Chapter 9
Assessing Transnational Human Mobility on a Global Scale

Emanuel Deutschmann, Ettore Recchi, and Michele Vespe

9.1 Introduction

Transnational mobility is the *sine qua non* of international migration. All international migrants have in common the basic fact of crossing (at least) one country border at some point of their migration trajectory. For a quantitative take on migration, thus, knowledge of the mobility flows of the world population amounts to a preliminary framing of global migration. Moreover, mobility data can contribute directly to understanding the scale of seasonal and other temporary forms of migration, which are hardly captured by official statistics (Gabrielli et al., 2019).

However, there is a surprising dearth of systematic information detailing the size of travel flows across countries worldwide. The Global Mobilities Project (GMP) at

E. Deutschmann (✉)
University of Flensburg, Flensburg, Germany

Migration Policy Centre, Robert Schuman Centre for Advanced Studies, EUI, Florence, Italy
e-mail: emanuel.deutschmann@uni-flensburg.de

E. Recchi
Sciences Po, Observatoire Sociologique du Changement (OSC), CNRS, Paris, France

Migration Policy Centre, Robert Schuman Centre for Advanced Studies, EUI, Florence, Italy

Institut Convergences Migrations, Paris, France
e-mail: ettore.recchi@eui.eu

M. Vespe
European Commission, Joint Research Centre (JRC), Ispra, Italy
e-mail: michele.vespe@ec.europa.eu

the European University Institute's Migration Policy Centre (MPC) aims to fill this gap by addressing different dimensions of transnational[1] mobilities (Recchi, 2017).

In this specific sub-project, we capitalize on two of the most comprehensive data sources on transnational human movements at a global scale:

1. Data on tourism, i.e., cross-border visits that include an overnight stay (*nota bene*: not necessarily for leisure), from the World Tourism Organization (UNWTO);
2. Data on cross-border air passenger traffic from Sabre, a private company that collects data directly from the airline industry.

Given that their data have been collected for different purposes, both sources, taken individually, have clear limitations when used in the attempt to provide insights into global human mobility. These limitations result in under-reporting of the scale of actual mobility across national borders. The data on tourism is incomplete in that people moving between countries for reasons other than tourism (in particular, returning residents) are not included. It is also distorted because visitors from some countries with few departures are not counted since their specific travel origin does not show up in the receiving country's tourism statistics. The data on air passenger traffic, in turn, does not factor in people who do not travel by airplane. In particular, journeys between neighboring countries, where cross-border mobility is particularly high (Deutschmann, 2016), are likely to be severely underestimated since people often use car, railway, or bus transportation rather than flights. We propose to remedy these systematic biases by combining and adjusting the two data sources, thereby producing more reliable estimates of cross-country human mobility globally. We describe the merging of these sources also as a possible precedent for similar endeavors for other types of country-to-country flows (like migration).

In the following sections, we firstly make general remarks about the composition of transnational mobility data in the two baseline sources and give an overview of the procedures followed to combine them (Sect. 9.2). We then describe these procedures in more detail in Sect. 9.3. Section 9.4 highlights some findings derived from the first explorations of the newly created dataset. In the conclusion (Sect. 9.5), we outline some pending limitations, advocate the use of this novel dataset to study transnational human mobility empirically in social science research and describe a set of general lessons from our project that might prove useful for other researchers embarking on similar endeavors.

---

[1]While we are aware that in the field of migration studies 'transnational' has a more demanding meaning that involves the regular movement of the *same* individuals across certain borders (Wimmer & Glick Schiller, 2002), we use the term 'transnational' in the meaning it has in the field of international relations, where it is employed to describe any movement by non-state actors that spans across national borders (Nye & Keohane, 1971).

9.2 Discerning the Composition of Transnational Mobility Flows

Our aim is to obtain robust estimates of the absolute number of yearly travels from and to every country worldwide. In formal terms, we set out to measure the volume of cross-border travels T across all pairs of sovereign states a, b, c, ... n on the planet. Such travels are carried out by both non-residents (NR) and residents (R) of receiving countries and take place by *air* (flights) or by *land/water* transportation (trains, buses, cars and other private road vehicles, boats, ferries and ships),[2] which we indicate by exponents ^{A} and ^{L}, respectively. Therefore:

$$T_{a \to b} = NR^A_{a \to b} + R^A_{a \to b} + NR^L_{a \to b} + R^L_{a \to b}$$

Unfortunately, no existing source contains information on all four components simultaneously. The original tourist files include only $NR^A_{a \to b} + NR^L_{a \to b}$, i.e., they register tourist *arrivals* in destination countries, but not tourists returning to their countries of origin.[3] Air traffic statistics include $NR^A_{a \to b} + R^A_{a \to b}$, i.e., air passengers only.[4] Thus, both datasets are suboptimal as they systematically exclude $R^L_{a \to b}$. Despite their differences, we expect the two datasets to be strongly correlated, because they share the same core component: $NR^A_{a \to b}$. They should diverge only when $R^A_{a \to b}$ and/or $NR^L_{a \to b}$ are large and/or not correlated.

The original UNWTO tourist files, however, also record residents of b going from b to a *with all transportation means*, that is $R^A_{b \to a}$ and $R^L_{b \to a}$. If we imagine that these people return to their country of residence in the same year of their outbound travel, we can count them as part of $R^A_{a \to b}$ and $R^L_{a \to b}$. We can thus assume that $R^A_{a \to b} + R^L_{a \to b} = R^A_{b \to a} + R^L_{b \to a}$. This assumption falls short of the travelers who: a) travel by the end of the year and come back in the following calendar year, or b) resettle abroad. As for a), we can maintain that these travelers are offset by similar travelers 12 months earlier. As for b), these travelers are migrants. A comparison of migration flows (in the most conservative estimate: Abel & Cohen, 2019, p. 8) and global tourist flows (in the conservative estimate of Deutschmann, 2016) shows a 1 to 98 relationship. That is, migrant travel corresponds to about 1% of tourist travels. Thus, 1% is the approximate maximum size of the error we introduce in our tourism

---

[2] Other statistically marginal forms of mobility (by foot or bike, for instance) are also included, provided they take place legally (i.e., they are registered). Unregistered or illegal border crossings are in fact left out by default from tourism and air traffic statistics, and, as a consequence, from our estimates.

[3] For a small number of receiving countries, returning residents are reported in the original UNWTO data, however without indication of where they are returning from. Thus, this information cannot be used in research interested in obtaining country-to-country flow estimates.

[4] Note that air traffic statistics do not allow us to distinguish between these two components since they are based on the location of the airport of origin and destination, not on the residence or nationality of travellers.

estimates through this assumption (see also Sect. 9.4). Conceptually, migration (be it voluntary or involuntary) is excluded from our estimates, even though we cannot rule out that some 'visitors' may overstay their travels and thus become migrants. More on this issue will be explored in the Conclusions (Sect. 9.5). We therefore revise the original UNWTO tourism data to build a yearly matrix of tourists/visitors travelling from a to b that also includes (returning) travellers from b who moved to a:

$$T_{a \to b}^{\text{revised}} = NR_{a \to b}^{A} + NR_{a \to b}^{L} + R_{b \to a}^{A} + R_{b \to a}^{L}$$

Hereafter, we will call this the GMP-revised tourism data [1]. Its creation is described in detail in Sect. 9.3.1.

As for the air passenger data, which we use in its KCMD-revised form [2] (see explanation below), we assume that they tend to be lower than the revised tourism data [1] because travelers also move by other means of transportation. However, [1] and [2] should converge progressively as the distance between origin and destination increases, given that air travel tends to become the exclusive means of transportation at long distances. This distance-mediated relationship between [1] and [2] leads us to transform the air passenger data. We compute an estimate of transnational mobility [3] that adjusts [2] by a factor that accounts for the distance between countries. The formal procedure to estimate [3] is described in Sect. 9.3.3.

In a final step, we combine the two revised sources, [1] and [3], to create an integrated dataset on global transnational mobility. As we hold that both [1] and [3] tend to underestimate actual mobility flows, our final estimate is always the largest of the two when we have both information—that is, either [1] or [3]. When we lack [3], we take [1], and vice versa.

Figure 9.1 provides an overview of this procedure. The individual steps are described in more detail in the following sections. The resulting final dataset covers 196 sender and receiving countries, generating a matrix of 38,220 cases (i.e., country pairs) per year. For the entire 2011–2016 period, about 9.5 billion trips (approx. 61%) are ultimately derived from [1] and 6 billion trips (approx. 38%) from [3]. Overall, 12.0% of cells are empty, which can mean either a total absence of transnational mobility between these countries (most likely in the case of pairs of small and distant nations) or missing data. The Global Transnational Mobility Dataset covers an estimated total of 15.7 billion trips.

9.3 Building the Dataset

In the following subsections, we outline in more detail how we handled the raw data and proceeded toward the production of the final Global Transnational Mobility Dataset. We first describe the creation of the GMP-revised tourism data (Sect. 9.3.1).
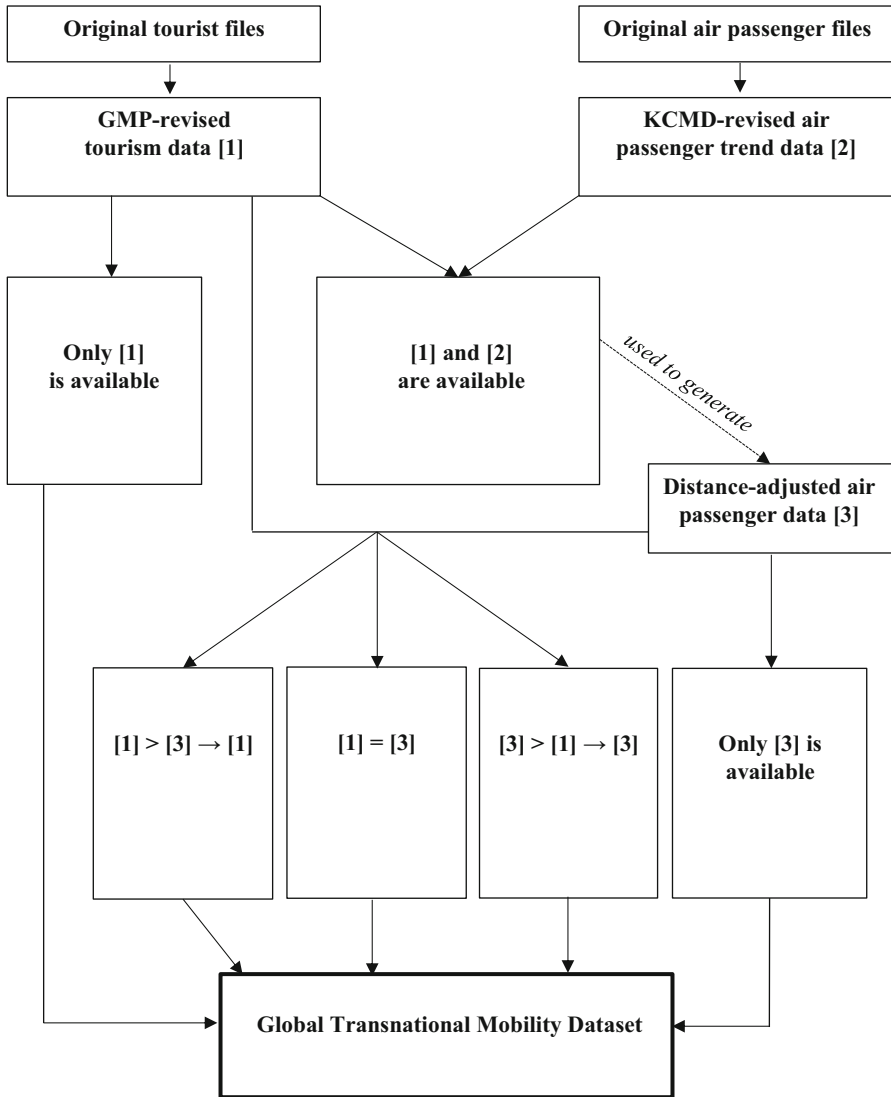
Fig. 9.1 Overview of the data composition

Second, we bring the KCMD-revised air passenger trend data in (Sect. 9.3.2). Third, we introduce the correction factor that adjusts the latter source, taking geographic distance into account (Sect. 9.3.3). Finally, we describe the merging and finalization of the dataset (Sect. 9.3.4).

9.3.1 Creating the GMP-Revised Tourism Data [1]

Our first source, the UNWTO tourism data, was obtained by the Global Mobilities Project (GMP) of the EUI's Migration Policy Centre (MPC) from the UNWTO as a set of files containing yearly flows from 1995 to 2016 for a global set of countries and territories worldwide (UNWTO, 2015).[5] While the harmonization and collection of national statistics on travels is part of the UNWTO mission, its online data are highly   aggregated   (see: https://www.unwto.org/unwto-tourism-dashboard, consulted December 18th, 2019). Therefore, we drew on the original country data kindly provided upon request by this organization. This dataset consists of 219 distinct files, one per receiving country/territory. To create a unified, standardized, and usable dataset (hereafter the GMP-revised tourism data), we took the following steps:

Step 1: Prioritizing the different UNWTO operationalizations of 'arrivals'

The country-to-country flow data on arrivals is reported in eight different categories in the UNWTO data (see Table 9.2 in the Appendix). The UNWTO defines arrivals—and describes its sources—as follows:

> *Arrivals* data measure the flows of international visitors to the country of reference: each arrival corresponds to one inbound tourism trip. If a person visits several countries during the course of a single trip, his/her arrival in each country is recorded separately. In an accounting period, arrivals are not necessarily equal to the number of persons travelling (when a person visits the same country several times a year, each trip by the same person is counted as a separate arrival).
>
> Arrivals data should correspond to *inbound visitors* by including both tourists and same-day non-resident visitors. All other types of travelers (such as border, seasonal and other short-term workers, long-term students and others) should be excluded, as they do not qualify as visitors. Data are obtained from different sources: administrative records (immigration, traffic counts, and other possible types of controls), border surveys or a mix of them. If data are obtained from accommodation surveys, the number of guests is used as estimate of arrival figures; consequently, in this case, breakdowns by regions, main purpose of the trip, modes of transport used or forms of organization of the trip are based on complementary visitor surveys. (UNWTO, 2015, p. 9).

To include as many cases as possible in the unified dataset, we use all eight 'arrivals' categories, in the order of preference shown in Table 9.2 in the Appendix.

Step 2: Creating a unified dataset

We then created a unified dataset that contains the relevant country-to-country flow data for all cases for which this information was available.[6] In doing so, we exclude

---

[5]At UNWTO, we thank Jacinta Mora for facilitating our access to these tourism statistics.

[6]There are 18 countries that are part of the UNWTO data collection that do not report arrivals by country of origin (which means they may be part of the full tourism dataset as senders of tourists but not as receivers): Afghanistan, Bonaire, Djibouti, Equatorial Guinea, Eritrea, Gabon, Ghana, Guinea-Bissau, Liberia, Libya, Mauritania, Saba, Sao Tome and Principe, Sint Eustatius, South Sudan, Syrian Arabic Republic, Turkmenistan, and United Arab Emirates.

several 'odd' sender categories, such as 'other countries of the world', which cannot readily be included in a country-to-country flow matrix. Details about this procedure and its consequences are described in Recchi et al. (2019a, Appendix).

Step 3: Adding returning residents

In line with the considerations made in Sect. 9.2, we add the returning residents $R^A_{b \to a} + R^L_{b \to a}$, to the incoming non-residents $NR^A_{a \to b} + NR^L_{a \to b}$ to obtain a more complete picture of human mobility across borders. In doing so, we effectively double the number of trips in the tourism dataset. Furthermore, the matrix becomes symmetric, i.e., mobility flows are now, by necessity, the same in both directions ($T^{\text{revised}}_{a \to b} = T^{\text{revised}}_{b \to a}$). Note that information is only added up if it was available in both directions. If one of the two values were missing (i.e., if information was available for the tie $a \to b$ but not for $b \to a$), the overall value was set to missing. This was done on the grounds that the overall information was considered unreliable when information in one direction was unavailable and that the other source (distance-adjusted air traffic data) is to be preferred.[7] After this step, we have obtained the GMP-revised tourism data [1].

9.3.2   Bringing in the KCMD-Revised Air Passenger Trend Data [2]

The second source is a dataset on global air passenger traffic collected by a private travel industry company, Sabre (2020). The dataset contains information on the total number of passengers flying between any two airports worldwide, regardless of whether the flights are direct or indirect. Here, we draw on a simplified and reduced version created by researchers at the European Commission's Knowledge Centre on Migration and Democracy (KCMD) that represents the yearly trend between countries (henceforth KCMD-revised air passenger trend data [2]). This version was generated through a time-series decomposition that dissects the raw overall air passenger flow between two countries into a trend component, a seasonal component, and a residual component (Gabrielli et al., 2019). In the KCMD-revised air passenger trend data [2] used here, the monthly trend data is aggregated to yearly averages. The data is available for the years 2011 to 2016.

   We merge the two datasets [1] and [2] using ISO 3166-1 alpha-3 country codes. In line with the considerations made in Sect. 9.2, we hypothesize:

(a) [1] to be on average larger than [2], as it includes both air passengers and land/ water travellers;

---

[7]We are indebted to Thomas Ginn for his close examination of our dataset, which led to the realization that this point was not clearly specified in a previous version of this manuscript.

(b) [1] and [2] to be highly correlated, since many travellers use flights to cross borders;

(c) [1] and [2] to be more strongly correlated as the distance between country pairs increases, since people are more likely to use air transportation at longer distances.

All three hypotheses hold empirically. As expected, tourism figures based on [1], where cross-border trips are reported with all transportation means, tend to be higher than air passenger figures based on [2], which report journeys that take place with flight transportation only. Table 9.1 shows the distribution of the deviations between the two data sources across cases (i.e., country pairs), by year. Negative values denote that there are more tourists than air passengers; positive values denote that there are more air passengers than tourists travelling between a pair of countries. The median (50th percentile) across years is −2410 trips, and even at the 75th percentile of cases, there are still more tourists than air passengers (−85 trips). Table 9.1 also reveals that, as the distribution is quite stable over time, the divergence between the two sources is no coincidence, but does indeed reflect the structural difference described above in hypothesis (a).

Figure 9.2 shows the relationship between the tourist-air passenger discrepancy and geographic distance (based on CEPII's GeoDist dataset [Mayer & Zignago, 2006]). A clear pattern emerges: there are only sizeable discrepancies at short geographic distances. The most extreme negative deviations (i.e., a lot more tourists than air passengers) are Hong Kong ↔ China (89–93 million, depending on year and direction), Macao ↔ China (37–43 million), United States ↔ Mexico (30–34 million), and Germany ↔ Poland (26–33 million). As Fig. 9.2 clearly shows, extreme cases consistently cluster together over time (different shapes represent different years). This suggests that these discrepancies are not random but systematic and meaningful. The inspection of specific cases with the highest negative[8] deviations helps to understand the rationales of the discrepancies, which can overlap and reinforce each other:

(a) Mobility between *nearby countries*: tourists exceed air passengers because many people move across borders with land (train, car, bus) or water (ferry, ship) transportation. Examples include the four extreme outlier country pairs tagged in Fig. 9.2.

(b) *Grand-tour tourism*: Here, people fly to one country (e.g., from the U.S. to the Netherlands), and then go by car or train to other countries (e.g., France). In these other countries, they are counted as tourists (e.g., through hotel registration data) but not as air passengers.

---

[8]In fact, there are few exceptional cases in which there are more air passengers than registered tourists. These are mostly distant country pairs with large contingents of migrants or returning nationals (who are not registered by tourism statistics) but relatively modest inflows of other visitors (e.g., India and Oman).

Table 9.1 Distribution of the difference between tourists and air passengers

Percentiles	2011	2012	2013	2014	2015	2016
Min	−89,300,000	−89,800,000	−89,400,000	−90,200,000	−92,400,000	−93,400,000
1%	−3,918,997	−4,064,395	−4,002,791	−4,361,469	−3,865,980	−4,136,718
5%	−514,371	−581,089	−661,828	−655,484	−569,920	−643,928
10%	−192,821	−212,287	−235,487	−232,265	−183,901	−218,354
25%	−22,009	−27,635	−30,651	−28,778	−24,436	−28,451
50%	−1997	−2536	−2924	−2493	−2189	−2323
75%	−63	−113	−126	−94	−56	−55
90%	1770	1220	998	1371	1480	4097
95%	11,824	10,775	8400	10,081	10,992	28,604
99%	131,253	140,405	109,720	113,494	140,005	257,340
Max	1,137,767	834,788	1,070,940	1,191,830	1,396,962	2,525,211
Obs.	5359	5771	5649	5653	5779	5262
Mean	−210,505	−219,209	−232,735	−232,250	−224,670	−243,573
Std. Dev.	2,175,910	2,132,248	2,178,926	2,221,919	2,251,131	2,393,686
Skewness	−30	−30	−28	−28	−29	−27
Kurtosis	1105	1131	1048	1020	1043	939

Note: Negative values indicate that there are more tourists than air passengers; positive values indicate that there are more air passengers than tourists travelling between a pair of countries
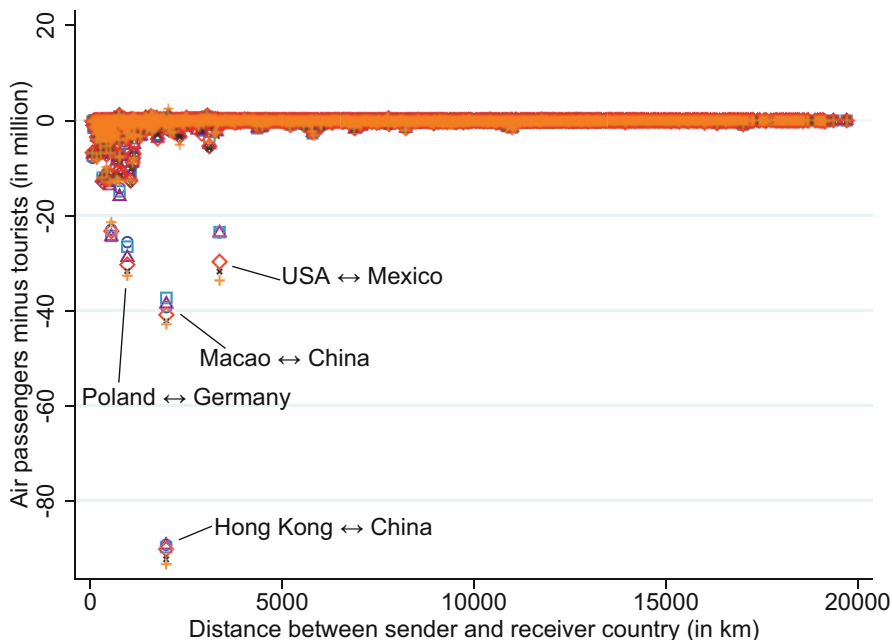
Fig. 9.2 The relation between geographic distance and divergences between the GMP-revised tourism dataset [1] and the KCMD-revised air passenger trend dataset [2]
Note: Different shapes denote different years. Distance is obtained from Mayer and Zignago (2006)

While rationale (b) is difficult to deal with but presumably marginal in statistical terms (see the remaining limitations described in Sect. 9.5), we treat rationale (a) by creating a correction factor that takes distance into account.

9.3.3 Creating the Distance-Adjusted Air Passenger Data [3]

The goal here is to adjust the KCMD-revised air passenger trend data [2] to correct for the fact that it underestimates mobility at short distances due to the use of alternative transportation means. To do so, we draw on the distance (in km) between country pairs. Our correction factor is specified as:

$$\left(\frac{k_{\max}}{k_{A \leftrightarrow B}}\right)^{1/c}$$

where $k_{\max}$ is the maximum possible distance between two countries, in this case 19,951.16 km (the distance between Paraguay and Taiwan), and $k_{A \leftrightarrow B}$ is the empirical distance between two countries A and B, based on CEPII's GeoDist dataset (Mayer & Zignago, 2006). The parameter c is chosen so that it maximizes the

correlation r between the GMP-revised tourism data [1] and the KCMD-revised air passenger trend data [2].[9] The rationale behind this is the assumption that [1] is not biased in terms of distance. Distance-adjusting [2] so that its correlation with [1] is maximized should thus lead to the best possible correction factor.

The result of this procedure is illustrated in Fig. 9.3a. After this adjustment, the correlation is $r(max) = 0.7282$. Higher and lower c's lead to lower correlations. Figure 9.3b illustrates how the size of the resulting correction factor (based on the c that maximized the correlation) decreases as geographic distance increases between countries. The relationship resembles a fat-tailed power-law curve that is almost universally found to describe the spatial structure of human and animal mobility well (see Deutschmann, 2016 for an overview). Figure 9.3c shows the empirical distribution of resulting correction factors. For most cases, the correction is relatively small (correction factor < 1.5).

Figure 9.4 shows, on a log-log plot, how the GMP-revised tourism data [1] and the distance-adjusted air passenger data [3] relate to each other for all cases in which data from both sources is available. It reveals that, despite the distance-adjustments, the tourism data is still larger in about 70% of cases (i.e., more data points are located below the diagonal [solid line]). The adjustment can thus be considered conservative overall. The correlation is strong and clear, in line with hypothesis (b) in Sect. 9.3.2.

9.3.4 Creating the Global Transnational Mobility Dataset

In the final step, we merge the two revised data sources. As we hold that both the GMP-revised tourism data [1] and the distance-adjusted air passenger data [3] individually tend to under-estimate actual mobility flows (see Sect. 9.2), our final estimate is always the largest of the two when we have both kinds of information—that is, either [1] or [3]. When we lack [3], we take [1]; and vice versa. As final steps, we:

– Round decimals (non-integer estimates can occur due to the time-series decomposition applied by Gabrielli et al., 2019 and the correction factor introduced above).
– Add missing full country names and information on the world region a country is located in, based on the United Nations classification (drawing on Duncalfe, 2018).
– Exclude countries for which, after the merging procedure, no information was available.[10] Consequently, the dataset is reduced to the set of 196 countries used when creating the unified UNWTO dataset.

---

[9]We combine data from all available years and exclude cases with more than ten million trips to reduce the influence of these outliers on the calculations. On average, 31 cases are ignored per year (0.08% of the total).

[10]Countries and territories excluded are: Aruba, Anguilla, Cocos Islands, Cook Islands, Christmas Islands, Western Sahara, Falkland Islands, Faroe Islands, Guadeloupe, Grenada, Greenland, French
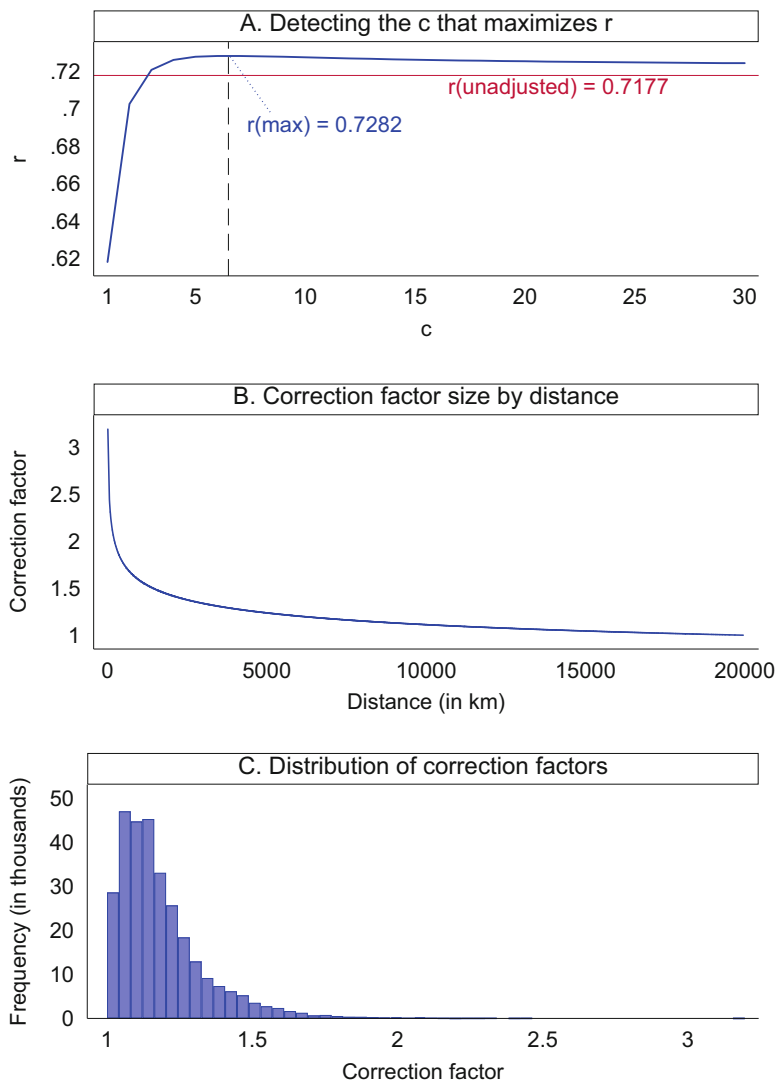
Fig. 9.3 Adjusting the distance-based correction factor for the KCMD-revised air passenger trend data to maximize the fit with the GMP-revised tourism data

The resulting Global Transnational Mobility Dataset can be explored on an interactive world map at the KCMD Dynamic Data Hub (https://bluehub.jrc.ec.europa.eu/migration/app/index.html; browse 'Datasets' – 'Mobility' – 'Global Transnational

---

Guiana, Montenegro, Northern Mariana Islands, Montserrat, Martinique, New Caledonia, Norfolk Islands, Pitcairn, Puerto Rico, French Polynesia, Reunion, Saint Helena, Saint Pierre and Michelon, Serbia, Tokelau, Taiwan, Wallis and Futuna Islands.
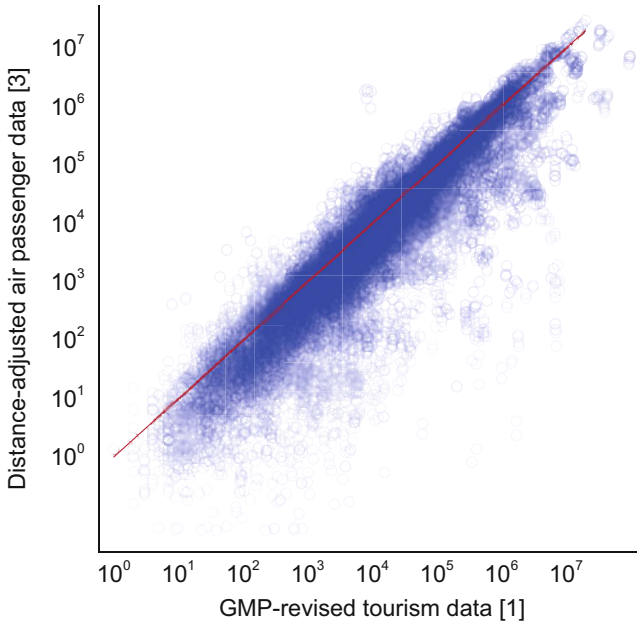
Fig. 9.4 The correlation between the distance-adjusted air passenger data [3] and the GMP-revised tourism data [1]

Mobility (KCMD-EUI)' – 'Estimated Trips'). More information can be found on the website of the Migration Policy Centre of the EUI (https://migrationpolicycentre.eu/projects/global-mobilities-project/), where the dataset can be downloaded. A list of the variables contained in the dataset can be found in the Appendix (Table 9.3).

9.4   Exploring the Dataset: Key Descriptive Findings

The Global Transnational Mobility Dataset covers 196 sender and receiving countries. Through the integration of two different sources, it is, to our knowledge, more comprehensive than all pre-existing information on worldwide cross-border mobility. Among other merits, its focus on transnational movements on a global scale helps to put migration in perspective, in both its geographical and demographic scope. The number of yearly migrant flows is very difficult to establish, and different alternative estimation methods have been proposed (Abel & Cohen, 2019; Abel & Sander, 2014; Azose & Raftery, 2019; Dennett, 2016). According to these methods, estimates range between 30 and 90 million migration episodes per year in the 2010–2015 period (Abel & Cohen, 2019, p. 8). Based on our new dataset, we estimate that, on average, about 2.55 billion yearly cross-border trips took place in the 2011–2015 period. Very crudely, thus, international migration episodes are

between 28 and 85 times (depending on migration estimates) less frequent than human movements across national borders in general. For specific regions, this ratio can be even higher. For example, in the European Union (for which actual yearly migration flow data is available), approximately 500–700 transnational trips occurred for every migratory move in 2016 (Deutschmann & Recchi, 2022).

While we leave to future research the full exploitation of the dataset's potential, also in conjunction with other datasets (not only on migration but also, for instance, on global trade, bilateral political relationships and many other potential predictors or predicted variables), the following pages offer a preliminary outline of several major takeaways.

9.4.1 Worldwide Transnational Mobility Is Rapidly Increasing Over Time

During the time frame under study, 2011 to 2016, transnational human mobility increased dramatically. In absolute terms, the number of estimated trips grew from about 2.3 billion in 2011 to about 2.9 billion in 2016. As Fig. 9.5a reveals, this growth is much larger than the growth in world population. This indicates that, collectively, humanity has indeed become more transnationally mobile. In this regard, transnational mobility is developing in line with cross-border communication, but in contrast to migration, which has not grown significantly faster than the world population (Czaika & De Haas, 2014; Deutschmann, 2021). This is also visible in Fig. 9.5b, which shows how, within the EU-28 (for which information on yearly migration flows is available), the number of transnational trips has grown much faster than both population and yearly migration flows (growth rates illustrated relative to the 2016 value). One important consequence of these diverging trends is that migration as a share of all transnational mobility is decreasing over time. In other words, temporary mobility has become more common relative to permanent migration.

The enormous increase in transnational mobility in a relatively short time frame raises questions on many grounds, like its environmental impact; its contribution to the spread of epidemics (Liu et al., 2020); its association with global systemic risks (Centeno et al., 2015); and, from a sociological perspective, social inequalities in access to these increased mobility opportunities. The latter issue is briefly touched upon in the following section.

9.4.2 Transnational Mobility Tends to Cluster Within World Regions

Figure 9.6a shows the mobility (in million trips) within world regions, using the United Nations M.49 Geoscheme as a base for assigning countries to regions. We
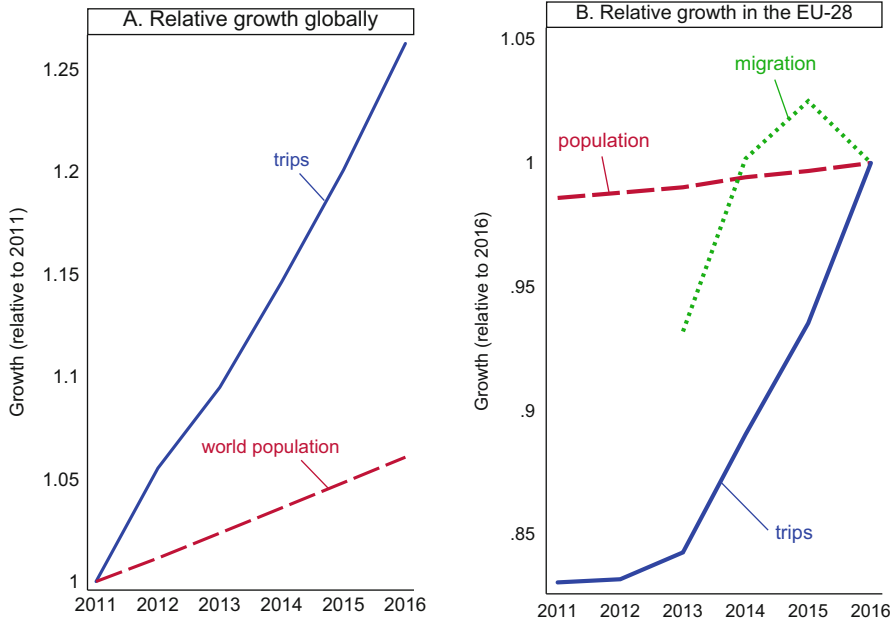
Fig. 9.5 Relative growth of mobility (and migration) globally and in the EU-28
Note: The graphs are based on the Global Transnational Mobility Dataset (trips), World Bank (2018) population data, and Eurostat migration data

find that Europe is the region with the highest number of intraregional trips, followed by Asia. The Americas are behind, and the smallest number of trips occur within Africa and Oceania.[11]

*Inter*regional mobility is far less common than intraregional mobility, with 80% of all mobility occurring within world regions in any given year (Deutschmann, 2020). However, there are differences between world regions in this regard (Fig. 9.6b): Intraregional mobility is more than five times more likely to occur than interregional mobility in the case of Europe; more than four times in the case of Asia; and almost three times in the case of the Americas. In the case of Africa, intraregional mobility is basically as likely as interregional mobility and in Oceania, intraregional mobility is half as likely as interregional mobility.

Note, however, that this comparison may be seen as 'unfair' since the pool of potential connections is obviously much larger in the case of interregional mobility

---

[11] Note that this simple measure may not be the best one to represent regional mobility. It is well possible that within Europe, for example, the high number of trips is driven by a subset of country pairs and that others participate very little in the intraregional network of transnational human mobility. Deutschmann (2021) proposes to use density-based measures as an alternative that allows to take into account between how many country pairs in a region meaningful amounts of mobility exist. Moreover, more sophisticated analyses would have to consider the varying population sizes of regions.
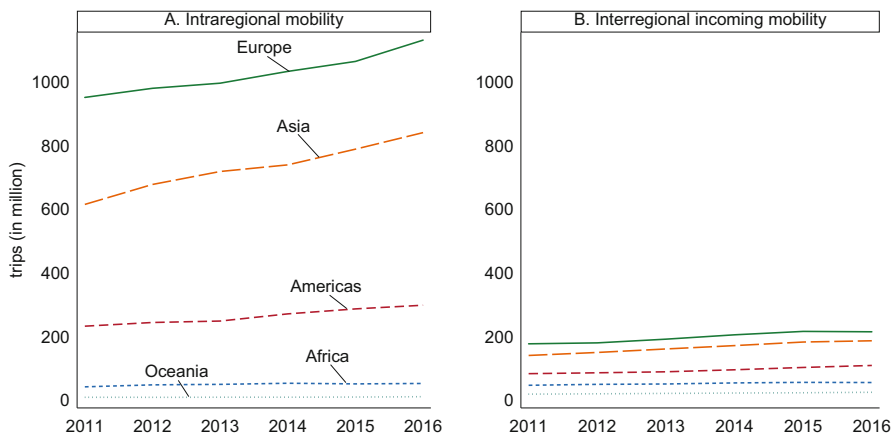
Fig. 9.6 Mobility within and between world regions

than in the case of intraregional mobility. A more sophisticated and 'just' comparison (which goes beyond the scope of this chapter) would be to compare intraregional mobility to mobility towards *specific* world regions. Past research has found that when this is done, mobility also tends to cluster within Africa and Oceania (Deutschmann, 2021).

In any case, Fig. 9.6a, b highlight the extreme stratification of opportunity to engage in transnational mobility at the global scale. Transnational mobility within Europe is about twenty times the amount of mobility within Africa, in spite of the much larger population of the latter continent. This global inequality in mobility chances has important sociological implications. For example, it has been shown that transnational human capital is an important resource that improves opportunities in life (Gerhards et al., 2017). Furthermore, transnational mobility shapes world views, attachment to other countries and cosmopolitan attitudes (Deutschmann et al., 2018; Helbling & Teney, 2015; Kuhn, 2015; Mau et al., 2008; Recchi, 2015). While these consequences of unequal access to transnational mobility chances have mainly been studied from a European viewpoint so far, a global perspective is largely missing. The Global Transnational Mobility Dataset may prove a good starting point for future analyses in this direction. The next section digs a little deeper into this global stratification by looking at the relationship between transnational human mobility and levels of prosperity.

9.4.3 Transnational Mobility Differs by Levels of Prosperity and Country Size

There is a relatively strong and significant relationship between a country's number of outgoing trips and the national level of prosperity, measured as GDP per capita in
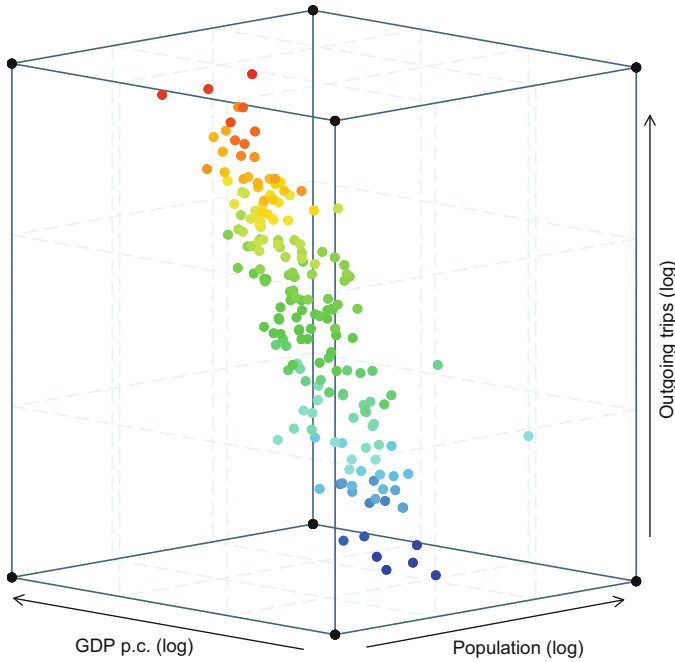
Fig. 9.7 The relation between mobility, population size, and GDP per capita

purchasing power parity based on World Bank data ($r = .63$). A similar pattern is found for the relationship between mobility and population size ($r = .58$). The three-dimensional graph in Fig. 9.7 illustrates the relationship between the three factors in combination. The distribution of dots, representing countries, follows a clear pattern, ranging from low GDP, small population and low mobility (bottom front corner) to high GDP, large population and high mobility (upper back corner). These insights are not entirely new but are showcased in a clear and robust way by this novel dataset. Future research may engage in more complex analyses, taking a larger set of factors into account and building more comprehensive multivariate models to study the antecedents and consequences of transnational human activity worldwide (see also Recchi et al., 2019b).

9.5   Discussion

A spate of migration and asylum-seeking crises has been hitting the world since the turn of the twenty-first century. The globe is on the move but, in spite of their salience in the media and public opinion, refugees and other migrants constitute only a tiny portion of the whole number of people crossing borders daily. According to various estimates, there were between 30 and 90 *million* migration episodes per year

in the early 2010s worldwide (Abel & Cohen, 2019). But according to our estimate, yearly border-crossings come close to 3 *billion* globally. By providing estimates of the amount of such transnational mobility beyond migration, the Global Transnational Mobility Dataset facilitates the study of the volume, directions and change of country-to-country human mobility on a worldwide scale.

This chapter described the procedures by which we have reached these estimates. While there is no single existing data source providing exact information on the number of people crossing national borders worldwide, we have argued that the two more complete and reliable sources (data on tourism and data on air passengers) show significant consistency and can be merged according to a few relatively simple combination rules.

We hope that our work will prove useful in two regards. First, we hope that the freely available GMP Global Transnational Mobility Dataset will be used to tackle questions related to mobility at the global scale. Potential applications are manifold and range from transnational mobility's unequal global structure and its social consequences to analyses that use the data to model the spread of infectious diseases. A first external study has already leveraged the dataset to model the spread of Covid-19 (Liu et al., 2020). Second, we hope that some aspects of our methodological approach can be transferred to other instances where researchers may consider merging two data sources, for instance, where regional migration data is available from several sources.

By focusing on yearly country-to-country flows of human *mobility* (whatever their duration), our dataset complements estimates of worldwide *migration* flows which refer to stays abroad longer than 12 months based on the conventional UN definition of migration. This dataset also improves upon previous usages of the UNWTO data (Deutschmann, 2016, 2021; Reyes, 2013) by capitalizing on an additional source and estimation methods. Finally, the Global Transnational Mobility Dataset parallels recent alternative attempts at measuring population mobility with digital sources (Fiorio et al., 2017; Hawelka et al., 2014; Messias et al., 2016; Rango & Vespe, 2017; Spyratos et al., 2018, 2019; State et al., 2013; Zagheni et al., 2017). Data triangulation across our data and digital estimates may prove useful to test the comparability of outcomes obtained through such different approaches.

Several important limitations remain. The first issue concerns the existence of grand-tour tourism and open-jaw flights. For instance, consider a traveler who goes on a round trip to Southeast Asia from Italy. She flies from Rome to Bangkok both on her way in and out and takes buses or rents a car to travel subsequently through Thailand, Vietnam, Laos, and Cambodia, before returning to Thailand to take her flight back home. According to the original UNWTO tourism data, there would be four trips: ITA → THA, ITA → VNM, ITA → LAO, and ITA → KHM. According to the GMP-revised tourism data [1], there would be eight trips: ITA → THA, THA → ITA, ITA → VNM, VNM → ITA, ITA → LAO, LAO → ITA, ITA → KHM, and KHM → ITA. According to the air passenger data (regardless of distance-adjustment), there would be two trips: ITA → THA, THA → ITA. In reality, however, there were six trips: ITA → THA, THA → KHM, KHM → VNM, VNM → LAO, LAO → THA, and THA → ITA. In this case, both sources and all

strategies lead to very different outcomes and none of them captures the transnational mobility that actually took place. This issue has no easy solution. Structurally, it should lead to a slight overestimation of long-distance mobility between world regions (which is most likely when such roundtrips are prone to occur). However, we argue that, compared to all global travels, these kind of journeys are rare and should not jeopardize the overall reliability of the dataset.

A second limitation consists of the following: by basing a substantial part of our mobility estimates on visitors who stayed overnight ('tourists' in the UNWTO terminology), we may be underestimating short-term border crossings. For instance by commuters who live in border regions and regularly go to the other side for work, leisure, or shopping. The following example is revealing in this regard: For the US, detailed data on land-border crossings are available (US Department of Transportation, 2018). Looking at mobility between the US and Canada, the distance-adjusted air passenger data (see Sect. 9.3.3) estimates about 20 million trips, while the GMP-revised tourism data (see Sect. 9.3.1) suggests around 33 million trips. The recorded land-border crossings, by contrast, are 103 million—98 million private car passengers alone. Many of these moves are not likely to be overnight stays. While it is hard to generalize from this example, it suggests that the mobility estimates in the Global Transnational Mobility Dataset (and the correction factor introduced in Sect. 9.3.3), although considerably larger than those provided by alternative global sources, are still quite conservative.

Finally, it is important to keep in mind that the Global Transnational Mobility Dataset contains mobility *estimates* rather than counts of actual, recorded trips. This is crucial. By applying a *statistical* approach to correct and adjust the data, we aimed to create a revised dataset that *on average* captures mobility between countries more accurately. In a minority of individual cases, this revision procedure might, however, lead to more inaccurate estimates. We would therefore like to remind that this dataset is well-suited to study structural features of transnational human mobility globally or for aggregates of countries. If the research interest is mobility between specific pairs of countries, the estimates in the Global Transnational Mobility Dataset should be taken with caution. Readers need to be aware of this limitation and should possibly compare the estimates to figures provided by alternative sources.

With these caveats in mind, we hope that this novel dataset will prove to be a valuable resource for all researchers interested in studying the human side of globalization. More particularly, this dataset can help embed migratory movements into the larger picture of transnational human mobility and better eschew the 'settlement bias' (Hugo, 2014) that recurrently weakens traditional migration studies. Attention to transnational mobility is especially needed to take into account less traditional and more reversible forms of migration (temporary, circular, shuttle, etc.). Also, it is needed more generally to remind us that international migrants are first of all people who cross borders – and therefore part and parcel of a mobile world.

Finally, there are several general lessons we want to share with readers who might be interested in following a similar merging strategy based on different datasets:

1. *Automatize!* The more the combination procedure is automatized, the easier it is to update datasets in the future. In our case, both the UNWTO and Sabre continuously update their datasets and it would be desirable to be able to quickly expand the time frame of the Global Transnational Mobility Dataset as these updates become available. Despite the availability of monthly air passenger volume projections for 2020 (Iacus et al., 2020), we were unfortunately not able to automatize the whole process (partly due to point [2] below). However, scholars interested in conducting a similar project with other sources are advised to automatize as much as possible to increase efficiency and facilitate future updates.

2. *Standardize!* One of the most time-consuming issues in combining mobility data from different sources is to bring the datasets into a mergeable format. One common obstacle occurs when only idiosyncratic country names rather than standardized country codes are available. For example, rather than using the standardized code COD, sources often only contain idiosyncratic names such as "Congo, Democratic Republic of", "D.R. Congo", "DR Congo", or "Congo, DR". We therefore appeal to all data-collecting organizations and individuals who publish such data to use standardized formats such as ISO 3166-1 alpha-3 country codes. Doing so will increase the potential for automatization and thereby increase efficiency. In our case, the UNWTO data files contain *numeric* ISO codes (e.g., "180" for D.R. Congo) while the air traffic data used the three-letter version (e.g., "COD"). While the conversion between the two is of course possible, it still constitutes one additional step that could be avoided by consistent usage of the more intuitive letter version. A further obstacle are changes in ISO country codes, such as the switch from ROM to ROU in the case of Romania due to a 2002 administrative decision. Such little changes often prevent flawless merging and lead to costly manual inspections. While we appreciate all existing standardization efforts, we believe there is still room for improvement.

3. *Annotate!* Good documentation is important, and we recommend annotating every single step in the procedure as clearly as possible and early on to increase inter-individual transparency.

4. *Keep it simple!* Several technically more sophisticated methods (e.g., multiple imputation) turned out not to lead to any useful information. We therefore developed the more straightforward correlation-maximizing approach presented above and the simple set of rules for combining the two sources. Often, in our view, it makes sense to stick to the classic KISS principle – Keep it simple, stupid!

5. *Globalize!* It is generally a good idea to start with the most comprehensive coverage and only drop data later on. It is always possible to make the dataset smaller but a lot more difficult to make it larger again.

6. *Be cautious!* In our view, merging data from different sources can bring advantages – and we see a lot of benefits of the GMP Global Transnational Mobility Dataset – but it may also carry risks, as already emphasized above. After all, different datasets are usually collected with different purposes in mind and are often based on different definitions and collection procedures. It is therefore

important to keep the resulting limitations in mind and reflect carefully on whether or not a certain combined dataset is well-suited for a specific research goal.

We hope that these recommendations can help other researchers who plan to embark on similar endeavors.

Appendix

Table 9.2 Categories in the UNWTO dataset

Code	Description	Preference
112	Arrivals of non-resident tourists at national borders, by country of residence	1st
111	Arrivals of non-resident tourists at national borders, by nationality	2nd
122	Arrivals of non-resident visitors at national borders, by country of residence	3rd
121	Arrivals of non-resident visitors at national borders, by nationality	4th
1912	Arrivals of non-resident tourists in all types of accommodation establishments, by country of residence	5th
1911	Arrivals of non-resident tourists in all types of accommodation establishments, by nationality	6th
712	Arrivals of non-resident tourists in hotels and similar establishments, by country of residence	7th
711	Arrivals of non-resident tourists in hotels and similar establishments, by nationality	8th

Table 9.3 Variables contained in the Global Transnational Mobility Dataset

Name	Description
source_name	Name of the country of origin
target_name	Name of the country of destination
source_iso3	ISO 3166-1 alpha-3 code of the country of origin
target_iso3	ISO 3166-1 alpha-3 code of the country of destination
year	Year, ranges from 2011 to 2016
estimated_trips	Estimated trips
dist	Geographic distance
source_region	Region of the country of origin
target_region	Region of the country of destination
source_subregion	Sub-region of the country of origin
target_subregion	Sub-region of the country of destination

Note: Geographic distance is obtained from CEPII's GeoDist dataset (Mayer & Zignago, 2006). Regions and subregions are based on the UN M.49 GeoScheme

References

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science, 343*(6178), 1520–1522.

Abel, G. J., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data, 6*(1), 1–13.

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, 116*(1), 116–122.

Centeno, M. A., Nag, M., Patterson, T. S., Shaver, A., & Windawi, A. J. (2015). The emergence of global systemic risk. *Annual Review of Sociology, 41*, 65–85.

Czaika, M., & De Haas, H. (2014). The globalization of migration: Has the world become more migratory? *International Migration Review, 48*(2), 283–323.

Dennett, A. (2016). *Estimating an annual time series of global migration flows–an alternative methodology for using migrant stock data* (pp. 125–142). Global Dynamics: Approaches from Complexity Science.

Deutschmann, E. (2016). The spatial structure of transnational human activity. *Social Science Research, 59*, 120–136.

Deutschmann, E. (2020). Visualizing the regionalized structure of mobility between countries worldwide. *Socius, 6*, 1–3.

Deutschmann, E. (2021). *Mapping the transnational world: How we move and communicate across borders, and why it matters*. Princeton University Press.

Deutschmann, E., Delhey, J., Verbalyte, M., & Aplowski, A. (2018). The power of contact: Europe as a network of transnational attachment. *European Journal of Political Research, 57*(4), 963–988.

Deutschmann, E., & Recchi, E. (2022). Europeanization via transnational mobility and migration. In M. Eigmüller, S. Büttner, & S. Worschech (Eds.), *Sociology of Europeanization*. De Gruyter Oldenbourg, 283–306.

Duncalfe, L. (2018). *ISO-3166 country and dependent territories lists with UN regional codes*. Available at: https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes. Last accessed 08 Jan 2019.

Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration. *Proceedings of the 2017 ACM on Web Science Conference*, 103–110.

Gabrielli, L., Deutschmann, E., Natale, F., Recchi, E., & Vespe, M. (2019). Dissecting global air traffic data to discern different types and trends of transnational human mobility. *EPJ Data Science, 8*. https://doi.org/10.1140/epjds/s13688-019-0204-x

Gerhards, J., Hans, S., & Carlson, S. (2017). *Social class and transnational human capital. How middle and upper class parents prepare their children for globalization*. Routledge.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science, 41*(3), 260–271.

Helbling, M., & Teney, C. (2015). The cosmopolitan elite in Germany. Transnationalism and postmaterialism. *Global Networks, 15*(4), 446–468.

Hugo, G. (2014). A multi sited approach to analysis of destination immigration data: An Asian example. *International Migration Review, 48*(4), 998–1027.

Iacus, S. M., Natale, F., Santamaria, C., Spyratos, S., & Vespe, M. (2020). Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. *Safety Science, 104791*.

Kuhn, T. (2015). *Experiencing European integration: Transnational lives and European identity*. Oxford University Press.

Liu, Q., Liu, Z., Zhu, J., Zhu, Y., Li, D., Gao, Z., ... Wang, Q. (2020). Assessing the global tendency of COVID-19 outbreak. *MedRxiv*. https://doi.org/10.1101/2020.03.18.20038224

Mau, S., Mewes, J., & Zimmermann, A. (2008). Cosmopolitan attitudes through transnational social practices? *Global Networks, 8*(1), 1–24.

Mayer, T., & Zignago, S. (2006). *GeoDist: The CEPII's distances and geo-graphical database* (MPRA paper no. 31243).

Messias, J., Benevenuto, F., Weber, I., & Zagheni, E. (2016). From migration corridors to clusters: The value of Google+ data for migration studies. In *Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 421–428).

Nye, J. S., & Keohane, R. O. (1971). Transnational relations and world politics: An introduction. *International Organization, 25*(3), 329–349.

Rango, M., & Vespe, M. (2017). *Big data and alternative data sources on migration: From case-studies to policy support. Summary report*. Joint Research Centre of the European Commission.

Recchi, E. (2015). *Mobile Europe: The theory and practice of free movement in the EU*. Palgrave Macmillan.

Recchi, E. 2017. *Towards a global mobilities database: Rationale and challenges.* Explanatory Note. MPC/EUI. Available at: http://www.migrationpolicycentre.eu/docs/GMP/Global_Mobilities_Project_Explanatory_Note.pdf. Last accessed 3 Mar 2019.

Recchi, E., Deutschmann, E., & Vespe, M. (2019a). *Estimating transnational human mobility on a global scale* (Robert Schuman Centre for Advanced Studies / Migration Policy Centre WP 2019/30). European University Institute.

Recchi, E., Deutschmann, E., & Chabriel, M. (2019b). *The global network of transnational mobility*. N-IUSSP, October. http://www.niussp.org/article/the-global-network-of-transnational-mobilityle-reseau-mondial-de-mobilite-transnationale/

Reyes, V. (2013). The structure of globalized travel: A relational country-pair analysis. *International Journal of Comparative Sociology, 54*(2), 144–170.

Sabre. (2020). *Market intelligence global demand data*. http://www.sabreairlinesolutions.com/home/software_solutions/airports/. Last accessed 28 Jun 2020.

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2018). *Migration data using social*. A European Perspective. JRC Technical Report. https://doi.org/10.2760/964282

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLoS One, 14*(10). https://doi.org/10.1371/journal.pone.0224134

State, B., Weber, I., & Zagheni, E. (2013). Studying inter-national mobility through IP geolocation. In *Proceedings of the sixth ACM international conference on web search and data mining* (pp. 265–274).

United Nations World Tourism Organization (UNWTO). (2015). *Methodological notes to the tourism statistics database*. UNWTO.

U.S. Department of Transportation. (2018). *Border crossing entry data*. Available at: https://data.transportation.gov/Research-and-Statistics/Border-Crossing-Entry-Data/keg4-3bc2. Last accessed 9 Jan 2019.

Wimmer, A., & Glick Schiller, N. (2002). Methodological nationalism and beyond: Nation-state building, migration and the social sciences. *Global Networks, 2*(4), 301–334.

World Bank. (2018). *Population, total*. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL. Last accessed 9 Jan 2019.

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review, 43*(4), 721–734.