WORKING PAPERS

N° 1728

March 2026

"Accounting for the full distribution of temperature to predict international migration"

Evangelina Dardati, Thibault Laurent, Paula Margaretic,
Ean Paredes and Christine Thomas-Agnan

Toulouse School of Economics

Accounting for the full distribution of temperature to predict international migration

Evangelina Dardati* Thibault Laurent† Paula Margaretic‡ Ean Paredes§
Christine Thomas-Agnan¶

Abstract

This paper evaluates the role of climate variables in predicting international migration by proposing two alternative modeling approaches: scalar-on-composition and scalar-on-density regressions. We compare them with the standard scalar-on-scalar approach. Although most studies rely on annual averages of daily temperatures, focusing solely on central measures can mask essential details, such as nonlinearities and threshold effects. Using the full temperature distribution, either by binning or smoothing, the proposed models achieve improved predictive performance out-of-sample. These gains highlight the importance of properly handling the compositional nature of daily temperature bin data to avoid misleading interpretation of the estimates and flawed inferences. Finally,

*Center for Applied Economics, Department of Industrial Engineering, University of Chile, Beauchef 851, Santiago, Chile, email: evangelina.dardati@uchile.cl

†Toulouse School of Economics, 1 Esp. de l'Université, 31000 Toulouse, France. Email: thibault.laurent@tse-fr.eu.

‡Universidad de Chile, Departamento de Administración, Facultad de Economía y Negocios, Av. Diagonal Paraguay 257, Santiago, Chile. Email: pmargareti@fen.uchile.cl.

§Universidad de Chile, Departamento de Administración, Facultad de Economía y Negocios, Av. Diagonal Paraguay 257, Santiago, Chile. Email: eanparedesbyrt@gmail.com.

¶Toulouse School of Economics, 1 Esp. de l'Université, 31000 Toulouse, France. Email: christine.thomas@tse-fr.eu.

‖We acknowledge financial support from the Belmont Forum and from CRA-MGC-2 NETWORK, IAI.This research was funded with support provided by the Inter-American Institute for Global Change Research (grant MGC 2023-2), which is supported by the US National Science Foundation (Award 2025226), and by the French National Research Agency (ANR) under grants ANR-17-EURE-0010 (Investissements d'Avenir program) and ANR-22-MIGR-0002. Paula Margaretic acknowledges financial support from ANID Fondecyt project 1240312. All remaining errors are ours.

we demonstrate how incorporating complete temperature distributions into alternative climate scenarios can substantially affect projected outmigration.

Keywords: compositional data, temperature, migration projections, climate change.

1 Introduction

Climate change (CC) ranks among the greatest threats to humanity in modern history (Intergovernmental Panel on Climate Change (IPCC), 2014, 2007). Climatologic evidence shows that climate change raises global temperatures, increases sea levels (Stocker, 2014), and intensifies the frequency and severity of natural disasters (Stott, 2016). CC undermines human well-being through multiple channels, including reductions in labor productivity, threats to food security, and constraints on access to water and natural resources (Dell et al., 2014; Carleton and Hsiang, 2016). However, these damages do not distribute evenly across time or space. By imposing heterogeneous effects across countries and regions, CC is likely to affect both voluntary migration and forced displacement in a geographically unequal manner throughout the 21st century.[1] As projections of CC become increasingly alarming, including those of global warming, questions on how many people will migrate, and where they will move from and to, have become more urgent and a source of controversy, drawing intense attention from both researchers and policymakers (Conte et al., 2021; Boas et al., 2019; McLeman, 2019).

Despite this growing interest, scholars still lack a clear understanding of the long-term effects of CC on international migration (Hoegh-Guldberg et al., 2018). This gap limits efforts to anticipate migratory pressures and to design effective interventions for vulnerable populations, especially those at risk of being trapped in extreme poverty. In this paper, we focus on rising temperature as a key manifestation of CC and propose an alternative empirical procedure to predict international migration induced by climate-related changes. Our approach treats the annual distribution of daily temperatures worldwide as compositional data, where each bin relative frequency represents a share of a fixed total (e.g., 365 days per year). This contrasts with the existing literature that relies on aggregated temperature

[1] Local socio-economic conditions—such as individuals' capacity and willingness to migrate, the strength of social networks, and the responsiveness of political institutions—shape these responses.

measures to capture climate effects (Beine and Parsons, 2015; Bosetti et al., 2021; Cai et al., 2016).[2] Because increasing frequency in one bin inherently reduces it in another, the histogram exhibits a unit-sum property. Compositional data analysis addresses this issue by applying transformations—such as logarithmic ratio transforms—that preserve the relative structure of the data and mitigate spurious correlations and biased inference (Hron et al., 2012).[3] Our methodology integrates the compositional structure of temperature data into regression models to predict international outmigration at country level worldwide, enabling us to improve prediction accuracy by detecting potentially non-linear relationships (Carter et al., 2018).

This paper makes three main contributions. First, it examines the relationship between CC and international migration by treating temperature data explicitly as compositional and by applying compositional data analysis methods that respect this structure. It also quantifies the resulting gains in predictive performance. Second, it produces country-level outmigration projections that reflect changes in the full temperature distribution under a range of climate change scenarios. Third, it reports point estimates of the impact of global warming on international migration and constructs confidence intervals around these estimates. We interpret these intervals as measures of model uncertainty.

Our empirical analysis relies on the international migration database compiled by the Organization for Economic Cooperation and Development (OECD). The database records annual bilateral migration flows from 155 origin countries to 35 OECD destination countries over 2000–2019. We construct the dependent variable as the annual outmigration rate from each origin country by summing flows across all OECD destination countries and scaling by

[2]These studies typically summarize daily temperature data (e.g., minimum, maximum, mean) into yearly or decade averages.

[3]In contrast, regression approaches that include temperature histograms typically create temperature bins, include each bin in the model, and omit one to avoid multicollinearity (Mullins and Bharadwaj, 2021; Deschênes and Greenstone, 2011), treating the bins as independent variables and ignoring the compositional nature of the histogram.

the origin-country population. Accordingly, the unit of observation is the origin-country–year, and our estimates and forecasts refer to total outmigration from each origin country to the set of OECD destinations.

To evaluate the predictive benefits of accounting for the compositional nature of origin-country temperature data, we compare three modeling approaches. First, we estimate a regression that constructs one-degree temperature bins, counts the number of days falling within each bin, and applies the transformations described in Thomas-Agnan et al. (2025). We refer to this approach as scalar-on-composition regression. Second, we fit linear combinations of B-splines to daily temperature histograms and then estimate a regression that uses the resulting density as a functional and distributional covariate (Talská et al., 2021). We refer to this approach as scalar-on-density regression. Third, we estimate a benchmark regression that aggregates daily temperature data into annual origin-country averages. This specification matches common practice in the literature, and we refer to it as scalar-on-scalar regression.

Each specification includes climate-category indicators and interactions between maximum temperature and climate categories, so the temperature–outmigration relationship varies across climate zones and captures heterogeneity in adaptation. We also control for origin GDP per capita and year effects, since income and liquidity constraints shape the ability to finance migration and year effects absorb common shocks.[4] Finally, we project international outmigration by origin country in 2100 under alternative global warming scenarios, using the Shared Socioeconomic Pathways developed in the climate modeling literature.

Our results show that incorporating the full temperature distribution substantially im-

[4]We use climate-category fixed effects rather than country fixed effects because the analysis targets prediction from cross-country differences in climate distributions and long-run distributional shifts under warming scenarios; country fixed effects would instead force identification to rely on short-run within-country deviations around each country's baseline climate. Because the dependent variable aggregates flows over destinations, we interpret predictions as total outmigration from each origin country to OECD destinations, not as bilateral destination choice.

proves out-of-sample prediction accuracy. Across performance metrics, both the scalar-on-density and the scalar-on-composition models perform best compared to the scalar-on-scalar. For example, the scalar-on-composition reduces the root mean squared error by 19.9% compared to our scalar-on-scalar specification, which follows standard specifications that rely most often on annual average temperature. These findings highlight the value of modeling temperature as a distribution rather than collapsing it into a single scalar measure.

Finally, we assess how projected climate change influences international outmigration in 2100 by comparing regional projections under different model specifications. We find substantial variation across regions of the world. Under the Shared Socioeconomic Pathways (SSP) 2.6, the distribution-based approaches imply larger increases in outmigration from Africa and Asia and sharper declines from Europe and the Americas than the scalar-on-scalar approach, relative to 2019. For example, Europe rises from 2.58 million in 2019 to 3.08 million in 2100 under scalar-on-scalar, but it falls to 2.34 million under scalar-on-composition. Asia falls from 2.79 million to 2.51 million under scalar-on-scalar, but it rises to 3.79 million under scalar-on-composition. The same ordering holds for outmigration rates.

This contrast reflects what each model treats as the relevant climate shock. The scalar-on-composition model uses the full binned distribution of daily maximum temperatures, so it responds to where within the distribution climate change adds or removes days. The scalar-on-scalar model instead summarizes temperature with a single annual average, so it smooths over those within-year reallocations. This distinction matters because migration incentives often depend on thresholds and tail risk. A small change in the annual mean can hide a large change in the number of very hot days or very cold days.

In Europe and the Americas, the scalar-on-composition projection turns lower because warming in many countries in these regions relaxes cold constraints more than it increases damaging heat exposure. The distribution shifts away from the cold tail and toward milder

temperature ranges. That shift can raise expected productivity and reduce climate-related risk in origin areas, which weakens the incentive to leave. The scalar-on-scalar approach cannot separate this "cold relief" channel from other changes because it collapses the full distribution into an average. The scalar-on-composition approach keeps the full set of distributional movements, so it can translate a shift toward milder days into lower predicted outmigration.

In Africa and Asia, the scalar-on-composition projection turns higher because projected climate change shows up as a stronger reallocation toward the hot tail, even under SSP-2.6. Additional very hot days can reduce agricultural yields, lower labor productivity, and increase health risk, especially where baseline temperatures already sit near physiological and agronomic thresholds. These damages raise the incentive to move. The annual-mean approach can understate this mechanism because it averages tail changes into a single number. The distribution-based approaches treat those tail movements as first-order inputs and therefore generate larger projected increases in outmigration from these regions. Taken together, these projection differences show that treating temperature as a distribution rather than an annual average can materially change both the magnitude and the regional pattern of climate-induced outmigration

The paper proceeds as follows. We first review the relevant literature on climate and migration. Section 2 describes the data, while Section 3 outlines the data processing steps and the modeling framework. Section 4 presents the main empirical results. Section 5 examines projected outmigration under different climate scenarios and compares our findings with those from the existing literature. Finally, Section 7 concludes.

Literature review

This paper contributes to the literature that examines the relationship between climate change and migration.[5] A set of papers quantifies the economic consequences of climate change and explores its long-term implications for migration and related outcomes (e.g., Desmet et al., 2018; Desmet and Rossi-Hansberg, 2015; Conte et al., 2021; Burzyński et al., 2022). These studies often model countries as spatially disaggregated units with heterogeneous sectoral structures to assess the dynamic economic effects of climate exposure. Unlike approaches grounded in structural modeling, our strategy is data-intensive and model-agnostic, designed to maximize predictive accuracy.

Our work aligns most closely with recent studies that offer forward-looking projections of international migration in response to climate change (e.g., Cattaneo et al., 2024; Missirian and Schlenker, 2017; Burzyński et al., 2022; Caballero Reina et al., 2024). We contribute to this literature by introducing a novel empirical method to improve migration forecasting. Drawing on data from more than 150 origin countries, our primary contribution lies in predicting international outmigration and in demonstrating how accounting for the compositional nature of temperature distributions improves the accuracy of these predictions. Our goal is not to identify causal effects of climate on migration, but to enhance predictive performance. Following the standard approach in this literature, we incorporate exogenous climate counterfactual scenarios to estimate future migration patterns. We provide confidence intervals for our projected migration impacts, offering a practical measure of model uncertainty—a critical input for both researchers and policymakers grappling with the long-run implications

[5] Within this strand of literature, there is no consensus on the existence or direction of association between climate change and migration: climatic shocks may induce, constraint or have no impact on migration depending on the particular characteristics of the climatic shock and the region of occurrence. These mixed results persist regardless of the data type—whether derived from cross-national household surveys, micro-censuses, or macro-level analyses. In particular, among the macro level studies of bilateral international migration, some find a positive relationship between rising temperatures and international migration (Backhaus et al., 2015; Cai et al., 2016; Cattaneo and Peri, 2016), while others find no statistically significant association (Beine and Parsons, 2015).

of climate change.[6]

2 Data

We use OECD migration flow data from the International Migration Database[7], which includes 155 origin countries and 35 OECD destination countries, covering the period from 2000 to 2019 on an annual basis. Our dependent variable is the log of the outmigration rate (OMR) for each origin country, which we compute as follows:

$$\log(OMR)_{ct} = \log\left(\frac{\sum_{d \in D_c} Flow_{cdt}}{Pop_{ct}} \times 100\right) \tag{1}$$

where $Flow_{cdt}$ represents the bilateral migration flow from origin country c to destination country d in year t, D_c is the total number of destination countries, and Pop_{ct} denotes the population of country c in year t.[8]

For temperatures, we use daily gridded maximum and minimum temperature data from MERRA-2 ($0.5^\circ \times 0.625^\circ$ resolution), for the period 1980 to 2020. For each grid cell, we construct one-degree temperature bins and count the number of days per year that fall within each bin. For example, the 3°C bin for maximum temperature includes all days when the maximum daily temperature lies between 3°C and 4°C.

Formally, let $s \in S = \{\max, \min\}$ denote the type of temperature variable. Define the temperature in grid cell g on day d as $\tau^s_{g,d}$. We then define the indicator function $H^{h,s}_{g,d}$ as:

$$H^{h,s}_{g,d} = \begin{cases} 1, & \text{if } \tau^s_{g,d} \in [h, h+1), \\ 0, & \text{otherwise,} \end{cases}$$

[6]We ensure accessibility by providing a transparent methodology and a user-friendly implementation in R (R Core Team, 2025), enabling other researchers to apply our approach with ease.

[7]Retrieved from `https://data-explorer.oecd.org/`

[8]The UN's World Population Prospects 2022 (WPP 2022) provides country-level population estimates from 1950 to 2021 and projections from 2022 to 2100 (retrieved from `https://population.un.org/wpp`).

where $\tau^s_{g,d}$ represents the temperature in grid g on day d for the temperature variable s. Therefore, this function equals one if the daily temperature falls within the one-degree bin $[h, h+1)$, and zero otherwise.

The total number of days in year t for which the temperature in grid g falls within bin h is given by:

$$D^{h,s}_{g,t} = \sum_{d \in D_t} H^{h,s}_{g,d},$$

where D_t represents the number of days in year t. We average the number of days for each temperature bin h of type s across all grid cells within a country c in a given year t as follows:

$$\bar{X}^s_{ct,h} = \frac{1}{G_c} \sum_{g \in G_c} D^{h,s}_{g,t}, \tag{2}$$

where G_c denotes the total number of grid cells within country c. Thus, for each country–year pair, we obtain a discretized representation of the distribution of daily maximum and minimum temperatures. The final dataset includes 3,088 observations over the period 2000–2019.[9] Figure A1 in the appendix presents the histograms for each temperature variable.

We classify origin countries by climate zone using the Köppen–Geiger climate classification maps (version of March 2017).[10] Each country is assigned the dominant climate type based on the largest area within its borders. We group countries into four categories: tropical, arid, temperate, and continental.[11] We define climate zones as follows: *tropical climates* have a coldest-month mean temperature above 18°C with abundant year-round rainfall; *arid climates* occur where annual precipitation is below the threshold defined by potential evapotranspiration; *temperate climates* have coldest-month means between –3°C and 18°C; and *continental climates* have coldest-month means at or below –3°C and at least one month

[9] Two countries have missing data: Serbia (2000–2005) and Montenegro (2000–2005).

[10] Retrieved from `https://koeppen-geiger.vu-wien.ac.at/present.htm`

[11] Mauritius and Cabo Verde, classified as oceanic, are grouped under tropical and arid, respectively. Iceland, classified as polar, is included in the temperate group due to the small number of countries in the polar and oceanic categories.

exceeding 10°C.[12] Finally, we obtain real gross domestic product per capita from the Penn World Tables version 10.01. This dataset provides annual data for 183 countries from 1950 to 2019.

Table 1 reports summary statistics of each scalar variable by climate category for the year 2019. Figure A2 in the appendix displays the climate classification of origin countries in the sample.

Table 1: Summary statistics by climate category in 2019

	Outmigration rate (%)				Population (Millions)				Real GDP/capita (2017, USD) (Thousands)			
Category	Mean	Min	Max	S.D.	Mean	Min	Max	S.D.	Mean	Min	Max	S.D.
Arid	0.19	0.01	1.34	0.31	67.80	0.58	1421.86	223.91	17.64	1.18	103.45	21.77
Continental	0.38	0.04	1.13	0.32	43.69	1.33	334.32	91.98	34.36	5.02	63.39	19.30
Temperate	0.51	0.00	2.17	0.56	21.22	0.36	125.79	28.53	32.37	2.60	102.35	20.37
Tropical	0.19	0.00	0.93	0.22	57.63	0.39	1383.11	185.59	10.13	0.25	67.56	11.27

Notes. The table exhibits summary statistics by climate category for the year 2019.

Table 1 shows that, on average, the outmigration rate in 2019 was highest in temperate climates and lowest in arid climates. Arid countries have the largest average populations, while temperate countries have the smallest. Real gross domestic product per capita of origin countries is highest in continental and temperate climates and lowest in tropical climates.

3 Methods

3.1 Preprocessing and methodological tools

In this section, we detail the procedure we use to prepare the data for the scalar-on-composition and scalar-on-density regression models. For the scalar-on-composition model, we summarize the temperature data using a histogram, and for the scalar-on-density model, we summarize the temperature data using a spline function represented by its coefficients in a B-spline basis.

[12] See Kottek et al. (2006) for full details on the quantitative criteria of the Köppen–Geiger classification.

3.1.1 Histogram preprocessing

We start with temperature bins ranging from -64°C to 64°C for daily maximum temperatures and from -70°C to 41°C for daily minimum temperatures. Next, we drop all bins with an average number of days below 0.1 in the global distribution (see Figure A1), as these are very infrequent. Finally, for the remaining bins, we replace zero values with 10^{-7} to avoid numerical issues in subsequent log-transformation steps.

In this step, we replace the original index h, which denoted the specific temperature bin, with the index b, which indicates its position in the vector. For country c in year t, let $U^s_{ct} = (u^s_{ct,1}, u^s_{ct,2}, \ldots, u^s_{ct,B_s})$ be the vector whose components $u_{ct,b}$ corresponds to the relative frequency $\bar{X}^s_{ct,h}$ from (2), where h is now represented by the positional index b, and B_s is the total number of temperature bins, which depends on the value of s.

3.1.2 Compositional data

Compositional data are vectors of positive components that convey relative rather than absolute information. Their components are positive and add up to a constant, which we may assume to be one without loss of generality. For the constant equal to 1, they lie in the unit simplex space

$$\mathcal{S}^D = \left\{ \mathbf{u} = (u_1, \ldots, u_D)^T : u_m > 0, m = 1, \ldots, D; \ \sum_{m=1}^{D} u_m = 1 \right\}. \tag{3}$$

For example, the vector U^s_{ct} introduced in Section 3.1.1 is a compositional vector: it is quite easy to see that the sum of its components is equal to the number of days in year t multiplied by the number of bins B_s. The following closure operator $\mathcal{C}(.)$ maps this compositional vector U^s_{ct} to Z^s_{ct} in the unit simplex:

$$Z_{ct}^s = \mathcal{C}(U_{ct}^s) = \left(\frac{u_{ct,1}^s}{\sum_{b=1}^{B_s} u_{ct,b}^s}, \frac{u_{ct,2}^s}{\sum_{b=1}^{B_s} u_{ct,b}^s}, \ldots, \frac{u_{ct,B_s}^s}{\sum_{b=1}^{B_s} u_{ct,b}^s} \right) \tag{4}$$

We have $Z_{ct}^s = (\zeta_{ct,1}^s, \zeta_{ct,2}^s, \ldots, \zeta_{ct,B_s}^s)$, with $\zeta_{ct,B_s}^s = \frac{u_{ct,B_s}^s}{\sum_{b=1}^{B_s} u_{ct,b}^s}$ and $\sum_{b=1}^{B} \zeta_{ct,j}^s = 1$.

The simplex space can be equipped with a Enclidean structure by defining a proper addition (called perturbation), substraction (called negative perturbation) and a multiplication by a scalar (called powering), see Pawlowsky-Glanh et al. (2015). These operations are obtained by pulling back to the simplex the classical operations of $\mathbb{R}^D$ by the inverse clr transformation. For inclusion of a compositional vector in a regression model, its doubly constrained nature (positivity and sum equal to one) requires a specific treatment. The prevalent approach relies on the use of log-ratio transformations which are scale invariant and map these vectors into an uncontrained space where classical regression can be used, see for example Van den Boogaart and Tolosana-Delgado (2013).

The simplest log-ratio transformation is the centered log-ratio (clr) transformation defined by:

$$\text{clr}(Z_{ct}^s) = \left(\ln(\zeta_{ct,1}^s), \ln(\zeta_{ct,2}^s), \ldots, \ln(\zeta_{ct,B_s}^s) \right) - \left(\frac{1}{B_s} \sum_{b=1}^{B_s} \ln(\zeta_{ct,b}^s) \right) \left(1, \ldots, 1 \right) \tag{5}$$

The clr transformed vector, which lies in $\mathbb{R}^D$, has lost the positivity constraint but still has a constraint since the sum of its components is now 0. For this reason, it is not easy to use the clr transformed vector within a regression model and people have introduced more elaborate transformations to get rid of this singularity. The family of isometric log-ratio (ilr) transformations (Pawlowsky-Glanh et al., 2015) addresses the singularity problem of the centered log-ratio transformation by generating orthogonal coordinates. The ilr-transformed vector $V_{ct}^s = \text{ilr}(Z_{ct}^s) := (v_{ct,1}^s, v_{ct,2}^s, \ldots, v_{ct,B_s-1}^s)$ lies in a space of smaller dimension $\mathbb{R}^{D-1}$. It can be expressed as the product of a so-called contrast matrix (see Pawlowsky-Glanh et al. (2015)) by the clr transformed vector. The precise expression of these matrices will not be

needed to understand the ensuing developments. Indeed, it is just a technical tool to allow the use of classical linear regression in the transformed space but the results transformed back to the simplex are the same independently of the chosen ilr transformation.

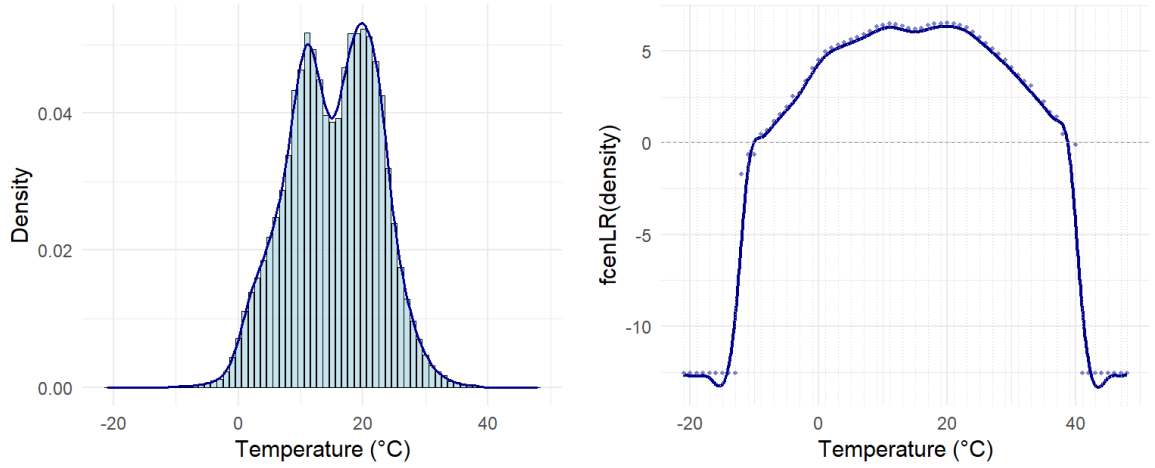
3.1.3 Spline preprocessing

We fit the clr-transformed histograms of daily maximum and daily minimum temperatures with cubic splines, to capture the continuous nature of the data and avoid loosing the bin's order in the further treatment. For each country and year, we estimate a cubic spline on the clr-transformed histogram, $\text{clr}(Z_{ct}^s)$, which yields a smoothed density function for temperature. To do so, due to the constrained nature of the clr vector, we need a specific kind of splines introduced in Machalova et al. (2016).

We place knots at the lower bound of evenly spaced intervals, selecting every second bin; if the total number of bins is odd, we include the upper bound of the last interval. This results in 36 knots for maximum and 32 for minimum daily temperatures. We choose uniform spacing because country-level temperature distributions vary widely, and using quantiles of the global distribution yields poor smoothing performance.

With this setup, we use cross-validation to select the optimal smoothing parameter, evaluating 100 values on a log-scale grid. This allows the optimal parameter to vary by country and year. Figure 1 illustrates the histogram, smoothed density, and centered log-ratio transformation of maximum daily temperatures for France in 2008, showing a strong fit of the spline to the observed data.

Figure 1: Density (left) and clr density (right) for France in 2008.



3.2 Model specifications

We compare three model specifications in an out-of-sample exercise. All models use the log of the outmigration rate as the dependent variable. They differ only in how they summarize temperature: an annual average (scalar), a one-degree histogram treated as a composition (discrete), and a smoothed distribution based on B-splines (density). In each specification, we include climate-category indicators and interactions between maximum temperature and climate categories. These interactions allow the temperature–outmigration relationship to vary across climate zones, which captures structural differences in adaptation. For example, tropical countries can respond differently to additional heat than temperate countries.

In all specifications, we also control for GDP per capita at the origin. Origin GDP per capita proxies income, development, and liquidity constraints that shape households' ability to finance migration and their exposure to climate-sensitive livelihoods. We include climate-category fixed effects rather than country fixed effects. Our goal is to predict outmigration using climate distributions across countries and over time. Country fixed effects would absorb all time-invariant cross-country differences in outmigration. They would also absorb much of the long-run climate variation, since temperature distributions differ sharply across countries.

With country fixed effects, identification would rely mainly on within-country year-to-year deviations. That design emphasizes short-run anomalies around a country's baseline climate. It does not align with our objective, which focuses on distributional shifts under long-run warming scenarios.

Climate-category fixed effects control for broad, persistent heterogeneity while preserving cross-sectional variation that supports prediction and scenario-based projections. Finally, because we aggregate flows over destinations, we do not model destination-specific conditions, such as destination-level policies or business cycles. We interpret our predictions as total outmigration from each origin country to the set of OECD destinations, not as bilateral destination choice.

3.2.1 Scalar-on-scalar regression

We begin with the scalar-on-scalar specification, widely used in the climate change literature. The independent variables are the annual averages of daily maximum and minimum temperatures. The baseline model is as follows.

$$\log(OMR)_{ct} = \overline{\tau}_{ct}^{max}\delta^{max} + \sum_{j\in J} I_{c,j}\eta_j + \sum_{j\in J} \overline{\tau}_{ct}^{max} I_{c,j}\gamma_j + \overline{\tau}_{ct}^{min}\delta^{min} + GDP_{ct}\phi + \theta_t + \epsilon_{ct}, \quad (6)$$

where $\overline{\tau}_{ct}^{max}$ and $\overline{\tau}_{ct}^{min}$ are the annual averages of daily maximum and minimum temperatures in country c in year t, δ^{max} and δ^{min} are their corresponding coefficients. $I_{c,j}$ is an indicator equal to 1 if c belongs to climate group $j \in \{\text{Tropical}, \text{Arid}, \text{Temperate}\}$ (zero for Continental), and η_j is the corresponding group coefficient. γ_j is the coefficient associated with the interaction between the annual average of daily maximum temperature and the indicator for climate group j. GDP_{ct} denotes GDP per capita (coefficient ϕ). It controls for time-varying economic development and resources that shape both migration incentives and the ability to pay migration costs. The relationship can be non-monotonic. Higher income can relax

liquidity constraints and raise outmigration, while very low income can limit mobility even under strong push factors. Finally, θ_t are year fixed effects for common shocks, and ϵ_{ct} is the error term.

3.2.2 Scalar-on-composition regression

The second model adopts the scalar-on-composition specification, using the discretized distribution of daily maximum and minimum temperatures (ilr-transformed histogram) as explanatory variables:

$$\begin{aligned}\log(OMR)_{ct} = & \sum_{b=1}^{B_{max}} v_{ct,b}^{max}\delta_b^{max} + \sum_{j\in J} I_{c,j}\eta_j + \sum_{j\in J}\sum_{b=1}^{B_{max}} v_{ct,b}^{max} I_{c,j}\gamma_{b,j} \\ & + \sum_{b=1}^{B_{min}} v_{ct,b}^{min}\delta_b^{min} + GDP_{ct}\phi + \theta_t + \epsilon_{ct}, \end{aligned} \tag{7}$$

where B_{max} (resp. B_{min}) is the number of bins of the histogram of T_{max} (resp. T_{min}), $v_{ct,b}^{\max}$ is the b-th component of $V_{ct}^{\max}$, the ilr transform of the maximum-temperature histogram for country c in year t, with corresponding coefficient $\delta_b^{\max}$. $\gamma_{b,j}$ denotes the interaction coefficient between that b-th ilr component and the indicator for climate group j. We define $v_{ct,b}^{\min}$ analogously for the minimum-temperature histogram, with coefficients $\delta_b^{\min}$. Note that the first term in (7) is simply the inner product of the vector $V_{ct}^{\max}$ with the vector of corresponding coefficients $\boldsymbol{\delta^{max}}$ and therefore by isometry is also equal to the inner product in the simplex of the maximum-temperature histogram with components $\bar{X}_{ct,h}^{max}$ with the ilr inverse of $\boldsymbol{\delta^{max}}$. This term can also be written as the inner product of the clr coefficients of the histogram vector with the clr coefficients of the ilr inverse of $\boldsymbol{\delta^{max}}$. Same can be done for the third term (interaction between temperature and climate type) and the fourth term for mimimum-temperature.

3.2.3 Scalar-on-density regression

Finally, we test whether incorporating continuous temperature densities improves outmigration predictions. This model is specified as follows:

$$\begin{aligned}\log(OMR)_{ct} = \; & \langle \beta^{\max}(\tau), f_{ct}^{\max}(\tau) \rangle_{\mathcal{B}^2([a_{\max}, b_{\max}])} + \sum_{j \in J} \eta_j \cdot I_{c,j} \\ & + \sum_{j \in J} I_{c,j} \cdot \left\langle \gamma_j^{\max}(\tau), f_{ct}^{\max}(\tau) \right\rangle_{\mathcal{B}^2([a_{\max}, b_{\max}])} \\ & + \left\langle \beta^{\min}(\tau), f_{ct}^{\min}(\tau) \right\rangle_{\mathcal{B}^2([a_{\min}, b_{\min}])} + GDP_{ct} \cdot \phi + \theta_t + \epsilon_{ct}, \end{aligned} \tag{8}$$

where $f_{ct}^{max}(\tau)$ denotes the estimated maximum daily temperature density function for country c in year t, defined over the interval $[a_{max}, b_{max}]$, and $f_{ct}^{min}(\tau)$, the equivalent for the minimum daily temperatures over the interval $[a_{min}, b_{min}]$. We obtain this density by fitting cubic splines to the centered log-ratio–transformed histogram of daily temperatures. The inner product

$$\langle \beta^s(\tau), f_{ct}^s(\tau) \rangle_{\mathcal{B}^2([a_s, b_s])}$$

represents the integral of the product of the coefficient function $\beta^s(\tau)$ and the density function $f_{ot}^s(\tau)$ over the interval $[a_s, b_s]$, which is the functional covariate in the regression model. This formulation allows the model to flexibly capture the potentially non-linear and heterogeneous effects of temperature across different segments of its distribution.

The coefficient functions $\beta^s(\tau)$ weight the contribution of the temperature density at each point within the specified interval, allowing a continuous assessment of the impact of temperature variations on the outmigration rate. This approach uncovers distributional patterns that average summaries obscure, offering a more detailed view of the climate–migration relationship by identifying how changes in specific temperature ranges affect migration.

3.3 Model evaluation

We compare all candidate models using three in-sample metrics: the adjusted R^2, which penalizes overfitting from additional parameters; and the root mean squared error (RMSE), which captures the average deviation between fitted and observed values.

Because a model may fit well in-sample due to overfitting, we also assess out-of-sample performance using the root mean squared prediction error (RMSPE)—the out-of-sample counterpart of the root mean squared error. We implement a modified cross-validation routine, based on a leave-one-year-out approach for panel data with time fixed effects (Lu and Su, 2020; Shao, 1993).

The procedure is as follows. First, for each year in the dataset, we randomly set aside 50% of that year's observations as the test set. Second, we train the model on the remaining data, which includes all other years as well as the non-held-out half of the current year. Third, we generate predictions for the held-out subset and compute the RMSPE to evaluate predictive accuracy. Finally, we repeat this process for each year and aggregate the resulting RMSPE values to assess the model's performance over time.

The rationale for this approach is that the inclusion of year fixed effects requires retaining part of each year's data in the training set to estimate the corresponding time effects. Excluding all observations from the target year would prevent the model from projecting the dependent variable for that year.

4 Results

Table 2 reports the results of the baseline models including dummy variables for the climate categories and interaction terms between the climate categories and the maximum temperature. The table compares three specifications: the scalar-on-scalar (first column), scalar-on-composition (second column), and scalar-on-density (third column).

Table 2 shows that, across the three specifications, GDP per capita enters with a positive and statistically significant coefficient. This pattern fits the idea that very low income constrains mobility. Previous evidence shows that households in poorer countries often lack the resources needed to finance international migration, so higher income can increase outmigration by relaxing liquidity constraints and by raising the expected returns to moving (Cattaneo and Peri, 2016; Barbosa-Alves and Britos, 2025). For the scalar-on-scalar specification, average daily maximum temperature is non-significant, while the average daily minimum temperature is positively associated with a higher outmigration rate. Intuitively, this suggests that zones with warmer temperatures are associated with more outmigration. Table A1 in the Appendix presents results from an alternative specification using annual averages of daily mean temperatures. The coefficient continues to be positive and statistically significant, indicating that higher temperatures are associated with increased outmigration rates.

For the three models, arid and temperate climates are associated with a higher outmigration rate. For the scalar-on-scalar model, the interaction of arid and temperate climate with the average daily maximum temperature is negative and significative, indicating that those climate zones may be more resilient to higher temperatures than other climates such as continental. Intuitively, arid regions have already adapted to higher temperatures, so a further increase in the maximum temperature may be less of a shock compared to other regions.

Table 2 also shows that in-sample, both the scalar-on-composition and scalar-on-density models outperform the scalar-on-scalar specification across all metrics. Compared to the scalar-on-scalar model, the scalar-on-composition increases the R^2 by 90%, the adjusted R^2 by 80%, and reduces the RMSE by 28%. The scalar-on-density model produces slightly smaller gains, with 80% increase in R^2, a 74% increase in adjusted R^2, and a 23% reduction in RMSE.

Table 2: Comparison of models for the log of outmigration rate using average daily maximum and minimum temperatures

	Scalar-on-scalar	Scalar-on-composition	Scalar-on-density
Log GDP/capita	0.193**	0.156**	0.189**
	(0.097)	(0.070)	(0.075)
Avg. daily max temp.	−0.083		
	(0.053)		
Avg. daily min temp.	0.175***		
	(0.044)		
Tropical dummy	−2.161	0.963	1.859
	(2.140)	(2.238)	(1.717)
Arid dummy	1.680**	1.908***	1.644***
	(0.816)	(0.622)	(0.612)
Temperate dummy	2.041***	2.149**	2.297***
	(0.576)	(0.924)	(0.752)
Tropical x max temp.	−0.014		
	(0.076)		
Arid x max temp.	−0.130***		
	(0.043)		
Temperate x max temp.	−0.153***		
	(0.046)		
Num.Obs.	3,088	3,088	3,088
R2	0.346	0.656	0.620
R2 Adj.	0.340	0.611	0.593
RMSE	1.16	0.84	0.89
Max and min temp histogram	No	Yes	Yes
Time FE	Yes	Yes	Yes

Notes. This table compares the interaction scalar-on-scalar model with the interaction scalar-on-composition and scalar-on-density specifications. The maximum and minimum temperature histograms use the full temperature distributions rather than the annual average for each variable. *p<0.1; **p<0.05; ***p<0.01.

Finally, Table 3 shows the results of our cross-validation approach to assess out-of-sample performance. Specifically, Table 3 presents the RMSPE for each excluded year, along with the average RMSPE for all years and the improvement relative to the corresponding scalar-on-scalar version. As in the in-sample evaluation, we observe improvements in terms of RMSPE when using the scalar-on-composition and scalar-on-density models. In particular, the RMSPE is lower for each specific year, and, on average, the reduction in RMSPE for the

scalar-on-composition (scalar-on-density) model is -19.8% (-21.1%).

Tables A1 and A2 in the Appendix report the in-sample and out-of-sample performance of the model using daily mean temperature as the climate metric. For reference, both tables also include the performance of the baseline model based on the distributions of the daily minimum and maximum temperatures. The results reveal a consistent pattern: models that account for the full distribution of temperatures—namely, the scalar-on-composition and scalar-on-density specifications—outperform the scalar-on-scalar model in terms of RMSE and RMSPE. Notably, the improvement is slightly larger for the baseline model that uses daily minimum and maximum temperatures compared to the model that uses daily mean temperature, as shown in Table A2.

Table 3: Modified leave-one-year-out cross validation (RMSPE)

Year out	Scalar-on-scalar	Scalar-on-composition	scalar-on-density
2000	1.129	1.265	1.009
2001	1.237	1.044	1.046
2002	1.185	0.903	0.897
2003	1.328	1.056	1.092
2004	1.131	1.007	1.011
2005	1.220	0.939	0.916
2006	1.188	0.988	0.942
2007	1.351	1.071	1.031
2008	1.208	0.927	0.863
2009	1.149	0.806	0.851
2010	1.208	0.881	0.890
2011	1.200	0.815	0.851
2012	1.048	0.845	0.830
2013	1.134	0.923	0.802
2014	1.100	0.969	0.937
2015	1.204	0.926	0.888
2016	1.082	0.845	0.918
2017	1.165	0.961	0.946
2018	1.254	0.950	1.002
2019	1.302	0.982	1.070
Average	1.191	0.955	0.940
Improvement (% chg.)	0.000	-19.810	-21.103

Notes. The table compares the scalar-on-scalar, scalar-on-composition, and scalar-on-density specifications based on their out-of-sample performance, measured by the Root Mean Squared Percentage Error (RMSPE), for each year excluded from the sample. We implement a leave-one-year-out cross-validation procedure, in which each year is sequentially held out as the test set while the model is trained on the remaining years.

5 The projected marginal impact of climate change on migration flows at the end of the century

Following Dargel and Thomas-Agnan (2024), we measure marginal effects of the compositional regressor in (7) by expressing a change in the temperature histogram as a relative change of the original composition. Let $\bar{X}^s_{ct} = (\bar{X}^s_{ct,1}, \ldots, \bar{X}^s_{ct,B_{max}})$ denote the vector of histogram shares. We represent the projected change in the temperature distribution by a compositional perturbation $\Phi^{\max}_{c,\text{ssp}}$ and map the baseline composition to the counterfactual one through the perturbation operator, $\bar{X}^s_{ct} \mapsto \bar{X}^s_{ct} \oplus \Phi^{\max}_{c,\text{ssp}}$. This construction ensures that the counterfactual histogram remains in the simplex and that the change has a linear representation in clr space.

For the projection exercise, we combine the estimated migration response function with projected changes in the distribution of daily temperatures in 2100 under alternative Shared Socioeconomic Pathways (SSPs). We obtain temperature projections from CMIP6 through the Copernicus Climate Data Store. In our baseline projections, we use the GFDL-ESM4 model to construct country-specific changes in the distribution of daily maximum temperatures between 2019 and 2100 under SSP1-2.6.[13] SSP1-2.6 represents a low-emissions pathway with strong mitigation and a focus on sustainability. We use it as a benchmark to provide conservative projections of climate change impacts. Our projections therefore condition on a given SSP and a given climate model. They capture uncertainty in the estimated migration model, but they do not incorporate uncertainty across climate models or uncertainty about which SSP path the world follows.

Let $\Phi^{\max}_{c,\text{ssp}} = (\varphi^{\max}_{c1,\text{ssp}}, \ldots, \varphi^{\max}_{c\{B\max\},\text{ssp}})$ denote the compositional shift between the projected 2100 distribution of daily maximum temperatures and the observed 2019 distribution

[13] Data obtained from `https://cds.climate.copernicus.eu/datasets/projections-cmip6?tab=overview` (GFDL-ESM4 model).

for country c under scenario ssp. We compute the implied change in outmigration by combining this shift with the estimated coefficients on the maximum-temperature histogram. The resulting impact is

$$\text{Impact}_{c,ssp} = \log(OMR)_{c,ssp} - \log(OMR)_{c,2019} = \sum_{b=1}^{B_{max}} \text{clr}(\Phi^{\max}_{c,ssp})_b \alpha_{cb}, \tag{9}$$

where $\text{clr}(\Phi^{\max}_{c,ssp})_b$ is the (additive) shift in clr space for bin b and origin country c and where α_{cb} are coefficients in clr space corresponding to the two terms involving the maximum-temperature histogram: $\text{clr}(\text{ilr}^{-1}(\boldsymbol{\delta^{max}} + \boldsymbol{\gamma^{max}_c}))$. We obtain estimated impacts by replacing the parameters in (9) with their corresponding estimates.

Table 4 reports the estimated log-changes in outmigration rates associated with projected changes in the distribution of daily maximum temperatures in 2100 under the SSP1-2.6 scenario, disaggregated by climate category. The results indicate a 39% increase in outmigration rates for arid countries (0.391), although it is not statistically significant. In contrast, for continental climates, the projection indicates a large reduction in outmigration rates of 136% (–1.360). Similarly, tropical climates are expected to see a 31% reduction in outmigration rates (–0.309). For temperate climates, the projected effect is modest and statistically insignificant—a 7% increase (0.067) in outmigration rates.

In addition, Table 4 reports the lower and upper bounds of the estimated impacts, as well as their standard deviations. These bounds summarize uncertainty in the estimated migration response function. They capture sampling variation in the regression parameters and propagate it into the implied impacts. We interpret these intervals as regression-model uncertainty conditional on the assumed warming path and the chosen climate model. They do not capture uncertainty across SSP scenarios or uncertainty across climate models within CMIP6. When we compute the coefficients of variation, we find that countries in the temperate region exhibit the largest relative dispersion. This suggests that, although the aggregate

impact for temperate countries appears modest, it masks substantial heterogeneity across individual countries. Recall that this category consists primarily of European countries. In contrast, the estimated impacts for countries in the continental region—followed by those in the tropical region—show the lowest relative variation, indicating more consistent effects within those groups.

Table 4: Impacts in outmigration rates in 2100 by climate category under SSP1-2.6 scenario in the scalar-on-compositional model

Category	Outmigration 2019 (%)	Impact (log change)	Lower bound	Upper bound	Std. Dev.	Number of countries
Arid	0.195	0.391	-0.498	1.279	0.453	40
Continental	0.375	-1.360***	-2.049	-0.671	0.352	14
Temperate	0.511	0.067	-0.599	0.734	0.340	43
Tropical	0.185	-0.309*	-0.673	0.054	0.185	58

Notes. The table reports outmigration rates in 2019 and the projected impacts on outmigration rates of changes in the distribution of maximum daily temperatures in 2100 under the SSP1-2.6 scenario, disaggregated by climate category. It expresses impacts as log changes, where positive values indicate increases in outmigration rates and negative values indicate decreases. The lower and upper bounds provide 95% confidence intervals around the impact point estimates. Std. Dev. stands for the standard deviation of the estimated impact. *p<0.1; **p<0.05; ***p<0.01.

Table 5 compares observed outmigration (in millions of people) in 2019 with projections for 2100 based on the scalar-on-scalar and scalar-on-composition models, disaggregated by region. Both projections combine estimated outmigration rates with population forecasts to express results in absolute terms. In 2019, 7.49 million people migrated to OECD countries. By 2100, projections indicate that between 7.5 and 8.3 million people will do so due to climate-induced changes in temperature distributions.

The table also highlights notable differences between the two modeling approaches. The scalar-on-composition model projects larger increases in outmigration from Africa, Asia, and Oceania, and larger declines from the Americas and Europe, compared to the scalar-on-scalar model. For example, in levels, the composition model projects Africa rising from 0.65 million (2019) to 1.07 million (2100) and Asia rising from 2.79 to 3.79 million. It also projects the

Americas falling from 1.40 to 0.98 million and Europe falling from 2.58 to 2.34 million.

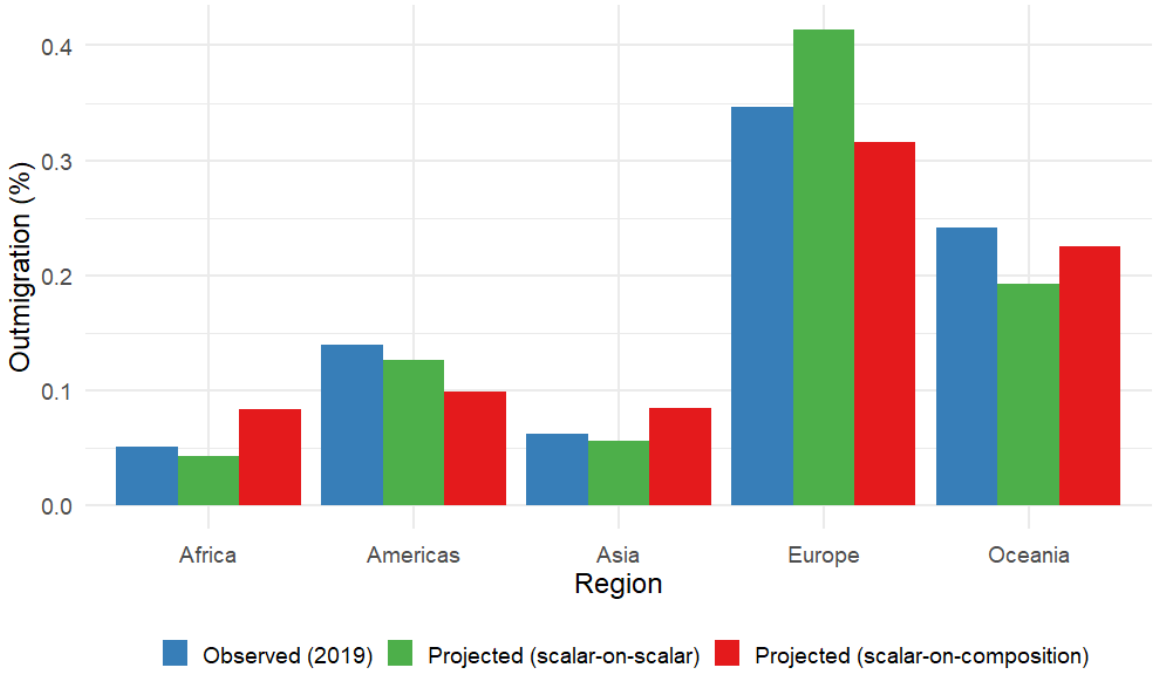
Table 5: Observed and projected outmigration (millions of people) under the SSP1-2.6 scenario for each model

Region	Observed (2019)	Projected (2100)	
		Scalar-on-scalar	Scalar-on-composition
Africa	0.65	0.55	1.07
Americas	1.40	1.26	0.98
Asia	2.79	2.51	3.79
Europe	2.58	3.08	2.34
Oceania	0.08	0.06	0.07
Total	7.49	7.46	8.26

Notes. The table shows observed outmigration in 2019 and projections for 2100 under SSP1-2.6 scenario by region (millions of people). The table compares two model specifications: scalar-on-scalar and scalar-on-composition.

Figure 2 complements these findings by comparing observed and projected outmigration rates—expressed relative to regional populations—under both modeling approaches and disaggregated by region. In Africa and Asia, the scalar-on-scalar model predicts a decline in outmigration rates, while the scalar-on-composition model projects an increase, highlighting its sensitivity to changes in the temperature distribution. In the Americas, both models project declining outmigration, with the composition-based model indicating a steeper reduction. In Europe, the scalar-on-scalar model forecasts an increase in outmigration, whereas the scalar-on-composition model suggests a decline. In Oceania, both models project reduced outmigration, though the scalar-on-scalar model anticipates a more pronounced decrease.

Figure 2: Observed (2019) and projected (2100) outmigration rates of the SSP1-2.6 scenario for each model by region



The economic intuition for the results in Table 5 and Figure 2 comes from what each model treats as *the relevant climate shock*. The scalar-on-scalar model summarizes temperature with a single annual average, so it smooths over where within the distribution the change occurs. In contrast, the scalar-on-composition model uses the full binned distribution of daily maximum temperatures, so it responds to shifts in specific ranges of the distribution. This difference matters because migration responses can depend on tail risk and on thresholds. A small movement in an annual mean can hide a large change in the number of very hot days, or a large reduction in very cold days.

In Europe and the Americas, the composition model projects lower outmigration because it assigns weight to how warming reallocates days across bins in climates where "cold relief" can dominate. Table 4 shows a large projected reduction for continental climates (–1.360) and only a modest average effect for temperate climates, and temperate climates consist primarily

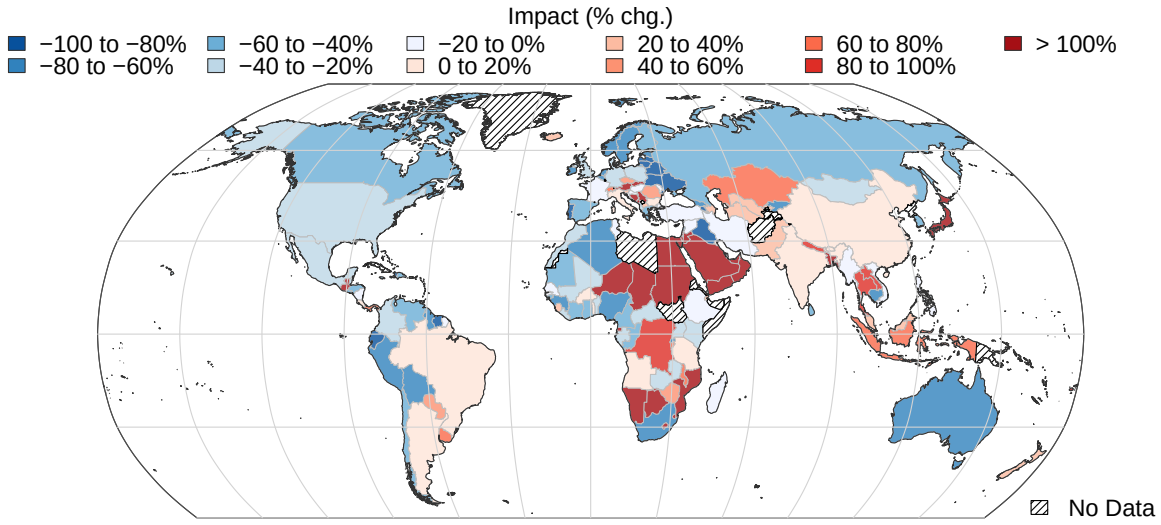
of European countries. Europe and North America include many temperate and continental areas, so a shift away from cold days and toward milder ranges can reduce climate-related push factors on net. This logic also fits Figure 2, where the scalar-on-composition model projects a decline for Europe and a steeper decline for the Americas than the scalar-on-scalar model. The scalar-on-scalar model can miss this channel because it reacts to annual averages. In our estimates, the average daily minimum temperature is positively associated with outmigration in the scalar-on-scalar specification, so warming can mechanically raise predicted outmigration even when the key change comes from fewer cold extremes rather than more harmful hot extremes.

In Africa and Asia, the composition model projects higher outmigration because warming often shows up as an increase in the frequency of very hot days, not just a higher mean. The composition approach picks up those upper-tail reallocations directly, so it projects increases in outmigration rates in both regions. This pattern matches the economic mechanism that heat extremes can reduce agricultural yields and labor productivity and can raise health risks, which increases incentives to move. The scalar-on-scalar model can instead predict declines in Africa and Asia because it compresses the entire distribution into a mean and because its interaction structure can dampen the marginal effect of higher maximum temperatures in some warm climate zones. In other words, the mean-based model can treat warming as a smooth change that societies absorb, while the distribution-based model treats the growth in extreme-heat days as the core shock that pushes people to leave.

Figure 3 illustrates the projected percentage change in outmigration rates for each country under the scalar-on-composition model using the SSP1-2.6 scenario. These changes reflect the effects of projected increases in maximum daily temperatures. Dark blue shades represent countries where outmigration rates are expected to decline by up to 100%, while dark red tones indicate countries where outmigration could more than double (100%). The largest

projected increases in outmigration to OECD countries appear in parts of Africa, Asia, and Europe—sometimes exceeding 100%. In contrast, other countries within the same regions exhibit declines or more moderate changes (ranging from –40% to +20%). This variation underscores the substantial heterogeneity in climate-driven migration responses, which reflects differences in baseline temperature distributions and region-specific projections.

Figure 3: Impact on outmigration rate of the SSP1-2.6 scenario by country for the scalar-on-composition model



Although our data, methodology, and projection horizons differ from those in related studies, our findings remain broadly consistent with previous work. For example, Cattaneo et al. (2024) use bilateral migration data and project higher migration flows from Africa, Asia, and Latin America, especially toward Europe and North America. This aligns with our results, which show that by the end of the century, the largest climate-induced increases in outmigration to OECD countries originate from these same regions (see Figure 2).

To facilitate a closer comparison with Cattaneo et al. (2024), we compute migration projections for the year 2030 under the SSP2-4.5 scenario (see Table A3 in the Appendix). Cattaneo et al. (2024) project annualized migration increases of approximately 3.6% from Africa, –1% from Asia, and 0.7% from the Americas to Europe and North America (a proxy

for OECD). In contrast, our SSP2-4.5 estimates yield annualized rates of 11.4%, 2.9%, and –3.7% for the same regional groups. Although the origin and destination sets differ, both studies point to substantial climate-related shifts in migration patterns, particularly from Africa and Asia.

In turn, Missirian and Schlenker (2017) project that, by the end of the century, asylum applications from non-OECD countries to the European Union will increase by 28% due to weather anomalies. Although our analysis does not restrict attention to asylum applications, our projections using the scalar-on-composition model show that total migration to OECD countries in 2100 will be approximately 10% higher than in 2019.

In terms of annualized changes, Missirian and Schlenker (2017) estimate long-term migration shifts (2070–2099) to the EU under RCP 4.5 of 0.3% from Africa, 0.5% from Asia, 0.2% from the Americas, and –0.2% from Europe. In comparison, our projections for 2100 under the SSP2-4.5 scenario show annualized changes in outmigration to OECD countries of 0.9% from Africa, 0.1% from Asia, –0.5% from the Americas, and –0.6% from Europe. Although our destination set is broader and the definitions of migration differ, both studies identify Africa as the region that contributes the largest increases in climate-driven migration.

Finally, Burzyński et al. (2022) estimate that climate change will induce approximately 57 million international migrants aged 30 to 60 over the course of the 21st century, with most originating from sub-Saharan Africa, Asia, and South America. Their projections, however, are based on a general equilibrium model that accounts not only for climate shocks but also for induced economic adjustments, including shifts in productivity, sectoral structure, and location-specific income effects.

6 Marginal effects of hypothetical changes in the temperature distribution

To better interpret the relationship between temperature and outmigration, we simulate how hypothetical changes in a country's temperature distribution would affect outmigration rates. Unlike the previous section where we projected impacts under specific scenarios derived from Shared Socioeconomic Pathways, this section evaluates how targeted changes in selected temperature bins influence outmigration outcomes at a given time horizon H.

Because compositional data are subject to a constant-sum constraint, an increase in one bin must be offset by a decrease in others. To account for this, we construct a hypothetical distribution that specifies the direction of change in the original temperature profile that we aim to analyze.

Following Dargel and Thomas-Agnan (2024), starting from the initial distribution of daily maximum temperatures of country c in year 2019 denoted by $\bar{X}^s_{c2019}$, we write the hypothetical temperature distribution at time horizon H obtained by applying a change of intensity κ along a specified direction vector Φ as

$$\bar{X}^s_{cH}(\kappa) = \bar{X}^s_{c2019} \oplus \kappa \odot \Phi = \big(\bar{X}^s_{c2019,1}\,\varphi_1^{\kappa},\ \bar{X}^s_{c2019,2}\,\varphi_2^{\kappa},\ \ldots,\ \bar{X}^s_{c2019,B_{max}}\,\varphi_{B_{\max}}^{\kappa}\big).$$

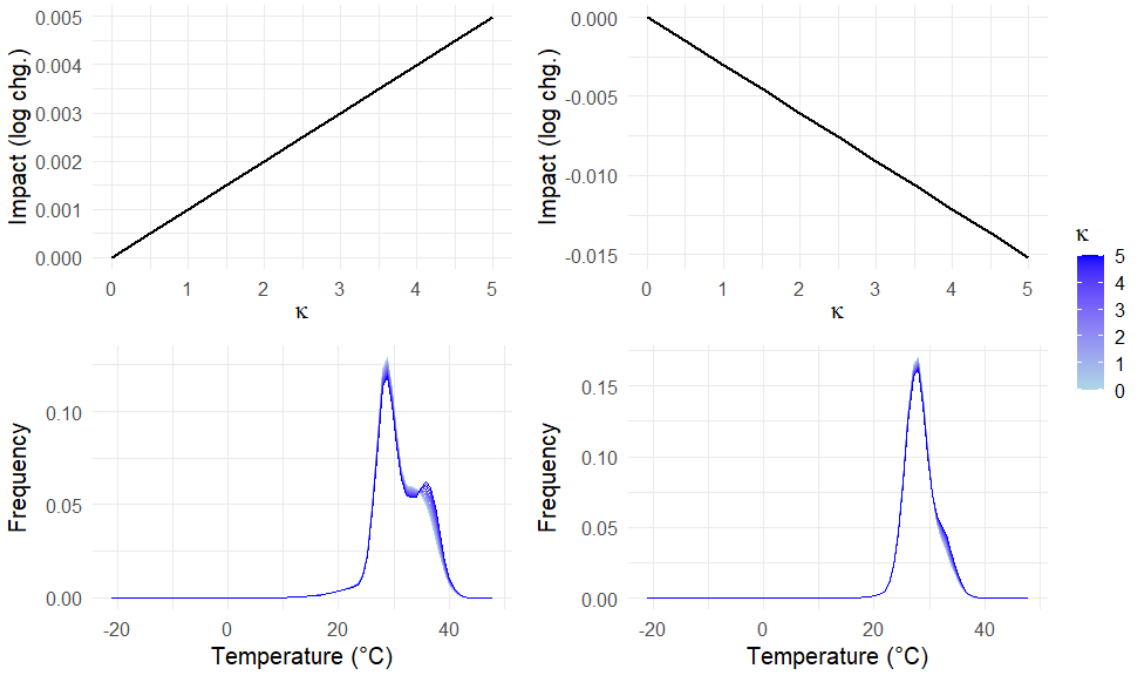
Each component φ_b of the direction vector indicates the relative change applied to bin b for an intensity of $\kappa = 1$, with $\kappa = 0$ corresponding to the original distribution (i.e., no change).

We construct a direction vector Φ that considers increases of the highest temperatures. Let *hot* denote the set of n_{hot} hottest bins, defined as those containing the top 10% of the temperature distribution mass. For each bin b, we set $\varphi_b = \exp(1/n_{\text{hot}})$ if $b \in hot$, and $\varphi_b = 1$ otherwise, and then close the vector.

The mass-based selection of the bins allows us to assess potential habituation effects: populations in relatively hot countries may be accustomed to high maximum temperatures, whereas populations in relatively cold countries may react more strongly to the same increase.

Figure 4 shows the impacts on the outmigration rate (top row) resulting from the distributional changes illustrated in the bottom row for Brazil (left column) and the Democratic Republic of Congo (right column). We focus on these two countries because they differ significantly in income levels, while sharing a similar support in their baseline temperature distributions. We observe that increased exposure to very high temperatures raises outmigration in Brazil but reduces it in the Democratic Republic of Congo. This pattern aligns with findings in the migration literature, suggesting that financial constraints in low-income countries often limit the ability to migrate in response to climate stressors. In contrast, people in middle-income countries such us Brazil are more likely to rely on migration as an adaptation strategy to extreme heat.

Figure 4: Impact in outmigration rate (top panel) of a extreme-heat change in the temperature distribution (bottom panel). Left: Brazil. Right: Democratic Republic of Congo



7 Conclusions

Climate change poses a significant threat to humanity, and rising global temperatures are recognized as a driver of increased human migration in the coming decades.

This paper presents a methodological framework for projecting future migration patterns. We conceptualize the global distribution of daily temperatures as compositional data, integrating this structure into a regression model. Our analysis demonstrates that incorporating the full temperature distribution significantly reduces prediction errors compared to models relying on simpler average temperature metrics.

It is important to acknowledge that our findings focus solely on the impact of temperature changes on international migration. Human migration is a complex phenomenon influenced by a multitude of factors, including but not limited to, migration policies, adaptation strategies, financial resources, and socio-political dynamics, all of which can either amplify or mitigate migration flows. The primary objective of this exercise is to demonstrate that more sophisticated and comprehensive measures of temperature can yield substantially more accurate migration predictions than those derived from univariate statistics such as average maximum or minimum temperatures, or median temperature.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data availability statement.

The data and code that support the findings of this study are available in the corresponding author's webpage. The repository includes replication code, the analysis datasets, and documentation describing all steps from raw inputs to final tables and figures.

References

Backhaus, A., I. Martinez-Zarzoso, and C. Muris (2015). Do climate variations explain bilateral migration? A gravity model analysis. *IZA Journal of Migration 4*, 1–15.

Barbosa-Alves, M. and B. Britos (2025). Climate change and international migration. Unpublished document.

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

Boas, I., C. Farbotko, H. Adams, H. Sterly, S. Bush, K. Van der Geest, H. Wiegel, H. Ashraf, A. Baldwin, G. Bettini, et al. (2019). Climate migration myths. *Nature Climate Change 9*(12), 901–903.

Bosetti, V., C. Cattaneo, and G. Peri (2021). Should they stay or should they go? Climate migrants and local conflicts. *Journal of Economic Geography 21*(4), 619–651.

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

Caballero Reina, J., J. Crespo Cuaresma, K. Fenz, J. Zellmann, T. Yankov, and A. Taha (2024). Gravity models for global migration flows: A predictive evaluation. *Population Research and Policy Review 43*(2), 29.

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.

Carleton, T. A. and S. M. Hsiang (2016). Social and economic impacts of climate. *Science 353*(6304), aad9837.

Carter, C., X. Cui, D. Ghanem, and P. Mérel (2018). Identifying the economic impacts of climate change on agriculture. *Annual Review of Resource Economics 10*, pp. 361–380.

Cattaneo, C., E. Massetti, S. Dasgupta, and F. Farinosi (2024). *Climate Variability and Worldwide Migration: Empirical Evidence and Projections.* IMF Working Papers. Washington, D.C: International Monetary Fund.

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

Conte, B., K. Desmet, D. K. Nagy, and E. Rossi-Hansberg (2021). Local sectoral specialization in a warming world. *Journal of Economic Geography 21*(4), 493–530.

Dargel, L. and C. Thomas-Agnan (2024). Pairwise share ratio interpretations of compositional regression models. *Computational Statistics & Data Analysis 195*, 107945.

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.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics 3*(4), 152–85.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018). The geography of development. *Journal of Political Economy 126*(3), 903–983.

Desmet, K. and E. Rossi-Hansberg (2015). On the spatial economic impact of global warming. *Journal of Urban Economics 88*, 16–37.

Hoegh-Guldberg, O., D. Jacob, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K. Ebi, F. Engelbrecht, J. Guiot, et al. (2018). Impacts of 1.5°C global warming on natural and human systems. In *Global Warming of 1.5 C.: An IPCC Special Report*, pp. 175–311. IPCC Secretariat.

Hron, K., P. Filzmoser, and K. T. and (2012). Linear regression with compositional explanatory variables. *Journal of Applied Statistics 39*(5), 1115–1128.

Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change 2007: The Physical Science Basis*. Cambridge University Press.

Intergovernmental Panel on Climate Change (IPCC) (2014). *Climate Change 2014 – Impacts, Adaptation and Vulnerability*. Cambridge University Press.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift 15*(3), 259–263.

Lu, X. and L. Su (2020). Determining individual or time effects in panel data models. *Journal of Econometrics 215*(1), 60–83.

Machalova, J., K. Hron, and G. S. Monti (2016). Preprocessing of centred logratio transformed density functions using smoothing splines. *Journal of Applied Statistics 43*(8), 1419–1435.

McLeman, R. (2019). International migration and climate adaptation in an era of hardening borders. *Nature Climate Change 9*(12), 911–918.

Missirian, A. and W. Schlenker (2017). Asylum applications respond to temperature fluctuations. *Science 358*(6370), 1610–1614.

Mullins, J. T. and P. Bharadwaj (2021). Weather, climate, and migration in the United States. Working Paper 28614, National Bureau of Economic Research.

Pawlowsky-Glanh, V., J. J. Egozcue, and R. Tolosana-Delgado (2015). *Modeling and Analysis of Compositional Data*. John Wiley & Sons.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association 88*(422), 486–494.

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Stott, P. (2016). How climate change affects extreme weather events. *Science 352*(6293), 1517–1518.

Talská, R., K. Hron, and T. M. Grygar (2021). Compositional scalar-on-function regression with application to sediment particle size distributions. *Mathematical Geosciences 53*, 1667–1695.

Thomas-Agnan, C., M. Simioni, and T.-H. Trinh (2025). Discrete and smooth scalar-on-density compositional regression for assessing the impact of climate change on rice yield in vietnam. Technical report, Toulouse School of Economics (TSE).

Van den Boogaart, K. G. and R. Tolosana-Delgado (2013). *Analyzing compositional data with R*, Volume 122. Springer.

A1 Appendix

Specification using daily mean temperatures

As an alternative, we estimate the 3 regression models specified in Section 2 but using daily mean temperatures instead of daily maximum and minimum temperatures. Thus, we estimate the following 3 models.

$$\log(OMR)_{ct} = \overline{\tau}_{ct}^{avg}\delta^{avg} + \sum_{j\in J} I_{c,j}\eta_j + \sum_{j\in J} \overline{\tau}_{ct}^{avg} I_{c,j}\gamma_j + GDP_{ct}\phi + \theta_t + \epsilon_{ct} \tag{10}$$

$$\log(OMR)_{ct} = \sum_{b=1}^{B_{avg}} v_{ct,b}^{avg}\delta_b^{avg} + \sum_{j\in J} I_{c,j}\eta_j + \sum_{j\in J}\sum_{b=1}^{B_{avg}} v_{ct,b}^{avg} I_{c,j}\gamma_{b,j} + GDP_{ct}\phi + \theta_t + \epsilon_{ct} \tag{11}$$

$$\log(OMR)_{ct} = \langle \beta_{avg}(\tau), f_{ct,avg}(\tau)\rangle_{\mathcal{B}^2([a_{avg},b_{avg}])} + W_{ct}\phi + \theta_t + \epsilon_{ct} \tag{12}$$

All of them include the climate categories dummies and interaction with the annual average of the daily mean temperature. Table A1 and Table A2 present the results for the coefficients of the models, and the cross validation, respectively.

Notably, when using the scalar-on-composition and scalar-on-density models, we obtain improvements in in-sample performance in terms of R^2, adjusted R^2 and RMSE. This is also true in the out-of-sample exercise, although the reductions in RMSE are lower than when using the complete distribution of maximum and minimum temperatures, regardless of whether the scalar-on-composition or scalar-on-density models are employed.

Aditional figures

Figure A1: Global histogram of daily temperatures.

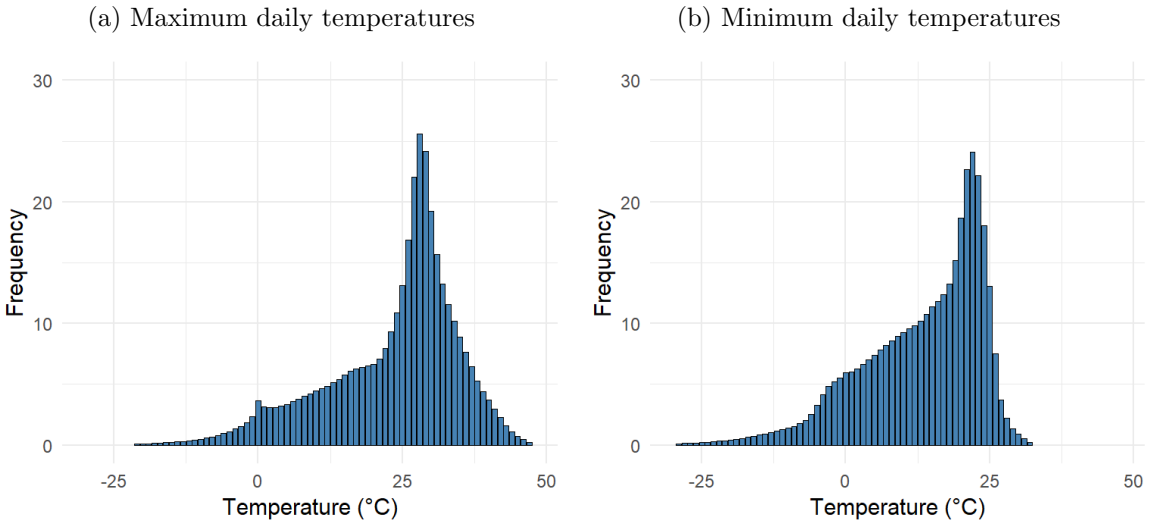


Figure A2: Climate categories for each country.

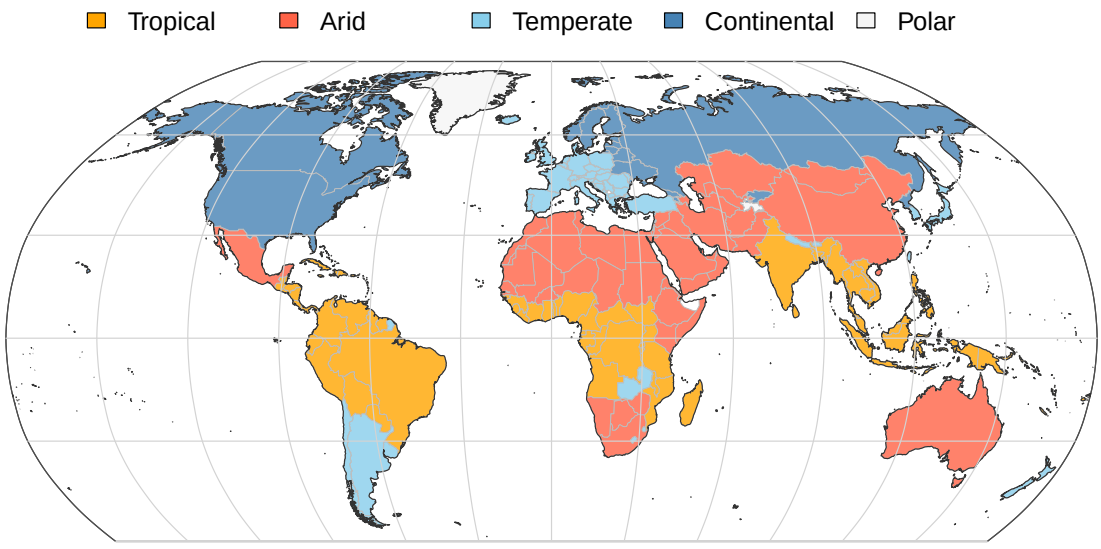


Figure A3: Histograms for each temperature variable by climate category (average across all countries and years).

(a) Maximum daily temperatures

(b) Minimum daily temperatures

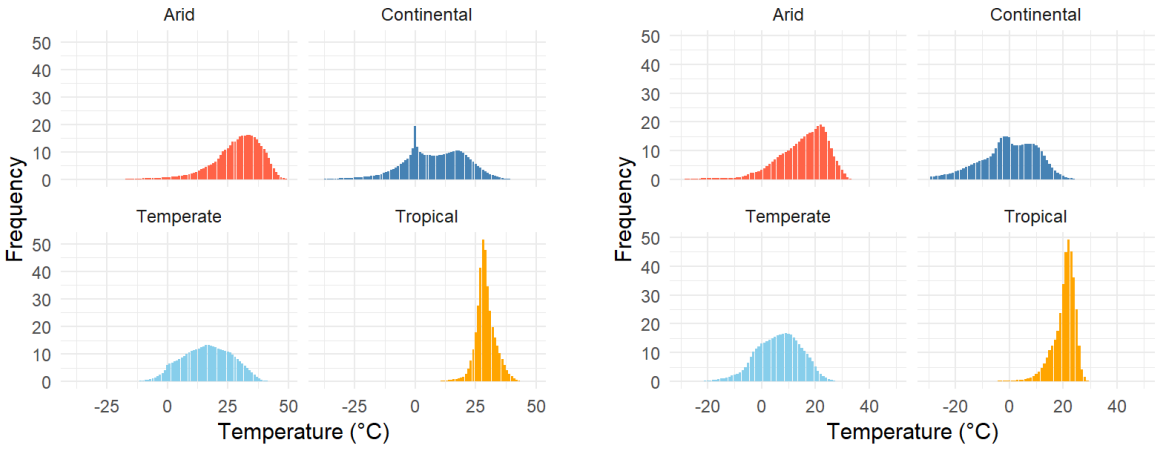
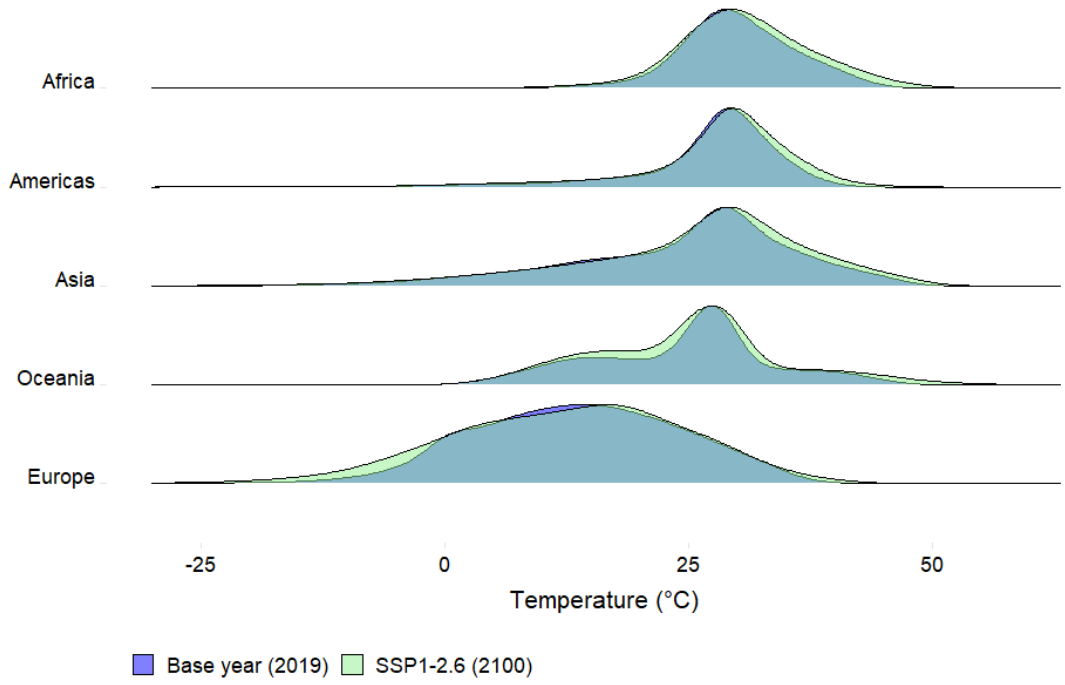


Figure A4: Observed (2019) and projected (2100) maximum daily temperature distribution under SSP1-2.6 by region



Aditional tables

Table A1: Comparison of models for the log outmigration rate using daily mean temperatures

	Scalar-on-scalar	Scalar-on-composition	Scalar-on-density
Log GDP/capita	0.341***	0.224**	0.228**
	(0.095)	(0.095)	(0.092)
Average daily mean temperature	0.084**		
	(0.035)		
Tropical dummy	−3.827*	1.825**	1.309
	(2.153)	(0.793)	(0.805)
Arid dummy	0.541	2.363***	2.238***
	(0.580)	(0.621)	(0.609)
Temperate dummy	1.855***	2.151**	1.688**
	(0.473)	(0.888)	(0.729)
Arid x mean temp.	−0.134***		
	(0.043)		
Tropical x mean temp.	0.068		
	(0.093)		
Temperate x mean temp.	−0.213***		
	(0.049)		
Num.Obs.	3,088	3,088	3,088
R2	0.296	0.579	0.557
R2 Adj.	0.290	0.537	0.533
RMSE	1.21	0.93	0.96
Mean temp histogram	No	Yes	Yes
Time FE	Yes	Yes	Yes

Notes. This table compares the scalar-on-scalar model with the scalar-on-composition and scalar-on-density specifications. Mean temp histograms means that the model use the full daily mean temperature distribution rather than the average. *p<0.1; **p<0.05; ***p<0.01.

Table A2: RMSPE for the model with daily maximum and minimum temperatures and the model with daily mean temperatures

	Scalar-on-scalar		Scalar-on-composition		scalar-on-density	
Year	Max & min	Mean	Max & min	Mean	Max & min	Mean
2000	1.129	1.119	1.265	1.008	1.009	0.948
2001	1.237	1.211	1.044	1.073	1.046	1.090
2002	1.185	1.152	0.903	1.006	0.897	0.997
2003	1.328	1.311	1.056	1.084	1.092	2.046
2004	1.131	1.108	1.007	1.046	1.011	1.006
2005	1.220	1.209	0.939	1.139	0.916	1.058
2006	1.188	1.168	0.988	1.026	0.942	1.007
2007	1.351	1.322	1.071	1.196	1.031	1.159
2008	1.208	1.197	0.927	0.893	0.863	0.930
2009	1.149	1.134	0.806	1.065	0.851	0.992
2010	1.208	1.176	0.881	0.957	0.890	0.893
2011	1.200	1.172	0.815	0.908	0.851	0.948
2012	1.048	1.034	0.845	0.924	0.830	0.902
2013	1.134	1.112	0.923	0.855	0.802	0.869
2014	1.100	1.084	0.969	0.982	0.937	1.019
2015	1.204	1.185	0.926	0.980	0.888	0.952
2016	1.082	1.041	0.845	0.888	0.918	0.897
2017	1.165	1.157	0.961	1.031	0.946	1.052
2018	1.254	1.242	0.950	1.223	1.002	1.181
2019	1.302	1.284	0.982	1.127	1.070	1.121
Average	1.191	1.171	0.955	1.021	0.940	1.053
Improvement (% change)	0.0	0.0	–19.8	–12.8	–21.1	–10.0

This table compares out-of-sample performance (RMSPE) of three model specifications—scalar-on-scalar, scalar-on-composition, and scalar-on-density—using daily maximum and minimum temperatures versus daily mean temperatures. We employ leave-one-year-out cross-validation: each year is held out as the test set while the model is trained on all other years.

Table A3: Observed and projected outmigration (millions of people) under the SSP2-4.5 scenario for each model

Region	Observed	Projected (2030)	
		Scalar-on-scalar	Scalar-on-composition
Africa	0.65	0.69	2.13
Americas	1.40	1.41	0.93
Asia	2.79	2.74	3.81
Europe	2.58	2.61	1.20
Oceania	0.08	0.07	0.11
Total	7.49	7.52	8.18

Notes. This table reports observed outmigration (millions of people) in 2019 and projections in 2030 under the SSP2-4.5 scenario, using both the scalar-on-scalar and scalar-on-composition specifications.