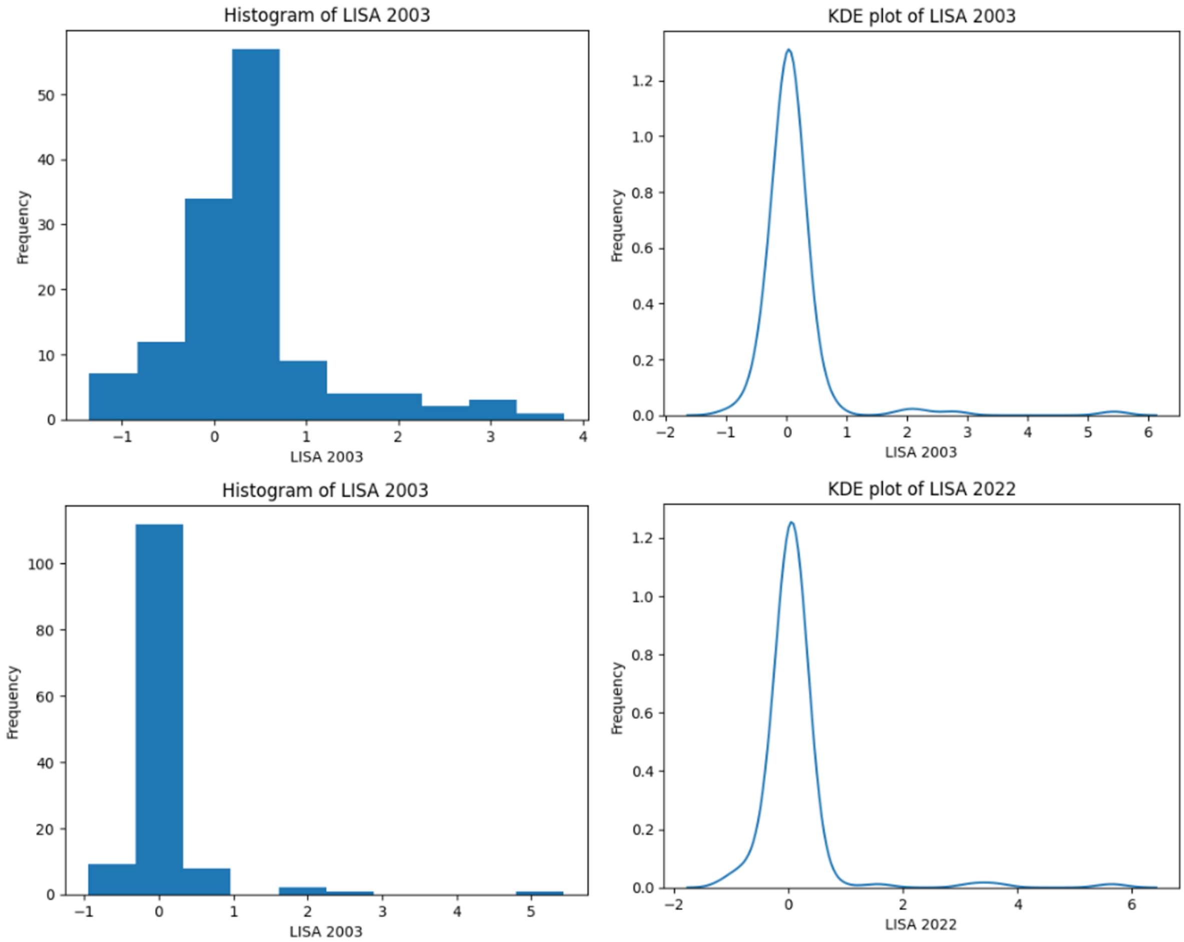


Figure C.2: Distribution of Drought LISA values

the plot in Figure C.8. The data is normally distributed (2003: KS statistic = 0.0595, p-value = 0.7021; 2022: KS statistic = 0.0453, p-value = 0.9196). (See also Figure C.9.) Therefore, in the analysis, the Pearson correlation coefficient and the paired t-test are used to compare the two data groups at different time instances.

C.3. Coastal floodings

This section shows supportive evidence for the coastal flooding analysis.

C.3.0.1. Statistical cluster changes

To examine statistically that coastal flooding clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used in this analysis. Therefore, assumptions are examined. The relation between the values look linear, without outliers, see Figure C.10. However, the LISA 2003 and 2022 values are not normally distributed, indicated by the Kolmogorov-Smirnov test (2003: KS statistic = 0.1554, p-value = 0.0029; 2022: KS statistic = 0.1554, p-value = 0.0028). This can also be seen in Figure C.11. Therefore, the Kendall correlation coefficient is used instead of the Pearson correlation. Earlier is reflected on why Spearman correlation is not chosen.

Also, assumptions for the paired t-test are not met, as one can see above. Therefore, the Wilcoxon Signed-Rank test is conducted to formally test differences. The results can be found in Chapter 5.

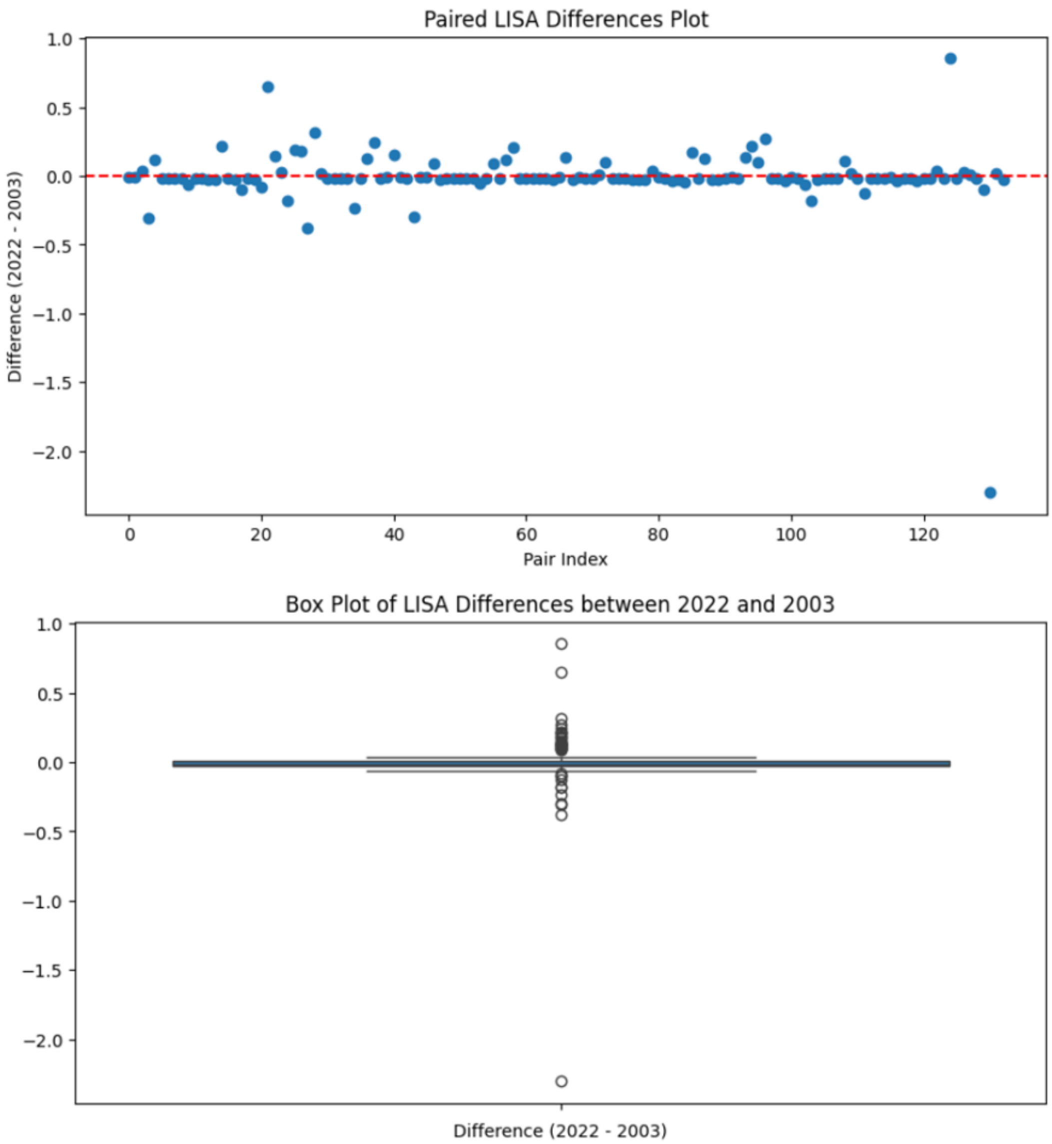


Figure C.3: Drought LISA scores spread

C.3.1. Coastal flooding changes over years

In this section, the original, regular coastal flooding intensity values are compared in 2003 and 2022, and not the changing clustering values (LISA values) are compared, as was done in the previous section. Before comparing the change in 2003 and 2022 coastal flooding intensity values, it needs to be examined which correlation coefficient and statistical test can be used. Plots are shown in Figure C.12. This indicates that the data is not normally distributed. That is also confirmed by the Kolmogorov-Smirnov test (2003: 0.2547, p-value = 3.2948e-08; 2022: KS statistic = 0.2624, p-value = 4.0592e-09).

C.4. Sea level rise

This section shows supportive evidence for the sea level rise analysis.

C.4.0.1. Statistical cluster changes

To examine statistically that sea level rise clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used. Therefore, assump-

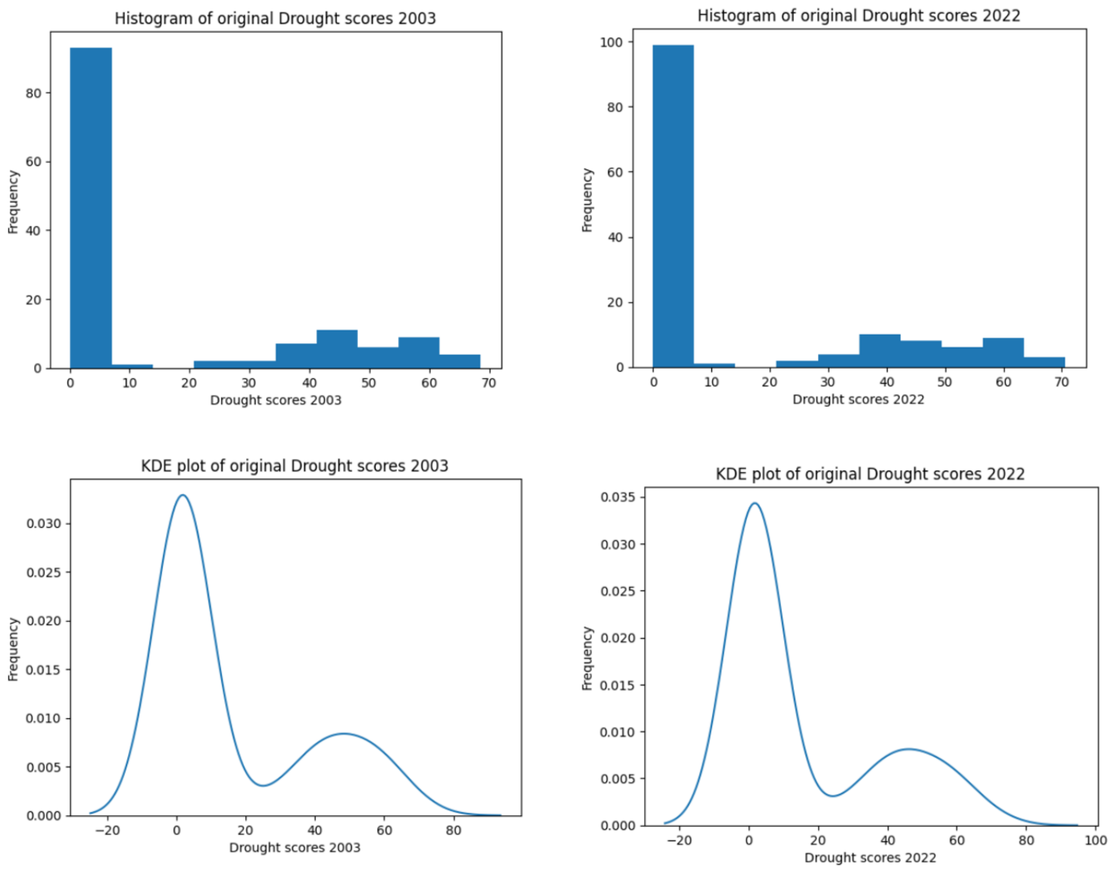


Figure C.4: Distribution of Drought original values

tions are examined. The relation between the values look linear, see Figure C.14.

However, the LISA 2003 and 2022 values are not normally distributed, indicated by the Kolmogorov-Smirnov test (2003: KS statistic = 0.1666, p-value = 0.0011; 2022: KS statistic = 0.1633, p-value = 0.0014). In addition, non-normality is shown in Figure C.15. Therefore, the Kendall correlation coefficient is used instead of the Pearson correlation coefficient. Earlier is reflected on why Spearman correlation is not chosen.

Also, assumptions of the paired t-test are not met, as one can see above (non-normality). Therefore, the Wilcoxon Signed-Rank test is conducted. The results can be found in Chapter 5.

C.4.1. Sea level rise changes over years

In this section, the original, regular sea level rise risk values are compared in 2003 and 2022, and not the changing clustering values (LISA values) are studied, as was done in the previous section. Before comparing the change in 2003 and 2022 sea level rise values, it needs to be examined which correlation coefficient and statistical test can be used. In Figure C.16 is shown that the data is not normally distributed. This is also indicated by the Kolmogorov-Smirnov test (2003: KS statistic = 0.1937, p-value = 6.5262e-05; 2022:KS statistic = 0.2025, p-value = 1.4025e-05).

C.5. Cyclones

This section shows supportive evidence for the cyclone analysis.

C.5.0.1. Statistical cluster changes

To examine statistically that cyclone intensity clusters do (not) change over time, statistical tests are conducted. First, it needs to be determined which tests and coefficients can be used. Therefore, sta-

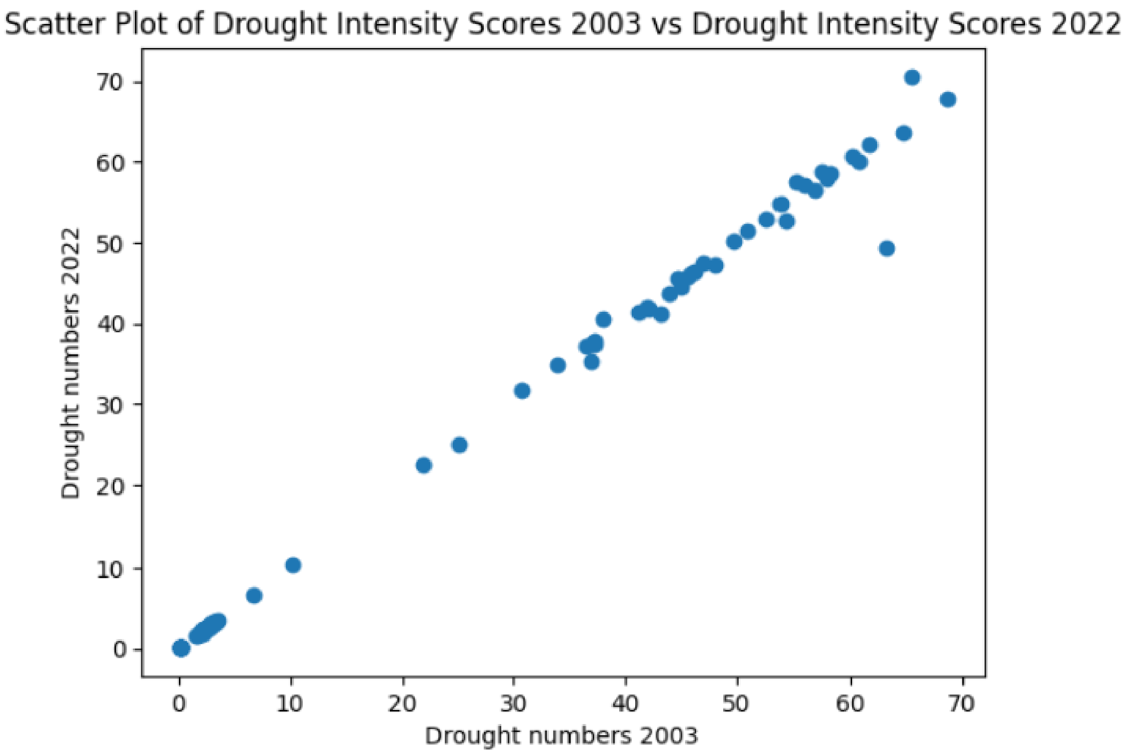


Figure C.5: Comparison original Drought scores

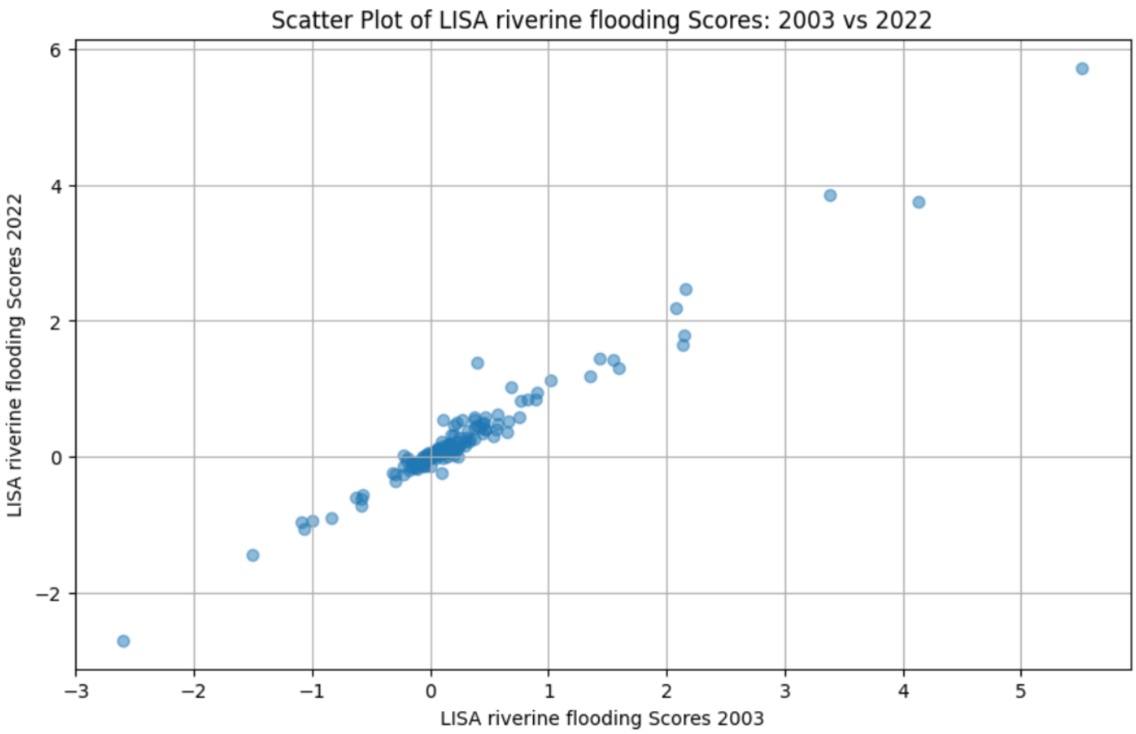


Figure C.6: Comparison LISA Riverine flooding scores

tistical assumptions are examined. The relation between the values looks linear, but with two 'outliers' in both 2003 and 2022, see the plot in Figure C.18.

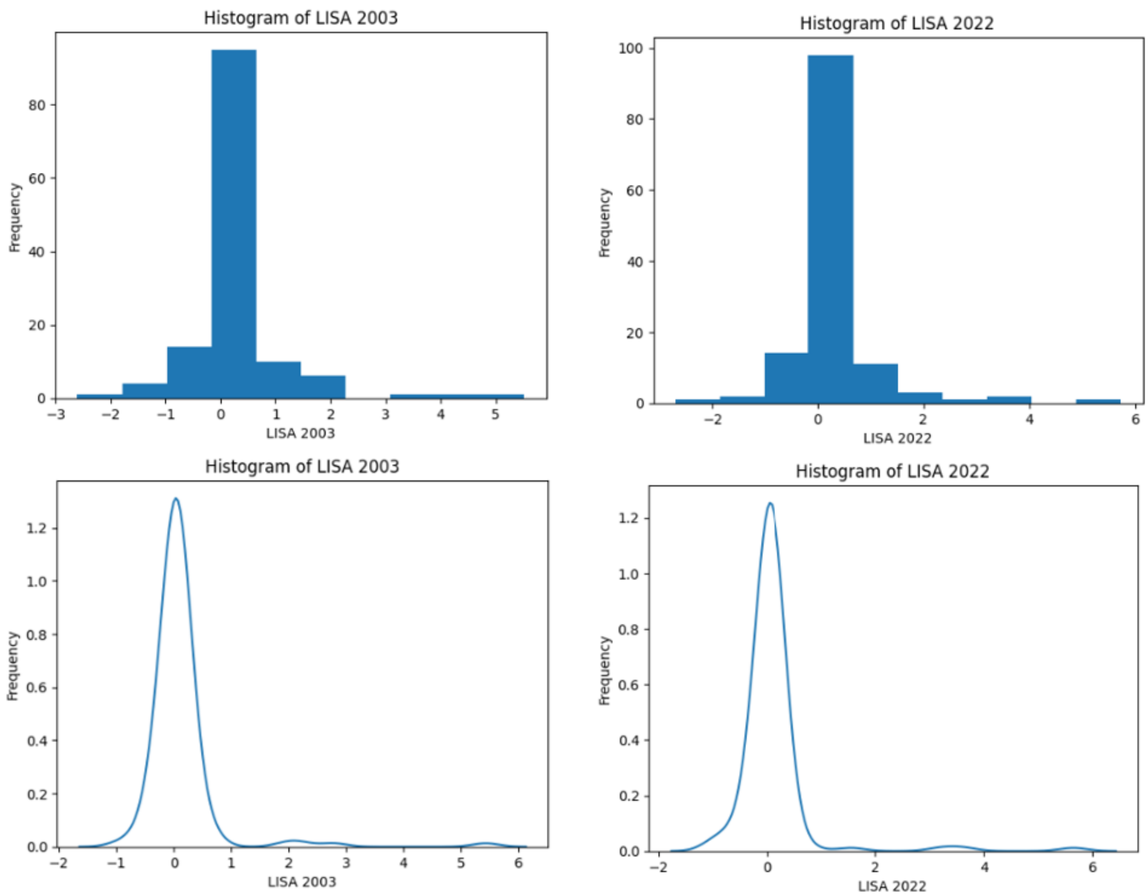


Figure C.7: Distribution LISA Riverine

However, the LISA 2003 and 2022 values are not normally distributed, indicated by the Kolmogorov-Smirnov tests (2003: KS statistic = 0.5006, p-value = 2.3630e-31; 2022: KS statistic = 0.4953, p-value = 1.1693e-30). This can also be seen in Figure C.19. Therefore the Kendall correlation coefficient is used instead of the Pearson correlation.

Also, assumption for the paired t-test are not met, as one can see above (non-normality). Therefore, the Wilcoxon Signed-Rank test is conducted. The results can be found in Chapter 5.

C.5.1. Cyclone changes over years

In this section, the original, regular cyclone intensity risk values are compared in 2003 and 2022, and not the changing clustering values (LISA values), as was done in the previous section.

Before comparing the change in 2003 and 2022 cyclone intensity values, it needs to be examined which correlation coefficient and statistical test can be used. This data is not normally distributed, indicated by the Kolmogorov-Smirnov test (2003: KS statistic = 0.4402, p-value = 2.4609e-24; 2022: KS statistic = 0.4508, p-value = 7.2813e-27) and plots (see Figure C.21).

C.6. Refugees fleeing from a country

This section shows supportive evidence for the refugee origin analysis.

C.6.0.1. Statistical cluster changes

To statistically establish whether the LISA values change over years, first is examined which correlation coefficient and statistical tests should be used. Histograms and Kernel Density Estimate plots (see Figure C.23), and the Kolmogorov-Smirnov tests (2003: KS statistic = 0.3554, p-value = 5.3023e-17;

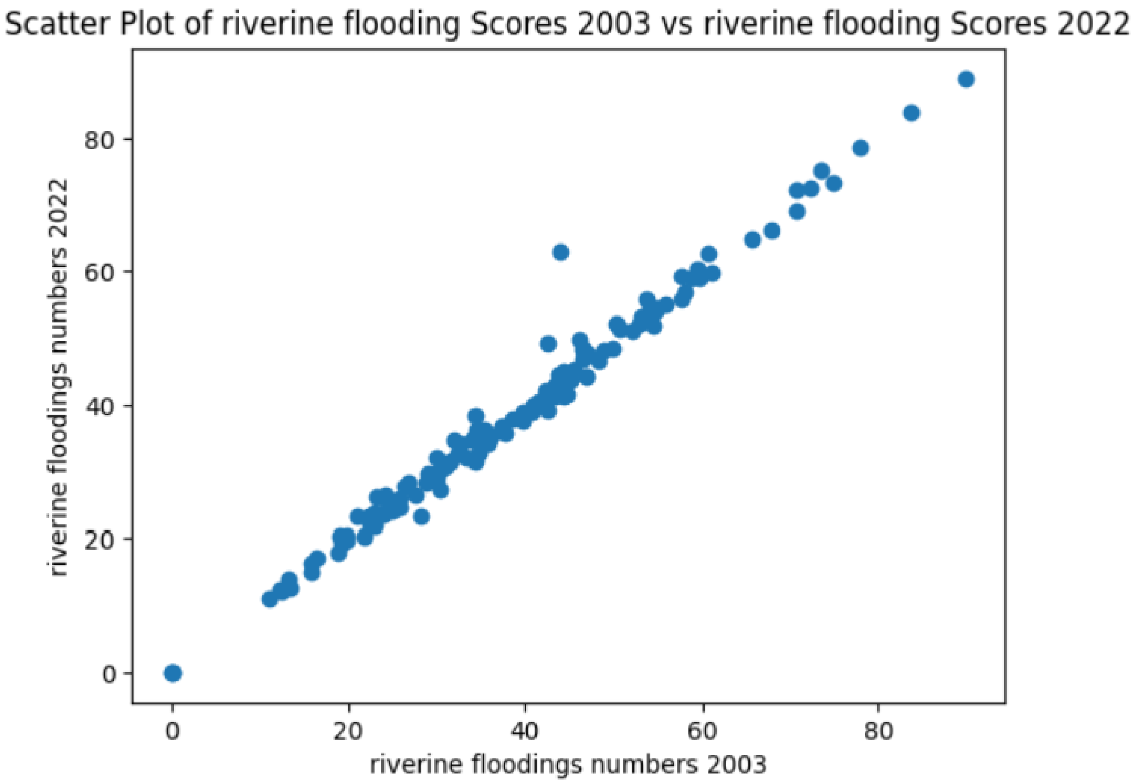


Figure C.8: Comparison original Riverine flooding scores

2022: KS statistic = 0.3223, p-value = 5.3567e-14), indicate that both the 2003 cluster data and the 2022 cluster data, are not normally distributed. Therefore, in this study is looked at the Kendall correlation coefficient and the Wilcoxon Signed-Rank test. Results can be found in Chapter 5.

C.6.1. Changes in refugee origin numbers over years

In the original data, to statistically compare 2003 and 2022 refugees fleeing, first normality is examined. It is clear from the histograms and the KDEs (Figure C.26) that the data is not normally distributed. Therefore, the Kendall correlation coefficient as well as the Wilcoxon Signed-Rank Test are used, as opposed to respectively a Pearson correlation coefficient and paired t-test.

C.7. Refugees fleeing to a country

This section shows supportive evidence for the refugee destination analysis.

C.7.0.1. Statistical cluster changes

The cluster data (LISA values) are not normally distributed as can be seen in Figure C.29. Also, the Kolmogorov-Smirnov tests indicate this (2003: KS statistic = 0.3607, p-value = 5.2389e-16; 2022: KS statistic = 0.3736, p-value = 3.5446e-17). So, to test changes over years, non-parametric tests are used.

C.7.1. Changes in refugee destination numbers over years

To do certain statistical tests, first certain assumptions should be examined. This data is not normally distributed (see Figure C.31). Therefore, the Kendall correlation coefficient is used and a Wilcoxon Signed-Rank Test is conducted.

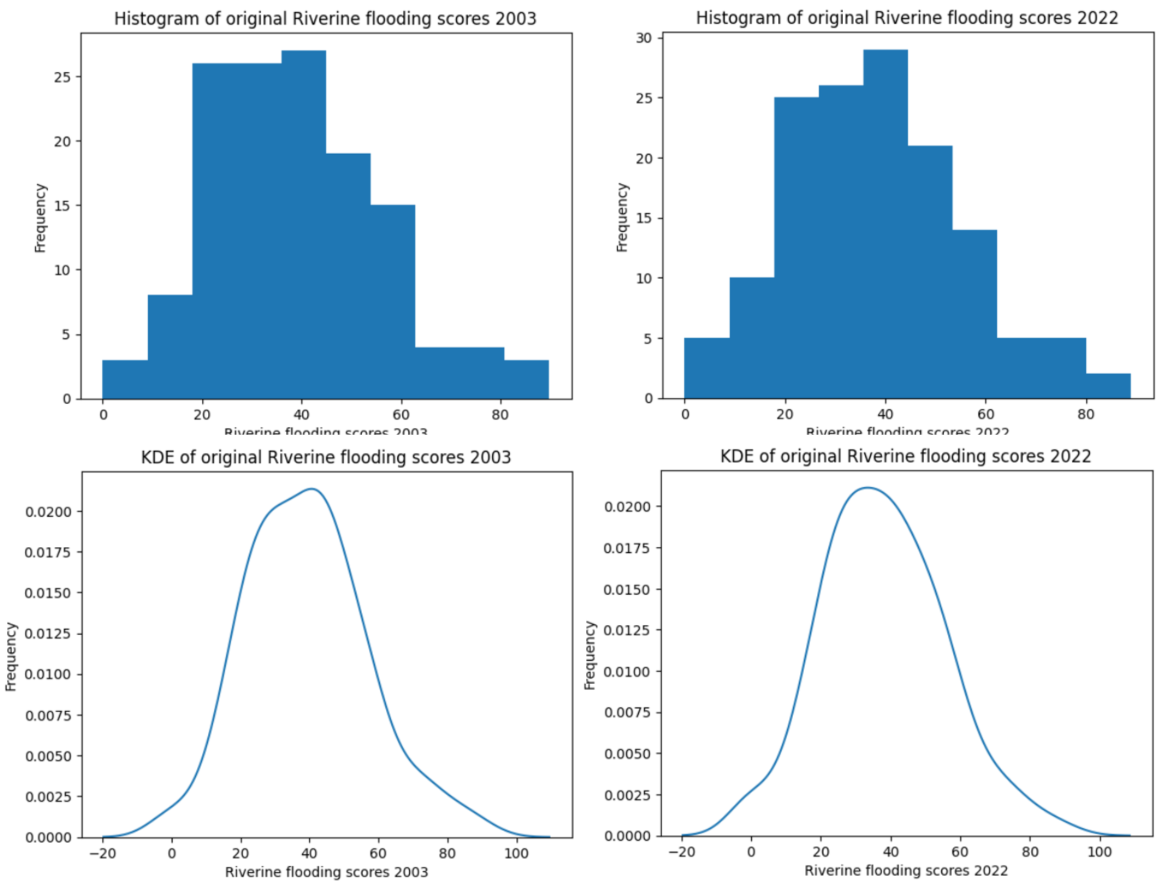


Figure C.9: Distributions original riverine flooding scores

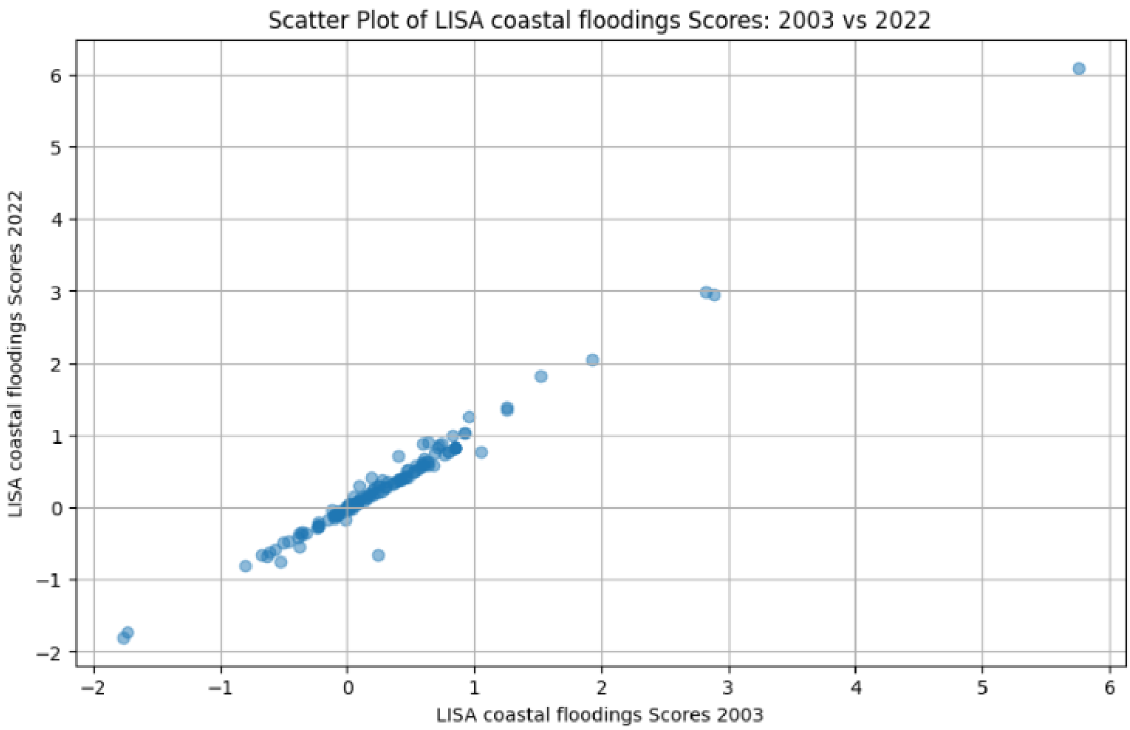


Figure C.10: Comparison LISA Coastal flooding scores

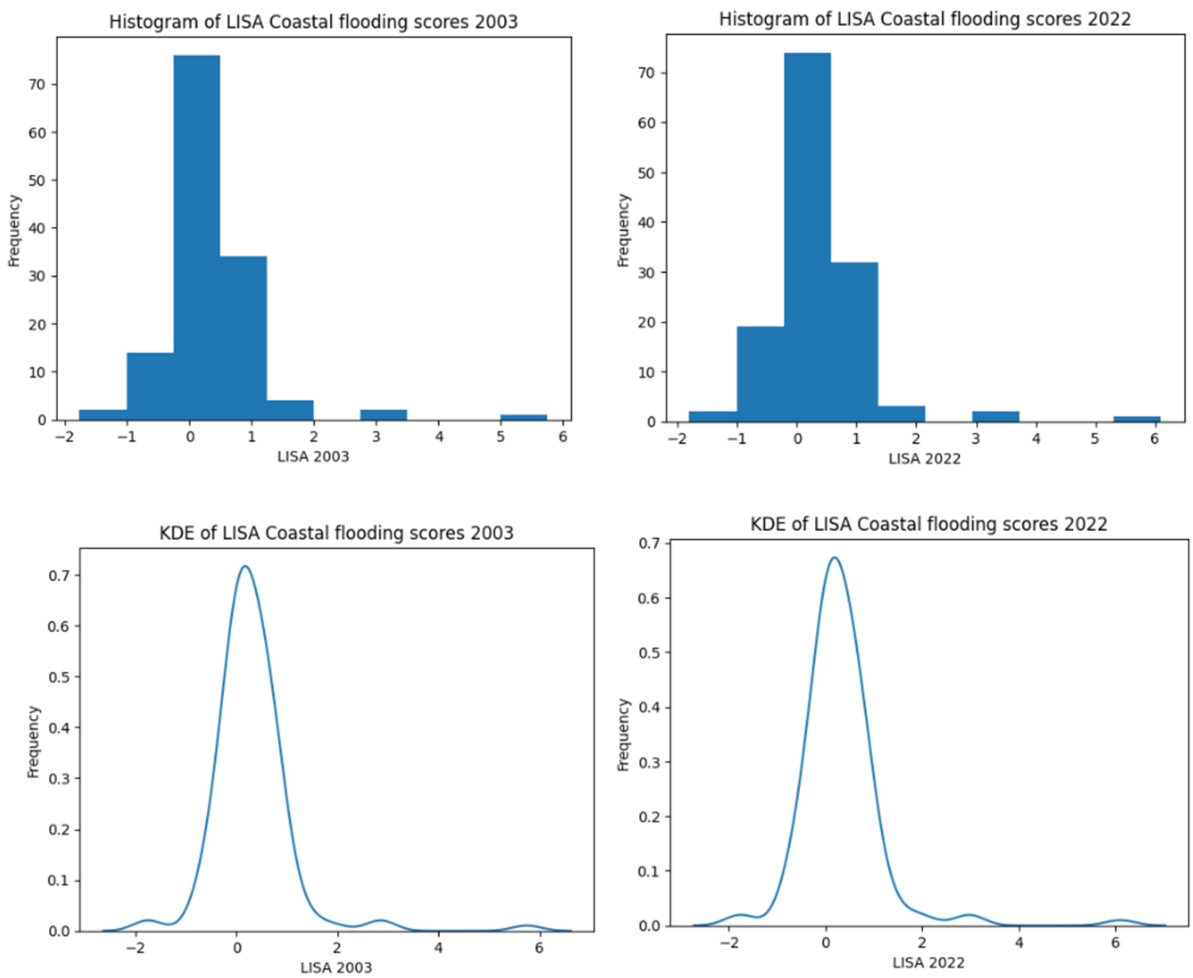


Figure C.11: Distribution LISA Coastal flooding scores

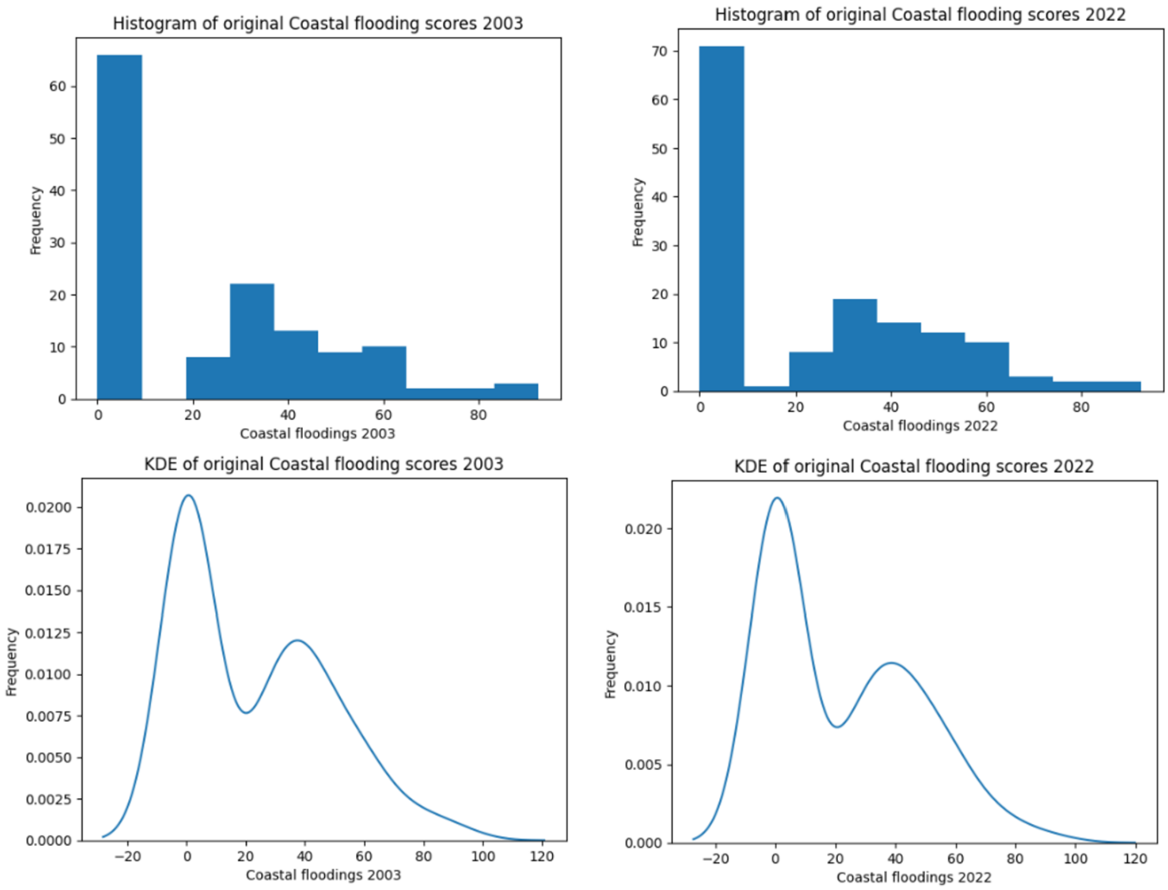


Figure C.12: Distributions original Coastal flooding scores

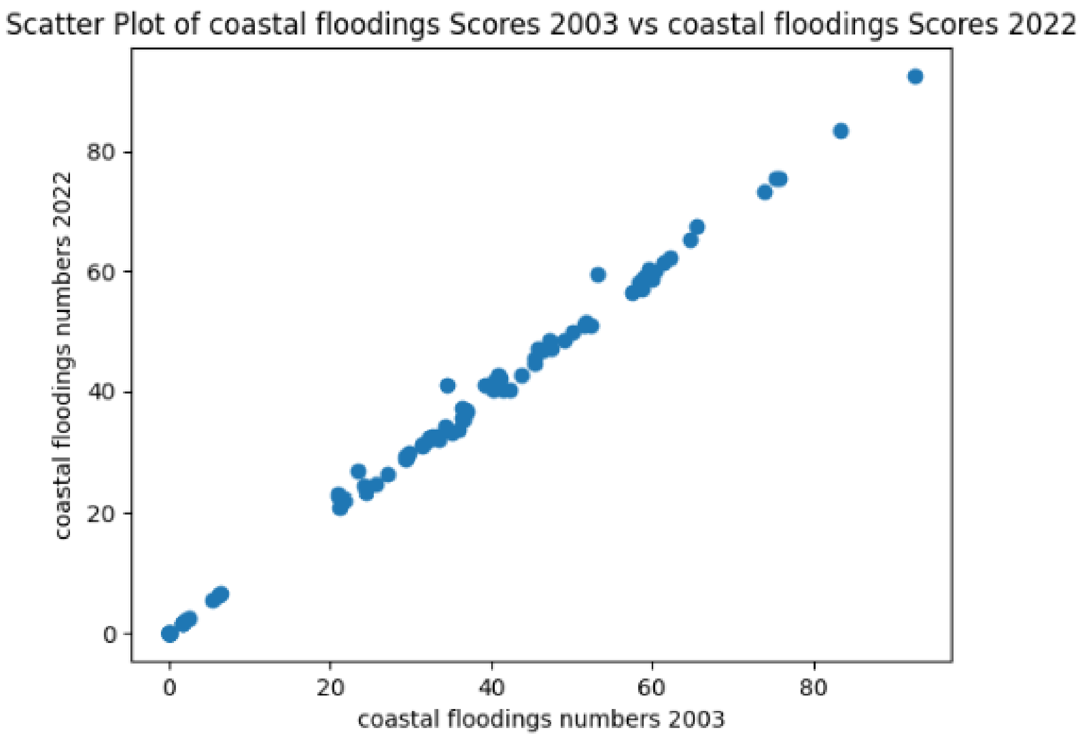


Figure C.13: Comparison original Coastal flooding scores

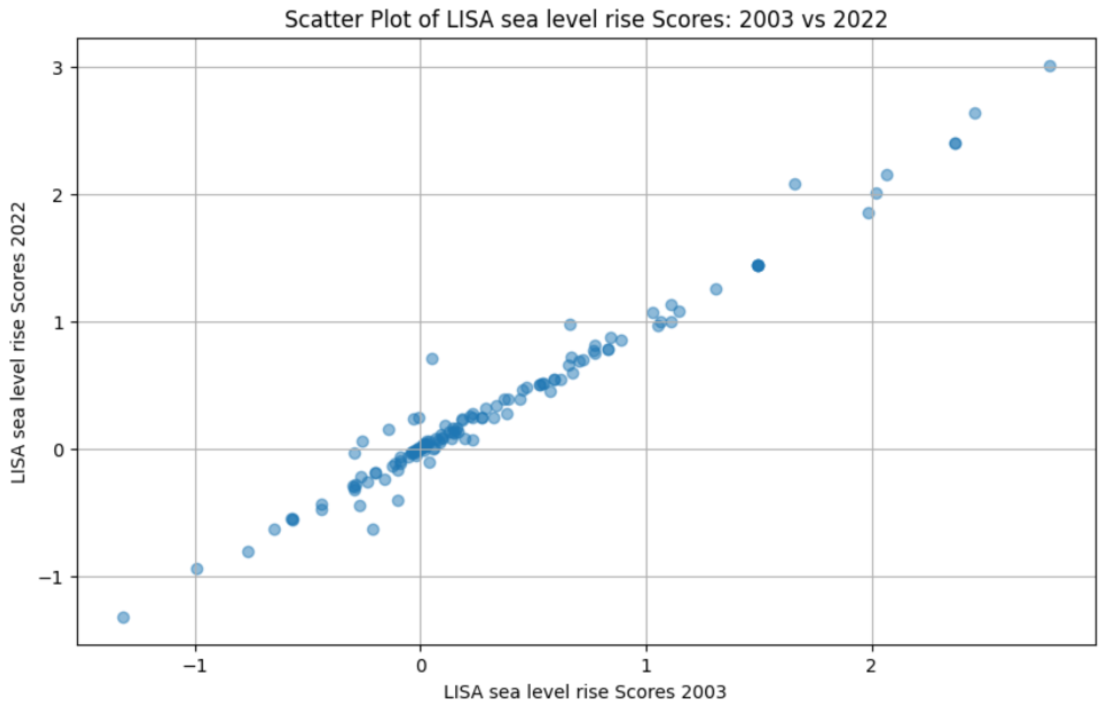


Figure C.14: Comparison LISA Sea level rise scores

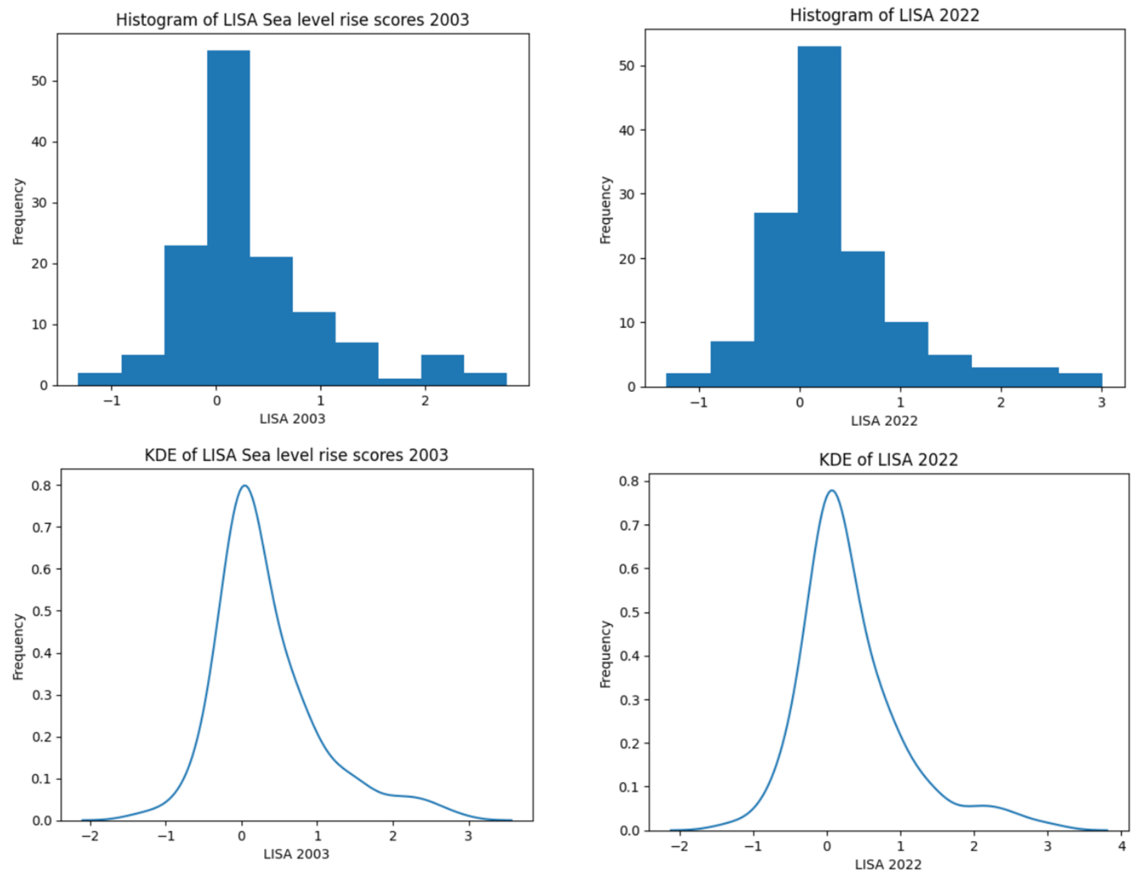


Figure C.15: Distributions LISA Sea level rise scores

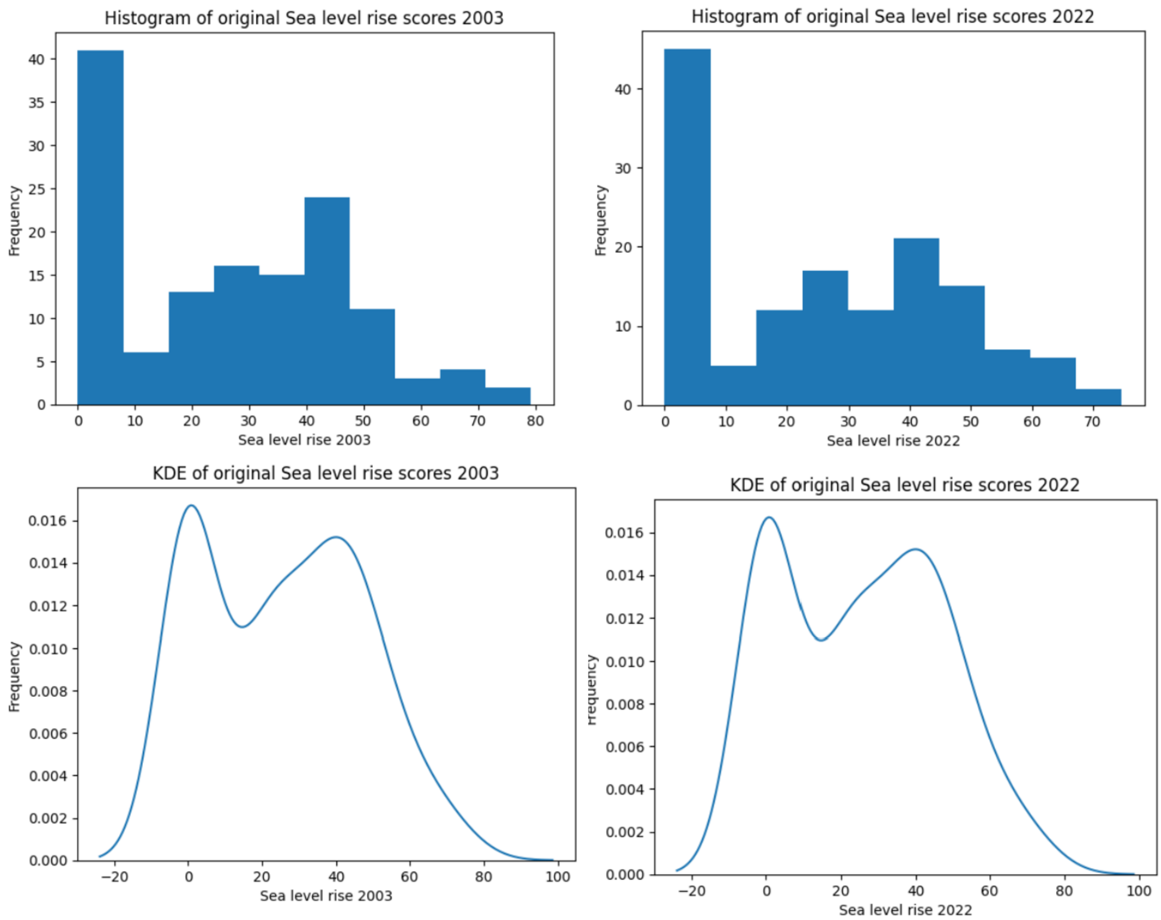


Figure C.16: Distributions original Sea level rise scores

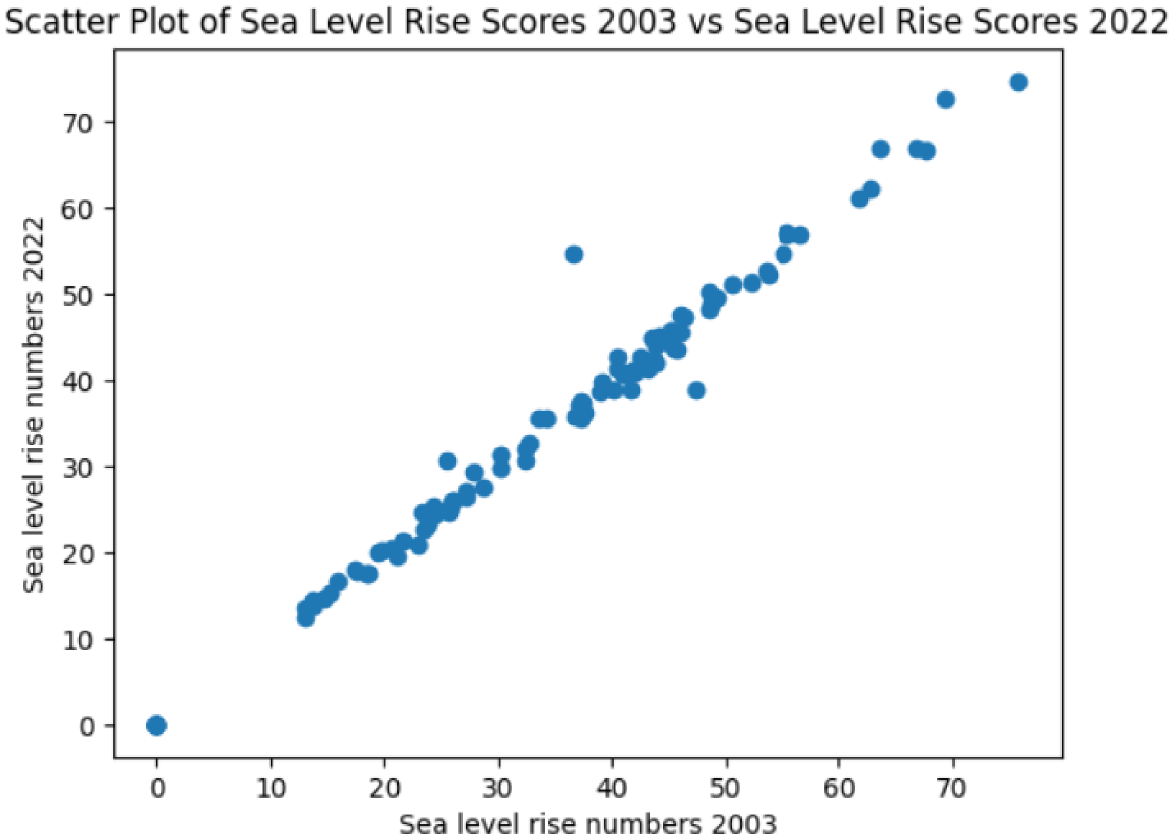


Figure C.17: Comparison original Sea level rise scores

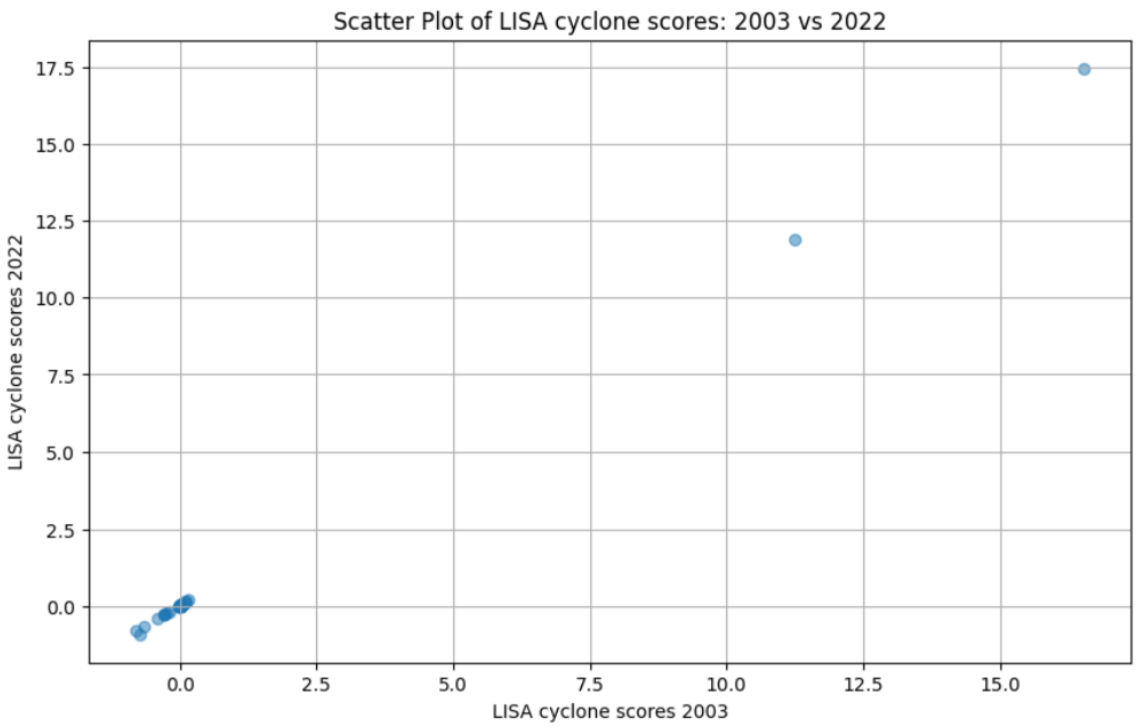


Figure C.18: Comparison LISA Cyclone scores

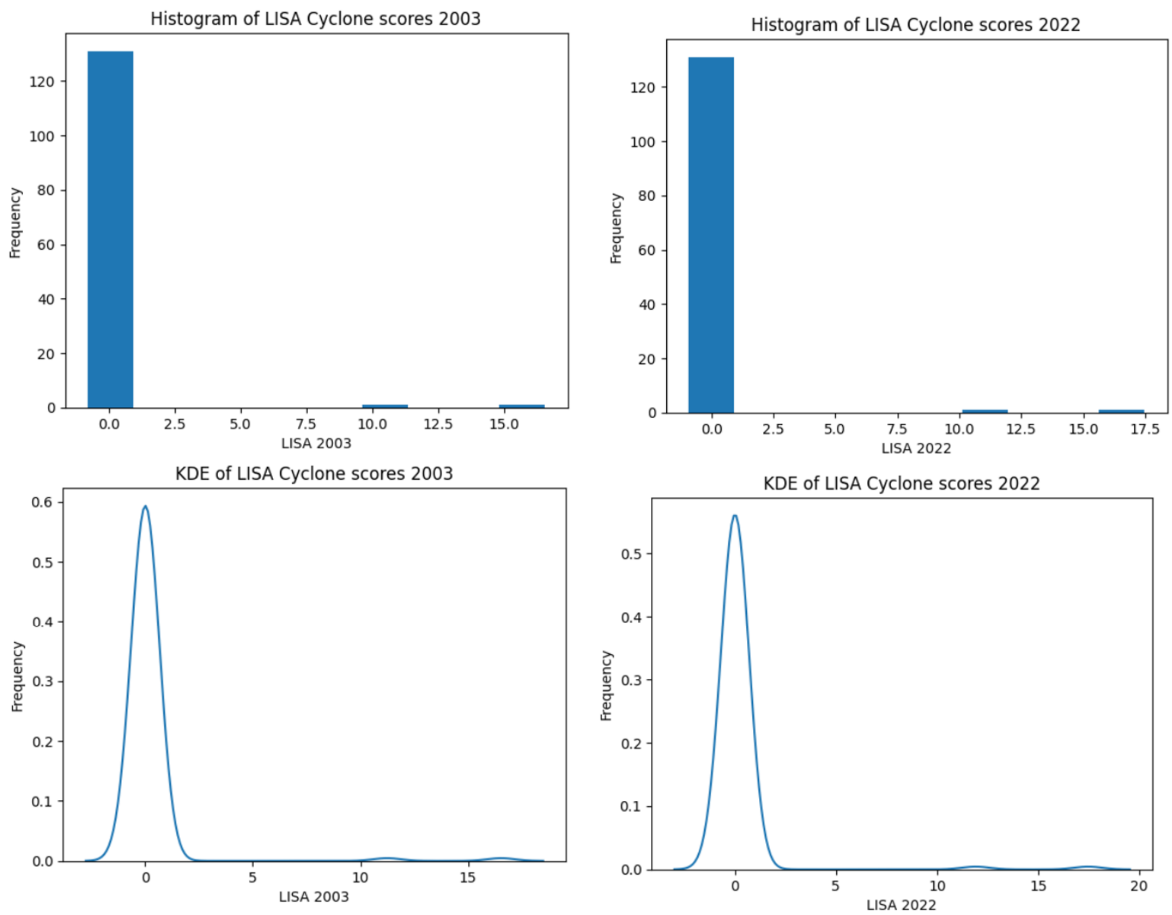
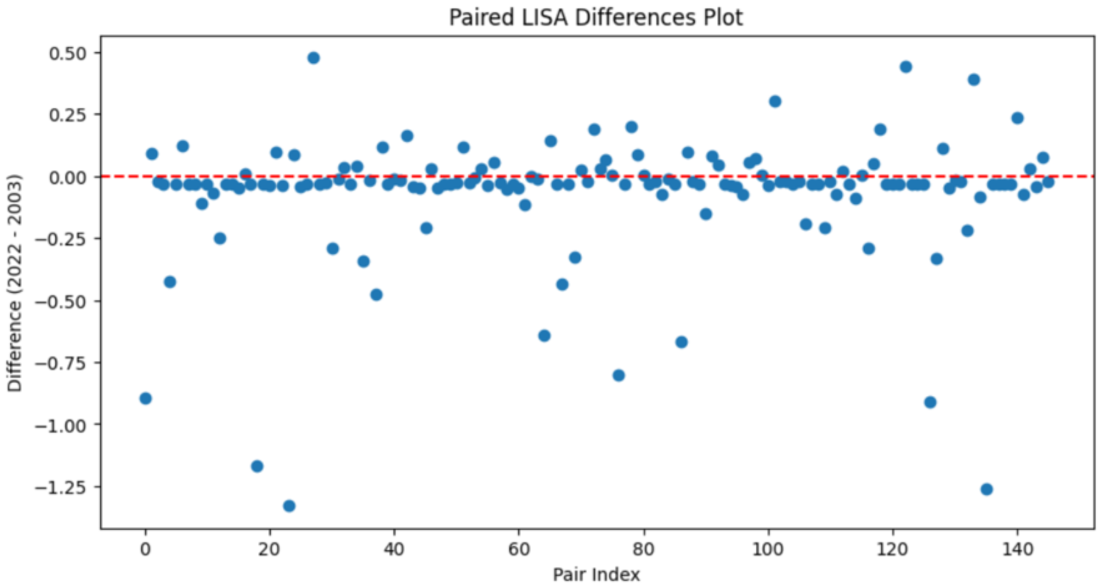


Figure C.19: Distributions LISA Cyclone scores



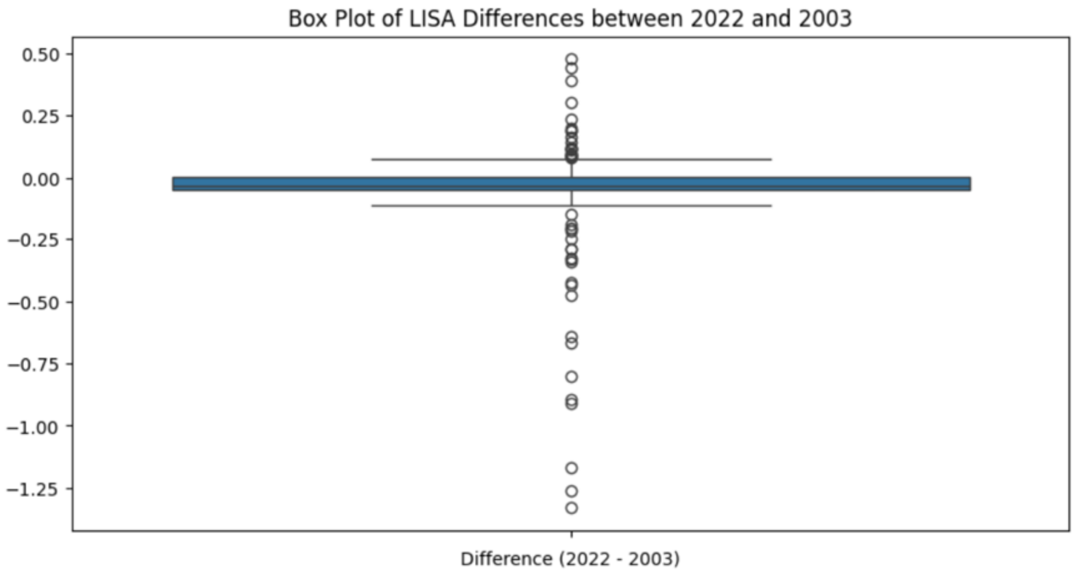


Figure C.20: Differences LISA Cyclones

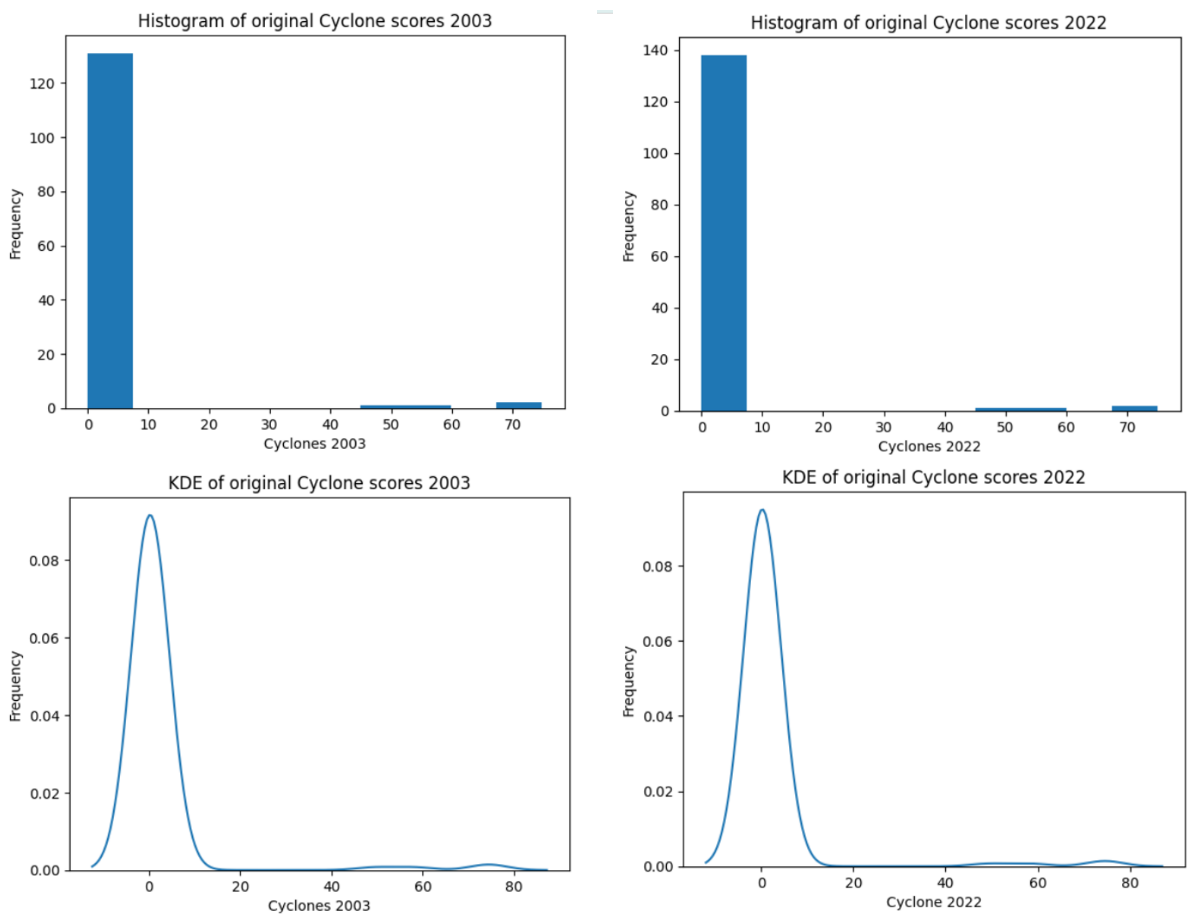


Figure C.21: Distributions original Cyclone scores

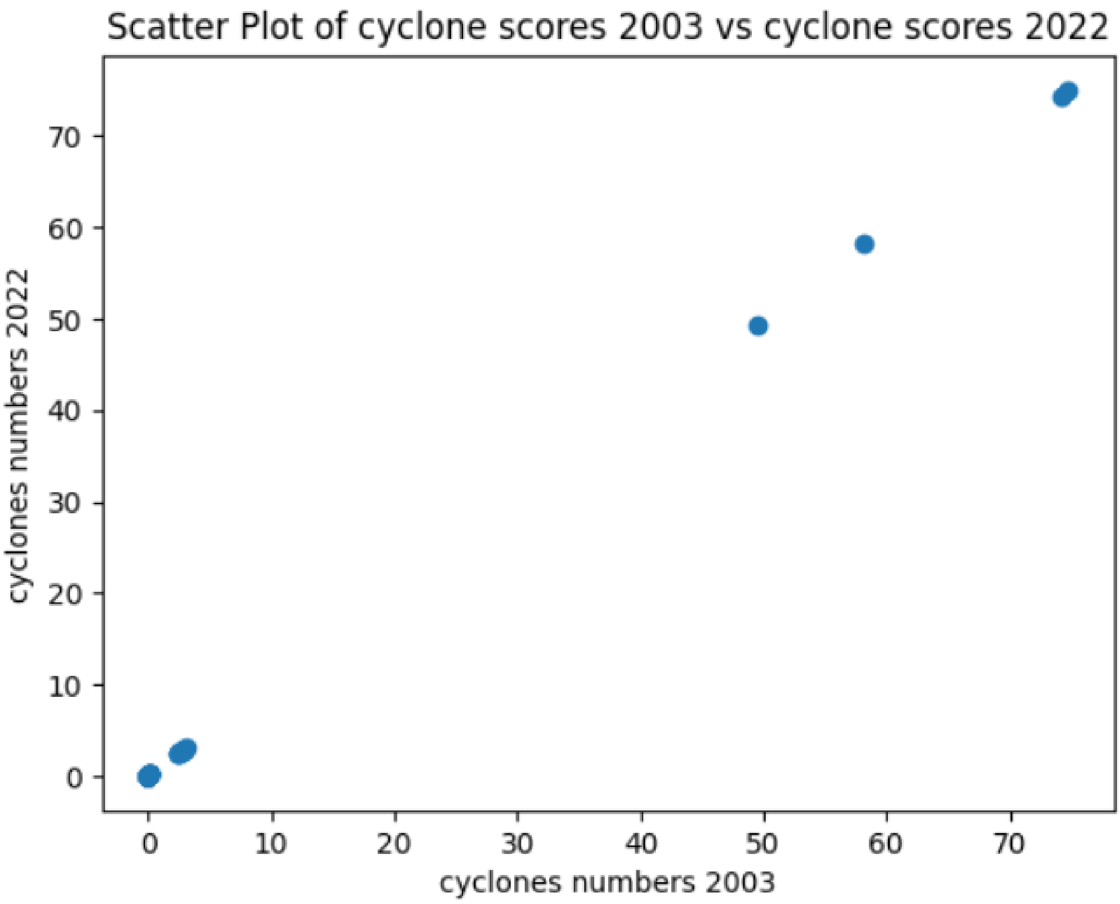


Figure C.22: Comparison original Cyclone scores

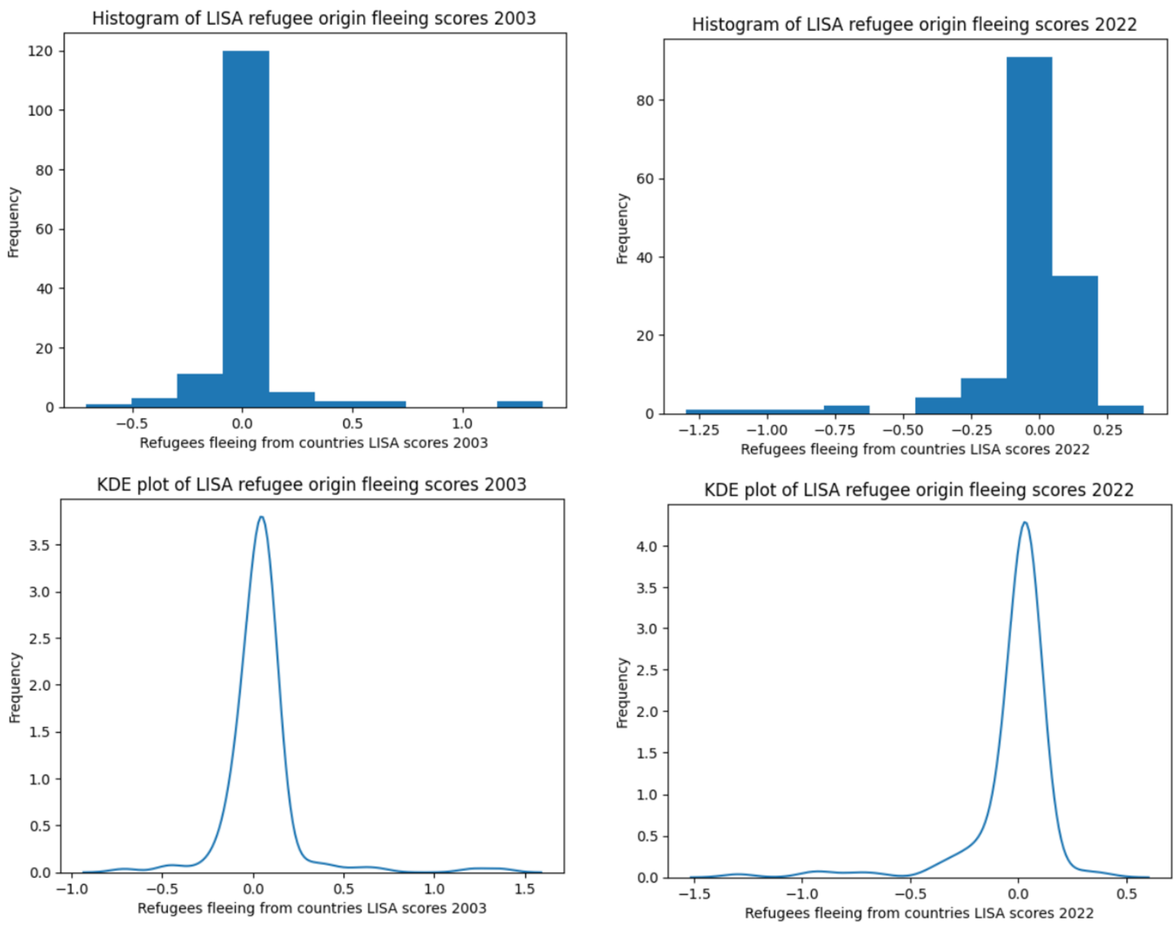


Figure C.23: Distributions LISA refugee origin scores

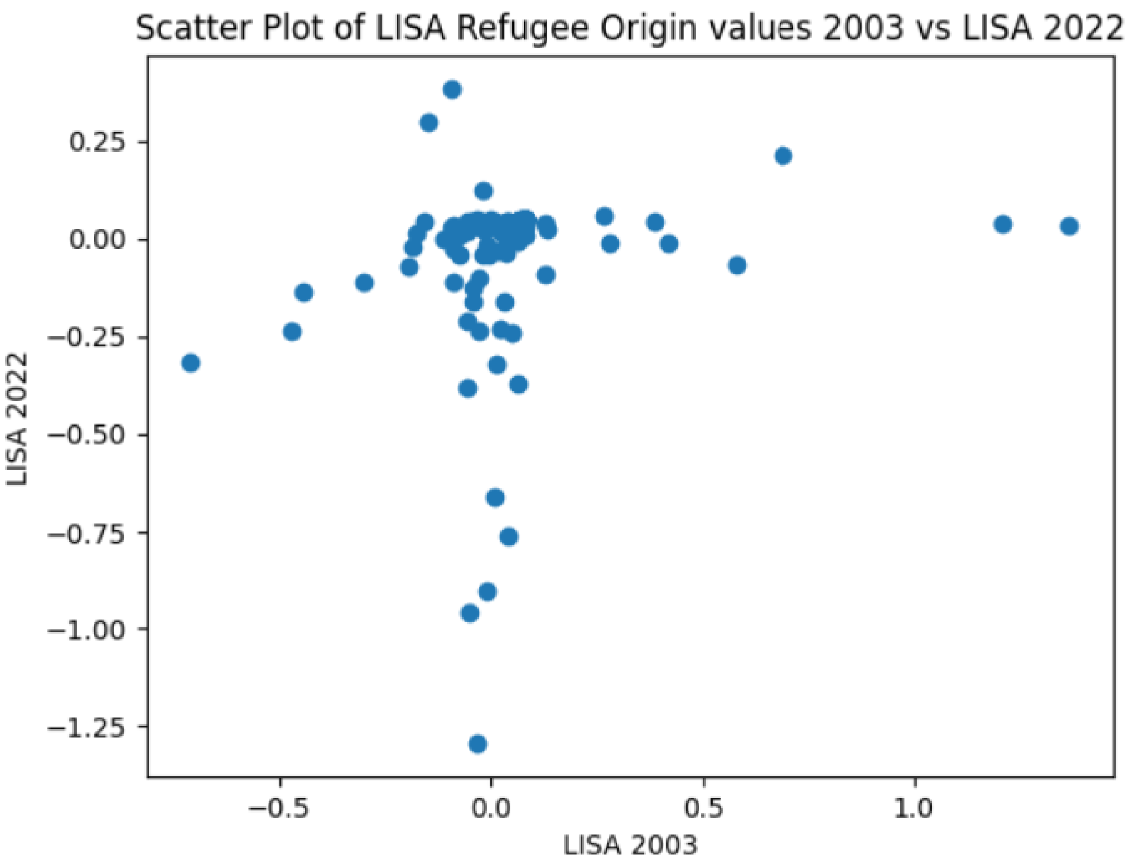
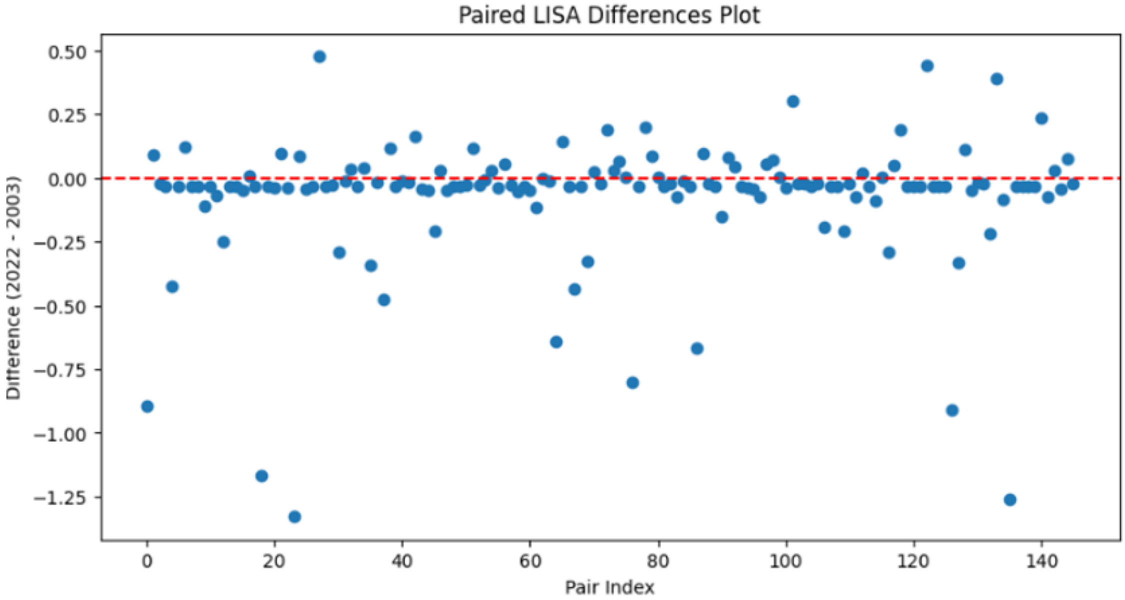


Figure C.24: Comparison LISA refugee origin scores



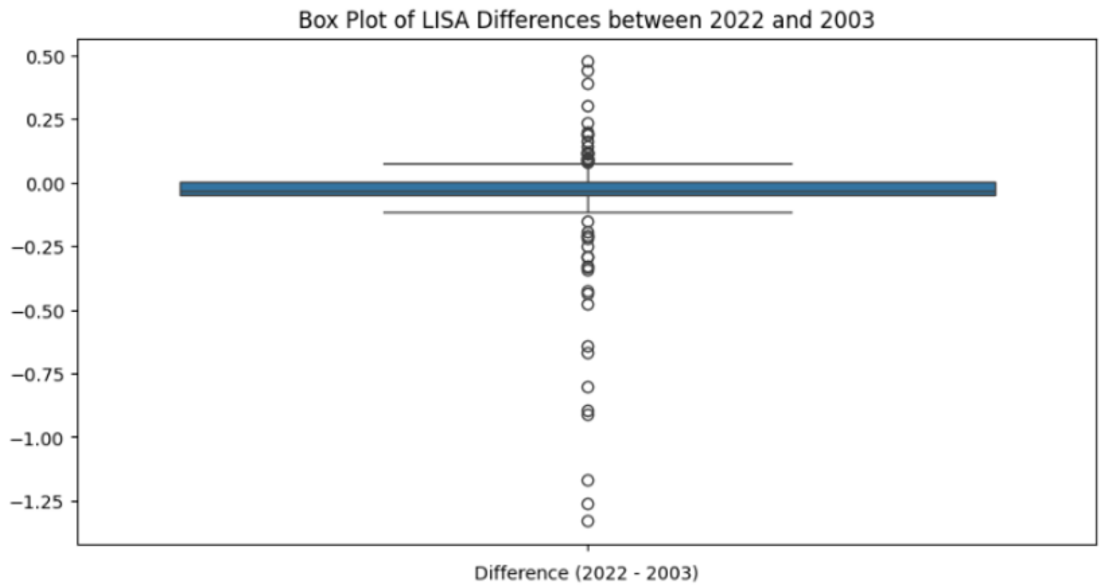


Figure C.25: Difference plot LISA values refugees origin

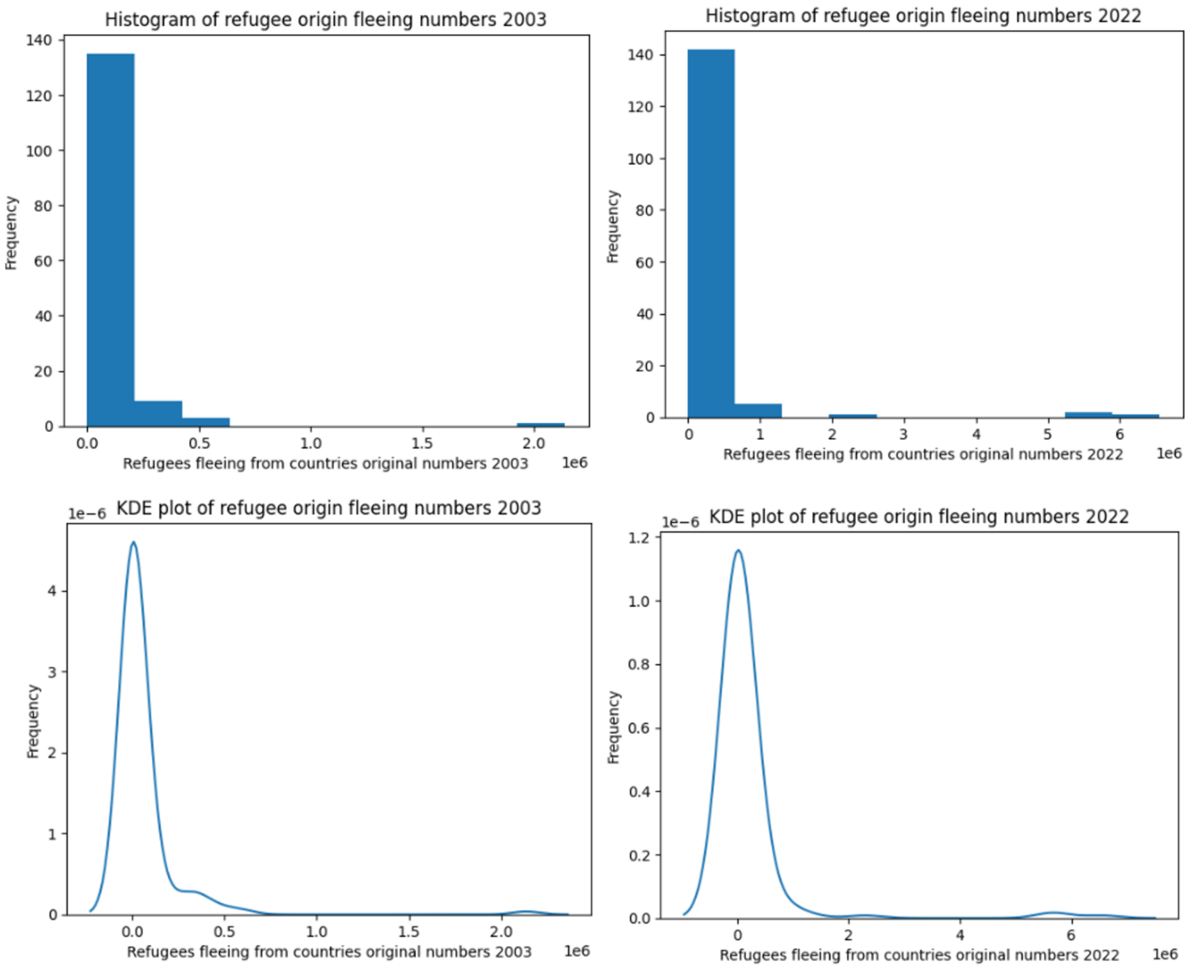


Figure C.26: Distributions original refugee origin values

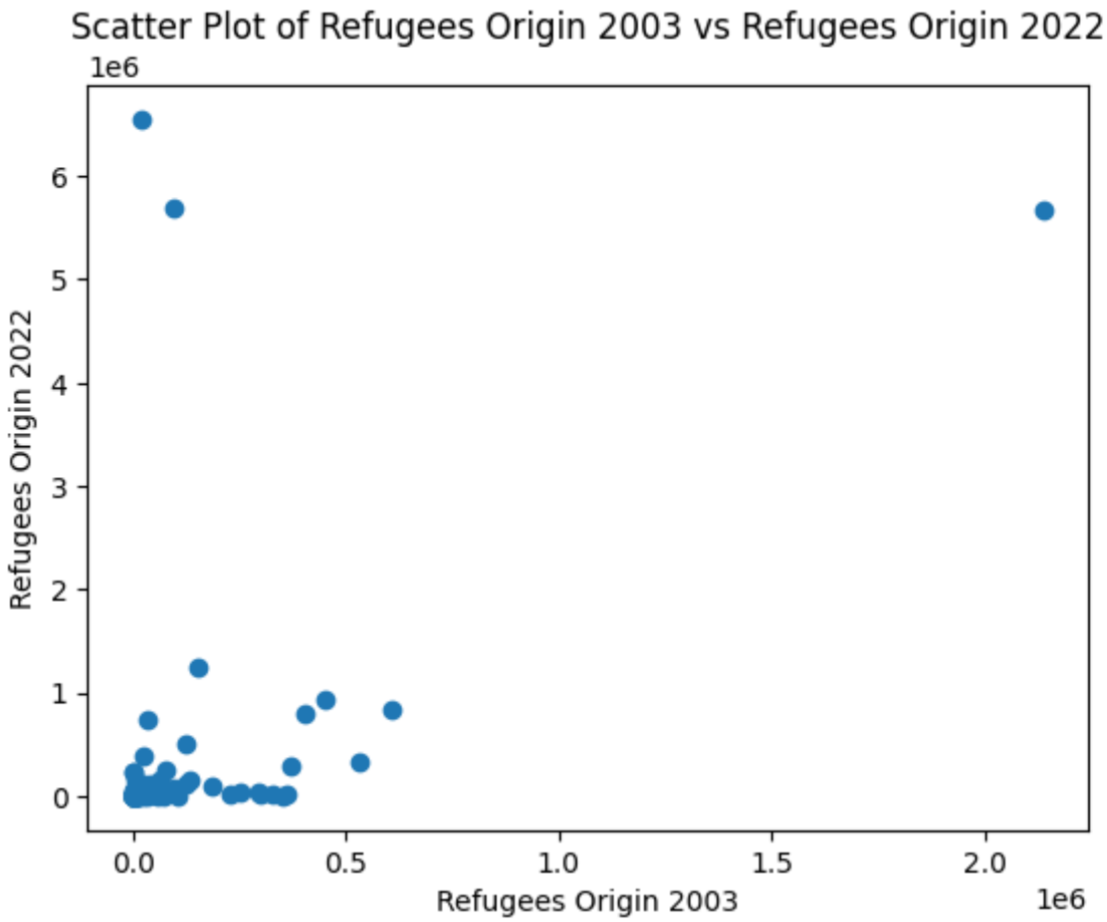


Figure C.27: Comparison of Refugees Origin 2003 and 2022

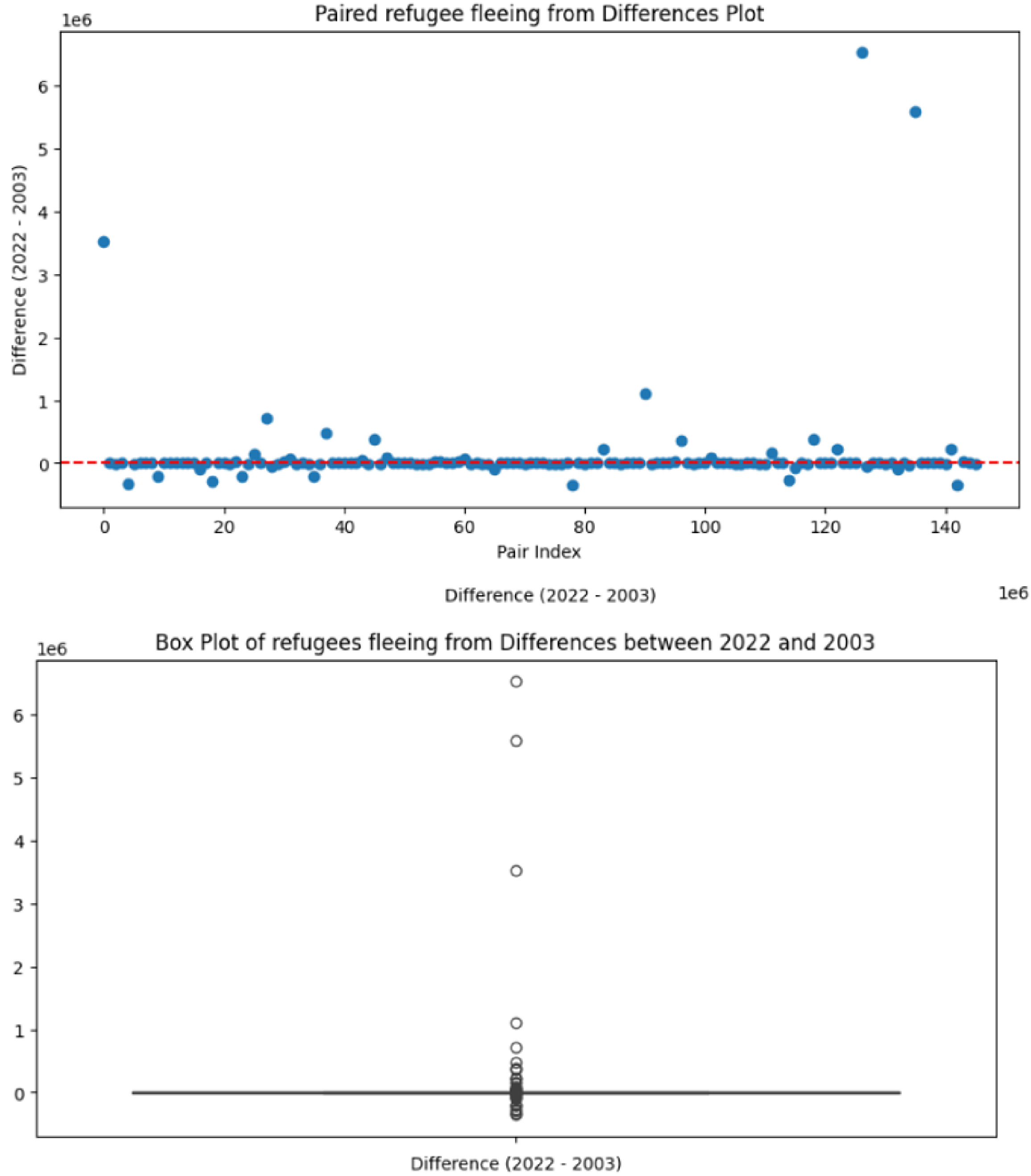


Figure C.28: Difference plot original refugee origin values

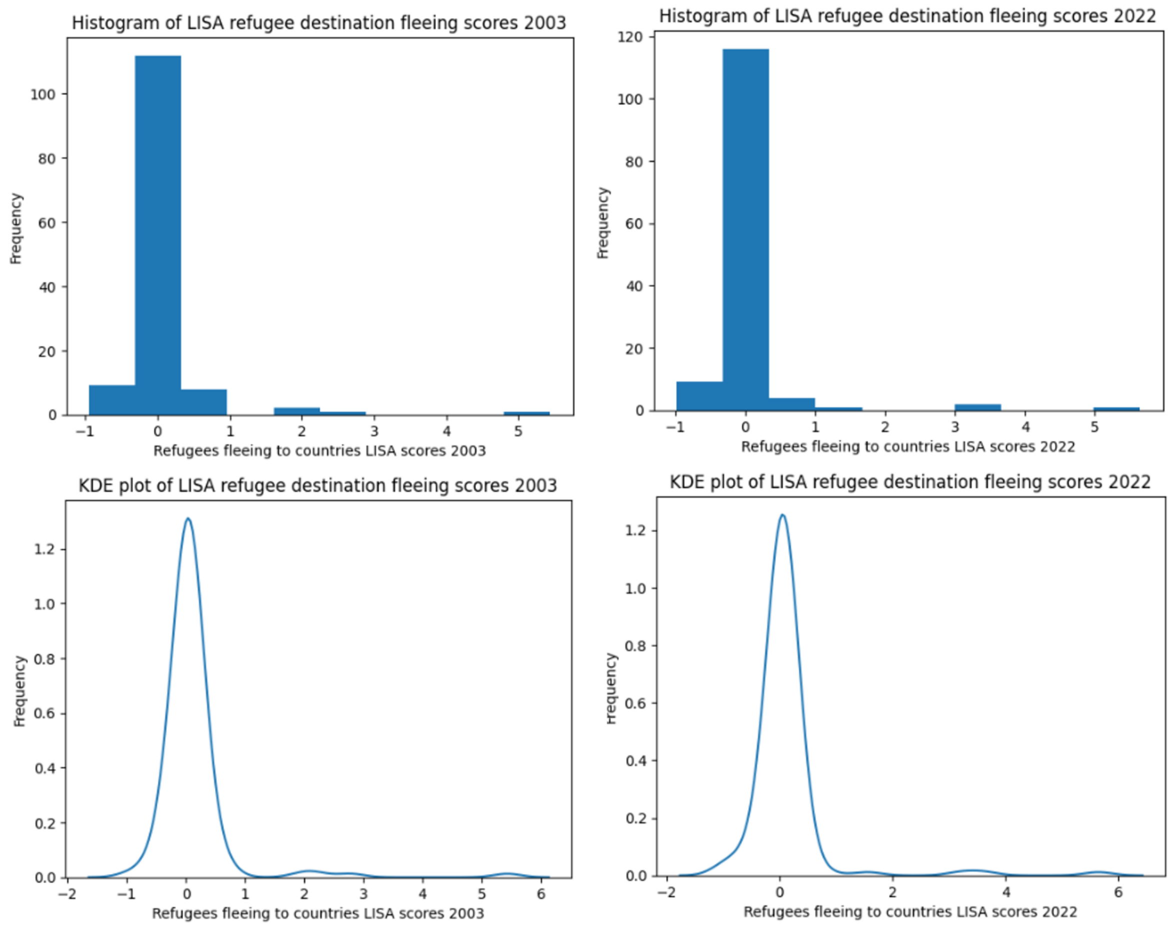


Figure C.29: Distributions LISA refugee destination values

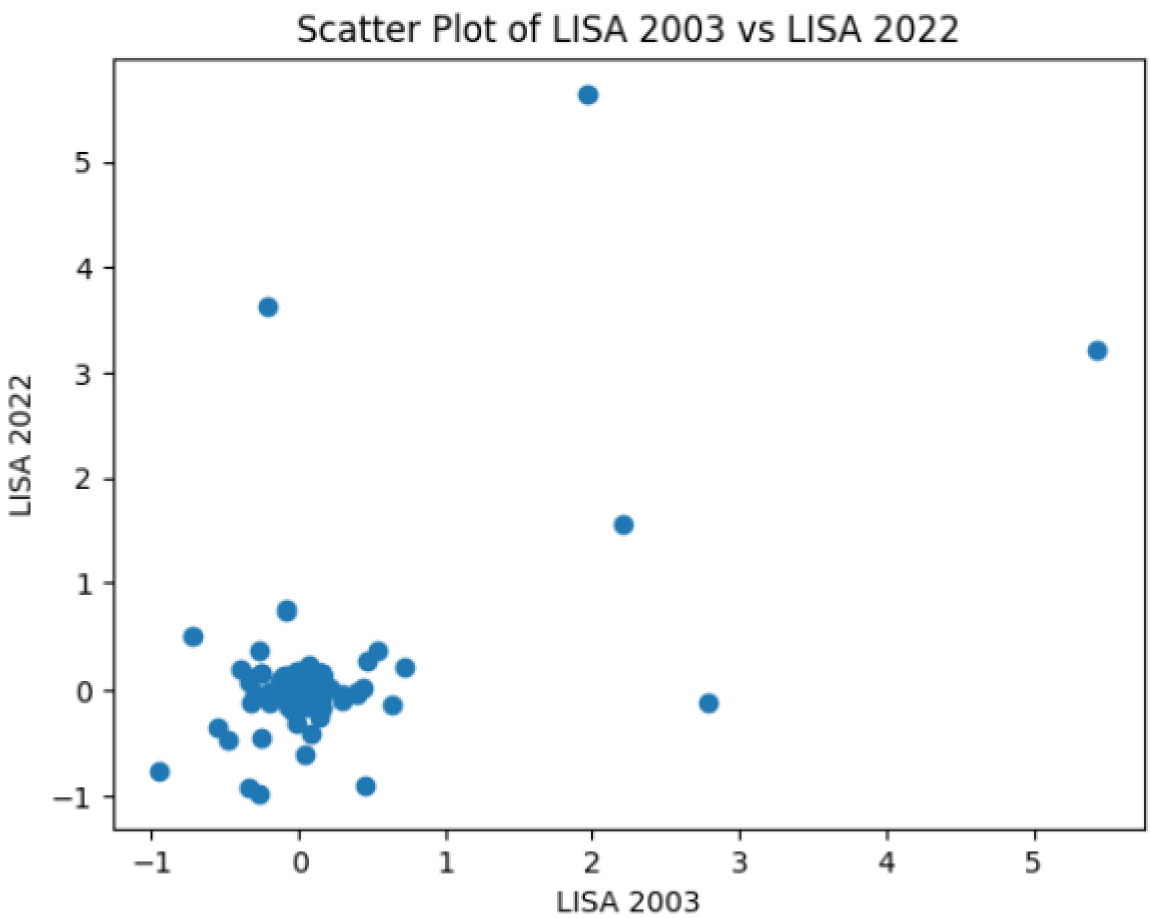


Figure C.30: Comparison LISA refugee destination values

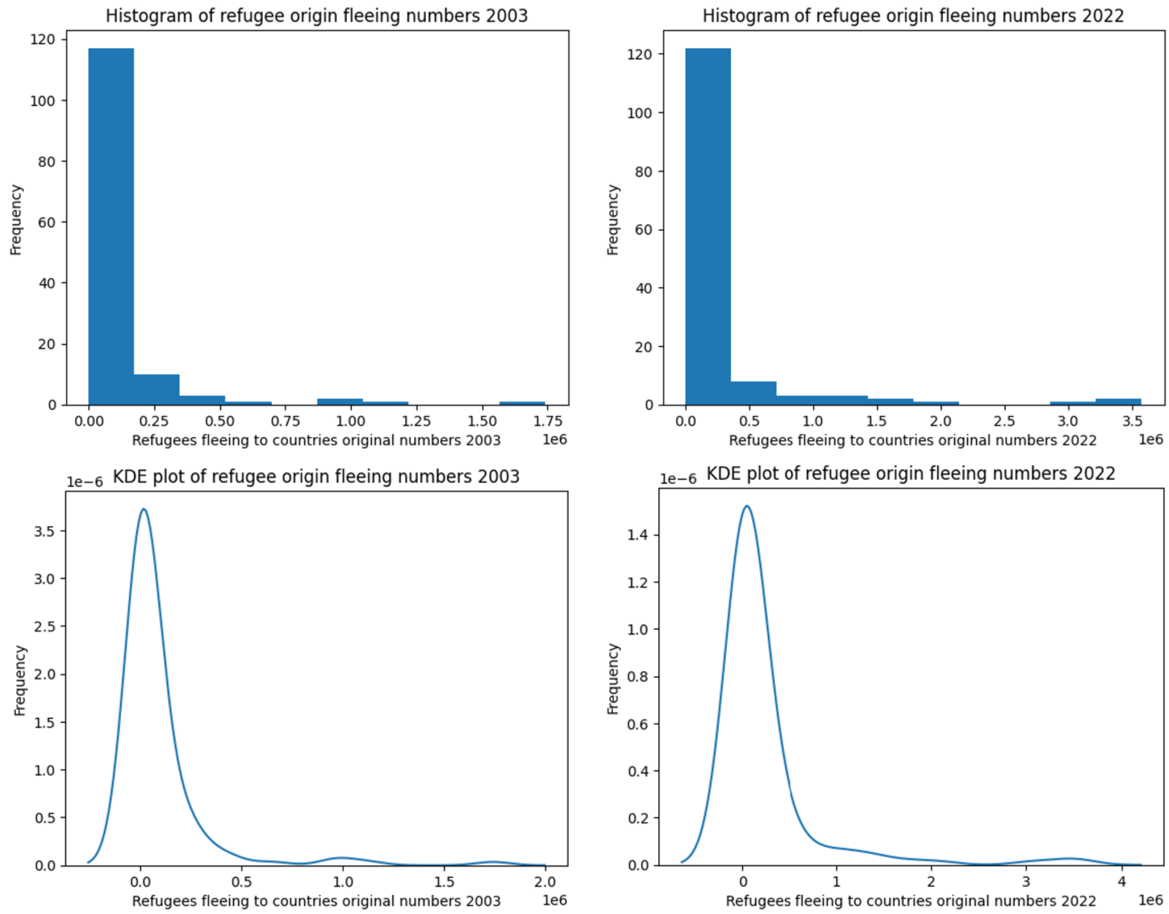


Figure C.31: Distribution original refugee destination values

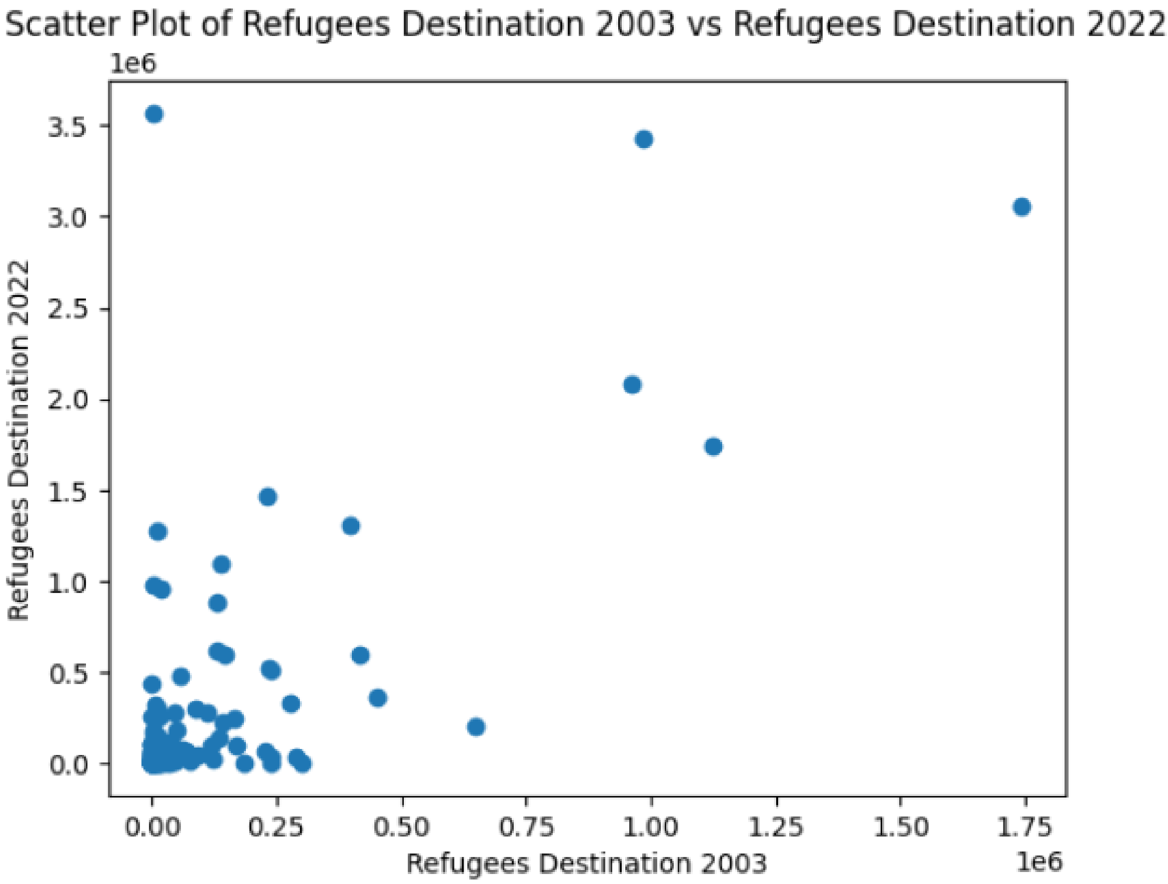


Figure C.32: Comparison original refugee destination values

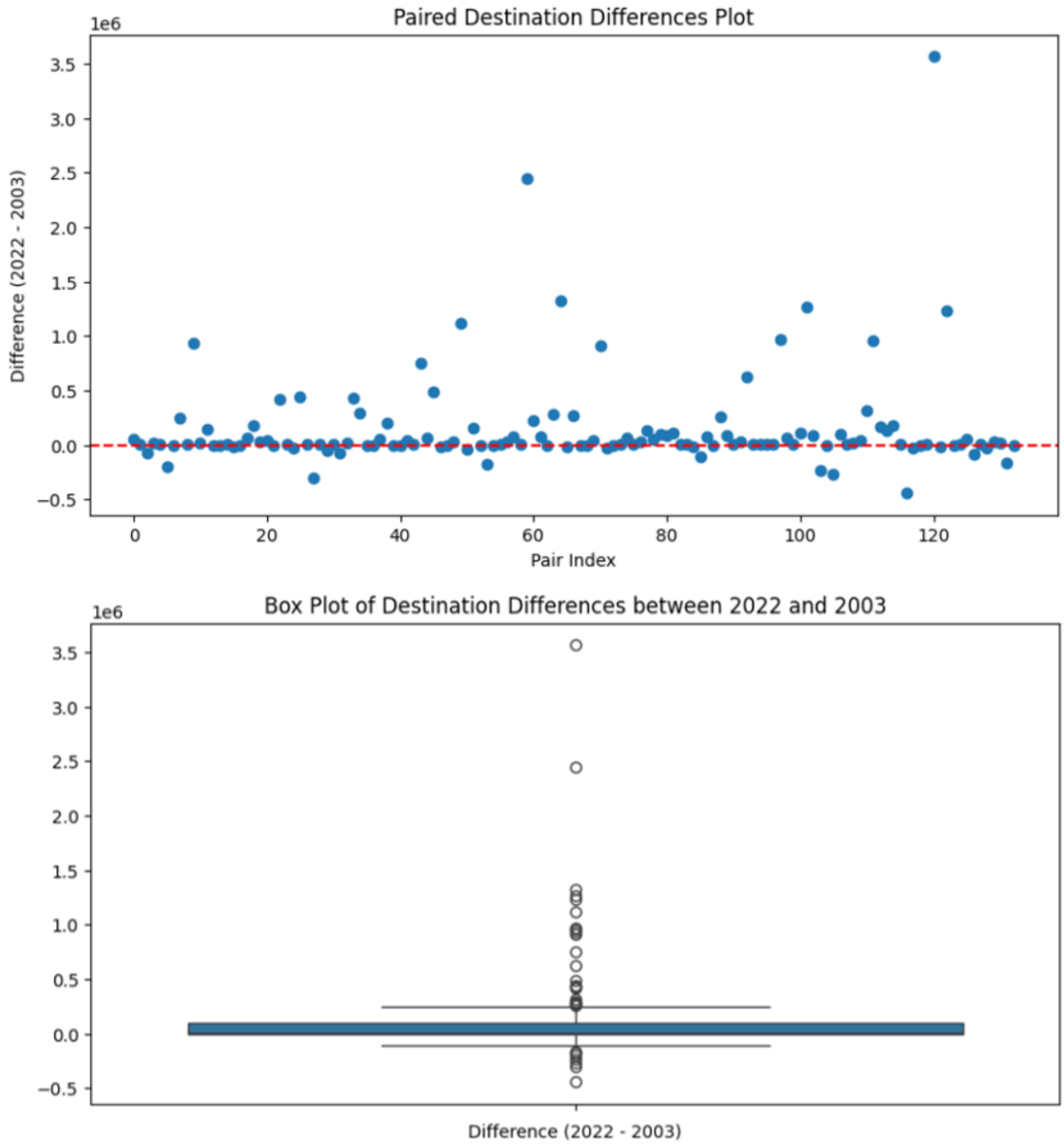


Figure C.33: Differences original refugee destination values

D

Supportive results Chapter 6: Comparing Origin and Destination locations of Refugees

D.1. Droughts

D.1.1. Droughts: Comparing origin-destination of refugees

D.1.1.1. Droughts: Statistical generic difference tests

Due to the overlap between the top origin and destination locations, it becomes difficult to conduct statistical tests to map out generic differences. However, two options remain possible (which themselves also consist of two options). Although these tests are not ideal, all four approaches are employed and combined to determine if consistent conclusions can be drawn.

First category of options
1) Testing differences of means
Compare the median differences of the means. The definition of this is that, for every year, one takes the mean drought score for all the countries in group origin and group destination, and gathers this over 20 years from 2003 till 2022. Then one can compare whether the means of these groups over year differ. One basically compares all years in one then, and threats the mean score as individual data point in the comparison. This is not specific, and you do not see this often, but one can identify with one statistic if the median of the means over time differs. One can then conclude if the drought mean scores over years are statistically different between top origin and destination locations of refugees. The main idea is that means do not overlap, as opposed to the original, raw values.

At first is shown that the mean drought scores of groups do not both follow the normal distribution and that therefore a non-parametric test has to be used. Not both groups have means that are normally distributed when gathered over years, so the Mann-Whitney U test is used instead of a t-test. A Shapiro-Wilk test is used to reflect on normality, because the sample size is smaller than 50. Otherwise, a Kolmogorov-Smirnov (KS) test should have been used (Mishra et al., 2019). (Origin means: W statistic = 0.8267, p-value = 0.0022; destination means: W statistic = 0.9116, p-value = 0.0683.) The origin means are not normally distributed, only the destination means are. That is why medians are compared instead of means (Mann-Whitney U vs independent t-test.) Results can be found in the main text in Chapter 6.

2) Testing differences of medians
The reasoning is the same as above. This test is conducted to determine whether the same conclusion holds when gathering all the medians over years and then comparing whether the median of those medians differs.

Also the medians of both groups over years are not both normally distributed. (Origin medians: Shapiro-Wilk test: W statistic = 0.5026, p-value = 3.3558e-07; destination medians: Shapiro-Wilk test: W statistic = 0.2978, p-value = 7.3749e-09.) Both are not normally distributed, the null hypotheses are rejected. So, again, the Mann-Whitney U test is used. Results can be found in the main text in Chapter 6.

Second category of options
In this section, the one sample/signed (non-)parametric test for both groups is conducted, over all years. That is to conclude also per year on the differences between groups.

What also can be done as a generic test, besides the individual comparison plots that are also shown in this chapter later, is taking one of the medians of the top 25 groups and test whether there is a difference with the values of the other group. Median is written here, because a one sample signed Wilcoxon rank test is used, due to not fulfilling conditions of parametric-test for the original, raw data. For example, the top 25 data in 2020 is not normally distributed in terms of Drought values, as one can conclude from the Shapiro-Wilk tests (origin values: W statistic = 0.7824, p-value = 0.0001; destination values: W statistic = 0.6597, p-value = 2.1906e-06). This can be done for every year. In this analysis, not two data columns are compared and conclusions about their difference in medians are drawn (because then the overlap assumption problem emerges), but the median of one of the groups is taken as given/constant, and the values of the other group are analyzed to conclude if the median of those values differs from the median that was treated as constant. That is how the non-parametric one sample Wilcoxon signed rank test works (King & Eckersley, 2019).

Results and interpretations are presented in the main text in Chapter 6.

D.2. Sea level rise

D.2.1. Sea level rise: Comparing origin-destination of refugees

D.2.1.1. Sea level rise: Statistical generic difference tests

Due to the above, it becomes difficult to do statistical tests to test generic differences between top 25 origin and destination countries. However, there are two options that are still available (which themselves also consist of two options). Those are not very elegant, so both categories (four tests) are conducted and results are combined to see if the same conclusions can be drawn.

First category of options
1) Testing differences of means
The means do not ‘overlap’ between groups. Therefore, the idea is to gather the mean values of every year groups and compare those to test the differences between them. So, instead of original values, means are used. Then, a Mann-Whitney U test or an independent t-test can (still) be conducted. To determine which test needs to be conducted, again, assumptions need to be tested. At first, normality is tested. A Shapiro-Wilk test is used to reflect on normality, because the sample size is smaller than 50. Otherwise, a Kolmogorov-Smirnov (KS) test should have been used (Mishra et al., 2019). (Origin means: W statistic = 0.8254, p-value = 0.0021; destination means: W statistic = 0.8982, p-value = 0.0382.) One can immediately conclude that the data is not normally distributed, as the null hypotheses cannot be rejected. Both groups have no means that are normally distributed, when gathered over years. Therefore, the Mann-Whitney U test will be conducted. This means that the median differences of the means are compared. The definition of this is that, for every year, one takes the mean sea level rise score of all the countries in group origin and group destination, and gather this over 20 years from 2003 and 2022. Then one can compare whether the means of these mean-groups over year differ. One basically compares all years in one then, and threat the mean score as individual data point in the comparison. This is not specific, and you do not see this often, but one can identify with one statistic if the median of the means over time differs between top origin and destination countries. One can conclude if the sea level rise risk mean scores over the years are statistically different between origin and destination groups. The test is basically used to determine if there are statistically significant differences between two groups (MacFarland, Yates, MacFarland, & Yates, 2016). It tests if two independent groups are from the same population. In other words, of the gathered means over years can come from the same population values. “For practical purposes, the user may think of it informally as testing whether or not the two distributions have the same median” (Riffenburgh, 2012).

Results can be found in Chapter 6.

2) Testing differences of medians
The reasoning is the same as above. This test is conducted to determine whether the same conclusion holds when gathering all the medians over years and then comparing whether the median of those medians differs.

Medians of both groups over years are not normally distributed (origin medians: W statistic = 0.7608, p-value = 0.0002; destination medians: W statistic = 0.8528, p-value = 0.0059). So, again, the Mann-Whitney U test is used. Results can be found in Chapter 6.

Second category of options
In this section, the one sample/signed (non-)parametric test over all years, for both groups, is conducted. That is, to conclude also per year on the differences between groups.

So, what also can be done as a generic statistical test, is taking one of the medians of the top 25 groups and test whether there is a difference with the values of the other group. Median is written here, because a one sample signed Wilcoxon rank test is used, due to not fulfilling conditions of parametric-tests. For example, the Shapiro-Wilk tests indicates in 2020 that the top 25 origin and destination locations are not normally distributed in terms of sea level rise risk (origin: W statistic = 0.8971, p-value = 0.0159; destination: W statistic = 0.9254, p-value = 0.0681). This can be done for every year. In this analysis, not two data columns are compared and conclusions about their difference in medians are drawn (because then there is the overlap assumption problem), but the median of one of the groups is treated as constant, and the values of the other group are analyzed to conclude if the median of those values differs from the median that was treated as constant. That is how the non-parametric one sample Wilcoxon signed rank test works (King & Eckersley, 2019).

Results can be found in Chapter 6.

D.3. Riverine floodings

D.3.1. Riverine floodings: Comparing origin-destination of refugees

D.3.1.1. Riverine floodings: Statistical generic difference tests

Due to overlap in top origin and destination locations it becomes difficult to do statistical tests to compare generic/overall differences between partly overlapping groups (top 25 origin and destination countries). However, there are two options that are still available (which themselves also consist of two options). All are not very elegant, so both categories (four tests) are conducted and combined to see if the same conclusions can be drawn.

First category of options
1) Testing differences of means
The means do not 'overlap' between groups. Therefore, the idea is to gather the mean values of groups for every years and compare those to test the differences between them. So, instead of original values, means are used. Then, a Mann-Whitney U test or an independent t-test can be conducted. To determine which test needs to be conducted, again, assumptions need to be tested. At first, normality is tested. A Shapiro-Wilk test is used to reflect on normality, because the sample size is smaller than 50. Otherwise, a Kolmogorov-Smirnov (KS) test should have been used (Mishra et al., 2019). (Origin means: W statistic = 0.8558, p-value = 0.0067; destination means: W statistic = 0.9251, p-value = 0.1240.) One can conclude that not all the data is normally distributed. The top 25 origin means are not normally distributed, when gathered over 20 years. Therefore, the Mann-Whitney U test will be conducted as alternative to the unpaired t-test. This means intuitively that the median differences of the means are compared. The definition of this is that, for every year, one takes the mean riverine flooding intensity score of all the countries in group origin and group destination, and gather this over 20 years from 2003 and 2022. Then one can compare whether the means of these mean-groups over year differ. One basically compares all years in one then, and threat the mean score as individual data point in the comparison. This is not specific, and you do not see this often, but one can identify with one statistic if the median of the means over time differs. One can conclude if the riverine flooding intensity mean scores over the years are statistically different between origin and destination groups.

The test is basically used to determine if there are statistically significant differences between two groups (MacFarland et al., 2016). It tests if two independent groups are from the same population. In other words, of the means can come from the same population values. “For practical purposes, the user may think of it informally as testing whether or not the two distributions have the same median” (Riffenburgh, 2012).

Results can be found in Chapter 6.

2) Testing differences of medians
The reasoning is the same as above. This test is conducted to determine whether the same conclusion holds when gathering all the medians over years and then comparing whether the median of those medians differs.

Also the median values of the two top 25 groups are both not normally distributed, when gathered over years (origin medians: W statistic = 0.8895, p-value = 0.0263; destination medians: W statistic = 0.8536, p-value = 0.0061.). So, again, the Mann-Whitney U test is used.

Results can be found in Chapter 6.

Second category of options
In this section, the one sample/signed (non-)parametric test over all years, for both groups, is conducted. That is, to conclude also per year on the differences between groups.

So, what also can be done as a generic test, and besides the individual comparison plots that are also shown in the section below, is taking one of the means of the top 25 groups and test whether there is a difference with the values of the other group. Means is written here, because a one sample t-test is used, due to the fact that conditions of parametric-tests are fulfilled, with the most significant one being the normality distribution. Those original values in both groups are normally distributed, based on the 2020 example. The null hypotheses cannot be rejected by the Shapiro-Wilk test (top 25 origin locations: W statistic = 0.9837, p-value = 0.9467; top 25 destination locations: KS statistic = 0.0939, p-value = 0.9654). This is remarkable, because many original values of other climatic indicators are not normally distributed. Therefore, the one sample t-test is here chosen instead of the one-signed Wilcoxon rank test (Ramachandran & Tsokos, 2021). In this analysis, not two data columns are compared and conclusions about their difference in means are drawn (because then there is the overlap assumption problem), but the mean of one of the groups is treated as given/constant, and the values of the other group are analyzed to conclude if the mean of those values differs from the mean that was treated as constant. That is how the parametric one sample t-test works.

Results can be found in Chapter 6.

D.4. Coastal floodings

D.4.1. Coastal floodings: Comparing origin-destination of refugees

D.4.1.1. Coastal floodings: Statistical generic difference tests

Due to the overlap in top origin and destination countries, it becomes difficult to do statistical tests to compare generic/overall differences. However, there are two options that are still available (which themselves also consist of two options). All are not the preferred tests, so both categories (four tests) are conducted and combined to see if the same conclusions can be drawn.

First category of options
1) Testing differences of means
The means do not ‘overlap’ between groups. Therefore, the idea is to gather the mean values of groups for every years and compare those to test the differences between them. So, instead of original values, means are used. Then, a Mann-Whitney U test or an independent t-test can be conducted. To determine which test needs to be conducted, again, assumptions need to be tested. At first, normality is tested. A Shapiro-Wilk test is used to reflect on normality, because the sample size is smaller than 50. Otherwise, a Kolmogorov-Smirnov (KS) test should have been used (Mishra et al., 2019). (Origin means: W statistic = 0.9511, p-value = 0.3846; destination means: W statistic = 0.8101, p-value =

0.0012.) One can conclude that not all the mean data is normally distributed. The top 25 destination means are not normally distributed, when gathered over 20 years.

Therefore, the Mann-Whitney U test will be conducted as alternative to the unpaired t-test. This means intuitively that the median differences of the means are compared. The definition of this is that, for every year, one takes the mean coastal flooding intensity score of all the countries in group origin and group destination, and gather this over 20 years from 2003 and 2022. Then one can compare whether the means of these mean-groups over year differ. One basically compares all years in one then, and threat the mean score as individual data point in the comparison. This is not specific, and you do not see this often, but one can identify with one statistic if the median of the means over time differs. One can conclude if the coastal flooding intensity mean scores over the years are statistically different between origin and destination groups. The test is basically used to determine if there are statistically significant differences between two groups (MacFarland et al., 2016). It tests if two independent groups are from the same population. In other words, of the means can come from the same population values. “For practical purposes, the user may think of it informally as testing whether or not the two distributions have the same median” (Riffenburgh, 2012).

Results can be found in Chapter 6.

2) Testing differences of medians
The reasoning is the same as above. This test is conducted to determine whether the same conclusion holds when gathering all the medians over years and then comparing whether the median of those medians differs.

Also the top 25 median values gathered over 20 years are both not normally distributed (origin medians: W statistic = 0.6583, p-value = 1.2606e-05; destination medians: W statistic = 0.7164, p-value = 6.2252e-05). So, again, the Mann Whitney U test is used.

Results can be found in Chapter 6.

Second category of options
In this section, the one sample/signed (non-)parametric test over all years, for both groups, is conducted. That is, to conclude also per year on the differences between groups.

So, what also can be done as a generic test, and besides the individual comparison plots that are also shown in this chapter later, is taking one of the medians of the top 25 groups and test whether there is a difference with the values of the other group. Median is written here, because a one sample signed Wilcoxon rank test is used, due to not fulfilling conditions of parametric-tests for the original data. For example, the top 25 data in 2020 is not normally distributed in terms of Coastal flooding values, as one can conclude from the Shapiro-Wilk tests (origin values: W statistic = 0.8732, p-value = 0.0050; destination values: W statistic = 0.7933, p-value = 0.0002). This can be done for every year. In this analysis, not two data columns are compared and conclusions about their difference in medians are drawn (because then there is the overlap assumption problem), but the median of one of the groups is treated as given/constant, and the values of the other group are analyzed to conclude if the median of those values differs from the median that was treated as constant. That is how the non-parametric one sample Wilcoxon signed rank test works (King & Eckersley, 2019).

Results can be found in Chapter 6.

D.5. Cyclones

D.5.1. Cyclones: Comparing origin-destination of refugees

D.5.1.1. Cyclones: Statistical generic difference tests

Due to the overlap between top origin and destination countries, it becomes difficult to do statistical tests to compare generic/overall differences. However, there are two options that are still available (which themselves also consist of two options). All are not the preferred tests, so both categories (four tests) are conducted and combined to see if the same conclusions can be drawn.

First category of options
1) Testing differences of means
The means do not 'overlap' between groups. Therefore, the idea is to gather the mean values of groups for every years and compare those to test the differences between them. So, instead of original values, means are used. Then, a Mann-Whitney U test or an independent t-test can be conducted to compare the groups of means. To determine which test needs to be conducted, again, assumptions need to be tested. At first, normality is tested. A Shapiro-Wilk test is used to reflect on normality, because the sample size is smaller than 50. Otherwise, a Kolmogorov-Smirnov (KS) test should have been used (Mishra et al., 2019). (Origin means: W statistic = 0.6011, p-value = 3.0260e-06; destination means: W statistic = 0.3244, p-value = 1.1577e-08.) One can conclude that the top 25 mean data over years is not normally distributed. This is concluded by the findings of the Shapiro-Wilk tests. The origin and destination means are not normally distributed, when gathered over years.

Therefore, the Mann-Whitney U test will be conducted as alternative to the unpaired t-test. This means intuitively that the median differences of the means are compared. The definition of this is that, for every year, one takes the mean cyclone intensity score of all the countries in group origin and group destination, and gather this over 20 years from 2003 and 2022. Then one can compare whether the means of these mean-groups over year differ. One basically compares all years in one then, and threat the mean score as individual data point in the comparison. This is not a specific/detailed analysis, and you do not see this often, but one can identify with one statistic if the median of the means over time differs. One can conclude if the cyclone intensity mean scores over the years are statistically different between top origin and destination groups. The test is basically used to determine if there are statistically significant differences between two groups (MacFarland et al., 2016). It tests if two independent groups are from the same population. In other words, of the means can come from the same population values. "For practical purposes, the user may think of it informally as testing whether or not the two distributions have the same median" (Riffenburgh, 2012).

Results can be found in Chapter 6.

2) Testing differences of medians
The reasoning is the same as above. This test is conducted to determine whether the same conclusion holds when gathering all the medians over years and then comparing whether the median of those medians differs.

Here one can see something interesting: all the medians, for origin as well as destination, for every year, always have the value 0.01 (see Figure D.1). Results and interpretations can be found in Chapter

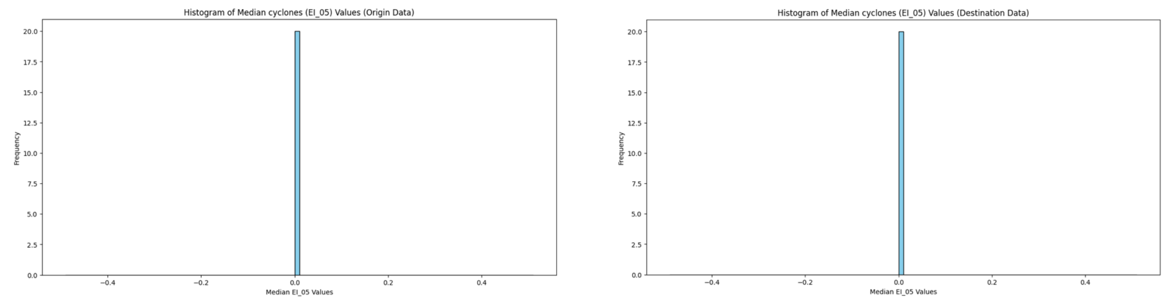


Figure D.1: No distribution of median values over years for cyclone intensity data

6.

Second category of options
In this section, the one sample/signed (non-)parametric test over all years, for both groups, is conducted. That is, to conclude also per year on the differences between groups.

So, what also can be done as a generic test, and besides the individual comparison plots that are also shown in this chapter later, is taking one of the medians of the top 25 groups and test whether there is a difference with the values of the other group. Median is written here, because a one sample signed

Wilcoxon rank test is used, due to not fulfilling conditions of parametric-tests for the original data. For example, the top 25 data in 2020 is not normally distributed in terms of Cyclone intensity values, as one can conclude from the Shapiro-Wilk tests (origin values: W statistic = 0.2343, p-value = 2.0476e-10; destination values: W statistic = 0.3363, p-value = 1.2750e-09). This can be done for every year. In this analysis, not two data columns are compared and conclusions about their difference in medians are drawn (because then there is the overlap assumption problem), but the median of one of the groups is treated as given/constant, and the values of the other group are analyzed to conclude if the median of those values differs from the median that was treated as constant. That is how the non-parametric one sample Wilcoxon signed rank test works (King & Eckersley, 2019).

Results can be found in Chapter 6.