communications physics

Article

https://doi.org/10.1038/s42005-024-01672-z

A generalized vector-field framework for mobility

Check for updates

Erjian Liu[1,2,3] ✉, Mattia Mazzoli[4], Xiao-Yong Yan[2] & José J. Ramasco[1] ✉

Given the identification with travel demand and its relevance for transportation and urban planning, the estimation of trip flows between areas is a fundamental metric for human mobility. Previous models focus on flow intensity, disregarding the information provided by the local mobility orientation. A field-theoretic approach can overcome this issue and handle both intensity and direction at once. Here we propose a general vector-field representation starting from individuals' trajectories valid for any type of mobility. We also show with simplified models how individuals' choices determine the mesoscopic properties of the mobility field. Distance optimization in long displacements and random-like local exploration are necessary to reproduce empirical field features observed in Chinese logistic data and in New York City Foursquare check-ins. Our framework is able to capture hidden symmetries in mesoscopic urban mobility and opens the doors to the use of field theory in a wide spectrum of applications.

Characterizing mobility patterns between locations of people and goods is not only a long-standing research topic in disciplines such as geography and spatial economics[1–8], but it has also critical practical applications for urban planning[9–12], building and expansion of transportation infrastructure[13–16], population distribution studies and city sprawl[17–21], urban socioeconomic development[22–24], location-based services[25] and infectious diseases spreading[26–32].

Early research works mainly used data from transportation surveys and census to analyze people travel and activity patterns[8,33]. The collection of such mobility data can be expensive and time-consuming, and it is hard to acquire a sufficient sample size[34,35]. With the development of information and communication technologies (ICTs), the availability of large-scale high-resolution mobility data, such as call detail records, GPS-located taxi data and online social network data, has notably increased. This enabled researchers to quantitatively study various mobility contexts and put forward new models to analyze the underlying mechanism of mobility[8,36–42].

In terms of models, mobility can be studied at two different levels: individual trajectories and aggregated flows between areas. Individual mobility models usually focused on characterizing the behavior of individuals in the process of selection of destinations, adding a certain degree of stochasticity to account for people heterogeneity and free will (see refs. 36,43–45 for some examples and[8] for a recent review). At the aggregated level, the earliest models fall into two families: the gravity[13,46] and the intervening opportunity model[14,47], which has later evolved into the so-called radiation model[48]. These models and their updated versions predict the trip flows between locations, describing travel distance distribution and defining locations attractiveness[48–53]. In fact, when compared in terms of performance at predicting trip flow intensity the gravity and the radiation models show only slight differences[50]. A difference emerges when the dominant local direction of the trips is considered[54].

Such combination of mobility flow intensity and orientation can be studied using a field theoretical framework. The idea of using fields and potentials for studying mobility emerged in the context of the gravity model, where these concepts appear in a natural way[55,56]. The lack of data prevented further advances on this direction, until a recent work[54] proved that a vector field framework could be used to characterize trips between home and work (commuting) in a number of cities in the world. Not only that, this framework was able to solve a controversy almost 80 years old on which of the two families of models performs best to describe commuting. The gravity model produces results that matches empirical commuting mobility patterns both in intensity and orientation of the flows. A field representation has lately been used in a machine learning context, where knowing the potential can significantly improve the performance of the model to predict traffic flows[57]. Later, some works translated the field approach to lower scales of mobility (single flows)[58–60], pedestrian route selection[61] or the mobility associated to the celebration of special events[62]. Nevertheless, it is not yet clear how the definition of mobility fields can be extended to any type of mobility starting from individual trajectories and what are the features that may permeate from the microscopic mobility information to the mesoscopic (field) scale.

[1]Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), 07122 Palma de Mallorca, Spain. [2]School of Systems Science, Beijing Jiaotong University, Beijing 100044, China. [3]Beijing Transport Institute, Beijing 100073, China. [4]ISI Foundation, 10126 Turin, Italy. ✉e-mail: erjian@ifisc.uib-csic.es; jramasco@ifisc.uib-csic.es

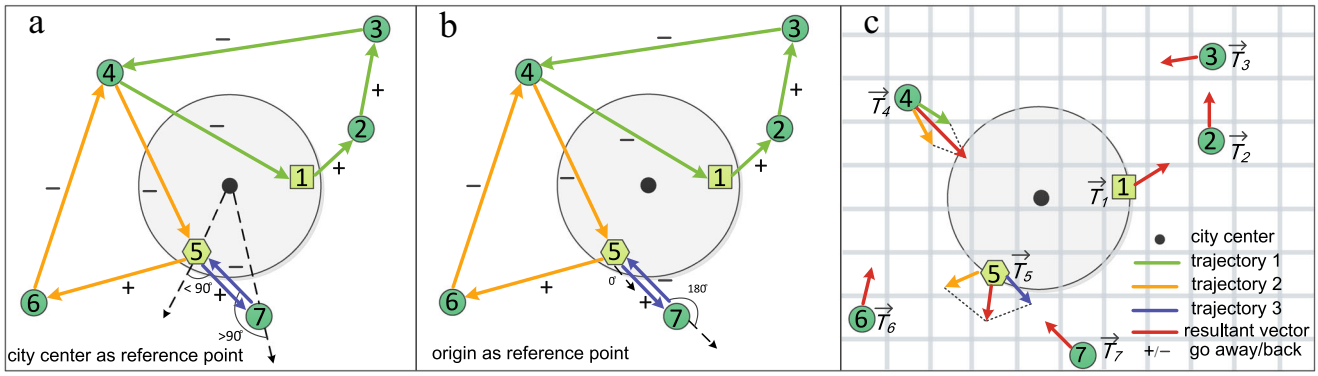


Fig. 1 | Definition of trajectories orientation and resulting vector field. The large gray circle stands for an idealized city (with the black point as the city center), the green numbered circles are the stops sequence of each trajectory (1-2-3-4-1, 5-6-4-5 and 5-7-5), while the square in 1 and the hexagon in 5 are the trajectories' origins. **a** City center as reference point (RP). Vectors of each color connect consecutive stops of a trajectory, for example, the trajectory 1-2-3-4-1. When the vector connecting X to Y forms an acute angle with the position vector $\vec{X}$ starting in the city center, we mark the vector $\overrightarrow{XY}$ as positive, meaning that the agent is moving away from the city center, like the vector $\vec{57}$. Vice versa, we mark the vector as negative when the agent moves towards the city center, e.g. the vector $\vec{75}$. **b** Trajectory-origin as RP. Identically to what we defined for the city center but we use the origins of trajectories 1 or 5, respectively, as RP for the position vectors. **c** Sketch of the method to build the vector field. The space is divided in a grid, vectors departing from stops in a cell i are normalized and summed vectorially to produce $\vec{T}_i$. Then we define the mobility field $\vec{W}_i$ dividing the vector $\vec{T}_i$ by the number of trips departing from i.

In this work, we introduce a definition of a vectorial framework for mobility encoded in individual trajectories, which is valid for all mobility types. To understand the origin of the fields, we also define a sign for the trips: those pointing toward the city (reference point) are negative, while those pointing away are positive. The sign of a full trajectory depends on the majority sign of the trips composing it. We empirically find a net imbalance, with more negative than positive trips, in trajectories extracted from logistic trips in 20 Chinese cities and Foursquare check-ins in New York City. The asymmetry of trip directions introduces a mobility field in the space. We propose then three individual mobility models to analyze what is the minimal set of ingredients needed to reproduce the observed imbalance. Our results show that a distribution of stops decaying with the distance to the city center, length optimization for long displacements and random-like local exploration are fundamental to reproduce empirical mobility fields. We finally focus on the properties of the mobility fields, empirical and model generated, and find that they fulfill the divergence (Gauss) theorem and that are mainly irrotational so that a potential can be defined at the mesoscopic scales for mobility. This work extends thus the mathematical amenability of mobility data and offers an alternative approach for individual mobility patterns analysis.

Results

From trips to vectors

The microscopic scale of mobility is the realm of individual trajectories where the main features are stop locations, traveling times between stops, idle times, etc. The global macroscopic scale refers to the whole considered region, for example, total number of displacements, total number of trajectories or average distance covered between stops. There is, nevertheless, an intermediate scale, which we call mesoscopic. Let us divide the space of the full region into grid cells of a given area, for instance, in a city these could have an arbitrary dimension of 1 km^2. The description of the mobility flows at the mesoscopic scale in these cells in terms of intensity and overall direction is the main goal of this work and for this purpose we introduce a vectorial framework.

A vector is a mathematical object characterized by both magnitude and direction, while a vector field is defined by an association of a vector to each point in the space. A sketch of the method to define a vector field from mobility data is displayed in Fig. 1, where we show an idealized "circular" city, its central point (black circle) and three trajectories: 1-2-3-4-1, 5-6-4-5 and 5-7-5. Most of our trajectories are closed, i.e. the origin and the final stop coincide, but this is not a requirement to define the field. The steps to define the vector field are as follows:

1. We define vectors between consecutive stops. This can be seen, for example, in Fig. 1a, b, where the vectors $\vec{12}, \vec{23}, \vec{34}$ and $\vec{41}$ can be extracted from the first trajectory. A vector pointing from one stop X to the next Y is called $\overrightarrow{XY}$ and is located on X.
2. The vectors $\overrightarrow{XY}$ are normalized to obtain the unit vectors $\vec{xy}$.
3. The space is divided in grid cells of equal area and all unit trip vectors $\vec{xy}$ within each cell i are vectorially summed to define the resulting vector $\vec{T}_i$, which informs on the average mobility direction in i (see Fig. 1c).
4. $\vec{T}_i$ is normalized by the total number of trips leaving cell i to obtain the mesoscopic mobility vector field $\vec{W}_i$. This process is analogous to defining the gravitational or electrical fields dividing the force by the mass or charge, respectively.

The vectors $\vec{W}$ constitute thus the mobility field. In simple words, we use vectors to represent the trips within each trajectory and it is the embedding of these vectors in the urban space which forms the mobility field.

From trip vectors to trajectory orientation

In order to characterize how individuals explore space we need to determine whether trips in each area head towards or away from a given reference point (RP). We identify two options for RP: the first one is the city geographical center (Fig. 1a), in this case the RP is absolute and equal for all the trajectories; the second option is to establish the origin of each trajectory as RP (Fig. 1b). This latter option implies that the RP is different for every trajectory. As we will see below, the trajectory-origin RP shows some useful features and most of the results here are, therefore, displayed using such RP unless otherwise stated.

With respect to the chosen RP, we can allocate a sign to each displacement vector $\overrightarrow{XY}$ (or $\vec{xy}$) of any trajectory. Remember that the vector $\overrightarrow{XY}$ sits on X, and that every stop X can be described by a position vector $\vec{X}$ from the RP to X. To understand whether the displacement $\overrightarrow{XY}$ occurs toward or away from the RP, we compute the angle θ in the range $(-180, 180]$ between the above two vectors (see Fig. 1a and b), where $|\theta| = 0$ means moving straight away from the RP and $|\theta| = 180$ means moving strictly toward the RP. We assign the vector $\overrightarrow{XY}$ a positive sign $(+)$ if $|\theta| < 90$, and a negative sign otherwise $(-)$. By convention, vectors $\overrightarrow{XY}$ are

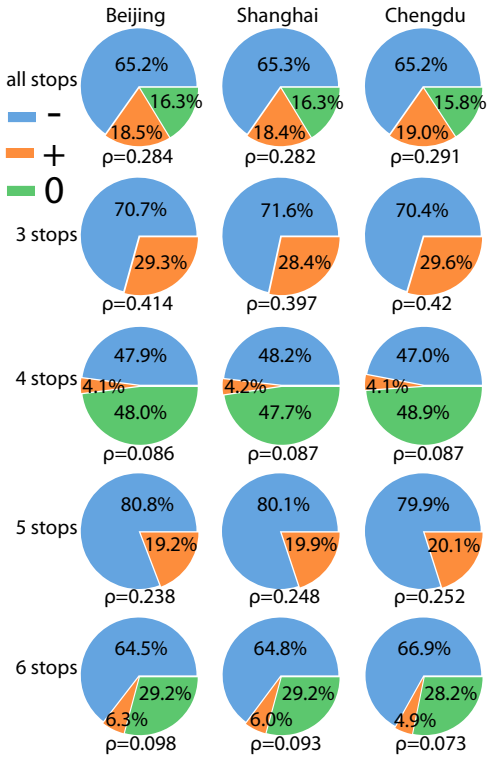


Fig. 2 | Unbalance of empirical trajectories. Fraction of positive (orange), negative (blue), and balanced (green) trajectories with 3, 4, 5, 6 stops and for all trajectories of trucks departing from Beijing, Shanghai and Chengdu, using the trajectory origin as reference point(RP). ρ stands for the ratio between the number of positive and negative trajectories.

positive if the stop X coincides with the RP. Some examples from Fig. 1 are the vectors $\overrightarrow{34}$ or $\overrightarrow{64}$ that are both negative pointing to the RP in both representations, or the vector $\overrightarrow{23}$ that is positive. Vectors exiting from the origins such as $\overrightarrow{57}$, $\overrightarrow{12}$ and $\overrightarrow{56}$ are by convention positive.

We characterize for every trajectory t whether trips are on average toward $(-)$ or away $(+)$ from the RP by summing the signs of all the displacement vectors. Dividing by the total number of vectors, we obtain the average orientation H_t laying in the range between -1 and 1, where $H_t = 1$ means that all the displacements are positive (i.e., away from the RP) and vice-versa for $H_t = -1$. Note that distance or duration of trips are not taken into account. Finally, $H_t = 0$ implies a full balanced trajectory. H_t is a microscopic observable that encodes individual behavioral features of mobility.

A priori, there is no reason to assume that there should be more displacements in one orientation than in the other. This means that overall there should be as many trajectories with positive or negative H_t. We count as N_+ the number of trajectories with $H_t > 0$, N_- the number of those with $H_t < 0$ and N_0 as the number of balanced trajectories. We finally introduce the unbalance ratio ρ as

$$\rho = \frac{N_+}{N_-}. \quad (1)$$

The existence and orientation of the mesoscopic field and the orientation of trajectories are mathematically interlinked. Since every displacement generates an unit trip vector contributing to the overall mobility field, a situation with ρ clearly deviating from one could generate a majority direction for displacements and, consequently, an average field (see Supplementary Note 4 (Supplementary Fig. 8 and Supplementary Tables 4, 5) for a demonstration that $\rho < 1$ implies the existence of a field).

Empirical trajectories orientation

Knowing that an unbalance in the trajectories may lead to the presence of a mobility field, it is important to test whether empirical trajectories are balanced or not. We consider two datasets: the first one, D1, refers to logistic trucks trajectories departing from the 21 largest Chinese cities, while the second one, D2, refers to Foursquare check-ins of individuals in New York City (NYC) (see Methods and Supplementary Note 1 for a detailed description of the two datasets (Supplementary Tables 1 and 2 and Supplementary Fig. 1 and 2)). We show the empirical results for ρ in Beijing, Shanghai and Chengdu of D1 using trajectory origins as the RP in Fig. 2. The pie charts show the fraction of positive, negative and balanced trajectories with a fixed number of stops and also the overall results. Trajectories with negative orientation (i.e., N_-) are dominant, hence $\rho < 1$ in all cases. Results are robust for trajectories of any number of stops and for all trajectories in all the 21 cities analyzed. Note that negative trajectories dominate independently from the length of trajectories and there is high similarity in the ρ values across cities, overall and for trajectories with a certain number of stops. This result does not hold if the RP is the city center (see Supplementary Information Note 3 and Supplementary Figs. 5–7). It seems that the analysis performed with the origins of trajectories as RP is able to absorb the details of the cities in terms of shape, streets and communication axes (e.g. highways), hence leaving only individual mobility behavior. This has two consequences: firstly, we can neglect the urban shape when modeling individuals' mobility behavior in terms of ρ; secondly, we can tune the model on an arbitrary city and perform out-of-sample accurate predictions. The results are robust for further cities (see below in Understanding the origin of the trajectory unbalance section and Supplementary Note 5 Supplementary Figs. 9–12).

Since the definitions of vector signs and trajectory orientations are general, we can apply it to any type of mobility. As a comparative, we perform the same analysis on D2, the check-in records of Foursquare, and find a similar pattern (see Supplementary Fig. 9). Regarding the Foursquare check-in data, trajectories with negative orientation (i.e., N_-) also exhibit dominance, resulting in $\rho < 1$, for trajectories of any number of stops and for all trajectories (see Supplementary Note 5). Note that D2 encodes a different type of mobility compared to D1.

Comparing the results obtained from the Foursquare check-ins and the freight data, the calculated ρ values differ, whether considering only a certain number stops or for all trajectories. For example, when considering all trajectories in the Foursquare trajectories, ρ is 0.266, while for the number of stops 3, 4, 5, and 6, the corresponding ρ values are 0.357, 0.09, 0.19, and 0.05, respectively. However, in both the Chinese logistic data and New York City Foursquare check-ins, we observe an overall imbalance with more negative than positive trajectories. This suggests the existence of a property that is valid for all cities, across mobility types and both datasets.

Understanding the origin of the trajectory imbalance

In order to understand the mechanisms leading to the empirically observed unbalance between the signs of the trajectories, we introduce four models in an increasing order of complexity, each with added mechanisms over the previous one to characterize the needed ingredients (see Methods for details). The simplest configuration includes a circular city of radius R. A generic trajectory is composed of a sequence of stops $\{\vec{x}_0, \vec{x}_1, \ldots \vec{x}_{s-1}\}$, with the origin $\vec{x}_0$ located randomly inside the circular city.

For the first model, called Rand, the other stops locations are selected completely at random in space. We have confirmed that the model trajectories tend to be balanced in the thermodynamic limit (large L) and that it does not generate a field (see Supplementary Note 6 and Supplementary Figs. 13–14). This was a sanity check before advancing to more elaborated models and we disregard this model from now on. The next three models are more relevant and are developed making simple behavioral assumptions about spatial navigation, a sketch with their description can be seen in Fig. 3.

The d-rand model follows the same logic but is informed with a spatial distribution of stops $D(r)$ decaying with the distance to the city center. This mimics a random-like exploration but with constraints on the spatial

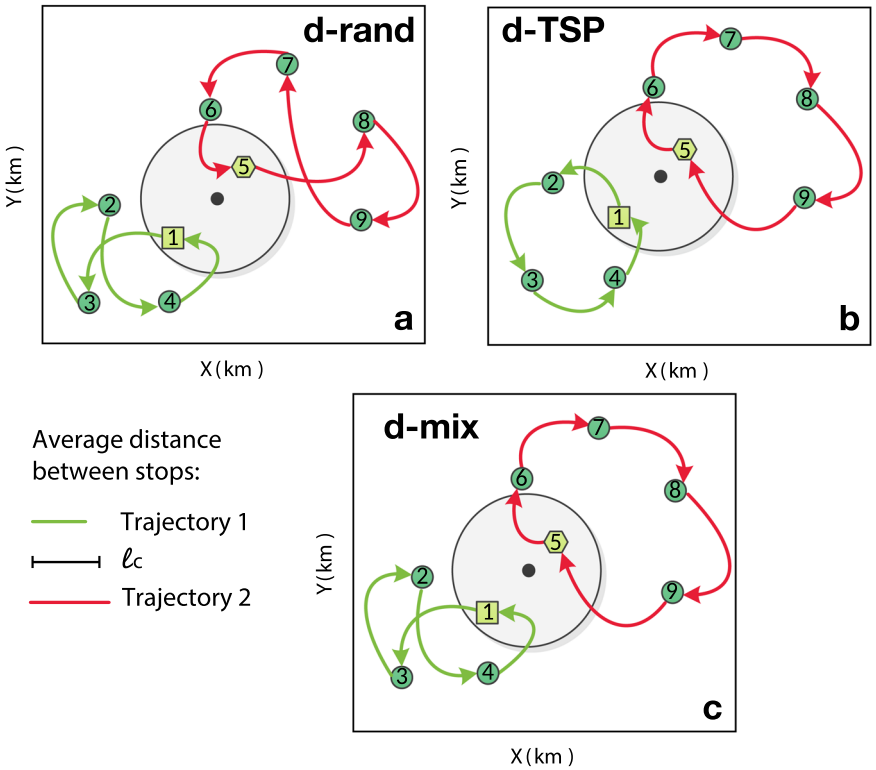


Fig. 3 | Schematic description of the models. The large gray circle represents the city area (the black circle is the city center), the numbered green circle represent stops while square and hexagon are the origins of two trajectories (red and green). These trajectories have different average trips distance: the green one is below ℓ_c, the red trajectory is above ℓ_c. Models (**a**) d-rand, (**b**) d-TSP and (**c**) d-mix.

distribution of stops. The next model, d-TSP, allows travelers to reorder the stops to minimize the total distance traveled (with a Traveling Salesman Problem (TSP) optimization algorithm). Finally, d-mix interpolates between the two previous models and the distance optimization only occurs if the average distance between trajectory stops goes over a threshold ℓ_c. Trajectories distances are not optimized otherwise. More details on the precise definition of the models are offered in the Methods section below.

The models are informed by the empirical statistics from Beijing (see Methods for further details on the models construction). We tune the d-mix model parameter ℓ_c by minimizing the mean absolute error on ρ in Beijing using all trajectories and the trajectory origin as RP. The best results are obtained for $\ell_c \approx 1.72\,km$, which is a reasonable threshold for route optimization (see Supplementary Note 12 and Supplementary Fig. 24).

Violin plots in Fig. 4a show the resulting distribution of ρ for the three stochastic models and the empirical distribution from the 21 cities in D1. We also show the distributions for trajectories with a fixed amount of stops only. The first observation of interest is that all the models produce trajectories that are consistently negative and hence holding $\rho < 1$. A second aspect is that d-rand generates the lowest values of ρ among all models. This is natural since the location of consecutive stops is random, without ordering them to reduce the distance traveled, and hence, the probability of crossing to the other side of the city (a negative sign for the trip) is high. Many negative trips contribute to an overall negative sign for the trajectories and a lower value of ρ. Models with stops reordering may reduce the number of displacements from one side to the other of the city, and the trajectory signs are less often negative (higher values of ρ). In contrast and following the same argument in the opposite way, d-TSP produces trajectories with the highest value of ρ. ρ for d-mix, on the other hand, lies between these two extremes as do also the empirical values of ρ. When the analysis is restricted to trajectories with a certain number of stops, the stochastic fluctuations are larger since the number of trajectories decreases. However, d-mix fits best to the empirical values of ρ.

In Fig. 4a we show the empirical ρ for trajectories from all the 21 cities in D1 together. The cities contributing to this violin plot are, nevertheless, very heterogeneous in population. This is why in Fig. 4b we depict the empirical values ρ as a function of population and see that there is no noticeable dependence. Moreover, all the empirical values fluctuate within the 95 % interval of the value of ρ obtained by the d-mix model (fitting ℓ_c only in Beijing). Finally, we can analyze the trajectories generated by the fitted d-mix model by their number of stops. In Fig. 4c, we see that, in general, the agreement for trajectories of a fixed number of stops aligns well with the empirical results from the three cities in Fig. 2. All these results have been confirmed using different values of the modeled city size R, the space considered L and, even at a qualitative level, for different $D(r)$ functions (see Methods for details on the models' parameters and Supplementary Notes 7, 8 and 9 (Supplementary Figs. 15–19) for the robustness check).

The d-mix model mobility field

Since the models produce unbalanced trajectories, i.e. $\rho < 1$, they also generate a net mobility field (see Supplementary Section 4 for a mathematical proof). Next we study the properties of the field $\overrightarrow{\mathbf{W}}$ generated by the fitted d-mix model (see Methods for the details of the modeling setting). We consider circular contours of radius r from the city center and analyze the flux of $\overrightarrow{\mathbf{W}}$ as a function of r. The flux is calculated in both as a surface integral over the circular contour and as the volume integral of the divergence of the mobility field, $\nabla \overrightarrow{W}$ (see Methods for the formal flux calculation). According to Gauss' Divergence Theorem, if the field is well behaved the two ways of calculating the flux should yield the same result. This is confirmed in Fig. 5a, where the two calculations of the flux Φ_W as a function of the distance are in agreement with a coefficient of determination $R_p^2 = 0.98$. Note that this does not apply to the city area where the estimated fluxes are close to zero. The fact that the Gauss' Theorem is fulfilled is important because it is related to the existence of a source for the field.

A second relevant feature to explore is the field curl. This is connected to the possibility of defining a potential for the field. Figure 5b displays the module of the curl, which lies in the z-axis perpendicular to the plot. One can distinguish the area of the city in the internal circle. There are some non-zero curl areas, with the highest values concentrated close to the city border. Actually, these values are small when compared with a null model (see Fig. 5c). In this null model, the direction of the d-mix vectors in each cell is randomly reoriented. We call this the "fully random" model and it is intended to assess the level of curl induced only by noise. The overall

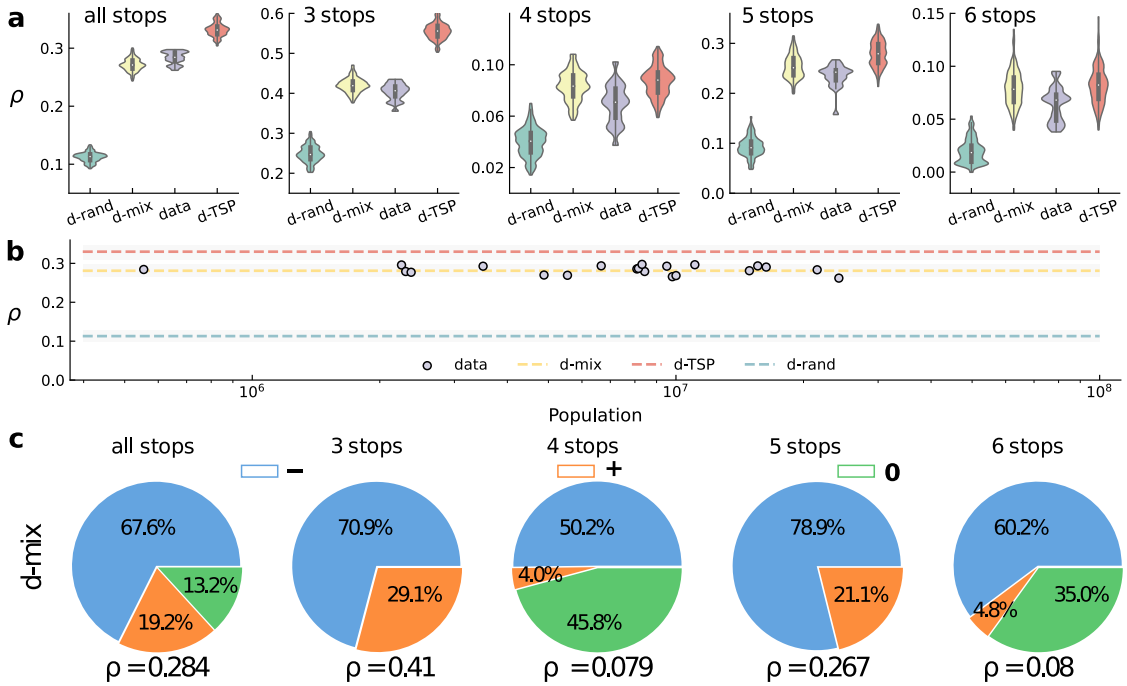
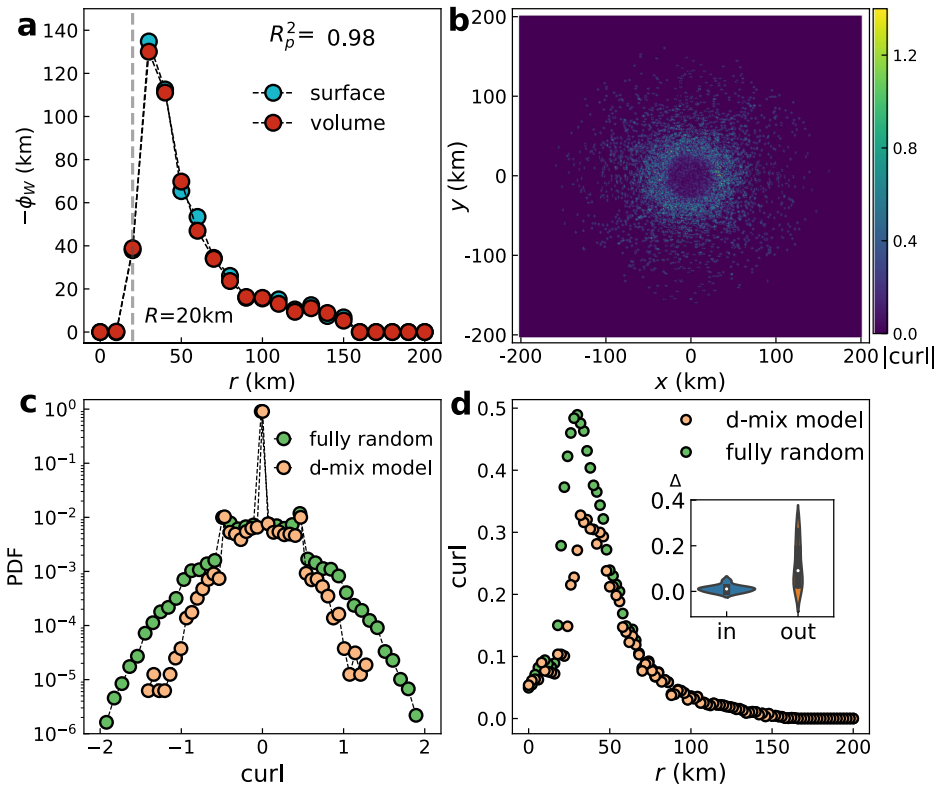


Fig. 4 | Models and trajectory unbalance. a Violin plots of the model-predicted and real values of the unbalnace ratio ρ for trajectories with different number stops. The violin plots depict the values of ρ from d-rand model (in green), d-mix model (in yellow), the data in D1 (in purple) and d-TSP model (in red) for trajectories with 3, 4, 5, 6 stops and for all trajectories. **b** Values of ρ for the different cities as a function of their population. The horizontal dashed lines correspond to the median values of ρ for all the modeled trajectories, shaded areas indicate the confidence interval between 5 % and 95 %, color codes as above. Purple dots represent the values of ρ for all empirical trajectories in each city. **c** Pie charts for the fraction of positive (orange), negative (blue) and balanced (green) trajectories generated with the fitted d-mix model. These sequence of charts should be compared with the empirical values observed in different cities in Fig. 2. For these simulations, we used a city radius of $R = 20\ km$ and bounding box $L = 400$ km and with 100,000 trajectories.

Fig. 5 | Properties of the d-mix mobility field. The mobility field is obtained from the d-mix model with the parameter of characteristic distance of $\ell_c = 1.72\ km$ with a circular city of radius $R = 20\ km$ in a space limited by a box of side $L = 400\ km$ and with 100 000 trajectories. **a** Comparison between the flux measured as the surface integral in blue, and as the volume integral of the divergence of the mobility field (see Methods for the calculation) in red, in both cases as a function of the distance to the city center r. The coefficient of determination R^2_p is obtained as the square of the Pearson correlation coefficient of both curves. The vertical dashed line marks the city radius R. **b** Module of the curl of the field, the colors represent $|\nabla \vec{W}|$ for each cell in km^{-1}. **c** Comparison between the curl of the field generated with the d-mix model and that of the fully random model obtained by randomly reassigning directions to $\vec{W}$ in each cell. **d** Average absolute value of the curl as a function of the distance to the city center for the d-mix and the fully random model. The inbox shows a violin plot for the difference, Δ, bin by bin in the main between the d-mix model and the fully random model predictions.



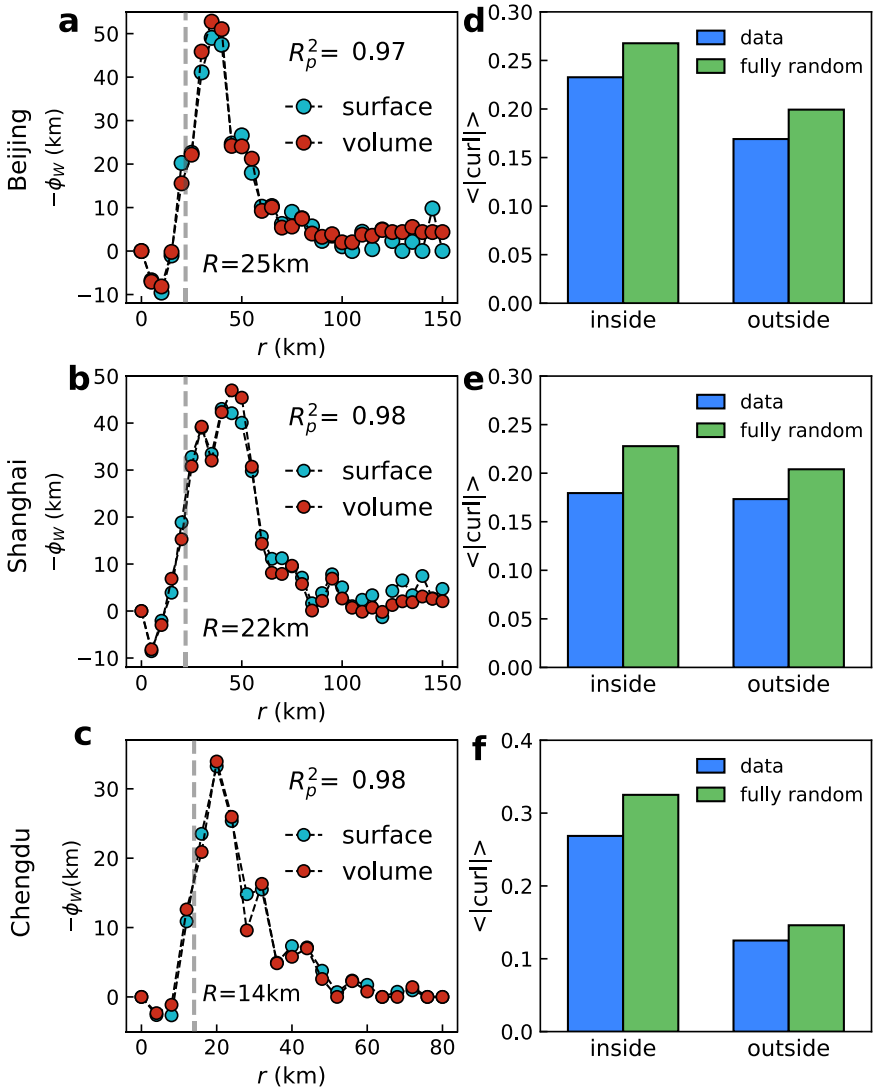


Fig. 6 | Empirical vector fields. In (**a**–**c**) fluxes of the empirical field $\vec{w}$ as a function of the distance to the city center r. In blue the surface integral flux, in red the volume integral flux, both as a function of r. R_p^2 is the coefficient of determination between the curves of the flux and the volume integral of the field divergence as a function of the distance. The verstical dashed lines mark the city radius R. **d**–**f** Average module of the curl, comparing the fully-random model and the empirical data. The separation between in and out of corresponds to the distance at which $\Phi_W = \Phi_{\max}/2$, r_c. The analysis in the outer side ends at $r = 50\ km$. After this distance, trip vectors in Beijing and Chengdu become sparse and the statistics is non-representative.

distribution of curl modules of the field generated by the d-mix has lower variance than the null model one. Furthermore, in Fig. 5d, we compare the average module of the absolute value of the curl enclosed by a circle of radius r from the city center. We see that both models coincide inside the city $r \le R = 20\ km$. However, beyond the city area the d-mix model has curl values systematically below the null model ones. This guarantees the possibility to define a potential out of the city for the d-mix mobility field.

Empirical mobility fields

In Fig. 6a–c, we see that the empirical fields generated in the three cities used as example fulfill the Gauss' Theorem as well. Results for further cities are included in Supplementary Note 8 (Fig. 17). The flux profiles as a function of r show some common features: a first r interval with positive Φ_W values (vector field pointing mainly outwards). This area is the core of the city, where the logistic trucks drop goods while most of the origins of the trajectories lie further in the peri-urban area. Despite its positive character, the existence of a net flux is relevant because the agreement between both integrals show that the Divergence Theorem is fulfilled in all the space. Further, the flux becomes negative (vector field pointing on average inwards) and it reaches the half height value at $r_c = 25$ km for Beijing, $r_c = 22\ km$ for Shanghai and $r_c = 14\ km$ for Chengdu. In the Foursquare check-in records, we also observed an initial interval (r) characterized by positive values of Φ_W. Subsequently, the flux becomes negative and it reaches the half height value at $r_c = 4\ km$ (see Supplementary Figs. 22 and 23 in Supplementary Note 11). Since for the d-mix model the city size approximately coincides with the point at which $\Phi_W(R) \approx \Phi_{\max}/2$, we will take r_c as arbitrary radii of the three cities logistic cores. The fluxes peak further away (~35 km in Beijing, ~50 km in Shanghai and ~20 km in Chengdu) and eventually Φ_W decays as r increases. These general features compare well with those of the d-mix model (Fig. 6a), except for the inside city fluxes, which are not well captured by the model.

We display in Fig. 6d–f the average absolute value of the curl enclosed and excluded by a circle of radius r_c, i.e. internal and external respectively, to the cities. We find that the empirical curl is systematically lower than the fully-random counterpart in all cases. This implies that empirical potentials can be defined in all the space. Similar results are attained with other cities of D1 (see Supplementary Note 8, Supplementary Fig. 17) and with the Foursquare check-in data in New York City (see in Supplementary Note 10 and Supplementary Figs. 20–21).

Potentials

Knowing that we can define a potential for the d-mix model (out of the city) and also for the empirical data everywhere, we plot next the equipotential curves on the maps (Fig. 7). The potential of the d-mix model shows a circular symmetry. This is due to the circular city shape introduced (Fig. 7a) and the isotropic assumption. The contours of the empirical urban areas are dependent on the city shape (Fig. 7b–d), adapting to geographical constraints as in the case of Shanghai with the sea and islands. It is also interesting how the potential highlights the presence of satellite cities as occurs for Beijing and Shanghai. The potential contours plotted extend some tens

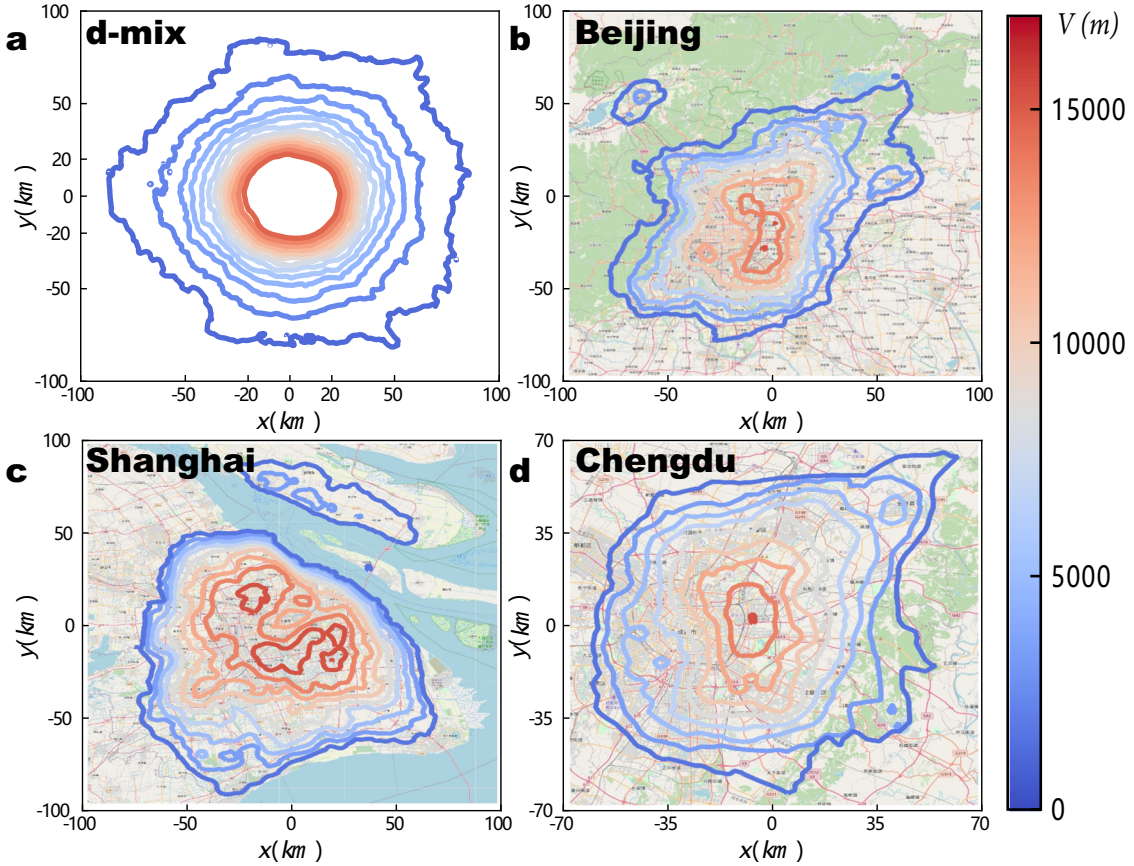


Fig. 7 | Equipotential contours. a Contours for the d-mix model with 100 000 trajectories in a circular city of radius $R = 20\ km$ and in a box of side $L = 400\ km$. Empirical equipotential contours for (**b**) Beijing, (**c**) Shanghai and (**d**) Chengdu. Maps in (**b**, **c** and **d**) were generated using Mapnik based on Openstreetmap layouts for the background under ODbL license.

of kilometers outside the cities, eventually becoming fuzzy. The field continues beyond that point, but given the lack of statistics it is hard to extract meaningful potential contours.

Hybrid d-mix model

We saw how the unbalance ratio of trajectories orientation ρ does not depend on the urban shape when using trajectory origins as RP. This is critical since it allowed us to study the origin of the mobility fields with simplified models. However, the spatial shape of the fields and potentials are inherently connected to the real city configuration. To explore further the validity of the model assumptions, we need to make another step and introduce a hybrid d-mix model. We consider the empirical trajectories one by one, e.g., $\{\vec{x}_0, \vec{x}_1, ..., \vec{x}_{s-1}\}$, keeping $\vec{x}_0$ as origin, but randomizing the order of the other stops. We then input these trajectories to the d-mix model, reordering them according to the model rules. The resulting trajectories are not necessarily equal to the empirical original ones, although if the model is doing a good work they should be similar. Indeed, we have a coincidence of 79% in Beijing, 71% in Shanghai and 79% for Chengdu, and for the Foursquare check-in records, we have a coincidence of 71% (see Supplementary Note 11). The question is thus whether this over 20% mismatch has mesoscopic effects on the field or not.

The potential estimated from the hybrid d-mix model and the empirical one are compared for the three cities in Fig. 8a–c. We find a good agreement with R_P^2 above 0.93 for all the three cities. 3D profiles of the potential are displayed in Fig. 8d–f for the empirical fields and in Fig. 8g–i for the field generated by the hybrid d-mix model. In the Foursquare check-in records, we found robust agreement with R_P^2 exceeding 0.9, we present the 3D profiles of the potential in Supplementary Fig. 23 (see Supplementary Note 11). One can clearly appreciate the similarity between modeled and empirical potentials of the same city. This implies that the hybrid d-mix model is capturing well the mechanisms behind the empirical mobility fields and its utility may go beyond its use as explanatory tool for individual behavior as above. The major deviations occur in the largest potential values, close to the maxima and the city centers where the d-mix potential is undefined and for which the model has not been fitted.

Discussion

In this work we have introduced a way to define a mobility field starting from individual trajectories. This is a generalization with respect to previous works based on Origin-Destination commuting matrices, and it allows us to study a wide range of mobility data. As illustrative examples, we have studied with this framework mobility data from two different sources: logistic routes of trucks around and across the 21 largest Chinese cities and Foursquare check-ins in NYC. In all cases, we have found a well-behaved field fulfilling the Gauss Divergence Theorem and with a curl value that it is in general smaller than the one expected by a fully random model. This implies that it is possible to define a potential almost anywhere in metropolitan areas and, consequently, to search for a source for the mobility field. This observation opens up intriguing possibilities for practical applications, suggesting that the mobility potential could serve as a foundation for optimizing mobility strategies within urban environments. The existence of a consistent and well-behaved field offers promising avenues for further exploration and refinement of mobility models.

Starting from individual behavioral assumptions of spatial exploration, we have advanced in the conceptual framework by analyzing the basic ingredients needed to generate mesoscopic mobility fields with features matching those of the empirical ones. We have introduced a metric, the unbalance ratio ρ, to characterize the fraction of displacements in

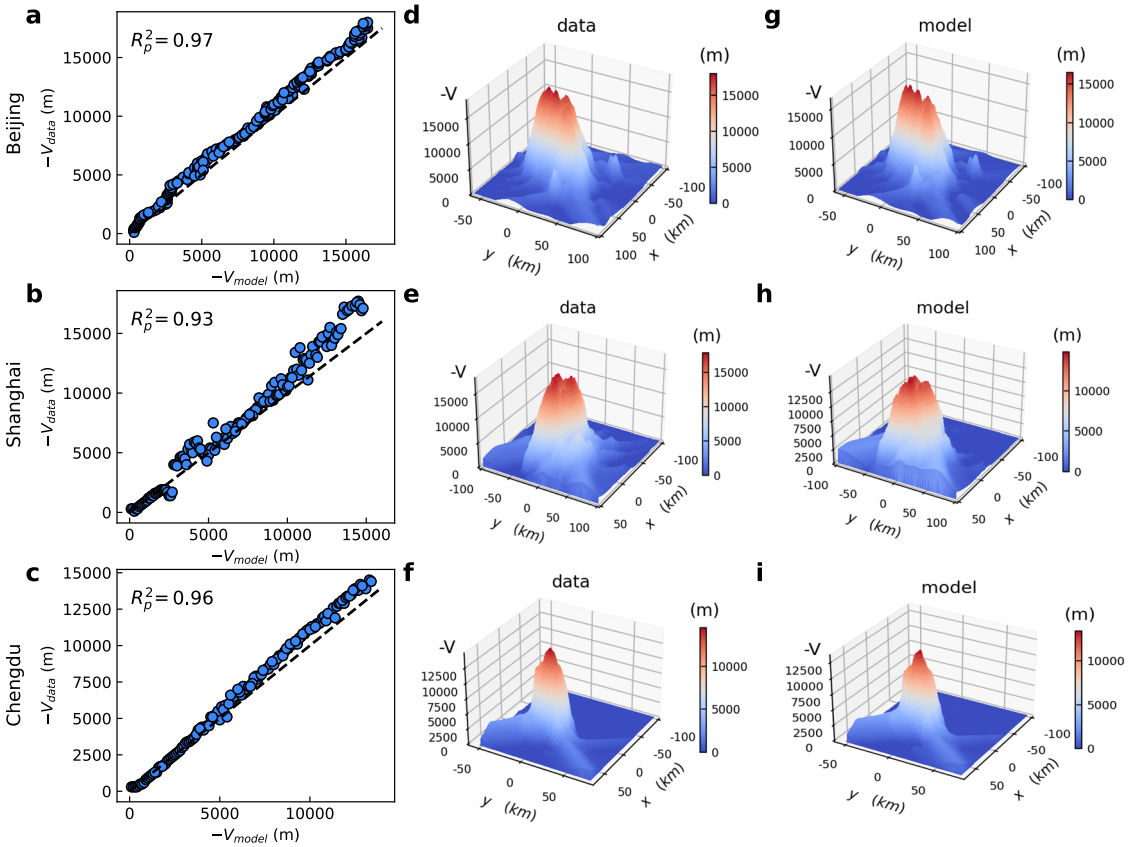


Fig. 8 | Hybrid d-mix model predictions of empirical potentials. a–c Correlation plots between the hybrid d-mix model and the empirical potentials, yielding $R_p{}^2 = 0.97$ for Beijing, $R_p{}^2 = 0.93$ for Shanghai and $R_p{}^2 = 0.96$ for Chengdu. **d–f** Empirical potential in the space and (**g–i**) the hybrid d-mix model predictions. Both models and empirical flows show the polycentric nature of Beijing and Shanghai, while Chengdu is more monocentric.

trajectories that move mostly towards or away from a reference point RP (city center or origin of the trajectory). The unbalance among these directions (ρ far from one) implies a net displacement direction and induces a mobility field. This metric allows us to quantify the strength of the factors leading to the formation of the field. We have then introduced a set of minimal models with growing complexity to explore what information is fundamental to generate the field. All the models are based on trajectories, and so the basic components are: an origin $\vec{x}_0$ and a sequence of stops $\{\vec{x}_1, \ldots \vec{x}_{s-1}\}$, with s representing the total number of locations visited in the trajectory. The simplest model, a random selection of the origin within the city and random location for the stops, is not able to generate a field. We have then added an ingredient: stops randomly extracted following a decaying distribution of the distance to the city center, and built the d-rand model. This model generates trajectories with an unbalance ratio less than the unit and, consequently, a mobility field. However, it is not able to reproduce the empirical values of ρ (its ρ is smaller). To approach realistic values, one must include the fact that individuals aim at optimizing their trajectories, reordering the sequence of stops $\{\vec{x}_0, \vec{x}_1, \ldots \vec{x}_{s-1}\}$ to reduce the total distance traveled. Mimicking this process, we added to d-rand a Traveling Salesman Problem solver to reduce total distance, and built the d-TSP model. This model generates a field with ρ closer to the empirical one than the produced by d-rand model, but still higher. The assumption of rational optimization of all trajectories does not hold for short distance trips. For this reason we introduced the d-mix model, which optimizes stops only if the average displacement between them is larger than a given threshold ℓ_c. d-mix model interpolates between both behaviors, and can be tuned on ℓ_c to generate trajectories with a ρ consistent with the empirical ones. d-mix model is not only able to reproduce ρ, but most of its field features are realistic as well. The Gauss Divergence Theorem is fulfilled and the curl of the field is smaller than in a fully random field. The model also generates a potential for the field, which, however is based on isotropic assumptions and hence, may differ from the empirical ones due to urban shape and natural constraints. Moreover, the model assumes an isotropic circular city setting. This limitation is overcome by the introduction of a hybrid d-mix model informed with real but randomized trajectories stops from the data. By letting the d-mix rules apply to reorder them, this hybrid model is able to reproduce the spatial shape of the empirical fields and potentials.

This work advances on the understanding of how the field theory can be applied to the mesoscopic scales of human mobility. Field theory is a fundamental tool in physics with a well equipped set of mathematical results developed for its use, which we hope can be translated to mobility studies in the near future. The established mathematical framework of field theory holds the potential to unlock new approaches for modeling and analyzing complex mobility patterns, contributing to a deeper and more nuanced comprehension of human movement dynamics. Additionally, urban poly-centrism and predominant patterns among mobility centers have been recently the focus of many studies due to their association to life quality indexes, city livability[12], walkability, sustainability, services accessibility and epidemic outbreak susceptibility[32]. The potential provides a clear representation of the structure of a city at a mesoscopic scale. It captures the spatial organization and connectivity patterns of mobility centers, offering insights into the distribution of activities, resources and flows. This information can help urban planners to take more informed decisions.

Methods

Mobility data and processing

The empirical results of this work are based on two datasets: the first is a truck travel records (D1)[63,64], and the second is the check-in records of Foursquare (D2)[65].

The D1 dataset includes data from over 20 Chinese cities, which can be downloaded from the National Road Freight Supervision and Service Platform (https://www.gghypt.net/). This platform is used to record the real-time geographic locations of all heavy trucks in China and monitor the potential traffic threats. The dataset contains >2.7 million travel records, spanning from May 18, 2018 to May 31, 2018. The attributes of one travel record include truck ID, timestamp, longitude, latitude and speed (see Supplementary Note 1 with Supplementary Figs. 1 and 2 and Supplementary Tables 1 and 2 for more details about the dataset used in this study). Details on the data treatment are provided in the references[63,64].

The D2 dataset is from New York city[65]. Foursquare is a location-based social network on which users share their coordinates when check-in in (see Supplementary Note 1 for more details about the dataset). This dataset contains 42035 individuals, in which 23520 users have trips among different analysis zones (here the space is divided according 2010 census areas, see https://www.census.gov/geo/mapsdata/maps/block/2010/).

Origin of the trajectories

In our datasets, we have records containing a series of times and stop (or checkin) locations for the different agents (trucks or individuals). Since trucks frequently return to their base point, due to the driver ending their work shift or reloading to start the next round of freight distribution. In view of this, we assume that the most visited location, the base point, is the origin of the trajectories that restart every time the agent visit this point. This location is also the RP in the origin-based framework. In D1 this is identified as the truck most commonly visited location with the longest stay times, since this is likely to be the logistic center of operations. In D2, we assign individuals' origins as the location with the largest number of check-ins. In both D1 and D2, a single trajectory is the defined as the sequence of stops occurring between the first and the next stop in the origin. The statistical description of the D1 trajectories in several cities of D1 are provided in Supplementary Note 2 (Supplementary Fig. 3 and Supplementary Table 3), and the same for D2 in NYC in the Supplementary Fig. 4 and Supplementary Table 3.

Models

The basic ingredients of our models will be inspired by the structure and statistics of the empirical trajectories. Supplementary Fig. 3 shows the complementary cumulative distributions of three variables associated to trajectories starting in a circle of radius 20 km centered at Beijing, Shanghai and Chengdu as paradigmatic examples. The first distribution, $D(r)$, refers to distance of the trajectory stops to the city centers (Supplementary Fig. 3a–c). The city centers are the barycenters of the areas considered (see Supplementary Table 2). As the figure shows, the location of the stops can be relatively far away from the city center, with the distribution falling slowly to the thousands of kilometers. We will adopt in the models a probability $D(r)$ of finding a stop at a certain distance r of the city center that on very first approximation will fall as $D(r) \sim r^{-\alpha}$. The next distribution, Supplementary Fig. 3d–f, refers to the number of stops per trajectory $N_s(s)$. The minimum number of stops is 2, because we are counting at least origin and final destination. The range of values is relatively limited, up to 40, and presents a decay that we will approach in the models by $N_s(s) \sim s^{-\beta}$. Finally, Supplementary Fig. 3g–i shows the distribution of the number of trajectories starting from the same origin $N_t(t)$. Truck fleets may have an operation center from which several vehicles leave, or simply the same truck appears in several trajectories starting always from the same origin. The distribution is wide, reaching more than one thousand trajectories and we will approximate it by $N_t(t) \sim t^{-\gamma}$. The distributions for the Foursquare check-in in New York City can be found in Supplementary Fig. 4.

If we consider as in Fig. 1a circular city of radius R inside of a space limited by a square of side L, we can build a first null model by selecting the location of the origin of a trajectory at random in the internal circle $\vec{x}_o$. We call this model the Rand model. The number of trajectories departing from $\vec{x}_o$ is then obtained as a stochastic extraction of $N_t(t)$, while for each of them the number of stops s can be extracted from $N_s(s)$. The location of the $s-1$ consequent stops is randomly chosen within the bounding square. We take a setting with a circular city of radius $R = 20\ km$ centered in a squared area of side $L = 400\ km$. This is to be considered the general setting on which we run all our models. The model is run to generate 100,000 trajectories and with them produce a field $\vec{W}$ following the recipe of Fig. 1. In each perimeter position, we observe that the flux of the field as a function of r fluctuates around zero inside the city circle R and it only gets negative as r reaches close to the bounding box L (see Supplementary Fig. 13). Such negative net flux is only a finite-size effect as can be seen by increasing the box size L and also by using periodic boundary conditions in the bounding box instead of open ones (Supplementary Fig. 14).

d-rand model. There are, therefore, missing ingredients in this basic model to be able to generate a stable mobility field. The first mechanism that we are going to consider is a spatial distribution of stops falling with the distance to the city center $D(r)$ as the one observed in Supplementary Fig. 3a ($D(r) = 1.2\, r^{-2.2}$ for $r \geq 1\ km$, if $r < 1$ km it is uniform ($D(r) = \frac{1}{R^2}$)). This model will be called d-rand (Fig. 3a) and it consists in randomly extracting, as before, a location in the circle containing the city for $\vec{x}_0$, the number of trajectories starting at $\vec{x}_0$ from $N_t(t)$ and the number of stops per trajectory s from $N_s(s)$. Then, for each trajectory we choose at random with $D(r)$ the radius of the location of the $s-1$ stops besides the origin, the directions from the center in which every stop lies is also randomly selected. A trajectory is thus formed by the origin and all the other stops $\{\vec{x}_0, \vec{x}_1, ..., \vec{x}_{s-1}\}$. As we will see, this model is able to produce unbalanced trajectories and a field.

d-rand has, however, a major caveat: consecutive stops can be at opposite sides of the city and it is unrealistic to have a driver passing back and forth through the city center without grouping nearby stops to reduce the total distance traveled and the fuel consumed.

d-TSP model. The next model to consider, called d-TSP (Fig. 3b), corresponds to the effect of a manager looking at the sequence $\{\vec{x}_0, \vec{x}_1, ... \vec{x}_{s-1}\}$ obtained as in the d-rand and reordering the sequence of stops from 1 to $s-1$ to minimize the total trajectory distance. This process can be mapped into the well known traveling salesman problem (TSP), in which a salesman needs to visit a set of locations, each location is visited once and only once, and finally must return to the starting position[66]. We employ in the d-TSP an heuristic algorithm (genetic algorithm[67]) developed to approximate the solution of the traveling salesman problem.

d-mix model. Finally, we introduce a model that interpolates between d-rand and d-TSP. We will call this model d-mix and the rules are as illustrated in Fig. 3c. The TSP reordering of stops is only allowed if the average travel distance of one trajectory between stops is larger than a threshold ℓ_c. The idea behind d-mix is that the driver will not invest the effort of optimizing the trajectory if the distance between consecutive stops is very short. The limit of d-mix model for small ℓ_c corresponds thus to d-TSP model, while for large ℓ_c it becomes d-rand model.

Numerical calculation of the flux

The definition of the flux is

$$\Phi_w^s = \oint d\ell\, \vec{\mathbf{n}}\, \vec{\mathbf{W}}, \tag{2}$$

where the integral is over the surface (perimeter S) enclosing the area of interest, $d\ell$ is the element of surface, $\vec{n}$ the unit vector normal to the surface and $\vec{W}$ the vector field. From a numerical perspective, the integrals are

calculated as

$$\Phi_w^s = \sum_{i\in S} d\ell\, \vec{\mathbf{n}}_i\, \vec{\mathbf{W}}_i, \tag{3}$$

where the index i runs over all the cells intersecting the perimeter S, $\vec{n}_i$ is the unit vector normal to the surface in i and $d\ell$ is approximated by the total perimeter of S divided by the number of intersecting cells.

The definition of the integral of the divergence is

$$\Phi_w^v = \int dV\, \nabla\vec{\mathbf{W}}, \tag{4}$$

where the integral is now of volume (surface in 2D of the enclosed area). From a numerical perspective, the integrals are calculated as

$$\Phi_w^v = \sum_{i\in V}\left(\frac{W_{x_{(\alpha+1,\beta)}} - W_{x_{(\alpha,\beta)}}}{\Delta x} + \frac{W_{y_{(\alpha,\beta+1)}} - W_{y_{(\alpha,\beta)}}}{\Delta y}\right)\Delta V, \tag{5}$$

where the index i runs over the cells in the volume V and ΔV is approximated by the area of the unit cell. W_x and W_y are the x and y components of the vector $\vec{W}$. The indices (α, β) refer to the position of cell i in the grid, in such a way that, for instance, $(\alpha \pm 1, \beta)$ are the positions of the adjacent cells to i in the x-direction. Δx and Δy are side sizes of the cells in the $x-$ and $y-$ directions.

Numerical calculation of the curl

The curl of $\vec{\mathbf{W}}$ in the cell i, whose indices in the $x-$ and $y-$ directions are (α, β), as above, is determined as:

$$\nabla \times \vec{\mathbf{W}} = \frac{W_{y_{(\alpha+1,\beta)}} - W_{y_{(\alpha-1,\beta)}}}{2\Delta x} - \frac{W_{x_{(\alpha,\beta+1)}} - W_{x_{(\alpha,\beta-1)}}}{2\Delta y}. \tag{6}$$

Numerical calculation of the potential

The potential is calculated by numerically solving the equations $-\nabla V = \vec{\mathbf{W}}$, taking into account that $\nabla \times \vec{\mathbf{W}} = 0$. For the computation of the empirical potential, we used conditions $V = 0$ in all the boundary regions of the box and then use the forward centered discretization formula for the gradient operator[54,68] starting from the city bounding box corner. In a cell i with indices (α, β)

$$\frac{dV_i}{dx} = \frac{V_{(\alpha+1,\beta)} - V_{(\alpha,\beta)}}{\Delta x} = W_{x_{(\alpha,\beta)}}, \tag{7}$$

and also

$$\frac{dV_i}{dy} = \frac{V_{(\alpha,\beta+1)} - V_{(\alpha,\beta)}}{\Delta y} = W_{y_{(\alpha,\beta)}}, \tag{8}$$

the procedure is iterated until all cells have been assigned a potential. To decrease the noise, we average the resulting potentials starting from the four corners of the bounding box .

Parameter estimation

In our d-mix model, the single free parameter is ℓ_c, which directly determines whether the agents optimize or not the order of the trajectory stops. For a given empirical dataset, we rely on ρ to estimate ℓ_c. To accomplish this, we define the following function

$$E(\ell_c) = |\rho_{data} - \rho(\ell_c)|, \tag{9}$$

where ρ_{data} is obtained from the dataset for all trajectories, and $\rho(\ell_c)$ is calculated through the d-mix model with parameter ℓ_c. The objective function can be minimized to yield an estimated value of ℓ_c needed to reproduce with d-mix the signs of the set of empirical trajectories.

Data availability

The D1 dataset on truck trajectories in China can be downloaded from the National Road Freight Supervision and Service Platform (https://www.gghypt.net/). The D2 dataset on Foursquare check-ins in New York City was obtained from the details given in ref. 65. The 2010 census divisions in the case of New York were downloaded from https://www.census.gov/geo/mapsdata/maps/block/2010/.

Code availability

The code used for this work is available at https://github.com/orgs/erjianliu.

Received: 25 August 2023; Accepted: 22 May 2024;
Published online: 13 June 2024

References

1. Wilson, A. *Entropy In Urban And Regional Modelling* 1st edn, Vol. 166 (Routledge, 1970).
2. Bergstrand, J. H. The gravity equation in international trade: some microeconomic foundations and empirical evidence. *Rev. Econ. Stat.* **67**, 474–481 (1985).
3. Karemera, D., Oguledo, V. I. & Davis, B. A gravity model analysis of international migration to North America. *Appl. Econ.* **32**, 1745–1755 (2000).
4. Roy, J. R. & Thill, J.-C. Spatial interaction modelling. *Pap. Reg. Sci.* **83**, 339–361 (2003).
5. Rouwendal, J. & Nijkamp, P. Living in two worlds: a review of home-to-work decisions. *Growth Change* **35**, 287–303 (2004).
6. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
7. Carra, G., Mulalic, I., Fosgerau, M. & Barthelemy, M. Modelling the relation between income and commuting distance. *J. R. Soc. Interface* **13**, 20160306 (2016).
8. Barbosa, H. et al. Human mobility: models and applications. *Phys. Rep.* **734**, 1–74 (2018).
9. Batty, M. *The New Science Of Cities* Reprint 520 end (MIT Press, Cambridge MA, USA, 2013).
10. Barthelemy, M. *The Structure And Dynamics Of Cities: Urban Data Analysis And Theoretical Modeling* 1st ed, Vol. 278 (Cambridge University Press, 2017).
11. Li, R. et al. Simple spatial scaling rules behind complex cities. *Nat. Commun.* **8**, 1841 (2017).
12. Bassolas, A. et al. Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 4817 (2019).
13. Zipf, G. K. The p 1 p 2/d hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* **11**, 677–686 (1946).
14. Stouffer, S. A. Intervening opportunities: a theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
15. Kitamura, R., Chen, C., Pendyala, R. M. & Narayanan, R. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation* **27**, 25–51 (2000).
16. Ortúzar, J. and Willumsen, L. *Modeling Transport* 4th edn, 607 (John Wiley and Sons Ltd., 2010).
17. Deville, P. et al. Dynamic population mapping using mobile phone data. *Proc. Natl Acad. Sci. (USA)* **111**, 15888–15893 (2014).
18. Louail, T. et al. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 5276 (2014).
19. Louail, T. et al. Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**, 6007 (2015).
20. Verbavatz, V. & Barthelemy, M. The growth equation of cities. *Nature* **587**, 397–401 (2020).
21. Pappalardo, L., Ferres, L., Sacasa, M., Cattuto, C. & Bravo, L. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Sci.* **10**, 29 (2021).

22. Xu, Y., Belyi, A., Bojic, I. & Ratti, C. Human mobility and socioeconomic status: analysis of singapore and boston. *Comp. Environ. Urban Syst.* **72**, 51–67 (2018).
23. Barbosa, H. et al. Uncovering the socioeconomic facets of human mobility. *Sci. Rep.* **11**, 8616 (2021).
24. Mimar, S. et al. Connecting intercity mobility with urban welfare. *PNAS Nexus* **1**, pgac178 (2022).
25. Scellato, S., Noulas, A. & Mascolo, C. Exploiting place features in link prediction on location-based social networks. In *Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 1046–1054,* https://doi.org/10.1145/2020408.2020575 (ACM, 2011).
26. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
27. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484–21489 (2009).
28. Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* **7**, 581 (2011).
29. Tizzoni, M. et al. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10**, e1003716 (2014).
30. Jia, J. S. et al. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
31. Mazzoli, M. et al. Interplay between mobility, multi-seeding and lockdowns shapes COVID-19 local impact. *PLoS Comput. Biol.* **17**, e1009326 (2021).
32. Aguilar, J. et al. Impact of urban structure on infectious disease spreading. *Sci. Rep.* **12**, 3816 (2022).
33. Boyce, D. E. & Williams, H. C. *Forecasting Urban Travel: Past, Present and Future* Reprint edn, 672 (Edward Elgar Publishing, 2015).
34. Levinson, D. & Kumar, A. Activity, travel, and the allocation of time. *J. Am. Plan. Assoc.* **61**, 458–470 (1995).
35. Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. & Haupt, T. Observing the rhythms of daily life: a six-week travel diary. *Transportation* **29**, 95–124 (2002).
36. González, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779 (2008).
37. Bagrow, J. P. & Lin, Y.-R. Mesoscopic structure and social aspects of human mobility. *PLoS One* **7**, e37676 (2012).
38. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PLoS One* **7**, e37027 (2012).
39. Lenormand, M. et al. Cross-checking different sources of mobility information. *PLoS One* **9**, e105184 (2014).
40. Hawelka, B. et al. Geo-located twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **41**, 260–271 (2014).
41. Lenormand, M., Gonçalves, B., Tugores, A. & Ramasco, J. J. Human diffusion and city influence. *J. R. Soc. Interface* **12**, 20150473 (2015).
42. Blondel, V. D., Decuyper, A. & Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **4**, 10 (2015).
43. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
44. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
45. Alessandretti, L., Aslak, U. & Lehmann, S. The scales of human mobility. *Nature* **587**, 402–407 (2020).
46. Carey, H. C. *Principles of Social Science*, vol 3 (JB Lippincott & Company,1867).
47. Ruiter, E. R. Toward a better understanding of the intervening opportunities model. *Transp. Res.* **1**, 47–56 (1967).
48. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
49. Yan, X.-Y., Zhao, C., Fan, Y., Di, Z. & Wang, W.-X. Universal predictability of mobility patterns in cities. *J. R. Soc. Interface* **11**, 20140834 (2014).
50. Lenormand, M., Bassolas, A. & Ramasco, J. J. Systematic comparison of trip distribution laws and models. *J. Transp. Geogr.* **51**, 158–169 (2016).
51. Liu, E. & Yan, X.-Y. A universal opportunity model for human mobility. *Sci. Rep.* **10**, 4657 (2020).
52. Simini, F., Barlacchi, G., Luca, M. & Pappalardo, L. A deep gravity model for mobility flows generation. *Nat. Commun.* **12**, 6576 (2021).
53. Yan, X.-Y. & Zhou, T. Destination choice game: a spatial interaction theory on human mobility. *Sci. Rep.* **9**, 9466 (2019).
54. Mazzoli, M. et al. Field theory for recurrent mobility. *Nat. Commun.* **10**, 3895 (2019).
55. Steward, J. Q. Empirical mathematical rules concerning the distribution and equilibrium of population. *Am. Geogr. Soc.* **37**, 461–485 (1947).
56. Mukherji, S. *The Mobility Field Theory of Human Spatial Behavior: A Behavioral Approach To The Study of Migration and Circulation in The Indian Situation* https://scholarspace.manoa.hawaii.edu/items/4319b924-6fe5-4486-b2ae2c8d317ec8fc (The University of Michigan, 1975).
57. Wang, J., Ji, J., Jiang, Z. & Sun, L. Traffic flow prediction based on spatiotemporal potential energy fields. *IEEE Transactions on Knowledge and Data Engineering* **35**, 9073–9087 (2022).
58. Aoki, T., Fujishima, S. & Fujiwara, N. Urban spatial structures from human flow by Hodge-Kodaira decomposition. *Sci. Rep.* **12**, 11258 (2021).
59. Shida, Y., Takayasu, H., Havlin, S. & Takayasu, M. Universal scaling laws of collective human flow patterns in urban regions. *Sci. Rep.* **10**, 21405 (2020).
60. Shida, Y., Ozaki, J., Takayasu, H. & Takayasu, M. Potential fields and fluctuation-dissipation relations derived from human flow in urban areas modeled by a network of electric circuits. *Scientific Reports* **12**, 9918 (2022).
61. Bongiorno, C. et al. Vector-based pedestrian navigation in cities. *Nat. Comput. Sci.* **1**, 678–685 (2021).
62. Yang, H., Li, M., Guo, B., Zhang, F. & Wang, P. A vector field approach for identifying anomalous human mobility. *IET Intell. Transp. Syst.* **n/a**, 1–18 (2022).
63. Yang, Y. et al. Identifying intracity freight trip ends from heavy truck GPS trajectories. *Transp. Res. Part C: Emerg. Technol.* **136**, 103564 (2022).
64. Yang, Y. et al. Identifying intercity freight trip ends of heavy trucks from GPS data. *Transp. Res. Part E: Logist. Transp. Rev.* **157**, 102590 (2022).
65. Bao, J., Zheng, Y. & Mokbel, M. F. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proc. 20th International Conference on Advances in Geographic Information Systems, Pages 199–208,* https://doi.org/10.1145/2424321.2424348 (ACM, 2012).
66. Bellman, R. Dynamic programming treatment of the travelling salesman problem. *J. ACM (JACM)* **9**, 61–63 (1962).
67. Weile, D. S. & Michielssen, E. Genetic algorithm optimization applied to electromagnetics: a review. *IEEE Trans. Antennas Propag.* **45**, 343–353 (1997).
68. Hyman, J. M. & Shashkov, M. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comp. Math. Appl.* **33**, 81–104 (1997).

Acknowledgements

We would like to thank Xin Lu for a critical reading and useful suggestions on the manuscript. E.L. and J.J.R. acknowledge funding from MICIU/AEI/10.13039/501100011033/FEDER/EU under project APASOS (PID2021-122256NB-C22) and from MICIU/AEI/10.13039/501100011033 under project Next4Mob (PLEC2021-007824) and the Maria de Maeztu Program for units of Excellence in R&D CEX2021-001164-M. X.-Y.Y. was supported by the National Natural Science Foundation of China (Grant No. 72271019).

E.L. also acknowledges funding from the National Natural Science Foundation of China (72271019, 72288101) and from the Central Leading Local Science and Technology Development Fund Project (236Z0805G). M.M. acknowledges support from the Lagrange Project of the Institute for Scientific Interchange Foundation (ISI Foundation) funded by Fondazione Cassa di Risparmio di Torino (Fondazione CRT). M.M. and J.J.R. also received partial support from the project EOSC-SIESTA (101131957) pof the EU Commission under the Horizon Europe Framework.

Author contributions

E.L., M.M. and J.J.R. designed the research and contributed new conceptual tools; E.L., M.M. and J.J.R. performed the research; X.-Y.Y. and E.L. collected and analyzed the empirical data; J.J.R. coordinated the study. All authors contributed to the discussion, to the writing and approved the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-024-01672-z.

Correspondence and requests for materials should be addressed to Erjian Liu or José J. Ramasco.

Peer review information *Communications Physics* thanks Vedran Sekara and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.