Radiation model for migration with directional preferences

Lucas Kluge,[1,2] Anders Levermann,[1,2,3] and Jacob Schewe[1]

[1]*Potsdam Insitute for Climate Impact Research, Telegrafenberg, 14473 Potsdam, Germany*

[2]*Institute of Physics and Astronomy, University of Potsdam, Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Germany*

[3]*Lamont-Doherty Earth Observatory, Columbia University, New York, New York 10964-1000, USA*

(Received 24 May 2022; revised 24 October 2022; accepted 13 December 2022; published 30 December 2022)

The radiation model is a parameter-free model of human mobility that has been applied primarily for short-distance moves, such as commuting. When applied to migration, it underestimates the number of long-range moves, such as between different US states. Here we show that it additionally suffers from a conceptual inconsistency that can have substantial numerical effects on long-distance moves. We propose a modification of the radiation model that introduces a dependence on the angle between any two alternative potential destinations, accounting for the possibility that migrants may have preferences about the approximate direction of their move. We demonstrate that this modification mitigates the conceptual inconsistency and improves the model fit to observational migration data, without introducing any fitting parameters.

DOI: 10.1103/PhysRevE.106.064138

I. INTRODUCTION

The radiation model (RM) was proposed as a universal model of human mobility [1]. It builds on the concept of intervening opportunities [2], where the flow between two locations depends not only on the properties of those origin and destination locations, but also on the locations in between. The RM has a number of advantages over earlier intervening opportunity models, as well as over the commonly used gravity models [3], in which the flow depends on the origin, destination, and distance between them. The RM can be derived analytically from a particle-diffusion process and is mathematically more self-consistent than the gravity model [1]. It is also parameter-free, which is an advantage for data-limited applications. Especially when it comes to migration, in many countries, data on a detailed spatial level are often unavailable or only partially available, impeding the parameter calibration necessary for gravity models.

After its proposal, the RM and variations of it have been applied to several problems, including commuting [4], urban mobility [5], freight transportation [1], forced migration [6], and the spreading of diseases [7]. Despite its advantages, however, it has been shown that the original RM does not fit the data well for all types of mobility, and several modifications have been proposed to generalize the model, often introducing one or several fitting parameters [8–11]. For instance, the RM underestimates longer-distance moves when applied to internal migration data, while this problem is alleviated in an extended model in which the intervening opportunities are downweighted through an exponent [8]. Another extension introduces two parameters that can be used to give the model either an exploratory or cautious behavior, depending on the application [9]. These parameters allow the user to fit the model to different spatiotemporal scales. Similarly, another generalization introduces a scaling exponent to make the model more applicable to different spatial scales [10].

Notably, both the original RM and the above-mentioned modifications are isotropic, in the sense that flow rates are independent of the orientation of a given destination relative to other locations. This may be true for short-distance and temporary forms of mobility such as commuting, but may be less plausible for permanent moves over longer distances. In this paper we focus on within-country migration and show that the RM yields implausible results when alternative destinations are not in the same direction. To mitigate this conceptual problem, we propose an extension of the model where the intervening opportunities are weighted depending on their direction relative to the destination. We further motivate this assumption by demonstrating the corresponding direction dependence in observational data and we show that the modified angle-dependent model outperforms the original RM when applied to large-scale migration data in several countries, without the need of additional fitting parameters.

II. MODEL AND DATA

A. Radiation model

For the purposes of this paper we will be using the RM as proposed by Simini *et al.* [1] but adding a normalization factor according to Masucci *et al.* [4] (discussed below). The RM can be derived from a particle diffusion process (see the Appendix for a detailed derivation): Each migrant originates from a specific origin, passing through an environment of intervening opportunities, in order to enter some specific destination. Assuming that the distribution of incomes or amenities, i.e., those factors that make a location attractive as a place of residence, is the same everywhere, the actual probability for a given level of income or amenities to be available in a given location depends on the location's total population size. In other words, one may imagine that a larger city offers better-paid jobs because it offers overall more jobs, so the probability of finding a job matching a migrant's income

expectations is higher. The underlying income or amenity distribution then cancels out and the probability of a given move to occur, as well as the overall flow rate between two locations, depends only on the population sizes of the origin, destination, and all locations in between.

We call the area including any number of intervening opportunities an opportunity cell. In the following we will use intervening opportunities and intervening population interchangeably. Contrary to gravity-type models, the distance between the origin and destination has no direct effect on the magnitude of a migration flow. Formally, each bilateral migration flow can be obtained through the equation

$$M_{ij} = M_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{1}$$

where M_{ij} is the number of migrants moving from origin i to destination j; m_i and m_j denote the population of the origin and destination, respectively; M_i is the total number of migrants leaving the origin i, which can also be expressed as $\frac{N_M}{N} m_i$, with $\frac{N_M}{N}$ the ratio of total people moving (N_M) and total population (N); and s_{ij} indicates the number of intervening opportunities and is defined as

$$s_{ij} = \sum_{k \forall d_{ik} < d_{ij}} m_k, \tag{2}$$

with d_{ij} the distance between two locations i and j.

Equation (1) is accurate only in the limit of large numbers of population units, while for finite systems it needs to be normalized by multiplying the right-hand side by $\frac{1}{1-\frac{m_i}{N}}$ [4]. Since we investigate countries of different sizes and the normalization can have a significant effect in some of the smaller countries, we will be using the normalized version

$$M_{ij} = M_i \frac{1}{1 - \frac{m_i}{N}} \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}. \tag{3}$$

B. Data

For our research, we will be using internal migration data from four different countries to test and evaluate our model approach. We choose USA, Mexico, Argentina, and Peru for our investigation. Most tasks will be performed using the internal migration data from the USA because it is obtained through tax return data rather than microcensus data and therefore offers the highest accuracy. Furthermore, the data offers demographic and geographic patterns that are useful for some of our investigations.

1. USA

The data on internal migration flows in the USA are obtained from the Internal Revenue Service that were recorded between 2007 and 2008 [12]. The migration estimates are obtained by evaluating the mailing addresses provided in the tax returns. A changing address, compared to the previous year, indicates that a person moved. Disadvantages of this data set include that people who are not required to file tax returns, like low-income and older people, are excluded. Additionally, tax returns submitted after September are excluded as well. These returns most often belong to high-income persons. Additional data include county and state borders, distances in between counties [13,14], and county population data [15]. In total, we consider migration between 3140 countries in 48 states (excluding Alaska and Hawaii) and the District of Columbia.

2. Argentina, Mexico, and Peru

Internal migration data for Argentina, Peru, and Mexico are obtained from the IPUMS International database [16]. IPUMS International provides microcensus data on bilateral migration flows, obtained through the survey data on a subsample of the total population. In addition to the survey data which only include a fraction of the total population, they include a set of weights which represent the number of persons in the population represented by one entry in the data sample. For Argentina and Peru, IPUMS provides the same weight for each data point. These flat weights are created by divided the sample size of the microcensus by the total population. In total, the microcensus data for both these countries include 1% of the total population. In Mexico the census bureau surveyed 10% of the population, which has been chosen to represent different geographies, population sizes, and living situations. Contrary to Argentina and Peru, the weights are not flat but are heterogeneous so that the sample matches, for example, the entirety of rural and urban areas. A more in depth description of this data set can be found in [17]. Using both these data sets, microcensus and weights, we are able to create bilateral migration flow estimates for all three countries.

The population stocks for Argentina, Peru, and Mexico originate from the national census, more specifically from the National Institute of Statistics and Geography for Mexico [18], the National Institute of Statistics and Censuses for Argentina [19], and the Instituto Nacional de Estadistica e Informatica for Peru [20]. To obtain geographical data for Mexico we use the borders of each municipality [21] to calculate its center. The centers are then used to calculate the distances between individual municipalities. The data for Mexico include 2448 municipalities within 31 states. The geographical data for Argenitna and Peru are obtained from Instituto Geografico Nacional (through the Humanitarian Data Exchange Website) [22]. For Argentina we only consider the departments, not municipalities, resulting in 24 departments. For Peru the data set consists of 196 provinces.

III. RESULTS

In the following we will first show a discontinuity we found in the original RM, followed by our modification, the introduction of an angle dependence in the s_{ij} term. Furthermore, we investigate whether the data indicate directional preferences and test the overall performance of our approach.

A. Discontinuity

The RM implies a discontinuity related to the way the intervening population s_{ij} is calculated. To show the discontinuity, we construct the following scenario. Consider a migration flow between locations i and j and between i and k. The locations j and k have a similar distance from i, with k being marginally closer than j, $\overline{ij} = \overline{ik} + \delta$ (δ being some short distance), so that $s_{ij} = s_{ik} + m_k$. A visualization of this scenario

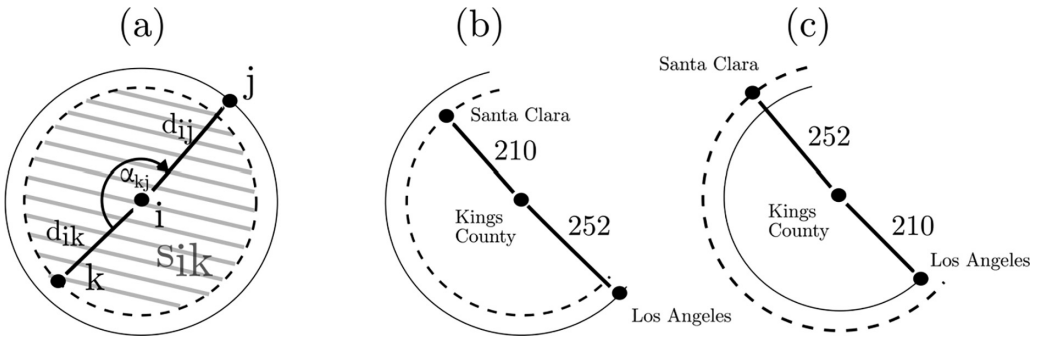

FIG. 1. (a) Migration scenario, with origin i and destinations k and j. Here d_{ik} and d_{ij} denote the respective distances, s_{ik} denotes the intervening opportunities for a migration flow between i and k, and α_{kj} is the angle between k and j. (b) and (c) Two migration scenarios originating from Kings County with the destination Los Angeles or Santa Clara. Each partial circle represents the radius indicating which population counts in the intervening opportunities and which do not. (b) The distance between KC and SC is 210 km and between KC and LA it is 252 km, making SC part of LA's intervening opportunities. (c) Swapped distances. The distance between KC and SC is now 252 km and that between KC and LA is 210 km, making LA part of SC's intervening opportunities.

is shown in Fig. 1(a). The flow rates r_{ij} ($M_{ij} = M_i \frac{1}{1-\frac{m_i}{N}} r_{ij}$) and r_{ik} can then be expressed as

$$r_{ik} = \frac{m_i m_k}{(m_i + s_{ik})(m_i + m_j + s_{ik})},$$
$$r_{ij} = \frac{m_i m_j}{(m_i + \underbrace{s_{ik} + m_k}_{s_{ij}})(m_i + m_j + \underbrace{s_{ik} + m_k}_{s_{ij}})}. \quad (4)$$

If we now assume that $m_k \approx m_j$,

$$r_{ik} = \frac{m_i m_k}{(m_i + s_{ik})(m_i + m_k + s_{ik})},$$
$$r_{ij} \approx \frac{m_i m_k}{(m_i + s_{ik} + m_k)(m_i + s_{ik} + 2m_k)}, \quad (5)$$

and compare both flow probabilities, we obtain

$$\frac{r_{ik}}{r_{ij}} \approx \frac{m_i + s_{ik} + 2m_k}{m_i + s_{ik}} = 1 + 2\frac{m_k}{m_i + s_{ik}}. \quad (6)$$

Considering the final expression, one can see that the predicted flow rates differ by a factor $1 + 2\frac{m_k}{m_i + s_{ik}}$. This factor is close to 1 when the destinations are much less populous than the origin and/or the surrounding region. However, in the case of large destinations and relatively sparse populated surroundings, the factor can be substantial: For example, assuming that all three locations have the same population size and that the circle around i of radius $\overline{ij}$ is empty except for k, the model predicts three times more migrants from i to k than from i to j, even if j and k are at almost the same distance from i.

To illustrate this with a real world example, we calculate the migration flows from Kings County (KC), California, to Los Angeles (LA), California and Santa Clara (SC), California. We choose the destination counties because their population sizes are large compared to their surrounding (small s_{ij}) and their origin and because KC has a similar distance to both LA (252 km) and SC (210 km). We predict the respective flow rates, first using the actual distances and second swapping the distances. The purpose of this exercise is to determine the impact of LA counting in the intervening population for migration from KC to SC [Fig. 1(b)] compared to not having it count in the population [Fig. 1(c)] and conversely the effect of SC counting or not counting in the intervening population for migration from KC to LA.

The results show that moving each of the cities only by a few kilometers can have a significant impact on the predicted migration rates. The migration rate from KC to SC differs by almost a factor of 10 and the migration rate from KC to LA differs by a factor of 2 between the two scenarios (Table I, column "RM").

B. Angle-dependent radiation model approach

We argue that a model predicting substantially different flow rates due to incrementally small differences in distance is potentially implausible. There are cases when it may be plausible. First, if j and k are located in almost the same direction from i, i.e., k is located immediately in front of j, then it is plausible that k intercepts many of the migrants who could potentially also move to j. Given that everything else is equal, migrants would tend to choose the closer destination even if the difference in distance is small. Second, the same argument could apply if j and k are located in different directions from i but the distances $\overline{ij}$ and $\overline{ik}$ are short, because in such a case the move to either destination would not displace the migrant

TABLE I. Impact of manipulated distances on the migration flows obtained through the RM and angle-dependent radiation model approach. The term "swapped distance" indicates that the distances of Santa Clara (SC) and Los Angeles (LA) to Kings County (KC) are swapped so that SC is slightly closer to KC than to LA. The column labeled "Census" indicates the observed migration flows, RM the values estimated by the original RM, and ADRM the estimates from the angle-dependent radiation model approach.

	Number of migrants		
Origin and destination	Census	RM	ADRM
KC → LA (normal distance)	277	68	137
KC → LA (swapped distance)	277	125	151
KC → SC (normal distance)	88	50	100
KC → SC (swapped distance)	88	6	23

from their larger region of residence and the direction could thus be considered unimportant for the decision to move. (This argument should apply even more if the move is not permanent but temporary, such as commuting, supporting the applicability of the RM for commuting problems.)

However, in the case of migration over longer distances and when potential destinations are in different directions, it appears implausible that marginal differences in distance should influence migration decisions to an extent that would justify substantially more migration to one destination than to another, as shown above. In other words, the isotropy of s_{ij} in the RM implies that migrants evaluate all potential destinations irrespective of their relative direction from the origin, and thus of the distance between each other; and give clear preference to a destination that is marginally more attractive than the next most attractive potential destination, even if the two are many hundreds or thousands of miles apart.

To solve this issue, we introduce a modification to calculating s_{ij}. Instead of considering the whole population within the circle equally, as in Eq. (2), we weight the population depending on the angle α_{kj} between the opportunity cell k and the destination under consideration j,

$$s_{ij} = \sum_{k \forall d_{ik} < d_{ij}} m_k \frac{b + \cos \alpha_{kj}}{b+1}, \tag{7}$$

with $b \geqslant 1$. This modification yields the original value of s_{ij} when k is in a direct line between i and j and a fraction $(b-1)/(b+1)$ of the original value when it is in the opposite direction. In this general form, the parameter b controls how strongly the influence of k declines with increasing angle. We will show below that the intuitive choice of $b = 1$,

$$s_{ij} = \sum_{k \forall d_{ik} < d_{ij}} \frac{m_k}{2}(1 + \cos \alpha_{kj}), \tag{8}$$

where the contribution of k to s_{ij} becomes zero at $\alpha_{kj} = \pi$, is also supported by the data.

It should be pointed out that the original RM is normalized such that the total number of migrants in a country is conserved:

$$\sum_{i,j} M_{ij}^{\text{original}} = N_M. \tag{9}$$

However, Eq. (7) implies that $\sum_{ij} M_{ij}^{\text{angle}} > N_M$, since $s_{ij}^{\text{angle}} < s_{ij}^{\text{original}}$. Therefore, in our angle-dependent radiation model (ADRM), we include an additional normalization

$$M_{ij} = \tilde{M}_{ij} \frac{N^M}{\sum_{kl} \tilde{M}_{kl}}, \tag{10}$$

with

$$\tilde{M}_{ij} = m_i \frac{1}{1 - \frac{m_i}{N}} \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}. \tag{11}$$

C. Direction dependence in migration data

To further motivate our approach, we provide empirical evidence of a direction dependence in internal migration patterns in the USA, using two different methods. First, we investigate the intervening opportunities by considering three examples of long-range (approximately 1000 km) migration, originating from Salt Lake City, Kansas City, and Minneapolis. We choose these cities because all of them have relatively heterogeneous surroundings: For Kansas City and Minneapolis, the western part of the surroundings is less densely populated than the eastern part and for Salt Lake City the coastal area, including, for example Los Angeles and San Francisco, has higher populations than the eastern surroundings (Fig. 2, top row). We identify potential destinations within a circle around the origin.[1] These locations, having a similar distance to the origin, all yield similar values of s_{ij} in the original RM (Fig. 2 second row).

We now rearrange Eq. (1) to calculate the s_{ij} that would be necessary for the original RM to perfectly match the migration flows given by the census data:

$$s_{ij} = -\frac{2m_i + m_j}{2} + \sqrt{\left(\frac{2m_i m_j}{2}\right)^2 + m_i^2 + m_i m_j - \frac{m_i^2 m_j}{M_{ij}^{\text{census}}}\left(1 - \frac{m_i}{N}\right)}. \tag{12}$$

The result is far from isotropic; higher hypothetical numbers of intervening opportunities arise in more populous parts of the ring than in less populous parts (Fig. 2, third row), suggesting that if populous areas lie near the connecting line between origin and destination, the number of migrants will decrease significantly more than if these high population areas would lie in the opposite direction of the destination.

In Fig. 2 (bottom row) we display the values of s_{ij} obtained through our ADRM. These match the spatial patterns of the hypothetical "optimal" s_{ij} in all three cases, providing illustrative support for an angle-dependent model.

In addition to the results shown before, we construct a second experiment to test for direction dependence in intervening opportunities. We select, from the US migration data, large samples of county triplets including an origin county i and two destination counties j and k with $0.6\, d_{ij} < d_{ik} < 0.9\, d_{ij}$, i.e., k is somewhat closer to i than j (and additionally does not belong to the same metropolitan area as j, which could happen if d_{ik} and d_{ij} were chosen too close to each other). All j and k have population sizes above 1×10^6. We measure the ratio of observed migration rates M_{ij}/M_{ik} dependent on the angle between j and k, α_{kj}. If the hypothesis of an angle dependence is correct, i.e., if intervening opportunities j intercept more migrants that could potentially move to k if j is located in approximately the same direction as k rather than in the opposite direction, then we expect to find a negative relation between M_{ij}/M_{ik} and α_{kj}.

This negative relation is indeed visible in the data, for both small and large origin counties (Fig. 3). It emerges despite a large spread owing to the idiosyncrasies of individual country triplets (population sizes and distances are not identical across

[1]The radius and width of the ring are chosen slightly different between the cities, to account for different county sizes as well as geographic boundaries (coasts and borders). Rings are partly curtailed by the Californian coastline, Canadian border, and Great Lakes.

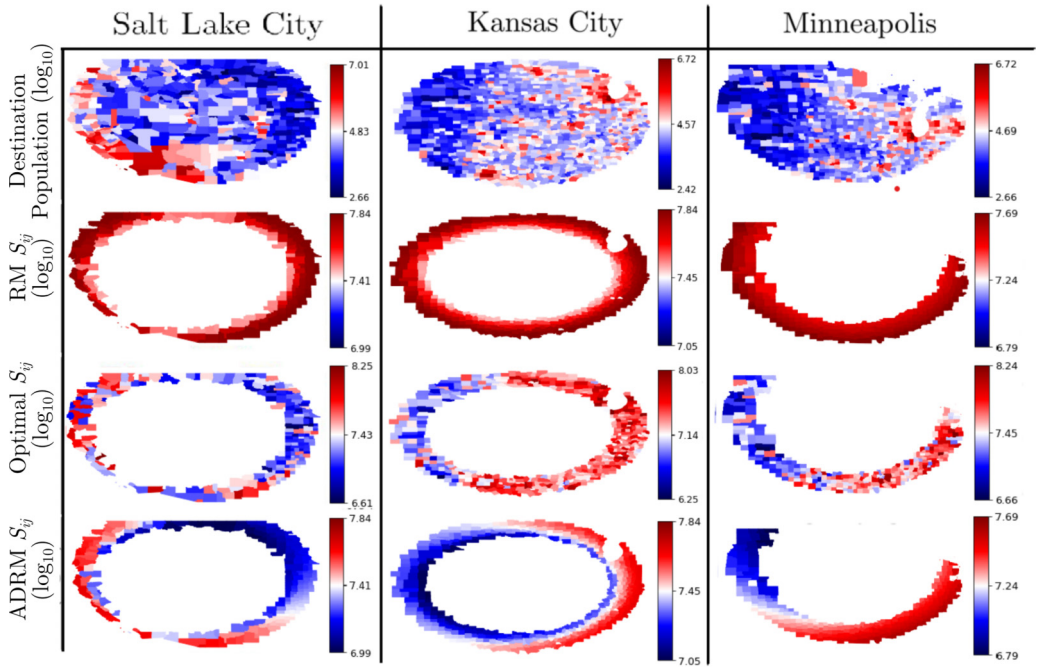

FIG. 2. Population count and s_{ij} for counties within a circle (excluding areas outside the USA) around Salt Lake City (the radius is 1100 km and the ring width 240 km) (left column), Kansas City (radius of 875 km and ring width of 210 km) (middle column), and Minneapolis (radius of 825 km and ring width of 210 km) (right column). The top row displays the destination population, the second row the s_{ij} of the original RM for each county in the outer circle being treated as a potential destination, the third row the optimal s_{ij}, calculated with Eq. (12), and the bottom row the s_{ij} of the ADRM [Eq. (7)].

the sample, but only satisfy the conditions mentioned above). In addition, a linear fit indicates that the median value of M_{ij}/M_{ik} approaches 1 at $\alpha_{kj} = \pi$. This is consistent with a value of 1 for the parameter b in Eq. (7): If $b = 1$, then the contribution of j to s_{ik} is zero at $\alpha_{kj} = \pi$ and in a case where $m_k \approx m_j$ and $d_{ik} \approx d_{ij}$, this implies equal migration rates M_{ij} and M_{ik} according to Eq. (1) [or Eq. (11)]. Figure 3 thus not only indicates the presence of a direction dependence of intervening opportunities in migration data, but also supports a choice of $b = 1$, i.e., the influence of j on migration rates from i to k is negligible if j and k are located in opposite directions.

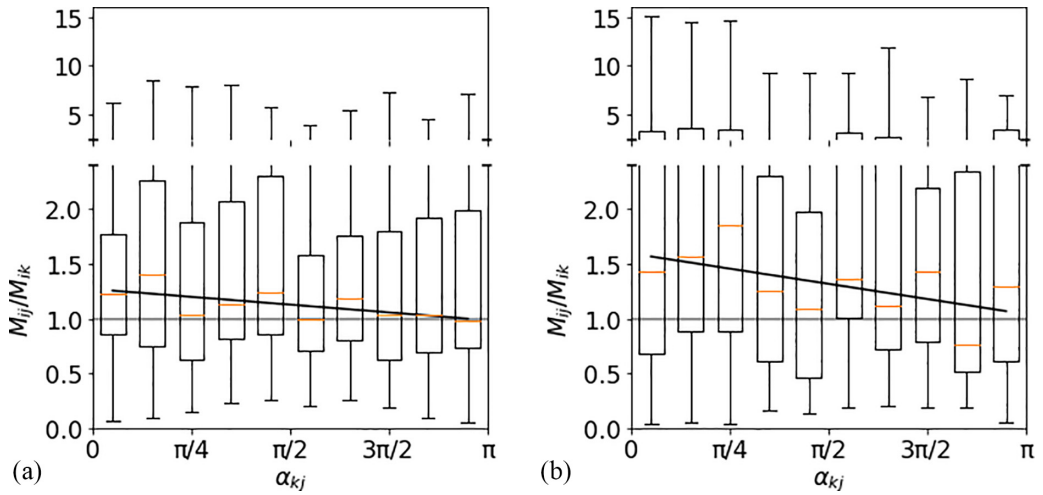

FIG. 3. Migration rate ratios for city triplets, consisting of one origin i and two destinations k and j in relation to the angle α_{kj}. The distances fulfill the condition $0.6d_{ij} < d_{ik} < 0.9d_{ij}$ as well as (a) the origin population less than or equal to 170 000, destination population greater than or equal to 1 000 000, and distance greater than 100 and less than 1300 and (b) origin population greater than or equal to 1 000 000, destination population greater than or equal to 1 000 000, and distance greater than 100 and less than 2000. The sample size is (a) 1094 and (b) 1140. Each panel shows the data binned into multiples of $\pi/10$. Boxes indicate the 25th and 75th percentiles, the whiskers indicate the 5th and 95th percentiles, the orange lines indicate the median values for each bin, black lines indicate the linear fit through the median values, and gray lines indicate a ratio of one.

TABLE II. The R^2 and Sørensen-Dice (SD) coefficients for the application of the RM and the ADRM on internal migration data sets in different countries.

Country	RM R^2	ADRM R^2	RM SD	ADRM SD
USA	−0.001	0.545	0.517	0.554
Mexico	−0.263	0.069	0.276	0.311
Argentina	0.367	0.410	0.508	0.533
Peru	0.135	0.333	0.380	0.417

D. Model evaluation

Introducing directional preferences into the radiation model, as described above and supported by empirical findings, mitigates the discontinuity discussed in Sec. III A. While incremental changes in distance between two alternative and equal destinations still induce a jump in migration rates, this jump is plausibly larger when the two potential destinations are in about the same direction from the origin and becomes small when they are in different directions. In the example of equally sized i, j, and k and $s_{ij} = m_k$, the angle-dependent model (with $b = 1$) predicts three times more migrants to k than to j only if k and j are in the same direction from i and it predicts equal rates of migration to both j and k if they are in opposite directions. Repeating the counterfactual calculation for the Kings County–Santa Clara–Los Angeles triangle using the angle-dependent approach yields differences of a factor 4 (instead of 10) for Santa Clara and 1.25 (instead of 2) for Los Angeles (compare columns "RM" and "ADRM" in Table I).

As a final step, we evaluate the overall performance of the ADRM [consisting of Eqs. (8), (10), and (11), i.e., $b = 1$] in reproducing observed migration flows in four different countries and compare it with the performance of the original RM. We also test the influence of the choice of b on the model performance. For this evaluation we use two measures, R^2 and the Sørensen-Dice coefficient.

The R^2 value is given by

$$R^2 = 1 - \frac{\sum_{ij} \left(M_{ij}^{\text{census}} - M_{ij}^{\text{model}}\right)^2}{\sum_{ij} \left(M_{ij}^{\text{census}} - \bar{M}^{\text{census}}\right)^2}, \tag{13}$$

with $\bar{M}^{\text{census}}$ the mean of all census flows. The Sørensen-Dice coefficient is given by [23,24]

$$E^{\text{Sørensen}} = \frac{2 \sum_{i,j} \min\left(M_{ij}^{\text{model}}, M_{ij}^{\text{census}}\right)}{\sum_{i,j} M_{ij}^{\text{census}} + \sum_{i,j} M_{ij}^{\text{model}}}. \tag{14}$$

Here $E^{\text{Sørensen}}$ can be interpreted as a similarity measure between simulations and observations. Zero indicates a total mismatch whereas $E^{\text{Sørensen}} = 1$ indicates a perfect match. A

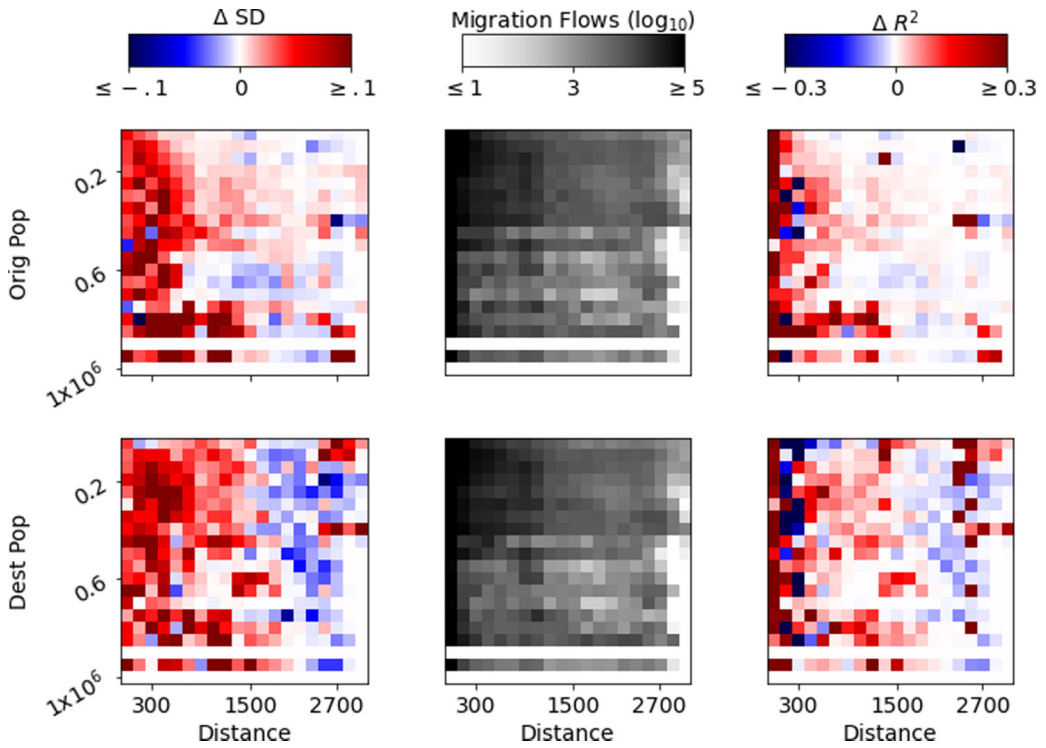

FIG. 4. Performance of the ADRM and RM tested on US internal migration data. The left column shows the difference between the Sørensen-Dice coefficient for the ADRM and RM depending on the origin population and distance (top) and destination population and distance (bottom). The middle column shows the number of migrants depending on the origin population and distance (top) and destination population and distance (bottom). The right column shows the difference between the R^2 score for the ADRM and RM depending on the origin population and distance (top) and destination population and distance (bottom). Each square indicates the difference between the performance measure of the original RM and the ADRM, red indicates that the ADRM performs better and blue that the RM performs better. The middle column is shown to indicate which population and distance regimes are most frequented. White lines in the plot indicate no sample data for these population and distance combinations.

TABLE III. The R^2 and Sørensen-Dice coefficients for different parameters b of the general approach.

	USA		Mexico		Argentina		Peru	
b	R^2	SD	R^2	SD	R^2	SD	R^2	SD
1	0.545	0.554	0.069	0.311	0.410	0.533	0.333	0.417
1.5	0.455	0.552	0.000	0.303	0.408	0.528	0.297	0.410
2	0.385	0.547	−0.045	0.298	0.404	0.524	0.271	0.405
3	0.293	0.540	−0.101	0.291	0.397	0.520	0.239	0.398
5	0.197	0.533	−0.157	0.286	0.388	0.388	0.205	0.380
10	0.108	0.526	−0.205	0.281	0.379	0.312	0.174	0.380

similarity measure is more useful than, e.g., a measure of correlation, since we are interested in the variation as well as in the magnitude of the flows. The Sørensen-Dice coefficient is comparable to some other similarity measures such as the Jaccard index [25]. It has been shown to have a high sensitivity even for heterogeneous data and to be relatively unaffected by outliers [26].

The ADRM improves both performance measures at the national scale for all four country data sets, compared to the original RM (Table II). It also performs best for $b = 1$, compared to higher values of b, in all four countries (Table III), which confirms this intuitive choice for b and the validity of the proposed model without fitting parameters [Eqs. (8), (10), and (11)]. To further investigate the performance of both the original RM and the ADRM, we calculate the difference between both performance measures for the US (Fig. 4) and Mexico (Fig. 5) dependent on traveled distance, origin, and destination population. In the Appendix we include the absolute performance measures for both models as well. Furthermore, we include the number of census flows depending on distance, origin, and destination population in the middle column of these plots (Figs. 4 and 5).

First, we will be discussing the performance in the US (Fig. 4). The performance is mostly independent of the origin and destination population but changes significantly with distance. For the origin population– and distance-dependent plots (top row), the results indicate that the ADRM improves the R^2 and Sørensen-Dice measures for short and intermediate distances. Furthermore, the ADRM yields slightly better results for larger populations as well. Considering the difference in Sørensen-Dice coefficients with respect to destination population and distance, the results indicate that the ADRM performs significantly better for distances below 1500 km and that the RM performs better for distances above 2000 km. Looking at the difference in R^2, we see mixed results for short distances (less than 300 km) and otherwise a superior ADRM for distances between 300 and 1500 km and a superior RM for distances above 2000 km.

Considering the absolute values of both measures, the ADRM outperforms the RM considering almost all

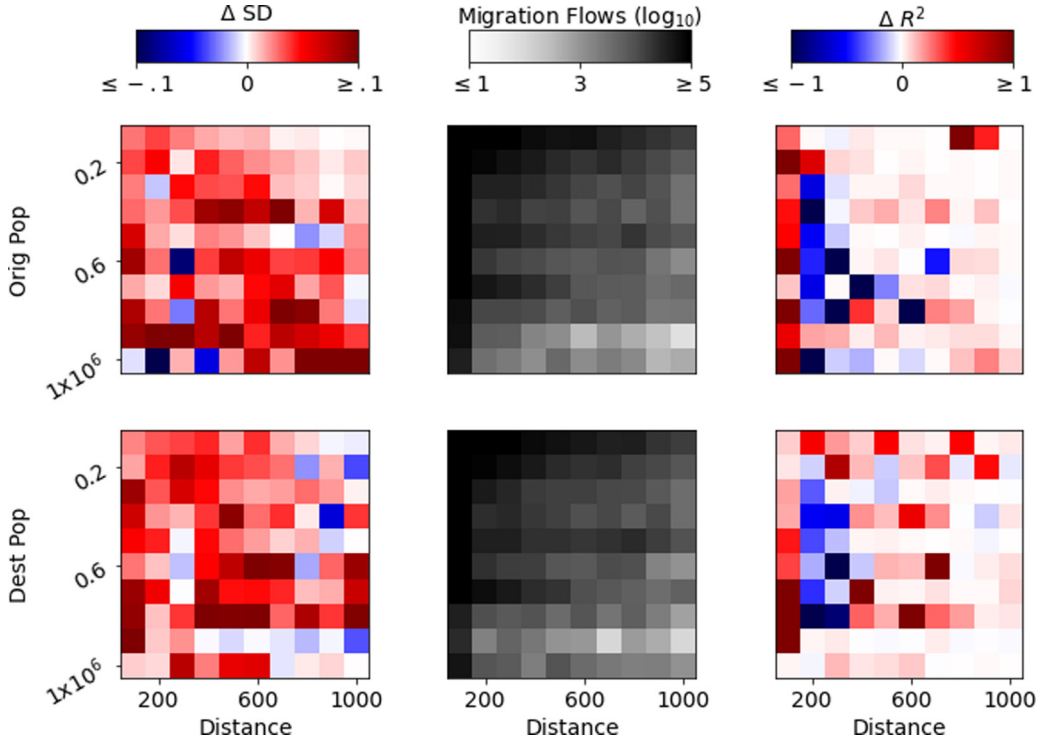

FIG. 5. Performance of the ADRM and RM tested on Mexican internal migration data. The left column shows the difference between the Sørensen-Dice coefficient for the ADRM and RM depending on the origin population and distance (top) and destination population and distance (bottom). The middle column shows the number of migrants depending on the origin population and distance (top) and destination population and distance (bottom). The right column shows the difference between the R^2 score for the ADRM and RM depending on the origin population and distance (top) and destination population and distance (bottom). Each square indicates the difference between the performance measure of the original RM and the ADRM, red indicates that the ADRM performs better and blue that the RM performs better. The middle column is shown to indicate which population and distance regimes are most frequented.

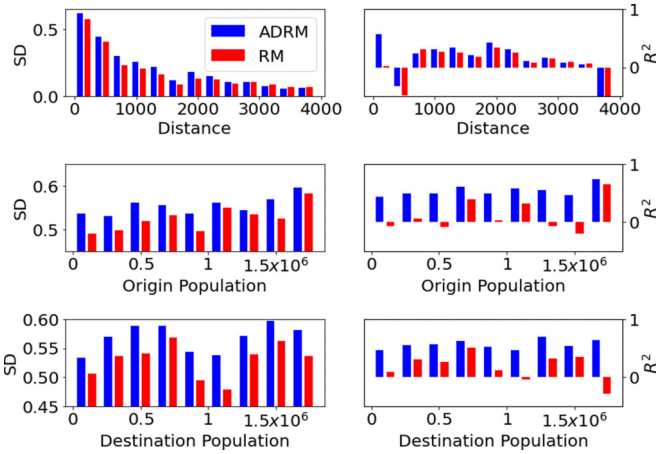

FIG. 6. Performance of the original radiation model and ADRM tested on US internal migration data, dependent on traveled distance (top), origin population size (middle), and destination population size (bottom). The left column shows the Sørensen-Dice (SD) coefficient and the right column the R^2 score. Each bar denotes the performance value for an interval of 300 km or 200 000 persons, respectively. The R^2 is clipped between -0.5 and 1.

investigated dependences (see the Appendix). In terms of distance, the ADRM shows significant improvements for migration flows within 300 km. For longer distances the ADRM still performs slightly better.

The ADRM also performs better than the original RM across all groups of origin and destination population size (see the Appendix). Considering the R^2 value, we can see that the ADRM shows positive values for all population sizes, whereas the original model shows multiple negative values.

Next we will be discussing the performance in Mexico (Figs. 5–7). Considering the Sørensen-Dice coefficient, we can see that the ADRM performs better than the original

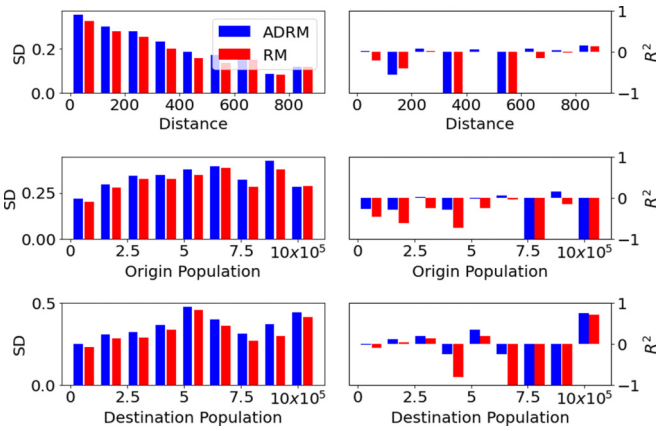

FIG. 7. Performance of the RM and ADRM on Mexican internal migration data. The left column shows the Sørensen-Dice coefficient and the right the R^2 score. The top row shows the performance depending on the traveled distance, the middle row depending on the origin population, and the bottom row depending on the destination population. Each marker denotes the performance value for an interval of 100 km or 120 000 persons, respectively. The R^2 is clipped between -1 and 1.

model for almost all distances and population sizes. The R^2 values show a slightly different pattern. Even though the ADRM performs better in most cases, the original RM shows good results for comparably small distances (100–300 km). We can find this pattern for most origin and destination population size groups. Interestingly, the ADRM still performs better for migration flows shorter than 100 km.

In terms of absolute performance in Mexico (see the Appendix), the ADRM shows improving Sørensen-Dice measures in all regimes. Considering the R^2 values, the ADRM performs better for most regimes.

IV. CONCLUSION

In summary, we have shown that the radiation model yields implausible (in the case of long-distance, permanent moves) results when alternative destinations are at similar distances but opposite directions from the origin. Furthermore, we have found evidence in migration data for the influence of intervening opportunities to be direction dependent. Evaluating the environment for a specific destination shows that large-population areas have a larger impact on the migration if they are located in between the origin and destination compared to when they have the same distance to the origin but are positioned in the opposite direction of the destination. In other words, when people make the decision to migrate, they may not consider every possible direction equally attractive.

We have proposed a modification of the RM that accounts for such a directional dependence in the way the intervening opportunities, i.e., the s_{ij}, are calculated. Using internal migration data for several countries, we have shown that this angle-dependent radiation model is capable of capturing the heterogeneous intervening opportunity patterns obtained from the data better than the original RM. Moreover, it mitigates potentially implausible differences in migration rates between destinations at similar distance but opposite direction that arise in the original RM. Finally, the ADRM matches observed migration data in four countries significantly better than the RM, as measured by R^2 scores and Sørensen-Dice coefficients. The fact that the best performance is always obtained when intervening opportunities at 180° are weighted down to zero, rather than some more moderate weighting factor, underlines that the direction dependence is indeed significant and allows us to abstract, from the more general model with a weighting parameter, a model without an explicit fitting parameter that works well in all the studied countries and may thus be of use in other locations too even when no calibration data are available.

We should note that we have only considered internal migration, as opposed to other forms of mobility, and the ADRM should not be considered another generalization of the RM that will be of use across all spatiotemporal scales. Nevertheless, the direction dependence in the concept of intervening opportunities may be relevant in other contexts too, for instance, in intraurban moves between different neighborhoods. Our interest in migration is related to the fact that gravity-type models are still the predominant approach in migration modeling, despite their conceptual and practical shortcomings [1,27], while radiation-type models have found less application in studies of migration, compared to, e.g., commuter

mobility. Improving the applicability of radiation-type models to migration data while maintaining their conceptual rigor and simplicity may help open new methodological avenues for migration research.

Considering the applications of this model version, we think it can be helpful in estimation scenarios without any calibration data, following the idea of the original radiation model. Furthermore, the general finding that when considering intervening opportunities their importance for the migrant differs depending on their relative location to the origin and destination might be of importance for other models that use approaches accounting for intervening opportunities or any kind of intercepting amenities.

ACKNOWLEDGMENT

The work was supported within the framework of the European Union Horizon 2020 programme, Grant Agreements No. 870649 (FUME) and No. 869395 (HABITABLE).

APPENDIX: RADIATION MODEL

In this Appendix we briefly summarize the derivation of radiation as it was provided by Simini *et al.* [1]. The derivation given here should only point out the origin and some motivation for the model and might deviate from the standard mathematical notation.

The starting point for the derivation originates from a particle diffusion process. Each migrant can be though of as a particle, released in some origin i, and the origin is associated with some level of amenities or benefits. These origin amenities can be thought of as an absorption threshold. The individual aims for an increase of amenities above the origin level but at the same time minimizing the migration distance and picking the closest destination that provides better benefits. The amenities are assumed to scale with the population of a certain area.

Mathematically, we start the derivation by introducing a random variable Z representing amenities. The underlying distribution for this variable is given by a probability density function $p(z)$. Next we formulate the probability of one person moving from i to j,

$$P(1|m_i, m_j, s_{ij}) = \int_0^\infty dz\, \mathrm{Prob}_{m_i}(Z = z)\mathrm{Prob}_{s_{ij}}(Z \leqslant z)\mathrm{Prob}_{m_j}(Z > z), \tag{A1}$$

where m_i and m_j are the origin and destination populations and s_{ij} describes the intervening population between i and j and can be calculated as the sum of people who live closer to i than j is to i. The expression consists of three parts: first, the origin term $\mathrm{Prob}_{m_i}(Z = z)$, expressing the probability that the maximum value extracted from $p(z)$ after m_i trials is equal to z; second, the surroundings term $\mathrm{Prob}_{s_{ij}}(Z \leqslant z)$, describing the probability that after s_{ij} trials all values extracted from $p(z)$ are smaller than or equal to z; and finally the destination term $\mathrm{Prob}_{m_j}(Z > z)$, representing the probability that after m_j trials at least one value extracted from $p(z)$ is larger than z. In the end, we integrate over all possible amenity values z.

To obtain the form of the radiation model used in the main text, we need to rewrite these three probabilities. First, we rewrite the surroundings term

$$\mathrm{Prob}_{s_{ij}}(Z \leqslant z) = \mathrm{Prob}(Z \leqslant z)^{s_{ij}} = \left(\int_0^z p(\tilde{z})d\tilde{z}\right)^{s_{sij}} \tag{A2}$$

and the destination term

$$\mathrm{Prob}_{m_j}(Z > z) = 1 - \mathrm{Prob}(Z \leqslant z)^{m_j}. \tag{A3}$$

Finally, we can express the probability density function $\mathrm{Prob}_{m_i}(Z = z)$ as the derivative of its cumulative distribution function

$$\mathrm{Prob}_{m_i}(Z = z) = \frac{d}{dz}\mathrm{Prob}_{m_i}(Z \leqslant z) = \frac{d}{dz}\mathrm{Prob}(Z \leqslant z)^{m_i} \tag{A4}$$

$$= m_i \mathrm{Prob}(Z \leqslant z)^{m_i - 1}\frac{d}{dz}\mathrm{Prob}(Z \leqslant z). \tag{A5}$$

Inserting all these expression into Eq. (A1) yields

$$P(1|m_i, m_j, s_{ij}) = \int_0^\infty dz \left(m_i \mathrm{Prob}(Z \leqslant z)^{m_i - 1}\frac{d}{dz}\mathrm{Prob}(Z \leqslant z)\right)\mathrm{Prob}(Z \leqslant z)^{s_{ij}}[1 - \mathrm{Prob}(Z \leqslant z)^{m_j}] \tag{A6}$$

$$= m_i \int_0^\infty dz \left(\frac{d}{dz}\mathrm{Prob}(Z \leqslant z)\right)[\mathrm{Prob}(Z \leqslant z)^{m_i + s_{ij} - 1} - \mathrm{Prob}(Z \leqslant z)^{m_i + m_j + s_{ij} - 1}] \tag{A7}$$

$$= m_i \left(\frac{1}{m_i + s_{ij}} - \frac{1}{m_i + m_j + s_{ij}}\right). \tag{A8}$$

The total number of migrants moving from origin i to destination j can be expressed as

$$M_{ij} = M_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{A9}$$

where M_i is the total number of migrants leaving the origin i.

Note that the notation, e.g., for $\text{Prob}_n(Z = z)$ for the probability of obtaining the maximum value z after n trials is unusual but is inspired by the derivations provided in [1,9,28,29].

[1] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, A universal model for mobility and migration patterns, Nature (London) **484**, 96 (2012).

[2] S. A. Stouffer, Intervening opportunities: A theory relating mobility and distance, Am. Sociol. Rev. **5**, 845 (1940).

[3] G. K. Zipf, The P_1P_2/D hypothesis: On the intercity movement of persons, Am. Sociol. Rev. **11**, 677 (1946).

[4] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows, Phys. Rev. E **88**, 022812 (2013).

[5] X. Liang, J. Zhao, L. Dong, and K. Xu, Unraveling the origin of exponential law in intra-urban human mobility, Sci. Rep. **3**, 2983 (2013).

[6] A. J. Garcia, D. K. Pindolia, K. K. Lopiano, and A. J. Tatem, Modeling internal migration flows in sub-Saharan Africa using census microdata, Migrat. Stud. **3**, 89 (2015).

[7] M. Tizzoni, P. Bajardi, A. Decuyper, G. K. Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, On the use of human mobility proxies for modeling epidemics, PLoS Comput. Biol. **10**, e1003716 (2014).

[8] L. Kluge and J. Schewe, Evaluation and extension of the radiation model for internal migration, Phys. Rev. E **104**, 054311 (2021).

[9] E.-J. Liu and X. Y. Yan, A universal opportunity model for human mobility, Sci. Rep. **10**, 4657 (2020).

[10] C. Kang, Y. Liu, D. Guo, and K. Qin, A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint, PLoS One **10**, e0143500 (2015).

[11] E. Liu and X. Yan, New parameter-free mobility model: Opportunity priority selection model, Physica A **526**, 121023 (2019).

[12] Internal Revenue Services (IRS) Bilateral Count to Count Migration Flows, https://www.irs.gov/pub/irs-soi/county0708.zip, accessed 21 September 2020.

[13] County Distance Database, National Bureau of Economic Research, https://data.nber.org/data/county-distance-database.html, accessed 19 April 2020.

[14] USA shape file, https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html, accessed 21 September 2020.

[15] Census US Intercensal County Population Data, National Bureau of Economic Research, https://data.nber.org/data/census-intercensal-county-population.html, accessed 19 April 2020.

[16] IPUMS International: Migration Data, Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D020.V7.2, accessed 2 June 2020.

[17] B. Jones, F. Riosmena, D. H. Simon, and D. Balk, Estimating internal migration in contemporary Mexico and its relevance in gridded population distributions, Data **4**, 35 (2019).

[18] Mexico Population Census, http://en.www.inegi.org.mx/temas/estructura/, accessed 5 June 2020.

[19] National Institute of Statistics and Censuses (INDEC), obtained through City Population, https://www.citypopulation.de/en/, accessed 19 May 2021.

[20] Instituto Nacional de Estadística e Informática, Peru, obtained through City Population, https://www.citypopulation.de/en/, accessed 19 May 2021.

[21] Mexico shape file, https://datacatalog.worldbank.org/dataset/mexico-municipalities-2012, accessed 2 June 2020.

[22] Instituto Geografico Nacional, obtained through the Humanitarian Data Exchange, https://data.humdata.org/, accessed 25 May 2021.

[23] T. A. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons, K. Dan. Videnk. Selsk. Biol. Skr. **5**, 1 (1948).

[24] L. R. Dice, Measures of the amount of ecologic association between species, Ecology **26**, 297 (1945).

[25] V. Verma and R. K. Aggarwal, A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: Empirical and theoretical perspective, Soc. Netw. Anal. Min. **10**, 43 (2020).

[26] B. McCune, J. B. Grace, and D. L. Urban, *Analysis of Ecological Communities* (MjM Software Design, Gleneden Beach, 2002), Vol. 28.

[27] R. M. Beyer, J. Schewe, and H. Lotze-Campen, Gravity models do not explain, and cannot predict, international migration dynamics, Human. Soc. Sci. Commun. **9**, 56 (2022).

[28] F. Simini, A. Maritan, and Z. Néda, Human mobility in a continuum approach, PLoS ONE **8**, e60069 (2013).

[29] X.-Y. Jia, E.-J. Liu, C.-Y. Chen, Z. He, and X.-Y. Yan, An interactive city choice model and its application for measuring the intercity interaction, Front. Phys. **10**, 850415 (2022).