Border crossings between 243 countries: The Global Transnational Mobility Dataset 2.0, 1995-2022*

Tobias Grohmann[a], Ettore Recchi[a,b], Luca Bernasconi[a]
[a] EUI, Florence, Italy; [b] Sciences Po, Paris, France
Correspondence: tobias.grohmann@eui.eu

This version 6 February 2025

Abstract

This paper introduces a new dataset that estimates the volume of human travel across country borders worldwide between 1995 and 2022. It builds and expands on pioneering work that presented estimates for 2011 to 2016 (Recchi et al., 2019). The dataset enables the study of the volume, directions, and changes in global human mobility. Our estimates reveal that total transnational mobility increased from 4.87 billion trips in 1995 to 9.64 billion in 2019, largely outpacing global population growth. Across the board, international migration constitutes a tiny fraction of transnational travel (less than 1% worldwide and as low as .15% in Europe). The rise of transnational mobility has been particularly sustained in East and South-East Asia. This region was, however, also the hardest hit by Covid-19 travel restrictions and their aftermath, which brought its flows in 2022 back to mid-1990s levels. Most border crossings are intra-regional, especially in Europe. Despite the widespread growth in volume, the global network of cross-border mobility has not significantly changed its overall configuration around nine major clusters in more than a quarter of a century. Germany stands out as the main hub in Europe and globally, followed by the US and China. However, some regional mobility clusters have split and others have merged, with individual countries shifting between clusters. The dataset may be used to study global-level phenomena in fields such as migration and tourism studies, sustainability, epidemiology, international economics, and international relations.

Keywords: human mobility; tourism; migration; transnationalism

* This research is part of the Global Mobilities Project based at the Migration Policy Centre (MPC) of the European University Institute. It is funded by the European Union through the MIGMOBS ERC Advanced Grant (101097240, ERC-2022-AdG).

1 Introduction

The globe is on the move. The size and scope of this phenomenon–which appears as one of the hallmarks of our age (Recchi & Safi, 2024)–have been extensively documented in numerous migration-related publications by international organizations, such as the OECD Migration Outlooks (e.g., OECD, 2024), and independent studies (e.g., Abel & Cohen, 2019; Akbaritabar et al., 2024; Raymer et al., 2022). This constant rise has also witnessed spikes of migration and asylum-seeking tied to events such as the wars in the Balkans, Syria, Sudan, and Ukraine, the political and economic downturns in Venezuela, the European enlargement to the East, growing demand of foreign workers in the Gulf States, among others. Simultaneously, there has been a significant increase in non-migratory cross-border mobility, much larger than migration-related mobility (Recchi et al., 2019). For instance, soaring numbers of tourists descend upon cities worldwide; seasonal workers especially in agriculture, hospitality, and construction, move abroad for short-term jobs; digital nomads have entered the international scene, driven by technological progress, changing work paradigms, and significant cost of living differences; and medical tourists travel internationally to access affordable, specialized, or advanced healthcare. Against this backdrop, the Covid-19 pandemic in 2020-21 led to an unprecedented crash in cross-border mobility, affecting countless individuals worldwide.

This paper aims to estimate the number of trips across each country-to-country border worldwide from 1995 to 2022. The Global Transnational Mobility Dataset 2.0 (GTMD2.0), created as an outcome of this research, facilitates the study of the volume, directions, and changes in cross-country human mobility worldwide.[1]

Our estimates indicate that total global transnational mobility amounted to 4.87 billion trips in 1995, rising without interruptions (except during the Covid-19 period) to 7.22 billion in 2022, and having peaked just before the pandemic to 9.64 billion. In comparison, the world population increased from 5.7 billion in 1995 to 7.9 billion in 2022, suggesting that mobility has grown faster than the global population. On average, 56.7% of all border crossings are to European countries, the largest share by far, followed by Eastern and South-eastern Asia at 10.9%. Oceania has the smallest share, with less than 1%.

Most border crossings are intra-regional, with Europe having the highest share of intra-regional mobility at 93% of all the trips reaching a European country. Only Northern America and Oceania exhibit higher inter-regional mobility than intra-regional mobility. International

[1] Like in Recchi et al. (2019), we adopt the term 'transnational' as used in international relations, to describe any movement by *non-state* actors that spans across national borders, thus being distinct from 'international', which refers to *state* actors (Nye & Keohane, 1971). We are also aware that in migration studies 'transnational' tends to imply movements of the *same* individuals across borders (Wimmer & Glick Schiller, 2002), which is not at all the case in the context of our work.

migration, defined as cross-border residence changes, constitutes only a small fraction of total border crossings. Central and Southern Asia has the highest share at 1.5%, followed by Oceania. In other regions, migration accounts for less than 1% of total mobility, with Europe having the lowest share at .15%.

The potential utility of this dataset extends beyond migration and tourism studies to other research areas such as epidemiology, sustainability, international economics, and international relations. For instance, to study globalization, our estimated border crossings could be compared to other flows, such as trade and capital flows.

This paper expands historically and methodologically an earlier work that devised an integrated estimate of country-to-country cross-border human mobility using global statistics on tourism and air passenger traffic (GTMD1.0, Recchi et al., 2019). That pioneering dataset, covering trips from 2011 to 2016, provided an unprecedented and comprehensive resource on transnational human mobility worldwide. It combined tourism and air traffic data to test their relative contributions and corrected for their limitations, resulting in a robust dataset that has been instrumental in understanding global mobility patterns. The present article considerably extends the annual coverage from 2011-2016 to 1995-2022 and updates the concepts and methods used in that pioneer paper.

Among the contributions closest to ours is the study by Llano et al. (2023), which estimates bilateral tourism flows among 74 countries from 1995 to 2018. Their analysis differentiates between one-day excursionists and overnight tourists and provides estimates of domestic tourism flows stratified by mode of transport. Our work differs in several ways: First, our objective is to estimate all types of mobility between countries, not just outgoing tourism flows. This means that our approach includes not only visits by non-resident tourists but also their return journeys and border crossings by migrants, i.e., trips of people who change their country of residence. Second, the data records on which we base our estimates do not only include tourism and migration flows but also data on air travel passenger volumes, which we use to validate and complement the estimates obtained from the tourism data. Third, our dataset has a more global outlook, including as many countries as possible. While Llano et al. (2023) estimate flows between 74 countries, our final dataset includes flows between 243 countries.

2 Definitions and conceptual framework

2.1 Who are international travelers?

In this section, we define broad categories of international travelers as included (or not) in our dataset. A traveler is a person who moves between different geographic locations for any purpose and duration (UNDESA, 2010, para. 2.4). Hence, if the locations are separated by international borders, the person moving between them is an international traveler. In our approach, we distinguish between two kinds of international travelers–visitors and migrants–depending on whether their border crossing also represents a change of country of usual residence.

Visitors. As visitors, we understand all people who travel across international borders *without* simultaneously changing their country of usual residence. These include both same-day and overnight visitors, whatever their reason for travel. Reasons may include holiday, leisure, and recreational; business and professional; visiting friends and relatives; education and training; health and medical care; religion and pilgrimage; shopping; transit; adventure, or any other subjective reason humans may attribute to their spatial mobility. This definition of visitors is similar but not entirely congruent with the one in the UN's Recommendations for Tourism Statistics (UNDESA, 2010, para. 2.41 ff.). For example, in contrast to the UN, we also include international commuters who live in one country but regularly travel across international borders to reach their workplace without changing their usual residence. Think of employees whose employer deploys them to work in a country other than their country of residence or cross-border workers who commute daily or weekly across international borders to their workplace. The latter category of international travel is more common in free-movement areas, such as the European Union, where non-residents can participate in a member state's labor market based on their EU citizenship rights. Cross-border commuting is less common–albeit not absent–in other world regions.

Migrants. As migrants, we understand all people who travel across international borders *and* change their place of usual residence by doing so. We sometimes refer to them as "residence changers" or "future settlers". This definition is similar to the UN's Recommendations on Statistics of International Migration (UNDESA, 1998), which additionally distinguish between long-term and short-term migrants.[2] Yet, our approach differs from most conventional categorizations as we do not differentiate between people by their *motive* of residence change. These motives may include work, study, family reunification, and threats to their life and health at home. Hence, this group of residence-changing international travelers includes all those commonly referred to as permanent migrants, refugees, temporary labor migrants, seasonal workers, and digital nomads.

Finally, let us define the notions of origin and destination country. We will adopt a country of residence perspective, as opposed to a country of birth or citizenship perspective, to determine

[2] According to the UN, migrants are international, long-term migrants when they move abroad for longer than 12 months, and their destination country effectively becomes their new country of usual residence. Short-term migrants are people who move abroad for less than 12 but more than 3 months. For the period abroad, their destination country becomes their new country of usual residence. So, this distinction is irrelevant for our approach as our criterion to differentiate migrants from visitors is the change of residence.

the origin country of a traveler. This means that a mobile person's current country of residence is their country of origin, as they are currently part of that country's population. Yet, the origin country of a traveler is not necessarily the same as the origin of an individual trip. So, imagine an A-country resident traveling between countries B and C. While country A is their origin country, country B is the origin of their current trip. However, most available migration and refugee statistics report the number of immigrants and refugees based on a country of birth or citizenship perspective. This has to do with how countries conduct population censuses, which is the typical source for international migration statistics, where migrant stocks and flows are usually reported in terms of foreign-born residents or foreign-born arrivals (OECD International Migration Database 2024; UNDESA International Migrant Stock 2020).

2.2 The composition and measurement of transnational mobility flows

Methodologically, our work is based on the idea that there are two main approaches to measuring the number of international border crossings. The first relies on administrative data about arrivals collected by receiving countries according to the broad categories outlined above. The second relies on passenger volumes in all possible modes of cross-border transport. We use elements of both to generate our estimates and discuss them in turn.

2.2.1 Trips as measured by type of traveler

Most receiving countries classify incoming travelers according to the above categories and keep records of these arrivals. Thus, in principle, the number of incoming persons from the different registries could be added to compute the annual total of trips between countries.

One difficulty with this approach is that receiving countries register arriving persons at different places and store this information in different registries. Typically, there is not one single reception desk for check-in upon arrival in a country. For instance, tourist visitors may be registered through their accommodation; cross-border workers may only be registered through their work permits; immigrants are registered by the respective government agencies (further distinguishing asylum seekers, for instance), and so forth. This means, however, that if one wanted to add up all arrival categories, first, one would need data from all the different registers, and second, these registry data would have to be comparable between countries (i.e., countries should classify arrivals based on the same criteria. Yet, although the UN issues recommendations on how to count tourists and other visitors (UNDESA, 2010), immigrants (UNDESA, 1998), and refugees (EU & UN, 2018), there is heterogeneity in how member states implement these recommendations. In particular, there is quite some difference in how receiving countries classify temporary and circular migration (e.g., ILO, 2022).

Notwithstanding these practical considerations, in principle, arriving visitors and migrants could be summed up to yield the total amount of trips between countries. We note that the persons traveling through the ij-corridor could be i-residents leaving their country of residence i, or j-country residents returning from i to their country of residence j, or k-country residents who transit through country i on their way to their destination j. Hence,

$$trips_{ijt}^{A} = out_{i,ijt} + return_{j,ijt} + transit_{k,ijt} \tag{1}$$

Where $out_{i,ijt}$ stands for outgoing trips carried out by i-country residents, $return_{j,ijt}$ stands for return trips carried out by j-country residents, and $transit_{k,ijt}$ stands for transit trips carried out by third-country k residents.

More specifically, out_{ijt} includes travelers from both broad categories introduced above: i-resident **visitors** on their way to country j and **migrants** who change their residence from country i to country j. In contrast, the return trips $return_{ijt}$ only include j-resident **visitors** returning from their stay in country i. So-called "return migrants", i.e., j-born individuals who return from their current or temporary country of residence i to their country of birth, are not present in the return trips. This is because migrants are residents in their host country and, therefore, are counted in the outgoing trips of i-residents. While they are called "return migrants" from a country-of-birth perspective, they are simply migrants, i.e., repeated residence changers, from a country-of-residence perspective. Finally, the transit trips, $transit_{ijt}$, include k-resident **visitors** and **migrants** who transit through country i on their way to the final destination of their current travel in country j. We will not consider this kind of travel in our dataset. Transiting travelers usually represent only technical border crossings, e.g., layovers at international airports, involving next to no meaningful engagement with a country or society.

Given these considerations, in this paper we use the following decomposition to operationalize our estimate of total cross border trips between all countries of the world:

$$trips_{ijt}^{A} = visitors_{i,ijt} + visitors_{j,ijt} + migrants_{i,ijt} \tag{2}$$

2.2.2 Trips as measured through transport volumes

The second approach to measuring border crossings relies on the volumes of passengers in various modes of transport. Typical forms of transport include plane, car, rail, and ship traffic, but in some border regions, crossings on foot, bike, etc., are also not entirely uncommon. So, if one had records on incoming passengers, one could simply add up their numbers across all modes of transport to arrive at the total volume of trips between two countries.

$$trips_{ijt}^{B} = air_{ijt} + land_{ijt} + water_{ijt} \tag{3}$$

Where air_{ijt} represents the volume of passengers crossing international borders between country i and country j by plane; $land_{ijt}$ those in cars, trains, and other land-based modes of transport; and $water_{ijt}$ all those traveling in ships and other water-based vessels.

One feature of this approach is that it does not allow us to distinguish between the two categories of transnational travelers discussed above (e.g., visitors and migrants). Moreover, this approach does not allow for distinction between outgoing, return, or transit trips if the passenger's records do not include the residence country of the traveling person.

2.3 Strategy to estimate transnational mobility

Our strategy to estimate total border crossings between countries relies on the assumption that, in principle, the two ways of measuring the number of international border crossings yield the same amount of border crossings. So, our procedure follows two main steps: first, we estimate $trips^A$ and $trips^B$ using three of the largest existing records of border crossings; second, we combine them into one coherent measure of transnational mobility. Figure 1 illustrates our workflow. Here, we briefly outline its main building blocks and discuss details in the following sections.

Our three data sources are:

1. Inbound tourism, 243 origin and 211 destination countries, 1995-2022, UN Tourism Data, World Tourism Organization (UNWTO).
2. Migration flow estimates, 230 origin and destination countries, 1990-2020, Guy Abel.[3]
3. Air passenger volumes, 254 origin and destination countries, 2010-2023, Market Intelligence Data, Sabre.[4]

We limit our timeframe to the 1995-2022 period, for which we maximize data availability from the three sources. To unlock on the full potential of the Sabre data, we back-cast its information on a yearly basis to cover the 1995-2009 period. In later sections we provide details and document on the robustness of this operation.

[3] Abel's migration estimates include refugee flows. More specifically, the UN migrant stock data, which is used to estimate migration flows, is adjusted to include refugee statistics.
[4] A more detailed description of each of these records follows in the sections below.

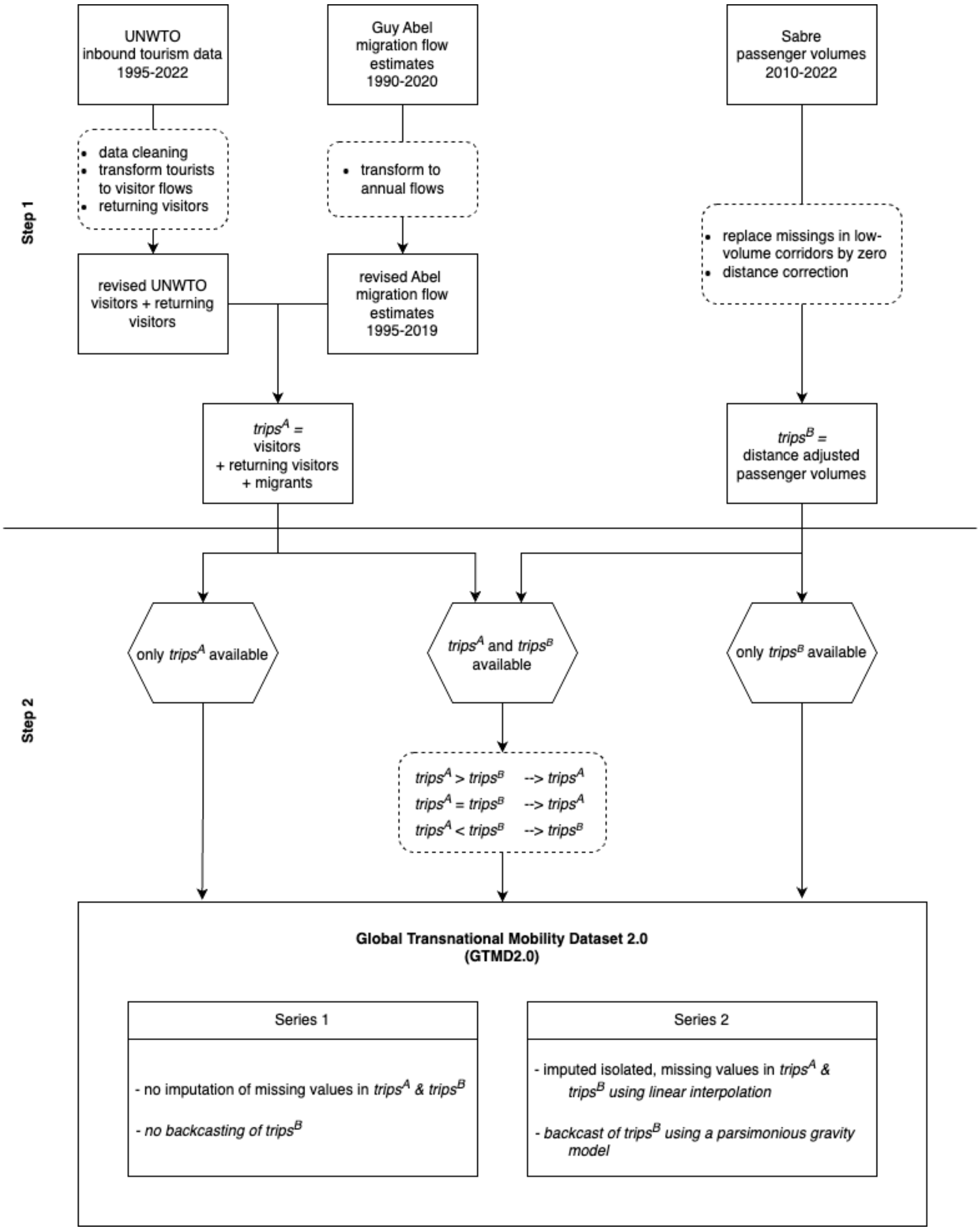


Figure 1 Procedure of dataset assemblage

In the first step, we revise the three data records such that they match our grid of origin-destination-year cells for all existing countries since 1995. Moreover, in the UNWTO data, which is reported in different tourism categories (e.g., arrivals at national borders, arrivals in hotels, arrivals by non-nationals vs non-residents), we generate a unified metric of 'non-resident visitors arriving at national borders'. Also, we add the same volume of visitors traveling in the ij-direction to those traveling in the ji-direction to approximate returning visitors. In the migration data, which is reported in 5-year intervals, we compute annual migrant flows by dividing by five. Finally, we add visitors, returning visitors, and migrant flows to obtain an estimate of $trips^A$.

Then, to estimate $trips^B$, we generate distance-adjusted air passenger volumes. The reasoning is that passenger volumes on airplanes will exactly approximate the number of travelers between countries the further they are apart; the closer these countries are, the more travelers will use other modes of transport. So, using the cases in which visitor data and air passenger data are available, we regress visitors on passengers and distance (and other accessory geographical information, such as the existence of land borders between countries), and use the coefficients to adjust the air passenger volumes to represent all modes of transport in corridors that are not only accessible by plane. This 'distance' adjustment yields $trips^B$.

In our second step, we combine $trips^A$ and $trips^B$ to create our dataset of global mobility flow estimates. We assume that, in principle, the two ways of measuring the number of international border crossings yield the same amount of border crossings. Therefore, we use the equality

$$trips^A_{ijt} = trips^B_{ijt} \tag{4}$$

to validate our $trips^A$ and $trips^B$ estimates, as well as to fill in values whenever $trips^A$ are missing. In some cases we only have $trips^A$ (12%), in others only $trips^B$ (73%), and in the rest, we have both (15%).[5] In cases we have only one of the two, we use this as an estimate of global transnational mobility. If we have both, we pick the record that yields the higher total volume in a corridor, assuming both will underestimate actual mobility flows. This maximizes the total amount of country-to-country estimates of human transnational mobility.

3 Data Records

3.1 Visitor flows

To represent bilateral flows of non-resident visitors, we use data from the UNWTO's *Tourism Data in Excel Format–Data from 1995 onwards* (UNWTO, 2024).[6] More specifically, we use the data on *Inbound Tourism*, which reports tourist arrivals in destination countries annually between 1995 and 2002. The data record is a collection of 219 Excel files, each corresponding to one destination country. Among other indicators, these destination country-specific files contain the annual volume of arrivals by origin country, which we use as the bilateral visitor flows in Equation (2).

---

[5] The percentages are based on Series 1 of our estimates. See details below. Using Series 2, we have only $trips^A$ in less than 1% of cases, only $trips^B$ in 84% of cases and both in 15%.
[6] This dataset is not openly accessible but must be purchased from the UN World Tourism Organization. We purchased and downloaded the data in May 2024.

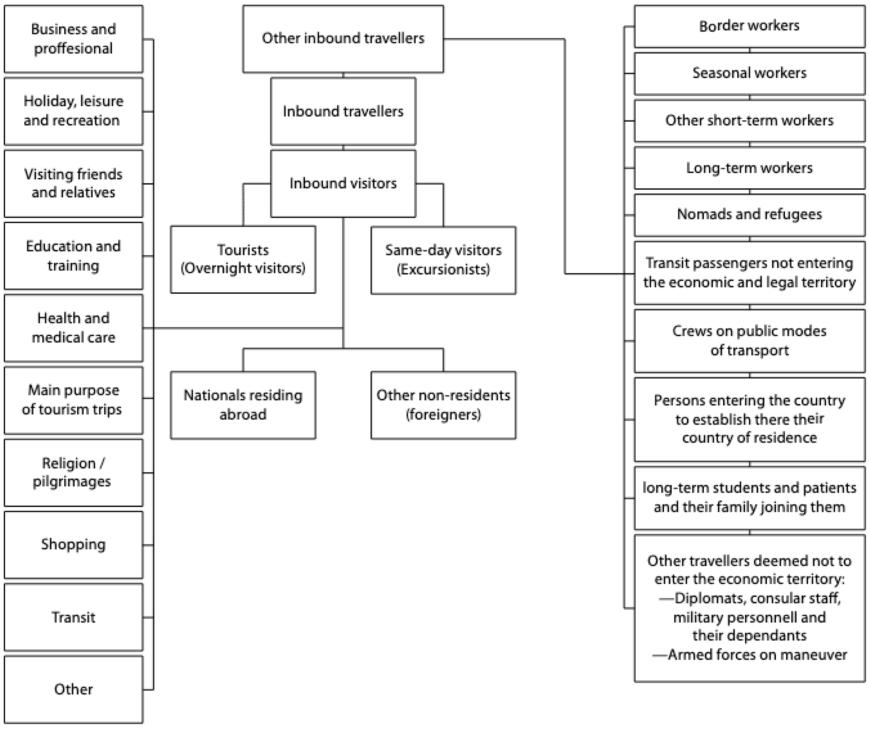


Figure 2 Classification of Inbound Visitors. Source: International Recommendations for Tourism Statistics, UNDESA (2008). Typos from original source.

Arrivals represent the flow of international visitors to the reference country, where each arrival represents one inbound tourism trip. When a person visits multiple countries within a single trip, the arrival in each country is documented separately; hence, within a given accounting period, the number of arrivals may exceed the number of unique travelers; for example, if a person visits the same country multiple times in a year.

In this data record, inbound visitors refer to overnight and same-day foreign visitors who travel for specific reasons, such as holidays, business, visits to friends and relatives, etc., as specified in the classification of inbound visitors according to the UN's International Recommendations for Tourism Statistics (UNDESA, 2010) (see Figure 2). Other categories of travelers—such as border, seasonal, and short-term workers—are excluded, as they do not fit the definition of visitors. Thus, the UN Tourism Data does not completely match our definition of visitors in Section 2.1, which is much broader and includes, for example, commuting workers who operate in a foreign country without this country becoming their usual residence. Yet, as suitable data on short-term cross-border workers is notoriously difficult to obtain (ILO, 2022; OECD, 2019), we do with the UN Tourism Data and contend that our estimates will at least represent a lower bound of international border crossings.

Finally, the reference countries in the UN Tourism Data record arrivals differently: some through border surveys, some through immigration offices, and some through accommodation surveys. Moreover, some countries count only overnight visitors but not same-day excursionists, and some define foreigners by nationality rather than residence. Overall, this results in eight arrival categories reported in the source data (see Table 1). As the visitor flow volumes cannot be compared across categories, we convert the arrival categories into one statistic, thus creating a unified measure of inbound visitors. Details are described in Section 4.2.

3.2 Migrant flows

To represent bilateral migration flows, i.e., number of residence changes, we use data from Guy Abel's *Bilateral international migration flow estimates for 200 countries* (Abel & Cohen, 2019).[7] This data record provides estimates of bilateral migration flows between 230 countries in six 5-year periods (1990-95, 1995-2000, ..., 2015-2020). To obtain annual data, we divide the 5-year flows by five and distribute them over the respective years. This means that we end up with annual migration flows from 1990 to 2019.

We use these estimates rather than actual migrant flow counts because there is simply no dataset of international migrant flows with the country coverage appropriate for our project. Many countries do not publish data on bilateral migration flows and the countries that collect such data use differing criteria for defining migrants, preventing detailed cross-country comparisons.

To address these problems, the estimates in Abel's database are based on bilateral migrant stocks from the UN's International Migrant Stock dataset (UNDESA, 2020). Migrant stock data measure the number of migrants living in each country at a given time by country of birth. Such measurements are reported regularly by all UN countries and much easier to collect than flows, as only questions about place of birth need to be asked in a country's census, population registers, or administrative data collection systems. The migrant stocks are adjusted to include refugee populations (UNDESA, 2020, p. 4f.).

'Demographic accounting', the estimation method used by Abel (Abel, 2013; Abel & Cohen, 2019; see also Azose & Raftery, 2019), transforms the birthplace-specific migrant stock data into flows between residence countries. This allows to distinguish between out-, return- and transit-migration. These estimates capture migrants as residence-changers and, therefore, correspond well with the definition of migrants we use in this paper. Each incoming migrant counts as one border crossing.

[7] This dataset is openly accessible via https://guyabel.com/publication/bilateral-international-migration-flow-estimates/, where the author also provides recent updates and extensions. We used Version 7 (October 2024).

As will become apparent later, the total number of migrants represents only a fraction of total border crossings. We like to emphasize that this is not an artefact of our choice of migration estimates. In fact, Abel's migration flow estimates yield the highest total annual flows compared to other migration estimates and are, on average, higher than actual, counted migration flows (Abel & Cohen, 2019 and updates on Guy Abel's website; Grohmann, 2024, Chapter 2).

3.3 Passenger volumes

To represent air passenger flows, we use the Sabre Market Intelligence dataset from 2010 onward.[8] Sabre is a private company providing market analyses to the travel industry; its Community Portal includes detailed information originating from airline reservation systems and airports reporting the numbers of passengers on commercial flights. We aggregate the data at the country and year level to obtain annual country-to-country flows of passengers that match our origin-destination-year grid. We obtain 642,679 non-missing observations in the 2010-2022 period from the Sabre dataset.[9]

The Sabre Market Intelligence dataset provides the number of passengers between embarkation and disembarkation points for flights under the same flight number. This means it includes traffic statistics for the origin and final destination city-pairs without reporting intermediate stops. For instance, passengers flying from Florence (Italy) to Perth (Australia) with a layover in, for example, Dubai (UAE) appear only once in this dataset, not twice. This feature makes the dataset useful for describing mobility flows between more distant country pairs and, therefore, is ideal for the objectives of this paper. Moreover, it also aligns well with our decision not to include transit visitors in the dataset.[10]

---

[8] Details can be found here https://www.sabre.com/products/suites/pricing-and-revenue-optimization/market-intelligence/. We also accessed ICAO data, another source of airline traveler volume information which provides data collection from 1995. We eventually decided to rely on Sabre only, given its higher data granularity and completeness.

[9] Sabre does not report zero passenger volumes when there are no passengers. Instead, the respective corridor-year observation is left empty. We identify corridors with missing observations over time that show low passenger-volumes (either when the total for the years 2010-2022 is <= 1,000 OR when preceding and succeeding values are <= 100). We assume that the missing values in such corridors are, in fact, zero and, consequently, replace them with zeros.

We do not apply this method to the other data records. In the UNWTO data, it is not always clear whether missing values are indeed missing or whether they indicate zeros. For instance, there are missing values that undoubtedly indicate a missing record (e.g., all UNWTO visitor flows Israel to Portugal between 2001 and 2019), whereas others may very likely indicate zero values (e.g., Fiji to Portugal since 1995). In the Abel migration data, zero values indicate corridors with no migration flows, while missing data indicate observations that are not covered by that dataset.

[10] We need to acknowledge that travelers between distant city pairs requiring a layover can be on separate flights with different flight numbers. This can happen when there is no code sharing on flights. Generally, the likelihood of different flight numbers increases when two or more flight companies are involved, or the carriers are not part of the same airline alliance.

4 Methods: Constructing the dataset

4.1 The origin-destination-year grid

We place each of the data records into a matrix whose cells are given by all origin-destination country combinations for all years between 1995 and 2022. Each cell can be identified by f_{ijt}, where the indices stand for origin country (i), destination country (j), and year (t). In line with our earlier considerations, origin and destination represent the beginning and end point of a cross-border movement; in particular, the origin country does *not* refer to the country of birth, citizenship, or ancestry of the people who cross borders.

The grid is organized as a panel of unidirectional corridors between countries. Each direction between countries A and B, i.e. the $A \rightarrow B$ corridor and the $B \rightarrow A$ corridor, constitutes a separate panel.

Throughout the paper we refer to "countries" as our units of analysis. To designate a country or area, we use the "statistical units" defined by the M49 standard of the UN Statistics Division. The three-digit numerical codes associated with each unit constitute a set of mutually exclusive statistical units each year. Our dataset contains 243 such countries. While 195 of them are sovereign states recognized by the UN, the others have more complex political status. Since 1995, several countries have come into existence or ceased to exist in their previous form and, therefore, entered or exited the UN universe.[11] So, not all 243 countries are present through all years and, consequently, the grid is non-symmetric across years as states or territories enter and exit (if newly created or dissolved). Accounting for these entries and exits, our grid has 1,542,430 corridor-year cells in the 27 years between 1995 and 2022. Appendix A1 provides a detailed overview of all countries in the dataset, including their start and end points.

Placing the data records into this grid requires several steps of data cleaning and harmonization. Details are given in Appendix A2.

4.2 A unified measure of international visitors

As a next step, we generate a unified measure of international visitors from the UNWTO dataset. This is necessary because the reporting destination countries use different ways of categorizing arrivals at their borders, which are not readily comparable. So, inbound tourism figures reported by destination countries differ, depending on (i) whether foreigners are defined by nationality or by country of residence, (ii) whether the statistic refers to tourists or visitors, (iii) whether travelers are counted at national borders or in accommodation establishments, (iv) whether accommodation establishments include hotels and similar or also other forms of accommodation. This yields in total eight different arrival categories (Table 1).

[11] For recent changes, see the UN Statistics Division https://unstats.un.org/unsd/methodology/m49.

Table 1 Tourist arrival categories: definitions

code	label	Description
112	TFR	Arrivals of non-resident tourists at national borders, by country of residence
111	TFN	Arrivals of non-resident tourists at national borders, by nationality
122	*VFR*	*Arrivals of non-resident visitors at national borders, by country of residence*
121	VFN	Arrivals of non-resident visitors at national borders, by nationality
1912	THSR	Arrivals of non-resident tourists in hotels and similar establishments, by country of residence
1911	THSN	Arrivals of non-resident tourists in hotels and similar establishments, by nationality
712	TCER	Arrivals of non-resident tourists in all types of accommodation establishments, by country of residence
711	TCEN	Arrivals of non-resident tourists in all types of accommodation establishments, by nationality

Notes: The numerical code is from the UNWTO inbound data files; label and description are taken from World Tourism Organization (2023) Methodological Notes to the Tourism Statistics Database, page 213.

As the UNWTO arrival categories are not comparable–for instance, the number of tourist arrivals in hotels does not represent the total number of border crossings–we transform all traveler counts into one category, the VFR (visitor) category. This matches our definition of border crossings by non-residence-changing international travelers.

Our transformation procedure consists of two main steps, as shown in Figure 3. The first step is based on the simplifying assumption that a person's country of citizenship is the same as their country of residence. This assumption allows us to reduce the eight arrival categories to four: VFR, TFR, THSR, and TCER.

The following step transforms traveler counts from the TFR, THSR, and TCER categories into VFR counts. To do so, we use the predicted values from a Poisson regression, in which we regress the VFR counts on the counts of the respective other arrival category k in the subsamples in which both the VFR counts and the category k counts are available. The exact procedure is described in Appendix A3.

This marks a departure from Recchi et al.'s (2019) GTMD1.0 which focused on the TFR (112, tourist) category as the preferred measure and left the arrival counts untransformed. As expected, our visitor flows are substantially larger in volume compared to the GTMD1.0-style tourist flows because they capture a broader set of border crossings (Appendix A3, Table A4).[12]

[12] Llano et al. (2023) also transform arrival counts to obtain comparable figures. However, their methods differ from ours. First, they base their transformations on the outbound data files of the UNWTO dataset. Second, they focus on both TFR and VFR as their target categories, while we are only interested in VFR counts.

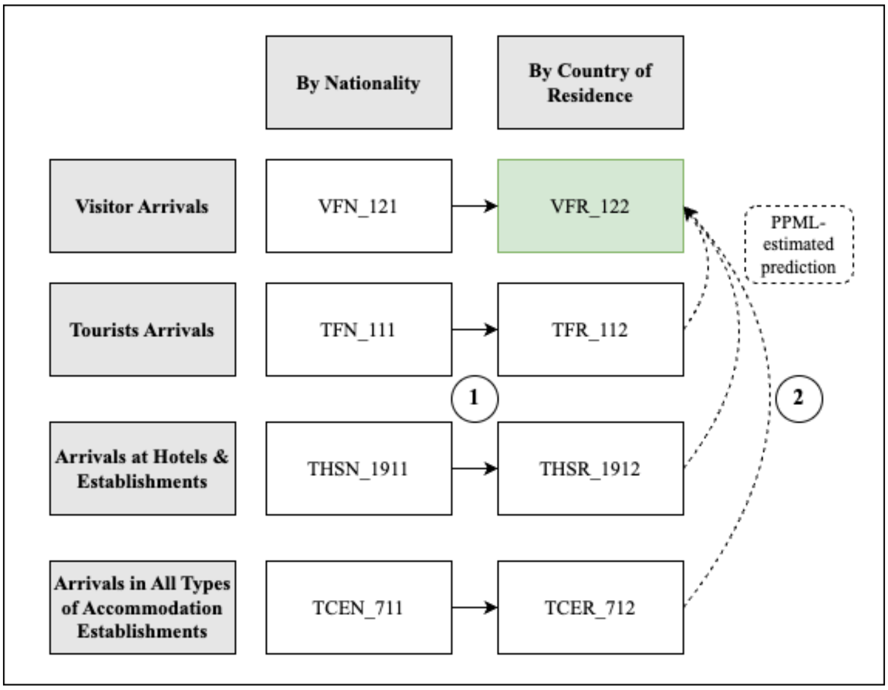


Figure 3 Conversion of tourism categories to non-resident visitors

4.3 Returning visitors

As discussed earlier, all border crossings also include visitors on their return journey home (cf. Equation (1)). Yet, the UNWTO data represents border crossings in one direction only, namely as arrivals of country i residents at the borders of country j. In line with our definition, we assume that every arriving visitor will return to their country of residence within the same year (see also Recchi et al., 2019). Thus, we generate returning visitors from our current UNWTO visitor flow data by reversing every incoming ijt-flow and adding that number to the corresponding jit-flow.

However, we add visitors and return visitors in a corridor-year only if both countries report tourism data. This has two related effects. First, it makes the visitors dataset symmetric with as many origin countries as destination countries. Second, it shrinks the dataset considerably. So, instead of the 391,195 non-missing observations we obtained from the transformations in the previous section, we now have 184,232 non-missing observations (see Table A4 in Appendix).

4.4 Adding migration data

In line with our general view of mobility, following Equation (2), we add Abel's migration flow estimates to the UNWTO visitor + returning visitor flows to obtain $trips^A$. Recall that the Abel data captures flows of people who change their place of usual residence–no matter their motive for changing residence. So, the data includes both voluntary and forced migration.

Abel's migration data covers a larger set of corridor-year observations than the visitor+return visitor flows ($visitors_{i,ijt} + visitors_{j,ijt}$) created in the previous step. However, not all corridor-years have corresponding migration data: first, the annual migration data goes only until 2019 (20,498 missing observations); second, some countries (e.g., Taiwan) are in our dataset but not in Abel's (1,263 missing observations). Yet, since the migration flow volumes are magnitudes smaller compared to the visitor flow volumes (e.g., <1% in 2018; see Table A4, Appendix A3) we use only the visitor flows to approximate $trips^A$ when the migration flows are missing.

Across all origin-destination-year combinations, there is a positive but relatively weak relationship between migrant flows and outgoing visitor flows (correlation coefficient of $\rho = 0.23$). This may be because typical immigration countries (i.e., migrant receiving countries) are usually large senders of international travelers such as tourists, excursionists, and other mobile populations. So, when restricting the sample to specific destination countries, the correlation is higher. For instance, the correlation is $\rho = 0.64$ across all incoming corridors to the United States in 2018.

This concludes our process of generating $trips^A$, the measure of border crossings based on the type of travelers who cross international borders. We now turn to $trips^B$, which uses data about the mode of transport to measure total border crossings per year.

4.5 Distance-corrected air travel passenger volumes

The volume of passengers on commercial flights must be taken with care as an indicator of human mobility. Air travel is notoriously more common the farther the distance travelled. To equate it to the number of transnational visitors, including those who travel on land or water, we correct the raw number as a function of travel distance between countries.

Following Equation (3), $trips^B_{ijt}$ should be the sum of passengers traveling from i to j using means of transport in the air, on land, and water. Since we do not have passenger volumes on land and water, we estimate $trips^B$ by implementing a geographic distance correction on the air passenger volumes. The idea is that air passenger volumes should be the same as visitor volumes in corridors between countries that are far apart. This is because faraway countries are almost exclusively accessible via airplane to most travelers, while closer countries can be reached using land or water-based means of transport, too. Indeed, the greater the distance between countries, the more similar air passenger volumes and $trips^A$ are (Appendix A4, Figure A2).

This intuition was first used by Recchi et al. (2019), who proposed a distance-correction factor that maximizes the correlation between tourists and air travel passenger volumes. In this paper, we develop a regression-based approach which allows to account for multiple factors–

not just geographic distance–to distance-correct the air passenger volumes. More specifically, we assume that several geographic factors determine the accessibility of country j from country i and thereby determine which share of $trips^A$ $trips^B_{ijt}$ is taken via air, land, or water. Thus, the number of passengers between i and j can be modelled as:

$$trips^B_{ijt} = \phi_{ijt}(air_{ijt} + land_{ijt} + water_{ijt}) \tag{5}$$

where ϕ_{ijt} represents the accessibility of country j from country i using air, and, or water-based means of transport.

As mentioned earlier, we do not have data on the volume of passengers by $land_{ijt}$ and $water_{ijt}$. Yet, in a large number of cases we have $trips^A$, which include not only air_{ijt} but also $land_{ijt}$ and $water_{ijt}$. Consequently, our strategy is to determine the share of air passengers, air_{ijt}, in these cases, *conditional* on the accessibility factor ϕ. So, leaning on the equality that $trips^A_{ijt} = trips^B_{ijt}$, we estimate the following equation using the Poisson pseudo maximum likelihood estimator:

$$trips^A_{ijt} = air^{\beta 1}_{ijt} * \phi^{\beta_2}_{ijt} \tag{6}$$

We then compute the predicted $\widehat{trips}^A_{ijt}$ from this estimation and use them as distance-corrected air passenger volumes. We test different empirical specifications and find that the one that operationalizes ϕ including distance, contiguity, landlocked-ness, and being an island state correlates the strongest with $trips^A$ without compromising the number of observations we can predict from it (see Appendix A4).

Overall, the correction increases the passenger volumes per corridor (cf. means, min, max, totals; Appendix A4, Table A6). However, the procedure also corrects some values downward (see Appendix A4, Figure A3). We use the corrected values when they are larger than the uncorrected values (634,886 observations) and *vice versa* (7,765 observations). In 720 of 72,361 of corridors (~1%) we have two sources over time, meaning that in these corridors we may have either corrected or uncorrected values from one year to the other.

4.6 Final estimates of border crossings

The distance correction completes our process of estimating $trips^A$ and $trips^B$. The next step is to combine $trips^A$ and $trips^B$ to generate our final estimates of border crossings. In some cases, we only have $trips^A$, in others only $trips^B$, and in others we have both. When we have only one of the two, we use this as an estimate of border crossings. If we have both, we pick the record that yields the higher total volume in a corridor, assuming both will underestimate actual mobility flows.

Eventually, we will generate two different series of border crossing estimates. These two series treat missing values differently: Series 1 uses the $trips^A$ and $trips^B$ estimates as described above without missing data imputation; Series 2 uses $trips^A$ and $trips^B$ estimates after imputing isolated missing values *and* after back-casting $trips^B$ to generate distance-corrected air passenger volumes from 1995 to 2009.

To generate Series 2, we first impute isolated missing values in $trips^A$ and $trips^B$ by linear interpolation with a maximum distance of two years. We do not impute in the years 2020-2022, as the pandemic may have substantially driven down the numbers, and the assumption of continuity behind our interpolation may be unfounded. Through the interpolation, we add 3,898 observations to $trips^A$ in 1995-2019 and 306 observations to $trips^B$ in 2010-2019.

The second imputation method concerns the missing values of $trips^B$ in 1995-2009. We impute these by back-casting from the non-missing values in the 2010-2019 period. The back-casting is done by estimating gravity models of $trips^B$ in the 2010-2019 period and then 'backwards predicting' its values to the years in the 1995-2009 period using the coefficients from the gravity estimations. Through back-casting, we add 631,918 observations to $trips^B$ in 1995-2009. Details of the procedure can be found in Appendix A5.

Table 2 documents the composition of our final border crossing estimates by source. As $trips^B$ are only available since 2010, Series 1 is composed solely of $trips^A$ from 1995 to 2009, and from 2010 to 2022 it is composed of either $trips^A$ or $trips^B$, depending on which is available or which is larger. In contrast, the back-casting procedure also generates observations of $trips^B$ before 2010. So, Series 2 is composed of either $trips^A$ or $trips^B$ in all years between 1995 and 2022. Moreover, since the corridor coverage of $trips^B$ is much larger, Series 2 will also have better coverage due to back-casting. Finally, while $trips^A$ only covers less than 5% of all corridor-year observations, it nonetheless supplies around 40% of the *volume* of transnational border crossings in our dataset. So, $trips^A$ is integral to cover the corridors that are large in volume. We offer Series 1 alongside Series 2, which contains our main estimates, to make our methodological decisions more transparent.

Table 2 Composition of Series 1 and Series 2 estimates

Panel A: Corridor-year statistics

	Series 1		Series 2	
	1995-2009	2010-2022	1995-2009	2010-2022
Possible corridor-year observations	803,130	744,732	803,130	744,732
Non-missing corridor-year observations	79,400	594,954	633,284	595,285
Source (in percent of non-missing observations)				
$trips^A$	100.00%	4.49%	3.31%	4.53%
$trips^A$, interpolated	-	-	0.04%	0.06%
$trips^B$	-	95.51%	-	95.43%
$trips^B$, interpolated	-	-	-	0.05%
$trips^B$, backcast	-	-	96.64%	-

Panel B: Volume statistics

	Series 1		Series 2	
	1995-2009	2010-2022	1995-2009	2010-2022
Total border crossings	48,687,614,832	99,319,271,712	87,102,324,943	99,339,768,950
Source (in percent of total border crossings)				
$trips^A$	100.00%	39.72%	43.11%	39.71%
$trips^A$, interpolated	-	-	0.05 %	0.01%
$trips^B$	-	60.27%	-	60.26%
$trips^B$, interpolated	-	-	-	0.01%
$trips^B$, backcast	-	-	56.84%	-

Table 3 Summary statistics of Series 1 and Series 2 estimates

Variable	Obs	Mean	Std Dev	Min	Max (in million)	Totals in 2019 (in million)
Border crossings, s1	674,354	219,480	2,371,958	0	151.5	9635.8
Border crossings, s2	1,228,569	151,756	1,836,110	1	151.5	9637.9

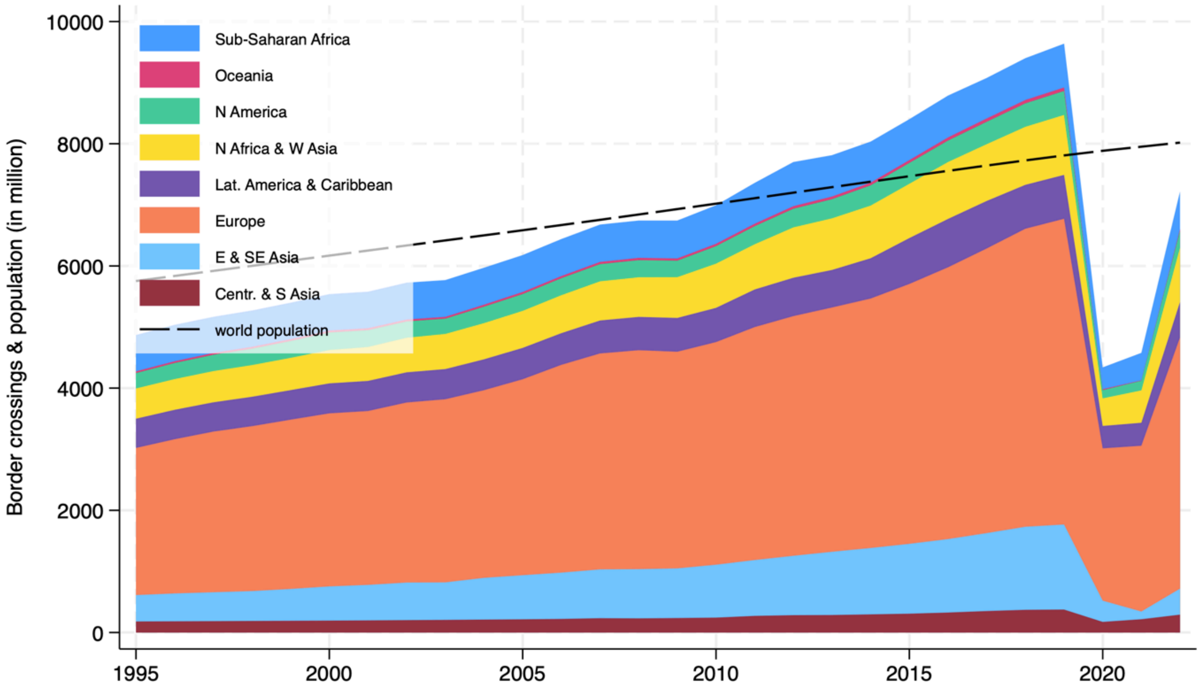


Figure 4 Global transnational mobility–GTMD2 estimates (Series 2) by destination world region, annual

5 Exploring the dataset

5.1 Development of global transnational mobility over time

According to our Series 2 estimates, total global transnational mobility amounted to 4.87 billion trips in 1995 and rose to 9.64 billion in 2019. dropping during the Covid-19 pandemic (4.34 in 2020) and recovering to 7.22 in 2022. In comparison, total world population was 5.73 billion in 1995, 7.74 billion in 2019 and 7.95 billion in 2022. In the quarter of a century between 1995 and 2019 (i.e., on the eve of the pandemic), the world population grew by an average of 1.28 percent a year; global transnational mobility grew by a yearly average of 2.88 percent. The slope of the growth curve became steeper from 2010 onwards, culminating with the abrupt decline provoked by travel restrictions during the Covid-19 pandemic.

The growth rate of transnational mobility was not uniform across world regions (Figure 5). East and South-East Asian countries experienced a faster pace of increase from the turn of the century onwards. On the eve of the pandemic, the number of border crossings in East and South-East Asia was 3.2 times higher than the one the region had in 1995, whereas it was less than two times larger in the rest of the world. In Sub-Saharan Africa–the region with the lowest increase–transnational trips amounted to only 1.2 times their number a quarter of a century earlier. In that same region, the Covid-19 pandemic pushed back the mobility volume to almost 1995 levels. However, the effect of the health emergency and restrictions to travel hit the strongest exactly where the growth rate had been the highest–in East and South-East Asia–,

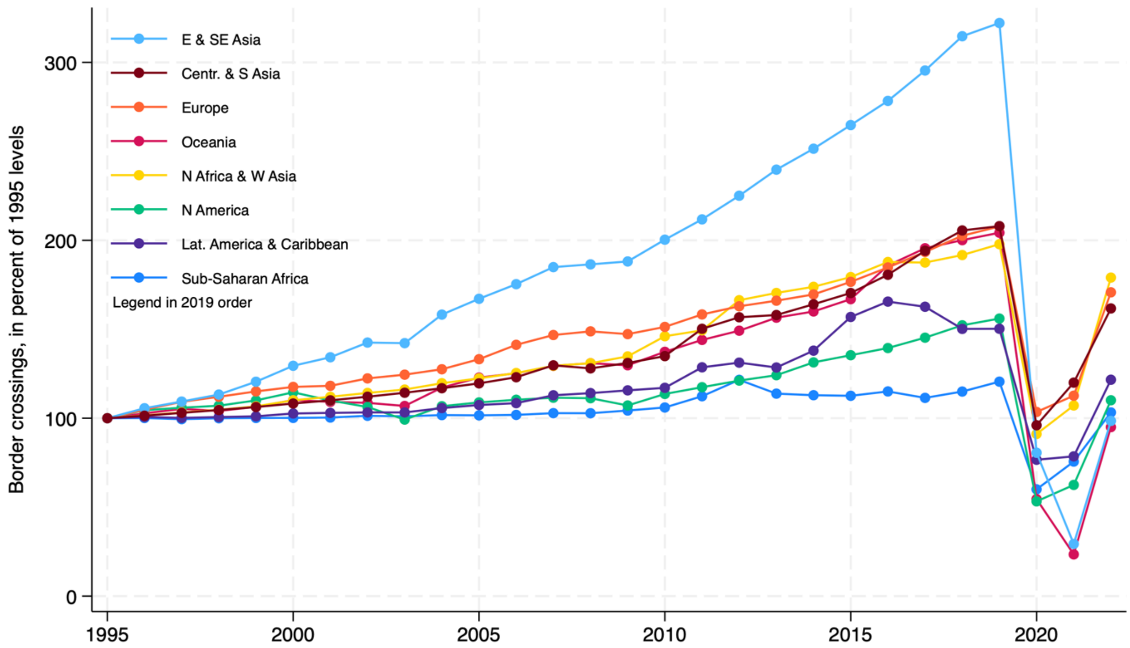


Figure 5 Growth rate of transnational mobility across world regions (1995=100)

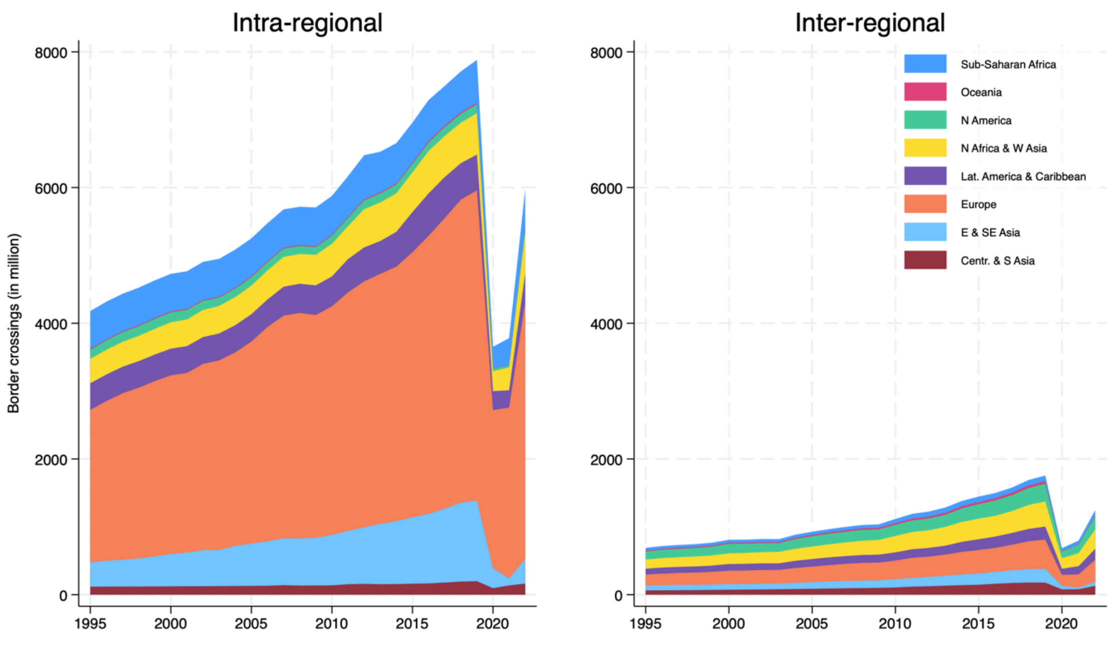


Figure 6 Intra and inter-regional mobility–GTMD2 estimates (Series 2) by destination world region, annual totals

Table 4 Composition of border crossings by region, pooled sample (1995-2022)

Destination regions	border crossings per region	of which intra-regional	of which inter-regional	share of migrants per region
Centr. and S Asia	3.04%	56.47%	43.53%	1.46%
E and SE Asia	10.94%	85.18%	14.82%	.28%
Europe	56.77%	93.41%	6.59%	.15%
Lat. America & Caribbean	7.77%	73.15%	26.85%	.24%
N Africa & W Asia	8.49%	65.17%	34.83%	.60%
N America	5.20%	42.97%	57.03%	.91%
Oceania	.53%	37.09%	62.91%	1.42%
Sub-Saharan Africa	7.25%	90.42%	9.58%	.50%

bringing border crossings in 2022 to even lower figures than in 1995. Post-Covid recovery, in terms of volumes of transnational mobility, was markedly more robust in Europe and North Africa/West Asia, attaining in 2022 the values of 2015.

As shown in Figure 6, the bulk of international mobility is driven by intra-regional mobility (see also, Deutschmann & Recchi, 2024). In Europe, in particular, 93% of transnational mobility is intra-regional. The number of intra-European border crossings alone is always more than twice as large as the global volume of inter-regional trips. The total size of inter-regional mobility is only about a fifth of that of total mobility.

Table 4 further illustrates mobility in the different world regions. The first column shows the average contribution of each region to global mobility across all years 1995-2022. Europe has the largest share with around 57%, while Oceania has the smallest share with less than 1%. The second column shows the average share of intra-regional mobility within each region over all years. The third column shows the average share of inter-regional mobility. Transnational mobility, as already observed, is dominated by intra-regional mobility. The exceptions are Northern America and Oceania where, due to the prevalence of geographically large countries, inter-regional mobility is larger than intra-regional mobility. International migration, i.e., cross-border residence changes, makes up only a fraction of total border crossings. Central and Southern Asia has the highest share with 1.46% followed by Oceania. In all other regions, migration constitutes less than 1% of total incoming cross-border mobility. It is the lowest in Europe, with 0.15% of all border crossings.

5.2 The global structure of transnational mobility

The entire configuration of human border crossings defines a global network that can be examined in structural terms, along the lines of social network analysis (Wasserman and Faust, 1994). How are mobility flows patterned along the 243 x 243 country corridors that constitute the global space of transnational travel? The previous descriptive analysis already suggests that we can hardly expect these flows to be scattered evenly. Indeed, a few corridors are highly crowded, while others are practically deserted. Table 5 lists the extremes of this distribution–

Table 5 Country pairs with highest and lowest mobility volumes in 1995, 2005, 2015 and 2022

year	Highest (descending)	Lowest (ascending)
1995	Germany–Austria	American Samoa–Bangladesh
	USA–Canada	Solomon Islands–Cabo Verde
	USA–Mexico	Timor-Leste–Trinidad and Tobago
	Germany–Poland	Vanatu–Senegal
	Germany–Netherlands	Cabo Verde–New Caledonia
	United Kingdom–Ireland	British Virgin Islands–Christmas Islands
	Germany–Switzerland	Solomon Islands–Guinea
	China–Hong Kong	Benin–Tuvalu
	Belgium–Netherlands	Guyana–Timor-Leste
	Germany–France	Solomon Islands–Sierra Leone
2005	Germany–Austria	San Marino–Saudi Arabia
	China–Hong Kong	Peru–San Marino
	USA–Mexico	Solomon Islands–Cabo Verde
	USA–Canada	Cabo Verde–New Caledonia
	United Kingdom–Ireland	British Virgin Islands–Christmas Island
	Germany–Switzerland	Malta–Tonga
	Germany–Netherlands	Cook Islands–Malta
	Belgium–Netherlands	Bahrain–French Polynesia
	Germany–Poland	Fiji–Sao Tome and Principe
	Germany–France	Cabo Verde–Micronesia
2015	Germany–Austria	Bhutan–Ecuador
	China–Hong Kong	Bhutan– Bahamas
	USA–Mexico	Bhutan–Guatemala
	Germany–Switzerland	Bhutan–Belarus
	United Kingdom–Ireland	Bhutan–Bolivia
	Germany–Netherlands	Panama–San Marino
	USA–Canada	Solomon Islands–Cabo Verde
	Belgium–Netherlands	British Virgin Islands–Christmas Island
	Malaysia–Singapore	Sao Tome and Principe–Samoa
	France–Switzerland	Cook Islands–Malta
2022	Germany–Austria	Macao– Saint Vincent and the Grenadines
	USA–Mexico	Myanmar–Dominica
	Germany–Netherlands	Solomon Islands–Cabo Verde
	Belgium–Netherlands	New Zealand–Guernsey
	United Kingdom–Ireland	Cabo Verde–New Caledonia
	Germany–Switzerland	British Virgin Islands–Christmas Island
	Germany–Poland	Sao Tome and Principe–Samoa
	USA–Canada	New Zealand–Isle of Man
	France–Switzerland	Cook Islands–Malta
	France–Spain	Malta–Tonga

that is, the ten country pairs with the highest and lowest volumes of travel in 1995, 2005, 2015 and 2022. This list is relatively stable over time.

Beyond the individual dyads, however, what matters is the full matrix of flows and their relationship in the global network of human transnational mobility. A way to synthetically describe–and represent visually–the overall configuration of the complete network of mobility flows is a Community Detection Algorithm (CDA). Simply put, a CDA detects clusters in a network by finding the nodes (here, countries) with significantly more connections than a random distribution of connections between nodes. Such algorithms have been used to inductively study human mobility and migration between countries before (e.g., Abel et al., 2021; Deutschmann et al., 2021).

While no universally agreed protocol of CDA has been established (Fortunato & Hric, 2016), we follow earlier studies of transnational mobility (e.g., Delhey et al., 2019; Deutschmann et al., 2021; X. Sun et al., 2016) who use the modularity-based Louvain method for community (cluster) detection (Blondel et al., 2008). We run our analysis in Gephi (Bastian et al., 2009), replicating the presets described in Deutschmann (2020). For the years 1995, 2005, 2015, 2022, representing the last four decades of transnational mobility, we determine clusters based on Series 2 of our estimates and produce corresponding network graphs (Figure 7). With a resolution of 1.0, the algorithm identifies clusters that align well with the network structure, according to the thresholds proposed by Blondel et al. (2008).[13]

In Figure 7, the width of the edges in the graphs is determined by the volume of mobility in each directional corridor, while their so-called weighted PageRank determines the size of the nodes. First developed by Page et al. (1999), the PageRank algorithm is a variation on eigenvector centrality, which goes beyond weighted degree (number of nodes) considerations also to include the degree of the connections to a particular node, ultimately providing a measure of influence of specific nodes. The location of the nodes corresponds to the capital city of each country. For each cluster, we highlight the most prominent hub, i.e., the country (node) with the highest weighted PageRank. In addition to the network graphs, we plot the weighted PageRank scores of the main hubs of each cluster (Figure 8).

Our network analysis suggests a strong element of continuity: over a quarter of a century, transnational mobility flows across countries constantly form eight or nine global clusters, broadly corresponding to geographic regions (e.g., the Americas, Eastern & South-Eastern Asia and Oceania, Western Africa, Eastern Africa, Europe, Middle East). However, stability masks some major ‘tectonic shifts’. We identify six of the most relevant changes over the period considered:

1. While Latin America originally constituted a separate cluster, from 2010 onwards, travel from this region tends to target North American destinations. The region has become integrated into the cluster, which revolves around the USA. In terms of transnational travel, the Americas have merged into a unified mobility system.
2. The cluster with the largest nodes revolves around Germany, due to the high levels of interconnectedness and proximity of EU member states (cf. Table 5 and Figure 8).

[13] Global modularities of 0.626, 0.599, 0.587, 0.549 are returned by the algorithm, each of which is well-above the minimum value of 0.4, which marks the level beyond which the detected communities can be deemed meaningful.

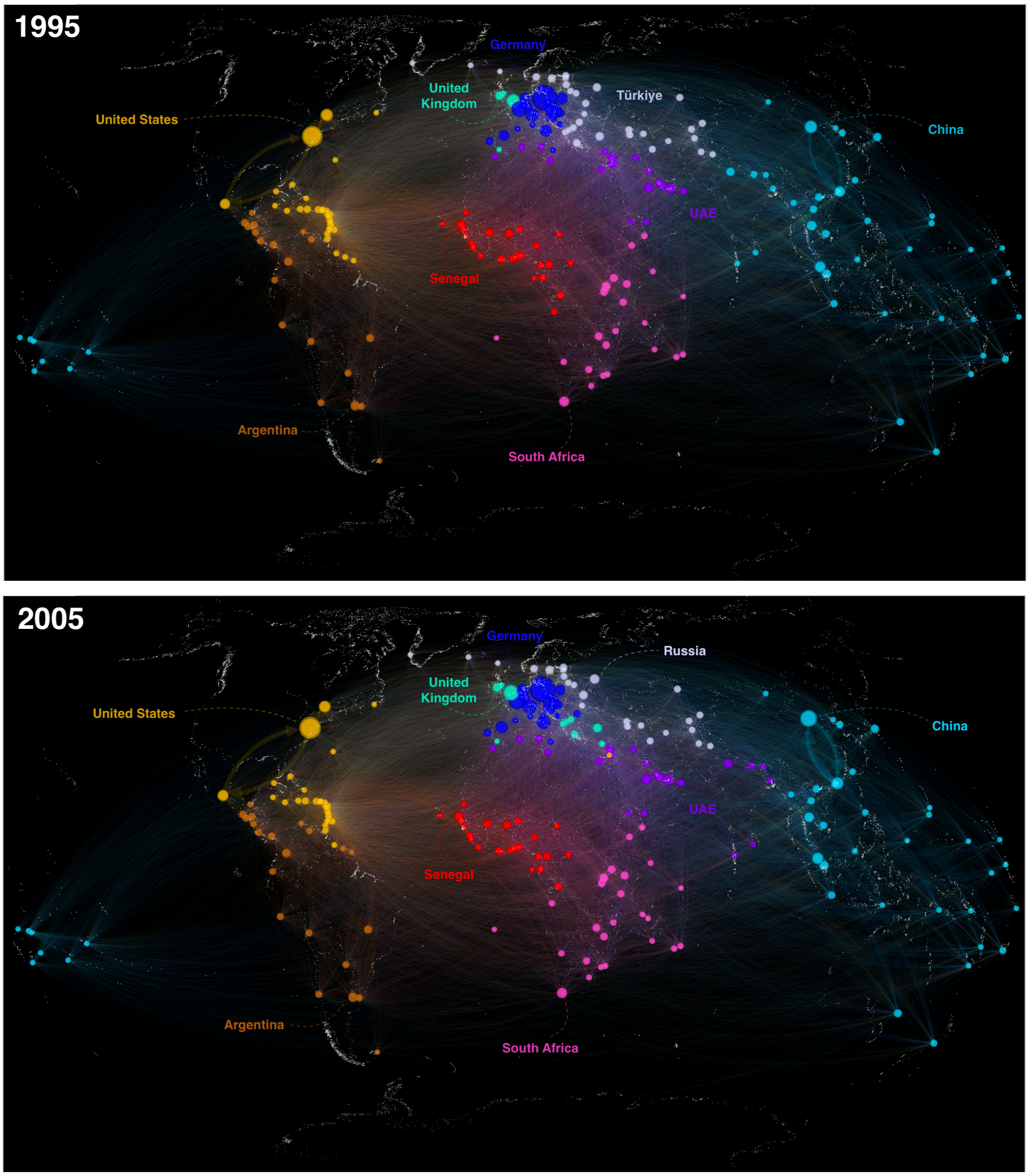
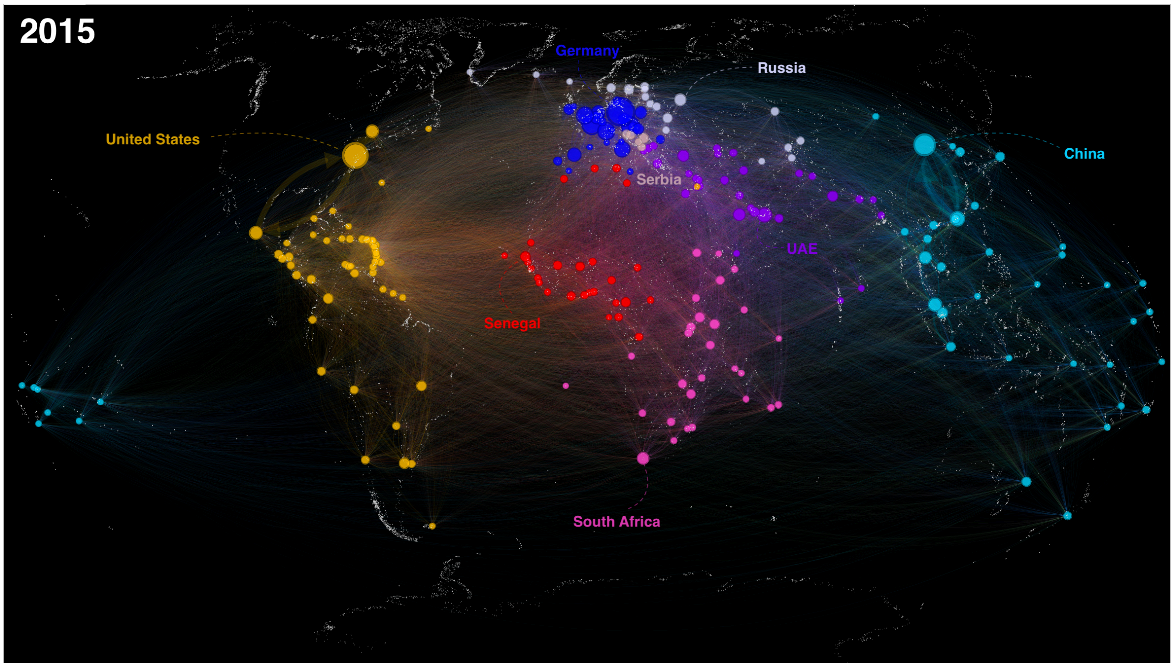


Figure 7 Clusters of human transnational mobility worldwide: 1995, 2005, 2015, 2022

However, Norway, Sweden, Finland and the Baltic States are never part of it, constituting a separate cluster that extends to Russia and some former USSR states (including Ukraine) until 2019. After Covid-19 and the Russia-Ukraine war, these Scandinavian-Baltic countries severed their ties with Russia and the other former USSR countries.

3. The cluster around the UAE, which continuously includes the Gulf states, Israel, Egypt, Iraq, Syria and Lebanon, expands east between 1995 and 2022, incorporating South Asian countries like India, Pakistan, Bangladesh, and Sri Lanka, but leaving countries



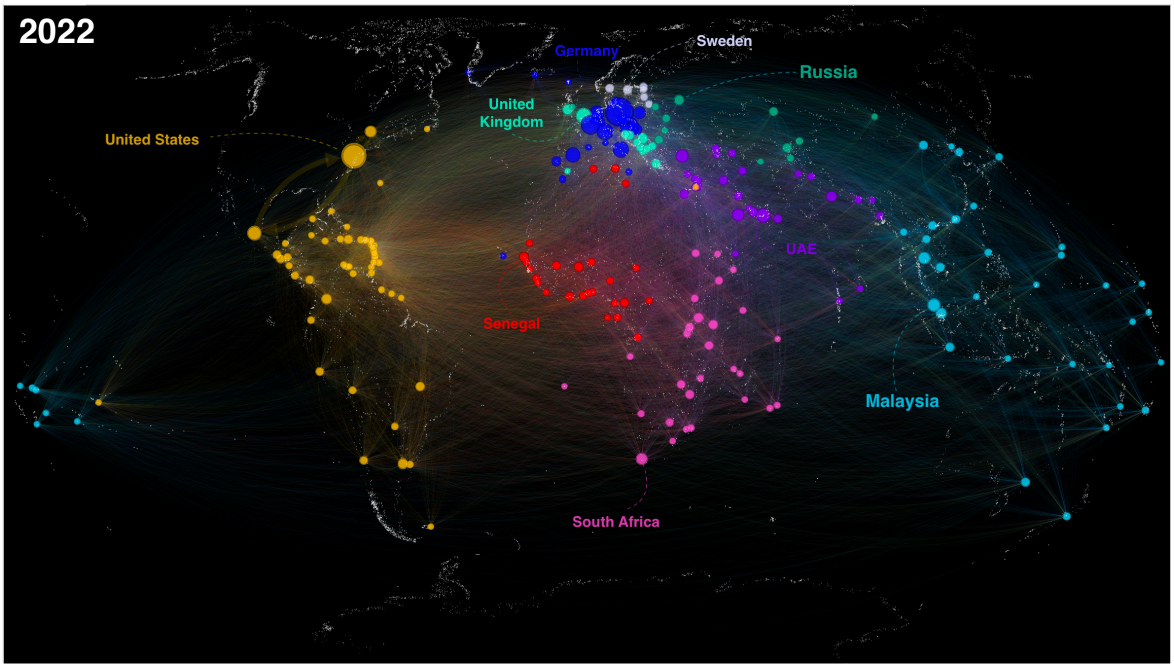


Figure 7 cnt'd: Clusters of human transnational mobility worldwide: 1995, 2005, 2015, 2022

of the Maghreb behind (Morocco, Algeria, Tunisia, Libya). Some Mediterranean European countries (e.g., Bulgaria, Greece, Albania, and Cyprus) and Türkiye are on the fringe of this cluster, entering and exiting it. Caucasian countries (Armenia, Georgia, and Azerbaijan) are progressively absorbed into its orbit as well.

4. UK and Ireland are always associated in the same cluster. While forming a cluster of their own in 1995 (including Isle of Man and Gibraltar), they are joined by Türkiye, Greece, and Albania in 2005. Then they are associated with the Continental European

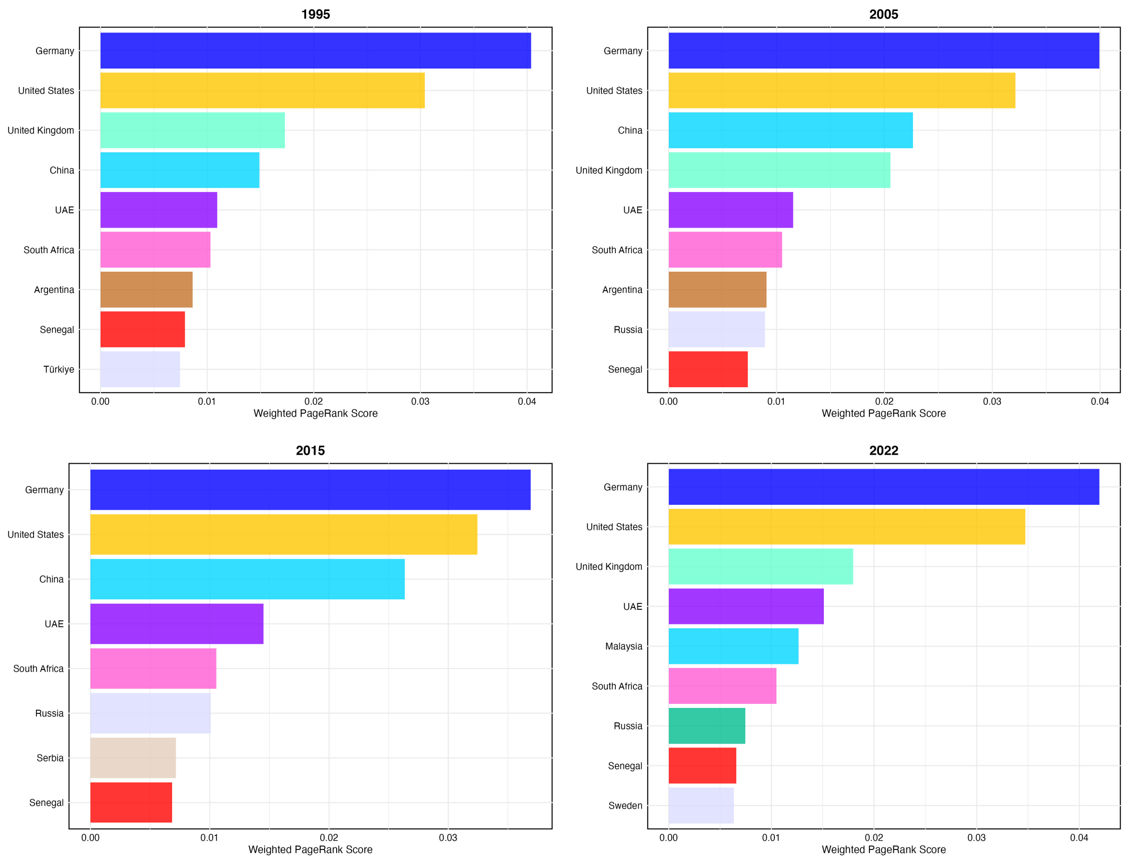


Figure 8 Weighted PageRank Scores of cluster hubs in 1995, 2005, 2015, 2022

cluster in 2015, but separated from it again in 2022, when they were joined by Greece and several Western Balkan states.

5. China forms the backbone of an Asian cluster. Japan, Korea, Australia, and New Zealand are part of it. The cluster remains stable with the exception that Southern Asian countries like India, Pakistan, and Bangladesh switch to the UAE-centered cluster over time. In 2022, Malaysia is the largest hub in this cluster (Figure 8) as Chinese transnational mobility was still recovering from the travel restrictions during the Covid-19 pandemic.
6. North-African countries (Morocco, Algeria, Tunisia, and Libya) move from the Middle Eastern-Asian cluster to the West African cluster in the 2010s. As the only one, Morocco joins the Continental European cluster in 2022.

6 Discussion and conclusion

This paper presents the Global Transnational Mobility Dataset 2.0 (GTMD2.0), which estimates the volume of human travel across country borders worldwide from 1995 to 2022. Our analysis reveals that transnational mobility has significantly increased globally, doubling in volume between 1995 and 2019 and outpacing global population growth over the same period. The dataset highlights that most border crossings are intra-regional, with Europe leading in both total and intra-regional mobility. International migration constitutes a small fraction of total border crossings.

Compared to an earlier version of the dataset (GTMD1.0: Recchi et al., 2019), we improve not only the historical timeframe but also the methodology in two important ways. First, on a conceptual level, GTMD2.0 includes border crossings not only by visitors and returning visitors but also by migrants, i.e., people who cross borders and, by doing so, change their country of residence. While our results show that this does not make a tangible difference because migration flows are extremely low compared to non-migratory border crossing events, it makes the dataset coverage conceptually exhaustive. Second, this paper introduces a new method to aggregate different indicators to form a unified metric of visitor flows. The result is that the current GTMD2.0 estimates of visitor volumes are almost twice as high as its predecessor's.

Our dataset draws on three different and extensive sources. As Bircan put it (2024, p. 6), "estimating cross-border mobilities and flows with big data is fraught with difficulties." In our case, the resulting dataset has two notable limitations. First, the reliance on multiple non-conventional sources, triangulating tourism and air travel passenger volumes, introduces potential inconsistencies and gaps. While we have employed methods to validate and complement these data, discrepancies in data collection and reporting standards across countries may affect the accuracy of the estimates (Recchi & Tittel, 2023). Second, the dataset primarily focuses on recorded travel data, potentially underrepresenting undocumented or irregular cross-border movements. This limitation is particularly relevant for regions with significant informal migration or travel, where official records may not capture the full extent of human mobility across countries. Ultimately, the dataset is designed to capture broad structural shifts in global mobility with the widest possible coverage, rather than achieving perfect accuracy at the corridor-year level. Indeed, while bias cannot be ruled out idiosyncratically, we are confident that the dataset fittingly represents global trends in human mobility.

GTMD2.0 is meant to serve a large spectrum of research, across several disciplines, with broad policy implications. In migration research, it can feed the burgeoning "predictive analytical" turn (Bircan, 2024; see also Bosco et al., 2022; Laczko et al., 2023; Rampazzo et al., 2023; Salah, 2022; Tjaden, 2021; Verhulst & Young, 2023), by offering a long time series of country-to-country movements that form the backdrop of more specific global population flows (for instance, student migration). The forecast of rising, declining, and newly emerging migration corridors can be addressed in a comparative and longitudinal way by combining this dataset with geographically more circumscribed figures or estimates (e.g., Barker & Bijak, 2025;

Melachrinos et al., 2020). Tourism studies can equally leverage the dataset to draw scenarios by projecting existing trends in border-crossing (Volo, 2020).

Global mobility data can support sustainability research as well (C.-L. Chang et al., 2020; Gössling et al., 2023; Scott, 2021). Given the climate impact of travel, future environmental policies may need to design taxes or travel caps on actual mobility trends. Our data permit to give a historical and comparative perspective to such trends, and also form the basis for a detailed assessment of their environmental externalities. Most recent research alerts on the global disparities that drive up the carbon emissions of travel, and points to the necessity of tailoring policy interventions to the specificity of mobility corridors (Y.-Y. Sun et al., 2024).

The global scope of the dataset is particularly congenial to epidemiological modeling, too. The spread of diseases depends largely on human mobility, which in turn follows established routes and network-based logics on a planetary scale (e.g., S. Chang et al., 2021; Pastor-Satorras & Vespignani, 2001). During the Covid-19 pandemic, several studies leveraged GTMD1.0 to assess aspects of the virus spread and impact (e.g., Clemens & Ginn, 2020; Klamser et al., 2024; Nemira et al., 2021; Recchi et al., 2022). Observational data on mobility are necessary to feed agent-based models in path-dependent and realistic ways (Frías-Martínez et al., 2011). Our dataset, spatially close to a full global coverage and temporally extending over more than a quarter of a century, can serve this purpose better than data with a less comprehensive and durable scope.

Finally, as social scientists, we plan to use the dataset to investigate the relationship between global travel, socioeconomic development, and inequalities, adding a comprehensive population movement dimension to a political economy and international relations perspective. What traditional research treats as ‘tourists’ or ‘visitors’– categories of apparent modest relevance politically and socioeconomically–conceal a plethora of social actors who grease the wheels of local economies and societies through their transnational practices even when they do not settle permanently at destination (Favell, 2022; Recchi, 2024). Existing studies have overlooked the possible contribution of population inflows beyond migration figures, which, as our data show, are only the tip of the iceberg of human mobility across national borders.

References

Abel, G. J. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, *28*, 505–546. https://doi.org/10.4054/DemRes.2013.28.18

Abel, G. J., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data*, *6*(82), 1–13. https://doi.org/10.1038/s41597-019-0089-3

Abel, G. J., DeWaard, J., Ha, J. T., & Almquist, Z. W. (2021). The form and evolution of international migration networks, 1990–2015. *Population, Space and Place*, *27*(3), e2432. https://doi.org/10.1002/psp.2432

Akbaritabar, A., Theile, T., & Zagheni, E. (2024). Bilateral flows and rates of international migration of scholars for 210 countries for the period 1998-2020. *Scientific Data*, *11*(1), 816. https://doi.org/10.1038/s41597-024-03655-9

Azose, J. J., & Raftery, A. E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(1), 116–122. https://doi.org/10.1073/pnas.1722334116

Barker, E. R., & Bijak, J. (2025). Mixed-frequency VAR: A new approach to forecasting migration in Europe using macroeconomic data. *Data & Policy*, *7*, e3. https://doi.org/10.1017/dap.2024.82

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Bircan, T. (2024). Augmentation or replication? Assessing big data's role in migration studies. *Data & Policy*, *6*(e51). https://doi.org/10.1017/dap.2024.57

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bosco, C., Grubanov-Boskovic, S., Iacus, S. M., Minora, U., & Sermi, F. (2022). Data innovation in demography, migration and human mobility: Opportunities and challenges of non-traditional data. *JRC Science for Policy Report*. https://doi.org/10.2760/027157

Chang, C.-L., McAleer, M., & Ramos, V. (2020). A Charter for Sustainable Tourism after COVID-19. *Sustainability*, *12*(9), Article 9. https://doi.org/10.3390/su12093671

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*(7840), 82–87. https://doi.org/10.1038/s41586-020-2923-3

Clemens, M. A., & Ginn, T. (2020). Global Mobility and the Threat of Pandemics: Evidence from Three Centuries. *IZA Discussion Paper Series*, *13947*. https://doi.org/10.2139/ssrn.3751840

Delhey, J., Verbalyte, M., Aplowski, A., & Deutschmann, E. (2019). Free to move: The evolution of the European migration network, 1960–2017. In M. Heidenreich (Ed.), *Horizontal Europeanisation: The Transnationalisation of Daily Life and Social Fields in Europe* (1st ed.). Routledge.

Deutschmann, E. (2020). Visualizing the Regionalized Structure of Mobility between Countries Worldwide. *Socius: Sociological Research for a Dynamic World*, *6*, 2378023120958564. https://doi.org/10.1177/2378023120958564

Deutschmann, E., & Recchi, E. (2024). How global is international mobility? In E. Recchi & M. Safi (Eds.), *Handbook of human mobility and migration* (pp. 94–112). Edward Elgar Publishing. https://doi.org/10.4337/9781839105784.00013

Deutschmann, E., Recchi, E., & Bicchi, F. (2021). Mobility hub or hollow? Cross-border travelling in the Mediterranean, 1995–2016. *Global Networks*, *21*(1), 146–169. https://doi.org/10.1111/glob.12259

EU & UN. (2018). *International recommendations on refugee statistics*. Publications Office of the European Union. https://doi.org/10.2785/517815

Favell, A. (2022). Immigration, integration and citizenship: Elements of a new political demography. *Journal of Ethnic and Migration Studies*, *48*(1), 3–32. https://doi.org/10.1080/1369183X.2022.2020955

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Frías-Martínez, E., Williamson, G., & Frías-Martínez, V. (2011). Agent-Based Modelling of Epidemic Spreading using Social Networks and Human Mobility Patterns. *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing*, 57–64.

Gössling, S., Balas, M., Mayer, M., & Sun, Y.-Y. (2023). A review of tourism and climate change mitigation: The scales, scopes, stakeholders and strategies of carbon management. *Tourism Management*, *95*, 104681. https://doi.org/10.1016/j.tourman.2022.104681

Grohmann, T. (2024). *Measurement, drivers, and consequences of international migration: Essays in migration economics*. University of Groningen. https://doi.org/10.33612/diss.904990049

ILO. (2022). *Temporary labour migration: Unpacking complexities—Synthesis Report*. International Labour Organization.

Klamser, P. P., Zachariae, A., Maier, B. F., Baranov, O., Jongen, C., Schlosser, F., & Brockmann, D. (2024). Inferring country-specific import risk of diseases from the world air transportation network. *PLOS Computational Biology*, *20*(1), e1011775. https://doi.org/10.1371/journal.pcbi.1011775

Laczko, F., Vidal, E. M., & Rango, M. (2023). *Measuring Global Migration: Towards Better Data for All*. Routledge. https://doi.org/10.4324/9781003266075

Llano, C., Pardo, J., Pérez-Balsalobre, S., & Pérez, J. (2023). Estimating multicountry tourism flows by transport mode. *Annals of Tourism Research*, *103*, 103672. https://doi.org/10.1016/j.annals.2023.103672

Melachrinos, C., Carammia, M., & Wilkin, T. (2020). Using big data to estimate migration “push factors” from Africa. In P. Fargues & M. Rango (Eds.), *Migration in West and North Africa and across the Mediterranean. Trends, risks, development and governance* (pp. 98–116). International Organization for Migration.

Nemira, A., Adeniyi, A. E., Gasich, E. L., Bulda, K. Y., Valentovich, L. N., Krasko, A. G., Glebova, O., Kirpich, A., & Skums, P. (2021). SARS-CoV-2 transmission dynamics in Belarus in 2020 revealed by genomic and incidence data analysis. *Communications Medicine*, *1*(1), 1–9. https://doi.org/10.1038/s43856-021-00031-1

Nye, J. S., & Keohane, R. O. (1971). Transnational Relations and World Politics: An Introduction. *International Organization*, *25*(3), 329–349.

OECD. (2019). Capturing the ephemeral: How much labour do temporary migrants contribute in OECD countries? In *International Migration Outlook 2019* (43rd ed., pp. 117–165). OECD Publishing, Organisation for Economic Co-operation and Development. https://doi.org/10.1787/c3e35eec-en

OECD. (2024). *International Migration Outlook 2024* (48th ed.). OECD Publishing, Organisation for Economic Co-operation and Development. https://doi.org/10.1787/50b0353e-en

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, *86*(14), 3200–3203. https://doi.org/10.1103/PhysRevLett.86.3200

Rampazzo, F., Rango, M., & Weber, I. (2023). New Migration Data: Challenges and Opportunities. In E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli, & M. Vespe (Eds.), *Handbook of Computational Social Science for Policy* (pp. 345–359). Springer International Publishing. https://doi.org/10.1007/978-3-031-16624-2_18

Raymer, J., Guan, Q., Shen, T., Wiśniowski, A., & Pietsch, J. (2022). Estimating international migration flows for the Asia-Pacific region: Application of a generation–distribution model. *Migration Studies*, *10*(4), 631–669. https://doi.org/10.1093/migration/mnac023

Recchi, E. (2024). A plea for a sociology of transnational human mobility: Existing evidence and prospective challenges. In W. Sievers, R. Bauböck, M. Czaika, & A. Kraler (Eds.), *Drawing Boundaries and Crossing Borders: Migration in Theorie und Praxis* (pp. 39–56). Verlag der Österreichischen Akademie der Wissenschaften. https://doi.org/10.1553/978OEAW95559

Recchi, E., Deutschmann, E., & Vespe, M. (2019). Estimating Transnational Human Mobility on a Global Scale. *EUI Working Papers*, *RSCAS 2019/30*.

Recchi, E., Ferrara, A., Rodriguez Sanchez, A., Deutschmann, E., Gabrielli, L., Iacus, S., Bastiani, L., Spyratos, S., & Vespe, M. (2022). The impact of air travel on the precocity and severity of COVID-19 deaths in sub-national areas across 45 countries. *Scientific Reports*, *12*(1), 16522. https://doi.org/10.1038/s41598-022-20263-y

Recchi, E., & Safi, M. (Eds.). (2024). *Handbook of Human Mobility and Migration*. Edward Elgar Publishing Limited. https://doi.org/10.4337/9781839105784

Recchi, E., & Tittel, K. (2023). The empirical study of human mobility: Potentials and pitfalls of using traditional and digital data. In Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, & Michele Vespe (Eds.), *Handbook of Computational Social Science for Policy*. Springer International Publishing. https://doi.org/10.1007/978-3-031-16624-2

Salah, A. A. (2022). Can big data deliver its promises in migration research? *International Migration*, *60*(2), 252–255. https://doi.org/10.1111/imig.12984

Scott, D. (2021). Sustainable Tourism and the Grand Challenge of Climate Change. *Sustainability*, *13*(4), Article 4. https://doi.org/10.3390/su13041966

Sun, X., Wandelt, S., Dzikus, N., & Linke, F. (2016). Air passenger flow communities between countries. *2016 35th Chinese Control Conference (CCC)*, 9276–9281. https://doi.org/10.1109/ChiCC.2016.7554832

Sun, Y.-Y., Faturay, F., Lenzen, M., Gössling, S., & Higham, J. (2024). Drivers of global tourism carbon emissions. *Nature Communications*, *15*(1), 10384. https://doi.org/10.1038/s41467-024-54582-7

Tjaden, J. (2021). Measuring migration 2.0: A review of digital data sources. *Comparative Migration Studies*, *9*(1), 59. https://doi.org/10.1186/s40878-021-00273-x

UNDESA. (1998). *Recommendations on Statistics of International Migration* (Rev. 1). United Nations, Department of Economic and Social Affairs. https://unstats.un.org/unsd/publication/seriesm/seriesm_58rev1e.pdf

UNDESA. (2010). *International recommendations for tourism statistics: 2008* (Rev.1). United Nations, Department of Economic and Social Affairs.

UNDESA. (2020). *International Migrant Stock 2020*. United Nations, Department of Economic and Social Affairs. https://www.un.org/development/desa/pd/content/international-migrant-stock

UNWTO. (2024). *UN Tourism Data in Excel Format—Data from 1995 onwards*. World Tourism Organization. https://www.e-unwto.org/toc/untourism1995/1/1

Verhulst, S. G., & Young, A. (2023). Operationalizing Data Collaboratives for Migration. In *Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners* (pp. 92–103). International Organization for Migration (IOM).

Volo, S. (2020). Tourism statistics, indicators and big data: A perspective article. *Tourism Review*, *75*(1), 304–309. https://doi.org/10.1108/TR-06-2019-0262

Wimmer, A., & Glick Schiller, N. (2002). Methodological Nationalism and Beyond: Nation-state Building, Migration and the Social Sciences. *Global Networks*, *2*(4), 301–334.

Appendix:

Border Crossings between 243 countries: The Global Transnational Mobilities Dataset, 1995-2022

Tobias Grohmann[a], Ettore Recchi[a,b], Luca Bernasconi[a]
[a] EUI, Florence, Italy; [b] Sciences Po, Paris, France
Correspondence: tobias.grohmann@eui.eu

This version 06 February 2025

A1 Country Overview

Table A1 Countries in the origin-destination-year grid

	Country name	3-letter code	M49 code	Region	Code start	Code end
1	Afghanistan	AFG	4	Central and Southern Asia	1970	2023
2	Albania	ALB	8	Europe	1970	2023
3	Algeria	DZA	12	Northern Africa and Western Asia	1970	2023
4	American Samoa	ASM	16	Oceania	1970	2023
5	Andorra	AND	20	Europe	1970	2023
6	Angola	AGO	24	Sub-Saharan Africa	1970	2023
7	Anguilla	AIA	660	Latin America and the Caribbean	1980	2023
8	Antarctica	ATA	10		2000	2023
9	Antigua and Barbuda	ATG	28	Latin America and the Caribbean	1970	2023
10	Argentina	ARG	32	Latin America and the Caribbean	1970	2023
11	Armenia	ARM	51	Northern Africa and Western Asia	1992	2023
12	Aruba	ABW	533	Latin America and the Caribbean	1986	2023
13	Australia	AUS	36	Oceania	1970	2023
14	Austria	AUT	40	Europe	1970	2023
15	Azerbaijan	AZE	31	Northern Africa and Western Asia	1992	2023
16	Bahamas	BHS	44	Latin America and the Caribbean	1970	2023
17	Bahrain	BHR	48	Northern Africa and Western Asia	1970	2023
18	Bangladesh	BGD	50	Central and Southern Asia	1971	2023
19	Barbados	BRB	52	Latin America and the Caribbean	1970	2023
20	Belarus	BLR	112	Europe	1970	2023
21	Belgium	BEL	56	Europe	1970	2023
22	Belize	BLZ	84	Latin America and the Caribbean	1970	2023
23	Benin	BEN	204	Sub-Saharan Africa	1970	2023
24	Bermuda	BMU	60	Northern America	1970	2023
25	Bhutan	BTN	64	Central and Southern Asia	1970	2023
26	Bolivia (Plurinational State of)	BOL	68	Latin America and the Caribbean	1970	2023
27	Bonaire, Sint Eustatius and Saba	BES	535	Latin America and the Caribbean	2011	2023
28	Bosnia and Herzegovina	BIH	70	Europe	1992	2023
29	Botswana	BWA	72	Sub-Saharan Africa	1970	2023
30	Brazil	BRA	76	Latin America and the Caribbean	1970	2023
31	British Virgin Islands	VGB	92	Latin America and the Caribbean	1970	2023
32	Brunei Darussalam	BRN	96	Eastern and South-Eastern Asia	1970	2023
33	Bulgaria	BGR	100	Europe	1970	2023
34	Burkina Faso	BFA	854	Sub-Saharan Africa	1970	2023
35	Burundi	BDI	108	Sub-Saharan Africa	1970	2023
36	Cabo Verde	CPV	132	Sub-Saharan Africa	1970	2023
37	Cambodia	KHM	116	Eastern and South-Eastern Asia	1970	2023
38	Cameroon	CMR	120	Sub-Saharan Africa	1970	2023
39	Canada	CAN	124	Northern America	1970	2023
40	Cayman Islands	CYM	136	Latin America and the Caribbean	1970	2023
41	Central African Republic	CAF	140	Sub-Saharan Africa	1970	2023
42	Chad	TCD	148	Sub-Saharan Africa	1970	2023
43	Chile	CHL	152	Latin America and the Caribbean	1970	2023
44	China	CHN	156	Eastern and South-Eastern Asia	1970	2023
45	China, Hong Kong Special Administrative Region	HKG	344	Eastern and South-Eastern Asia	1970	2023
46	China, Macao Special Administrative Region	MAC	446	Eastern and South-Eastern Asia	1970	2023
47	Christmas Island	CXR	162	Oceania	1970	2023
48	Cocos (Keeling) Islands	CCK	166	Oceania	1970	2023
49	Colombia	COL	170	Latin America and the Caribbean	1970	2023
50	Comoros	COM	174	Sub-Saharan Africa	1970	2023
51	Congo	COG	178	Sub-Saharan Africa	1970	2023
52	Cook Islands	COK	184	Oceania	1970	2023
53	Costa Rica	CRI	188	Latin America and the Caribbean	1970	2023
54	Croatia	HRV	191	Europe	1992	2023
55	Cuba	CUB	192	Latin America and the Caribbean	1970	2023
56	Curaçao	CUW	531	Latin America and the Caribbean	2011	2023
57	Cyprus	CYP	196	Northern Africa and Western Asia	1970	2023
58	Czechia	CZE	203	Europe	1993	2023
59	Côte d'Ivoire	CIV	384	Sub-Saharan Africa	1970	2023
60	Democratic People's Republic of Korea	PRK	408	Eastern and South-Eastern Asia	1970	2023
61	Democratic Republic of the Congo	COD	180	Sub-Saharan Africa	1970	2023
62	Denmark	DNK	208	Europe	1970	2023
63	Djibouti	DJI	262	Sub-Saharan Africa	1970	2023
64	Dominica	DMA	212	Latin America and the Caribbean	1970	2023
65	Dominican Republic	DOM	214	Latin America and the Caribbean	1970	2023
66	Ecuador	ECU	218	Latin America and the Caribbean	1970	2023
67	Egypt	EGY	818	Northern Africa and Western Asia	1970	2023

68	El Salvador	SLV	222	Latin America and the Caribbean	1870	2023
69	Equatorial Guinea	GNQ	226	Sub-Saharan Africa	1970	2023
70	Eritrea	ERI	232	Sub-Saharan Africa	1993	2023
71	Estonia	EST	233	Europe	1992	2023
72	Eswatini	SWZ	748	Sub-Saharan Africa	1970	2023
73	Ethiopia	ETH	231	Sub-Saharan Africa	1993	2023
74	Falkland Islands (Malvinas)	FLK	238	Latin America and the Caribbean	1970	2023
75	Faroe Islands	FRO	234	Europe	1970	2023
76	Fiji	FJI	242	Oceania	1970	2023
77	Finland	FIN	246	Europe	1970	2023
78	France	FRA	250	Europe	1970	2023
79	French Guiana	GUF	254	Latin America and the Caribbean	1970	2023
80	French Polynesia	PYF	258	Oceania	1970	2023
81	Gabon	GAB	266	Sub-Saharan Africa	1970	2023
82	Gambia	GMB	270	Sub-Saharan Africa	1970	2023
83	Georgia	GEO	268	Northern Africa and Western Asia	1992	2023
84	Germany	DEU	276	Europe	1991	2023
85	Ghana	GHA	288	Sub-Saharan Africa	1970	2023
86	Gibraltar	GIB	292	Europe	1970	2023
87	Greece	GRC	300	Europe	1970	2023
88	Greenland	GRL	304	Northern America	1970	2023
89	Grenada	GRD	308	Latin America and the Caribbean	1970	2023
90	Guadeloupe	GLP	312	Latin America and the Caribbean	1970	2023
91	Guam	GUM	316	Oceania	1970	2023
92	Guatemala	GTM	320	Latin America and the Caribbean	1970	2023
93	Guernsey	GGY	831	Europe	2005	2023
94	Guinea	GIN	324	Sub-Saharan Africa	1970	2023
95	Guinea-Bissau	GNB	624	Sub-Saharan Africa	1974	2023
96	Guyana	GUY	328	Latin America and the Caribbean	1970	2023
97	Haiti	HTI	332	Latin America and the Caribbean	1970	2023
98	Honduras	HND	340	Latin America and the Caribbean	1970	2023
99	Hungary	HUN	348	Europe	1970	2023
100	Iceland	ISL	352	Europe	1970	2023
101	India	IND	356	Central and Southern Asia	1970	2023
102	Indonesia	IDN	360	Eastern and South-Eastern Asia	1970	2023
103	Iran (Islamic Republic of)	IRN	364	Central and Southern Asia	1970	2023
104	Iraq	IRQ	368	Northern Africa and Western Asia	1970	2023
105	Ireland	IRL	372	Europe	1970	2023
106	Isle of Man	IMN	833	Europe	1982	2023
107	Israel	ISR	376	Northern Africa and Western Asia	1970	2023
108	Italy	ITA	380	Europe	1970	2023
109	Jamaica	JAM	388	Latin America and the Caribbean	1970	2023
110	Japan	JPN	392	Eastern and South-Eastern Asia	1970	2023
111	Jersey	JEY	832	Europe	2005	2023
112	Jordan	JOR	400	Northern Africa and Western Asia	1970	2023
113	Kazakhstan	KAZ	398	Central and Southern Asia	1992	2023
114	Kenya	KEN	404	Sub-Saharan Africa	1970	2023
115	Kiribati	KIR	296	Oceania	1979	2023
116	Kuwait	KWT	414	Northern Africa and Western Asia	1970	2023
117	Kyrgyzstan	KGZ	417	Central and Southern Asia	1992	2023
118	Lao People's Democratic Republic	LAO	418	Eastern and South-Eastern Asia	1970	2023
119	Latvia	LVA	428	Europe	1992	2023
120	Lebanon	LBN	422	Northern Africa and Western Asia	1970	2023
121	Lesotho	LSO	426	Sub-Saharan Africa	1970	2023
122	Liberia	LBR	430	Sub-Saharan Africa	1970	2023
123	Libya	LBY	434	Northern Africa and Western Asia	1970	2023
124	Liechtenstein	LIE	438	Europe	1970	2023
125	Lithuania	LTU	440	Europe	1992	2023
126	Luxembourg	LUX	442	Europe	1970	2023
127	Madagascar	MDG	450	Sub-Saharan Africa	1970	2023
128	Malawi	MWI	454	Sub-Saharan Africa	1970	2023
129	Malaysia	MYS	458	Eastern and South-Eastern Asia	1970	2023
130	Maldives	MDV	462	Central and Southern Asia	1970	2023
131	Mali	MLI	466	Sub-Saharan Africa	1970	2023
132	Malta	MLT	470	Europe	1970	2023
133	Marshall Islands	MHL	584	Oceania	1991	2023
134	Martinique	MTQ	474	Latin America and the Caribbean	1970	2023
135	Mauritania	MRT	478	Sub-Saharan Africa	1970	2023
136	Mauritius	MUS	480	Sub-Saharan Africa	1970	2023
137	Mayotte	MYT	175	Sub-Saharan Africa	2002	2023
138	Mexico	MEX	484	Latin America and the Caribbean	1970	2023
139	Micronesia (Federated States of)	FSM	583	Oceania	1991	2023
140	Monaco	MCO	492	Europe	1970	2023
141	Mongolia	MNG	496	Eastern and South-Eastern Asia	1970	2023
142	Montenegro	MNE	499	Europe	2006	2023
143	Montserrat	MSR	500	Latin America and the Caribbean	1970	2023

144	Morocco	MAR	504	Northern Africa and Western Asia	1970	2023
145	Mozambique	MOZ	508	Sub-Saharan Africa	1970	2023
146	Myanmar	MMR	104	Eastern and South-Eastern Asia	1970	2023
147	Namibia	NAM	516	Sub-Saharan Africa	1970	2023
148	Nauru	NRU	520	Oceania	1970	2023
149	Nepal	NPL	524	Central and Southern Asia	1970	2023
150	Netherlands (Kingdom of the)	NLD	528	Europe	1970	2023
151	Netherlands Antilles	ANT	530	Latin America and the Caribbean	1986	2010
152	New Caledonia	NCL	540	Oceania	1970	2023
153	New Zealand	NZL	554	Oceania	1970	2023
154	Nicaragua	NIC	558	Latin America and the Caribbean	1970	2023
155	Niger	NER	562	Sub-Saharan Africa	1970	2023
156	Nigeria	NGA	566	Sub-Saharan Africa	1970	2023
157	Niue	NIU	570	Oceania	1970	2023
158	Norfolk Island	NFK	574	Oceania	1970	2023
159	North Macedonia	MKD	807	Europe	1992	2023
160	Northern Mariana Islands	MNP	580	Oceania	1991	2023
161	Norway	NOR	578	Europe	1970	2023
162	Oman	OMN	512	Northern Africa and Western Asia	1970	2023
163	Pakistan	PAK	586	Central and Southern Asia	1970	2023
164	Palau	PLW	585	Oceania	1994	2023
165	Panama	PAN	591	Latin America and the Caribbean	1982	2023
166	Papua New Guinea	PNG	598	Oceania	1975	2023
167	Paraguay	PRY	600	Latin America and the Caribbean	1970	2023
168	Peru	PER	604	Latin America and the Caribbean	1970	2023
169	Philippines	PHL	608	Eastern and South-Eastern Asia	1970	2023
170	Poland	POL	616	Europe	1970	2023
171	Portugal	PRT	620	Europe	1970	2023
172	Puerto Rico	PRI	630	Latin America and the Caribbean	1970	2023
173	Qatar	QAT	634	Northern Africa and Western Asia	1970	2023
174	Republic of Korea	KOR	410	Eastern and South-Eastern Asia	1970	2023
175	Republic of Moldova	MDA	498	Europe	1992	2023
176	Romania	ROU	642	Europe	1970	2023
177	Russian Federation	RUS	643	Europe	1992	2023
178	Rwanda	RWA	646	Sub-Saharan Africa	1970	2023
179	Réunion	REU	638	Sub-Saharan Africa	1970	2023
180	Saint Barthélemy	BLM	652	Latin America and the Caribbean	2007	2023
181	Saint Helena	SHN	654	Sub-Saharan Africa	1970	2023
182	Saint Kitts and Nevis	KNA	659	Latin America and the Caribbean	1980	2023
183	Saint Lucia	LCA	662	Latin America and the Caribbean	1970	2023
184	Saint Martin (French Part)	MAF	663	Latin America and the Caribbean	2007	2023
185	Saint Pierre and Miquelon	SPM	666	Northern America	1970	2023
186	Saint Vincent and the Grenadines	VCT	670	Latin America and the Caribbean	1970	2023
187	Samoa	WSM	882	Oceania	1970	2023
188	San Marino	SMR	674	Europe	1970	2023
189	Sao Tome and Principe	STP	678	Sub-Saharan Africa	1970	2023
190	Saudi Arabia	SAU	682	Northern Africa and Western Asia	1970	2023
191	Senegal	SEN	686	Sub-Saharan Africa	1970	2023
192	Serbia	SRB	688	Europe	2006	2023
193	Serbia and Montenegro	SCG	891	Europe	1992	2006
194	Seychelles	SYC	690	Sub-Saharan Africa	1970	2023
195	Sierra Leone	SLE	694	Sub-Saharan Africa	1970	2023
196	Singapore	SGP	702	Eastern and South-Eastern Asia	1970	2023
197	Sint Maarten (Dutch part)	SXM	534	Latin America and the Caribbean	2011	2023
198	Slovakia	SVK	703	Europe	1993	2023
199	Slovenia	SVN	705	Europe	1992	2023
200	Solomon Islands	SLB	90	Oceania	1970	2023
201	Somalia	SOM	706	Sub-Saharan Africa	1970	2023
202	South Africa	ZAF	710	Sub-Saharan Africa	1970	2023
203	South Sudan	SSD	728	Sub-Saharan Africa	2011	2023
204	Spain	ESP	724	Europe	1970	2023
205	Sri Lanka	LKA	144	Central and Southern Asia	1970	2023
206	State of Palestine	PSE	275	Northern Africa and Western Asia	1999	2023
207	Sudan	SDN	729	Northern Africa and Western Asia	2011	2023
208	Sudan (old)	SDN (old)	736	Northern Africa and Western Asia	1970	2011
209	Suriname	SUR	740	Latin America and the Caribbean	1970	2023
210	Svalbard and Jan Mayen Islands	SJM	744	Europe	1970	2023
211	Sweden	SWE	752	Europe	1970	2023
212	Switzerland	CHE	756	Europe	1970	2023
213	Syrian Arab Republic	SYR	760	Northern Africa and Western Asia	1970	2023
214	Taiwan, Province of China	TWN	158	Northern Africa and Western Asia	1970	2023
215	Tajikistan	TJK	762	Central and Southern Asia	1992	2023
216	Thailand	THA	764	Eastern and South-Eastern Asia	1970	2023
217	Timor-Leste	TLS	626	Eastern and South-Eastern Asia	1970	2023
218	Togo	TGO	768	Sub-Saharan Africa	1970	2023
219	Tonga	TON	776	Oceania	1970	2023

220	Trinidad and Tobago	TTO	780	Latin America and the Caribbean	1970	2023
221	Tunisia	TUN	788	Northern Africa and Western Asia	1970	2023
222	Turkmenistan	TKM	795	Central and Southern Asia	1992	2023
223	Turks and Caicos Islands	TCA	796	Latin America and the Caribbean	1970	2023
224	Tuvalu	TUV	798	Oceania	1978	2023
225	Türkiye	TUR	792	Northern Africa and Western Asia	1970	2023
226	Uganda	UGA	800	Sub-Saharan Africa	1970	2023
227	Ukraine	UKR	804	Europe	1970	2023
228	United Arab Emirates	ARE	784	Northern Africa and Western Asia	1970	2023
229	United Kingdom of Great Britain and Northern Ireland	GBR	826	Europe	1970	2023
230	United Republic of Tanzania	TZA	834	Sub-Saharan Africa	1970	2023
231	United States Minor Outlying Islands	UMI	581	Oceania	1986	2023
232	United States Virgin Islands	VIR	850	Latin America and the Caribbean	1970	2023
233	United States of America	USA	840	Northern America	1776	2023
234	Uruguay	URY	858	Latin America and the Caribbean	1970	2023
235	Uzbekistan	UZB	860	Central and Southern Asia	1992	2023
236	Vanuatu	VUT	548	Oceania	1970	2023
237	Venezuela (Bolivarian Republic of)	VEN	862	Latin America and the Caribbean	1970	2023
238	Viet Nam	VNM	704	Eastern and South-Eastern Asia	1976	2023
239	Wallis and Futuna Islands	WLF	876	Oceania	1970	2023
240	Yemen	YEM	887	Northern Africa and Western Asia	1990	2023
241	Zambia	ZMB	894	Sub-Saharan Africa	1970	2023
242	Zimbabwe	ZWE	716	Sub-Saharan Africa	1970	2023
243	Åland Islands	ALA	248	Europe	2002	2023

A2 Data cleaning

Upon importing the three data records into the grid, we noticed that especially the UNWTO Tourism Data required extensive data cleaning. First, while originally there were 219 Excel sheets, 7 countries (Afghanistan, Djibouti, Equatorial Guinea, Liberia, Mauritania, Nauru, and South Sudan) do not provide data on inbound tourism by origin. This means that these 7 countries will not appear as destinations in our tourism data. Second, we correct several mistakes and inconsistencies regarding country names and codes (Section A2.1). Third, the list of origin countries contains "odd" travel origin categories. Whenever possible, we split or aggregate these anomalous cases to translate them into our units (Section A2.2). Finally, we deal with "overhanging" observations where values are reported although an origin or destination country may not exist anymore/yet in our grid (Section 0).

A2.1 Correction of errors in UNWTO dataset

Regarding the mistakes and inaccuracies reported in Table A2, we take the following actions:

- We assume that the country *names* are correct but that the assigned codes are faulty.
- We replace the faulty codes of Curaçao, Ethiopia, Germany, and Sint Maarten (Dutch Part) with the correct M49 codes.
- As they "occupy" the wrong codes, we assign fictitious country codes to Bonaire, Sint Eustatius, Saba, and Yap State as the origin or destination country. Later, we will aggregate these observations to form the proper statistical units. So, for example, we will aggregate Bonaire, Sint Eustatius, and Saba (BES, 535) or Kosrae, Truk, Pohnpei, and Yap State (Federated States of Micronesia, 583) into their proper statistical units. Further details are described below.
- We drop observations with the origins "Hawaii, USA", "Wake Island", "Dubai", "Midway Islands", and "Johnston Island", because we cannot be sure whether travelers from, e.g., Dubai or Hawaii, are already included in the number of travelers from the UAE or USA, respectively.
- We leave disentangling the correct corridors involving origin countries Czechoslovakia and Sudan to a later stage.

At this stage, after the above operations, there are 212 destinations countries left for which we have reported tourism/travel inflows from 317 origins.

Table A2 Mistakes and inconsistencies in UNWTO (2024) dataset

Among origin countries	Among destination countries
First, Bonaire is listed as an origin country. Yet, Bonaire is not an independent country with a separate code. It is part of BES. Second, the code used for Bonaire is the M49 code of Sint Maarten (Dutch Part).	Same as on the left.
Whenever Curaçao was listed as the origin country, it was assigned code 535, the M49 code of BES. However, the correct M49 code of Curaçao is 531.	When listed as destination country, Curaçao is assigned the correct code.
The country name given to the statistical unit with code 200 is "Czech Republic/ Slovakia". The correct name would be "Czechoslovakia". Note that Czechia	Neither "Czech Republic/ Slovakia" nor "Czechoslovakia" appear as destinations. Instead,

and Slovakia are separated since 1993. Both Czechia and Slovakia are among the listed origin countries too.	Czechia (formerly Czech Republic) and Slovakia are listed as separate destinations as should be.
The code used for Ethiopia is the M49 code of old Ethiopia previous to secession of Eritrea in 1993.	Same as on the left.
The code used for Germany is the M49 code of old West Germany previous to reunification in 1991.	Same as on the left.
Sint Eustatius is listed as an origin country. Yet, Sint Eustatius is not an independent country with a separate code. It is part of BES. Second, the code used is the M49 code of (old) Saint Kitts-Nevis-Anguilla, which does not exist anymore since 1980.	Same as on the left.
The code used for Sint Maarten (Dutch Part) is the M49 code of Saint Martin (French Part).	Same as on the left.
The code used for Sudan is the M49 code of old Sudan before 2011 when South Sudan split from Sudan. South Sudan is listed separately as origin country.	Same, except that South Sudan does not report as a destination country (see above).
Yap State is listed as an origin country. Yet, Yap State is not an independent country with a separate code. Moreover, the code used is the current M49 code of Panama.	Yap State is not listed as a destination country.
"Saba", "Hawaii, USA", "Wake Island", "Kosrae State", "Dubai", "Johnston Island", "Truk State", "Midway Islands", and "Pohnpei State" are listed as origin countries. Yet, they are not independent states with a separate code. The codes used are currently not assigned to any country.	Except Saba, none of the other "countries" are listed as destinations.

A2.2 Odd origin countries in UNWTO dataset

Our next step is to address the issue that the list of origin countries contains "odd" travel origin categories. Odd categories of travel origin refer to cases where (i) either no single statistical unit is specified as an origin country but aggregates and groups (e.g., USSR or ASEAN countries) or where (ii) origins are specified that do not constitute a statistical unit themselves but are part of a statistical unit (e.g., constituent countries of Netherlands Antilles). Figure A1 shows the distribution of these categories in our data. To solve these issues, we largely follow the steps outlined by Recchi et al. (2019). However, we extend their approach by also aggregating origins that are merely parts of a statistical unit.

Category 1 (normal cases):

Fortunately, a majority of observations are "normal" cases, i.e., where cross-border travels are assigned to a pair of individual countries with one destination country (the reporting country) and one origin country. The normal cases cover 90.2% of all cross-border travel in our data.

Category 2 (country pairs): A second category of travel origin refers to country pairs. These make up 2.2% of the cross-border travel in our data. This category includes:

> "Australia, New Zealand", "Belgium / Luxemburg", "Canada, United States", "China + Hong Kong, China", "United Kingdom/Ireland", "Spain, Portugal", "India, Pakistan"

To increase the number of non-missing observations, we impute "normal" cases from Category 2 (country pairs). Recchi et al. (2019) suggest splitting the reported number of arrivals using weights that depend on the countries' population size and general propensity to get involved in tourism. Yet, in contrast to Recchi et al., we no longer weigh by the propensity but only by population size. The reasoning is that we treat each travel category separately, so we would have to compute a tourism propensity for an origin country within each category. However, this implies that the propensities depend on which destination countries report travel from that origin in the category. The computed category-specific propensities would not represent the overall travel demand of a country's population. So, we weigh only by population size and split the reported arrivals from a country pair accordingly. Let

- A_j= Total number of arrivals reported for the origin country pair p (e.g., USA and Canada) in the destination country j
- P_i= Population of country i (e.g., USA) in the corresponding year
- P_k= Population of "the other" country k (e.g., Canada) in the corresponding year

A_{ij}, the portion of arrivals in destination j attributed to origin country i are then computed as

$$A_{ij} = A_{pj} * \frac{P_i}{P_i + P_k}$$

A_{kj}, the portion of arrivals attributed to the other origin k are computed analogously.

There were 41,237 cases of category 2 observations. Hence, the imputation creates an additional 41,237 observations because the origin country pairs a split in two. Yet, because some countries list country pairs (i.e., category 2) as well as the countries separately, this creates duplicate observations. After the imputation, we had 82, 085 excess copies, and we had to decide which to keep and which to drop. First, we dropped all the excess observations created through the imputation that had missing tourist flows. In other words, we dropped the empty new observations. This left us with 1,221 excess observations. Second, we replaced the original observations with missing tourist flows by non-missing imputed observations. This left us with 488 excess copies, where one copy is the original reported value and the other is the imputed value. So, third, we dropped the observations with the imputed values, placing more trust in the ones originally reported. After this, there are no further duplicate observations in the dataset.

Category 3 (nationals residing abroad):

This category refers to the inward travel of nationals who reside abroad. Here, origins are aggregated into one category. Yet, because we cannot assign them to the correct travel corridors, we drop the observations with the origin “nationals residing abroad”.

Category 4 (defined groups):

The fourth category includes defined groups of countries with more than 2 countries. These include:

> "USSR", "Scandinavia", "Yugoslavia", "Benelux", "Baltic countries", "ASEAN countries", "Commonwealth Independent States", "Windward Islands"

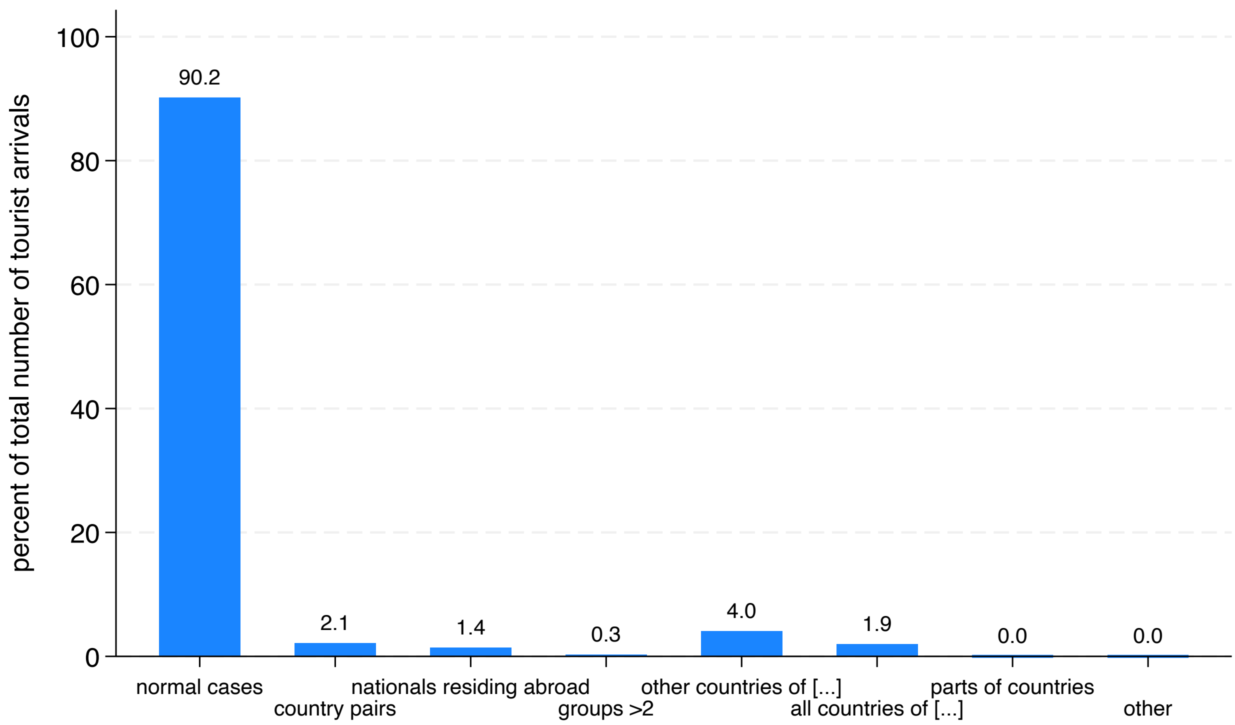


Figure A1 "Odd" travel origin categories (percent of the number of arrivals)

In this category, there are “convenience” groups, such as Benelux, Baltic countries, or ASEAN countries. Still, it also contains former unions of countries, e.g., USSR or Yugoslavia, that are now separate countries. Although it is, in principle, possible to impute values similar to what we did for the corridors in Category 2, we refrain from doing so here because the imputation will become more imprecise the more origin countries are grouped together, e.g., 15 countries for the former USSR.

Category 5 (other countries of [...])

This category refers to groupings of other origin countries in a world region, e.g., South East Asia, Central America, etc., or "the World". We drop observations with "other countries of ... " origins because we cannot assign them to the correct travel corridors.

Category 6 (all countries of [...])

This category refers to aggregates of all countries in a world region, e.g., all countries in the Caribbean, all countries in Northern Europe, etc. We drop observations with "other countries of ... " origins because we cannot assign them to the correct travel corridors and because they are aggregates, which will lead to double counting.

Category 7 (parts of countries)

This category refers to origins that are not a statistical unit in themselves but form a statistical unit together with others. In the UNWTO data, there are two such cases:

- Bonaire, Sint Eustatius and Saba: Bonaire, Sint Eustatius, Saba
- Micronesia (Federated States of): Kosrae, Yap, Truk, Pohnpei

Neither the unit Bonaire, Sint Eustatius and Saba (BES, 535) nor the unit Micronesia (FSM, 583) appear in the list of origin countries.

We aggregate corridors that have one of the above constituting units as origins. More specifically, for each destination country, when data is available for arrivals from each of the constituting units, we aggregate the number of arriving visitors per year and keep this aggregate tourism flow instead of the separate ones. If data is missing for arrivals from one or more of the constituting units, we drop all the respective observations.

Bonaire, Sint Eustatius, and Saba are also among the *destination* countries. Therefore, we also aggregate corridors with these units as destinations if all three report arrivals from the same origin. If only one or two of the three report travelers from an origin, we drop these observations.

Category 8 (other)

This category refers to corridors whose origin is given as "Hawaii, USA", "Wake Island", "Dubai", "Midway Islands", or "Johnston Island". Since we cannot know whether the visitors who arrive from these origins are already counted as arrivals from the USA or UAE, we drop the respective observations.

A2.3 Overhang observations

Finally, UNWTO occasionally reports "overhanging" observations–that is, tourist arrivals from origins or destinations that do not exist anymore/yet in the year in question. So, the corridor series are longer than expected. To make the inbound files consistent with the universe of recognized UN states, we attempt to merge corridors or split them as we did for the anomalous travel origins/destinations above.

In the Sabre air passenger data and Abel's migration flow estimates, we only encounter a handful of overhanging observations but none of the other issues. We solve the overhanging issue as described for the UNWTO data.

A3 UNWTO unified visitor flows details

Overall, we have 512,559 observations across all arrival categories. Yet, countries may report data in more than just one category. This means that there is some overlap between arrival categories and we have 391,195 origin-destination-year observations in our dataset for which *at least* one arrival category is reported. Table A3 shows the number of non-missing observations per category; the diagonal cells show total non-missing observations per category; the off-diagonal cells show the overlap of non-missing observations between two categories. An overview of which country reports which categories can be found in Table A5.

Table A3 Non-missing observations by tourism travel category

	122	**121**	**112**	**111**	**1912**	**1911**	**712**	**711**
122	86,948							
121	6,285	86,714						
112	16,586	903	129,753					
111	134	17,931	8,197	92,877				
1912	5,340	2,636	4,629	189	38,311			
1911	0	6,473	607	4,870	0	16,070		
712	6,331	3,876	7,931	1,300	26,488	0	41,068	
711	0	5,768	1,916	8,802	0	12,010	0	20,818

As described in the main text, our transformation procedure consists of two steps. The first is the simplifying assumption that a person's country of citizenship is their country of residence. This assumption allows us to reduce the eight arrival categories to four: VFR, TFR, THSR, and TCER. Then we transform these remaining four categories into VFR arrival counts. To do so, we first regress the VFR counts on each of the other categories k when VFR counts and category k counts are available. We also include several control variables; distance, contiguity, landlocked-ness, and being an island state.

$$v_{ijt}^{VFR} = \exp\left(\beta_0 + \beta_1 \ln v_{ijt}^{k} + \beta_2 \ln dist_{ij} + \beta_3 contig_{ij} + \beta_4 ll_i + \beta_5 ll_j + \beta_6 is_i + \beta_7 is_j\right)$$

Using the resulting parameter estimates we then predict what would be the VFR counts in all the cases where only category k counts are available. This yields our estimates of visitor flows. Descriptive statistics of the resulting $trips^A$ measure can be found in Table A4; alongside intermediate variables in our procedure and GTMD1.0-style tourist flows, which we assembled using our current data.

Table A4 Descriptive statistics for trips[A]

Variable	Obs	Mean	Std. Dev.	Min	Max (in million)	Totals in 2018 (in million)
tourist flows	391195	68488	831297	0	81.1	1542.9
visitor flows	391195	145820	1649026	0	134.8	3049.6
return visitor flows	391195	145820	1649026	0	134.8	3049.6
visitor + return visitors	184232	577607	4162892	1	151.5	5803.5
Abel migration flows	1551486	340	4913	0	0.6	22.0
trips[A]	184232	579217	4166201	1	151.5	5817.6

Table A5 UNWTO arrival categories by country

	TFR (112)	TFN (111)	VFR (122)	VFN (121)	THSR (1912)	THSN (1911)	TCER (712)	TCEN (711)
country name destination								
Albania	.	.	.	1	.	.	.	.
Algeria	.	.	.	1	.	.	.	.
American Samoa	.	1	.	.	.	.	.	.
Andorra	1	.	.	.	.	.	.	.
Angola	1	.	.	.	.	.	.	.
Anguilla	1	.	1	.	.	.	.	.
Antigua and Barbuda	1	.	.	.	.	.	.	.
Argentina	1	.	.	.	.	.	.	.
Armenia	1	.	.	.	.	.	.	.
Aruba	1	.	.	.	.	.	1	.
Australia	.	.	1	.	.	.	.	.
Austria	.	.	.	.	1	.	1	.
Azerbaijan	.	.	1	.	.	.	1	.
Bahamas	1	.	.	.	.	.	.	.
Bahrain	.	.	1	1	.	.	.	.
Bangladesh	.	1	.	.	.	.	.	.
Barbados	1	.	.	.	.	.	.	.
Belarus	.	1	.	1	.	.	.	.
Belgium	.	.	.	.	1	.	1	.
Belize	.	1	.	1	.	.	.	.
Benin	1	.	.	.	.	.	.	.
Bermuda	1	.	.	.	.	.	.	.
Bhutan	.	1	.	.	.	.	.	.
Bolivia (Plurinational State of)	.	1	.	.	.	.	.	1
Bonaire	1	.	.	.	.	.	.	.
Bosnia and Herzegovina	.	.	.	.	1	.	.	.
Botswana	1	.	.	.	.	.	.	.
Brazil	1	.	.	.	.	.	.	.
British Virgin Islands	1	.	1	.	.	.	.	.
Brunei Darussalam	.	1	.	1	.	.	.	.
Bulgaria	.	.	1	.	1	.	1	.
Burkina Faso	.	.	.	.	.	.	1	1
Burundi	.	1	.	.	.	.	.	.
Cabo Verde	.	.	.	.	.	.	1	.
Cambodia	1	.	.	.	.	.	.	.
Cameroon	.	.	.	1	.	.	.	1
Canada	1	.	1	.	.	.	.	.
Cayman Islands	1	.	.	.	.	.	.	.
Central African Republic	.	1	.	.	.	.	.	.
Chad	.	1	.	.	.	.	.	1

Chile	.	1	.	.	.	.	.	.
China	.	.	.	1	.	.	.	.
China, Hong Kong SAR	1	.	1	.	.	.	.	.
China, Macao SAR	.	.	1	1	.	.	1	.
Colombia	1	.	.	1	.	.	.	.
Comoros	.	1	.	.	.	.	.	.
Congo	.	.	1	.	.	.	1	.
Cook Islands	1	.	.	.	.	.	.	.
Costa Rica	.	1	.	.	.	.	.	.
Croatia	.	.	.	.	1	.	1	.
Cuba	.	.	1	.	.	.	.	.
Curaçao	1	.	.	.	.	.	1	.
Cyprus	1	.	.	.	1	.	1	.
Czechia	.	.	.	.	.	1	.	1
Côte dIvoire	1	.	1	.	.	.	.	.
Democratic Republic of the Congo	1	1	.	.	.	.	.	.
Denmark	.	.	.	.	1	.	1	.
Dominica	1	.	.	.	.	.	.	.
Dominican Republic	1	.	.	.	.	.	.	.
Ecuador	.	.	1	1	.	.	.	.
Egypt	.	.	.	1	.	.	.	.
El Salvador	.	1	.	.	.	.	.	1
Eritrea	.	.	.	1	.	.	.	.
Estonia	.	.	.	.	1	.	1	.
Eswatini	.	.	1	.	.	.	1	.
Ethiopia	1	.	.	.	.	.	.	.
Fiji	1	.	.	.	.	.	.	.
Finland	.	.	1	.	1	.	.	.
France	1	.	.	.	1	.	1	.
French Guiana	1	.	.	.	.	.	.	.
French Polynesia	1	.	.	.	.	.	1	.
Gabon	.	1	.	.	.	.	.	.
Gambia	.	1	.	.	.	.	.	.
Georgia	.	.	1	.	.	.	1	.
Germany	.	.	.	.	1	.	1	.
Ghana	.	1	.	.	.	.	.	.
Greece	1	.	.	.	1	.	1	.
Grenada	.	1	.	.	.	.	.	1
Guadeloupe	1	.	.	.	.	.	1	.
Guam	1	.	.	.	.	.	.	.
Guatemala	.	.	1	1	.	.	.	.
Guinea	1	1	.	.	.	.	1	.
Guinea-Bissau	.	1	.	.	.	.	.	.
Guyana	1	.	.	.	.	.	.	.
Haiti	1	.	.	.	.	.	.	.
Honduras	.	1	.	.	.	.	.	.
Hungary	.	1	.	1	.	1	.	1
Iceland	.	1	.	.	.	1	.	1
India	.	1	.	1	.	.	.	.
Indonesia	.	.	1	1	.	.	1	.
Iran (Islamic Republic of)	.	.	.	1	.	.	.	.
Iraq	.	.	.	1	.	.	.	.
Ireland	1	.	.	.	.	.	.	.
Israel	1	.	1	.	.	.	1	.
Italy	.	1	.	1	.	1	.	1
Jamaica	1	.	.	.	.	.	.	.
Japan	.	.	.	1	.	.	.	.
Jordan	.	1	.	1	.	.	.	.

Kazakhstan	.	.	1	.	.	.	.	.
Kenya	.	.	1	.	.	.	.	.
Kiribati	.	1	.	.	.	.	.	.
Kuwait	.	.	.	1	.	.	.	.
Kyrgyzstan	1	.	1	.	.	.	.	.
Lao People's Democratic Republic	.	.	.	1	.	.	.	.
Latvia	.	.	1	.	1	.	.	.
Lebanon	.	1	.	.	.	.	.	.
Lesotho	.	.	1	.	.	.	.	.
Libya	.	.	.	1	.	.	.	.
Liechtenstein	.	.	.	.	1	.	1	.
Lithuania	1	.	.	.	1	.	1	.
Luxembourg	.	.	.	.	1	.	1	.
Madagascar	.	1	.	.	.	.	.	.
Malawi	1	.	.	.	.	.	.	.
Malaysia	1	1	.	.	.	.	1	.
Maldives	.	1	.	.	.	.	.	.
Mali	1	1	.	.	.	.	1	1
Malta	1	.	.	.	.	.	1	.
Marshall Islands	1	1	.	.	.	.	.	.
Martinique	1	.	.	.	1	.	1	.
Mauritius	1	.	.	.	.	.	.	.
Mexico	1	1	.	.	.	.	.	.
Micronesia (Federated States of)	1	.	.	.	.	.	.	.
Monaco	.	.	.	.	.	.	.	1
Mongolia	.	1	.	1	.	.	.	.
Montenegro	.	.	.	.	.	1	.	1
Montserrat	1	.	1	.	.	.	.	.
Morocco	.	1	.	.	.	.	.	1
Mozambique	.	.	1	.	.	.	.	.
Myanmar	.	1	.	.	.	.	.	.
Namibia	.	1	.	1	.	.	.	.
Nepal	1	1	.	.	.	.	.	.
Netherlands (Kingdom of the)	.	.	.	.	1	.	1	.
New Caledonia	1	.	.	.	.	.	1	.
New Zealand	.	.	1	.	.	.	.	.
Nicaragua	.	1	.	.	.	.	.	.
Niger	.	1	.	.	.	.	.	.
Nigeria	.	.	.	1	.	.	.	.
Niue	1	.	.	.	.	.	.	.
North Macedonia	.	.	.	.	.	1	.	1
Northern Mariana Islands	.	.	.	1	.	.	.	.
Norway	1	.	.	.	1	.	1	.
Oman	.	.	1	1	.	.	.	.
Pakistan	.	1	.	.	.	.	.	.
Palau	.	1	.	.	.	.	.	.
Panama	.	.	1	.	.	.	.	.
Papua New Guinea	1	.	1	.	.	.	.	.
Paraguay	.	1	.	.	.	.	.	.
Peru	1	.	.	.	.	.	.	1
Philippines	1	.	.	.	1	.	1	.
Poland	.	.	.	1	1	.	1	.
Portugal	1	1	1	1	1	.	1	.
Puerto Rico	1	.	.	.	.	.	1	.
Qatar	.	.	.	1	.	.	.	.
Republic of Korea	.	.	.	1	.	.	.	.
Republic of Moldova	.	.	.	.	.	1	.	.
Romania	.	.	1	.	1	.	1	.

Russian Federation	.	.	.	1	.	.	.	.
Rwanda	.	.	.	1	.	.	.	.
Réunion	1	.	.	.	.	.	.	.
Saba	1	.	.	.	.	.	.	.
Saint Kitts and Nevis	1	.	.	.	.	.	.	.
Saint Lucia	1	.	.	.	.	.	.	.
Saint Vincent and the Grenadines	1	.	1	.	.	.	.	.
Samoa	1	.	1	.	.	.	.	.
San Marino	.	.	.	1	.	.	.	.
Sao Tome and Principe	.	1	.	.	.	.	.	.
Saudi Arabia	1	.	.	.	.	.	.	.
Senegal	.	.	.	.	.	.	.	1
Serbia	.	.	.	.	.	1	.	1
Seychelles	1	.	.	.	.	.	.	.
Sierra Leone	1	.	.	.	.	.	.	.
Singapore	.	.	1	.	.	.	.	.
Sint Eustatius	1	.	.	.	.	.	.	.
Sint Maarten (Dutch part)	.	1	.	.	.	.	.	.
Slovakia	.	.	.	.	.	1	.	.
Slovenia	.	.	.	.	.	1	.	1
Solomon Islands	1	.	.	.	.	.	.	.
South Africa	1	.	1	.	.	.	.	.
Spain	1	.	.	.	1	.	1	.
Sri Lanka	1	1	.	.	.	.	.	.
State of Palestine	.	.	.	.	.	.	.	1
Sudan (old)	.	1	.	.	.	.	.	.
Suriname	1	.	.	.	.	.	.	.
Sweden	1	.	1	.	1	.	1	.
Switzerland	.	.	.	.	1	.	1	.
Syrian Arab Republic	1	.	.	1	.	1	.	1
Taiwan, Province of China	.	.	1	.	1	.	.	.
Tajikistan	.	.	1	.	.	.	.	.
Thailand	1	1	.	.	.	.	.	1
Timor-Leste	1	.	.	.	.	.	.	.
Togo	.	.	.	.	.	.	1	.
Tonga	1	.	.	.	.	.	.	.
Trinidad and Tobago	1	1	.	.	.	.	.	.
Tunisia	.	1	.	.	.	.	.	1
Turkmenistan	.	1	.	.	.	.	.	.
Turks and Caicos Islands	1	.	.	.	.	.	.	.
Tuvalu	.	1	.	.	.	.	.	.
Türkiye	.	1	.	1	.	1	.	1
Uganda	1	.	.	.	.	.	.	.
Ukraine	1	.	.	.	.	.	.	.
United Arab Emirates	.	.	.	1	.	1	.	1
United Kingdom of Great Britain and Northern Ireland	.	.	1	.	.	.	.	.
United Republic of Tanzania	.	.	1	.	.	.	.	.
United States Virgin Islands	.	.	.	.	.	.	.	1
United States of America	1	.	.	.	.	.	.	.
Uruguay	.	.	.	1	.	.	.	.
Uzbekistan	.	.	1	.	.	.	.	.
Vanuatu	1	.	.	.	.	.	.	.
Venezuela (Bolivarian Republic of)	.	1	.	1	.	.	.	.
Viet Nam	.	.	1	.	.	.	.	.
Yemen (old)	.	1	.	.	.	.	.	.
Zambia	1	.	.	.	.	.	.	.

A4 Distance-corrected air passenger volumes

We test four different distance corrections of the Sabre passenger volumes. The first is the procedure described by Recchi et al. (2019) which we call RDV19 correction. See their paper for details.

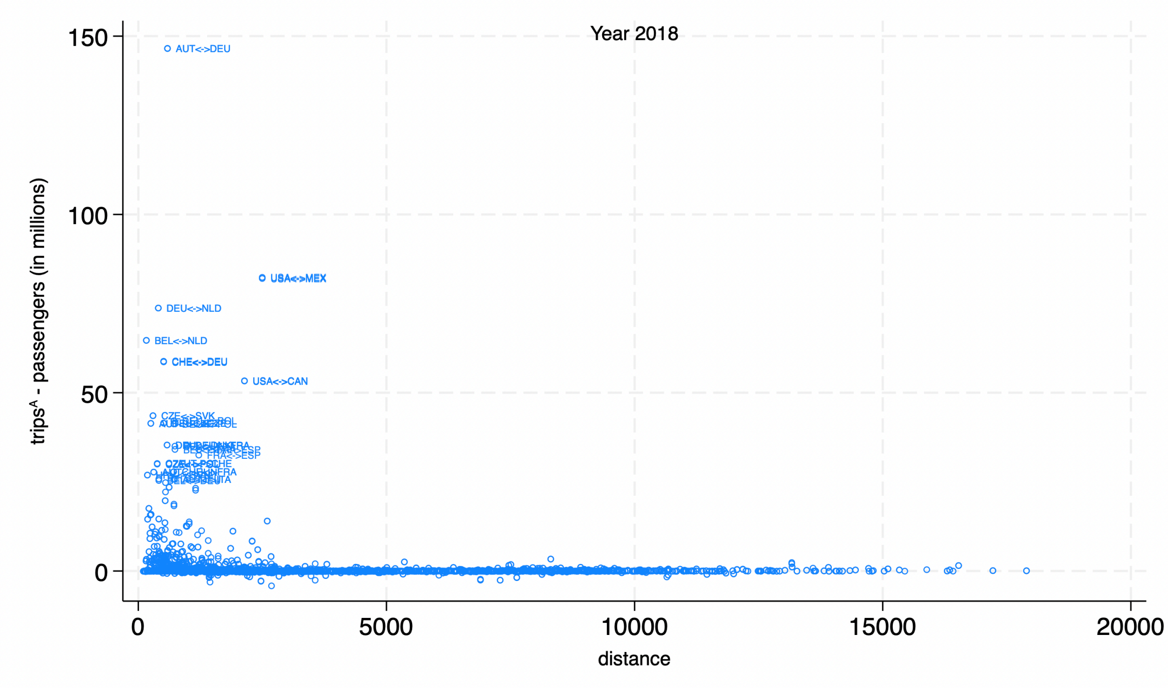


Figure A2 The greater the distance between countries, the more similar air passenger volumes and trips A

The other three are all based on Poisson regressions using pseudo maximum likelihood estimation following the considerations in Section 4.4 of our paper. We tested the following specifications

a) $trips_{ijt}^{A} = \exp(\beta_0 + \beta_1 \ln air_{ijt} + \beta_2 \ln dist_{ij})$

b) $trips_{ijt}^{A} = \exp(\beta_0 + \beta_1 \ln air_{ijt} + \beta_2 \ln dist_{ij} + \beta_3 contig_{ij} + \beta_4 ll_i + \beta_5 ll_j + \beta_6 is_i + \beta_7 is_j)$

c) $trips_{ijt}^{A} = \exp(\beta_0 + \beta_1 \ln air_{ijt} + \delta_{it} + \delta_{jt} + \delta_{ij})$

where δ_{it}, δ_{jt}, and δ_{ij} are fixed effects.

Table A6 shows a comparison between the results obtained by the different methods. We use “ppml correction b)” as our preferred method of correcting the air passenger data because it maximizes the correlation with $trips^A$ without compromising the number of observations we can predict from it.

Figure A3 Uncorrected and distance-corrected air passenger volumes vs. trips^A

Table A6 Results of distance correction

Panel A: Descriptive Statistics

Variable	Obs	Mean	Std Dev	Min	Max
$trips^A$	184232	579217.09	4166201.5	1	1.515e+08
Passengers, uncorrected	642679	24588.907	258068.91	0	21113634
passengers, RDV19 correction	642651	193456.43	2451616.5	0	1.967e+08
passengers, ppml correction a)	642651	151813.98	1672440.2	28.656	2.375e+08
passengers, ppml correction b)	642651	131164.32	1386604.2	42.006	79554786
passengers, ppml correction c)	101873	567204.24	4218394.3	.255	1.527e+08

Panel B: Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) $trips^A$	1.000					
(2) passengers, uncorrected	0.437	1.000				
(3) passengers, RDV19 correction	0.477	0.891	1.000			
(4) passengers, ppml correction a)	0.496	0.597	0.824	1.000		
(5) passengers, ppml correction b)	0.758	0.410	0.568	0.702	1.000	
(6) passengers, ppml correction c)	0.999	0.436	0.477	0.497	0.760	1.000

A5 Missing data imputation for Series 2

As described in the paper, we apply two imputation methods to produce Series 2 of our estimates of border-crossings.

A5.1 Linear Interpolation

First, we use linear interpolation to impute data to both $trips^A$ and $trips^B$ in order to minimize the number breaks in the respective time series. This is not a large-scale, complex exercise in imputation; rather, the aim is to fill gaps and obtain estimates for a small subset of the missing values in the dataset. Specifically, based on the findings of simulation studies in interpolation (Liu et al. 2019), we impute so-called isolated missing values using linear interpolation. In our case, isolated missing values are missing cells within a corridor for which there is data in the preceding and succeeding years; for example, we may have data for the years 2008 and 2010 in the ITA->JPN corridor but not for 2009. The known values are then leveraged to provide a gross estimation for the time series gap. We thus define a time threshold to be $\mathrm{Th}_t = 2$, meaning there can be no more than two missing values in a row in the time series of a given corridor *i* for them to be interpolated. Concretely, for an IMV in year *t* and corridor *i*, the estimation is

$$\hat{x}_{i,t} = \frac{1}{2}\left(x_{i,t-1} + x_{i,t+1}\right)$$

when the two directly neighboring values are known and

$$\hat{x}_{i,t} = x_{i,t-1} + \frac{1}{3}\left(x_{i,t+2} - x_{i,t-1}\right)$$
$$\hat{x}_{i,t+1} = x_{i,t-1} + \frac{2}{3}\left(x_{i,t+2} - x_{i,t-1}\right)$$

when two values are missing consecutively in the time series. Any observation that are consecutively missing for a period longer than the above threshold are left as missing. This intervention is minimally invasive, and adds smoothness to the various time series.

A5.2 Back-casting of $trips^B$

To impute the missing $trips^B$ values between 1995 and 2009, we backwards predict these values from the non-missing observations in the period 2010-2019. We don't use the years 2020-22 as flight travel was severely affected by the COVID-19 pandemic. For the back-casting we use coefficients from a gravity model of $trips^B$ which is specified as:

$$trips_{ijt}^{B} = \exp\left(\beta_0 + \beta_1 \ln trips_{ijt+1}^{B} + \cdots + \beta_5 \ln trips_{ijt+5}^{B} + \beta_6 \ln \sum_{ij} trips_{ijt}^{A} + \beta_7 \ln dist_{ij} + \beta_8 contig_{ij} + \beta_9 ll_i + \beta_{10} ll_j + \beta_{11} is_i + \beta_{12} is_j\right) * \epsilon_{ijt}$$

The model includes the following elements:

- $trips^B_{ijt+1}, \dots, trips^B_{ijt+5}$ represent forward lags of $trips^B$ to capture the respective 5-year trend of $trips^B$ in the 2010-2019 period
- $\sum_{ij} trips^A_{ijt}$represent annual totals of $trips^A$ to capture global levels of international travel
- The other variables are geographical variables: distance, contiguity, landlocked-ness and being an island state

The procedure to back-cast $trips^B$ is as follows. The first backwards prediction of $trips^B$ in 2009 is obtained by estimating the model using the 2010-2014 sample. This ensures that there is a full 5-year trend for each year in the estimation sample, i.e. that also for $trips^B$ in 2014, we have the forward lags in 2015, 2016, 2017, 2018, and 2019. Then, we compute $\widehat{trips}^B_{2009}$ using the estimated coefficients and the observed values of our righthand-side variables in the years 2010-2014.

We repeat this process until we reach the year 1995. That is, to predict values of $trips^B$ in 2008 we estimate the model on the 2009-2013 sample, which now includes the predicted 2009 values from the previous round of estimation, and so on.

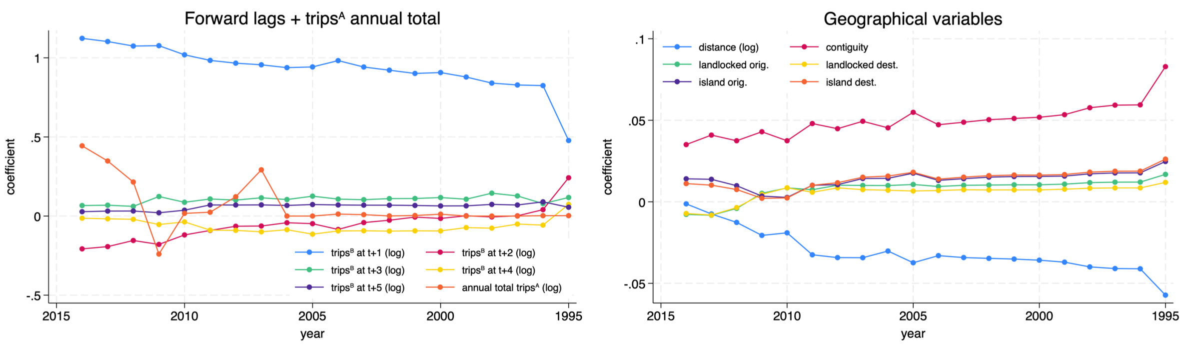


Figure A4 Results of back-casting estimations

Figure A4 shows the size of the β-coefficients from each round of estimations. The first forward lag has the largest effect size among the lags. The coefficient around +1 indicates that levels of international travel within corridors remain fairly stable from year to year. The negative coefficient on the second lag indicates the steepness of the series. Distance has a negative effect, which get bigger in size, throughout, and contiguity between countries has a positive effect.

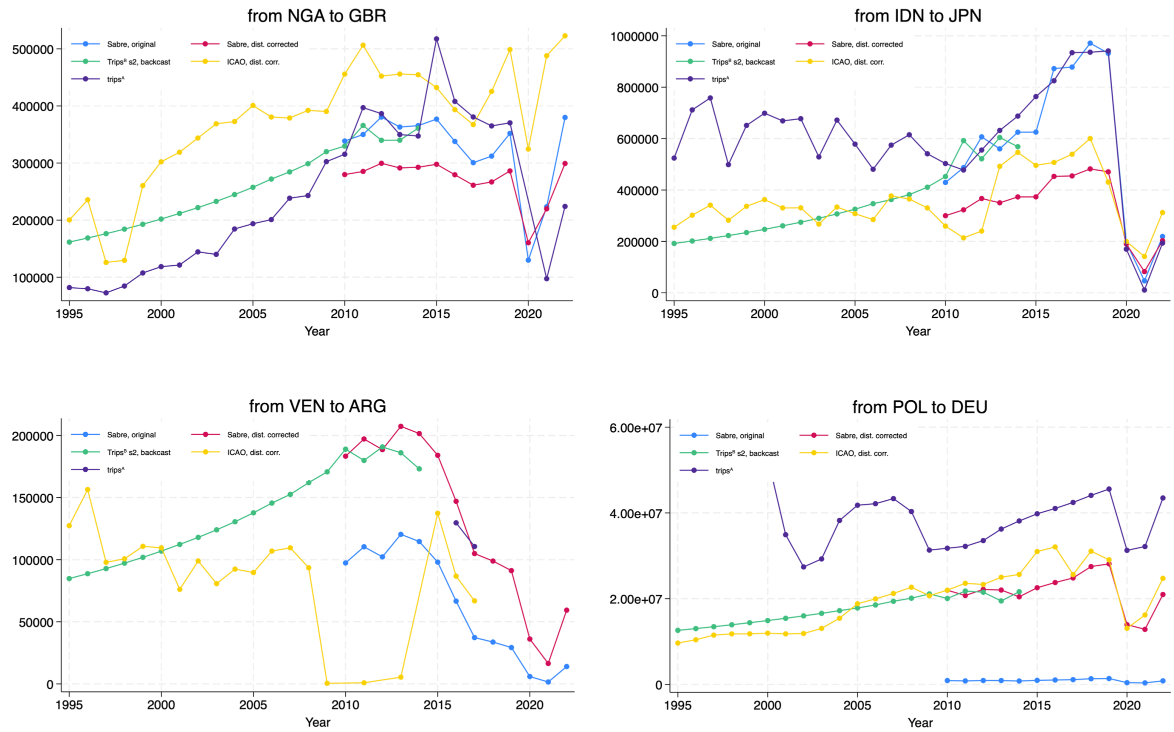


Figure A5 Back-casted $trips^B$ in selected corridors

We test and verify our predictions in three ways. First, we conduct spot-checks in selected corridors to gauge the plausibility of our estimates. The plots in Figure A5 suggests that our back-casted $trips^B$ values are plausible compared to $trips^A$ (purple) and ICAO data (yellow), which is an alternative data source for air passenger data available in the 1995-2009 period, albeit with much lower country coverage.

Second, we predict $trips^B$ values not just for 1995-2009 but also in the years 2010-2014 and compare them to the observed values in those years. The scatterplots in Figure A6 suggest well-behaved and largely accurate predictions.

Finally, we compare our 1995-2009 back-casted values to our $trips^A$ measure. Again, the scatterplots in Figure A7 suggest well-behaved and largely accurate predictions.

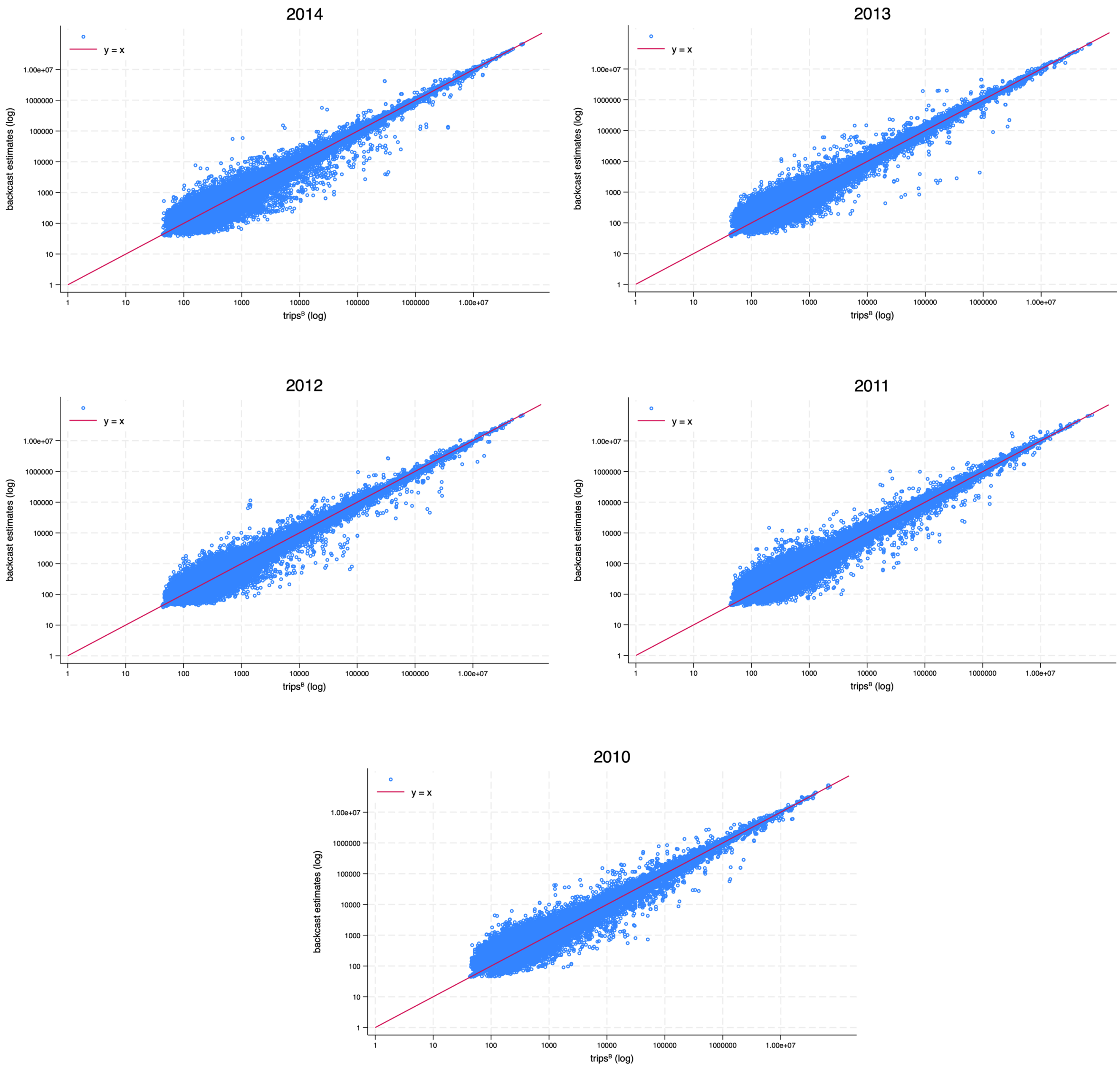


Figure A6 Back-cast trips^B vs. observed trips^B

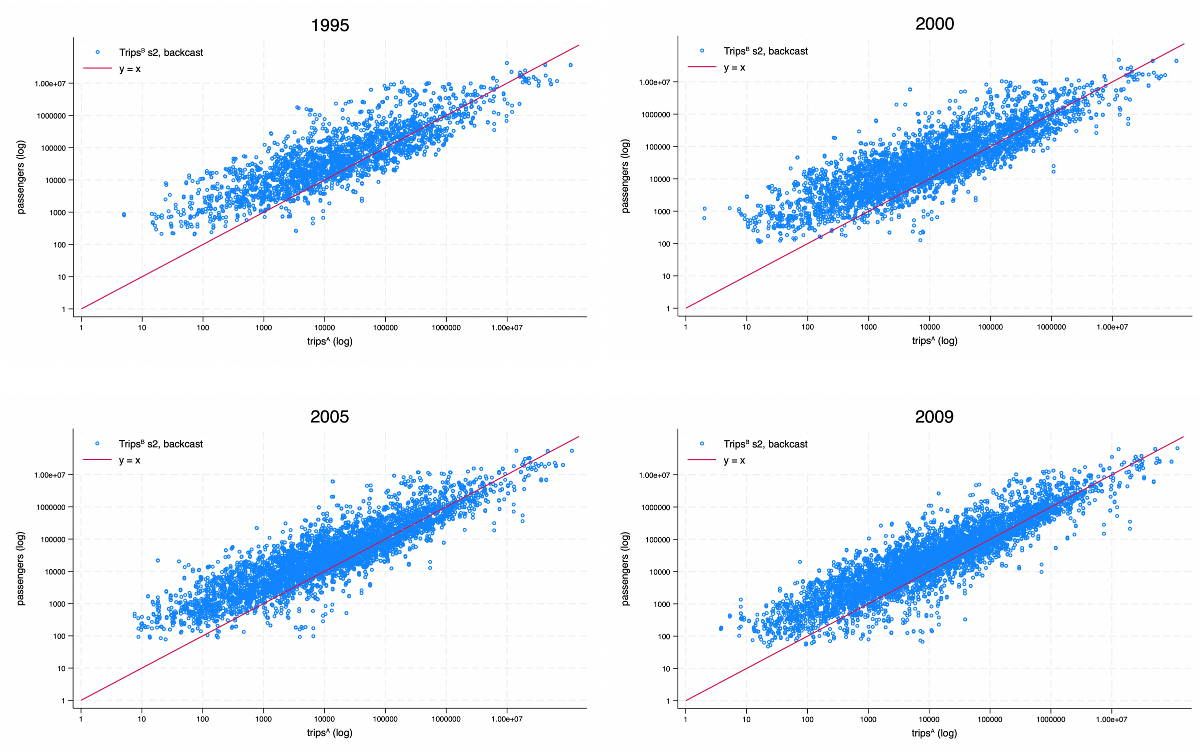


Figure A7 Back-cast trips^B vs trips^A

A6 Variables in the Dataset

Variable name	Description	Type
code_i	M49 code origin country	numerical
code_j	M49 code origin country	numerical
country_i	country name origin	string
country_j	country name destination	string
year	year	numerical
iso3code_i	ISO-alpha3 code destination	string
iso3code_j	ISO-alpha3 code destination	string
iso3code_ij	ISO-alpha3 code corridor	string
code_start_i	first year of origin country	numerical
code_end_i	last year of origin country	numerical
code_start_j	first year of destination country	numerical
code_end_j	last year of destination country	numerical
gtmd2_trips_s1	border crossings estimates, Series 1	numerical
gtmd2_trips_s1_source	source of Series 1 estimates: trips^A or trips^B	string
gtmd2_trips_s2	border crossings estimates, Series 2	numerical
gtmd2_trips_s2_source	source of Series 2 estimates: trips^A or trips^B, backcast, imputed	string
gtmd2_strregion_i	world region of origin country	string
gtmd2_strregion_j	world region of destination country	sting

References

Recchi, Ettore, Emanuel Deutschmann, and Michele Vespe. 2019. "Estimating Transnational Human Mobility on a Global Scale." *EUI Working Papers* RSCAS 2019/30.