FIELDS AND FOREIGN LANDS: PRE-INDUSTRIAL CLIMATE RISK AND INTERNATIONAL MIGRATION

Miriam Manchin* Alex Newnham† Elena Nikolova‡

January 14, 2026

Abstract

We study how pre-industrial climate risk during 1500–1800 shaped historical bilateral inward migration and present-day international migration stocks in Europe. Using high-resolution data, we find that a one-unit increase in historical precipitation variability reduces the share of today's migrants by 0.48 percentage points and lowers historical migration. The effects hold only in historically rural areas and are driven by climate variability during growing seasons, pointing to an agricultural channel. We show that persistence of historical climate risk in current migration patterns is closely related to differences in historical prosperity and migrant networks, highlighting links between climate, migration, and development.

JEL codes: F22, Q54
Keywords: International Migration, Climate Risk, Historical Migration, Long-Run Development

The authors would like to thank and acknowledge anonymous referees and assistance from the Alexander and Diviya Magaro Peer Pre-Review Program at Harvard's Institute for Quantitative Social Science. We also thank Cevat Aksoy, Jan David Bakker, Sascha Becker, Michelle Brock, Italo Colantone, Tomas Cvrcek, Ralph De Haas, Elodie Douarin, Sergei Guriev, Tim Hatton, Morgan Kelly, Ulrich Muller, Alexander Plekahnov, Gianmarco Ottaviano, Ekaterina Zhuravskaya and seminar participants at EBRD, Bocconi, UCL, the Baltic Connections Conference, the Social Science History Association Annual conference, IOS-Regensburg, GLO, the RES meeting, LSE, World Trade Institute, Universita Cattolica di Milano, and Zayed University for useful comments. We also thank Sultan Orazbayev for his indispensable initial work with the data. Mahreen Munir, Gabriel Basques Soares and Arthur Diniz De Souza provided excellent research assistance.

*Corresponding author. University College London, Politecnico di Milano, Italy; e-mail: m.manchin@ucl.ac.uk
†Université Paris-Saclay, France
‡Parami University, UCL, IOS Regensburg and Global Labor Organization; e-mail: elenanikolova@parami.edu.mm

1 Introduction

Climate change, defined as long-run changes in average climate conditions and their variability, poses one of the most serious challenges confronting humanity today.[1]. Climate change is particularly detrimental for the world's poor, who are more likely to experience climate shocks and less likely to recover from them quickly (Triyana et al., 2024). The literature has identified migration as a critical coping strategy for dealing with the severe impacts of climate change. According to projections by Burzyński et al. (2022), present-day climate change in the origin location will increase the number of migrants to 100- 200 million for the remainder of the 21st century. However, only around 57 million migrants will migrate internationally, leading to many people being trapped in extreme poverty and conflict arising from climate change.

Even so, despite an increasing number of studies, the environment-migration nexus and its underlying mechanisms remain poorly understood (Hoffmann et al., 2020). First, even though the literature agrees that the effect of climate on migration varies depending on the context and time frame (Cattaneo et al., 2019; Berlemann and Steinhardt, 2017; Millock, 2015), we know little about how and why historical factors affect current migration patterns. Such an investigation is even more relevant since some studies focusing on the contemporary link between climate and migration find a positive relationship, while others find no relationship or a negative one (Beine and Jeusette, 2021; Burzyński et al., 2022). In addition, because recent climate changes are anthropogenic while historical climate variations were natural, examining the relationship between historical climate and migration avoids the problem of reverse causality. Second, the existing literature has predominantly focused on (contemporary) climate events in the *origin* location and *out-migration* (Beine and Parsons, 2015; Cai et al., 2016; Mastrorillo et al., 2016; Coniglio and Pesce, 2015), while largely overlooking those in the *destination* location and *in-migration*, which are equally important. Third, much of the literature has focused solely on climate *averages*, thereby neglecting climate *variability*, an essential dimension of climate change.

This paper fills these gaps by investigating the long-run impact of pre-industrial climate risk, proxied by the standard deviations in temperature and precipitation, on both present-day and historical international migration patterns in Europe between 1500 and 1800. Our identification exploits fine-grained sub-national variation in pre-industrial climate risk within countries, conditional on rich geographic controls and regional fixed effects. We use highly disaggregated locality-level data (with cells of approximately 28 square kilometers) in eight European countries (France, Germany, Ireland, Italy, Netherlands, Portugal, Spain and the UK)[2]. While we focus on the role of historical climate risk on in-migration in the *destination* location, which has been understudied, our genealogical specification (described in more detail below) also accounts for origin climate risk in two different ways.[3]

[1]The IPCC definition of climate change is "change in the state of the climate that can be identified (e.g., by using statistical tests) by changes in the mean and/ or the variability of its properties and that persists for an extended period, typically decades or longer."(Intergovernmental Panel on Climate Change, nd)

[2]The equivalent is about 0.25 degree grids for our contemporary migration data and 0.5 degrees for the historical migration data, which we describe in more detail below. While the contemporary migration data only covers the eight countries on which we focus, the historical migration data has a wider geographical coverage with a larger area in Europe as shown in Figure A.3 in the Annex.

[3]The genealogical specification includes origin cell-time fixed effects, and we also present robustness checks using origin-cell climate as an explanatory variable, with our main results on the importance of destination climate risk remaining the same.

Our conceptual framework and analyses emphasize the link between historical climate risk, historical agriculture, historical migration, and present-day migration. International migration in the pre-industrial era in Europe largely consisted of individuals seeking occasional employment, with a significant portion being agricultural migrants (Klein and Van Lottum, 2020). Seasonal agricultural migration played a central role in crop production throughout European history (Moch, 2003; Oris, 2003). In pre-industrial times, a significant number of seasonal agricultural migrants were willing to relocate frequently, sometimes to further away regions and even across countries (Hochstadt, 1983, 1999; Bade, 2002), with permanent settlement often emerging from repeated temporary mobility. For example, a considerable portion of the population in a German region neighboring Dutch areas would spend each summer in Holland engaging in agricultural labor (Hochstadt, 1983). According to Bade (2002), long-distance seasonal workers often followed established migratory routes that had been developed and refined over generations. Sometimes, these temporary agricultural migrants also settled permanently in their destination locations (Moch, 2003), often lured by economic or legal incentives. For example, in Scotland, seasonal workers were given a cottage with garden, in exchange for one woman from the family working for free during the harvest (Devine, 1984).

Economic historians concur that before the industrial era, fluctuations in climate represented a substantial economic risk in Europe because climate risk significantly affected crop yields and food availability (Campbell, 2010; van Bath, 1963). This implies that climate risk, through its impact on crop yields, influenced the demand for and supply of international migrant labor, much of which was seasonal and followed recurring routes, and which often resulted in permanent settlement in destination locations(Galloway, 1986; Deschacht and Winter, 2015a; Klein and Van Lottum, 2020). During poor harvests, demand for migrant labor was low as there was no need for additional workers, and the low yields could not feed both the local and migrant populations. Conversely, during good harvests, the demand for migrant agricultural workers increased. Moreover, climate risk influenced the attractiveness of a particular destination for would-be migrants: high climate variability meant more uncertainty, and thus potentially a lower supply of migrants. Climate variability, therefore, likely discouraged migration due to the potentially uncertain (or reduced) welfare in the destination location, both for short-term labor movements and for longer-term settlement decisions.

Since these supply and demand forces can work in opposite directions, identifying the overall direction of the effect of past climate variation on international migration is ultimately an empirical question. To answer it, we utilize two complementary, and high-resolution migration datasets. Each dataset accomplishes a specific and necessary goal in testing our theoretical framework. The genealogical migration time-series dataset allows us to identify the relationship between historical climate risk and historical international migration flows, while the contemporary cross-sectional migration dataset - which covers only the eight countries on which we focus - allows us to study the relationship between historical climate risk and present-day international migration stocks and to identify suggestive mechanisms. We focus on international migration for two reasons. First, as we discuss in the conceptual framework, seasonal agricultural migration across borders was particularly important in pre-industrial Europe. Second, the point of overlap between the contemporary migration dataset (which includes only international migrants) and the genealogical migration dataset (which includes domestic and international migrants) is international migration. However, our results are not sensitive to excluding domestic migrants. In the empirical section, we run specifications with the genealogical data

including both domestic and international migrants, and our results remain unchanged.

The first migration dataset captures historical genealogy-based bilateral migration flows and covers the period 1500-1800. More precisely, we use the FamiLinx data, a publicly available dataset derived from the genealogical website Geni.com, which has extensive individual-level birth and death location information. An important benefit of the Familinx dataset is that it provides unparalleled and uninterrupted temporal and geographical coverage, compared to alternative data, such as parish records (Blanc, 2024b). While this dataset is non-representative and may be subject to potential biases, we conduct rigorous tests to show that it can be used for reliable inferences on the link between historical climate risk and migration. Specifically, we show that migration patterns in the genealogical data are not systematically related to local population size or development, that results are robust to restricting the sample to regions with high coverage, and that estimates are stable across time periods with differing data quality.

In order to calculate migration flows, we assign individuals to specific cells based on their birth and death locations. With this information, we identify migrant individuals as those who were born in a particular cell (in a particular country) but who then passed away in a different cell (in another country). Aggregating migrants over 100-year intervals provides a measure of bilateral migration between two cells over time.[4] The dataset has a time dimension, and allows us to control (using fixed effects) for all origin cell-time varying factors, all bilateral cell-specific factors, and all destination cell-specific factors in each cell, in addition to our main variable of interest, climate risk. We also control for destination cell-level population density - an important pull factor - calculated from this data, and for mean temperature and precipitation.

The second dataset provides information on current international migration patterns (i.e, stocks) in the destination locations at the level of localities (0.25 degree grids) based on high-quality census data. While the data is extremely detailed, it is available only for a single year, 2011. Using this data, we regress the share of migrants in a particular locality in 2011 on pre-industrial climate variation, measured by the standard deviation in precipitation and temperature over the period 1500-1800 (focusing on the growing season of major crops). We control for other location characteristics, including additional climate factors, such as mean precipitation and temperature over the same period, while geographical controls include land suitability for agricultural activities, coastal region proximity, river proximity, altitude, area size, and latitude and longitude. To proxy for economic activity, we include population density in a given cell, and account for whether the cell was a city or rural area in pre-industrial times. To make sure that our results are not confounded by region-specific factors (such as economic conditions, or cultural and historical factors, including past colonial ties), we also include region (NUTS-1) fixed effects.[5]

Our historical and current migration analyses reveal a negative relationship between pre-industrial climate risk and inward migration. Using our genealogical dataset, we find that there was less bilateral inward migration to locations with higher climate variability in pre-industrial times. Next, using the contemporary migration dataset, we find that a one-unit (1 cm) increase in the standard-deviation of historical precipitation decreases the share of migrants in a given cell by 0.48 percentage points, which is equivalent to 6.6%

[4]Results are robust to using periods of different duration (50 or 70 years).

[5]Even though this specification lacks information on climate variation in the sending location, our specifications with historical migration allow us to control for origin conditions via origin-destination fixed effects, with our results remaining consistent across the two specifications.

relative to the mean of the dependent variable. The observed effects hold in historically rural locations and during periods corresponding to the growing season of major crops. In addition, in locations with higher potato suitability (a crop that is more resilient to climate risk than other crops), the impact of climate risk is mitigated. This evidence points to agriculture as an important context through which climate risk affected migration decisions. Our analysis withstands a number of robustness checks, including accounting for spatial correlation.

We explore several potential mechanisms through which historical climate risk could have had a persistent effect on today's migration patterns. Given historical data limitations, these results should be treated as exploratory, since it is challenging to pin down the precise channels and mechanisms in the historical context we examine. First, today's migrants may prefer locations where their ancestors or their ancestors' peers were located due to network effects. Historical evidence shows that seasonal migrants occasionally settled down (Collins, 1976), and that peer effects played a significant role in driving historical migration in Europe (Klein and Van Lottum, 2020). Second, climate risk may lead to conflict,[6] which in turn could have deterred migrants from conflict-prone destination locations. On the other hand, more volatile climate conditions may have fostered a culture of cooperation and potentially more accepting attitudes towards foreigners (Buggle and Durante, 2021)[7], both in the short and long term[8], thus attracting more migrants to such locations. Conversely, more close-knit societies could have also developed hostility toward migrants. In the same vein, high climate risk could have led to higher levels of social capital, and the development of migrant-friendly institutions and policies. Past climate risk might have also reduced both past and present economic development and growth (Kotz et al., 2022), and worse development outcomes in turn might have reduced in-migration. We explore the importance of these mechanisms in our conceptual framework and in the empirical analysis. We find patterns consistent with persistence related to past migration networks and historical economic conditions (proxied by GDP and population density). The latter are captured using two closely related historical proxies: GDP and population density. Higher pre-industrial climate risk was associated with lower historical income levels and lower historical migration, which in turn are correlated with lower current migration.

Our work builds on and contributes to an important literature arguing that historical environmental and agricultural factors can have long-run consequences for economic growth and development (Alesina et al., 2013; Easterly, 2007; Galor and Özak, 2016; Hornbeck, 2012; Nunn and Qian, 2011). Our paper makes a number of important contributions to the literature. Our recently assembled and highly detailed datasets allow us to pinpoint the relationship between historical climate variation and both contemporary and past inward migration, which has not been studied before. While much of the previous work focuses on temperature and precipitation levels alone (Dell et al., 2014, 2012), we extend the analysis by considering both the levels and standard deviations of temperature and precipitation, showing that it is variation, rather than levels, that ultimately matter for migration. By adopting this approach, inspired by Kahn et al. (2021), we explicitly account for changes in the distribution of weather patterns, encompassing not only aver-

[6]See Burke et al. (2015) for a review of relevant empirical evidence.

[7]Similarly, Giuliano and Nunn (2017) show that societies where temperature is more variable today value traditions less and are more open towards change.

[8]Evidence shows that cultural attitudes tend to persist over the long-run, see for example Voigtländer and Voth (2012)

age values but also their variability. Using two complementary and very detailed spatial datasets enables us to move beyond country-level determinants and exploit very fine-grained spatial heterogeneity within countries, along with being able to include regional fixed effects, controlling for all region-specific factors. Our paper further contributes to a small but important literature on the link between climate and migration, which, however, has largely ignored the long-run effect of climate and the European context due to the lack of data. A further innovation of this work is that we study climate risk as a driver of *inward* migration, as opposed to *outmigration*, the latter being the focus of most existing studies.[9] Since outward and inward migration may have different drivers and consequences, mapping the link between climate variation and inward migration is essential for fully understanding how climate affects migration. [10]

On the policy front, our paper highlights the importance of historical climate risk for in-migration. It also offers a preview of how ongoing climate change and climate risk are likely to affect prosperity and migration in the future, emphasizing that these impacts are long-term. The findings also underscore the heightened vulnerability of developing economies that rely heavily on agriculture and do not have appropriate coping strategies for attracting migrants, such as increased fiscal spending. Developing crops that are resilient to climate risk, such as the potato in the historical setting, may be a potential solution for such lower-income countries.

In the next section, we outline a conceptual framework linking climate variation to past and current migration patterns. This is followed by a description of the data we use. Next, we discuss our empirical specification and the results, while our final section offers concluding remarks.

2 Conceptual framework: migration and climate variation

In this section, we outline how historical climate risk, measured by climate variability, could have impacted inward migration patterns in the past and how this relationship could have persisted. We also discuss how this effect is likely to have been related to agricultural production.

2.1 Climate variation in pre-industrial times

In Europe, industrialization and urbanization were not fully felt until the mid-1800s, when the industrial revolution had a real effect.[11] For centuries before, the vast majority of the continent remained predominantly rural, with agriculture serving as the primary means of subsistence for the population (Ladurie, 1971). As a result, climate variation constituted an important economic risk, since it was linked to crop yields and the availability of food (Parker, 2001; Davis, 1973). Most agricultural activity in Europe involved

[9] For examples of studies focusing on outmigration, see Cai et al. (2016), Boustan et al. (2012) and Beine and Parsons (2015)

[10] While we show that past climate risk is an important determinant of today's migration patterns in Europe, as we discuss in the next section, climate risk is, of course, not the only determinant of migration, and the findings in this paper should not be interpreted as supporting environmental determinism (Pei et al., 2022).

[11] See, for example, Bairoch et al. (1988).

crops whose growing season is generally during spring and summer,[12] thus weather variation, particularly in these periods, had important consequences.[13] For example, Galloway (1986) notes that a two-week period of intense rain during autumn could destroy an entire harvest, and shows that there is a significant correlation between seasonal weather fluctuations and yearly yields in the pre-industrial period.

2.2 Climate variation, labor demand and labor supply

Seasonal migrants played an indispensable role in agricultural production throughout European history. An example from 1607 illustrates its significance: Baron de Sancy asked how the Spanish could work the land without the French, if the French did not come to help and settle. Who would harvest if more troops of French did not come? And who would serve the Spanish in the fields and cities if these same French did not? (as cited in Moch (2003)). Oris (2003) discusses how different locations can be thought of as a system where there is a mutual symbiosis between the native population and temporary migrants. Some locations consistently required additional migrant workers, while other origin locations regularly supplied migrants. Similarly, Bade (2002) observes that there were geographical areas which depended on an additional workforce periodically to reduce risks, and that long-distance seasonal migrants regularly migrated along the same routes.[14] Hochstadt (1983) writes that *'Each spring the roads of Europe came alive as peasants streamed out of their villages and fell in with the bands from other villages to seek work in distant places, and the roads teemed again in late autumn when the workers began their homeward trek. From Spain to Russia migratory peasants linked disparate economic systems.'* Lucassen and Lucassen (2009) estimate that in the second part of the seventeenth century about 8% of Europeans (excluding those in Russia) could be considered an international migrant. Notably, Van Lottum (2007) indicates that around 1 in every 10 individuals born in Scotland around 1650 was living abroad. Hahn (2000) affirms that even before industrialization, seasonal labor migration across national borders held significance. An important factor contributing to the popularity of migration in Europe was the relatively high proportion of unmarried men and women (estimated at 10-20%), who were free to move compared to regions such as Asia, where marriage was nearly universal (Moch, 2003).[15]

Seasonal agricultural workers offered a practical solution to meet labor demand by selectively employing workers during periods of high agricultural output. This approach

[12]These crops include barley, wheat, rapeseed/canola, sugar beets, potatoes and oats (Northern Europe); wheat, barley, maize, rapeseed/canola, sugar beets and grapes (Western Europe); and wheat, barley, maize, sunflower, pulses, potatoes and olives (Southern Europe) (Leff et al., 2004).

[13]For instance, according to Davis (1973) referring to pre-industrial times, *"The changing climate, apparent in the slow worsening and the slow improvement, over many decades, of the 'average' weather conditions around which each year's weather fluctuated, was a factor of the utmost economic importance."*

[14]In the same vein, Grantham (1993) emphasizes the important dependence in pre-modern agricultural economies on temporary/seasonal workers. Similarly, Hochstadt (1999) demonstrates that migration frequency in pre-industrial Germany was comparable to contemporary levels.

[15]Lucassen and Lucassen (2009) argue that both permanent and temporary migration held immense importance in early modern Europe, with a notable demand for large numbers of seasonal migrants over long distances, driven by the prospect of higher wages during harvest periods. Klein and Van Lottum (2020) also note that long-distance (several hundreds of kilometers) migration was not rare in Europe. Moch (2003) discusses the broad movement of workers who traveled to the Netherlands during this era, originating from an arc spanning from Calais in France to the German states along the North Sea.

might have reduced the need for sustaining a larger population during times of diminished yields, thereby potentially lessening the economic and social consequences of lower yields. Deschacht and Winter (2015b) look at the 1840 crisis in Flanders triggered by poor yields, leading to famine and disease. The study highlights a decrease in the number of seasonal migrants during this period, while local residents did not engage in out-migration, despite having the freedom to relocate within the country or abroad.[16] Similarly, Dribe (2003) finds that in pre-industrial Sweden out-migration was not a prevalent response to harvest failures and resulting economic stress, even among landless individuals. These findings suggest that, in the case of weather shocks, the adjustment process might have involved reductions in seasonal migrants rather than widespread outmigration.

Historical climate risk affected not only the demand for workers, but also their supply. Traditional migration models typically involve a one-time decision for individuals to migrate based on evaluating the net return of moving to a new location, accounting for costs, compared to the return of remaining in their home location.[17] One could modify such a framework to incorporate not only migrants' wages in the sending and receiving locations, but also climate risk as an important measure of wage uncertainty over time. When climate risk is high, potential migrants' earnings in the new location become uncertain. Conversely, with low climate risk, wage uncertainty decreases, thereby increasing the expected utility of migrating.[18] The type of migration that prevailed in Europe during pre-industrial times often involved temporary agricultural workers moving periodically along the same route, with permanent settlement frequently emerging from repeated mobility.[19] Importantly, this seasonal migration was heavily influenced by fluctuations in agricultural yields.[20] A Roy model incorporating climate variation could be extended to take into account the characteristics of seasonal, repeated migration decisions, whereby as people learn more about the destination and the cost of migration, they keep adjusting their expectations about the utility of migration.[21] Similarly, origin climate could also be incorporated in such a model. While we expect the origin climate to potentially influence migration, as has been shown in the literature, this is not the focus of this paper. Therefore, we control for origin climate and climate risk with time-varying origin cell fixed effects in the empirical analysis.

In sum, the discussion in this sub-section illustrates that the *net* effect of *past* climate

[16] See also Jacquemyns (1928) and Vandenbroeke (1979) on locals not responding by moving out during the crisis. Mueller et al. (2020) also find that in present-day Africa, climate variability does not significantly impact rural out-migration.

[17] See, for example, Roy (1951), Borjas (1987), and Grogger and Hanson (2011).

[18] One example where climate is integrated into this framework as a factor influencing the individual decision to migrate is Beine and Parsons (2015).

[19] For example, Klein and Van Lottum (2020) note that international migration in the pre-industrial era largely consisted of individuals seeking occasional employment, with a significant portion being agricultural migrants.

[20] As Auffhammer and Schlenker (2014) point out, weather variability is extremely important for agricultural output, since weather is a direct input in the production function. Similarly, Colmer (2021) shows that in present-day India, weather-induced agricultural shocks (due to an increase in temperature) are associated with a reduction in agricultural production and, in turn, the employment and wages of agricultural workers. While agricultural shocks do not induce migration of the local population, the paper highlights an alternative coping mechanism that was not available in the historical context which we study: an offsetting movement of workers in the manufacturing and services sectors.

[21] See for example Stark and Fan (2007) who develop a theoretical framework explaining repeated seasonal migration.When there is a substantial variation in yields, factors such as available work, wages, and food will also fluctuate accordingly, thus influencing the attractiveness of the location for seasonal workers. In locations where migrants repeatedly observe greater uncertainty due to fluctuating climate, inward migration and settlement will become less desirable.

variation on *past* migration will depend on demand and supply forces. While demand for migrant labor fluctuated due to climate risk, long-term migration supply and settlement likely gravitated towards locations with less climate risk. Since these supply and demand forces may go in opposite directions, identifying the overall direction of the effect of past climate variation on migration is ultimately an empirical question, which we answer in the following sections.

2.3 Explaining today's migration

In this section, we discuss why historical climate risk could have had a persistent impact on today's migration. Multiple channels could underlie the observed patterns, and the existing literature offers some clues on what these may be. Note that not all of the proposed mechanisms might be relevant, and we will provide empirical tests in the next section for each of these potential channels.

Past migration patterns (either in the form of permanent settled migrants or seasonal migrants regularly returning to the same location) could have had a permanent impact on current migration patterns through network effects, whereby today's migrants are attracted to locations to which their ancestors or their ancestors' peers went, thus reinforcing the direct impact of climate variation. Occasionally, pre-industrial seasonal migrants also settled down in the destination location. As documented by several historians (see for example Bade (2002), Collins (1976)), these migrants often returned to the same locations, possibly due to social ties.[22] According to Collins (1976) *"An integral feature of many village economies in central England was the travelling "bands" of workmen who created their own special circuits and trod year after year the same elliptic path"*, noting that some of these seasonal workers, often Irish, remained in England. There are also examples of legal or economic incentives to retain permanently seasonal workers (see Grantham (1993)). For example, in Scotland seasonal workers were given a cottage with garden, in exchange for one woman from the family working for free during the harvest (Devine, 1984). Klein and Van Lottum (2020) look at migration flows at the period 1700-1800 using maritime data and find that past migration is among the most robust and quantitatively important driver of cross-country historical migration.[23]

A second possible channel involves past *conflicts*. The literature finds that regions experiencing higher climate variation were more prone to conflicts (see Hsiang et al. (2013) and Burke et al. (2015) for a review of this literature). In turn, this could have led to lower inward migration to these locations, thus potentially shaping the migration patterns that we observe today.

Moreover, historical climate risk could have increased the propensity to *trade*, as trade could have been a coping mechanism for hard times. In turn, increased trade could have led to increased in-migration due to closer economic ties and cooperation. Buggle and Durante (2021) review a large literature in history and anthropology which indeed confirms the importance of trade for overcoming environmental risks in the European and non-European context (see, for instance, Bateman (2011) and Nettle (1998)). In addition,

[22]Network effects could be driven by assimilation, information transmission, or mutual support. However, we do not have appropriate historical data to distinguish among these possible drivers.

[23]For contemporary migration, Beine et al. (2011) find that social networks explain about 71% of the variation of the observed variability in migration flows (see also Manchin and Orazbayev (2018) or Manchin (ming)). Similarly, Mahajan and Yang (2020) find that hurricanes abroad increase migration to the US, with the effect higher for origin countries with prior migrant stocks in the US, with a key role played by existing immigrant networks in sponsoring relatives for legal migration.

Buggle and Durante (2021) find that climate risk does lead to market integration in the prices of wheat following climate shocks.

Furthermore, climate risk could have had a persistent effect on current migration through its negative impact on historical *income* levels. With limited or no coping mechanisms (such as propensity to trade), climate variation could have reduced economic growth, thus discouraging inward migration both throughout history and today. As Braudel (1973) emphasizes, after humans had transitioned to agriculture 'the rhythm, quality, and deficiency of harvests ordered all material life'. There is also emerging quantitative evidence related to the link between climate risk and income. Damania (2020) reviews the recent literature and concludes that variations in rainfall and water availability have significant negative impacts on sectors such as agriculture, human capital and conflict. However, the impact of climate risk on overall economic activity remains uncertain, possibly due to issues related to spatial correlation and measurement. In fact, Damania et al. (2020) use annual subnational gross domestic product data to show a concave relationship between precipitation and local gross domestic product growth between 1990 and 2014, while the impact vanishes using larger spatial scales. Likewise, Kotz et al. (2022) use a global panel of subnational economic output for 1,554 regions worldwide over the past 40 years to show that monthly standardized deviations of rainfall have a concave relationship with economic growth, while economic growth rates are reduced by increases in the number of wet days and in extreme daily rainfall.

Social norms are a further potential channel.[24] For instance, Buggle and Durante (2021) demonstrate that pre-industrial experiences of cooperation among subsistence farmers in response to climate risk contributed to the development of norms of generalized trust. These social norms persisted over time, even as the economic significance of climate diminished. In the same vein, Litina (2016) shows that natural land productivity in the past, and its effect on the desirable level of cooperation in the agricultural sector, had a persistent effect on the evolution of social capital, the process of industrialization and comparative economic development across the globe. Similarly, Ager and Ciccone (2018) demonstrate that in the nineteenth-century US, counties with higher agricultural risk, proxied by rainfall risk, had a higher share of religious communities, as a way to insure against such risk. It is therefore possible that regions with lower climate risk may have also exhibited pro-migrant attitudes alongside norms of trust, creating a welcoming environment for migrants today. Bursztyn et al. (2023) further show that the long-term presence of a given foreign ancestry in a US county leads to more generous behavior specifically toward that group's ancestral country. This insight is highly relevant in the context we study, since, as discussed above, in-migrants (either temporary or permanent) came regularly from the same location, which likely made it easier for pro-migrant attitudes to take shape.

Finally, another possible mechanism could have included the adoption of inclusive economic and political *institutions* in areas with historically high climate risk, with these (formal) institutions persisting until today. For example, Grantham (1993) discusses the importance of evolving institutions in pre-industrial times in order to ease the hiring of seasonal agricultural workers. Moreover, the literature highlights a strong link between norms of cooperation and trust, and the quality of institutions (see for example Tabellini (2010), Tabellini (2008), Guiso et al. (2016), Greif and Tabellini (2010), Nannicini et al.

[24]It could be that in-migration also brings a dynamic of cultural convergence between the origin and the destination countries, which itself fuels further migration. Unfortunately, our data does not permit us to examine this mechanism.

(2013)). Such inclusive institutions may have interacted with the pro-migrant norms to once again encourage contemporary migration.

We explore the importance of the mechanisms discussed in this subsection in subsection 4.4, and find suggestive evidence that historical income levels (captured by historical GDP per capita and population density) are most likely to explain the persistent effects we find.

3 Data

Our analysis exploits several high-resolution datasets, which allow us to undertake the empirical analysis at a disaggregated level ranging between 0.25 degree and 0.5 degree grids.

Historical genealogical migration

Data on bilateral historical migration come from the genealogy-based dataset FamiLinx, obtained from a comprehensive compilation of birth and death records of individuals from the genealogical website geni.com by Kaplanis et al. (2018). Users of the genealogical website geni.com search for their ancestors in the parish records and, for each generation, indicate precise information regarding birth, marriage, and death. Blanc (2024b) provides further details of the data, along with examples of family trees from the data. The full dataset includes 86 million individuals, out of whom there are approximately 16 million individuals for whom the location of birth and death can be identified. Given our time and geographical restrictions, we end up with close to 600,000 individuals in the sample.

We restrict the dataset to individuals born and dead in locations with available historical climate data. To align the migration data with the 0.5 degree resolution of the climate data (described below), we assign individuals to specific cells based on their birth and death locations.[25] We identify international migrants over one-hundred-year interval, an approximation of a human's life span. We also conducted our analyses using 50- and 70-year periods, and results remained very similar. In particular, an individual is considered to be an international migrant if she/he was born in a 0.5 degree cell in a given country but passed away in a different country (and cell). We calculate the number of international migrants based on this definition over one hundred-year periods, taking into account the date when the individual passed away to include the individual as a migrant in the hundred-year period.[26] We use records where both the birth and death of the individual took place between 1500 and 1800 in the European region. We provide further details on the geographic coverage and further information of the historical genealogical data set in Appendix A.1.

We conduct several important validation checks with the data. First, we check for differential attrition related to recorded population and differential levels of development using highly disaggregated data from Paprotny and Mengel (2023) on European population and GDP in 1870 (which is the earliest year available from Paprotny and Mengel

[25]With this genealogical dataset, we choose 0.5 degree cells. Unlike in the case of the contemporary migration specification (discussed in the next subsection), here we do not have explanatory variables at 0.25 degree disaggregation due to the use of fixed effects.

[26]We use one-hundred-year interval, as we do not have data regarding the date of migration, only the date and location of individuals' birth and death. Thus, we define migrants as individuals who die in a different country and cell compared to where they were born over a one-hundred-year period.

(2023)). Second, since the quality of observations is more likely to be worse in earlier years (from 1500 to 1650/1700, see Blanc (2024b)), we drop this period from our analysis. Third, we run regressions using only the regions with the best coverage of the data (located in the present-day Benelux region), and in additional specifications exclude those regions from the sample. These checks do not reveal attrition issues either over time or geographically. We provide further details on these validation checks in Appendix A.1.

The genealogical data identify individuals who relocated internationally between birth and death, and therefore capture long-run migration outcomes rather than temporary or circular movements. We do not observe intermediate (including seasonal) moves directly. In the pre-industrial European context, however, repeated temporary mobility and seasonal labor migration often created persistent corridors and information networks and, in many cases, culminated in permanent settlement. Our estimates should therefore be interpreted as capturing how destination climate risk shaped the geography of long-run settlement outcomes arising from broader mobility patterns.

It is also reassuring that the existing scholarship confirms that the dataset we use is robust. Crowd-sourced online genealogical data sources, such as the Familinx data, have recently been used by economics and demography scholars to understand population demographics before modern censuses became available. An important benefit of these datasets is that they provide significantly larger geographical coverage, and continuous long time periods compared to alternative data, such as parish records (Blanc, 2024b). While acknowledging the inherent biases in crowd-sourced data, particularly the over-representation of individuals with higher socioeconomic status in earlier periods and locations with low coverage (Stelter and Alburez-Gutierrez., 2022), recent validation studies provide strong support for the utility of Familinx data for historical demographic analysis. Blanc (2024b) conducted a comprehensive validation exercise across 30 European countries, including those examined in this study, comparing Familinx data to representative sources on urbanization, fertility, and longevity. Their findings demonstrate that, from the seventeenth century to the early twentieth century, selection bias in the Familinx data is limited, and the millions of individuals within these genealogies effectively represent the broader population trends in their respective countries. This conclusion is further supported by Blanc (2024a), who focuses specifically on France and find similar results for fertility (from 1675) and mortality (from 1725). While it is crucial to acknowledge potential limitations, such as the potential underrepresentation of women, married couples, and mortality rates observed in some studies (Cozzani et al., 2023; Corti et al., 2023; Colasurdo and Omenti, 2024), the robust validation evidence suggests that Familinx data can provide valuable insights into the demographic dynamics of pre-modern Europe.[27]

Contemporary migration

For current migration patterns, we use a dataset from the European Commission (EC). Constructed from high-quality census data, the dataset contains information on the country of origin of the population at 100m by 100m resolution, with accompanying latitude and longitude coordinates for eight European countries (France, Germany, Ireland, Italy,

[27]It is instructive to also examine validation of other papers using different genealogical data which suffer from similar issues related to unrepresentativeness. For instance, Clark (2023) uses genealogical data from England to study social status between 1600 and 2022 and acknowledges issues of data selectivity. The author then validates the genealogical data using representative data starting in 1800 (two hundred years after the year 1600, the start of the analysis).

Netherlands, Portugal, Spain and the UK) for the year 2011.[28] We aggregate these data up to a 0.25 degree grid which corresponds to the level of geographical and other controls used in the analysis. The uniqueness of the dataset stems not only from the high level of spatial resolution, but also from the extensive geographical coverage that includes almost 45,000 localities. Using this dataset, we calculate the share of migrants and population density for each location.

Historical climate risk

Our main independent variables of interest capturing climate risk over the last five centuries were obtained from the European Seasonal Temperature and Precipitation Reconstruction (ESTPR) and contain seasonal (spring/summer /autumn/winter) precipitation and temperature for the period between 1500 and 2000 (Pauling et al., 2006; Luterbacher et al., 2004). The grids in the data have a width of 0.5 degrees (approximately 56 kms near the equator). The precipitation and temperature data are not based on actual weather station records, but instead have been reconstructed using a sophisticated process through a complex set of indirect proxies (such as ice cores, tree rings, lake/ocean sediments, documented evidence etc.).

Using the seasonal weather data, we constructed measures capturing the variation in precipitation and temperature. More specifically, focusing on the growing seasons (spring and summer), we calculate the standard deviation in weather conditions (temperature and precipitation) across all years for each cell. Our contemporary specification then uses these two standard deviations at the cell level over the entire time period 1500 to 1800. In the genealogical specification instead these standard deviations were calculated over hundred year periods.[29] Additionally, we conducted robustness checks by examining variability across different time periods, and using absolute deviations instead of the standard deviation.

We present the cell-level changes in precipitation variability (showing differences between the 18th century and 17th century and between the 17th century and 16th century) in Appendix Figure A.1, and the cell-level changes in temperature variability (showing differences between the 18th century and 17th century and the 17th century and 16th century) in Appendix Figure A.2. The figures show that there was substantial variation across time for our two climate risk variables.

While our main variables of interest are the two climate variation variables, we also control for average climatic conditions by including the average level of temperature (in Celsius) and rainfall (in centimeter) over our sample period. These average climatic conditions could have had an independent impact on economic development over time, as well as on agricultural activities and methods. In the spirit of Kahn et al. (2021), such a specification allows us to explicitly model changes in the distribution of weather patterns, including not only averages of climate variables but also their variability.[30]

A potential concern could be the reliability of the climate data, which is not derived

[28] Alessandrini et al. (2017) describe the construction of the dataset.

[29] Measurement error in climate risk might be stronger when using shorter time periods, hence we use one-hundred year periods., which are longer periods, but still allowing us to use the required fixed effects. Concerns regarding measurement error in these data should be less salient after 1700, when data quality improves significantly. Limiting our genealogical analysis to the period 1700-1850 produces very similar results (Table A.4, column 5).

[30] As Kahn et al. (2021) explain, while the level of temperature in Canada is low, the country is warming up twice as quickly as the rest of the world, implying that the effect of climate risk in this case on outcomes such as infrastructure, human health, and ecosystems will be higher than for a country where climate variation is low.

from actual weather stations. Unfortunately, data based on weather stations is only available from 1901. However, Buggle and Durante (2021) show that there is very high correlation for climate variability (precipitation and temperature) constructed from the (Pauling et al., 2006; Luterbacher et al., 2004) data and from weather station-based data over the overlap period (1900-2000). The dataset has been used extensively by many other scholars (see, for example, (Waldinger, 2022; Buggle and Durante, 2021; Giuliano and Nunn, 2021)).

Geographical characteristics

Geographical characteristics can influence climate risk, agricultural activities and migration patterns. To isolate the specific effect of climate risk on migration, we account for these confounding factors by incorporating relevant controls that have been identified in the literature as important for agricultural activities and also potentially for migration patterns from Henderson et al. (2017). We include elevation (in meters), latitude, longitude, and variables to capture access to water transport (which has importance for trade, economic and agricultural activity, and migration in a region). In particular, we have data on the distance to coast (in kilometers), being located on a river, and the cell being a coastal area. Furthermore, to account for the current level of economic activity, we include controls for population density (calculated from our respective migration datasets). As a robustness check, we also utilize night light intensity data (also obtained from Henderson et al. (2017)). Lastly, we control for land suitability for agriculture (from the same source), which plays a vital role in agricultural activities and potentially influences migration patterns as well. We also obtain data on potato cultivation suitability from Malpede et al. (2023)[31], with the index variable ranging from zero to one (higher values indicating more suitability).

Trust, institutions, conflicts, historical economic activity

In order to uncover the potential mechanisms behind our results, we add additional variables from four sources. The first is a variable which controls for locations that experienced different types of conflicts in the past, from Dincecco and Onorato (2018). The conflict dummy variable takes the value of 1 if at least one conflict took place in the cell during the period from 1000 to 1800, regardless of the conflict's type or duration.

Second, we use historical GDP per capita, for the year 1800 obtained from Buggle and Durante (2021) to measure past, pre-industrial level of development. In addition, as an alternative measure, we use population density in 1800, calculated from our genealogical dataset.

Furthermore, we introduce a dummy variable to control for the location being rural or a city in pre-industrial times. This variable takes a value of 1 if the location was historically situated in a city with a population larger than 10,000 inhabitants. The dataset used for this variable is sourced from Bosker et al. (2008).

Moreover, we incorporate two additional variables derived from the European Social Surveys conducted in 2012. These variables are available at the regional level, specifically at NUTS 1, 2, or 3 depending on the country. The first variable captures the respondent's attitudes towards migrants, while the second variable measures trust in parliament.[32] Both variables provide insights into societal attitudes and beliefs related to migration

[31]The raw data come from the GAEZ 2012 database of the IIAS/FAO.

[32]The survey questions are: "Do immigrants make the country a better or a worse place?", and "Do you trust the parliament?".

and governance.

For further details on the dataset construction see section in the Annex, and for descriptive statistics see Annex Table A.1 and A.2.

4 Empirical specification and results

We start the empirical analysis by first investigating the relationship between pre-industrial climate risk and pre-industrial migration flows using the genealogical data (constructed from genealogical birth–death locations and aggregated over 100-year periods). Next, we look at the long-run impact of pre-industrial climate risk on current migration patterns using the contemporary migration dataset, followed by various robustness tests. Finally, we also present empirical results exploring potential mechanisms which are likely responsible for the long-run effects.

4.1 Historical migration flows

In this section, we present results using our genealogy based dataset. The main advantage of this dataset is that it has a time dimension, which allows us to investigate the contemporaneous effects of climate variation. In addition, we exploit the bilateral dimension of the data, which means that our empirical specification maps into the gravity model. While the gravity model has been extensively used to empirically estimate trade flows since Tinbergen (1962), and the theoretical foundations have been linked to different trade models (see an overview in Head and Mayer (2014)), it has also been applied to other types of flows between countries, including migration flows.[33] We follow the literature and use a Poisson pseudo-maximum likelihood model (Silva and Tenreyro, 2006), since it is consistent in the presence of fixed effects and allows our dependent variable to take on zero values.

Therefore, our empirical specification is:

$$M_{odt} = exp[\beta_1 V_{dt} + \beta_2 P_{dt} + \emptyset_{ot} + \rho_{od} + e_{odt}], \tag{1}$$

where M_{odt} is our dependent variable, which is the number of individuals born in origin cell o who died in destination cell d, aggregated within 100-year period t [34] with migration being limited to international migration.[35] Our main variable of interest is V_{dt}, the climate variation in the destination cell, calculated as the average climate variation over 100 year periods, and lagged by 25 years compared to the migration periods. Given our theoretical framework, we expect that migrants learn about destination climate risk over time, including through repeated temporary and seasonal mobility.. Such learning might be particularly relevant in environments where information exchange was relatively

[33] Beine et al. (2016) provide a good overview of the gravity model's application to international migration flows and its theoretical basis.

[34] As a robustness check, we also run regressions with shares (calculated as the share of origin cell population), and we find similar results. Results are available in Table A.3 in the Annex.

[35] To distinguish international from national migration, we use the current borders as a proxy for country borders. While this might not be exactly accurate for some historical border cells, we also include cell-time specific fixed effects for origin cells. In addition, when we run regressions with all migrants (i.e., including domestic migration), our results hold, indicating a negative relationship between climate variation and inward migration. See column 4 of Table A.4

slow, such as pre-industrial Europe. Therefore, we lag climate variation, consistent with a setting in which information about destination conditions diffuses gradually and migration decisions, whether repeated temporary moves or longer-run settlement, respond with delay. In addition we also provide a robustness check without such a lag (see Table A.6). In addition, we control for origin cell-time specific factors with origin-time fixed effects ($\varnothing_{ot}$), and for origin-destination pair cell specific factors with pair fixed effects (ρ_{od}). In order to control for changing economic activity in the destination cell over time, we include population density in the destination (P_{dt}), an important pull factor. In the economic history literature, population density is often used as a proxy for economic activity, particularly for time periods and locations for which there is no available GDP data (e.g. Acemoglu et al. (2002), Maddison (2007), Henderson et al. (2017)). Because the genealogy data is very detailed and goes back as early as 1500, we use population density calculated from this data. Standard errors are clustered at the level of destination cells.

We believe that there is little evidence to suggest that spatial correlation would be a concern in the panel setting of the genealogical data. According to Müller (2023), spatial correlation models posit that units that are close to each other are potentially strongly correlated, whereas units that are far apart are assumed nearly independent. However, in our setting, it is unclear what the location of a pair of units, or, more generally, the distance between two pairs is. This distance would not govern unmodelled residual correlation in a gravity model. In the genealogical-based regressions, we therefore cluster the errors at the level of destination cells, given that our main variable of interest are destination-time specific. Even so, in Table A.4, column 3, we collapse our data to the destination-time level and include Conley standard errors to deal with spatial autocorrelation, with results remaining very similar.

In the first column of Table 1, we present results using our main specification, and find that the coefficients on both climate risk variables are negative and significant. [36] The results in column 1 suggest that a one unit increase in precipitation risk decreases inward migration flows by 8%. The estimated effect of temperature variability is substantially larger in magnitude, indicating a very strong negative association. Considering the coefficient on population density (which is a proxy for historical income levels), a one percent increase in this variable increases historical bilateral inward migration flows by 0.6%. Mean precipitation has a positive and significant effect on migration, while mean temperature is insignificant. To compare the magnitudes of the coefficients, we turn to column 2 of Table 1, where we show standardized beta coefficients, and we find that temperature risk had a much more pronounced negative effect on historical migration than precipitation risk. In column 3, to test whether temperature and precipitation risk matters jointly, we add an interaction term between temperature variability and precipitation variability, which is not significant. In column 4, to control for potential non-linear effects of climate risk, we add quadratic terms for both precipitation variability and temperature variability. Precipitation variability and its quadratic term are now insignificant, but there seems to be a U-shaped relationship for temperature variability. The marginal effect for the standard deviation of temperature is negative (- 6.205), which is calculated at the sample mean and takes into account the joint effect of the standard deviation of temperature and its quadratic term.

One important word of caution is that the genealogy data is not representative, since

[36] Note that the number of observations shown in the table does not include observations dropped from the regressions due to being perfectly predicted by fixed effects.

it is based on a self-selected sample of adults who looked for their ancestors at some point in time. Thus our definition of migration likely significantly underestimates the rate of migration. Given these caveats, we believe that the genealogy data is useful for identifying trends and less useful for identifying the exact magnitudes of the coefficients.

Table 1: Genealogy-based bilateral migration flows

	(1) Levels	(2) Standardized	(3) Interactions	(4) Non-linear
Precipitation variability	-0.085	-0.175	-0.304	-0.033
	(0.040)**	(0.082)**	(0.234)	(0.103)
Temperature variability	-2.593	-2.306	-2.812	-13.656
	(0.624)***	(0.555)***	(0.603)***	(2.330)***
Mean precipitation	0.066	0.428	0.060	0.084
	(0.031)**	(0.204)**	(0.033)*	(0.032)***
Mean temperature	-1.774	-4.500	-1.493	-2.205
	(1.258)	(3.192)	(1.341)	(1.264)*
Log of destination pop	0.621	1.229	0.619	0.645
	(0.066)***	(0.130)***	(0.066)***	(0.061)***
Precipitation*temperature var.			0.047	
			(0.050)	
Squared precipitation var.				-0.003
				(0.006)
Squared temperature var.				0.870
				(0.181)***
N	20,528	20,528	20,528	20,528
FE	Yr x orig pair	Yr x orig pair	Yr x orig pair	Yr x orig pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the number of migrants from an origin cell to a destination cell over 100-year periods. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons with a lag of 25 years. Standard errors are clustered at the level of destination cells. All specifications include origin cell-year and destination-origin cell pair fixed effects.

4.2 Contemporary migration

After having established that historical climate risk negatively affected historical migration flows, we now turn to the impact of climate risk on current migration patterns using our contemporary migration data. Based on the outlined conceptual framework, we estimate the following empirical specification using OLS:

$$M_{c,r} = \beta_0 + \beta_1 V_{c,r} + \beta_2 X_{c,r} + v_r + e_{c,r} \tag{2}$$

where the outcome variable $M_{c,r}$ is the share of migrants in a given cell c and region r in year 2011.[37] The variable of interest is $V_{c,r}$, which is our proxy of climate risk, captured by the standard deviation in precipitation and temperature over the period 1500-1800. In addition, we control for other factors specific to the location ($X_{c,r}$). More specifically, climate controls include mean precipitation and mean temperature over the same period, while geographical controls include land suitability for agricultural activities, whether the cell is located in a coastal region, on a river, its (log) distance to coast, its altitude, its (log) area size, and its latitude and longitude. In order to control for economic activity

[37] As a robustness test we also use the (log) of number of migrants in a given cell (A.9).

in the cell, we use population density measured by (log) total population in the cell (and as robustness we use light intensity). Finally, we include a dummy variable for whether the cell is a historical city or rural area. All specifications also include region-specific fixed effects (v_r) at the NUTS 1 level. The fixed effects capture time-invariant region-specific factors influencing the number of migrants, such as economic conditions, as well as cultural and historical factors (for example, past colonial ties).

Table 2 presents three versions of Equation 2, with all specifications including fixed effects at the regional (NUTS) level, and standard errors clustered at the same level. In the next section, we also discuss results with standard error corrections controlling for spatial dependence and show that are our results remain robust. The first column shows results for a linear specification, where both variability in precipitation and temperature enter linearly. We find that while historical temperature variation is negative, but not significant, variability in precipitation has a negative impact on the share of migrants at 5% significance. A one-unit increase in the standard deviation of historical precipitation (measured in cm) decreases the share of migrants in a given cell by approximately half a percentage point (0.48). Given that the mean share of migrants in our sample is 7.28%, and the median is 5.71%, these effects correspond to 6.6% and 8.4% decrease relative to the mean and median of the dependent variable, respectively. Regarding other control variables, we find that there is a higher share of migrants in locations which are at lower altitudes, have smaller sizes, and are located on a river and in cities. Using standardized beta coefficients shown in column 2 to compare the relative effects, we see that pre-industrial precipitation risk has a relatively more important role in driving present day migration patterns in Europe compared to the other control variables. In column 3, we also include an interaction term between precipitation variability and temperature variability. The interaction effect is significant and negative, with the marginal effects for both precipitation and temperature remaining negative, indicating a joint relationship. In column 4, we include quadratic terms for both precipitation variability and temperature variability but these quadratic terms are not significant.

4.3 Robustness checks

4.3.1 Historical genealogical migration data

In Table A.3, we run regressions with the dependent variable calculated as the share of migrants relative to the origin cell population, and we find that our results hold. In Table A.4 we implement additional robustness checks using the genealogical data. In column (1), we include climate variation from previous periods to make sure our results are not driven by the persistent nature of climate and climate variability. This specification only allows the inclusion of origin-year fixed effects (as there is no time variation now in previous period's climate variables). In column (2), we also control for origin climate thus this specification does not include origin-time fixed effects. In column (3), we collapse the data to the destination-time level (we also include Conley standard errors to address spatial autocorrelation), while in column (4) we include both international and national migrants. Our results remain robust across specifications.

Moreover, in Table A.5 we conduct two falsification tests with the genealogical migration regressions. First, we regress bilateral historical migration flows on future climate (moving climate 300 years forward), and climate variation becomes insignificant as expected. (Using climate variability shifted 300 years forward provides a placebo that

Table 2: Contemporary migration: cross-sectional analysis

Dependent var : Migrants	(1) Levels	(2) Standardized	(3) Interactions	(4) Non-linear
Precipitation variability	-0.480	-0.131	2.656	-0.942
	(0.194)**	(0.053)**	(0.873)***	(0.409)**
Temperature variability	-0.068	-0.006	3.311	-2.574
	(0.970)	(0.092)	(1.136)***	(4.731)
Mean Precipitation	0.064	0.073	-0.005	0.057
	(0.037)*	(0.042)*	(0.038)	(0.038)
Mean Temperature	-0.117	-0.049	-0.194	-0.118
	(0.116)	(0.049)	(0.111)*	(0.106)
Land suitability	0.249	0.010	0.305	0.229
	(0.706)	(0.028)	(0.758)	(0.671)
Coastal region	0.331	0.019	0.288	0.305
	(0.391)	(0.022)	(0.402)	(0.394)
Distance to coast	0.326	0.068	0.324	0.341
	(0.222)	(0.046)	(0.232)	(0.230)
Altitude	-1.480	-0.077	-1.270	-1.540
	(0.582)**	(0.030)**	(0.566)**	(0.560)***
On a river	1.959	0.047	1.930	1.945
	(0.512)***	(0.012)***	(0.516)***	(0.508)***
Area size	-1.017	-0.097	-1.128	-1.011
	(0.470)**	(0.045)**	(0.466)**	(0.466)**
Total population	0.184	0.075	0.187	0.181
	(0.138)	(0.056)	(0.138)	(0.138)
City	0.908	0.044	0.925	0.924
	(0.337)***	(0.016)***	(0.327)***	(0.335)***
Longitude	0.249	0.035	0.286	0.244
	(0.181)	(0.026)	(0.163)*	(0.170)
Latitude	-0.385	-0.055	-0.327	-0.366
	(0.278)	(0.039)	(0.262)	(0.262)
prec.*temp. variability			-0.662	
			(0.195)***	
Precipitation var. Sq.				0.033
				(0.020)
Temperature var. Sq.				0.262
				(0.529)
R^2	0.35	0.35	0.35	0.35
N	7,097	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes	Yes
Prep. var. marginal eff.			-0.308	-0.575
Temp. var. marginal eff.			-0.382	-0.229

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects. Marginal effects for precipitation and temperature variability are reported in the last two rows.

preserves spatial correlation but breaks any causal link with historical migration.) Since using climate 300 years forward restricts our sample, to show that climate variation does not become insignificant simply due to the reduced sample size, in the second column we rerun our main specification using the same sample as in column 1 but with contemporaneous climate variation, and climate variation remains significant. The second placebo exercise is shown in the third column, where we use climate variation during the non-growing season. If the results are indeed driven by agricultural activity, non-growing season climate variation would be insignificant, which is what we find, reinforcing our results that it is past growing season climate variation what matters.

Finally, results are also robust to using no lags in climate variation, which is shown in Table A.6, and to using 50-year periods instead of 100 years (Table A.7).

4.3.2 Contemporary migration data

Next, we probe the robustness of our results by running additional regressions using our current migration data. First, we run a specification to test if the link between climate variation and current in-migration is indeed driven by agricultural activity. In Table 3, in the first two columns we run separate regressions for areas which were rural and city in pre-industrial times. If the impact of climate risk on migration is driven by agricultural activities, we would expect the impact of climate variation to be smaller or negligible in historical cities, and higher in rural areas. The results in columns 1 and 2 are based on split-sample estimates, with column 1 using a sample which is restricted to historical cities, while column 2 includes only historical rural areas. Our results are in line with the expectations, with a significant negative relationship between climate variation and migration only observed in the rural sample. The magnitude of the precipitation variability coefficient is slightly higher (but very close to the previous results) for rural areas, while for historical city locations it is much lower and insignificant, with a t-test ruling out equality of the city and rural coefficients.

In the last column of Table 3 we introduce a cell-specific measure of potato suitability as a proxy for potato cultivation (while actual potato cultivation might be endogenous, suitability itself is not). Potatoes are known to be more resilient to climate variation than other crops (McNeill (1999)). If agriculture activity is the channel through which climate risk affects migration, we expect the relationship to be attenuated in areas more suitable for potato cultivation. In the last column of the table, we interact our two climate variability measures with potato suitability (ranging from zero to one). The results provide further support for the agricultural channel: higher potato suitability significantly mitigates the negative impact of climate risk on migration.

We undertake further robustness checks for which results are presented in Annex Table A.8. More specifically, we drop cells which are at country borders, as in such cases non-natives might not represent actual migrants (this can be the case when historical borders shifted and the local population remained). Furthermore, we also use light intensity as a measure of economic activity instead of population density. Moreover, we re-estimate our main specification using climate variation calculated with alternative periods (i.e., 1500-1700, 1500-1750, 1500-1850), since industrialization started at slightly different times in different locations. This also addressed an additional concern. The recent literature has suggested that there might be reverse causality between crop expansion and rainfall/temperature (Mueller et al., 2016), although this being a recent phenomenon

Table 3: Contemporary migration, rural versus city destinations, potato suitability

Dependent var : Migrants	Only city	Only rural	Potato suitability
Precipitation variability	-0.013	-0.513	-0.658
	(0.288)	(0.202)**	(0.227)***
Temperature variability	-1.712	0.106	-1.158
	(1.281)	(1.072)	(0.825)
Mean Precipitation	-0.196	0.088	0.031
	(0.086)**	(0.037)**	(0.043)
Mean Temperature	-0.842	-0.073	-0.244
	(0.277)***	(0.117)	(0.113)**
Land suitability	-1.685	0.239	0.177
	(1.558)	(0.784)	(0.592)
Coastal region	0.627	0.356	0.312
	(0.829)	(0.466)	(0.380)
Distance to coast	1.247	0.210	0.274
	(0.453)***	(0.217)	(0.204)
Altitude	-2.845	-1.502	-1.603
	(2.039)	(0.629)**	(0.598)***
On a river	0.965	2.140	2.025
	(1.528)	(0.665)***	(0.542)***
Area size	-1.366	-1.032	-1.056
	(1.159)	(0.498)**	(0.487)**
Total population	0.039	0.218	0.214
	(0.258)	(0.129)*	(0.143)
Longitude	0.272	0.259	0.271
	(0.272)	(0.193)	(0.185)
Latitude	-0.479	-0.347	-0.453
	(0.384)	(0.295)	(0.277)
Potato suitability			-24.847
			(9.384)**
Prec.var * Potato suit.			0.825
			(0.421)*
Temp.var * Potato suit.			4.892
			(2.243)**
R^2	0.49	0.31	0.35
N	1,003	6,092	6,967
Regional FE	Yes	Yes	Yes
Test for equality of			
coeff. of Precipitation variability		3.89	
p value		0.049	

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects. The first column uses a sample restricted to historical cities, while the last column uses a sample with cells only in rural areas.

driven by irrigation and modern agricultural practices. Modern irrigation and agriculture techniques did not start until the late 18th century in Europe, and even later (in the 1850s) in the US (Leibundgut and Kohn, 2014). Therefore, before 1700, agriculture was subsistent and the issue of reverse causality would not be relevant. Thus these results where the cut-off for preindustrial climate variation is 1700, is prior to the period when such reverse causality would be an issue. Throughout all these various specification, our results hold, with precipitation variability remaining significant at 5% with a coefficient similar in size.

Several historians argue that the local population did not emigrate as a result of weather/yield fluctuations in pre-industrial times (see for example Deschacht and Winter (2015b), Jacquemyns (1928), Vandenbroeke (1979), Dribe (2003)). If contrary to this, climate variation affected native out-migration, our immigration rate measure could also be capturing changes in local population movements. Therefore, as a robustness, we use the total number of migrants (in logs) as dependent variable instead of using shares, which might be affected by changes in local population numbers (Annex Table A.9). Our main results hold, with the sign and significance of our main variables of interest remaining the same.

Next, we explore three alternative measure of climate variation: a principal component index measuring joint precipitation and temperature variability, absolute mean deviation of precipitation and temperature, and a joint principal component based index combining the separate absolute deviation measures (Annex Table A.10). The joint measure based on standard deviations is negative, and significant at 10%, confirming our main results. Using the alternative absolute mean deviation, only the joint measure is significant, and negative. These results again indicate a negative relationship between historical climate risk and inward migration.

Given that our dependent variable is a share, as a further robustness test we replicate our main results presented in Table 2 using a fractional logit estimation. We again find that historical climate risk in the form of precipitation variability reduces the share of current migrants, with the effect of this being also non-linear. The interaction of precipitation and temperature variability has a significant joint negative effect as well (Annex Table A.11).

Climate in a particular location can be correlated with climate in other locations, possibly also with those located far away. Given the context of our study - where pre-industrial migration did not involve movements across continents, but rather remained constrained to Europe - we are less worried about climate being correlated with climate in very distant locations outside Europe. Nevertheless, to account for such potential correlation we undertake standard error correction following Conley (1999) with different distance cut-off values and a Bartlett kernel (which assigns a higher weight to nearer points than to points that are further away). In addition, we also run regressions with a spatial autocorrelation model which assumes that the climate variables are spatially lagged and where we correct for spatial autocorrelation in the climate variables with a spatial weighting matrix using the inverse of the distance. In Annex Table A.12, we replicate Table 2 using a 150 km cut-off with Conley standard error correction. Significance increases, and becomes 1% for our main variables of interest (Annex Table A.12). In columns 1-5 of Table A.13, we run our main contemporary regression specification with with different cut-off distances (from 200 to 2000 km), while in the last column of the same table we present results using the spatial autocorrelation model where the climate variables are assumed to be spatially lagged. In all specifications, our results on

precipitation variation remain significant at 1%.

In Table A.14 we test the sensitivity of the results for the contemporary migration specifications to including different controls by including the explanatory variables gradually. Our results remain robust.

It is important to note that our specification with contemporary migration data looks at somewhat different historical temperature and precipitation variability measures compared to our specification with genealogical migration data, as the former is calculated over the whole time period while the latter is calculated for hundred-year periods. If variability was increasing or decreasing over these three centuries, the variability measure would be a larger compared to the century-long variability measures. Tables 5 and 6 in the paper's appendix show that this concern is not borne out by the data, since climate variability is very similar for these different periods. In the genealogy analysis, the mean of temperature variability is 4.85, with a standard deviation of 0.89, while the corresponding figures for temperature variability in the contemporary data are 4.43, with a standard variation of 0.66. The respective statistics for precipitation variability are 5.87 (mean) and 2.05 (standard deviation) for the genealogy analysis, and 5.54 and 1.87 for the contemporary data.

Finally, we also run our main specification with contemporary migration data on the three 100 year periods separately. While this addresses also the previous concern to some extent, it also provides further reassurance about the quality of the climate data. There could variations in proxy coverage, calibration, and data quality across the 1500–1800 period, in addition there could be some particular periods driving our results. The results in Annex Table A.15 show that this is unlikely to be the case, as the main results remain robust throughout the different centuries.

4.4 Potential mechanisms

Thus far, we have shown evidence consistent with historical climate risk affecting historical migration through agricultural conditions. However, it is still unclear how these effects persisted and influenced today's migration patterns. In this section, we explore the potential mechanisms discussed in the conceptual framework. Testing mechanisms is inherently complex, particularly in contexts where climate variability can simultaneously influence multiple outcomes. Accordingly, this section is exploratory and aims to assess consistency with broad channels rather than to provide a causal decomposition. As such, the results presented here should be interpreted cautiously as indicative of potential mechanisms rather than as definitive evidence. In order to test the role of the mechanisms, we first run regressions where we include each potential mechanism as a dependent variable to see if it is correlated with historical climate variation. The results are presented in Annex in Table A.16.[38] Next, we include each potential mechanism as an explanatory variable in our earlier specification using the contemporary migration data. These results are presented and discussed below. When the mechanism variable is significant and attenuates the reduced-form association between climate risk and migration,

[38]In the 18th century the industrial revolution was taking place and there was considerable internal migration from rural to city areas. Thus in the specifications with past migrants (calculated over the period 1700-1800), GDP per capita (covering 1800) and population density (covering 1800), in addition to our standard control variables, in these specifications we also include a control for percent of the area covered by cropland in 1870, the earliest available year. As expected, we find a negative correlation between cropland and GDP per capita/population density in 1800.

we interpret this pattern as consistent with that channel playing a role, while recognizing that this does not constitute a causal decomposition.

One possibility is that past migrants attracted more migrants to locations with lower climate risk through network effects (Cattaneo, 2019). In order to test this potential mechanism, using our genealogy dataset, we calculated the number of international migrants in each location and from all origins in the one-hundred-year period 1700-1800 as a proxy for historical migrant networks. Column 1 in Table 4 includes this measure of past international migrants as an additional control variable. Consistent with our conceptual framework, the magnitude of the precipitation variation coefficient drops and becomes insignificant, which is consistent with a role for historical migrant networks in the persistence we observe.[39]

[39]Results are robust if we instead use international migrants over the period 1800-1900.

Table 4: Contemporary migration, testing potential mechanisms

Dep. var: Migrants	(1) Past migrants	(2) Conflict	(3) Trade	(4) GDPperCap	(5) Pop.Density	(6) Attitudes	(7) Institutions
Precipitation var.	-0.228	-0.484	-0.307	-0.177	-0.223	-0.558	-0.473
	(0.144)	(0.195)**	(0.068)***	(0.070)**	(0.145)	(0.068)***	(0.068)***
Temperature var.	-0.125	-0.083	-0.203	0.197	-0.077	-0.195	-1.095
	(1.204)	(0.976)	(0.312)	(0.322)	(1.199)	(0.310)	(0.318)***
Mechanism	0.008	0.510	-0.439	5.381	0.482	3.411	3.612
	(0.003)**	(0.292)*	(0.083)***	(0.534)***	(0.151)***	(0.305)***	(0.289)***
R^2	0.38	0.35	0.18	0.17	0.38	0.21	0.21
N	5,337	7,097	7,056	6,816	5,337	7,063	7,063
Regional FE	Yes	Yes	No	No	Yes	No	No
Country FE	No	No	Yes	Yes	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Specifications where the mechanism variable is available at local level include regional fixed effects, for those where the mechanism variable is available only at regional level country fixed effects are included (in which case robust standard errors are shown). The full list of control variables (as in Table 2) are included, but not presented in the table.

In column 2, we test the hypothesis that locations with past conflicts drove away migrants, while still having an impact on today's migration patterns. We find no significant effect of past conflicts on today's migration and the impact of climate variation remains unchanged. In addition, we find that when including trade in the specification, the effect of precipitation risk remains significant and negative (column 3).

In columns 4 and 5, we look at whether economic activity might be driving our results. We use historical GDP per capita (from 1800),[40] and historical population density (calculated for 1800 from our genealogy dataset). When including these measures, the coefficient size on precipitation drops to less than half, and with population density even becomes insignificant. We interpret this pattern as consistent with historical economic conditions being an important correlate of the persistent effect on current migration. Higher climate risk is associated with lower historical GDP per capita (see these results in the Annex in Table A.16 where we run each of these mechanism variables on climate risk), which are in turn correlated with lower current migration.

The next two columns in the table test the importance of two additional channels. First, it could be that past climate variation led to more welcoming attitudes towards migrants, as well as more migrant-friendly institutions, which persisted until today. In column 6, we include current attitudes towards migrants, and in column 7, current trust in parliament (as a proxy for more open institutions). While these variables measure today's attitudes and institutions (at regional level) and not past attitudes, we expect that these factors are highly persistent over time (see (Buggle and Durante, 2021) and Voigtländer and Voth (2012) on the long-term persistence of similar attitudes). When including either attitudes towards migrants or trust in parliament, climate variability remains significant and negative. In other words, even though we find that the correlation between attitudes towards migrants and trust in parliament is positive for current migrants, the impact of climate variation remains the same, indicating that its impact is not going through these factors. In Table A.17, we replicate the specifications in Table 4, but using Conley standard errors with a 150 km cutoff, with results being very similar across the two tables.

5 Conclusion

Using two novel and highly disaggregated sub-national datasets, we examine the impact of historical climate risk on inward migration in Europe. First, using our bilateral genealogy dataset, we find that historical climate risk (captured by precipitation and temperature variability) negatively influenced historical migration flows. Next, using a different dataset containing the stock of migrants in 2011, we find that this negative relationship between the historical variability in precipitation and the share of current migrants persisted until today. A one-unit increase in the standard deviation of historical precipitation decreases the share of migrants in a given cell by 0.48 percentage points (6.6% relative to the mean of the dependent variable). In addition, the combination of historical temperature and precipitation variability has a joint negative effect on today's migration stocks. We find that these results are concentrated in localities that were historically rural and that they are strongest for climate variability during the growing season of major crops, consistent with agriculture playing a central role in the long-run relationships we identify. Moreover, we find suggestive evidence that historical prosperity levels and past migration are closely related to the persistence of climate

[40] Data were obtained from Buggle and Durante (2021).

effects on today's migration patterns. As predicted by our theory, migration network effects were also behind the persistence we uncover. Past climate risk affected negatively past migration, which in turn is negatively associated with today's migration patterns. These results withstand multiple rigorous robustness checks, including the consideration of spatial correlation using Conley standard errors.

Our results have important implications for the understanding of both the historical and contemporary relationship between climate risk and migration, in both developed and developing countries. Our historical analysis shows that climate risk in the *destination* location matters, that the relationship between historical climate risk and international in-migration today is negative, and that these patterns are consistent with an important role for historical income levels. These results suggest that long-run and current climate risk are relevant considerations when designing policies aimed at attracting migrants to a particular location. In addition, climate-mitigating policies may be particularly important in agriculture-dependent localities.

Compared to much of the existing literature, the spatial resolution of our data and the use of rich fixed effects strengthen the credibility of a causal interpretation and reduce concerns about omitted variables and measurement noise. Our findings provide novel insight on the importance of historical climate variation for migration. Although our analysis does not differentiate among the different types of migration (e.g., temporary/seasonal versus permanent settlement), future research could also examine how different types of migration are affected by historical climate risk, both in Europe and in other contexts. Digging deeper into the mechanisms that we have identified - including through historical and archival research - might also help illuminate the long-term impact of climate risk on current migration.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics 117*(4), 1231–1294.

Ager, P. and A. Ciccone (2018). Agricultural risk and the spread of religious communities. *Journal of the European Economic Association 16*(4), 1021–1068.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alessandrini, A., F. Natale, F. Sermi, and M. Vespe (2017). High resolution map of migrants in the eu. Technical report, JRC Technical Reports EUR 28770 EN, doi: 10.2760.

Auffhammer, M. and W. Schlenker (2014). Empirical studies on agricultural impacts and adaptation. *Energy Economics 46*, 555–561.

Bade, K. J. (2002). *L'Europe en mouvement: la migration de la fin du XVIIIe siècle à nos jours*. Seuil.

Bairoch, P., J. Batou, and C. Pierre (1988). *Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats (la)*. Librairie Droz.

Bateman, V. N. (2011). The evolution of markets in early modern europe, 1350–1800: a study of wheat prices. *The Economic History Review 64*(2), 447–471.

Beine, M., S. Bertoli, and J. Fernández-Huertas Moraga (2016). A practitioners? guide to gravity models of international migration. *The World Economy 39*(4), 496–512.

Beine, M., F. Docquier, and C. Ozden (2011). Diasporas. *Journal of Development Economics 95*(1), 30–41. Symposium on Globalization and Brain Drain.

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

Beine, M. A. and L. Jeusette (2021). A meta-analysis of the literature on climate change and migration. *Journal of Demographic Economics*, 1–52.

Berlemann, M. and M. F. Steinhardt (2017). Climate change, natural disasters, and migration—a survey of the empirical evidence. *CESifo Economic Studies 63*(4), 353–385.

Blanc, G. (2024a). The cultural origins of the demographic transition in france. *Working paper*.

Blanc, G. (2024b). Demographic transitions, rural flight, and intergenerational persistence: evidence from crowdsourced genealogies. *Working paper*.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. Technical report, National Bureau of Economic Research.

Bosker, E. M., E. Buringh, and J. Luiten van Zanden (2008). From baghdad to london: The dynamics of urban growth in europe and the arab world, 800-1800. *Vol*.

Bosker, M., E. Buringh, and J. L. Van Zanden (2013). From baghdad to london: Unraveling urban development in europe, the middle east, and north africa, 800–1800. *Review of Economics and Statistics 95*(4), 1418–1437.

Boustan, L. P., M. E. Kahn, and P. W. Rhode (2012). Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review 102*(3), 238–44.

Braudel, F. (1973). *Capitalism and material life, 1400-1800*. HarperCollins.

Buggle, J. C. and R. Durante (2021). Climate risk, cooperation, and the co-evolution of culture and institutions. *The Economic Journal*.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Climate and conflict. *Annu. Rev. Econ 7*, 577–617.

Bursztyn, L., T. Chaney, T. A. Hassan, and A. Rao (2023). The immigrant next door: Long-term contact, generosity, and prejudice. *American Economic Review Forthcoming*.

Burzyński, M., C. Deuster, F. Docquier, and J. D. Melo (2022). Climate change, inequality, and human migration. *Journal of the European Economic Association 20*(3), 1145–1197.

Cai, R., S. Feng, M. Oppenheimer, and M. Pytlikova (2016). Climate variability and international migration: The importance of the agricultural linkage. *Journal of Environmental Economics and Management 79*, 135–151.

Campbell, B. M. (2010). Nature as historical protagonist: environment and society in pre-industrial england. *The Economic History Review 63*(2), 281–314.

Cattaneo, C. (2019). Migrant networks and adaptation. *Nature Climate Change 9*(12), 907–908.

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

Clark, G. (2023). The inheritance of social status: England, 1600 to 2022. *Proceedings of the National Academy of Sciences 120*(27), e2300926120.

Colasurdo, A. and R. Omenti (2024). Using online genealogical data for demographic research: An empirical examination of the familinx database. *Center for Open Science Working Paper*.

Collins, E. J. T. (1976). Migrant labour in british agriculture in the nineteenth century. *The Economic History Review 29*(1), 38–59.

Colmer, J. (2021). Temperature, labor reallocation, and industrial production: Evidence from india. *American Economic Journal: Applied Economics 13*(4), 101–24.

Coniglio, N. D. and G. Pesce (2015). Climate variability and international migration: an empirical analysis. *Environment and Development Economics 20*(4), 434–468.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics 92*(1), 1–45.

Corti, G., S. Minardi, and N. Barban (2023). Trends in assortative mating in the united states, 1700–1910. evidence from familinx data. *The History of the Family 29*(4), 461–481.

Cozzani, M., S. Minardi, G. Corti, and N. Barban (2023). Birth month and adult lifespan. *Demographic Research 49*, 201–218.

Damania, R. (2020). The economics of water scarcity and variability. *Oxford Review of Economic Policy 36*(1), 24–44.

Damania, R., S. Desbureaux, and E. Zaveri (2020). Does rainfall matter for economic growth? evidence from global sub-national data (1990–2014). *Journal of Environmental Economics and Management 102*, 102335.

Davis, R. (1973). *The Rise of the Atlantic Economies*. Cornell University Press.

Dell, M., B. F. Jones, and B. Olken (2012). Temperature shocks and economic growth: Evidence from the last half centur. *American Economic Journal: Macroeconomics 4*(3), 66–95.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Deschacht, N. and A. Winter (2015a). Rural crisis and rural exodus? local migration dynamics during the crisis of the 1840s in flanders (belgium). *Explorations in Economic History 56*, 32–52.

Deschacht, N. and A. Winter (2015b). Rural crisis and rural exodus? local migration dynamics during the crisis of the 1840s in flanders (belgium). *Explorations in Economic History 56*, 32–52.

Devine, T. (1984). Women workers, 1850-1914. *Farm servants and labour in lowland Scotland, 1770-1914/edited by TM Devine*.

Dincecco, M. and M. G. Onorato (2018). *From Warfare to Wealth*. Cambridge University Press.

Dribe, M. (2003). Migration of rural families in 19th century southern sweden. a longitudinal analysis of local migration patterns. *The History of the Family 8*(2), 247–265.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of development economics 84*(2), 755–776.

Galloway, P. R. (1986). Long-term fluctuations in climate and population in the preindustrial era. *Population and Development Review 12*(1), 1–24.

Galor, O. and Ö. Özak (2016). The agricultural origins of time preference. *American Economic Review 106*(10), 3064–3103.

Giuliano, P. and N. Nunn (2017). Understanding cultural persistence and change. Technical report, National Bureau of Economic Research.

Giuliano, P. and N. Nunn (2021). Understanding cultural persistence and change. *Review of Economic Studies 88*, 1541–1581.

Grantham, G. W. (1993). Divisions of labour: agricultural productivity and occupational specialization in pre-industrial france 1. *The economic history review 46*(3), 478–502.

Greif, A. and G. Tabellini (2010). Cultural and institutional bifurcation: China and europe compared. *American economic review 100*(2), 135–40.

Grogger, J. and G. H. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics 95*(1), 42–57.

Guiso, L., P. Sapienza, and L. Zingales (2016). Long-term persistence. *Journal of the European Economic Association 14*(6), 1401–1436.

Hahn, S. (2000). Inclusion and exclusion of migrants in the multicultural realm of the habsburg" state of many peoples". *Histoire sociale/Social History*.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, Volume 4, pp. 131–195. Elsevier.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics 133*(1), 357–406.

Hochstadt, S. (1983). Migration in preindustrial germany. *Central European History 16*(3), 195?224.

Hochstadt, S. (1999). *Mobility and modernity: migration in Germany, 1820-1989*. University of Michigan Press.

Hoffmann, R., A. Dimitrova, R. Muttarak, J. C. Cuaresma, and J. Peisker (2020). A meta-analysis of country-level studies on environmental change and migration. *Nature Climate Change 10*(10), 904–912.

Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review 102*(4), 1477–1507.

Hsiang, S. M., M. Burke, and E. Miguel (2013). Quantifying the influence of climate on human conflict. *Science 341*(6151), 1235367.

Intergovernmental Panel on Climate Change (n.d., Accessed: 2025-06-18). Frequently asked questions. *https://www.ipcc.ch/help/frequently-asked-questions*.

Jacquemyns, G. (1928). Histoire de la crise économique des flandres (1845-1850): Mémoires de la classe des lettres/académie royale de belgique: Collection in-8.

Kahn, M. E., K. Mohaddes, R. N. Ng, M. H. Pesaran, M. Raissi, and J.-C. Yang (2021). Long-term macroeconomic effects of climate change: A cross-country analysis. *Energy Economics 104*, 105624.

Kaplanis, J., A. Gordon, T. Shor, O. Weissbrod, D. Geiger, M. Wahl, M. Gershovits, B. Markus, M. Sheikh, M. Gymrek, et al. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science 360*(6385), 171–175.

Klein, A. and J. Van Lottum (2020). The determinants of international migration in early modern europe: Evidence from the maritime sector, c. 1700–1800. *Social Science History 44*(1), 143–167.

Kotz, M., A. Levermann, and L. Wenz (2022). The effect of rainfall changes on economic productio. *Nature 601*(7892), 223–227.

Ladurie, E. (1971). Times of feast, times of famine: A history of climate since the year 1000.(transl. b. bray.) doubleday. *Garden City, New York*, 1–426.

Leff, B., N. Ramankutty, and J. A. Foley (2004). Geographic distribution of major crops across the world. *Global biogeochemical cycles 18*(1).

Leibundgut, C. and I. Kohn (2014). European traditional irrigation in transition part i. irrigation in times past - a historic land use practice across europe. *Irrigation and Drainage 63*(3), 273–293.

Litina, A. (2016). Natural land productivity, cooperation and comparative development. *Journal of Economic Growth 21*(4), 351–408.

Lucassen, J. and L. Lucassen (2009). The mobility transition revisited, 1500?1900: what the case of europe can offer to global history. *Journal of Global History 4*(3), 347?377.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004). European seasonal and annual temperature variability, trends, and extremes since 1500. *Science 303*(5663), 1499–1503.

Maddison, A. (2007). *Contours of the world economy 1-2030 AD: Essays in macro-economic history*. Oxford University Press.

Mahajan, P. and D. Yang (2020). Taken by storm: Hurricanes, migrant networks, and us immigration. *American Economic Journal: Applied Economics 12*(2), 250–277.

Malpede, M., G. Falchetta, and S. Shayegh (2023). Mosquitoes and potatoes: How local climatic conditions impede development. *Environmental and Resource Economics 86*(4), 851–892.

Manchin, M. (2024, forthcoming). Global evidence on the relative importance of nonfinancial drivers of international migration intentions. *International Migration Review*.

Manchin, M. and S. Orazbayev (2018). Social networks and the intention to migrate. *World Development 109*, 360–374.

Mastrorillo, M., R. Licker, P. Bohra-Mishra, G. Fagiolo, L. D. Estes, and M. Oppenheimer (2016). The influence of climate variability on internal migration flows in south africa. *Global Environmental Change 39*, 155–169.

McNeill, W. H. (1999). How the potato changed the world's history. *Social Research*, 67–83.

Millock, K. (2015). Migration and environment. *Annu. Rev. Resour. Econ. 7*(1), 35–60.

Moch, L. (2003). *Moving Europeans: migration in Western Europe since 1650*. Indiana University Press.

Mueller, N. D., E. E. Butler, K. A. McKinnon, A. Rhines, M. Tingley, M. Holbrook, and P. Huybers (2016). Cooling of us midwest summer temperature extremes from cropland intensification. *Nature Climate Change 6*(3), 317–322.

Mueller, V., G. Sheriff, X. Dou, and C. Gray (2020). Temporary migration and climate variation in eastern africa. *World Development 126*, 104704.

Müller, U. K. (June 15, 2023). Email communication. *Email communication with the authors*.

Nannicini, T., A. Stella, G. Tabellini, and U. Troiano (2013). Social capital and political accountability. *American Economic Journal: Economic Policy 5*(2), 222–50.

Nettle, D. (1998). Explaining global patterns of language diversity. *Journal of anthropological archaeolog 17*(4), 354–374.

Nunn, N. and N. Qian (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics 126*(2), 593–650.

Oris, M. (2003). The history of migration as a chapter in the history of the european rural family: An overview. *The history of the family 8*(2), 187–215.

Paprotny, D. and M. Mengel (2023). Population, land use and economic exposure estimates for europe at 100 m resolution from 1870 to 2020. *Scientific Data 10*(1), 372.

Parker, G. (2001). *Europe in Crisis: 1598-1648, 2nd Edition*. Wiley.

Pauling, A., J. Luterbacher, C. Casty, and H. Wanner (2006). Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation. *Climate dynamics 26*(4), 387–405.

Pei, Q., Y. Long, and X. Lin (2022). Climate change in historical perspective: Violence, conflict, and migration. *Handbook of Labor, Human Resources, and Population Economics*.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Stark, O. and C. S. Fan (2007). The analytics of seasonal migration. *Economics Letters 94*(2), 304–312.

Stelter, R. and D. Alburez-Gutierrez. (2022). Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics. *Proceedings of the National Academy of Sciences 119*(10), e2120455119.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics 123*(3), 905–950.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of europe. *Journal of the European Economic association 8*(4), 677–716.

Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy.

Triyana, M. M., A. Weicheng Jiang, Y. Hu, and M. S. Naoaj (2024). Climate shocks and the poor: A review of the literature. *World Bank Policy Research Working Paper Series 10742 (2024)*..

van Bath, B. S. (1963). The yields of different crops (mainly cereal) in relation to the seed c. 810–1820. *Acta Hist. Neederlandica 2*, 26–106.

Van Lottum, J. (2007). *Across the North Sea: the impact of the Dutch Republic on international labour migration, c. 1550-1850*, Volume 1. Amsterdam University Press.

Vandenbroeke, C. (1979). *Sociale en konjunkturele facetten van de linnennijverheid in Vlaanderen (late 14e-midden 19e eeuw)*. Maatschappij voor Geschiedenis en Oudheidkunde te Gent.

Voigtländer, N. and H.-J. Voth (2012). Persecution perpetuated: the medieval origins of anti-semitic violence in nazi germany. *The Quarterly Journal of Economics 127*(3), 1339–1392.

Waldinger, M. (2022). The economic effects of long-term climate change: Evidence from the little ice age. *Journal of Political Economy 130*(9), 2275–2314.

Xoplaki, E., J. Luterbacher, H. Paeth, D. Dietrich, N. Steiner, M. Grosjean, and H. Wanner (2005). European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophysical Research Letters 32*(15).

Annex

A.1 Additional robustness checks and results

The following tables report robustness checks grouped by estimator choice, climate variable construction, spatial correlation, and sample restrictions.

Table A.1: Contemporary migration data, descriptive statistics for the main variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Share of migrants (in %)	7097	7.364	7.101	0	100
Precipitation variability (Cm)	7097	5.577	1.884	2.112	15.598
Temperature variability (C°)	7097	4.474	.634	2.79	5.981
Mean precipitation (Cm)	7097	18.905	7.982	4.532	65.002
Mean temperature (C°)	7097	13.026	2.798	.086	20.414
Temperature variability, NGS (C°)	7097	4.119	.729	2.233	5.583
Precipitation variability, NGS (Cm)	7097	4.56	1.725	1.787	14.307
Land suitability	7097	.619	.282	.001	.999
Coastal region	7097	.125	.331	0	1
Ln distance to coast (1000km)	7097	-2.756	1.469	-13.99	-.8
Altitude (km)	7097	.359	.375	-.003	2.743
On a river	7097	.033	.179	0	1
Ln area size (km2)	7097	5.706	.603	.952	6.305
Ln total population	7097	8.632	2.805	-7.013	14.876
City	7097	.142	.349	0	1
Light intensity	7097	19.843	28.709	0	337.524
Conflict	7097	.121	.326	0	1

Land suitability is the predicted value of the propensity of a given parcel of land to be under cultivation based on four measures of climate and soil, light intensity is the sum of the pixels brightness encoded in 64-bit, coastal region, on a river, city and conflict are dummies.

Table A.2: Genealogical migration data, descriptive statistics for the main variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Flow international migration (in levels)	20528	1.15	3.056	0	126
Precipitation variability (Cm)	20528	5.867	2.049	1.54	20.665
Temperature variability (C°)	20528	4.849	.889	2.599	8.074
Mean precipitation (Cm)	20528	19.23	6.487	3.468	70.823
Mean temperature (C°)	20528	11.112	2.536	-.561	22.463
Log of destination pop	20528	4.832	1.979	0	8.989

The number of observations in this table is based on the numbers in the Poisson Pseudo Maximum Likelihood regressions (where singletons are dropped). Note that regressions with shares (next table) are run with OLS, and the full sample of observations is used there.

Table A.3: Genealogy-based bilateral migration shares

	(1)	(2)	(3)
Precipitation variability	-0.000	0.001	0.001
	(0.000)	(0.002)	(0.001)
Temperature variability	-0.007	-0.006	-0.059
	(0.004)*	(0.004)*	(0.017)***
Mean precipitation	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Mean temperature	-0.004	-0.004	-0.008
	(0.007)	(0.007)	(0.007)
Log of destination pop	0.001	0.001	0.002
	(0.000)***	(0.000)***	(0.000)***
Precipitation*temperature variability		-0.000	
		(0.000)	
Squared precipitation variability			-0.000
			(0.000)
Squared temperature variability			0.004
			(0.001)***
R^2	0.43	0.43	0.43
N	8,809,365	8,809,365	8,809,365
FE	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is migrants as a percentage share of origin population over 100-year periods. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons with a lag of 25 years. Standard errors are clustered at the level of destination cells. All specifications include origin cell-year, origin-destination cell fixed effects. The regressions are run with OLS, and the full sample of observations is used here.

Table A.4: Genealogical migration data robustness

	(1) Lagged climate variation	(2) Origin climate	(3) Destination-level data	(4) All migrants	(5) 1700-1850
Temperature variability	-3.458	-1.913	-10.262	-1.795	-1.408
	(1.530)**	(0.579)***	(5.138)**	(0.402)***	(0.275)***
Precipitation variability	0.021	-0.051	0.118	-0.024	-0.020
	(0.073)	(0.042)	(0.438)	(0.020)	(0.027)
L.Temperature var.	2.923				
	(4.964)				
L.Precipitation var.	-0.152				
	(0.095)				
L2.Temperature var.	0.919				
	(3.772)				
L2.Precipitation var.	0.066				
	(0.117)				
Log of destination pop	0.901	0.561	2.359	0.770	0.737
	(0.080)***	(0.081)***	(0.478)***	(0.031)***	(0.045)***
Mean precipitation		0.031	-0.026	0.014	-0.019
		(0.025)	(0.265)	(0.016)	(0.021)
Mean temperature		-0.953	-14.595	0.319	-0.196
		(1.015)	(6.872)**	(0.729)	(0.119)*
Precipitation var., origin		0.001			
		(0.003)			
Temperature var., origin		1.849			
		(0.552)***			
Mean precipitation, origin		-0.007			
		(0.002)***			
Mean temperature, origin		-2.913			
		(0.737)***			
N	3,080,127	26,961	4,644	83,140	44,781
FE	Origin-year	Pair, Year	Destination,Year	Yr x orig, pair	Yr x orig, pair
R^2			0.69		

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the genealogical data. The dependent variable is the number of migrants from an origin cell to a destination cell over hundred-year periods (columns 1-4) and fifty-year periods (column 5, due to the short coverage of the data). Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons with a lag of 25 years. Standard errors are clustered at the level of destination cells in the first, second, fourth and fifth columns. Standard errors in the third column are corrected for spatial correlation with the Conley method.

Table A.5: Genealogical migration data, placebo specifications

	(1) Future growing season	(2) Contemporanous growing season	(3) Non-growing season
Precipitation variability	-0.117	-0.204	
	(0.126)	(0.085)**	
Temperature variability	0.590	-0.893	
	(1.001)	(3.759)	
Mean precipitation	-0.216	0.074	
	(0.098)**	(0.042)*	
Mean temperature	0.737	1.928	
	(0.953)	(2.663)	
Log of destination pop	1.049	1.110	0.660
	(0.095)***	(0.104)***	(0.063)***
Precipitation variability, NGS			0.001
			(0.005)
Temperature variability, NGS			1.240
			(2.186)
Mean precipitation, NGS			0.021
			(0.012)*
Mean temperature, NGS			-1.517
			(1.036)
N	4,874	4,874	20,528
FE	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the genealogical data. The dependent variable is the number of migrants from an origin cell to a destination cell over 100-year periods. Column (1) presents specifications with the climate variation calculated for the growing season 300 years ahead. Column (2) uses the same sample but instead includes the climate variation for the contemporaneous growing season. Column (3) uses climate variation calculated over the non-growing seasons with a lag of 25 years. Standard errors are clustered at the level of destination cells.

Table A.6: Genealogical migration data, no climate lags

Dependent var : Migrants	(1) Levels	(2) Standardized	(3) Interactions	(4) Non-linear
Precipitation variability	-0.034	-0.068	0.073	-0.010
	(0.048)	(0.097)	(0.286)	(0.122)
Temperature variability	-1.574	-1.443	-1.467	-7.615
	(0.382)***	(0.350)***	(0.433)***	(1.673)***
Mean precipitation	0.096	0.630	0.099	0.112
	(0.026)***	(0.167)***	(0.027)***	(0.026)***
Mean temperature	-3.759	-9.563	-3.801	-3.582
	(0.647)***	(1.646)***	(0.653)***	(0.627)***
Log of destination pop	0.643	1.273	0.646	0.649
	(0.070)***	(0.139)***	(0.070)***	(0.068)***
Precipitation*temperature variability			-0.022	
			(0.057)	
Squared precipitation variability				-0.001
				(0.007)
Squared temperature variability				0.499
				(0.140)***
N	20,528	20,528	20,528	20,528
FE	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the genealogical data. The dependent variable is the number of migrants from an origin cell to a destination cell over 100-year periods. The climate variables are not lagged. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons. Standard errors are clustered at the level of destination cells.

Table A.7: Genealogical migration data, 50 year periods

Dependent var : Migrants	(1) Levels	(2) Standardized	(3) Interactions	(4) Non-linear
Precipitation variability	-0.032	-0.069	-0.311	0.098
	(0.032)	(0.068)	(0.164)*	(0.073)
Temperature variability	-1.216	-1.116	-1.639	-5.416
	(0.332)***	(0.305)***	(0.382)***	(1.427)***
Mean precipitation	0.038	0.248	0.034	0.032
	(0.020)*	(0.132)*	(0.020)*	(0.020)
Mean temperature	-0.124	-0.318	-0.109	-0.012
	(0.191)	(0.488)	(0.193)	(0.200)
Log of destination pop	0.745	1.366	0.741	0.744
	(0.064)***	(0.116)***	(0.064)***	(0.062)***
precipitation*temperature variability			0.056	
			(0.033)*	
Squared precipitation variability				-0.008
				(0.003)**
Squared temperature variability				0.378
				(0.121)***
N	34,110	34,110	34,110	34,110
FE	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair	Yr x orig, pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the genealogical data. The dependent variable is the number of migrants from an origin cell to a destination cell over 50-year periods. The climate variables are not lagged. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons. Standard errors are clustered at the level of destination cells.

Table A.8: Contemporary migration data, further robustness: no border cells, light intensity, different periods

Dependent var : Migrants	Without border cells	Light intensity	1500-1700	1500-1750	1500-1850
Precipitation variability	-0.536	-0.435	-0.447	-0.473	-0.496
	(0.217)**	(0.194)**	(0.187)**	(0.194)**	(0.198)**
Temperature variability	0.011	0.022	-0.032	-0.060	-0.076
	(0.936)	(0.974)	(0.995)	(0.979)	(0.967)
Mean Precipitation	0.082	0.057	0.051	0.064	0.066
	(0.032)**	(0.040)	(0.034)	(0.039)	(0.038)*
Mean Temperature	-0.173	-0.119	-0.124	-0.116	-0.114
	(0.127)	(0.119)	(0.116)	(0.115)	(0.115)
Land suitability	0.131	0.093	0.183	0.242	0.238
	(0.548)	(0.710)	(0.704)	(0.698)	(0.710)
Coastal region	0.273	0.303	0.367	0.335	0.325
	(0.406)	(0.356)	(0.391)	(0.389)	(0.392)
Distance to coast	0.338	0.267	0.339	0.327	0.322
	(0.244)	(0.215)	(0.220)	(0.222)	(0.223)
Altitude	-2.017	-1.471	-1.439	-1.482	-1.476
	(0.744)***	(0.530)***	(0.578)**	(0.580)**	(0.582)**
On a river	2.183	1.612	1.957	1.967	1.955
	(0.441)***	(0.486)***	(0.512)***	(0.513)***	(0.512)***
Total population	0.181		0.185	0.183	0.184
	(0.147)		(0.138)	(0.138)	(0.138)
City	0.879	0.243	0.910	0.910	0.905
	(0.355)**	(0.369)	(0.336)***	(0.337)***	(0.337)***
Longitude	0.330	0.263	0.250	0.250	0.245
	(0.186)*	(0.174)	(0.177)	(0.181)	(0.179)
Latitude	-0.403	-0.368	-0.388	-0.381	-0.389
	(0.261)	(0.276)	(0.294)	(0.283)	(0.277)
Area size		-0.929	-1.046	-1.017	-1.016
		(0.448)**	(0.470)**	(0.469)**	(0.471)**
Light intensity		0.031			
		(0.010)***			
R^2	0.36	0.35	0.35	0.35	0.35
N	6,529	7,097	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects.

Table A.9: Contemporary migration data, log of total migrants in levels as dependent variable

Dependent var : # of Migrants	(1)	(2)	(3)
Precipitation variability	-0.128	0.637	-0.321
	(0.060)**	(0.356)*	(0.127)**
Temperature variability	0.003	0.828	0.632
	(0.307)	(0.441)*	(1.197)
Mean Precipitation	0.022	0.005	0.021
	(0.008)***	(0.009)	(0.007)***
Mean Temperature	-0.042	-0.061	-0.037
	(0.026)	(0.031)*	(0.026)
Land suitability	0.108	0.122	0.087
	(0.112)	(0.118)	(0.101)
Coastal region	0.073	0.062	0.068
	(0.111)	(0.112)	(0.108)
Distance to coast	0.015	0.015	0.010
	(0.058)	(0.060)	(0.062)
Altitude	-0.386	-0.334	-0.393
	(0.135)***	(0.140)**	(0.130)***
On a river	0.265	0.258	0.266
	(0.068)***	(0.068)***	(0.065)***
Area size	-0.123	-0.150	-0.116
	(0.067)*	(0.072)**	(0.068)*
Total population	0.844	0.845	0.843
	(0.034)***	(0.034)***	(0.034)***
City	0.265	0.269	0.269
	(0.065)***	(0.064)***	(0.066)***
Longitude	0.096	0.105	0.088
	(0.058)	(0.055)*	(0.055)
Latitude	-0.078	-0.064	-0.069
	(0.077)	(0.075)	(0.072)
prec.*temp. variability		-0.162	
		(0.081)*	
Precipitation var. Sq.			0.013
			(0.006)**
Temperature var. Sq.			-0.065
			(0.138)
R^2	0.84	0.84	0.84
N	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the log of the total number of migrants in levels instead of shares. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects.

Table A.10: Contemporary migration data, alternative measures of climate variation

Dependent var : Migrants	(1)	(2)	(3)
Precipitation-temperature var.	-0.929		
	(0.521)*		
Mean Precipitation	0.031	0.038	0.036
	(0.042)	(0.055)	(0.051)
Mean Temperature	-0.087	-0.116	-0.124
	(0.128)	(0.139)	(0.157)
Land suitability	0.418	0.350	0.340
	(0.640)	(0.690)	(0.662)
Coastal region	0.390	0.278	0.278
	(0.396)	(0.369)	(0.371)
Distance to coast	0.411	0.280	0.274
	(0.224)*	(0.242)	(0.253)
Altitude	-1.244	-1.773	-1.775
	(0.612)**	(0.723)**	(0.727)**
On a river	1.905	1.860	1.866
	(0.512)***	(0.503)***	(0.500)***
Area size	-1.020	-1.107	-1.105
	(0.473)**	(0.459)**	(0.452)**
Total population	0.180	0.177	0.177
	(0.138)	(0.138)	(0.138)
City	0.908	0.945	0.945
	(0.335)***	(0.334)***	(0.335)***
Longitude	0.335	0.203	0.198
	(0.191)*	(0.224)	(0.197)
Latitude	-0.347	-0.214	-0.213
	(0.262)	(0.220)	(0.215)
Temperature variability, abs. mean dev.		0.758	
		(1.004)	
Precipitation variability, abs. mean dev.		-0.044	
		(0.025)*	
Precipitation-temperature var., abs. mean dev.			-0.887
			(0.406)**
R^2	0.34	0.34	0.34
N	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects. The first column uses a joint precipitation and temperature variability measure calculated with principal component analysis, based on the standard deviation of precipitation and temperature for growing seasons between 1500-1800. The second column shows results with an alternative measure, absolute mean deviation, while the third column uses the joint measure of precipitation and temperature absolute mean deviation constructed by principal component analysis.

Table A.11: Contemporary migration data, fractional logit regressions

Dependent var : Migrants	(1)	(2)	(3)
Precipitation variability	-0.078	0.582	-0.177
	(0.039)**	(0.207)***	(0.088)**
Temperature variability	0.028	0.688	-0.522
	(0.268)	(0.306)**	(1.481)
Mean Precipitation	0.010	-0.002	0.009
	(0.008)	(0.007)	(0.008)
Mean Temperature	-0.024	-0.037	-0.024
	(0.021)	(0.020)*	(0.018)
Land suitability	0.017	0.038	0.015
	(0.116)	(0.122)	(0.110)
Coastal region	0.053	0.043	0.053
	(0.069)	(0.072)	(0.069)
Distance to coast	0.045	0.047	0.049
	(0.039)	(0.041)	(0.039)
Altitude	-0.271	-0.235	-0.277
	(0.114)**	(0.113)**	(0.107)***
On a river	0.193	0.182	0.191
	(0.056)***	(0.056)***	(0.054)***
Area size	-0.133	-0.158	-0.131
	(0.060)**	(0.060)***	(0.058)**
Total population	0.031	0.032	0.031
	(0.023)	(0.023)	(0.023)
City	0.121	0.123	0.124
	(0.046)***	(0.045)***	(0.045)***
Longitude	0.049	0.061	0.050
	(0.040)	(0.039)	(0.038)
Latitude	-0.076	-0.064	-0.068
	(0.049)	(0.046)	(0.045)
prec.*temp. variability		-0.139	
		(0.047)***	
Precipitation var. Sq.			0.007
			(0.004)*
Temperature var. Sq.			0.056
			(0.165)
N	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (i.e. NUTS) level. All specifications include regional fixed effects.

Table A.12: Contemporary migration data, main specification-spatial correlation correction

Dependent var : Migrants	(1)	(2)	(3)
Precipitation variability	-0.496	2.662	-1.001
	(0.145)***	(0.708)***	(0.400)**
Temperature variability	-0.076	3.289	0.036
	(0.525)	(0.706)***	(0.026)
Mean Precipitation	0.066	-0.004	0.059
	(0.027)**	(0.030)	(0.026)**
Mean Temperature	-0.114	-0.194	-0.115
	(0.113)	(0.107)*	(0.113)
Land suitability	0.238	0.310	0.213
	(0.538)	(0.521)	(0.533)
Coastal region	0.325	0.284	0.299
	(0.341)	(0.350)	(0.344)
Distance to coast	0.322	0.323	0.335
	(0.150)**	(0.159)**	(0.151)**
Altitude	-1.476	-1.273	-1.536
	(0.686)**	(0.663)*	(0.683)**
On a river	1.955	1.926	1.942
	(0.390)***	(0.388)***	(0.387)***
Area size	-1.016	-1.129	-1.009
	(0.497)**	(0.491)**	(0.500)**
Total population	0.184	0.187	0.181
	(0.080)**	(0.081)**	(0.081)**
City	0.905	0.923	0.922
	(0.316)***	(0.314)***	(0.313)***
Longitude	0.245	0.284	0.237
	(0.074)***	(0.077)***	(0.075)***
Latitude	-0.389	-0.332	-0.372
	(0.152)**	(0.146)**	(0.155)**
prec.*temp. variability		-0.670	
		(0.146)***	
Precipitation var. Sq.			-2.354
			(2.809)
Temperature var. Sq.			0.239
			(0.312)
R^2	0.35	0.35	0.35
N	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: Standard errors are corrected with the Conley method (with 150km cutoff). The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. All specifications include regional fixed effects.

Table A.13: Contemporary migration data, specifications with different distance cutoffs (Conley errors) and spatial lags

Dependent var : Migrants	(1) 200 cutoff	(2) 500 cutoff	(3) 1000 cutoff	(4) 1500 cutoff	(5) 2000 cutoff	(6) SAR
Precipitation variability	-0.480 (0.148)***	-0.480 (0.108)***	-0.480 (0.070)***	-0.480 (0.057)***	-0.480 (0.050)***	-1.979 (0.735)***
Temperature variability	-0.068 (0.512)	-0.068 (0.287)	-0.068 (0.208)	-0.068 (0.170)	-0.068 (0.147)	2.320 (1.650)
Mean Precipitation	0.064 (0.027)**	0.064 (0.015)***	0.064 (0.010)***	0.064 (0.009)***	0.064 (0.007)***	0.421 (0.178)***
Mean Temperature	-0.117 (0.097)	-0.117 (0.078)	-0.117 (0.057)**	-0.117 (0.046)**	-0.117 (0.040)***	0.222 (0.512)
Land suitability	0.249 (0.496)	0.249 (0.266)	0.249 (0.183)	0.249 (0.149)*	0.249 (0.129)*	-0.299 (0.391)
Coastal region	0.331 (0.326)	0.331 (0.255)	0.331 (0.180)*	0.331 (0.147)**	0.331 (0.127)***	0.317 (0.340)
Distance to coast	0.326 (0.151)**	0.326 (0.076)***	0.326 (0.057)***	0.326 (0.047)***	0.326 (0.040)***	0.077 (0.126)
Altitude	-1.480 (0.646)**	-1.480 (0.661)**	-1.480 (0.481)***	-1.480 (0.393)***	-1.480 (0.340)***	-0.654 (0.469)
On a river	1.959 (0.390)***	1.959 (0.272)***	1.959 (0.190)***	1.959 (0.155)***	1.959 (0.134)***	1.973 (0.441)
Area size	-1.017 (0.538)*	-1.017 (0.512)**	-1.017 (0.365)***	-1.017 (0.298)***	-1.017 (0.258)***	-1.279 (0.232)
Total population	0.184 (0.082)**	0.184 (0.053)***	0.184 (0.036)***	0.184 (0.030)***	0.184 (0.026)***	0.195 (0.026)
City	0.908 (0.313)***	0.908 (0.195)***	0.908 (0.136)***	0.908 (0.111)***	0.908 (0.096)***	0.859 (0.210)
Longitude	0.249 (0.066)***	0.249 (0.047)***	0.249 (0.032)***	0.249 (0.026)***	0.249 (0.023)***	
Latitude	-0.385 (0.148)***	-0.385 (0.080)***	-0.385 (0.059)***	-0.385 (0.048)***	-0.385 (0.041)***	
R^2	0.35	0.35	0.35	0.35	0.35	
N	7,097	7,097	7,097	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: All specifications use the contemporary migration data. The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are corrected with the Conley method. Column (1) uses 200km as the distance cut-off value for the spatial correlation, column (2) 500 km, (3) 1000 km, (4) 1500 km, (5) 2000 km, while column (6) presents results using a spatial autocorrelation model where the climate variables are assumed to be spatially lagged. All specifications include regional fixed effects.

Table A.14: Contemporary migration data, robustness including the independent variables sequentially

Dependent var : Migrants	(1)	(2)	(3)	(4)
Precipitation variability	-0.479	-0.497	-0.496	-0.442
	(0.198)**	(0.202)**	(0.198)**	(0.162)***
Temperature variability	-0.440	-0.218	-0.076	-0.330
	(1.036)	(0.945)	(0.967)	(1.106)
Mean Precipitation	0.062	0.070	0.066	-0.596
	(0.030)**	(0.033)**	(0.038)*	(0.406)
Mean Temperature	0.140	-0.152	-0.114	-0.518
	(0.085)	(0.112)	(0.115)	(0.738)
Longitude	0.274	0.264	0.245	0.245
	(0.162)*	(0.170)	(0.179)	(0.175)
Latitude	-0.350	-0.469	-0.389	-0.274
	(0.261)	(0.282)	(0.277)	(0.244)
Land suitability		0.337	0.238	0.231
		(0.691)	(0.710)	(0.748)
Coastal region		0.407	0.325	0.301
		(0.391)	(0.392)	(0.362)
Distance to coast		0.306	0.322	0.337
		(0.222)	(0.223)	(0.201)*
Altitude		-2.105	-1.476	-1.486
		(0.531)***	(0.582)**	(0.579)**
On a river		2.135	1.955	1.861
		(0.558)***	(0.512)***	(0.499)***
Area size			-1.016	-0.994
			(0.471)**	(0.485)**
Total population			0.184	0.182
			(0.138)	(0.137)
City			0.905	0.937
			(0.337)***	(0.317)***
$MeanPrecipitation^2$				0.020
				(0.013)
$MeanPrecipitation^3$				-0.000
				(0.000)
$MeanTemperature^2$				0.063
				(0.074)
$MeanTemperature^3$				-0.002
				(0.002)
R^2	0.33	0.34	0.35	0.35
N	7,756	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the contemporary migration data. The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Standard errors are clustered at regional (NUTS) level. All specifications include regional fixed effects.

Table A.15: Contemporary migration data, robustness using climate data from different centuries

Dependent var : # of Migrants	1500-1600	1600-1700	1700-1800
Precipitation variability	-0.041	-0.034	-0.044
	(0.021)*	(0.014)**	(0.018)**
Temperature variability	0.078	-0.009	-0.063
	(0.950)	(1.023)	(0.918)
Mean precipitation	0.003	0.004	0.006
	(0.005)	(0.003)	(0.004)
Mean Temperature	-0.139	-0.118	-0.127
	(0.122)	(0.116)	(0.117)
Land suitability	0.295	0.181	0.337
	(0.679)	(0.709)	(0.712)
Coastal region	0.374	0.377	0.287
	(0.392)	(0.385)	(0.387)
Distance to coast	0.318	0.364	0.313
	(0.228)	(0.217)*	(0.227)
Altitude	-1.526	-1.438	-1.579
	(0.573)**	(0.588)**	(0.594)**
On a river	1.884	1.967	1.937
	(0.507)***	(0.511)***	(0.510)***
Area size	-1.078	-1.025	-0.999
	(0.467)**	(0.477)**	(0.475)**
Total population	0.183	0.184	0.183
	(0.138)	(0.138)	(0.138)
City	0.900	0.925	0.903
	(0.337)**	(0.336)***	(0.340)**
Longitude	0.253	0.254	0.258
	(0.178)	(0.188)	(0.192)
Latitude	-0.367	-0.331	-0.347
	(0.301)	(0.269)	(0.246)
R^2	0.34	0.34	0.34
N	7,097	7,097	7,097
Regional FE	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the contemporary migration data. The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1600 in the first column, 1600-1700 in the second column, and 1700-1800 in the last column. Standard errors are clustered at regional (NUTS) level. All specifications include regional fixed effects.

Table A.16: Mechanisms/channels as outcomes

Dependent var :	Past migrants	Conflict	Distance to trade routes	GDP per Capita	Population Density	Trust in parliament	Attitudes towards migrants
Precipitation var.	0.105	0.137	0.010	-0.022	-0.001	0.043	0.068
	(0.151)	(0.060)**	(0.015)	(0.002)***	(0.013)	(0.004)***	(0.004)***
Temperature var.	-2.216	0.105	-0.065	-0.053	-0.103	0.248	0.024
	(0.957)**	(0.389)	(0.051)	(0.013)***	(0.056)*	(0.018)***	(0.020)
Mean Precipitation	-0.049	-0.046	-0.011	0.007	-0.005	-0.005	-0.002
	(0.035)	(0.014)***	(0.004)***	(0.001)***	(0.003)*	(0.001)***	(0.002)
Mean Temperature	0.040	-0.073	0.124	0.024	-0.007	0.001	0.036
	(0.160)	(0.068)	(0.020)***	(0.003)***	(0.008)	(0.007)	(0.008)***
Land suitability	-1.061	1.058	-0.867	0.116	0.000	0.004	0.077
	(0.451)**	(0.340)***	(0.063)***	(0.010)***	(0.048)	(0.019)	(0.025)***
Coastal region	0.723	-0.041	-0.269	-0.024	0.125	-0.064	-0.078
	(0.362)**	(0.206)	(0.062)***	(0.011)**	(0.160)	(0.020)***	(0.022)***
Distance to coast	-0.095	0.078	-0.396	-0.002	0.029	-0.007	-0.019
	(0.114)	(0.101)	(0.021)***	(0.003)	(0.022)	(0.005)	(0.005)***
Altitude	-3.012	-1.258	0.606	0.091	-0.076	-0.003	0.082
	(1.182)**	(0.620)**	(0.103)***	(0.020)***	(0.068)	(0.037)	(0.047)*
On a river	0.026	0.681	0.302	0.127	0.034	0.186	0.112
	(0.188)	(0.315)**	(0.094)***	(0.015)***	(0.075)	(0.018)***	(0.021)***
Area size	-0.346	0.017	0.119	0.015	-0.074	0.020	0.091
	(0.260)	(0.125)	(0.049)**	(0.006)**	(0.039)*	(0.014)	(0.027)***
City	1.387	0.896	-0.194	0.031	0.169	0.045	0.026
	(0.172)***	(0.127)***	(0.043)***	(0.006)***	(0.070)**	(0.013)***	(0.016)
Longitude	0.194	-0.009	0.026	0.015	-0.016	-0.070	-0.044
	(0.144)	(0.068)	(0.008)***	(0.002)***	(0.009)*	(0.003)***	(0.003)***
Latitude	-0.270	0.050	-0.060	0.030	0.014	-0.017	0.045
	(0.154)*	(0.079)	(0.016)***	(0.002)***	(0.018)	(0.004)***	(0.006)***
Cropland 18th century	-0.003			-0.000	-0.002		
	(0.004)			(0.000)***	(0.001)*		
R^2	0.40		0.39	0.79	0.30	0.80	0.60
N	5,058	7,410	7,369	6,595	5,058	7,376	7,376
Region FE	Yes	No	No	No	Yes	No	No
Country FE	No	Yes	Yes	Yes	No	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: The dependent variable is the share of migrants in the total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Past migrants is the number of international migrants over the one-hundred-year period 1700-1800 using the geneological dataset (results are robust to using the number of international migrants over the one-hundred-year period 1800-1900). Conflict is a dummy for locations with historical conflicts. Distance to trade routes indicates the physical distance to the closest historical trade route. GDP per capita and population density are for 1800. Trust in parliament and attitudes towards migrants is based on current survey data. Cropland indicates the % of total cell being cropland in 1870 from Paprotny and Mengel (2023), the earliest available data. Specifications where the mechanism variable is available at the local level include regional fixed effects, for those where the mechanism variable is available only at the regional level, country fixed effects are included (in which case robust standard errors are shown).

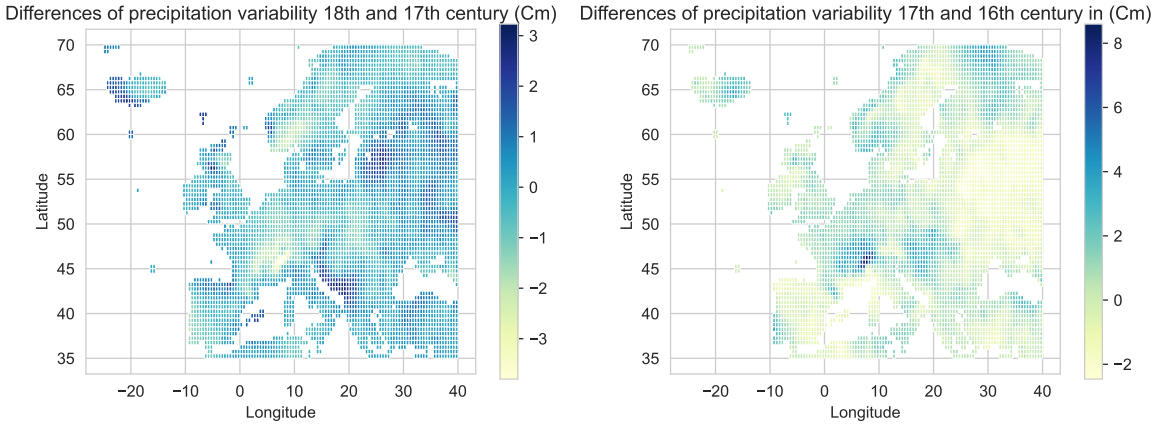
Table A.17: Testing potential mechanisms- Conley

Dep. var: Migrants	Past migrants	Conflict	Trade	GDPperCap	Pop.Density	Attitudes	Institutions
Precipitation var.	-0.228	-0.484	-0.307	-0.177	-0.223	-0.558	-0.473
	(0.158)	(0.140)***	(0.139)**	(0.153)	(0.157)	(0.179)***	(0.150)***
Temperature var.	-0.125	-0.083	-0.203	0.197	-0.077	-0.195	-1.095
	(0.757)	(0.528)	(0.710)	(0.733)	(0.756)	(0.629)	(0.665)*
Mechanism var.	0.008	0.510	-0.439	5.381	0.482	3.411	3.612
	(0.003)***	(0.250)**	(0.345)	(2.098)**	(0.146)***	(1.271)***	(0.848)***
R^2	0.38	0.35	0.18	0.17	0.38	0.21	0.21
N	5,337	7,097	7,056	6,816	5,337	7,063	7,063
Regional FE	Yes	Yes	No	No	Yes	No	No
Country FE	No	No	Yes	Yes	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

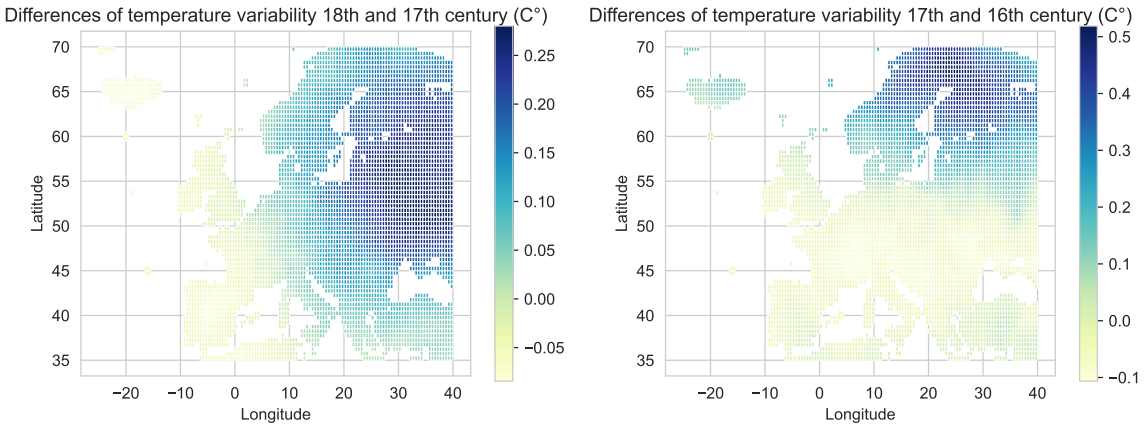
Note: Standard errors are corrected with the Conley method with a 150 km cutoff. The dependent variable is the share of migrants in total population in a given location. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons between 1500-1800. Specifications where the mechanism variable is available at local level include regional fixed effects, for those where the mechanism variable is available only at regional level country fixed effects are included. The full list of control variables (as in Table 2) are included, but not presented in the table.

Figure A.1: Genealogical data: Differences in precipitation variability over centuries



Note: The figure uses the genealogical data. The precipitation variability is computed using the standard deviation of the precipitation in the growing season during the century. For each cell, the figure plots the difference in precipitation variability between the 18th and 17th centuries, and between the 17th and 16th centuries.

Figure A.2: Genealogical data: Differences in temperature variability over 18th, 17th and 16th centuries



Note: The figure uses the genealogical data. The temperature variability is computed using the standard deviation of the temperature in the growing season during the century.For each cell, the figure plots the difference in temperature variability between the 18th and 17th centuries, and between the 17th and 16th centuries.

Supplementary Annex - Additional data description

The final dataset combines information from several datasets, described below. Table A.18 lists key sources and Figure A.3 summarizes the geography of the data.

Table A.18: An overview of the data used.

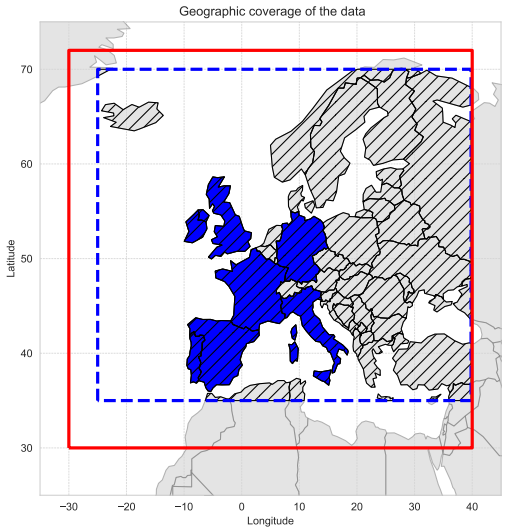
Source	Description	Format
Pauling et al. (2006)	Precipitation, 1500–2000	0.5 degree grid
Luterbacher et al. (2004)	Temperature, 1500–2000	0.5 degree grid
Kaplanis et al. (2018)	Genealogical information, 1500–1900	point
European Commission	Current migration, 2011	100 meter x 100 meter grid
Henderson et al. (2017)	Geographic measures, nightlights	0.25 degree grid
Dincecco and Onorato (2018)	Historical battle indicator	point
Bosker et al. (2008)	Historical city indicator	point
GAEZ 2012 database of FAO	Potato suitability	0.5 degree grid
Buggle and Durante (2021)	Pre-industrial GDP per capita	NUTS level
European Social Surveys	Attitudes and trust	NUTS level

A.1 Genealogical data - historical migration data

This section assesses whether selection and coverage issues in the genealogical data could bias the estimated relationship between historical climate risk and migration. We examine (i) whether recorded migration flows are mechanically related to local population size or development, (ii) whether results depend on regions with particularly high or low data coverage, and (iii) whether findings are sensitive to early periods with lower data quality. Across these checks, we find no evidence that data limitations drive our results.

Information on births and deaths of individuals is taken from Kaplanis et al. (2018), who compile the data from genealogical records. More specifically, Kaplanis et al. (2018) extracted data initially for 86 million individuals from the geni.com website. From this,

Figure A.3: Geographic coverage of the data



Note: the red (solid line) rectangle denotes the area for which precipitation data is available, the blue (dashed line) rectangle covers the area for which temperature data is available, the countries shaded in blue (dark) color have high-resolution migrant data. The striped areas on the map represent the geographical coverage kept for the historical migration.

about 16 million individuals were placed into longitude and latitude coordinates. The individuals in the dataset are distributed across the world, with about half of the geo-located individuals being based in Europe (about 55%), as shown in figure A.4. The authors undertake a number of validation exercises showing that the data has similar quality to traditionally collected data, but can provide a significantly larger scale. The data reflects historical events and trends, for example that death rates are higher during military conflicts. Comparing the average life span from the genealogical data (over the years 1840 to 2000) to the Human Mortality Database leads to a 98% concordance. To check the robustness of the data against potential socioeconomic biases, the data was compared to the Vermont Department of Health certificates for 80,000 cases over the period between 1985 and 2010. This showed a higher than 98% concordance, indicating a high level of consistency against socioeconomic ascertainment (see Kaplanis et al. (2018)).

We use the records of individual where both their birth and death took place during the 1500 - 1800 time period in the European region of interest (Fig A.3 shows the geographical coverage). This produces a sample with information on 576,000 individuals who were born and died within the region of interest and have information on birth/death year and locations. Using the observations with complete information, we restrict our analysis to individuals who died in a different country compared to the one they were born in, who we define as international migrants. Since we do not have the actual year of migration, we use 100-year periods over which we calculate the number international migrants (i.e. individuals born and dead in different countries over the period).

Table A.19 provides information on our sample, including the full sample, and the reduced sample used by the ppml estimator. The reduction of observations in the latter sample is due to the fact that perfectly predicted / singleton observations are dropped using the ppml estimator. Referring to column 1, the total number of individuals in

Figure A.4: Geographic distribution of individuals' location data



Source: Kaplanis et al. (2018) Figure 2D

the data with information on the location of their death is 576,005, while the number of origin-destination cell pairs with international migrants is 10,875. The number of international migrants is 27, 463.

Table A.19: Number of observation

Observation	**Full sample**	**Without singleton observations**
Number of Individuals	576,005	576,005
Number of cell-pairs with international migration	10,875	9,018
Died in a Different Country	27,463	23,608

The record for each individual's birth and death also comes with the respective longitude and latitude information. We match these locations to the climate data cells at 0.5 degrees. Figure A.5 shows the spatial distribution of the births and deaths of international migrants by the 0.5° cells used in the analysis.

We undertake a number of validation exercises. First, we test for differential attrition in the genealogical data. Using highly disaggregated data from Paprotny and Mengel (2023) on the total population in Europe in 1870 (the earliest year available), for each cell in the genealogical data, we calculate the rate of the genealogy population relative to the Paprotny and Mengel (2023) data, and correlate this variable with the different climate variables. These results are summarized graphically below in Figure A.6. In addition, we also correlate the genealogy/Paprotny Paprotny and Mengel (2023) population ratio with data capturing GDP per capita calculated from Paprotny and Mengel (2023) data to test for attrition based on differential levels of development across different locations. This is shown in Figure A.7.

We conclude that there is little evidence that climate variations or level of development affects the fraction of individuals recorded in the genealogical data.

Next, we run the main genealogical specification excluding all years prior to 1700, since data after 1700 is more reliable (Blanc, 2024b) Specifically, we restrict our sample to the period between 1700 to 1850. In order to be able to use the same set of fixed

Figure A.5: Log of international migrants per cell

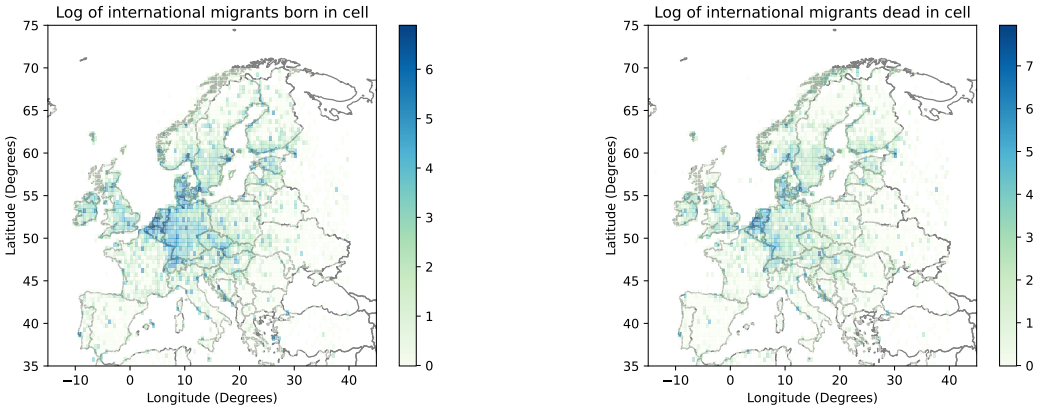
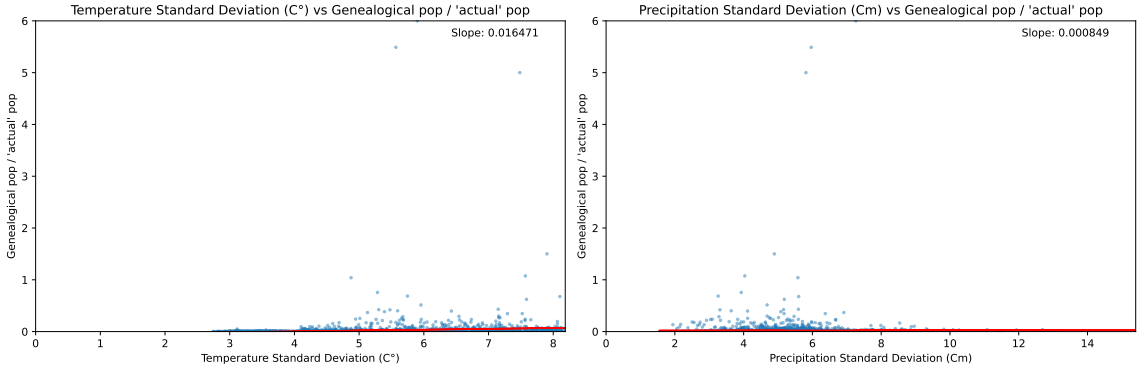


Figure A.6: Correlation between climate variability and genealogical population divided by total population from Paprotny and Mengel (2023)



Note: The figure shows the correlation between climate variability and genealogical population divided by total population from Paprotny and Mengel (2023). The figure uses the genealogical data, along with total population in each cell from Paprotny and Mengel (2023).

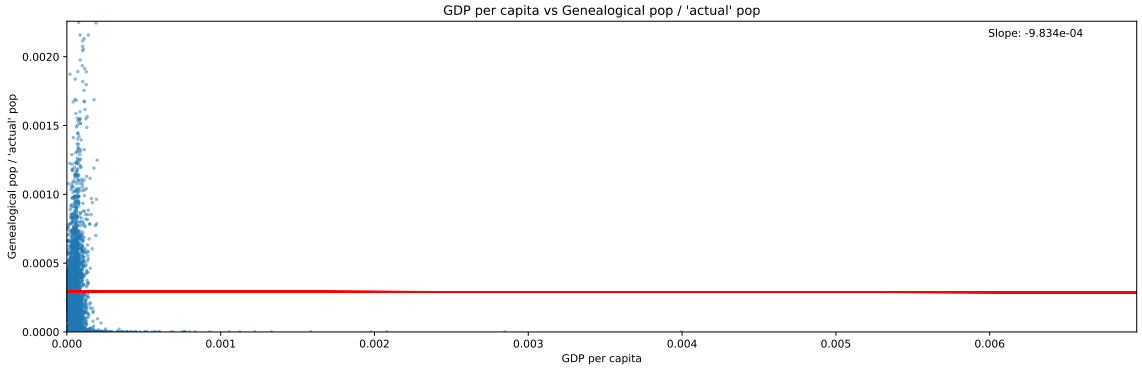
effects as before, instead of using 100-year periods, we use now 50-year periods. Results are shown in column 5 of Table A.4. Again, we find a significant negative impact of climate variation on bilateral migration flows.

Finally, we also rerun our main specification, splitting the sample into two. The sample in the first column in Table A.20 is restricted to the Benelux countries, which is the region with best data availability (as can be seen on Figure A.4), while the second column excludes those countries. Both results are qualitatively the same, again confirming our previous findings. While the coefficient on precipitation variability in column 1 is marginally significant, this is most likely driven by the significantly reduced number of observations in this restricted sample.

A.2 European Commission data on population

Data from the European Commission (EC) contains high-quality census-based information on country of origin (birth or citizenship) of population at 100m by 100m resolution, with accompanying latitude and longitude coordinates. Alessandrini et al. (2017) describe the construction of the dataset. The definition of country of origin varies between the countries in the sample, but in our work we use 'nationals' to describe all persons

Figure A.7: Correlation between GDP per capita and genealogical population divided by total population from Paprotny and Mengel (2023)



Note: The figure shows the correlation between GDP per capita (calculated from Paprotny and Mengel (2023) and genealogical population divided by total population from Paprotny and Mengel (2023).

Table A.20: Restricting sample to only Benelux and without Benelux countries

	(2) Only Benelux	(2) Without Benelux
Precipitation variability	-1.378 (0.662)**	-0.099 (0.040)**
Temperature variability	-2.134 (12.336)	-2.486 (0.664)***
Mean precipitation	0.391 (0.144)***	0.074 (0.034)**
Log of destination pop	1.053 (0.142)***	0.592 (0.061)***
Mean temperature		-2.192 (1.213)*
N	1,945	17,362
FE	Yr x orig, pair	Yr x orig, pair

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: All specifications use the genealogical data. The dependent variable is the number of migrants from an origin cell to a destination cell over hundred-year periods. Precipitation and temperature variability are measured by the standard deviation of precipitation and temperature during growing seasons with a lag of 25 years. Standard errors are clustered at the level of destination cells.

recorded as having country of origin that is the same as country of destination, while persons whose country of origin differs from the country of residence (as recorded in EC data) will be referred to as 'migrants'.

A.3 Climate data

The two historical climate datasets used in this paper are Pauling et al. (2006) for precipitation and Luterbacher et al. (2004) for temperature. The datasets were downloaded from `https://crudata.uea.ac.uk/cru/projects/soap/data/recon/`. Pauling et al. (2006) reconstruct historical precipitation data using various proxies (tree-ring chronologies, ice cores, etc.) and a range of regression techniques. They are able to reconstruct the data over a long time span, from 1500 to 1983, with distinction for four seasons (spring, summer, fall, winter). The data is gridded with coordinates of a centre of a 0.5 x 0.5 degrees grid for the latitude range from 30.25N to 70.75N and longitude range from 29.75W to 39.75E.

Luterbacher et al. (2004); Xoplaki et al. (2005) develop a proxy-based spatial dataset of historical temperature. The dataset also covers a long time span, from 1500 to 2002 (with the caveat that data after 1901 is based on observational estimates, while all other values are reconstructed). The grid dimensions are the same as those in the precipitation data, but the grid ranges from 35.25N to 69.75N (latitude) and from 24.75W to 39.75E (longitude).

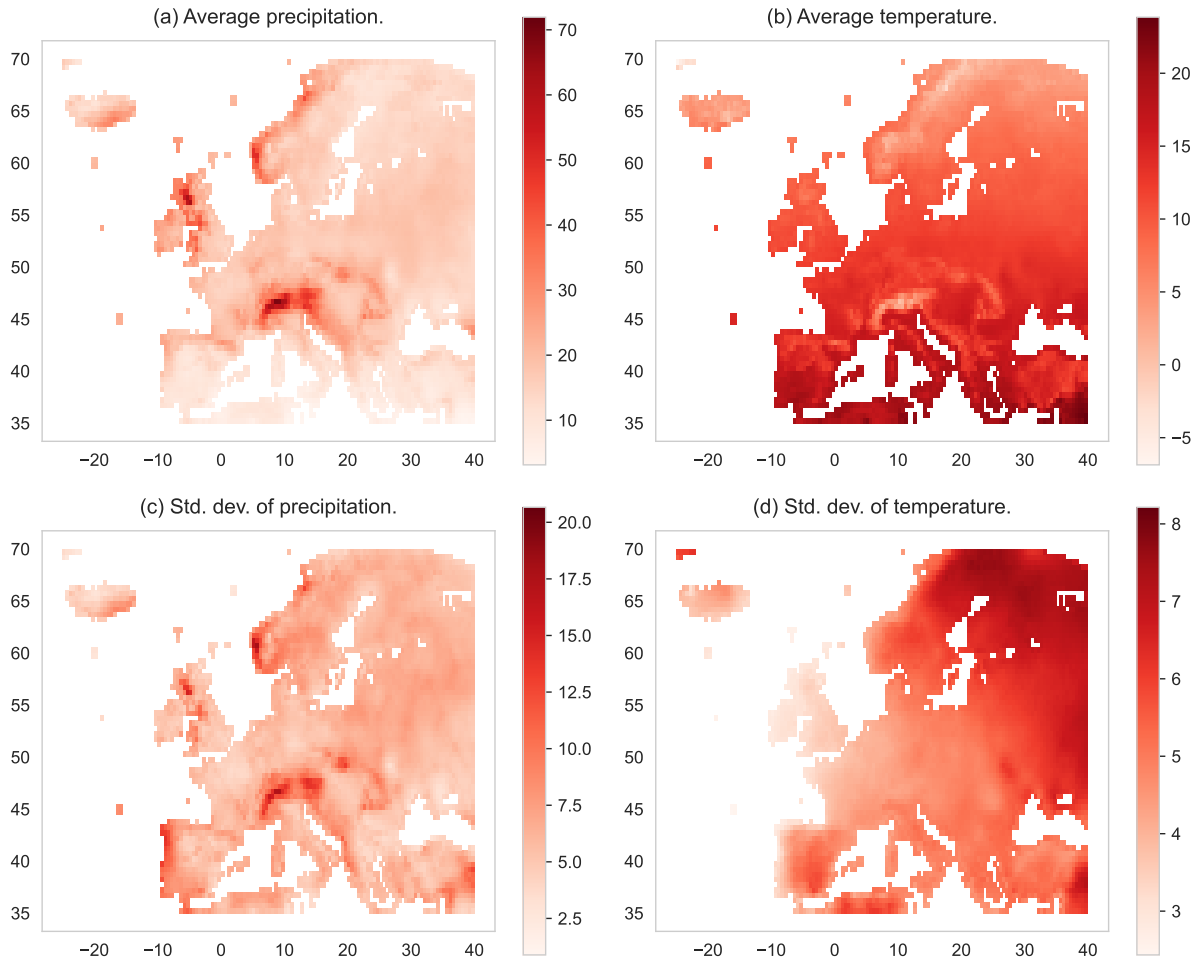
In line with Buggle and Durante (2021), we take the values of temperature and precipitation in summer and autumn of each year from 1500 to 1800 and calculate the mean and the standard deviation of temperature and precipitation. Figure A.8 shows the values of these variables.

A.4 Merging the EC and climate data

Given the differences in resolution between the EC population data, the geographic measures data (obtained from Henderson et al. (2017)) and the climate data, we chose the aggregation which was the most disaggregated level among the control variables. In other words, we aggregated the EC data (initially 100 meters x 100 meters) to the aggregation level of the geographic measures data at 0.25 degrees. This was done by assigning each of the EC population cells to a corresponding cell based on whether the EC cell centroid is within the boundaries (boundaries included) of a geographic measures data cell, summing up the population numbers to the corresponding aggregations.In order to take account of borders, a grid cell on a border of two or more countries (e.g. Germany and France) is represented by two entries in the EC data: one for population within the German part of the grid cell and one for the French part of the grid cell. Next, keeping the aggregation at 0.25 we merged into the data the climate cells (which were at 0.5 degrees) using the centroids of the cells.

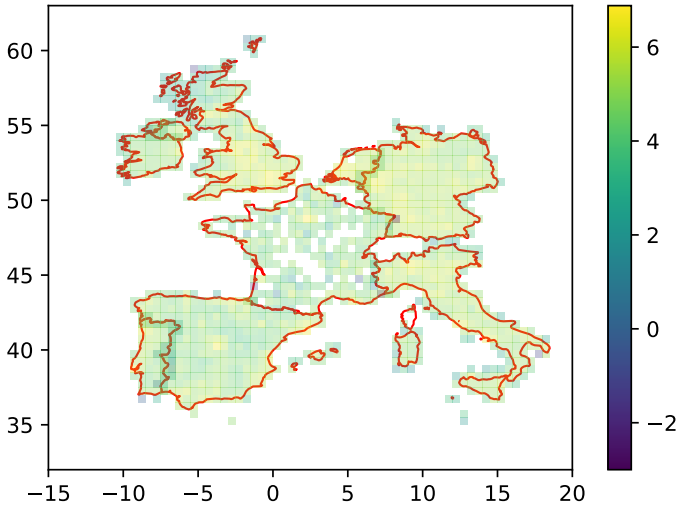
Figure A.9 shows the spatial distribution of total population across the grid. Total population includes population of all categories recorded in the EC data (own-country nationals, EU nationals, non-EU nationals and others). Figure A.10 shows the share of migrants in a given cell. Note that in the case of France, there are missing data in several cells.

Figure A.8: Climate variables for the growing seasons of 1500–1800.



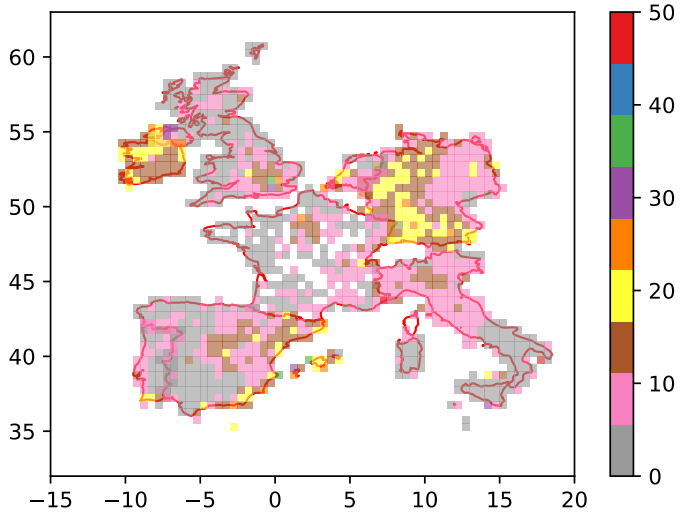
Note: Source: authors' calculations. Precipitation is measured in cm, and temperature is measured in Celsius.

Figure A.9: Total population per cell (log)



Note: Population was aggregated to 0.25 degree resolution. The displayed colors correspond to log_{10} of the population, numbers below 0 imply less than 1 person which is due to the disaggregation procedure used by EC, see Alessandrini et al. (2017).

Figure A.10: Share of migrants per cell



Note: The map shows the share of migrants in the total population.

A.5 City data

The dataset originates from Bosker et al. (2013), The data indicates the name and location of cities with more than ten thousand inhabitants that existed between 800 and 1800 covering Europe, the Middle East, and North Africa. We created a dummy variable which takes the value of 1 if at least one city with more than ten thousand inhabitants was located in the cell during the time period considered.

A.6 Conflict data

The dataset originates from Dincecco and Onorato (2018) and gives us the locations of different types of conflicts that took place in Europe from 1000 to 1900. These conflicts are divided into different types and durations. The conflict dummy variable takes the value of 1 if at least one conflict took place in the cell, regardless of its type and duration.