Evaluation and extension of the radiation model for internal migration

Lucas Kluge *

Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, D-14412 Potsdam, Germany and Institute of Physics and Astronomy, University of Potsdam, Karl-Liebknecht-Strasse 24/25, 14476 Potsdam, Germany

Jacob Schewe

Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, D-14412 Potsdam, Germany

(Received 16 February 2021; revised 11 June 2021; accepted 11 November 2021; published 29 November 2021)

Human migration is often studied using gravity models. These models, however, have known limitations, including analytic inconsistencies and a dependence on empirical data to calibrate multiple parameters for the region of interest. Overcoming these limitations, the radiation model has been proposed as an alternative, universal approach to predicting different forms of human mobility, but has not been adopted for studying migration. Here we show, using data on within-country migration from the USA and Mexico, that the radiation model systematically underpredicts long-range moves, while the traditional gravity model performs well for large distances. The universal opportunity model, an extension of the radiation model, shows an improved fit of long-range moves compared to the original radiation model, but at the cost of introducing two additional parameters. We propose a more parsimonious extension of the radiation model that introduces a single parameter. We demonstrate that it fits the data over the full distance spectrum and also—unlike the universal opportunity model—preserves the analytical property of the original radiation model of being equivalent to a gravity model in the limit of a uniform population distribution.

DOI: 10.1103/PhysRevE.104.054311

I. INTRODUCTION

Two fundamental approaches to modeling human migration have found widespread application. The *gravity model* estimates flows between two locations depending on their relative distance and their respective population sizes [1]. In the *intervening opportunities model*, on the other hand, the migration flow depends not on the distance but on the population sizes of the locations in between origin and destination [2]. Conceptually, in both cases, the population size of a potential destination is thought of as a proxy indicator of attainable incomes or amenities.

More recently, building on the idea of intervening opportunities, the *radiation model* was developed [3]. In contrast to the gravity model, as well as earlier intervening opportunities models, it can be analytically derived from a simple particle emission-absorption model and is parameter free, making it attractive for data-limited applications.

Both gravity and radiation models have been applied to different problems, ranging from commuting patterns [4,5] to urban and public transport movement [6,7] to the spreading of diseases [8–10]. Several studies have compared the two approaches for commuter flows and other short-term trips; often finding that the gravity model matches the data better but the radiation model performs acceptably given its parameter-free nature [3,11–13]. However, there is no systematic comparison for longer-term, internal migration flows. As gravity models remain the most widespread tools to model migration [14–17], it is of interest to test whether the radiation model offers a viable or even preferable alternative.

In this paper, we compare the performance of four different models for internal (within-country) migration: the gravity model, the radiation model, and two extensions of the radiation model: the universal opportunity model [18], and a more parsimonious extension that we call the damped opportunity model. In the following we briefly introduce the models. We apply them to data sets of internal migration flows in the USA and Mexico and show that the radiation model captures short-range moves well but underestimates longer-distance migration. Both extensions of the radiation model remedy this shortcoming. In addition, the damped opportunity model transforms into a gravity model in the limit of uniform population distribution, like the original radiation model, though with more realistic parameter values, thus reconciling both approaches.

II. MODELS

A. Gravity model

The first model we describe here is the widely used gravity model (GM), so named because of its mathematical similarity with Newton's law of gravity [1,19,20]. One of the most general forms in which it can be expressed is

$$M_{ij} = A\frac{m_i^a m_j^b}{d_{ij}^g}. \tag{1}$$

*kluge@pik-potsdam.de

Here M_{ij} describes the total number of migrants moving from origin i to destination j. The population sizes of the origin and destination sites are given by m_i and m_j, with d_{ij} representing the distance between them, and A, a, b, and g are fitting parameters. Sometimes, the distance dependence is expressed through an exponential decay term, instead of a power law, i.e., $e^{-g d_{ij}}$ in the denominator. The exponential approach has, however, been shown to perform less accurately in many cases [3]. We focus on the power-law approach. The GM model used for our comparison has been introduced in Ref. [21]. It consists of a superposition of two simple GMs as described in Eq. (1). They use a cutoff distance enabling one GM to estimate short-distance migration, whereas the second GM estimates migration over longer distances. Both the USA and Mexico can be divided into several states which then consist of smaller counties or municipalities, respectively. Therefore, we decide to separate long- and short-distance migration by using intrinsic state borders. As a result, we differentiate between inter- and intrastate migration, yielding

$$\tilde{M}_{ij} = A_1 \frac{m_i^{a_1} m_j^{b_1}}{d_{ij}^{g_1}} + A_2 \frac{m_i^{a_2} m_j^{b_2}}{d_{ij}^{g_2}}, \tag{2}$$

where the parameters subscripted with 1 and 2 are associated with intrastate and interstate migration, respectively.

B. Radiation model

The second model we discuss is the radiation model (RM) [3]. It can be derived (see supplementary information of Ref. [3]) from a particle diffusion process, where each migrant is represented as a particle X, emitted in origin i and having an absorption threshold $z_X^{(i)}$. This threshold can be thought of as the maximum level of income or amenities that the migrant may be able to attain when staying in i. When migrating, he or she will choose (will be "absorbed" by) a destination j that offers a higher level of income or amenities. The threshold $z_X^{(i)}$ is defined as the maximum number obtained after m_i random extractions from a distribution $p(z)$, where m_i is the population size of i, and $p(z)$ is an underlying distribution of incomes or amenities whose exact choice is not relevant. Similarly, any hypothetical destination j has a certain probability to absorb the particle X, with the absorbance $z_X^{(j)}$ defined as the maximum of m_j random extractions from $p(z)$.

The probability of a move from i to j to occur can then be expressed as

$$P(1|m_i, m_j, s_{ij}) = \int_0^\infty dz P_{m_i}(z) P_{s_{ij}}(<z) P_{m_j}(>z), \tag{3}$$

where s_{ij} is the total population size within a circle of radius d_{ij} around i, excluding m_i and m_j. $P_{m_i}(z)$ is the probability that the absorption threshold is exactly z, i.e., that the maximum value extracted from $p(z)$ after m_i trials is equal to z:

$$P_{m_i}(z) = \frac{dP_{m_i}(<z)}{dz} = m_i p(<z)^{m_i - 1} \frac{dp(<z)}{dz}. \tag{4}$$

Similarly, the probability that all locations within the radius d_{ij} have absorbances lower than z, i.e., are not acceptable to the migrant, is given by the probability of obtaining a maximum value smaller than z after s_{ij} random extractions:

$$P_{s_{ij}}(<z) = p(<z)^{s_{ij}}. \tag{5}$$

The probability that the next location, j, has an absorbance greater than z is given by the probability that among m_j extractions from $p(z)$ at least one is greater than z:

$$P_{m_j}(>z) = 1 - p(<z)^{m_j}. \tag{6}$$

Thus, Eq. (3) becomes

$$\begin{aligned} P(1|m_i, m_j, s_{ij}) &= m_i \int_0^\infty dz \frac{dp(<z)}{dz} [p(<z)^{m_i + s_{ij} - 1} \\ &\quad - p(<z)^{m_i + m_j + s_{ij} - 1}] \\ &= m_i \left(\frac{1}{m_i + s_{ij}} - \frac{1}{m_i + m_j + s_{ij}} \right). \end{aligned} \tag{7}$$

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

$$M_{ij} = M_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{8}$$

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

It has been shown that Eq. (8) is accurate only in the limit of very large numbers of location units with uniform population distribution, while for real systems a normalization factor $\frac{1}{1 - \frac{m_i}{M}}$ applies, with $M = \sum_i m_i$ [11]. For our application, given that the numbers of counties in the USA and of municipalities in Mexico are relatively large and the population distribution is relatively homogeneous, the normalization factor is small and we neglect it for simplicity.

Compared to the GM, the RM offers several advantages [3]: it is parameter free and, therefore, does not require any calibration; it can be analytically derived from a theoretical framework; and it shows reasonable limits for large destination population sizes. Moreover, a GM with parameters $a + b = 1$ and $g = 4$ can be derived from the RM in the limit of a uniform population distribution, offering a way to reconcile both approaches.

C. Universal opportunity model

The universal opportunity model (UOM) [18] is a generalization of the RM, allowing us to modify the relative weight given to opportunities in the origin and the surrounding:

$$\begin{aligned} M_{ij} &= M_i \int_0^\infty dz P_{m_i + \alpha \cdot s_{ij}}(z) P_{\beta \cdot s_{ij}}(<z) P_{m_j}(>z) \\ &= M_i \frac{(m_i + \alpha s_{ij}) m_j}{[m_i + (\alpha + \beta) s_{ij}][m_i + (\alpha + \beta) s_{ij} + m_j]}, \end{aligned} \tag{9}$$

with $\alpha + \beta \leqslant 1$. Here, the parameter α (termed the "exploratory tendency") can be thought to represent a migrant's level of ambition regarding incomes or amenities in the destination, while β (the "cautious tendency") can be thought to represent the extent to which a migrant evaluates the intervening opportunities s_{ij}, before deciding to move to j. For $\alpha = 0$, $\beta = 1$ the RM is obtained, while for $\alpha > 0$ and $\beta < 1$, more samples are drawn from $p(z)$ at the origin, meaning a larger probability for higher absorption thresholds. At the same time, fewer samples are drawn in the surrounding d_{ij}, meaning a

smaller probability that any location there is found acceptable. This results in larger flows to more distant destinations compared to the radiation model. As shown further below, this model, in its general form, does not simplify into a gravity-type model in the limit of uniform population distribution.

D. Damped opportunity model

In addition to the three existing models, we propose a more parsimonious (compared to the UOM) extension of the radiation model, which we call the damped opportunity model (DOM):

$$M_{ij} = M_i \int_0^\infty dz P_{m_i}(z) P_{s_{ij}^\kappa}(<z) P_{m_j}(>z) = M_i \frac{m_i m_j}{\left(m_i + s_{ij}^\kappa\right)\left(m_i + m_j + s_{ij}^\kappa\right)}. \quad (10)$$

Again, for $\kappa = 1$, the original RM is obtained. A choice of $\kappa < 1$ means that the number of intervening opportunities evaluated by the migrant grows more slowly with distance from the origin than does the population number s_{ij}. An intuitive motivation for such a choice is that complete knowledge about all intervening opportunities becomes harder to obtain the larger s_{ij} grows.

III. DATA AND CALIBRATION

A. Data

1. USA

Data on internal migration flows in the USA between 2007 and 2008 are obtained from the Internal Revenue Service (IRS) [22]. The IRS estimates the number of internal migrants based on the mailing addresses given in the tax returns. If a person's address has changed compared to the previous year, the person is considered a migrant. One of the major disadvantages of this data set is that it only captures movements of people who are required to file income tax returns, leading to an under-representation of poor and older people. Furthermore, tax returns after September of the filling year are excluded. This small percentage of tax returns mostly consists of complex tax returns, reportedly belonging to high-income persons. For comparison of intra- and interstate migration [Fig. 4(a)], data on migration flows between 2005 and 2017 were obtained from the U.S. census [24]. Both data sets include 3140 counties within 50 states (excluding Alaska and Hawaii). Additional data sets include county and state borders [25], county population data [26], and distances in between counties [27].

2. Mexico

Internal migration data for Mexico is obtained from the IPUMS-International database [28]. Since IPUMS only provides microcensus data, we use the method introduced in Ref. [29] to calculate a full set of bilateral migration flows. The population data were taken from the National Institute of Statistics and Geography [23]. Lastly, we used the borders of each municipality [30] to calculate its center. These centers were then used to calculate the distances. The data set includes 2448 municipalities within 31 states.

TABLE I. Gravity model coefficients for the USA and Mexico datasets.

Type	State	A	a	b	g
Inter	All (USA)	5.94×10^{-10}	1.20	1.03	0.75
Intra	All (USA)	4.62×10^{-4}	0.74	0.60	0.66
Inter	Texas	7.74×10^{-6}	1.06	0.94	1.58
Intra	Texas	5.33×10^{-3}	0.64	0.78	1.30
Inter	All (MEX)	2.03×10^{-7}	1.05	0.80	0.72
Intra	All (MEX)	7.45×10^{-4}	0.92	0.40	0.91
Inter	San Luis Potos	2.95×10^{-9}	1.33	0.93	0.90
Intra	San Luis Potos	7.29×10^{-4}	1.45	0.33	3.51

B. Calibration

All RMs are normalized with the corresponding number of migrants obtained from the observed data.

1. USA

For the GM we use all bilateral flows to estimate the fit parameters. All parameters are estimated using a least mean square optimization algorithm and are displayed in Table I. Table II shows the total number of migrants estimated by the GM, compared to the census data. These differ because unlike the RM, the GM explicitly models the number of moves between a given pair of locations. The implications of this are discussed further below.

For the DOM we used $\kappa = 0.92$. For the UOM we chose $\alpha = 0.2$ and $\beta = 0.7$. All parameters were obtained by maximizing the related R^2 value.

2. Mexico

For the GM we use all bilateral flows to estimate the fit parameters. All parameters are estimated using a least mean square optimization algorithm and are displayed in Table I. The total numbers of migrants obtained by the census and estimated by the GM are given in Table II. For the DOM we used $\kappa = 0.75$. For the UOM we chose $\alpha = 0.2$ and $\beta = 0.1$. All parameters were obtained by maximizing the related R^2 value.

IV. RESULTS

A. Data analysis

In this section we evaluate all four models against empirical data of internal migration flows in the USA and Mexico.

We evaluate models with respect to the distribution of migrant flow size over travel distance and over destination

TABLE II. Number of migrants from census and gravity for the USA and Mexico data.

Type	Census	Gravity
All States (USA)	16 020 712	24 810 250
Texas	4 385 540	5 098 056
All States (MEX)	8 515 260	6 181 710
San Luis Potos	103 311	97 029

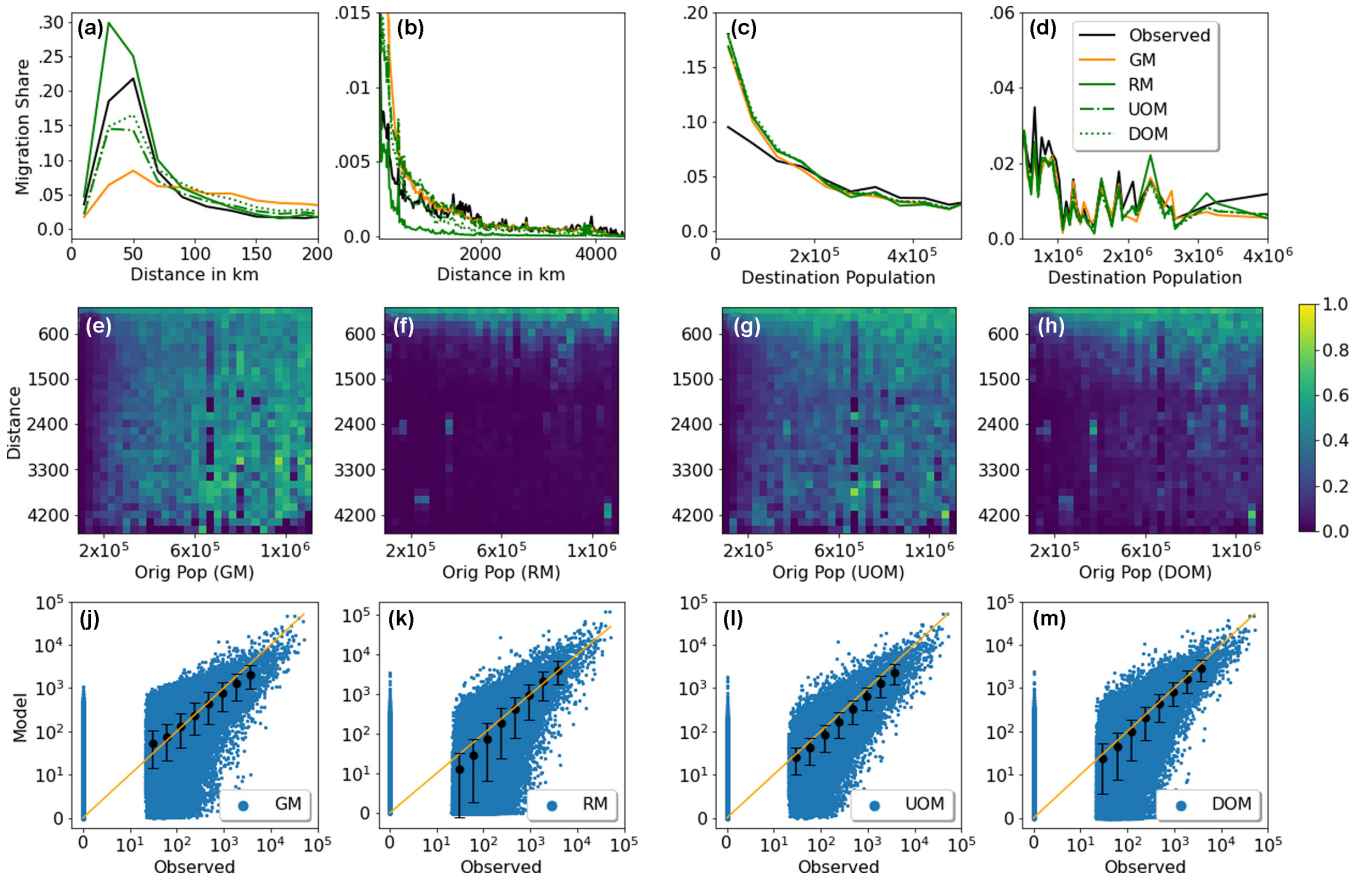

FIG. 1. Observed (tax return based) and model data for internal migration in the USA on a county level [22]. Panels (a) and (b) show the distribution of migrant flows over travel distance (in 20-km bins; y axis indicates share of total migrant flow), for distances below (a) and above 200 km (b). Analogously, panels (c) and (d) display the distribution of migrant flows over destination population size. Panels (e)–(h) display the Sørensen-Dice coefficient for each model as a function of traveled distance (150-km bins) and origin population size (33 333-inhabitant bins). Panels (j)–(m) compare the census and the model value of each individual flow. The orange line is the identity line. Black circles indicate the average, and error bars indicate the 25th and 75th percentile, across a range of different size classes.

population size, and with respect to individual bilateral (county-to-county or municipality-to-municipality) flows. Furthermore, to quantitatively compare the model performance for different data subsets, we calculate the Sørensen-Dice coefficient [31,32] which is given by

$$E^{\text{Sørensen}} = \frac{2\sum_{i,j}\min\left(M_{ij}^{\text{model}}, M_{ij}^{\text{census}}\right)}{\sum_{i,j} M_{ij}^{\text{census}} + \sum_{i,j} M_{ij}^{\text{model}}}. \tag{11}$$

$E^{\text{Sørensen}}$ 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. Since we are interested not only in the variation but also the magnitude of the flows, a similarity measure is more useful than, e.g., a measure of correlation. The Sørensen-Dice coefficient is similar to some other similarity measures such as the Jaccard index [33]. It has been shown to have a high sensitivity even for heterogeneous data and to be relatively unaffected by outliers [34]. In the past, it has been used extensively in ecological research, but has recently also been applied to human mobility [11].

Regarding the traveled distance [Figs. 1 and 2, panels (a) and (b)], the RM fits the data well for intermediate distances (between approximately 50 and 200 km), but slightly overestimates flows at shorter distances and substantially underestimates flows at larger distances. The GM, on the other hand, tends to underestimate short-distance migration and slightly overestimate migration at intermediate distances, but performs comparatively well at predicting longer-distance migration (in particular beyond approximately 500 km).

In relation to the destination population size [Figs. 1 and 2, panels (c) and (d)], both models fit the data relatively well, apart from an overestimation of flows to smaller destination counties or municipalities (approximately 50 000 inhabitants and below) and a slight underestimation of flows to middle-sized destinations (200 000 to 500 000 inhabitants). The Sørensen-Dice coefficients [Figs. 1 and 2, panels (e) and (f)] confirm that the RM yields good results for short and intermediate travel distances, whereas for larger distances (above approximately 500 km), the GM generally matches the data more closely. This difference matters, since observed flows at these larger distances are still substantial [Figs. 3(a) and 3(b)]. Both models have difficulty representing migration from smaller origins.

Finally, we compare individual origin-destination flows estimated by each model to the data [Figs. 1 and 2, panels (j)

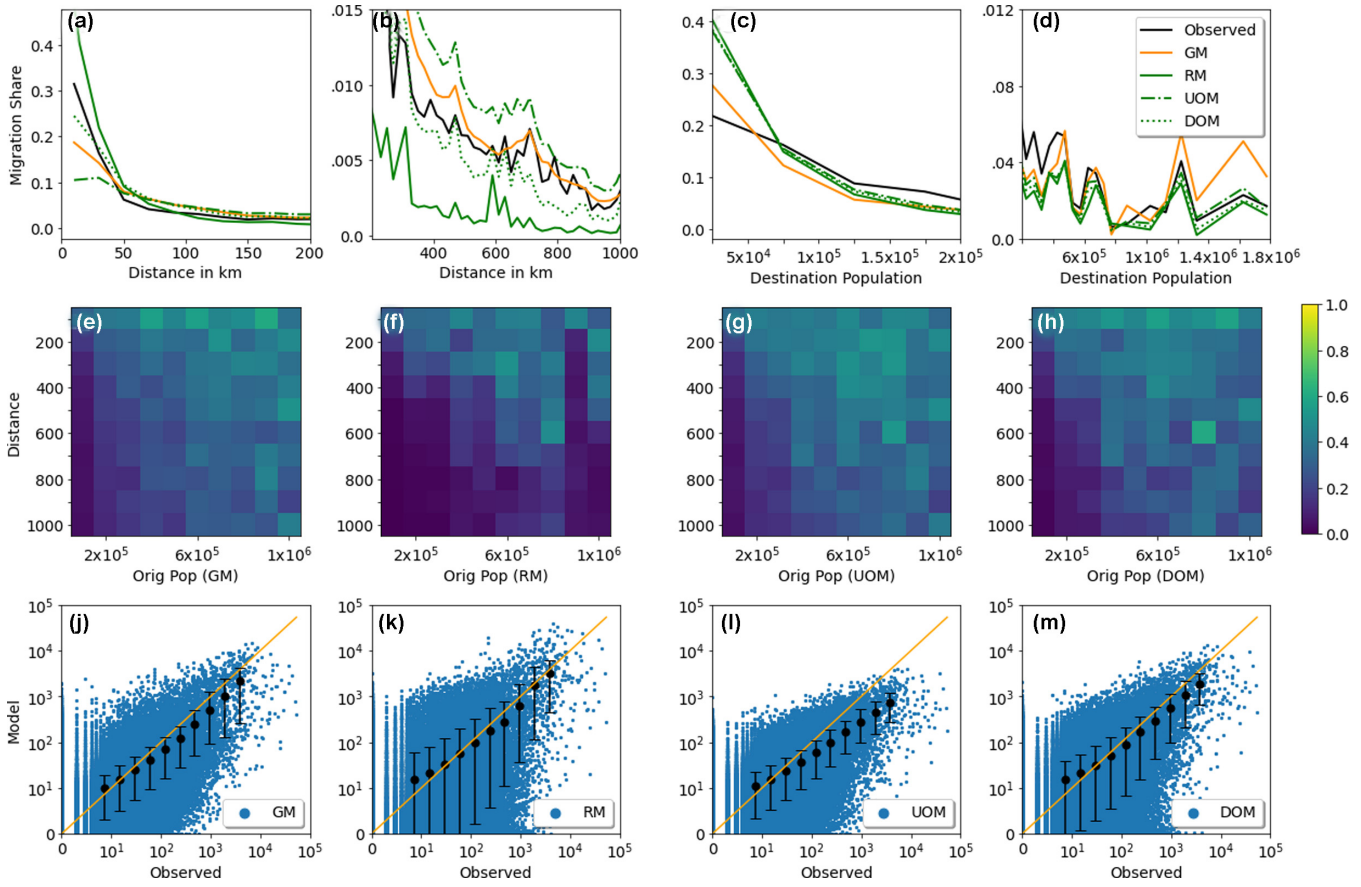

FIG. 2. As Fig. 1 but for Mexico, with observations based on microcensus data [23]. Binning distance in (a) and (b) is 20 km. Sørensen-Dice coefficients in panels (e)–(h) are shown for 100 km and 100 000 inhabitants bins.

and (k)]. While the RM fits the data well on average for most flow size classes, the spread is considerably larger than for the GM.

Next, in order to examine the spatial distribution of migration flows more closely, we analyze migration originating from counties in a single state. We choose Texas with its 254 counties in the USA and San Luis Potos with its 58 municipalities in Mexico. We distinguish between moves within the state and moves to counties in different states, and we compare the distribution of arriving migrants. For this analysis, we set M_i in Eqs. (8) to (10) to the respective total number of migrants reported in the data, separately for intrastate and interstate migration. This allows us to focus on the models' performance in simulating the spatial distribution of flows, rather than the total flow size. We consider this important, since the previous analysis already showed that, given the countrywide total number of migrants, the radiation model would underestimate the share of interstate moves, which tend to be longer distance.

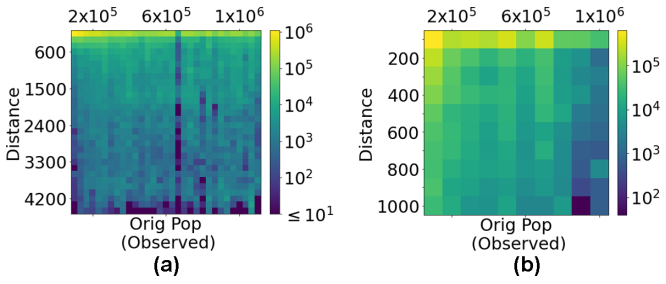

FIG. 3. Observed migration flow density in (a) the USA (IRS) and (b) Mexico (IPUMS), in terms of traveled distance and origin population population size. The colorbar indicates the number of migrants.

The spatial distribution of migrants arriving from Texas to other U.S. states is matched relatively well by the GM. The RM shows a concentration of moves to close-by destinations and underestimates the magnitude of longer-distance moves, for instance, to the northwest coast and the New England states [Fig. 4(a), panels A–C)]. A similar behavior is found in Mexico; however, the GM there overestimates the magnitude of flows from San Luis Potos state to distant municipalities [Fig. 4(b), panels A–C)]. For intrastate migration, on the other hand, the RM performs similarly well as the GM in San Luis Potos and better than the GM in Texas [Figs. 4(a) and 4(b), panels F–H). These results are generally in line with our findings from Figs. 1 and 2 that the RM performs well at intermediate distances but underestimates long-distance migration.

When comparing the GM and the RM, one has to keep in mind that one fundamental difference between the two model approaches is that the RM models exclusively the spatial distribution of moves out of a given origin i. The total number of moves, M_i, enters in Eq. (8) as a separate parameter that

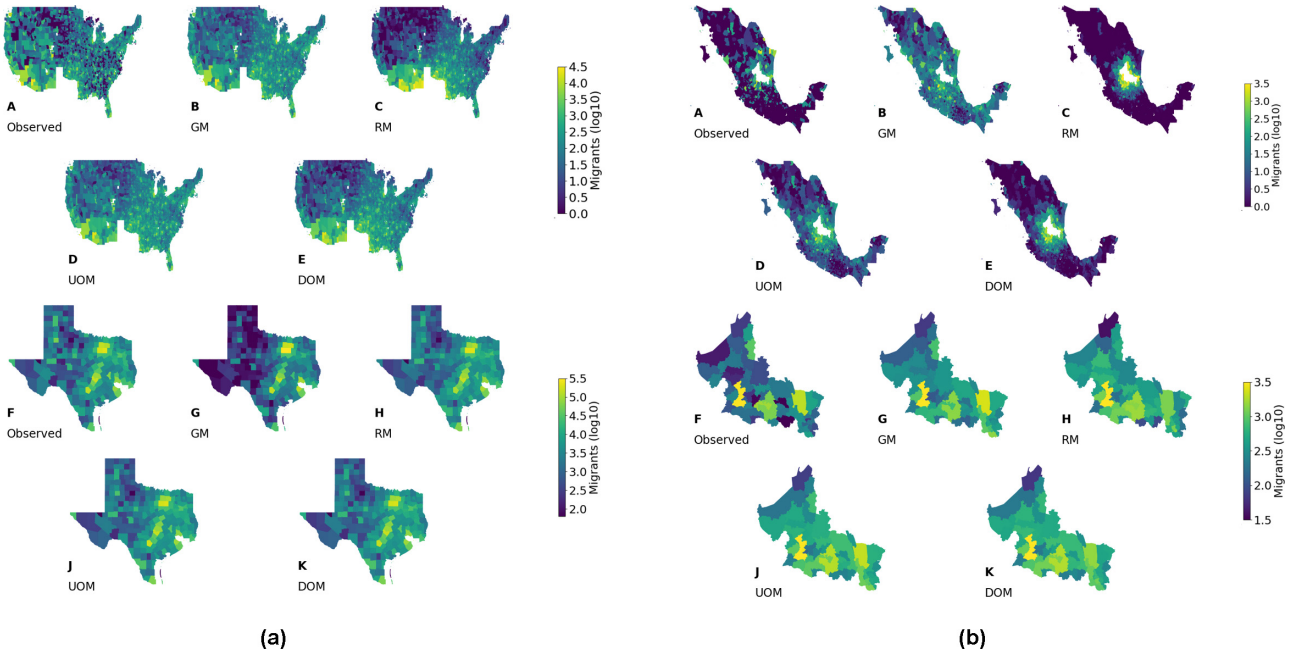

FIG. 4. Observed and model data for inter- and intrastate migration originating in (a) Texas and (b) San Luis Potos. Observed flows were taken from census data [23,24]. Panels A–E display the spatial distribution of arriving intrastate migrants and panels F–K display the distribution of arriving interstate migrants.

can be set to match the observed number from census data. In contrast, the GM models the absolute number of moves between any pair of locations i and j. As a result, the total number and the spatial distribution of moves out of i cannot be clearly separated. Therefore, estimates of the total number of migrants differ between the census data and the GM, as displayed in Table II.

This could lead to the GM comparing unfavorably to the RM, because the RM is provided, and therefore matches exactly, the total number of moves. To test the implications of this, we perform an additional calibration of the GM, in which we force the total number of moves to match the observed one (Supplemental Material [35]). The performance of the resulting, "normalized" GM is very similar to that of the initial GM, thus confirming the results discussed in this section (see Supplemental Material [35]; Figs. 1 and 2, panels (a)–(e) and (j)).

We now turn to the UOM and the DOM. In terms of flow size versus travel distance, in the USA, both the UOM and the DOM fit the data better than the RM for most of the distance spectrum [Figs. 1(a) and 1(b), dash-dotted and dotted lines]. The share of short-distance flows is slightly underestimated. In Mexico, the DOM fits the data even better than the UOM for distances below 50 km and between 150 and 700 km [Figs. 2(a) and 2(b)]. No significant difference is found between all three radiation-type models in terms of flow size versus destination population size [Figs. 1 and 2, panels (c) and (d)].

Considering the Sørensen-Dice coefficients for the UOM and the DOM, one sees a clear improvement compared to the RM. In the USA, both models show a similar performance for travel distances up to approximately 1500 km. At even larger distances, as well as for migration from smaller origin counties, the UOM tends to achieve higher coefficients [Figs. 1(g) and 1(h)]. It should be noted that observed flows are generally small at such very large distances [compare Figs. 3(a) and 3(b)]. For Mexico, both the UOM and the DOM produce major improvements compared to the RM across most of the travel distance—origin population space, and perform similarly well as, or even better than, the GM [Figs. 2(g) and 2(h)]. For small distances (0–100 km bin), the UOM achieves somewhat smaller coefficients than the RM, while the DOM achieves a performance similar to that of the RM.

In terms of individual bilateral flows, both the UOM and the DOM match the mean flows in the USA very well across all size classes and have a spread smaller than that of the RM. The UOM shows a spread even smaller than that of the DOM and the GM for small flows [Figs. 1(l) and 1(m)]. Similar results can be observed for Mexico, although here the DOM clearly matches the data better than the UOM for large flows [Figs. 2(l) and 2(m)].

Both the UOM and the DOM match the spatial distribution of arriving migrants well, for both interstate and intrastate migration from/in Texas [Fig. 4(a), panels D, E and J, K]. For interstate migration originating in San Luis Potos, both the DOM and the UOM come close to the observed pattern, with the UOM showing somewhat more and the DOM somewhat less long-distance migration than observed [Fig. 4(b), panels D and E]. As such, both models produce a pattern more realistic than that of either RM or GM. For intrastate migration in San Luis Potos, there is little difference between both models and, in fact, between any of the models [Fig. 4(b), panels F–K].

B. Theoretical background to the damped opportunity model

As we have seen above, the DOM fits empirical data better than the RM and as well as the UOM, or even slightly better

in the case of the Mexico dataset. We propose the DOM as a useful extension of the RM for the case of intrastate migration. It is more parsimonious than the UOM, in having only a single parameter κ. Moreover, this parameter can be readily interpreted as representing incomplete information, from an individual migrant's perspective, about the totality of intervening opportunities between i and j. The larger the distance between i and j, the less likely it becomes that a potential migrant will obtain and evaluate information about all jobs or amenities within the surrounding. Indeed, this consideration may be less relevant for short-distance commuter mobility, but for longer-distance migration, it is plausible that the weight of the intervening opportunities scales sublinearly with the population s_{ij}.

The DOM, moreover, preserves an analytical property of the RM that may be desirable. A conceptual difference between radiation- or intervening-opportunities-type models and gravity-type models is that the former account for an effect of inhomogeneities in the spatial distribution of population on average travel distance, while the latter do not. In the special case of a uniform population distribution (implying $m_i = m_j = m$ and $s_{ij} \propto m\, d_{ij}^2$), the original RM can be transformed into a gravity model with parameters $a + b = 1$ and $g = 4$, which agrees well with experimental values for a gravity model of commuting trips (supplementary information of Ref. [3]). In the absence of spatial inhomogeneities, the two modeling approaches can thus be reconciled.

We can apply the same transformation to the DOM. If the distribution of population over space is uniform, and spatial units have equal area F, then $s_{ij} = \frac{m}{F}\pi d_{ij}^2$. Plugging these assumptions into the DOM [Eq. (10)] we obtain

$$\begin{aligned}\frac{M_{ij}}{M_i} &= \frac{m_i m_j}{\left(m_i + s_{ij}^{\kappa}\right)\left(m_i + m_j + s_{ij}^{\kappa}\right)} \\ &= \frac{m}{\left(1 + \frac{s^{\kappa}}{m}\right)\left(2 + \frac{s^{\kappa}}{m}\right)} \\ &= \frac{m}{\left[1 + m^{\kappa-1}\left(\frac{\pi}{F}d^2\right)^{\kappa}\right]\left[2 + m^{\kappa-1}\left(\frac{\pi}{F}d^2\right)^{\kappa}\right]} \\ &\approx \frac{m}{\left(\frac{\pi}{F}\right)^{2\kappa} m^{2\kappa-2}\, d^{4\kappa}} \\ &= (F/\pi)^{2\kappa}\frac{m^{3-2\kappa}}{d^{4\kappa}}, \end{aligned} \tag{12}$$

where we have omitted subscripts after the first line for clarity. The result is equivalent to a gravity model with $a + b = 3 - 2\kappa$ and $g = 4\kappa$. With $\kappa < 1$, this makes the GM parameters more similar to the ones we estimate for internal migration, which are substantially different from those estimated for commuting [3].

The UOM, on the other hand, does not generally transform into a gravity-type model in the limit of uniform population distribution. Applying the same transformation of variables to the UOM [Eq. (9)] yields

$$\frac{M_{ij}}{M_i} \approx \left(\frac{F}{(\alpha+\beta)\pi}\right)^2 \frac{m}{d^4} + \frac{\alpha F}{(\alpha+\beta)^2\pi}\frac{m}{d^2}, \tag{13}$$

which is equivalent to a superposition of two GMs, one with $g = 4$ and one with $g = 2$. If β is set to zero, then the UOM transforms into a gravity-type model with parameters $a = b = 0$ and $g = 2$, i.e., the predicted flow distribution becomes independent of the origin and destination population sizes.

V. DISCUSSION AND CONCLUSIONS

We have, for the first time, systematically evaluated the radiation model for the case of internal migration and compared it to the gravity model, for two different, large countries: the USA and Mexico. Given its simplicity, we find the radiation model to perform relatively well overall and partly even better than the gravity model when it comes to short-distance migration. However, the radiation model features a fast decay of flow size with distance and thus severely underestimates long-distance migration, i.e., to destinations more than a few hundred kilometers away. Unlike commuting flows, which are very small at such large distances, long-distance moves beyond 300 km (500 km) make up 22% (19%) of all observed migration flows in the USA and 26% (22%) in Mexico [see Figs. 3(a) and 3(b)]. The systematic underestimation of such moves by the radiation model is therefore a serious drawback for applications related to migration. Using the analogy with a radiation process, the absorption of particles by the surrounding matter is too strong compared with empirical data.

We have evaluated two extensions of the radiation model that offer the flexibility to modulate the absorptive strength. The universal opportunity model (UOM) has been shown to work well for a variety of different forms of human mobility [18] and, according to our analysis, works better than the radiation model at reproducing long-distance migration. This, however, comes at the cost of poorer performance for short distances, at least in Mexico, and at the expense of two parameters that need to be calibrated. The damped opportunity model (DOM), suggested here as a more parsimonious extension of the radiation model, performs well over almost the entire travel distance spectrum in both countries, matches the spatial distribution of flows well both at the country level and for interstate and intrastate migration originating in selected states, and has only a single free parameter that needs calibration. Moreover, unlike the UOM, the DOM preserves the property of the radiation model of being equivalent to a gravity model in the limit of uniform population distribution. This allows users to reconcile the different modeling paradigms and choose the model that best suits the structure of the considered system and the availability of empirical data.

Considering the complexity of a phenomenon such as human migration, both the original gravity model and the radiation model are very simplistic, representing a given location only by its population size. Although agglomeration effects are a strong driver of migration [36], and population size can generally be considered to reflect, to some extent, the attractiveness of a given location, it is only a proxy for various factors such as job opportunities, social amenities, public infrastructure, etc., that attract migrants [37–39]. Economists and demographers now routinely use modified gravity models that include some of these factors and their influence on migration in more direct ways, for instance, representing economic opportunities through wage levels or employment rates, and representing the migration decision process through utility maximization frameworks [40–42].

Nevertheless, these modified gravity models suffer from some of the same drawbacks as the basic gravity model; notably, a lack of mathematical consistency and a high degree of parametrization [3]. The findings from our study suggest that modified radiation models, as an alternative modeling approach, may have potential for future practical applications in migration research.

To conclude, in this paper we have shown that the radiation model strongly underestimates long-distance migration and, as a result, overestimates short-range migration flows. We have proposed an extended version of the radiation model, which aims at solving the previously mentioned problems by including a parameter which dampens the growth of the intervening opportunities with increasing distance. We have compared this approach to the universal opportunity model, an existing extension of the radiation model including two fitting parameters, and have shown that our "damped opportunity model" yields a similar performance despite only using a single parameter and also preserves a desirable analytical property of the original radiation model. Our paper demonstrates that the gravity model yields better results than the original radiation model for within-country migration, but that radiation-type models can be adapted to offer competitive performance, while being analytically more consistent and more parsimonious than the gravity model.

ACKNOWLEDGMENT

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

[1] G. K. Zipf, The P_1P_2/D hypothesis: On the intercity movement of persons, Am. Sociol. Rev. **11**, 677 (1946).
[2] S. A. Stouffer, Intervening opportunities: a theory relating mobility and distance, Am. Sociol. Rev. **5**, 845 (1940).
[3] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, A universal model for mobility and migration patterns, Nature (London) **484**, 96 (2012).
[4] M. Stefanouli and S. Polyzos, Gravity vs radiation model: two approaches on commuting in Greece, Transp. Res. Procedia **24**, 65 (2017).
[5] G. McNeill, J. Bright, and S. A. Hale, Estimating local commuting patterns from geolocated twitter data, EPJ Data Sci. **6**, 24 (2017).
[6] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, A tale of many cities: Universal patterns in human urban mobility, PLoS ONE **7**, e37027 (2012).
[7] I. Hong and W.-S. Jung, Application of gravity model on the Korean urban bus network, Phys. A (Amsterdam, Neth.) **462**, 48 (2016).
[8] J. R. Gog, S. Ballesteros, C. Viboud, L. Simonsen, O. N. Bjornstad, J. Shaman, D. L. Chao, F. Khan, and B. T. Grenfell, Spatial transmission of 2009 pandemic influenza in the US, PLoS Comput. Biol. **10**, e1003635 (2014).
[9] Y. Xia, O. N. Bjørnstad, and B. T. Grenfell, Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics, Am. Nat. **164**, 267 (2004).
[10] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, Modelling cholera epidemics: the role of waterways, human mobility and sanitation, J. R. Soc. Interface **9**, 376 (2012).
[11] 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).
[12] V. Palchykov, M. Mitrović, H.-H. Jo, J. Saramäki, and R. K. Pan, Inferring human mobility using communication patterns, Sci. Rep. **4**, 6174 (2014).
[13] M. Lenormand, A. Bassolas, and J. J. Ramasco, Systematic comparison of trip distribution laws and models, J. Transport Geogr. **51**, 158 (2016).
[14] D. Karemera, V. I. Oguledo, and B. Davis, A gravity model analysis of international migration to North America, Appl. Econ. **32**, 1745 (2000).
[15] D. A. Plane, Migration space: Doubly constrained gravity model mapping of relative interstate separation, Ann. Assoc. Am. Geogr. **74**, 244 (1984).
[16] D. Bunea *et al.*, Modern gravity models of internal migration: the case of Romania, Theor. Appl. Econ. **4**, 127 (2012).
[17] K. K. Rigaud, A. de Sherbinin, B. Jones, J. Bergmann, V. Clement, K. Ober, J. Schewe, S. Adamo, B. McCusker, S. Heuser, and A. Midgley, Groundswell: Preparing for Internal Climate Migration (World Bank, Washington, DC, 2018).
[18] L. Er-Jian and Y. Xiao-Yong, A universal opportunity model for human mobility, Sci. Rep. **10**, 4657 (2020).
[19] I. S. Lowry, *Migration and Metropolitan Growth: Two Analytical Models* (Chandler, San Francisco, 1966).
[20] A. J. Garcia, D. K. Pindolia, K. K. Lopiano, and A. J. Tatem, Modeling internal migration flows in sub-Saharan Africa using census microdata, Migr. Stud. **3**, 89 (2015).
[21] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, Synchrony, waves, and spatial hierarchies in the spread of influenza, Science **312**, 447 (2006).
[22] IRS Migration Flows, https://www.irs.gov/pub/irs-soi/county0708.zip, id=212695, 00.html, accessed 2020-09-21.
[23] Mexico Population Census, http://en.www.inegi.org.mx/temas/estructura/, accessed 2020-06-05.
[24] U.S. Census County to County Flows, https://www.census.gov/topics/population/migration/guidance/county-to-county-migration-flows.html, accessed 2020-03-28.
[25] USA shape file, https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html, accessed 2020-09-21.
[26] Census U. S. Intercensal County Population Data, National Bureau of Economic Research, https://data.nber.org/data/census-intercensal-county-population.html, accessed 2020-04-19.
[27] County Distance Database, National Bureau of Economic Research, https://data.nber.org/data/county-distance-database.html, accessed 2020-04-19.

[28] Mexico 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: 2020-06-02.

[29] 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**, 50 (2019).

[30] Mexico shape file, https://datacatalog.worldbank.org/dataset/mexico-municipalities-2012, accessed 2020-06-02.

[31] Th. A. Sorensen, 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, Biol. Skr. **5**, 1 (1948).

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

[33] 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. Network Anal. Min. **10**, 43 (2020).

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

[35] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.104.054311 for a gravity model constrained to produce the same total number of moves as given by the census data.

[36] S. P. Kerr, W. Kerr, Ç. Özden, and C. Parsons, High-skilled migration and agglomeration, Annu. Rev. Econ. **9**, 201 (2017).

[37] D. Acemoglu, A microfoundation for social increasing returns in human capital accumulation, Q. J. Econ. **111**, 779 (1996).

[38] B. Jones, Assessment of a gravity-based approach to constructing future spatial population scenarios, J. Popul. Res. **31**, 71 (2014).

[39] J. R. Harris and M. P. Todaro, Migration, unemployment and development: A two-sector analysis, Am. Econ. Rev. **60**, 126 (1970).

[40] M. J. Greenwood, Internal migration in developed countries, in *Handbook of Population and Family Economics* (Elsevier, Amsterdam, 1997), Chap. 12, pp. 647–720.

[41] J. Grogger and G. H. Hanson, Income maximization and the selection and sorting of international migrants, J. Dev. Econ. **95**, 42 (2011).

[42] F. Docquier, J. Machado, and K. Sekkat, Efficiency gains from liberalizing labor mobility, Scand. J. Econ. **117**, 303 (2015).