



**HAL**  
open science

## Dissecting global air traffic data to discern different types and trends of transnational human mobility

Ettore Recchi, Emanuel Deutschmann, Natale Fabrizio, Michele Vespe

### ► To cite this version:

Ettore Recchi, Emanuel Deutschmann, Natale Fabrizio, Michele Vespe. Dissecting global air traffic data to discern different types and trends of transnational human mobility. *EPJ Data Science*, 2019, 8 (26), 10.1140/epjds/s13688-019-0204-x . hal-03567054

**HAL Id: hal-03567054**

**<https://sciencespo.hal.science/hal-03567054>**

Submitted on 11 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Dissecting global air traffic data to discern different types and trends of transnational human mobility

Lorenzo Gabrielli<sup>1,5</sup>, Emanuel Deutschmann<sup>2,3\*</sup> , Fabrizio Natale<sup>1</sup>, Ettore Recchi<sup>3,4</sup> and Michele Vespe<sup>1</sup>

\*Correspondence:  
[emanuel.deutschmann@uni-goettingen.de](mailto:emanuel.deutschmann@uni-goettingen.de)

<sup>2</sup>University of Göttingen, Göttingen, Germany

<sup>3</sup>Migration Policy Centre, European University Institute, Florence, Italy  
Full list of author information is available at the end of the article

## Abstract

Human mobility across national borders is a key phenomenon of our time. At the global scale, however, we still know relatively little about the structure and nature of such transnational movements. This study uses a large dataset on monthly air passenger traffic between 239 countries worldwide from 2010 to 2018 to gain new insights into (a) mobility trends over time and (b) types of mobility. A time series decomposition is used to extract a trend and a seasonal component. The trend component permits—at a higher level of granularity than previous sources—to examine the development of mobility between countries and to test how it is affected by policy and infrastructural changes, economic developments, and violent conflict. The seasonal component allows, by measuring the lag between initial and return motion, to discern different types of mobility, from tourism to seasonal work migration. Moreover, the exact shape of seasonal mobility patterns is extracted, allowing to identify regular mobility peaks and nadirs throughout the year. The result is a unique classification of trends and types of mobility for a global set of country pairs. A range of implications and possible applications are discussed.

**Keywords:** Mobility; Migration; Air traffic; Migrant seasonal workers; Tourism

## 1 Introduction

Human mobility across national borders is one of the key phenomena of our time. Yet, at the global scale, our knowledge about the structure and nature of such transnational<sup>a</sup> movements is still limited. Conventional studies often rely on administrative data and surveys gathered by international organizations such as the United Nations, the World Bank, or the World Tourism Organization. However, these sources tend to focus on specific types of mobility such as migration, tourism, or asylum-seeking and have some limitations regarding coverage and accuracy. For the study of many issues related to planet-scale human mobility, from the spread of epidemics [1] and microbial mass movement [2] to climate change [3] and the emergence of global systemic risks [4], more fine-grained information on cross-border mobility in all its facets is needed [5].

A new strand of research has drawn on mobile-phone data to detect patterns of human mobility [6–11]. However, such studies are usually restricted to the local or, at best, national scale. A global dataset based on such data is hard to obtain due to the limited

coverage of single mobile-phone service providers and the confined dissemination of specific smartphone-based navigation apps among the global population. Several studies have used email traffic [12] or social networks such as Facebook [13, 14], Twitter [15, 16], Google+ [17], or LinkedIn [18] to make inferences about global mobility. While these approaches are certainly valuable and have led to important new insights, the reliability of such digital trace data to obtain a comprehensive picture of global mobility can be questioned in a world where smartphone user penetration will only reach 38.5 percent in 2020 [19], half the world population is still offline today [20], and social media usage rates vary across countries depending on age and sex [14]. It is thus desirable to gather information from further sources to see whether different approaches lead to similar estimates.

Here, we propose to use air traffic data as a valuable alternative resource to make inferences about trends in and types of transnational human mobility. We use a large dyadic dataset that covers monthly air passenger traffic between 239 countries and territories worldwide from January 2010 to March 2018. This dataset is quite unique in both its temporal granularity, spatial scope and completeness. It provides estimates of the number of passengers on commercial flights operated globally and covers more than four million ties of transnational human mobility. While we acknowledge that this is still a partial take on global cross-border mobility, as people employ other transportation means (trains, buses, private vehicles, ships, ferries, or even their own feet) to move from one country to another, recent research comparing this source with world tourism data indicates that 44 percent of registered cross-border travels occur through commercial flights, and that this proportion increases at rising distances between countries [21]. Differently from studies relying on digital trace data, which are unable to estimate which part of the universe of journeys they cover, we know that we include almost half of all international travels worldwide, and the whole of air-based travels. Moreover, the significant decline of flying costs relative to other transportation means (particularly, ships, trains and private vehicles) since the late 20th century [22, 23] entails an ongoing democratization of access to air travel.

While earlier studies have already drawn on such air passenger data [24, 25], we argue that past research has not made full use of what is perhaps the most specific and valuable feature of this data: its fine-grained chronological resolution, *i.e.*, the fact that it is available monthly. Our study exploits this unique asset and presents two major innovations that promise to advance the field. This is possible through a time-series decomposition that is used to extract a trend and a seasonal component. The trend component permits—in more temporal detail than previous sources—to examine the development of mobility between countries and to test how it is affected by policy and infrastructural changes, economic crises, violent conflicts, and similar events and processes. The second innovation is possible through the seasonal component, which allows, by measuring the lag between initial and return motion, to discern different types of mobility, from tourism to seasonal work migration, that dominate between specific country pairs. The exact shape of seasonal mobility patterns permits to identify, for each pair of countries, regular mobility peaks and nadirs throughout the year. The result is a unique classification of trends and types of mobility for a global set of country pairs that is publicly accessible via the KCMD Dynamic Data Hub, which we invite interested readers to explore: <https://bluehub.jrc.ec.europa.eu/migration/app/?state=5d6005b30045242cabd750a2>.

This article is structured as follows. We first further describe the data and methods used. Next, we present the results, focusing first on the analysis of the trend component and then

on the analysis of the seasonal component. We close with a summary and discussion of the implications and limitations of the study.

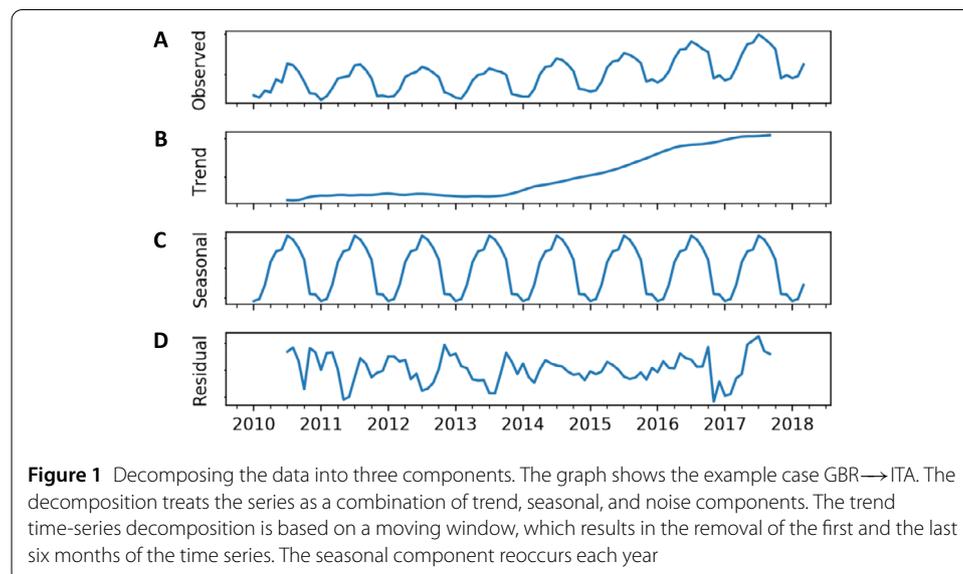
## 2 Data and methods

The analysis is based on an air passenger dataset provided by Sabre, a private company that collects data directly from the airline industry [26]. The dataset contains information on air traffic, in terms of the number of passengers per month, traveling from one country to another.<sup>b</sup> It covers worldwide air traffic from January 2010 to March 2018 between 239 countries and territories, resulting in a network of 45,444 country pairs. Overall, we are thus able to analyze the size and structure of more than four million ties of transnational human mobility.

The dataset covers historical origin-destination traffic as a measure of the total number of passengers flying between any two points, regardless of whether the flights are direct or indirect. Sabre produces the data using proprietary algorithms and does not report potential biases or incompleteness of the data. The commercially available dataset was accessed by the European Commission's Joint Research Centre, which pre-processed the data and ran the time-series decomposition as described in the next section. While the raw data is not freely available, a set of derived indicators is accessible online (see Availability of Data and Materials section at the end of this article). Furthermore, there is an open-access dataset of annual country-to-country travel estimates that are derived from the air passenger trend data we describe below, merged with tourism data and adjusted for distance [21]. This data can be openly accessed for further analysis and research.

### 2.1 Data preparation and time-series decomposition

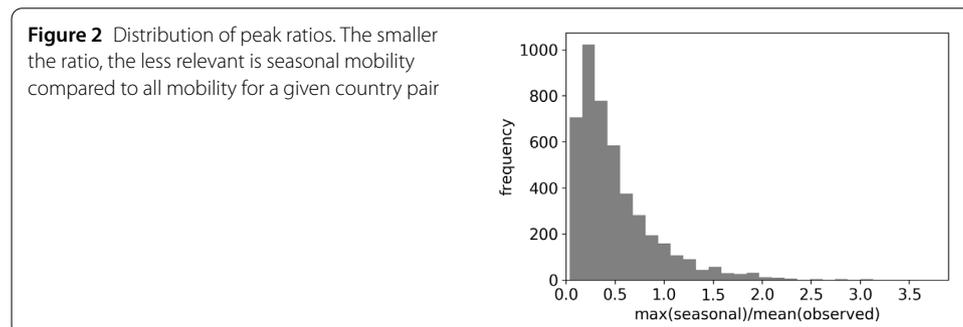
The main innovations of this study are possible through a time-series decomposition that allows to dissect the raw overall air passenger flow between two countries into a trend component and a seasonal component. This technique, implemented via the Python library *statsmodels* [27], allows to separate and extract clean signals from the raw data. Figure 1 illustrates the outcome of this method in the example case of air traffic from



Great Britain to Italy (GBR→ITA). Figure 1(A) shows the original time series, which is characterized by both an increase in mobility over time and a certain regularity that re-occurs every year.<sup>c</sup> The decomposition of the time series allows to separate and extract these trend (Fig. 1(B)) and seasonal (Fig. 1(C)) components, together with a residual component (Fig. 1(D)). Due to the nature of the technique, which uses a moving window, the first and the last six months of the trend time series are removed. The seasonal component consists of 12 data points, one for each month, representing the average across years. The residuals are white noise (*i.e.*, their distribution is Gaussian), which indicates that the decomposition correctly approximates the actual time series.<sup>d</sup> Throughout this study, we will ignore the residuals and focus on the trend and seasonal components to gain new insights into over-time developments and types of mobility.

With respect to the seasonal component, it has to be considered that between most country pairs globally no meaningful seasonal patterns exist, since the number of people moving between these countries is not high enough. Therefore, there is a need to define thresholds from which seasonal components are actually considered to be meaningful due to their size, both in absolute and relative terms. With respect to the *absolute* size, we treat a seasonal component as meaningful if the highest peak (*i.e.*, the month in the year during which most people move regularly) is  $\geq 500$  passengers. This threshold corresponds to the 90<sup>th</sup> percentile of the rank-ordered distribution of peaks. While this limit is of course arbitrary by nature, manual inspection of cases just above the 500-passenger threshold shows that such seasonal patterns are meaningful (see *e.g.*, the cases of BOL→PER and ARG→THA in Fig. 9 and their discussion below). For cases below the 500-passenger threshold, we argue that they are too susceptible to fluctuations to be considered meaningful and stable seasonal patterns. With regard to the *relative* size of the seasonal effect, we compute the ratio between the number of people moving during the peak month and the mean overall number of passengers per year (henceforth *peak ratio*), to get an understanding of the relevance of the seasonal component.

Figure 2 shows the resulting distribution of peak ratios. We exclude a small number of odd outliers with a peak ratio  $\geq 3.9$ .<sup>e</sup> The remaining cases in the distribution are meaningful, with larger values indicating that the seasonal effect is dominant compared to the overall amount of mobility occurring between two countries, while smaller values suggest that the seasonal component constitutes only a small share of the total air passenger flow. After applying the two thresholds (*i.e.*, a peak  $\geq 500$  passengers and a peak ratio  $\leq 3.9$ ), around 4500 country pairs (out of 45,444) remain in the analysis of the seasonal component.



## 2.2 Methods used to analyze the trend component

In the analysis of the trend component, we aim to address the following questions:

1. Does the mobility between countries increase or decrease over time?
2. How strong is this change typically?
3. Can we detect how sociopolitical and economic events and processes affect mobility flows?
4. How is air passenger traffic related to migration flows?

To answer question (1), we describe and classify the trend using the Spearman's rank correlation  $r_s$  between the trend (*i.e.*, the number of air passengers in a given year) and time (*i.e.*, the year). This allows to see if the trend is monotonously increasing (values close to  $r_s = 1$  with  $p < 0.05$ ), decreasing (values close to  $r_s = -1$  with  $p < 0.05$ ), or rather stable or fluctuating (with  $p \geq 0.05$ ) across the years under study [28].

To answer question (2), we compute the difference between the number of passengers travelling between  $A$  and  $B$  in the first and the last year under study (*i.e.*, the 2016 value minus the 2011 value), divided by the population size of the country of origin (to make numbers comparable across countries of different sizes). Values greater than 0 denote an increase; values smaller than 0 denote a decrease, and values = 0 denote no change between the two years.

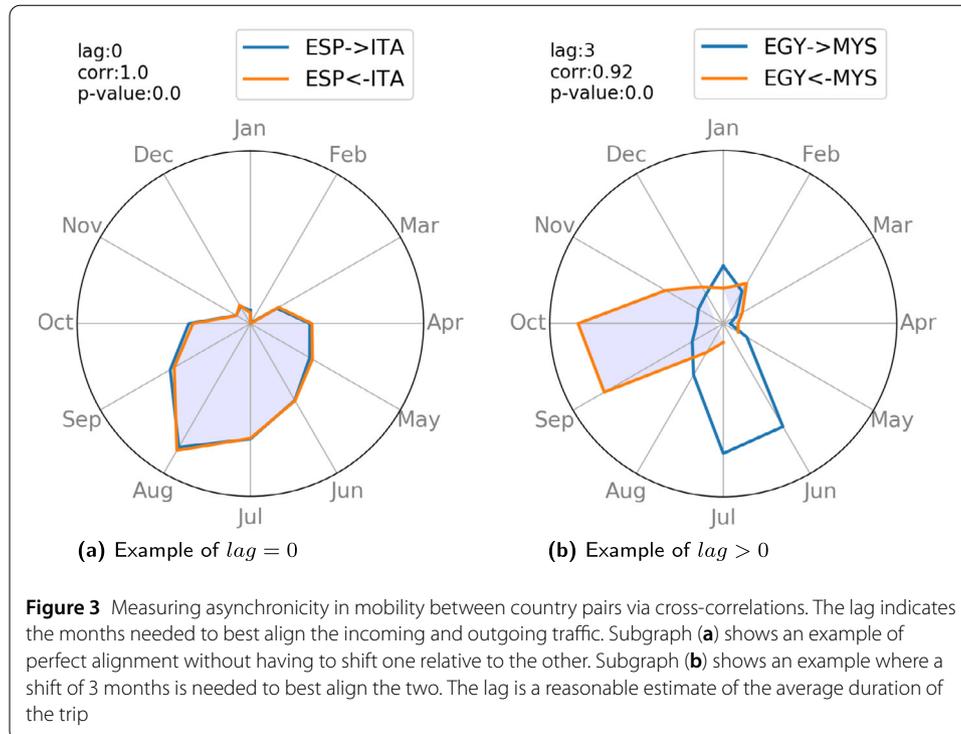
To answer question (3), we draw on a more qualitative approach, using external information to look at several example cases to study how policy and infrastructural changes, economic crises, and conflicts affect mobility between countries. While this analysis is not exhaustive, we hope it will provide valuable starting points for further research in this direction.

To answer question (4), we make use of the bi-directional nature of the data. In specific, we compute the difference in the cumulative sum of passengers over the years for both directions ( $A \rightarrow B$  minus  $A \leftarrow B$ ). We test whether a divergence between the two might indicate that some of the passengers stay in the surplus country to become migrants. If successful, this step may contribute to better estimating the size of "shadow populations" [29], *i.e.* the amount of illegal, undocumented migration, which is by definition hard to measure. In addition, we test whether there is a correlation between the trend in air passenger traffic and the trend in migrant flows between countries across the years under study. To make these comparisons, we draw on OECD migration flow data [30].

## 2.3 Methods used to analyze the seasonal component

To analyze structural similarities and chronological divergences between incoming and outgoing seasonal traffic between countries in a systematic way, we draw on cross-correlations. This method allows to test the similarity between two series, taking the displacement of one relative to the other into account. Cross-correlations are applied in many scientific contexts, from signal processing in echolocating mammals [31] and radar systems [32] to electron tomography [33] and biometric fingerprint matching [34]. Here, we use this method to measure how many months are needed to shift one seasonal time series in such a way that the incoming and outgoing traffic between a pair of countries are best aligned.

Figure 3 illustrates how this works. Subgraph (a) shows an example case (ESP  $\longleftrightarrow$  ITA) where both time series (*i.e.*, mobility from Spain to Italy and from Italy to Spain) are perfectly aligned (corr = 1.0,  $p < 0.05$ ) without having to shift one time series by one or



more months relative to the other ( $lag = 0$ ). By contrast, Fig. 3(b) shows an example case ( $EGY \longleftrightarrow MYS$ ), where the two time series (*i.e.*, mobility from Egypt to Malaysia and from Malaysia to Egypt) best align only after one of the two series is shifted (or “turned” to use a term that better fits the round radar-plot illustration) relative to the other by 3 months. We argue that the lag that corresponds to the highest cross-correlation for a given country pair provides an indication of the average duration of the trip, with higher lags suggesting longer stays abroad. In the example, trips between Egypt and Malaysia would thus typically last longer than trips between Spain and Italy. Furthermore, we argue that this lag allows to classify seasonal components into different patterns that correspond to different types of mobility. In specific, we will demonstrate that a lag of 0 months indicates mass tourism, a lag of 1 month suggests individual tourism over longer distances, and a longer lag hints to seasonal work migration dominating the seasonal mobility between the two countries.

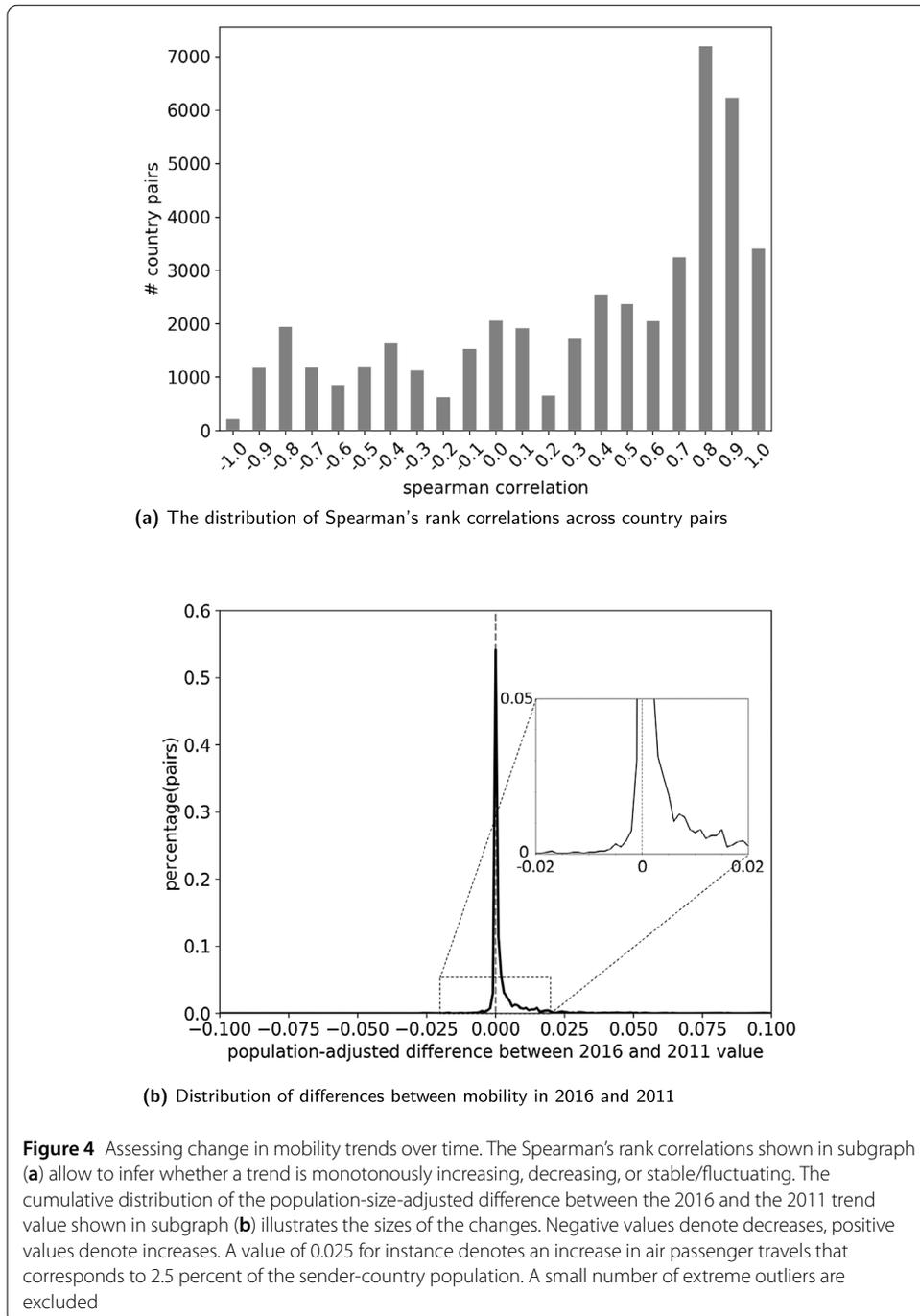
### 3 Results

In the following, we subsequently discuss the analysis of the trend component and the seasonal component.

#### 3.1 Trend analysis

##### 3.1.1 Identifying ascending and descending trends

Did transnational human mobility typically increase or decrease in the time frame under study? To answer this question, Fig. 4(a) shows the distribution of Spearman’s rank correlations, from  $r_s = -1$  (perfect monotonous decrease) to  $r_s = 1$  (perfect monotonous increase) in steps of 0.1. They reveal that significant monotonous increases (with  $r_s$  in between 0.8 and 1 and  $p < 0.05$ ) occur far more frequently than monotonous decreases (with  $r_s$  in between  $-1$  and  $-0.8$  and  $p < 0.05$ ). The cases between (with  $r_s$  in between  $-0.7$  and



0.7) hint to stable or fluctuating trends. In specific, around 13,000 country pairs experienced a significant ( $p < 0.05$ ) monotonous increase but only around 2300 country pairs saw a significant monotonous decrease. Thus, monotonous increases occurred 5.7 times more often than monotonous decreases during the observed time frame.<sup>f</sup>

To get an idea of the typical *size* of the change, Fig. 4(b) shows the distribution of the population-size-adjusted difference between the number of passengers travelling between *A* and *B* in the first and the last year under study (*i.e.*, the 2016 value minus the 2011 value). There are around 30,000 country pairs that experienced more air passenger mo-

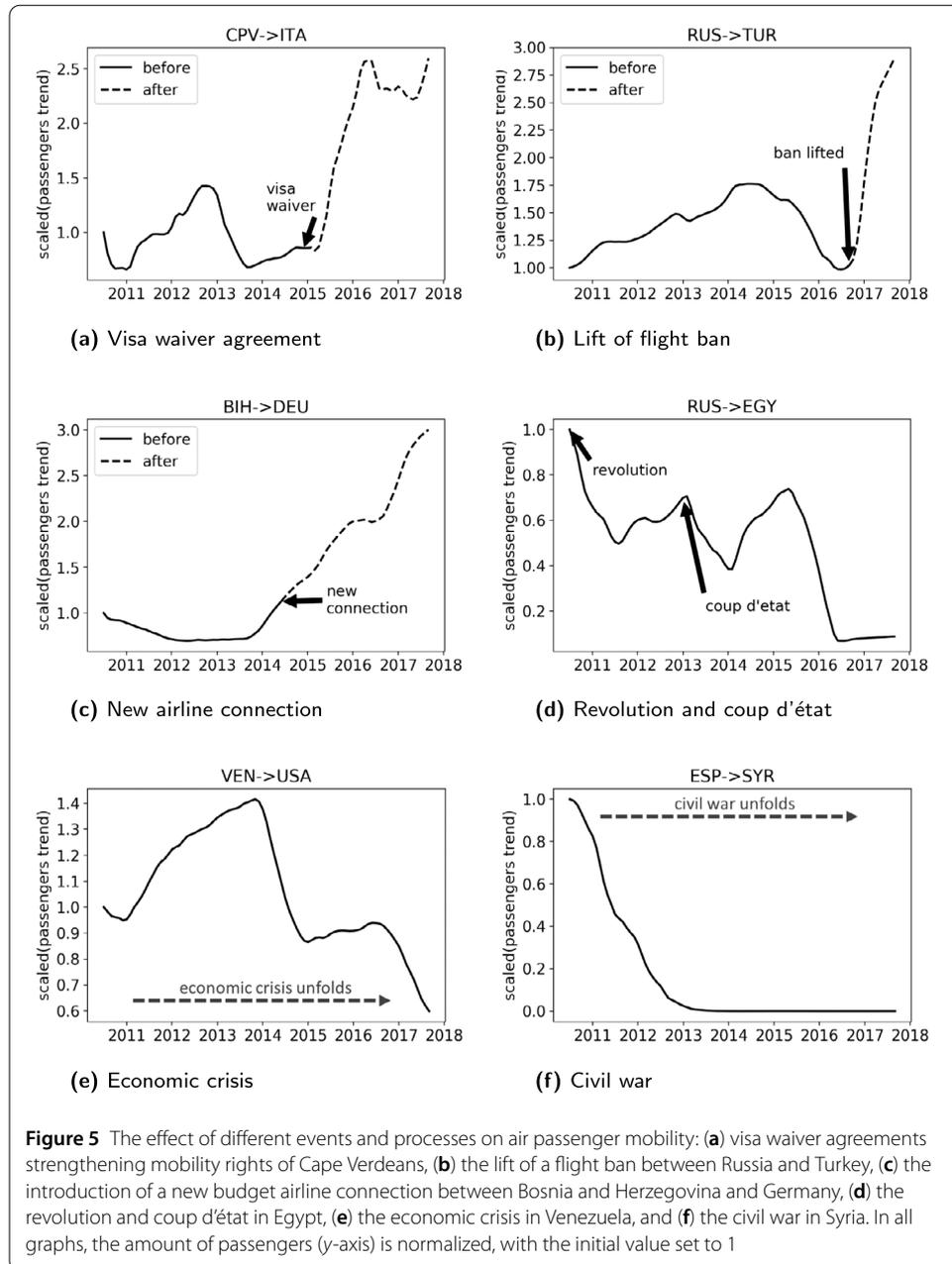
bility in 2016 than in 2011 (difference larger than 0), while around 12,000 country pairs experienced less (difference smaller than 0), and around 3000 country pairs had the same number of passengers.

As the figure reveals, the distribution is highly unequal. For the vast majority of country pairs, changes in the amount of passengers were rather minor (as can be seen from the strong clustering of cases around 0 in Fig. 4(b)). However, a small number of country pairs experienced very large increases over time, and an even smaller number of country pairs saw considerable decreases in mobility. For example, among the strong increases, there are 269 observations with a value of  $\sim 0.01$ , which denotes an increase in air passenger travels that corresponds to around 1 percent of the sender-country population. Moving even further towards the top end, there are also five observations with a value of  $\sim 0.1$ , or around 10 percent of the sender-country population. Moreover, the range in the graph is artificially set to  $[-0.1, 0.1]$  to be able to better visualize that the distribution is skewed to the right (which makes sense, since increases dominate). This means that a small number of extreme outliers are excluded from the graph. On the negative end, 12 cases with values smaller than  $-0.1$  are omitted, while at the top end, 35 cases with values greater than  $0.1$  are not shown. The largest decrease corresponds to 23 percent of the sender-country population (U.S. Virgin Islands  $\rightarrow$  United States), while the largest increase corresponds to 4 times the sender-country population (Turks and Caicos Islands  $\rightarrow$  United States).

Table 1 shows the 10 country pairs with the highest decreases and those with the highest increases in mobility in absolute terms (in million passengers) between 2011 and 2016. It reveals that many of the strongest decreases involve passenger traffic to and from Syria, a clear result of the civil war (cf. Fig. 5(f) and its discussion below). The strongest decrease in absolute numbers, however, occurred between Great Britain and Egypt, possibly a result of increased perceived threats of terrorism in Egypt that deter British (and Russian, see ranks 7 and 8) tourists. The UK Government’s Foreign Travel Advice for Egypt, for example, states that “Terrorists are very likely to try to carry out attacks in Egypt” [35]. A more detailed analysis of the decrease of mobility between Russia and Egypt follows below (cf. Fig. 5(d)). Notably, most of the highest increases of cross-border mobility between 2011 and 2016 occurred in Asia, specifically between China and Thailand, China and Korea, and Taiwan and Japan. Mobility between Mexico and the U.S. also increased starkly, as did travel between Great Britain and Spain.

**Table 1** The highest decreases and increases in mobility between 2011 and 2016 (in million passengers)

| Rank | Highest decrease |             |            | Highest increase |             |            |
|------|------------------|-------------|------------|------------------|-------------|------------|
|      | Origin           | Destination | Difference | Origin           | Destination | Difference |
| 1    | EGY              | GBR         | -0.40      | CHN              | THA         | 5.6        |
| 2    | GBR              | EGY         | -0.40      | THA              | CHN         | 5.6        |
| 3    | SYR              | ARE         | -0.36      | GBR              | ESP         | 5.3        |
| 4    | ARE              | SYR         | -0.35      | ESP              | GBR         | 5.2        |
| 5    | SAU              | SYR         | -0.35      | MEX              | USA         | 4.6        |
| 6    | SYR              | SAU         | -0.33      | KOR              | CHN         | 4.5        |
| 7    | RUS              | EGY         | -0.29      | USA              | MEX         | 4.2        |
| 8    | EGY              | RUS         | -0.29      | CHN              | KOR         | 4.2        |
| 9    | ARE              | CIV         | -0.22      | TWN              | JPN         | 3.6        |
| 10   | YEM              | SAU         | -0.19      | JPN              | TWN         | 3.6        |



### 3.1.2 Effects of policy and infrastructural changes, economic crises, and conflict

The trend data can also be used to look at the effect of specific events such as policy changes (e.g., the introduction of visa waiver agreements or the lift of a travel ban), alterations in travel infrastructure (e.g., new airline connections), sudden political instability (e.g., a coup d'état), as well as longer processes such as economic crises (e.g., depressions) and violent conflicts (e.g., civil wars) on the mobility of people. Six examples are shown in Fig. 5. The first four look at the effect of specific events that can be pinpointed to a precise date (indicated by solid arrows in the graphs). The last two examples investigate longer processes that evolve over several years (indicated by dashed horizontal arrows). While the plots alone do not allow us to *prove* causal effects, we believe that they *are* able to illustrate several plausible relations and provide a good starting point for further, more

systematic analyses across a larger number of country pairs as well as in-depth analyses of specific cases.

The first example (Fig. 5(a)) shows the effect of a visa waiver agreement between Cape Verde and the European Union on mobility from the African island state to Italy (CPV→ITA). The agreement entered into force on December 1st, 2014 [36] and its effect is clearly visible, with a stark increase of passenger traffic to Italy almost immediately after the liberalization. The second example (Fig. 5(b)) demonstrates the effect of a ban of Russian charter flights to Turkey being lifted at the end of August 2016. The ban had been imposed in 2015 after a Russian combat plane had been shot down by the Turkish air force on the border to Syria, triggering a diplomatic crisis between the two countries [37, 38]. After the ban was lifted, a steep increase in passenger mobility occurred, as visible in the figure.

The next two cases are a bit more complex. The third example (Fig. 5(c)) looks at the possible effect of a new airline connection. In June 2014, budget airline Wizz Air introduced new inexpensive flights between Dortmund in Germany and Tuzla in Bosnia and Herzegovina [39]. This event appears to have had a positive effect on the number of passengers, as Fig. 5(c) reveals. However, the causal relations between demand and supply may be less straightforward in this case: there was already an increase in mobility from 2013 to 2014, *before* the new connection was introduced (possibly due to a visa liberalization introduced in 2010, see below). It is thus well possible that the airline reacted to rising demand when introducing the new connection. This new connection then led to broader and cheaper supply, which subsequently spurred further demand. Such self-reinforcing effects are well-documented in research on road infrastructure expansion [40].

The fourth example, mobility between Russia and Egypt, illustrates the effects of several violent events in combination (Fig. 5(d)). The initial drop in 2011 may be connected to the Arab spring and the Egyptian revolution, which took place on January 25th, 2011. The second drop in the middle of the time series seems to be due to the coup d'état, which took place on July 3rd, 2013. Two years later, on October 31st, 2015, an Airbus A321-231, operated by the Russian airline Metrojet, exploded shortly after it departed from Sharm El Sheikh, Egypt. The crew and all 224 passengers, most of whom were Russian tourists, died. The consequences of this attack on the perceived security of Egypt as a destination for Russian travellers may partly explain the third sharp drop in Fig. 5(d), when air traffic from Russia to Egypt fell to the lowest levels in the time frame under study. This link is not entirely clear though, since mobility already started to drop *before* the date in question.

As mentioned above, the final two cases relate to processes rather than events. The fifth example (Fig. 5(e)) shows the effect of the economic depression in Venezuela on air passenger traffic to the United States. The country's crisis began in 2010, when President Chávez declared "economic war" on speculators in response to rising food prices [41]. As one can see, air passenger mobility from Venezuela to the United States first increased from 2011 to 2013, possibly due to well-off citizens leaving the country via plane. However, when the Venezuelan economy started to rapidly deteriorate and eventually collapse in 2013, the amount of air passengers flying from Venezuela to the U.S. began to drop quickly in 2014 and 2015 and reached very low levels in 2018. This does not mean, of course, that fewer people left Venezuela after 2013 to escape economic hardship. However, rather than flying to more distant countries such as the U.S., many of the recent migrants and refugees, more than 2 million, drove or even walked to neighboring countries in South America [42–44].

A second explanation for the stark decline is the fact that foreign visitors to the country (whose return flights we should see in this graph) may have been increasingly deterred by the unfolding crisis.

The sixth and final example (Fig. 5(f)) illustrates the consequences of a violent conflict, the civil war in Syria, on transnational human mobility. It shows the case of air passenger mobility from Spain, which rapidly decreases as the Syrian civil war escalates from March 2011 onward until it ceases entirely in late 2013.

These example cases show that, compared to conventional sources of global mobility with their yearly or even decadal updates, the high temporal resolution of the air passenger data (*i.e.*, its monthly updates) facilitates the analysis of how specific events and processes affect mobility flows between countries. This highlights the potential of air traffic data to reveal specific sociopolitical and economic drivers behind human mobility across nation-state borders. Potential applications are manifold and range from testing the effects of policies (*e.g.*, changes in visa requirements) on transnational mobility flows to examining the economic consequences of violence (*e.g.*, via the deterrence of tourism after terrorist attacks).

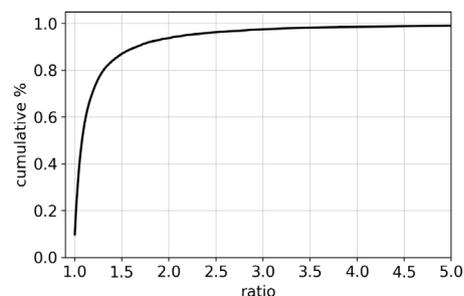
### 3.1.3 What is the relation between air passenger traffic and migration?

To answer this question, we first look at the cumulative sum of passengers over the years for both directions ( $A \rightarrow B$  and  $A \leftarrow B$ ), to test whether a divergence between the two might indicate that some of the passengers stay in the country with the surplus in passengers to become migrants. Figure 6 shows the cumulative distribution of the discrepancies expressed as ratios ( $A \rightarrow B$  over  $A \leftarrow B$ ).<sup>8</sup> It indicates that deviations are quite substantial overall. For about 80 percent of cases, the deviation lies below 25 percent (= a ratio of 1.25), but in about 20 percent of the cases, the difference is even larger, with a maximum ratio of 150.

This implies that it is indeed the case that not all passengers who travel via plane to one country return directly to the country of origin. But does this also mean that these people stay on to become migrants? To answer this question, we combined this information with data on migration flows between OECD countries [30]. Results do, however, not show a clear correlation and we conclude that this surplus in mobility in one direction over the other cannot easily be used as an indicator of migration.

We propose three possible explanations for why it is not feasible to carve out such a clear correlation. First, it is difficult to align temporally the inbound and outbound movements to calculate the surplus. Since air passengers data does not allow to track individual passengers or specific cohorts on the basis of their date of entry, the surplus method relies on assumptions about the length of stay of the passengers. The surplus between arrivals

**Figure 6** Discrepancies between the cumulative sum of mobility in the two opposite directions ( $A \rightarrow B$  vs  $A \leftarrow B$ ). The upper bound in the graph is set to a ratio of 5 to better illustrate the relation; the actual maximum ratio is 150



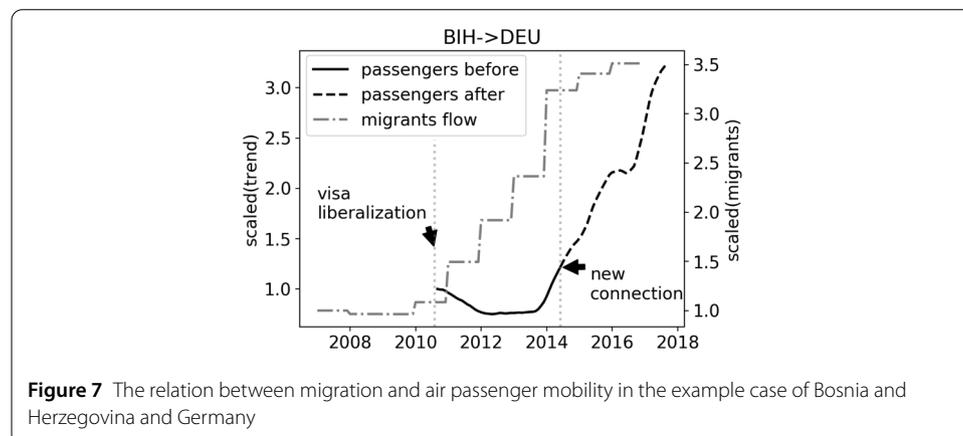
and departure at the aggregate level may represent migrants with arbitrary combinations of length of stay rather than corresponding to the fixed length of stay above 12 months used as threshold to define a migrant in the official statistics. Also using the differences in cumulative sums like in our case does not solve entirely the problem since it leaves open the question on how to account for the passengers arriving and leaving at the end-points of the period considered for the cumulative sums.

Second, there is a distortion in the data due to existence of hub airports and open jaw flights: people may fly to one central airport (e.g., to Paris, France) and then change air carrier, go by car or train to visit other countries (e.g., in Europe) or even return to their home country from a different hub airport (e.g., Moscow). This phenomenon may also result in discrepancies between the cumulative sum in the two opposite directions ( $A \rightarrow B$  vs  $A \leftarrow B$ ) if some hubs are popular entry gates and others are popular exit gates.

The third, and probably most important issue, is linked to the high difference in magnitude between the flows of air passengers and migration. Since the number of air passengers between two countries is several orders of magnitude higher than the number of migrants, any noise or bias in the estimate of the surplus from the air passengers' data could have disproportionate consequences on the estimates of much smaller migration flows.

The negative result for the use of the surplus method to estimate migration, however, does not mean that there is no relation between air passenger mobility and migration at all. In fact, at the aggregate level of annual bilateral data, we found a strong correlation between the overall number of air passengers moving between a pair of countries (in both directions) and the flows of migrants from OECD statistics ( $r = 0.77, p < 0.001$ , using logarithmic transformed values in both cases).

To illustrate this association, we draw again on the example case of Bosnia and Herzegovina and Germany that was introduced in the previous section. Figure 7 shows the trend in air passenger mobility that we saw earlier, combined with information on migration flows between the two countries. As can be seen, the shapes of the two trends are extremely similar, with the growth in migration flows preceding the proportional growth in air passenger traffic by about three years. The growth of migration flows seems to be (at least partially) attributable to a visa liberalization agreement between Bosnia and Herzegovina and the EU that entered into force in late 2010 [45]. The lagged parallel increase of air passenger traffic three years later seems to demonstrate that many of these migration flows do not take the form of 'classic' international migration, consisting of a one-time move to



a destination country into which the migrant then fully integrates without ever returning to his or her place of origin. Rather, it seems to follow the newer ideal type of *transnational* migration, where migrants continue to move back and forth between country of origin and destination, with experiential and cognitive horizons (what philosophers and sociologists call 'life-worlds' after Edmund Husserl and Alfred Schütz) spanning across a transnational space that encompasses both countries [46–49]. Hence, the rise in airline traffic after the rise in migration took place is likely due to people from Bosnia and Herzegovina who migrated to Germany returning to their country of origin to visit family and friends, thereby increasing air passenger traffic between the two countries (which may then lead airline companies to install new connections, which reinforces the trend, see above).

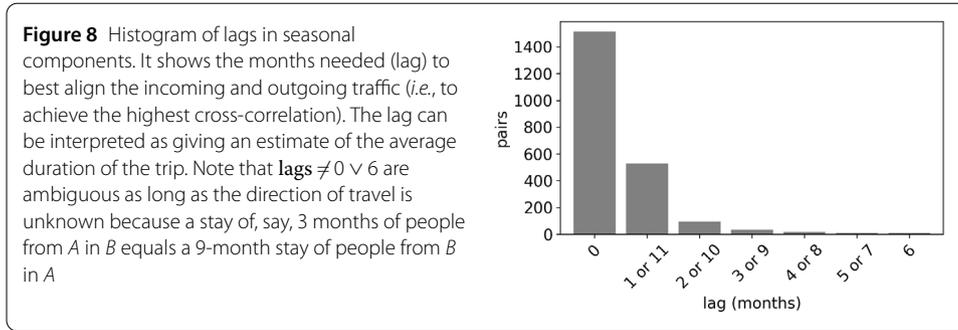
The high correlation between the flows of air passengers and of migrants, as reported in OECD statistics, could be determined by the presence of migrants among air passengers but also indicative of the fact that both the flows of air passengers and the flows of migrants between given combinations of countries are influenced by the same geographic and socioeconomic and structural factors that define preferential ties between countries. The anecdotal example of Bosnia and Herzegovina and Germany shows the establishment of these stronger relations through the two events of a visa liberalization at the end of 2010 and the opening of new airline routes a few years later. Although with different lags, these events, which are probably interconnected, determine both the increase of flows of migrants and of air passengers. The strong correlation for the whole dataset and the example of Bosnia and Herzegovina and Germany show the importance of considering migration and transnational human mobility as part of a unique phenomenon shaped by the same forces of modernization and globalization that are at the basis of the established network of international relations between countries.

The fact that both the flows of migrants and of air passengers tend to follow the same preferential bilateral pathways between countries leaves open the possibility of exploiting global air passenger data to get to an estimate of migratory flows especially in the areas of the world less covered by official statistics on migration. One conclusion of our attempt to estimate migration from the air passenger data is that rather than trying to extract the absolute number of migrants (understood as travellers who stay) from the surplus between inbound and outbound flows at different lags, the air passenger data can work as a method of estimation by allowing to quantify the strength of bilateral corridors which both define the movement of migrants and other forms of international mobility such as tourism, business trips and temporary work migration.

### 3.2 Seasonal analysis

The cross-correlations between the seasonal pattern from *A* to *B* and from *B* to *A* allow to discern three different patterns that appear to hint to different types of mobility dominating the movements between the country pair in question. The classification into the three patterns depends on the significance of the correlation (p-value), the degree of (a-)synchronicity (lag in months), and the peak ratio (*i.e.*, the relative importance of the seasonal component, see above). The higher the correlation, the clearer the case, in all three patterns.<sup>h</sup>

Figure 8 shows the distribution of lags for all country pairs with a meaningful seasonal effect (in line with the two thresholds introduced above). In the majority of cases, the series are aligned (lag = 0), but for some country pairs, they are not (lag > 0). We will see



that despite being relatively rare, these asynchronous cases are very interesting because they appear to be connected to different kinds of mobility than the synchronous ones. A peculiarity is the fact that only lags of 0 and 6 months are unambiguous. All other case have two interpretations: a lag of 1 month equals a lag of 11 months when the opposite direction of mobility is considered, 2 months correspond to 10, 3 to 9, 4 to 8, and 5 to 7.<sup>i</sup> From the seasonal air passenger data alone, it is not possible to understand which interpretation is more meaningful. Only further, external information about the direction of the flows may help to make such interpretations. We will look at several examples below, including a case where a lag of 1 or 11 months is connected to mobility where individuals tend to spend one month abroad (DEU  $\longleftrightarrow$  VNM) and another where people actually stay 11 months abroad (PHL  $\longleftrightarrow$  SWE). In the following, we first describe the three patterns and give three country-pair examples for each (cf. Fig. 9). Then, we present the systematic, overall picture, looking at structural differences between these patterns globally.

3.2.1 *Pattern I: mass tourism* ( $p < 0.05$ , lag = 0)

The first pattern occurs when there is a significant correlation between the two directions ( $p < 0.05$ ) and a high synchronicity (lag = 0). The seasonal pattern of country pairs falling into this category appears to be dominated by mass tourism, visits, and business travel *within* world regions. Three examples are visible in Fig. 9(A). The first example case is mobility between Great Britain and Italy (GBR  $\longleftrightarrow$  ITA) on the left-hand side. The correlation is high ( $r = 1$ ,  $p < 0.05$ ) and synchronous (lag = 0). The peak season takes place during the European summer months and the flow from Great Britain to Italy is slightly ahead of the reverse flow from Italy to Great Britain (visible from the small divergence in April/May on the one hand and September/October on the other). Both factors taken together seem to suggest that this seasonal pattern is dominated by holiday tourists from Great Britain travelling to Italy during the summer.

The second example (center graph in Fig. 9(A)) shows the flows between Germany and the Dominican Republic (DEU  $\longleftrightarrow$  DOM). The pattern is very similar to the one observed for GBR  $\longleftrightarrow$  ITA, except for the fact that the peak season now occurs during the European winter months. Taking common knowledge about the geographic and socio-economic specificities of the two countries into account, we can assume that this pattern is predominantly driven by German tourists travelling to the Dominican Republic to catch some sun during the cold winter months. Note that this flow spans two world regions and is thus rather exceptional for this pattern.

The third example (right-hand side in Fig. 9(A)) shows the flows between Bolivia and Peru (BOL  $\longleftrightarrow$  PER). While the season is spread relatively evenly across the year, the peak



months are now during the spring of the southern hemisphere (September–November). The flow from Bolivia to Peru seems to slightly precede the reverse flow from Peru to Bolivia, indicating that the majority of the movements likely originate there.

In all three examples for Pattern I, the peak ratio is relatively low (between 0.16 for BOL ↔ PER and 0.32 for DEU ↔ DOM), indicating that while these seasons exist and are meaningful in absolute terms, they constitute a relatively small share of all mobility be-

tween the countries involved. For example, while the main season for travelling between the UK to Italy is during the summer months, there is also a lot of mobility all year round.

### 3.2.2 *Pattern II: individual tourism* ( $p < 0.05$ , lag = 1)

The second pattern requires that there is a significant correlation between the two directions ( $p < 0.05$ ) and a slight asynchronicity (lag = 1). The seasonal pattern of country pairs falling into this category seems to be dominated by individual tourism over large distances. Three examples are visible in Fig. 9(B).

The first example is Germany and Vietnam (DEU  $\longleftrightarrow$  VNM) on the left-hand side. Apart from being slightly asynchronous (lag = 1), the pattern is also different from the ones observed so far in that it has *two* seasons instead of one: a first one peaking in January and March and a second one with peaks in July and August. The January–March peak corresponds to one of the two seasons that are best to travel Vietnam, with moderate temperatures and low rainfall. The second peak is in line with the main holiday season in Germany, but also partially overlaps with the second pleasant season, weather-wise, in Vietnam (August–October) [50].

The second example is Germany and Nepal (DEU  $\longleftrightarrow$  NPL, center graph in Fig. 9(B)). Again, there is a lag of 1 month and, since the mobility from Germany to Nepal precedes the return mobility from Nepal to Germany, this seasonal pattern appears to be driven by German tourists going to Nepal in October to return in November and in March to return in April. In theory, it could of course also be possible that the pattern is driven by Nepalese going to Germany for an 11-month stay, but Germany is not a typical destination for Nepalese migrant workers (in contrast to Qatar, see below). The October–November peak coincides with “the best season for trekking in Nepal” [51], while the March–April peak corresponds to the second tourist season in Nepal, when excursions to Mount Everest take place [51]. This case is thus quite similar to the seasonal pattern between Germany and Vietnam. However, it is notable that whereas the seasonal component for DEU  $\longleftrightarrow$  VNM is rather small relative to the overall mobility between the two countries (peak ratio: 0.33), the seasonal component for DEU  $\longleftrightarrow$  NPL is extremely dominant (peak ratio: 1.43). This means that Germans almost only travel to Nepal during the two peak seasons, but their travels to Vietnam are more evenly spread throughout the year.

The third example is Argentina and Thailand (ARG  $\longleftrightarrow$  THA) on the right-hand side of Fig. 9(B). This is a case of long-distance travel with a lag of 1 but only one season (in contrast to the previous two examples). Argentinians tend to travel to Thailand during the summer of the Southern hemisphere (mainly January–March), which coincides with the cool and dry season in Thailand that is considered the “best time to travel” [52], and return to Argentina with one month delay. In this case, too, a large share of the overall mobility between the two countries occurs during this main season (peak ratio: 0.96).

### 3.2.3 *Pattern III: seasonal work migration* ( $p < 0.05$ , lag > 1)

The third pattern exists when there is a significant correlation between the two directions ( $p < 0.05$ ) and a major asynchronicity (lag > 1). The seasonal pattern of country pairs falling into this category seems to be dominated by temporary migrant workers, usually over longer geographic distances. Three examples are given in Fig. 9(C). The first one is Spain and Senegal (ESP  $\longleftrightarrow$  SEN) on the left-hand side. The correlation is high ( $r = 0.87$ ,  $p < 0.05$ ) and asynchronous with a lag of 4 or 8 months. Senegalese form an increasingly

large part of the seasonal migrant worker population in Spain as Eastern Europeans become less willing to do the temporary, poorly paid jobs. Following a 2007 memorandum of understanding between Spain and Senegal, they are often hired directly in Senegal to work in the Spanish fishing industries and agriculture, *e.g.*, to harvest non-mechanized crops such as strawberries. Most of the Senegalese hired are women with social obligations (*e.g.*, with young children), who are expected to be more likely to return to Senegal after the seasonal work is done, and not to stay on to become irregular migrants [53, 54].

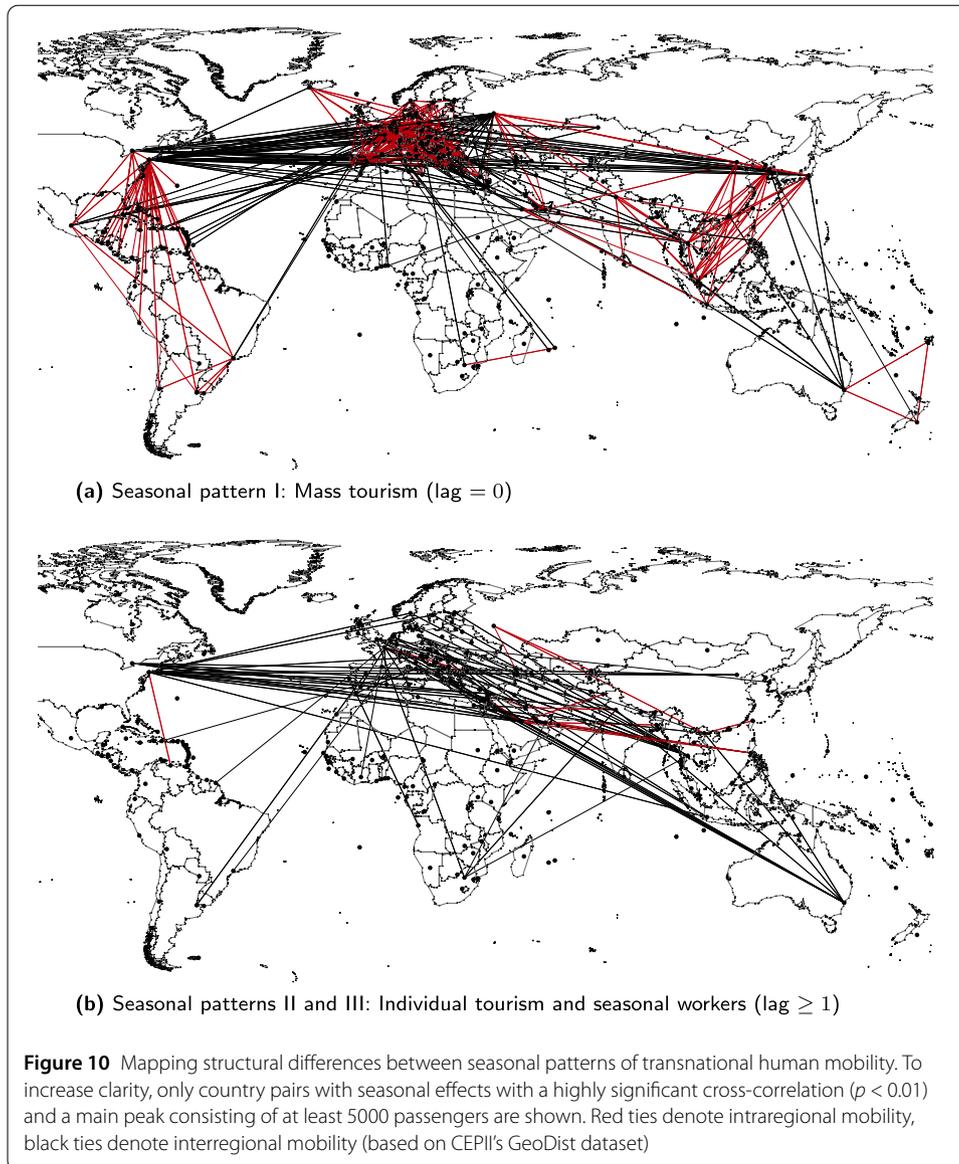
The second example is the Philippines and Sweden (PHL  $\longleftrightarrow$  SWE, center graph). Ten percent of the Filipino population work abroad and 28 percent of seafarers worldwide are Filipinos [55]. Furthermore, from the early 2000s, there has been an increased interest among shipping companies to change to Swedish flag due to Sweden's shipping policy and, simultaneously, Sweden increased the number of work visas it gives to Filipinos [56]. According to a spokesman of the Swedish Shipowners Association, "shipping is an international business where Swedish ship owners turn to the world's largest sailor manpower pool, the Philippines, to man their vessels" [56]. It is thus highly plausible that the seasonal pattern observed here is predominantly driven by Filipino seafarers who fly to Sweden in January to sign up on a ship under Swedish flag, stay on the ship for most of the year and return to their home country only in December, for a short holiday. The result is the particularly asynchronous pattern with a lag of 11 (or 1, depending on the perspective) month(s), and a high correlation ( $r = 0.97$ ,  $p < 0.01$ ).

The third example is Nepal and Qatar (NPL  $\longleftrightarrow$  QAT) on the right-hand side. Since the mid-1990s, the Number of Nepalese workers in Qatar increased steadily and in March 2005, a labour agreement was signed between the two countries. It is estimated that between 1.3 and 2.6 million Nepalese migrants work in the Gulf countries, and Qatar hired 34 percent of all Nepalese migrants in 2008 [57]. More than 95 percent of Nepalese migrants are unskilled or semiskilled workers, and even the Nepalese embassy in Doha advertises their cheap hiring costs as an important asset: "Nepalese workers are comparatively cost effective and their hiring cost is lower as compared to other labour exporting countries" [57]. The correlation is again relatively high ( $r = 0.7$ ,  $p < 0.01$ ) and asynchronous with a lag of 10 (or 2) months.

Overall, Fig. 9 thus reveals that a single indicator, the lag between initial and return mobility, allows to discern three different types of mobility that drive the cross-border movement of people around the world: mass tourism, individual tourism, and seasonal work migration. Going beyond these single case studies, we now attempt to outline the resulting global picture.

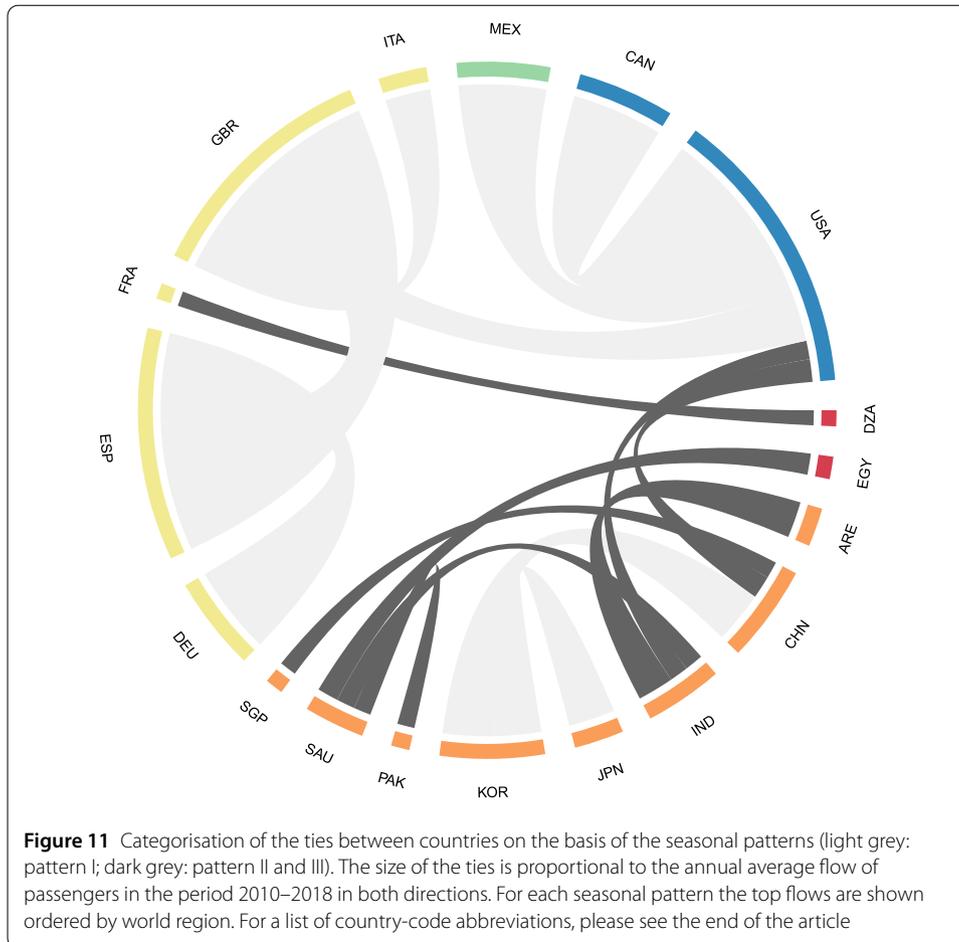
#### 3.2.4 Toward a global classification of country pairs by mobility type

The two maps in Fig. 10 show where the country pairs dominated by these different types of mobility are predominantly situated and how the seasonal patterns lead to structurally different networks at the global scale. Red lines indicate *intra*regional mobility, while black lines stand for *inter*regional mobility (based on a classification from CEPII's GeoDist dataset [58]). There is a clear contrast between the two maps. The network of ties classified as falling under Pattern I (lag = 0), *i.e.* country pairs with seasonal patterns dominated by mass tourism, visits and business travel (Fig. 10(a)) resembles a densely interconnected, decentralized web, with many ties occurring within world regions (visible by the clusters of red ties). There are particularly strong intraregional clusters within Eu-



rope and within Southeast Asia. By contrast, the network of ties with  $\text{lag} \geq 1$  (Patterns II and III, Fig. 10(b)), *i.e.*, of individual tourism and temporary work migration, is more organized in a star structure with many ties going from countries in the Global South on one end to Europe, the U.S., the Middle East, and Australia on the other, but with almost no connections within the Global South or within the Global North (visible through an overwhelming dominance of black ties).<sup>1</sup>

While the former structure (Fig. 10(a)) is thus more regionalized, the latter (Fig. 10(b)) resembles the global core-periphery structure that has been described in world-systems theory [59, 60]. Comparing the two figures, this is clearly discernible through the dominance of red ties in Fig. 10(a) and their absence in Fig. 10(b). Even the very few red lines in Fig. 10(b) are mostly long-distance ties across Asia that could equally be classified as “interregional”. The difference is particularly visible when looking at Southeast Asia: whereas almost all Southeast Asian countries are connected to each other in Fig. 10(A), none of them are in Fig. 10(B). This visual impression can be backed up with statistics. Using the



assignment of countries to world regions (America, Asia, Africa, Europe, and Pacific) from CEPII’s GeoDist dataset [58], it can be established that while 54.7 percent of all ties of Pattern I connect countries that are situated in the same region, this is true for only 19.4 percent of all ties of Patterns II and III. The vast majority of ties in Patterns II and III is thus interregional, which holds only for a minority of ties in Pattern I. This confirms our classification and fits the idea that most mass tourism occur over short distances [61, 62], whereas individual tourism and the market for cheap labor migrants are globalized.

Figure 11 shows the mobility ties between country pairs with the largest seasonal components, ordered by world region (denoted by different colors). Synchronous ties (with lag = 0) are shown in light grey, whereas asynchronous ties (with lag ≥ 1) are shown in dark grey. The size of the ties is proportional to the annual average flow of passengers in the period 2010–2018 in both directions. The graph reveals that the largest flows with no lag take place between countries that are comparatively well-off economically and where one country is indeed a traditional tourism destination (e.g., Spain or Italy), and/or between neighboring countries with large populations (Mexico ↔ USA, Canada ↔ USA, South Korea ↔ Japan). The largest ties characterised by a prolonged stay (lag ≥ 1, dark grey), by contrast, are country pairs that can plausibly be connected to flows of seasonal workers (e.g., flows between Pakistan, India, and Egypt as providers of cheap labor and the United Arab Emirates and Saudi Arabia as buyers of that labor). Individual tourism with

a lag of one month, however, is not clearly visible in Fig. 11, likely because its volume is smaller than the largest seasonal work migration flows.

#### 4 Summary and discussion

Despite its significance as a driver of globalization, with major socioeconomic, cultural, political, and environmental effects, transnational human mobility is rarely empirically grappled with on a global scale. In this article, we dissected monthly data on air passenger flows between countries worldwide from 2010 to 2018 to get new insights into trends and types of human cross-border mobility.

Regarding *trends* in transnational mobility, the main substantive findings are:

- In the bulk of country pairs, the volume of international travels was on the rise in the period under consideration. Persistently growing flows were almost six times more frequent than persistently declining flows.
- The uniquely detailed, monthly passenger data allows to link increases and declines in mobility between countries to specific events, from changes in visa regulations to economic downturns to socio-political crises and violent conflict.
- Mobility between country pairs is typically not identical in size in both directions. However, our hypothesis that a surplus in passengers in one direction could be related to migration-oriented travels could not be corroborated with migration figures from other sources.

The latter finding demands further inquiry with different data and methods. Regarding *types* of transnational mobility, the monthly nature of the dataset allowed us to extract a seasonal component of flows. Analyzing this seasonal component, we found that:

- The lag between incoming and outgoing air passenger traffic indicates the type of mobility (mass tourism, individual tourism, or seasonal work migration) that dominates the flow of people between a given country pair.
- While country pairs connected by mass tourism tend to lie on the same continent, those connected by individual tourism and seasonal work migration overwhelmingly span across continents.
- A novel global typology of seasonal mobility was created that could serve as an important resource for further exploration in future research, through case- or area-studies and matching with other global information.

Taken together, these findings promise to advance the field by creating a more detailed account of how and why people move around the world. Yet, several caveats remain. A first limitation of this study is inherent in the data that was used. Air passenger traffic covers a substantial part, but not the entirety of transnational human travels. By relying on this source we are thus prone to underestimate mobility between geographically close countries (and adjacent countries in particular), since a lot of this mobility occurs through road and railway traffic. Water transportation can also be significant in some cases (*e.g.*, ferries between Estonia and Finland).

A second limitation inherent in the data is that we lack individual-level information—who travelled, for what reasons, for how long. We also lack information on whether mobility events were one way or return flights. While this ensures that the privacy of specific individuals is protected and no inferences about the itineraries of concrete persons can be made, this also limits the capacity to use the dataset to infer migration flows. Nonetheless, the time-lagged component of the data permitted us to discern country pairs in which

seasonal transnational mobility equates mostly with mass tourism from others in which either individual tourism or temporary work migration stand out. We could single out these cases in a typology and map their occurrences globally. Such a data-driven classification is not exhaustive and entirely clear-cut, being based on a time-based component only, which is an imperfect indicator of the underlying types, and we employed external qualitative knowledge to validate it. This classification must therefore be seen as a mere first step to be enriched introducing additional information.

Overall, our analysis has several meaningful implications and points to directions of future research. The trend analysis corroborated existing evidence on the continuing, uninterrupted rise of international travels, in particular using flight transportation. This raises important questions about the environmental sustainability of this trend. Furthermore, the inequality in this regard, with mobility increasing between many country pairs, but stagnating or even dropping between others, calls for further inquiries into the global stratification of access to cross-border mobility and its opportunities.

In the analysis of the seasonal component, one of our more specific findings was the asynchronicity of movements on a monthly basis in a relevant proportion of cases. This may have implications in terms of equity between passengers, the environmental impact of air traffic, and the spread of epidemics. If mobility is synchronized in both directions, it is likely not a problem for airlines to fill seats on both ways. However, if mobility is highly seasonal and occurs with a large lag, the question arises of whether passengers are charged for the empty return seat on top of their own mobility costs. In other words, is it possible that seasonal migrant workers (high lag between initial and return mobility), in contrast to tourists (no or very small lag), have the financial burden of the asynchronicity of their transnational movements? Environmentally, this also raises the question of whether asynchronicity entails suboptimally used flights with negative externalities in terms of overall airplane emissions. A further question concerns possible implications of such lags for pandemic spreading. While the global air travel network has been used in past research to explore the relation between human cross-border mobility and the spread of diseases [63–65], to the best of our knowledge no studies have examined the role of such seasonal patterns and structural asymmetries in the movement of pathogens. All these questions could be scrutinized in future research.

Furthermore, the methods used in this article may encourage to explore: (a) the sub-national dimension of air passenger travels, which might reveal different trends and types of mobility; (b) the interaction of aviation-based mobility with mobility based on other transportation means, to dissect their relative weight and geographical distribution; (c) a larger time-frame, to highlight continuities and changes in transnational human mobility.

Given its observed and substantial rise, transnational human mobility will likely continue to be a salient topic in the coming decades, demanding further explorations of where people move for which reasons, and what the social, political, and environmental consequences are for the planet. We hope that the methods, data and analyses presented in this paper will be helpful for future research that may aim at tackling some of these questions.

#### **Acknowledgements**

We would like to thank Fabio Mazzarella, Elena Navajas Cawood, and Panayotis Christidis for the support provided in the extraction and exploratory analysis of the data. We are also grateful for feedback from participants in the MPC Seminar at the European University Institute's Migration Policy Centre, where a previous version of this paper was presented, and for extremely helpful comments from two anonymous reviewers at *EPJ Data Science*, which improved the manuscript a lot.

### Abbreviations

ARG, Argentina; BGR, Bulgaria; BIH, Bosnia and Herzegovina; BOL, Bolivia; CAN, Canada; CHN, China; CPV, Cape Verde; DEU, Germany; DOM, Dominican Republic; DZA, Algeria; EGY, Egypt; ESP, Spain; FRA, France; GBR, Great Britain; IND, India; ITA, Italy; JPN, Japan; KOR, South Korea; MEX, Mexico; MNE, Montenegro; MOZ, Mozambique; MYS, Malaysia; NPL, Nepal; PAK, Pakistan; PER, Peru; PHL, Philippines; QAT, Qatar; RUS, Russia; SAU, Saudi Arabia; SEN, Senegal; SGP, Singapore; SVK, Slovakia; SWE, Sweden; THA, Thailand; TUR, Turkey; UKR, Ukraine; USA, United States of America; VEN, Venezuela; VNM, Vietnam.

### Availability of data and materials

Given the private nature of the air passenger data it is not possible to release the full information, *i.e.*, the raw time series. However, we generated a set of indicators that allow other researchers to carry out further studies and explore relations between specific sets of countries that were not highlighted in this article (available at: <https://bluehub.jrc.ec.europa.eu/migration/app/?state=5d6005b30045242cabd750a2>). Two indicators describe the trend of mobility between countries (for detailed information about the indicators see the Data and Methods section above): • *Spearman's rank correlation*. The first indicator concerns the trend of passengers between countries. The trend is obtained through the time series decomposition of the data and is released at annual granularity for the period 2011 to 2016. The Spearman's rank correlation provides information on the monotonicity of the increase or decrease in the trend over time. • *Difference between 2016 and 2011 trend value*. The second measure indicates the intensity of the change in mobility between the first and last year under study, based on the annual trend time series. This indicator was normalized with respect to the population of the countries of origin. The other two indicators describe the seasonal pattern of mobility between countries: • *Lag*. For each pair of countries, the cross-correlation of seasonal mobility in both directions was calculated. The lag indicates the asynchronicity of seasonal flows between the country pair. It allows to understand the type of mobility that dominates seasonal mobility between countries, with a **lag** = 0 suggesting mass tourism and a **lag** > 0 indicating long-distance individual tourism or seasonal work migration. • *Peak ratio*. The last indicator is the ratio between the number of people moving during the peak month of the seasonal component and the mean overall number of passengers per year. It describes the prominence of the seasonal mobility relative to all mobility between a given country pair. Sender and receiver countries are identifiable through ISO 3166-1 alpha-3 country codes in the dataset. We invite interested researchers to download and explore the data.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

LG prepared the dataset. LG and ED performed the data analysis. ED and LG wrote the manuscript. ER and FN commented on the analyses and contributed to various sections. MV conceived and supervised the work. All authors read and approved the final manuscript.

### Author details

<sup>1</sup>European Commission, Joint Research Centre (JRC), Ispra, Italy. <sup>2</sup>University of Göttingen, Göttingen, Germany. <sup>3</sup>Migration Policy Centre, European University Institute, Florence, Italy. <sup>4</sup>Sciences Po, Observatoire Sociologique du Changement (OSC), CNRS, Paris, France. <sup>5</sup>Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy.

### Endnotes

- <sup>a</sup> In this study, we use the term "transnational" in the meaning it has in the field of international relations, where it is used to describe any movement by non-state actors that spans across national borders [66]. 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 [47].
- <sup>b</sup> The original data is at the airport level. Here, we draw on an aggregated country-to-country flow level version.
- <sup>c</sup> We removed the labels from the y-axis in this graph and re-scaled axes in following graphs through normalization for the sake of parsimony as the analysis in this paper focuses on trends and patterns rather than absolute numbers of passengers.
- <sup>d</sup> Note that in some cases, the residual may actually contain meaningful information. This occurs when exceptional (*i.e.*, unique or irregular) events have an effect on mobility between countries. Examples include the soccer world cup, which takes place in a different location every four years, or the Hajj. In this latter case, Saudi Arabia attracts air passengers from Muslim countries (*e.g.*, Nigeria) during this pilgrimage, but since the date of the Hajj changes each year in the Gregorian calendar, its effect is not well-captured in the seasonal component. To study the effect of such exceptional events, one would thus have to examine the residual or the raw observations. Doing so would, however, go beyond the scope of this article.
- <sup>e</sup> These excluded cases (VEN → MOZ, TUR ↔ SVK, MNE → UKR, and BGR ↔ SVK) are all structurally similar in that they have one extreme peak in the monthly time series that is mistakenly classified as a seasonal trend by the time-series decomposition.
- <sup>f</sup> As an additional test, we reran this analysis taking the development of the population size during the time span under study into account. Overall the results (available upon request) look very similar to those presented in Fig. 4(a). However, increases are slightly less common and decreases slightly more common in the population-size-adjusted version. This is, of course, highly plausible, since the population also grew in most countries during this time span. Thus, if we adjust for population size, the growth in mobility has to be larger than the growth in population to count as an 'actual' increase. In any case, the general picture is unaffected: increases dominate, regardless of whether mobility trends are considered in absolute terms or relative to the population size.
- <sup>g</sup> For this calculation, we removed all pairs (*a, b*) where *a* or *b* = 0, since the ratio is undefined when the denominator is zero.

- <sup>h</sup> Cases with a large and meaningful seasonal component but a low and non-significant cross-correlation might be a consequence of open-jaw flights and multi-lateral connections between countries (see above).
- <sup>i</sup> In theory, lags  $\geq 12$  months are also possible if people stayed a year or longer. Thus, a lag of 0 could also mean 12, 6 may actually mean 18. However, given the tailed distribution in Fig. 8, we assume there is a decreasing number of occurrences as the time lag increases, even beyond 6 months. In other words, we assume that time lags greater than 6 months are less likely to be observed with respect to **lags**  $\in [0, 6]$ , and for this reason we only consider time **lags**  $\in [0, 11]$ .
- <sup>j</sup> Due to the difficulty of discerning seasonal patterns with a lag of 1 month from those with a lag of 11 months from within the data itself, we present Patterns II and III combined in a single graph. Looking at the structure of the global networks of ties with a lag of 1 and 11 months versus those with a lag of 2 to 10 months shows that these networks are structurally rather similar, with a large share of long-distance mobility in both cases.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 February 2019 Accepted: 19 August 2019 Published online: 30 August 2019

## References

1. Bogoch II, Creatore MI, Cetron MS, Brownstein JS, Pesik N, Miniota J, Tam T, Hu W, Nicolucci A, Ahmed S et al (2015) Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 West African outbreak. *Lancet* 385(9962):29–35
2. Zhu Y-G, Gillings M, Simonet P, Stekel D, Banwart S, Penuelas J (2017) Microbial mass movements. *Science* 357(6356):1099–1100
3. Chapman L (2007) Transport and climate change: a review. *J Transp Geogr* 15(5):354–367
4. Centeno MA, Nag M, Patterson TS, Shaver A, Windawi AJ (2015) The emergence of global systemic risk. *Annu Rev Sociol* 41:65–85
5. Hui P, Mortier R, Piórkowski M, Henderson T, Crowcroft J (2010) Planet-scale human mobility measurement. In: Proceedings of the 2nd ACM international workshop on hot topics in planet-scale measurement. ACM, New York, p 1
6. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779
7. Csáji BC, Browet A, Traag VA, Delvenne J-C, Huens E, Van Dooren P, Smoreda Z, Blondel VD (2013) Exploring the mobility of mobile phone users. *Phys A, Stat Mech Appl* 392(6):1459–1473
8. Scherrer L, Tomko M, Ranacher P, Weibel R (2018) Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Sci* 7(1):19
9. Mizzi C, Fabbri A, Rambaldi S, Bertini F, Curti N, Sinigardi S, Luzi R, Venturi G, Davide M, Muratore G et al (2018) Unraveling pedestrian mobility on a road network using icts data during great tourist events. *EPJ Data Sci* 7(1):44
10. Alessandretti L, Sapiezynski P, Sekara V, Lehmann S, Baronchelli A (2018) Evidence for a conserved quantity in human mobility. *Nature Human Behaviour* 1
11. Ruktanonchai NW, Ruktanonchai CW, Floyd JR, Tatem AJ (2018) Using Google location history data to quantify fine-scale human mobility. *Int J Health Geogr* 17(1):28
12. Weber I, Zagheni E et al (2013) Studying inter-national mobility through ip geolocation. In: Proceedings of the sixth ACM international conference on web search and data mining. ACM, New York, pp 265–274
13. Zagheni E, Weber I, Gummadi K (2017) Leveraging Facebook's advertising platform to monitor stocks of migrants. *Popul Dev Rev* 43(4):721–734
14. Spyrtatos S, Vespe M, Natale F, Weber I, Zagheni E, Rango M (2018) Migration data using social media: a European perspective. <https://doi.org/10.2760/964282>
15. Hawelka B, Sitko I, Beinart E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located Twitter as proxy for global mobility patterns. *Cartogr Geogr Inf Sci* 41(3):260–271
16. 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. In: Proceedings of the 2017 ACM on web science conference. ACM, New York, pp 103–110
17. 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. IEEE Press, New York, pp 421–428
18. Rango M, Vespe M (2017) Big data and alternative data sources on migration: from case-studies to policy support summary. Report
19. Statista (ed) Smartphone User Penetration as Percentage of Total Global Population from 2014 to 2021. <https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/>
20. White E, Pinsky O Half the World's Population Is Still Offline. Here's Why that Matters. <https://www.weforum.org/agenda/2018/05/half-the-world-s-population-is-still-offline-heres-why-that-matters/>
21. Recchi E, Deutschmann E, Vespe M (2019) Estimating transnational human mobility on a global scale. MPC/RSCAS Working Paper Series. Florence: European University Institute (2019/30)
22. Hummels D (2007) Transportation costs and international trade in the second era of globalization. *J Econ Perspect* 21(3):131–154
23. Fu X, Oum TH, Zhang A (2010) Air transport liberalization and its impacts on airline competition and air passenger traffic. *Transp J* 49(4):24
24. Wandelt S, Sun X (2015) Evolution of the international air transportation country network from 2002 to 2013. *Transp Res, Part E, Logist Transp Rev* 82:55–78
25. Sun X, Wandelt S, Dzikus N, Linke F (2016) Air passenger flow communities between countries. In: Control conference (CCC), 2016 35th Chinese. IEEE, New York, pp 9276–9281

26. SABRE (2014) Aviation Data Intelligence, Leg\_Flow Tables. Retrieved from [https://www.sabreairlinesolutions.com/home/software\\_solutions/product/intelligence\\_exchange\\_apps\\_microapps](https://www.sabreairlinesolutions.com/home/software_solutions/product/intelligence_exchange_apps_microapps)
27. McKinney W, Perktold J, Seabold S (2011) Time series analysis in Python with statsmodels. Jarrodmillman Com, 96–102
28. Holtz P, Deutschmann E, Dobewall H (2017) Cross-cultural psychology and the rise of academic capitalism: linguistic changes in ccr and jccp articles, 1970–2014. *J Cross-Cult Psychol* 48(9):1410–1431
29. Peel M, Brunsden J, Migration The Riddle of Europe's Shadow Population. <https://www.ft.com/content/58f2f7f8-c7c1-11e8-ba8f-ee390057b8c9>
30. OECD International Migration Database. <https://stats.oecd.org/Index.aspx?DataSetCode=MIG>
31. Holderied MW, Baker CJ, Vespe M, Jones G (2008) Understanding signal design during the pursuit of aerial insects by echolocating bats: tools and applications. *Integr Comp Biol* 48(1):74–84
32. Vespe M, Jones G, Baker CJ (2009) Lessons for radar. *IEEE Signal Process Mag* 26(1):65–75
33. Frank J, McEwen BF (1992) Alignment by cross-correlation. In: *Electron tomography*. Springer, Berlin, pp 205–213
34. Karna DK, Agarwal S, Nikam S (2008) Normalized cross-correlation based fingerprint matching. In: *Computer graphics, imaging and visualisation, 2008. CGIV'08. Fifth international conference on. IEEE, New York*, pp 229–232
35. UK Government: Foreign Travel Advice: Egypt. <https://www.gov.uk/foreign-travel-advice/egypt>
36. European Union: Agreement Between the European Union and the Republic of Cape Verde on Facilitating the Issue of Short-stay Visas to Citizens of the Republic of Cape Verde and of the European Union. [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:22013A1024\(01\)&from=EN](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:22013A1024(01)&from=EN)
37. Hürriyet: Russian Charter Flights to Turkish Resorts Resume After Ties Mended. <http://www.hurriyetdailynews.com/russian-charter-flights-to-turkish-resorts-resume-after-ties-mended--103538>
38. Russian Aviation Insider: Russian Carriers Allowed to Resume Charter Flights to Turkey. <http://www.rusaviainsider.com/russian-carriers-allowed-to-resume-charter-flights-to-turkey/>
39. Sarajevo Times: "Wizz Air" in June Will Introduce Cheap Flights for Germany and the Netherlands from Tuzla Airport. [http://tumor.informatics.jax.org/cancer\\_links.html](http://tumor.informatics.jax.org/cancer_links.html)
40. Duranton G, Turner MA (2011) The fundamental law of road congestion: evidence from US cities. *Am Econ Rev* 101(6):2616–2652
41. Carroll R Venezuela's private retailers targeted as Hugo Chávez declares 'economic War'. <https://www.theguardian.com/world/2010/jun/27/venezuela-hugo-chavez-private-retailers>
42. BBC: Venezuela's Migration crisis: is enough being done? <https://www.bbc.com/news/world-latin-america-45546650>
43. Hanson R 4 charts show Venezuela's worsening migrant crisis. <http://theconversation.com/4-charts-show-venezuelas-worsening-migrant-crisis-102464>
44. Freier L, Parent N (2019) The regional response to the Venezuelan exodus. *Current History* 118(805):56–61
45. Council of the European Union: Visa Liberalisation for Albania and Bosnia and Herzegovina. [https://www.consilium.europa.eu/uedocs/cms\\_data/docs/pressdata/en/jha/117555.pdf](https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/jha/117555.pdf)
46. Schiller NG, Basch L, Blanc-Szanton C (1992) Towards a definition of transnationalism. *Annals of the New York Academy of Sciences* 645(1)
47. Wimmer A, Glick Schiller N (2002) Methodological nationalism and beyond: nation–state building, migration and the social sciences. *Glob Netw* 2(4):301–334
48. Levitt P, Jaworsky BN (2007) Transnational migration studies: past developments and future trends. *Annu Rev Sociol* 33:129–156
49. Deutschmann E (2017) Mapping the transnational world: towards a comparative sociology of regional integration. Unpublished Doctoral Dissertation, Jacobs University/University of Bremen
50. Crane Z When is the best time to visit Vietnam? <https://www.insider-journeys.com/blog/best-time-to-visit-vietnam>
51. Highland N Treks: Best time to visit Nepal. <https://www.nepalhighlandtreks.com/best-time-to-visit-nepal.html>
52. Audley When is the best time to visit Thailand? <https://www.audleytravel.com/thailand/best-time-to-visit>
53. Díaz-Diego J, Márquez-Domínguez J (2012) Poor hands in far lands: Senegalese seasonal workers in the strawberry fields of the Huelva province (South-western Spain). *Analele Universitatii din Oradea—Seria Geografie* 22(1):89–111
54. Sinatti G (2011) 'Mobile transmigrants' or 'unsettled returnees'? Myth of return and permanent resettlement among Senegalese migrants. *Population Space and Place* 17(2):153–166
55. Lindgren N, Nilsson J (2012) Filipinos sailing on the seven seas—a qualitative study of Filipino seafarers working on international vessels
56. Nilson C More and more Filipino sailors get Swedish work visas. <https://scandasia.com/170-more-and-more-filipino-sailors-get-swedish-work-visas/>
57. Bruslé T (2009) Who's in a labour camp? A socio-economic analysis of Nepalese migrants in Qatar. *Eur Bull Himal Res* 35–36:154–170
58. Mayer T, Zignago S (2011) Notes on cepii's distances measures (geodist). Paris: Centre d'Etudes Prospectives et d'Informations Internationales. (cepil working paper 2011-25)
59. Wallerstein I (2011) The modern world-system I: capitalist agriculture and the origins of the European world-economy in the sixteenth century, vol 1. University of California Press, Berkeley
60. Galtung J (1971) A structural theory of imperialism. *J Peace Res* 8(2):81–117
61. McKercher B, Chan A, Lam C (2008) The impact of distance on international tourist movements. *J Travel Res* 47(2):208–224
62. Deutschmann E (2016) The spatial structure of transnational human activity. *Soc Sci Res* 59:120–136
63. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A (2010) Modeling the spatial spread of infectious diseases: the global epidemic and mobility computational model. *J Comput Sci* 1(3):132–145
64. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The modeling of global epidemics: stochastic dynamics and predictability. *Bull Math Biol* 68(8):1893–1921
65. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci* 103(7):2015–2020
66. Nye JS, Keohane RO (1971) Transnational relations and world politics: an introduction. *Int Organ* 25(3):329–349