SI: Decomposing geographical and universal aspects of human mobility

Louis Boucherie[1,2], Benjamin F. Maier[1,2], and Sune Lehmann[1,2]

[1]DTU Compute, Technical University of Denmark
[2]Center for Social Data Science, University of Copenhagen

May 3, 2024

Contents

List of Figures

List of Tables

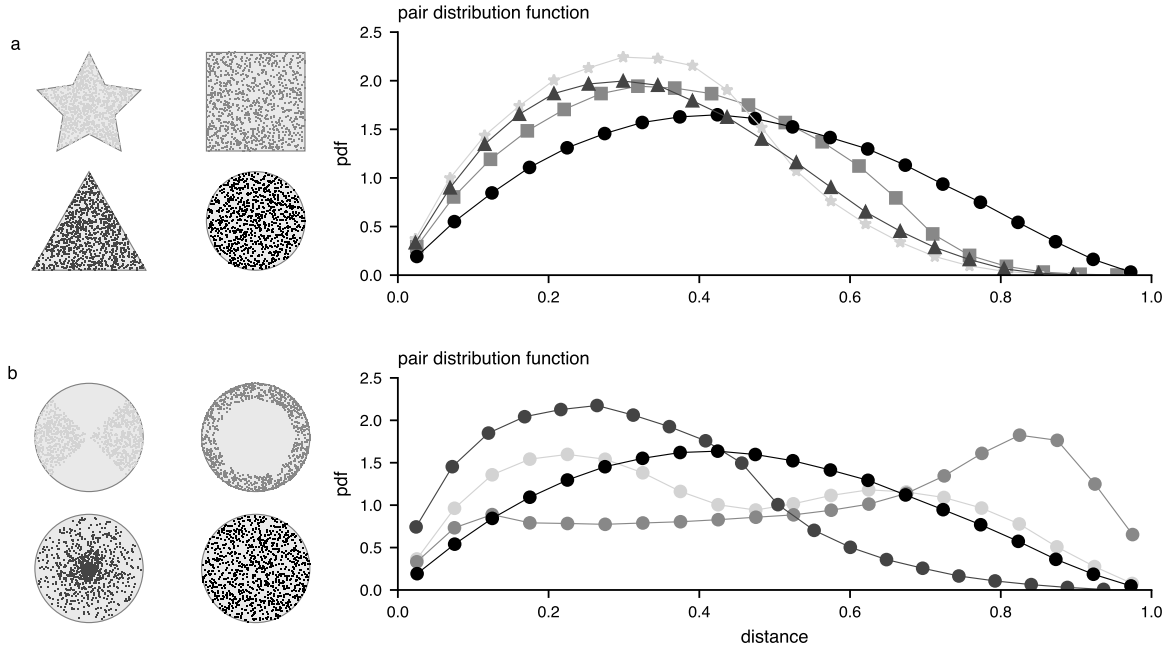

Figure 1: Example of pair distribution function for different toy geographies and density of points. **(a)** Four different shapes of landmasses and their respective pair distribution functions (same maker shape and greyscale). **(b)** For the circle shape, four different toy layouts of the built environment and their respective pair distribution function (same greyscale).

1 The concept of pair distribution function distribution

1.1 Pair distribution examples

This section presents additional examples of toy geographies that vary both the shape of the landmasses and the density distribution of points. These examples intend to demonstrate that the pair distribution distribution can encode the geography of the built environment. Figure 1a, illustrates how the pair distribution function captures the shape of uniformly distributed points. Figure 1b illustrates the impact of spatial point distribution on a fixed shape.

1.2 The pair distribution function emerges as a normalization for human mobility

Whenever points $x_i \in \Omega$ are distributed in a d-dimensional subspace $\Omega \subseteq \mathbb{R}^d$ according to some density $p^d(x)$, their pair distribution function $p(r)$ is one-dimensional and determined by both the geometry of the space as well as $p^d(x)$, given that there exists a (pseudo-) metric $r \equiv R(x, y)$ that determines a distance between two points x and y. This pair distribution is

given by

$$p(r) = \int_\Omega dx \int_\Omega dy\, p^d(x) p^d(y) \delta(r - R(x,y)). \tag{1}$$

When observations are made between two distinct points (i.e. when a quantity is counted) in the domain of interest, a distance-dependent distribution of the observed entities is observed. This distribution is hereafter referred to as the function $f(r)$. These observations can be quantified as the number of packages sent from location A to location B at distance r, the number of people moving from A to B at distance r, or the number of power line connections that connect substations at distance r in a power grid. The observed distribution function $f(r)$ is the result of two factors: the number of pairs of points that exist in the domain of interest, Ω, with associated point density $p^d(x)$, and the one-dimensional probability $\pi(r)$ with which pairs of distance r are manifested in the real world to be observed. In short, the number $f(r)$ of observations of distance r, is determined by the number of possible pairs $p(r)$ at distance r and the probability $\pi(r)$ that they would exist at this distance, i.e.

$$f(r) \propto \pi(r) p(r), \tag{2}$$

bar a normalization constant.

When we measure $f(r)$, we always measure with it the geometry of the subspace Ω as well as the density $p^d(x)$, manifested in the pair distribution function $p(r)$. The "universal", i.e. geometry-independent law that encapsulates the behavior of the system we want to study, is encoded in $\pi(r)$. To properly deduce this geometry-independent behavior of our system, we need to adjust our observation $f(r)$ by the geometry encoded in $p(r)$, i.e.

$$\pi(r) \propto \frac{f(r)}{p(r)}. \tag{3}$$

A pertinent question is which reference topology (geometry and point distribution) would result in an observation f that is directly proportional to the behavior π. This question holds practical significance, as considering behavior benefits greatly from a conceptual framework for the distribution of points within a space. Indeed, our goal is to understand the mechanisms that link two points within this space. Without defining these points or the space itself, this endeavor becomes somewhat pointless.

In our formalism, this implies that we are seeking a topology where,

$$f(r) \propto \pi(r), \tag{4}$$

which implies

$$p(r) = \text{const.} \tag{5}$$

Looking at Eq. (1), we have to find Ω and $p^d(x)$ such that $p(r) = \text{const}$. While there might be a multitude of solutions, the simplest one is a one-dimensional ring, which can be conceptualized as a box of length L and domain $x \in [0, L)$ with periodic boundary conditions, i.e. an associated distance

$$R(x,y) = \begin{cases} |x-y|, & 0 \leq |x-y| \leq L/2, \\ L - |x-y| & L/2 \leq |x-y| < L \end{cases} \tag{6}$$

and uniformly distributed points, i.e.

$$p^{d=1}(x) = \begin{cases} L^{-1}, & 0 \leq x < L, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

then the pair-wise distance distribution evaluates to,

$$p(r) = L^{-2} \int_0^L dx \left(\int_{x-L/2}^{x} dy \delta(r - (x-y)) + \int_x^{x+L/2} dy \delta(r - (y-x)) \right) \tag{8}$$

$$= \begin{cases} L^{-2} \int_0^L dx \, (1+1), & 0 \leq r \leq L/2, \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$= \begin{cases} (L/2)^{-1} & 0 \leq r \leq L/2, \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

for an illustration see Fig. 2. Hence, the observed distance distribution $f(r)$ occurring between pairs of locations at distance r on this topology will be proportional to the geometry-independent behavioral part $\pi(r)$. This technique has been used in [2, 3, 4, 5].

1.3 Conditions for the pair distribution function to identify a set of points uniquely

According to [6], in general, point configurations can be uniquely determined by their pair distribution distributions, up to a rigid transformation. However. there are counterexamples, for example when two distances are equal, as in figure 3. The counterexample starts with four points (triangles are uniquely identifiable). This counterexample can be extended to any number of points, as one can add infinitely many points on the dotted line of the two examples and their respective pair distribution would be the same.

In the context of geographic analysis, for a set of 2D points representing addresses, some distances between points will inevitably be repeated. If we consider 1×10^6 points located in a 1000 km $\times$ 1000 km square and whose coordinates are known with a 1 m precision, by a simple combinatorial argument, the number of pair distances is 1×10^{12}. Although the maximum distance in the square is $\sqrt{2} \times 1000$ km, or 1.414×10^6 m, which means that the set of possible distances contains 1.414×10^6 different values (due to the limited precision),

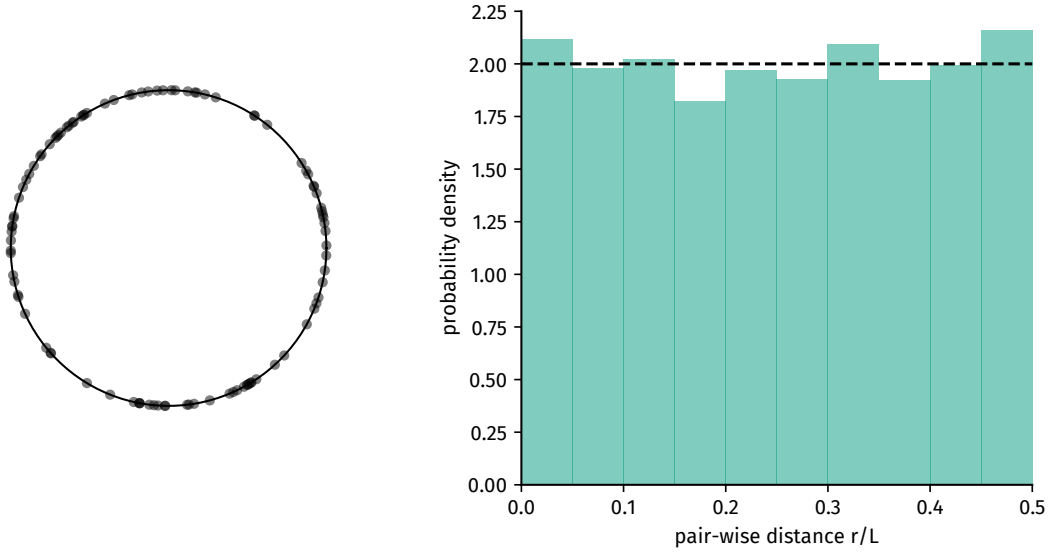

Figure 2: Given a ring with uniformly distributed points, the pair distance distribution of points will be constant.

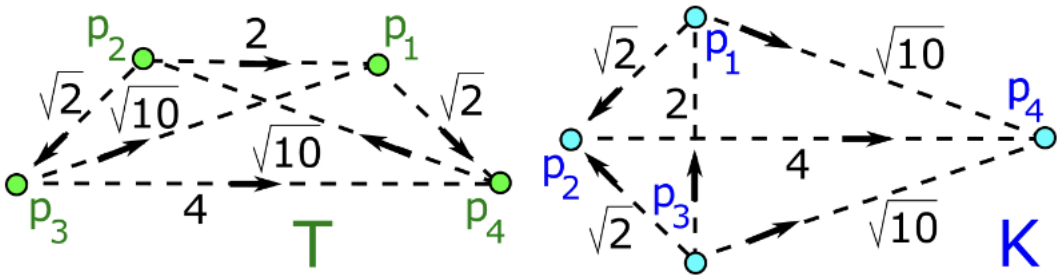

Figure 3: Example in which the pair distribution function does not uniquely identify a set of 2D points.

but there are 1×10^{12} instances, so some distances must be equal, and the pair distribution function does not uniquely identify a set of 2d points.

However, in the case of geography, due to the regularities and scaling law of the pair distribution, one can adopt a coarse-grained view of the problem. Instead of considering each point individually, we can first examine the pair distribution function between urban centers or other areas and iteratively reconstruct the geography (set of 2d points) in iterative. First, each urban center position should be a unique configuration, then independently on the local geography around each city. This should lead to a unique configuration (up to isometries that are of the second order for the pair distribution function). The coarse-grained construction would be similar to the quad-tree partitioning or the HDSCAN clustering (see section 4).

Figure 4: Map of 34,041,910 individual addresses in France **(a)**, and zoom on the Paris area **(b)**.

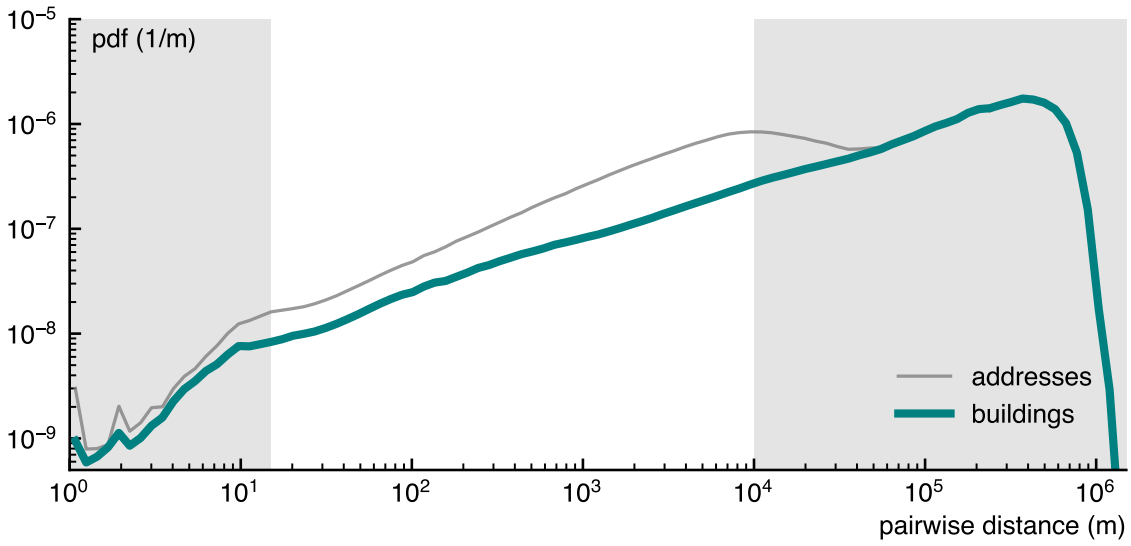

Figure 5: pair distribution of buildings and addresses in France. Neighborhood-scale with a linear onset and an oscillating modulation for the immediate neighborhood, then city-scale showing a power-law growth with exponent $\alpha = 0.67$, and finally country-scale with a slower growth and eventual fast decay.

1.4 Pair distribution function for France

The pair distribution function between residential buildings in France shows similar patterns as the one of Denmark, as shown in Fig.5. On the micro-scale (i.e. within distances of $r \lesssim 25\,\mathrm{m}$, the "neighborhood"). We observe a linear onset of neighborhood density, oscillatory modulated. For larger distances (mesoscale), this growth assumes a scaling of approximately $p(r) \propto r^{0.67}$ between 25 m and 200 m, which is even more pronounced in the distribution of address distances (up to 10000 m). The mesoscale has a larger amplitude for France than for Denmark (10 km vs 4 km)due to larger urban areas (Paris 2853 km^2 vs 526 km^2 for Copenhagen). After that, at the macro scale, this growth slows down, decays rapidly, and finally approaches zero as we reach the limit of the finite system.

2 Models for pair distribution function

2.1 Pair distribution functions

2.1.1 Generalized pair distribution function

In the main text, we defined the generalized pair distribution function model as

$$p_m(r, R) = C\frac{r}{R^2}\exp\left(-\left(\frac{r}{R}\right)^m\right). \tag{11}$$

It has a linear onset and a tail that falls as a stretched exponential and is a special case of the generalized Gamma distribution. We find C as follows. Begin with the Gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)\,\mathrm{d}t \tag{12}$$

and substitute $t = (r/R)^m$ such that $\mathrm{d}t = (m/R)(r/R)^{m-1}\mathrm{d}r$ and

$$\Gamma(z) = \frac{m}{R}\int_0^\infty (r/R)^{mz-m+m-1}\exp\left[-(r/R)^m\right]\mathrm{d}r. \tag{13}$$

Now we demand $mz - 1 = 1$, i.e. $z = 2/m$, to find

$$\Gamma(2/m) = mC^{-1}\int_0^\infty C\frac{r}{R^2}\exp\left[-(r/R)^m\right]\mathrm{d}r. \tag{14}$$

Due to the normalization condition, we have

$$C = \frac{m}{\Gamma(2/m)}. \tag{15}$$

To obtain the first moment, we demand $mz - 1 = 2$ such that $z = 3/m$ so we find

$$\Gamma(3/m) = \frac{mC^{-1}}{R}\int_0^\infty C\frac{r^2}{R^2}\exp\left[-(r/R)^m\right]\mathrm{d}r \tag{16}$$

$$= \frac{\Gamma(2/m)}{mR}\langle r\rangle \tag{17}$$

$$\langle r\rangle = \frac{mR\Gamma(3/m)}{\Gamma(2/m)}. \tag{18}$$

We can fit this distribution to data by using the per-sample log-likelihood

$$\mathcal{L} = \frac{1}{n}\ln L = \ln m - \ln\Gamma(2/m) + 2\ln w + \langle\ln r\rangle - w^m\langle r^m\rangle. \tag{19}$$

where $w = 1/R$ and we have an observational set $\{r_i\}$ of empirical pairwise distances with sample size n. From $\partial\mathcal{L}/\partial w = 0$ we find that the inverse city scale that maximizes the likelihood is given by

$$\hat{w}(m) = \left(\frac{2}{m\,\langle r^m\rangle}\right)^{1/m}. \tag{20}$$

With $\hat{w}(m)$, we can find the zero of

$$\frac{\partial\mathcal{L}}{\partial m} = \frac{1}{m} + \frac{2}{m^2}\psi(2/m) - \left\langle [\hat{w}(m)r]^m \ln[\hat{w}(m)r]\right\rangle \tag{21}$$

numerically, which gives $\hat{m}$. Here, $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

2.1.2 Circle

The pair distribution of a uniform distribution of random points within a disk of radius R is given by

$$p(r,R) = \begin{cases} \frac{4r}{\pi R^2}\arccos\left(\frac{r}{2R}\right) - \frac{2r^2}{\pi R^3}\sqrt{1-\frac{r^2}{4R^2}}, & r \le 2R, \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

see [7].

2.1.3 Parabola

Looking at Eq. (22), we see that this distribution looks somewhat close to a parabola with zeros at $r = 0$ and $r = 2R$. A parabolic pair distribution with such properties is given as

$$p(r,R) = \begin{cases} -\frac{3}{4R}\left(\frac{r}{R}\right)^2 + \frac{3r}{2R^2} & r \le 2R, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

Note that this is the Beta distribution with $\alpha = 2$ and $\beta = 2$ for random variable $x = r/R$.

2.1.4 Building locations by external and interaction potentials

Consider the location of a city as the literal center of interest, for example, the "central business district". Since it might be attractive for individuals to reach this center as quickly as possible (because amenities will be close to the center), we assume that there is an increased cost of living at a distance $|\boldsymbol{x}|$ from the center. For example, consider having to commute to a job within the central business district, which has a cost that increases with $|\boldsymbol{x}|$. At the same time, two simple mechanisms will prevent buildings from accumulating in the exact

center of the city. First, buildings have a certain average radius z, so they cannot be too close together (at a distance $< 2z$). There is an advantage to buildings not being too far apart, as they can share local amenities. Instead of modeling this explicitly, we simply assume that there is an inherent temperature T in the system, according to which house locations are distributed following an interaction and an external potential.

From a statistical physics point of view, we describe the system with the simplest external potential, which increases linearly with distance

$$V^{\text{ext}}(\boldsymbol{x}) = \gamma|\boldsymbol{x}| \tag{24}$$

where we assume that the origin of the coordinate system is in the center of the city. To model repulsion and attraction between houses, we also assume a Lennard-Jones interaction potential

$$V^{\text{int}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varepsilon \left[\left(\frac{2z}{|\boldsymbol{x}_j - \boldsymbol{x}_i|} \right)^{12} - 2 \left(\frac{2z}{|\boldsymbol{x}_j - \boldsymbol{x}_i|} \right)^{6} \right], \tag{25}$$

which is commonly used to model simultaneous attraction and repulsion between molecules in chemical solutions [8, 9].

In total, a system with these properties evolves according to the Hamiltonian

$$\mathcal{H}(\{\boldsymbol{x}, \boldsymbol{p}\}) = \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{p}_i^2 + \gamma \sum_{i=0}^{N} |\boldsymbol{x}_i| + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} V^{\text{int}}(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{26}$$

with two-dimensional momenta $\boldsymbol{p}_i$ and locations $\boldsymbol{x}_i$. In a canonical-ensemble formulation of the system, i.e. at constant inverse temperature β, the probability of finding a configuration $\{\boldsymbol{x}, \boldsymbol{p}\}$ in volume-element $d^2\boldsymbol{x}d^2\boldsymbol{p}$, is given by

$$\varrho[\{\boldsymbol{x}, \boldsymbol{p}\}] d^2\boldsymbol{x} d^2\boldsymbol{p} = \exp[-\beta \mathcal{H}(\{\boldsymbol{x}, \boldsymbol{p}\})] d^2\boldsymbol{x} d^2\boldsymbol{p}. \tag{27}$$

For now, we restrict ourselves to an ideal gas with $V^{\text{int}} = 0$, which will allow us to say something about the density of particles around the center of the city, i.e., we want to find the probability of a particle being present $p(r)dr$. Without loss of generality, we set $N = 1$, because due to ergodicity the trajectory of a particle will eventually follow the density of the whole distribution (think of it as taking the Nth root of the N particle density). Integrating over the momenta yields

$$\varrho[\boldsymbol{x}] d^2\boldsymbol{x} = \exp\left(-\beta\gamma|\boldsymbol{x}|\right) d^2\boldsymbol{x} \tag{28}$$

Changing the variables to polar coordinates, where r is the distance of the particle from the center, we find

$$\varrho[r, \phi] r dr d\phi = r \exp\left[-\beta\gamma r\right] dr d\phi, \tag{29}$$

i.e.

$$p(r)dr = r \exp\left[-\beta\gamma r\right] dr. \tag{30}$$

This means that if there are no interactions, the distribution of houses around the city center should follow an Erlang distribution with scale parameter $\lambda = \beta\gamma = 1/R_0$, where R_0 is half the city radius. Note that we have postulated that all particles have the same mass $m = 1$, so γr has the dimension of energy. The definition of the inverse temperature $\beta = 1/T$ implies that the temperature also has an energy dimension.

Relying on the arguments of the kinetic theory of gases, we can relate the temperature to the momenta of a particle with the identity

$$K = N_f N \frac{T}{2} \tag{31}$$

where $K = (1/2)\sum_i p_i^2$ is the kinetic energy and N_f is the degree of freedom of each particle, i.e. for single-atom particles in two dimensions, $N_f = 2$ (two translational degrees of freedom, no rotations, no oscillations). Note that this relates to the root-mean-square velocity of a single particle as

$$v_0 \equiv \sqrt{\langle v^2 \rangle} = \sqrt{T/2}. \tag{32}$$

In this sense, the instantaneous temperature plays the role of a particle's ability to overcome the potential energy. If $V^{\text{ext}}(r) = \gamma r$ represents a *cost* of being at distance r from the center, the temperature gives a measure of how well particles in the system can overcome that cost. If the temperature is low, particles cannot overcome this cost and the density in the city center will be high. If the temperature is high, particles can overcome the cost easily because of the larger amount of kinetic energy available in the system.

Increasing complexity by going back to the Lennard-Jones perspective, we want to obtain an intuition about how the radial particle density changes when they strongly repel each other. In the limit of $\varepsilon/T \gg 1$, particles will have a strong tendency to be found in their respective potential minimum, i.e. at distance $2z$ from each other. Effectively, we can think of them as hard disks of radius z with a tendency to form clusters. If the temperature is low, the effective radius of the city will be small (remember that the Erlang shape parameter is $\lambda = \gamma/T = 1/R_0$). In this case, it may happen that the number of particles that we would expect to lie within radius r from the center (according to the ideal gas) will be greater than the number of hard disks that can fit within a circle of radius r. When this happens, we expect a crystal to form at the center.

The maximum number of disks that can fit within a circle of size r can be approximated by

$$N_{\text{disks}}(r) = \theta A_{\text{circle}}(r)/A_{\text{disk}}. \tag{33}$$

Here, $A_{\text{circle}}(r)$ is the area of a circle of radius r, A_{disk} is the area of a circle of radius z, and θ is an optimal packing fraction, where we can approximately assume $\theta \approx 0.9$ for optimal hexagonal packing. Then,

$$N_{\text{disks}}(r) = \theta r^2/z^2. \tag{34}$$

From the ideal gas distribution, we expect to find

$$\begin{aligned} N_{\text{id}}(r) &= NP(r) && (35)\\ &= N\left(1 - \mathrm{e}^{-\gamma r/T}\left(1 + \frac{\gamma r}{T}\right)\right) && (36) \end{aligned}$$

within radius r (where $P(r)$ is the Erlang cumulative density function).

Following the aforementioned argumentation, we expect a nucleation effect when $N_{\text{id}}(r) > N_{\text{disks}}(r)$ for $gr/T \ll 1$. Linearizing the exponential factor, that happens when

$$\begin{aligned} N_{\text{id}}(r) &> N_{\text{disks}} && (37)\\ N\left(1 - \left(1 - \frac{\gamma r}{T}\right)\left(1 + \frac{\gamma r}{T}\right)\right) &> \theta\frac{r^2}{z^2} && (38)\\ N\left(1 - \left(1 - \frac{(\gamma r)^2}{T^2})\right)\right) &> \theta\frac{r^2}{z^2} && (39)\\ \frac{N(\gamma r)^2}{T^2} &> \theta\frac{r^2}{z^2} && (40)\\ \frac{N\gamma^2 z^2}{\theta T^2} &> 1, && (41) \end{aligned}$$

or in terms of the reference velocity v_0 and the Lennard-Jones distance d,

$$\frac{N\gamma^2 d^2}{8\theta v_0^2} > 1. \tag{42}$$

The radius of this cluster will approximately be given by the solution to the equation $N_{\text{id}}(r) = N_{\text{disks}}(r)$, which can be obtained numerically.

2.1.5 Molecular Dynamics simulation

We are interested in finding configurations $\mathcal{C}$ that accurately represent the canonical ensemble with number of particles N, constant but irrelevant volume [1], and average-constant temperature T with total potential energy of

$$V = \sum_{i=1}^{N}\left[V^{\text{ext}}(\boldsymbol{x}_i) + \frac{1}{2}\sum_{j\neq i} V^{\text{int}}(\boldsymbol{x}_i, \boldsymbol{x}_j)\right]. \tag{43}$$

[1] because we force the particles to be confined within a radially symmetric external potential, the total volume containing the particles does not matter.

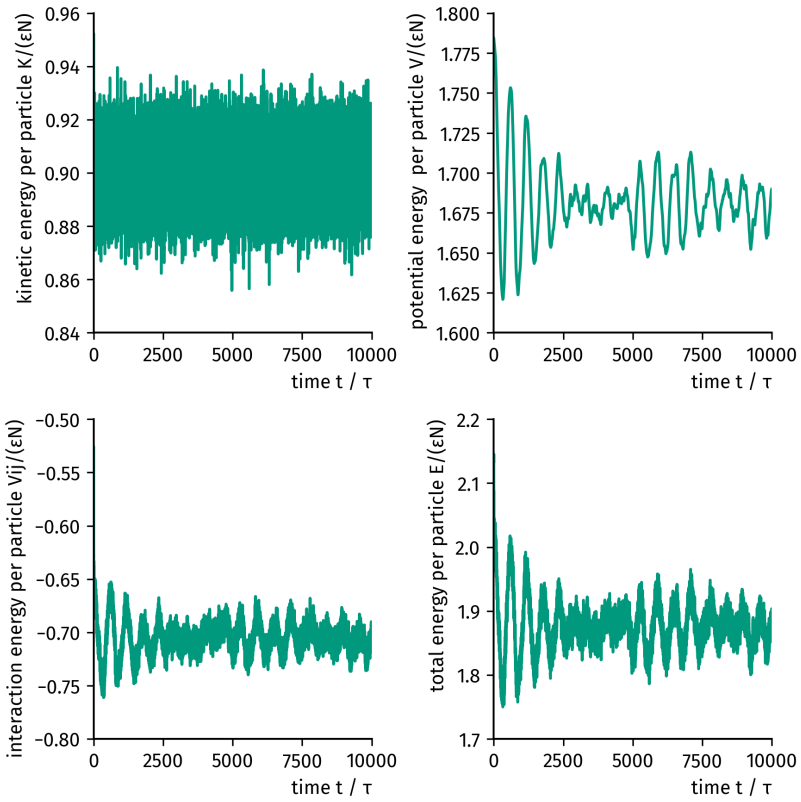

Figure 6: Energy evolution of a single MD simulation of a canonical ensemble of LJ particles in a linear external potential, thermalized by a stochastic Berendsen thermostat. Energies are displayed per particle in units of the LJ potential depth ε. Time is units of the stochastic Berendsen relaxation time τ.

To this end, we integrate the equations determined by the system's Hamiltonian numerically using the velocity-Verlet algorithm [10]. We also rescale the particle velocities according to the stochastic Berendsen thermostat [11] with relaxation time τ. To initiate the system in a state of sufficiently low potential energy, we assign initial particle positions according to the corresponding ideal gas ensemble. Then we run the collision algorithm described in Sec. with collision strength $u = 1$. Moreover, we set the temperature by defining the root-mean-square initial velocity v_0 per particle and assigning a random velocity vector drawn from a two-dimensional Gaussian distribution with standard deviation v_0.

As parameters, we choose $v_0 = 6$, $\gamma = 0.08$, $N = 10^4$, $\Delta t = 0.01$, $\varepsilon = 20$, $z = 3$, $\tau = 100\Delta t$. To speed up the numerical integration, we only consider pairs of particles that lie within distance $r \leq 6z$, found by constructing and querying a k-d-tree for each time step. When calculating the interaction energies, we therefore shift the potential V_{LJ} so that $V_{\mathrm{LJ}}(r = 6z) = 0$. We integrate the equations of motion until $t = 10^4\tau$. The energy time series for a single run can be seen in Fig. 6. The final configuration of this run is shown in Fig. 7.

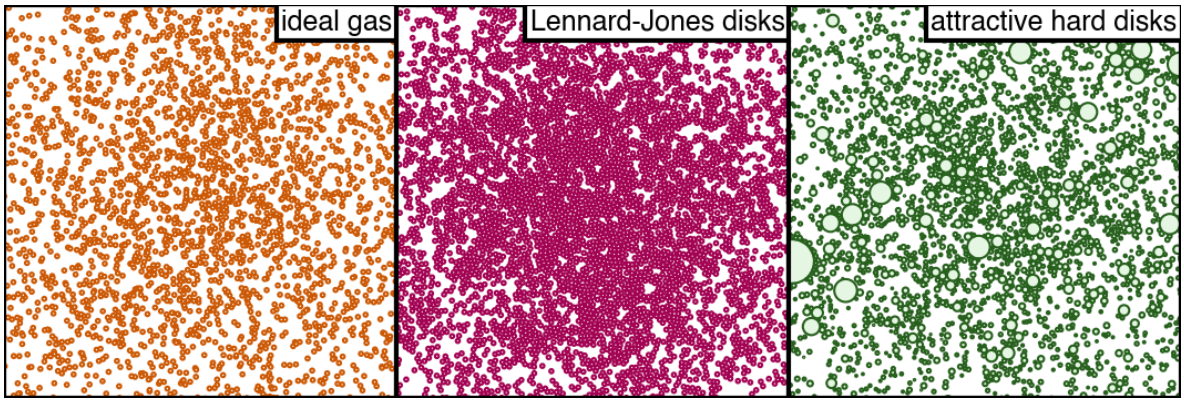

Figure 7: Three models for positioning buildings within an external potential of shape $V(\boldsymbol{x}) = \gamma|\boldsymbol{x}|$. (Left) An ideal gas, forming with radial density $p(|\boldsymbol{x}|) = (|\boldsymbol{x}|/R_0^2)\exp(-|\boldsymbol{x}|/R_0)$. (Middle) A configuration of Lennard-Jones disks. (Right) A configuration of heterogeneously sized hard disks with abstract attractive force.

2.1.6 Emulating attractive hard disks

We generate a configuration of hard disks with an unspecified attractive force (i.e. we do not explicitly integrate the equations of motion for a hard-sphere interaction potential with an additional attractive force). To do this, we first generate $N = 10^4$ random positions according to a radial Erlang distribution with scale parameter $R = 675$ (i.e. an ideal gas). Then, we draw a random radius for each position from a heterogeneous distribution with power-law tail. We first draw values $\hat{z}_i$

$$p(\hat{z}) = \frac{a^2-1}{2a} \times \begin{cases} \hat{z}^a, & \hat{z} \leq 1 \\ \hat{1}/z^a, & \hat{z} > 1 \end{cases} \tag{44}$$

and then assign disk radius $z_i = 3\hat{z}_i / \langle \hat{z} \rangle$. We choose $a = 4$ to obtain a heterogeneous distribution with non-finite variance in disk area. Afterward, we run the collision algorithm outlined in Sec.2.2 with collision strength $u = 1$. This leads the initially overlapping disks to take positions where their boundaries touch, i.e. an unlikely configuration to be found in the absence of an attractive force that would cause the disks to lie right at each other's boundaries.

An example configuration of this method is displayed in Fig. 7.

2.2 Collision-resolving algorithm

We have N disks with initial positions $\boldsymbol{x}_i$, radii R_i, and diameters $D_i = 2R_i$, respectively. With the distance vector $\boldsymbol{r}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$ and distance $r_{ij} = |\boldsymbol{r}_{ij}|$, let

$$\mathcal{E} = \{(i,j) : (i<j) \wedge (r_{ij} < \max(D_i, D_j)) \wedge (r_{ij} < R_i + R_j)\} \tag{45}$$

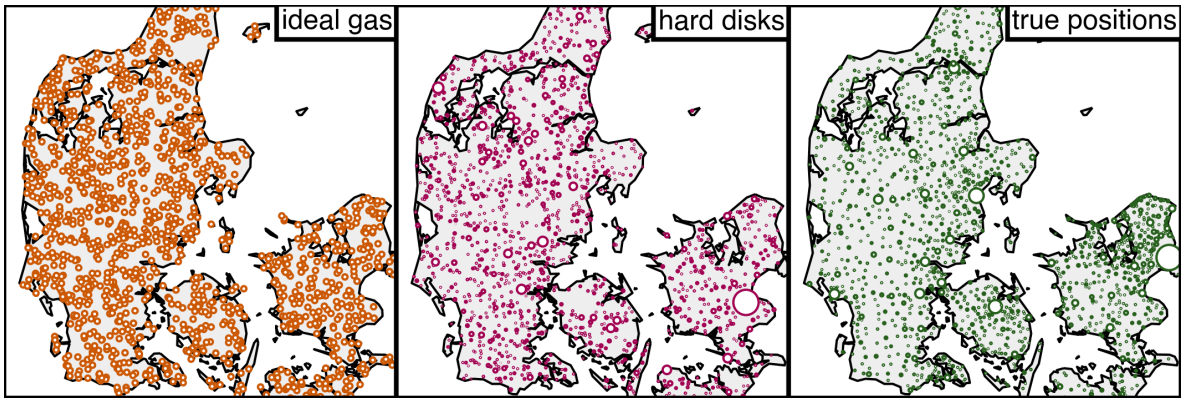

Figure 8: Three models for city positioning within Denmark. (Left) An ideal gas, i.e. positions chosen uniform at random within the administrative boundaries. (Middle) Hard disks are placed in a non-overlapping, random manner (see Sec. 2.3). Radiuses are equal to the inferred R values from the pair distribution MLE of Eq. 11 to Denmark's towns with more than 30 buildings. (Right) Cities placed at the centroid of their respective administrative boundaries with radius equal to R inferred from the respective city's empirical pair distribution .

be the set of pairs of disks that overlap. This set can be found by iterating over all disks i, finding all neighbors within distance D_i, for instance using a k-d-tree. With this definition, let

$$\mathcal{J}_i = \{j : (i,j) \in \mathcal{E} \vee (j,i) \in \mathcal{E}\} \tag{46}$$

be the set of disks that overlap with disk i. Then,

$$\tilde{\Delta}_i = u \sum_{j \in \mathcal{J}_i} \frac{\boldsymbol{r}_{ij}}{r_{ij}} \left(R_i + R_j - r_{ij}\right) \frac{m_j}{m_i + m_j}. \tag{47}$$

is the demanded initial displacement, with the displacement rate $0 < u \leq 1$ (imagine two overlapping disks—with $u = 1$, the collision would be resolved after one update). The masses m_j control the strength of the displacement. Suppose that disk i has s small mass and a colliding disk j has a great mass. The colliding disk j should move less. Hence, the influence on disk i from disk j should be proportional to m_j. Assuming homogeneous density of all disks, we can set

$$m_i = R_i^2. \tag{48}$$

We also want to avoid that disks move too far per one single update to avoid large jumps. Therefore, we move each disk by the final displacement vector

$$\Delta_i = \tilde{\Delta}_i \times \min(1, R_i / |\tilde{\Delta}_i|). \tag{49}$$

i.e. the disk shouldn't move more than its radius per update.

We update the whole ensemble of disks step by step until either

$$|\mathcal{E}| = 0 \tag{50}$$

or

$$\max_i |\Delta_i| < \epsilon. \tag{51}$$

With a default value of $\epsilon = 10^{-10}$ m.

If all disks are of equal radius R, we instead choose the second stop condition as

$$\min_{(i,j)\in\mathcal{E}} r_{ij} > (1-\epsilon)2R \tag{52}$$

and $\epsilon = 10^{-3}$.

2.3 Random non-overlap positioning algorithm of disks

We want to randomly distribute N hard disks of radii R_i in shape Ω. We start with the largest disk of radius $R_{\max}$ and iterate over all disks in decreasing order of size. For each disk j, we generate a random position $\boldsymbol{x} \in \Omega$ until the condition $|\boldsymbol{x} - \boldsymbol{x}_i| > R_j + R_i \; \forall j < i$ is satisfied, i.e. drawing new random positions until there are no overlaps with other, already placed disks. Then, assign $\boldsymbol{x}_j \leftarrow \boldsymbol{x}$ and continue with the next disk.

2.4 Influence of building location definition on pair distribution

In the dataset, the location of a building is defined as the location of the building's entrance door. We want to check how the Lennard-Jones ensemble's pair distribution and the pair-correlation function $g(r)$ change when the location of a building is not associated with its center. To this end, we take the configuration of Lennard-Jones disks as shown in Fig. 2d in the main text and redefine a disk's location to be (i) randomly within the disk and (ii) randomly on the rim of the disk. The resulting pair distribution and $g(r)$ are shown in Fig. 9. We see that the sharp peak almost disappears for both. At the same time, the onset of $g(r)$ becomes less abrupt, approaching a shape similar to that observed in the data.

2.5 Measuring the pair distribution function for large dataset

Calculating the pair distribution function for large datasets requires a substantial amount of memory. The number of different pairs among N points is $N(N-1)/2$. Consequently, for the dataset of 34 million address coordinates within France (Fig.4), we need to store 10^{15} distance values in memory, which is equivalent to 4,000 terabytes of data using 32-bit floats. Such a large amount of data storage is practically unfeasible. To overcome this computational hurdle, we proceed in two steps using a k-d-tree.

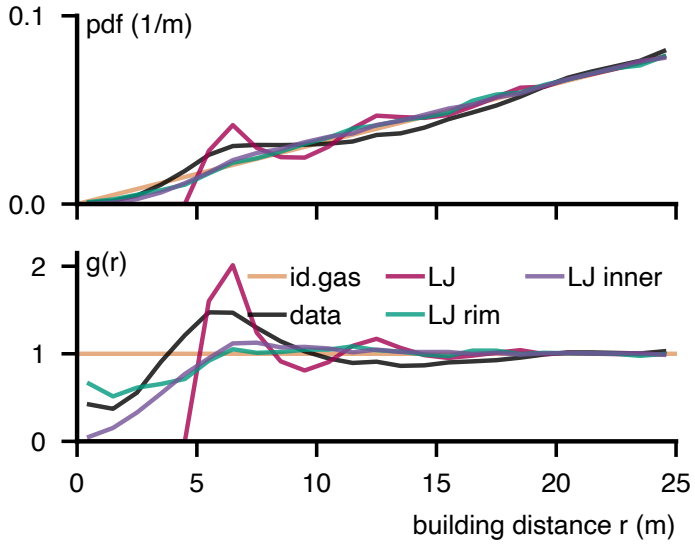

Figure 9: How the LJ pair distribution and pair-correlation function changes when building location is redefined to (i) randomly within the disk of radius z and (ii) randomly on the rim of disk of radius z.

2.5.1 Small distances

Let $\mathcal{D}$ be the two-dimensional, non-contiguous shape containing every building (or address, respectively) in Denmark and δ the set of buildings (addresses) i with locations $\boldsymbol{x}_i \in \mathcal{D}$. For each predefined subset of this form $\Omega \subseteq \mathcal{D}$ (for example, the official boundaries of a city), we iterate over all building (address) positions $\omega = \{i \in D : x_i \in \Omega\}$ located within Ω and compute the distances to all buildings (addresses) $j \in \delta \wedge j \neq i$ with $0 < |x_i - x_j| < r_{\max} = 200\,\mathrm{m}$. That is, for each predefined subset of $\mathcal{D}$ (e.g. city), we find all distances of every one of its buildings with respect to *all* buildings (addresses) that lie within radius $r_{\max}$, not just to those that also lie within its boundaries. We compute this histogram with a resolution (i.e. bin width) of 1 m. For this task, we use a k-d-tree on all locations of δ.

For each city i in Denmark, let Ω_i be the shape that is defined by its administrative boundaries. Note that none of the Ω_i overlap. We define as

$$\overline{\Omega} = \mathcal{D} \setminus \left(\bigcup_i \Omega_i \right) \tag{53}$$

the shape that includes land that is not associated with a city. Thus, iteration over all Ω_i and $\overline{\Omega}$ allows us to analyze the small-scale structure of every city, every building (address) that is not located in a city, and - by combining all these histograms - of every building (address) in Denmark.

We inferred the mesoscale scaling parameter α by fitting $p(r) = Cr^{\alpha}$ against the respective empirical pair distribution with 1m resolution in the range $r \in (25\,\mathrm{m}, 200\,\mathrm{m})$, using least

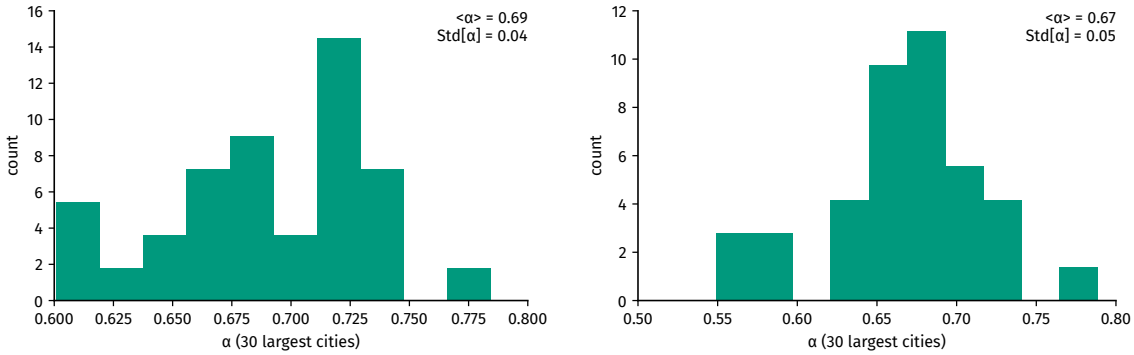

Figure 10: Inferred values of the sub-linear scaling exponent α in $p(r) \propto r^{\alpha}$ for $r \in (25\,\mathrm{m}, 200\,\mathrm{m})$ for the 30 largest Danish cities. (Left) buildings. (Right) addresses.

squares. We find $\langle \alpha \rangle = 0.69$ and $\mathrm{Std}[\alpha] = 0.04$ for the building pair distribution s of the 30 largest cities as well as $\langle \alpha \rangle = 0.67$, and $\mathrm{Std}[\alpha] = 0.05$ for addresses, cf. Fig. 10. Here, "largest" refers to the number of registered residential buildings with locations within the administrative boundaries of the city. For all buildings in Denmark, we find $\alpha = 0.67$ and for all addresses, we have $\alpha = 0.69$. Example analyses can be seen in Fig. 11 for buildings and in Fig. 12 for addresses. Note that the local environment around buildings and addresses that are not within a city/town boundary grows much slower with $\alpha = 0.30$ for both.

2.5.2 Larger distances

To compute the pair distribution for larger distances, we proceed as follows. For each shape Ω_i (and $\mathcal{D}$, respectively), we find the set of its building (address) locations ω_i (and δ, respectively). From this set we sample $n = \min(|\omega_i|, 3 \times 10^4)$ unique locations (without replacement). Then, we find the pairwise distances of all pairs of these sampled locations and bin them to find histograms.

Note that to obtain the country-wide pair distribution displayed in Fig. 2a in the main text, we combine the respective pair distributions from the small-distance and large-distance analyses by requiring that they take the same value at $r = 184.5\,\mathrm{m}$.

We show the pair distributions for buildings and addresses for the 100 largest Danish cities in Fig. 13 and 14. In Fig. 15 we show the empirical and fit pair distributions of all Danish cities with more than 30 buildings as well as the distributions of the inferred values of R and m, respectively. The tail of the radius distribution R scales as $p(R) \propto 1/R^{3.29}$, with $x_{\mathrm{min}} = 824\,\mathrm{m}$, inferred by the MLE technique in refs. [12, 13]. The tail decay $(r/R)^m$ has values of m with $\langle m \rangle = 1.61$ and $\mathrm{Std}[m] = 0.44$.

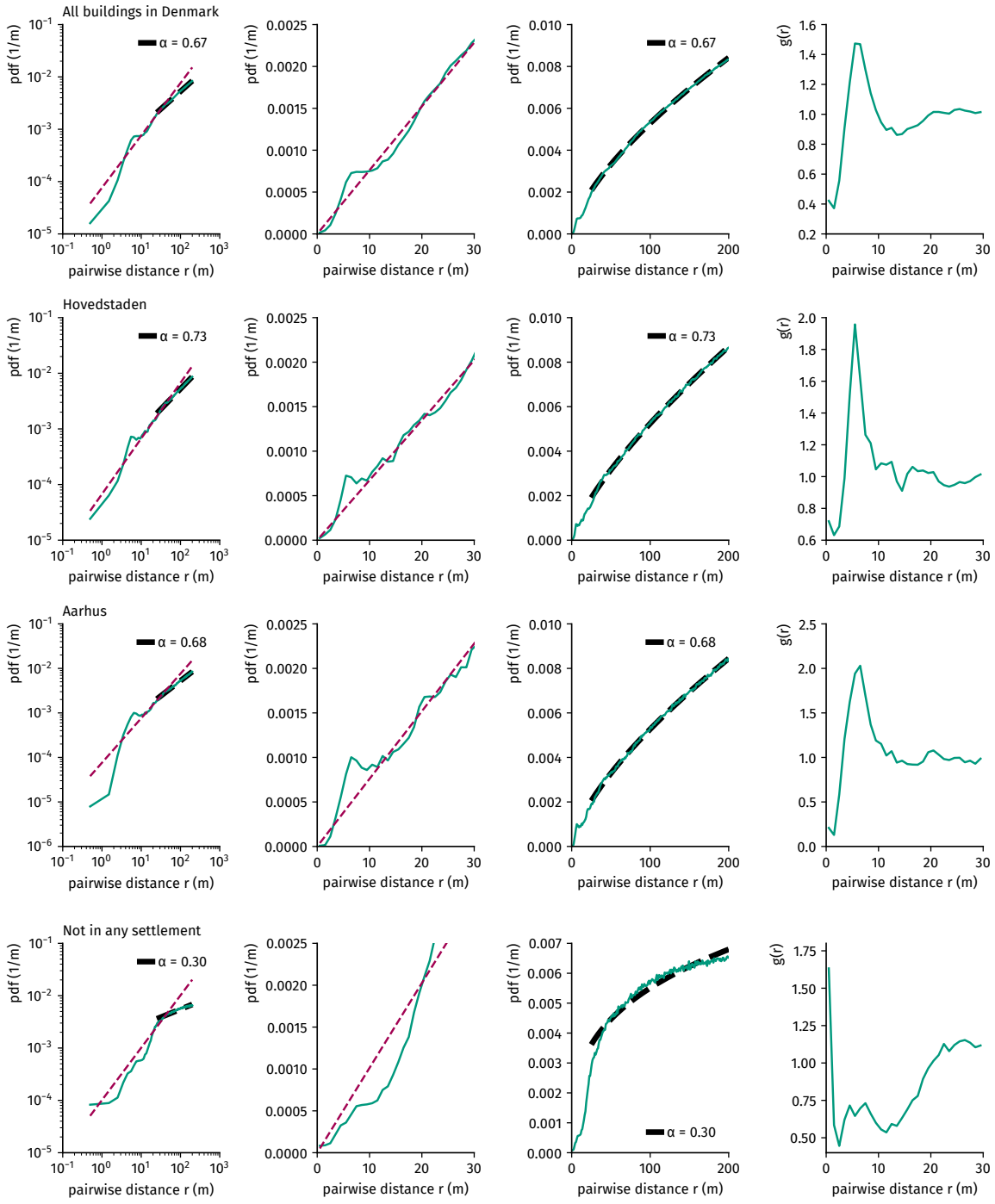

Figure 11: Building pair distribution functions and pair-correlation functions $g(r)$ with $r < 200\,\mathrm{m}$ for (top) every building of Denmark, (second from top) Hovedstadens (capital region), (second from bottom) Aarhus, and (bottom) every building not located with the administrative boundaries of any city/town.

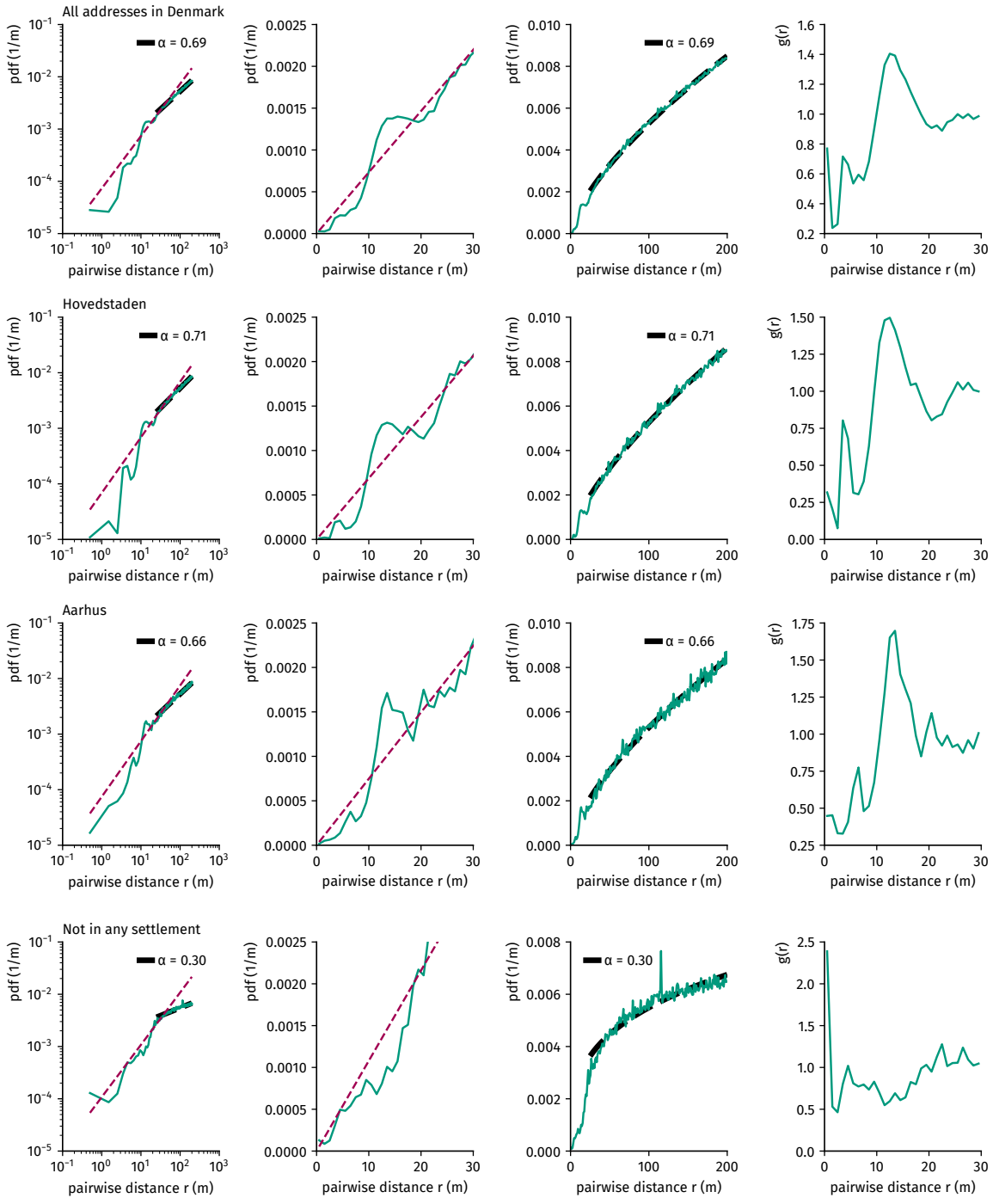

Figure 12: Address pair distribution functions and pair-correlation functions $g(r)$ with $r < 200\,\mathrm{m}$ for (top) every building of Denmark, (second from top) Hovedstadens (capital region), (second from bottom) Aarhus, and (bottom) every address not located with the administrative boundaries of any city/town.

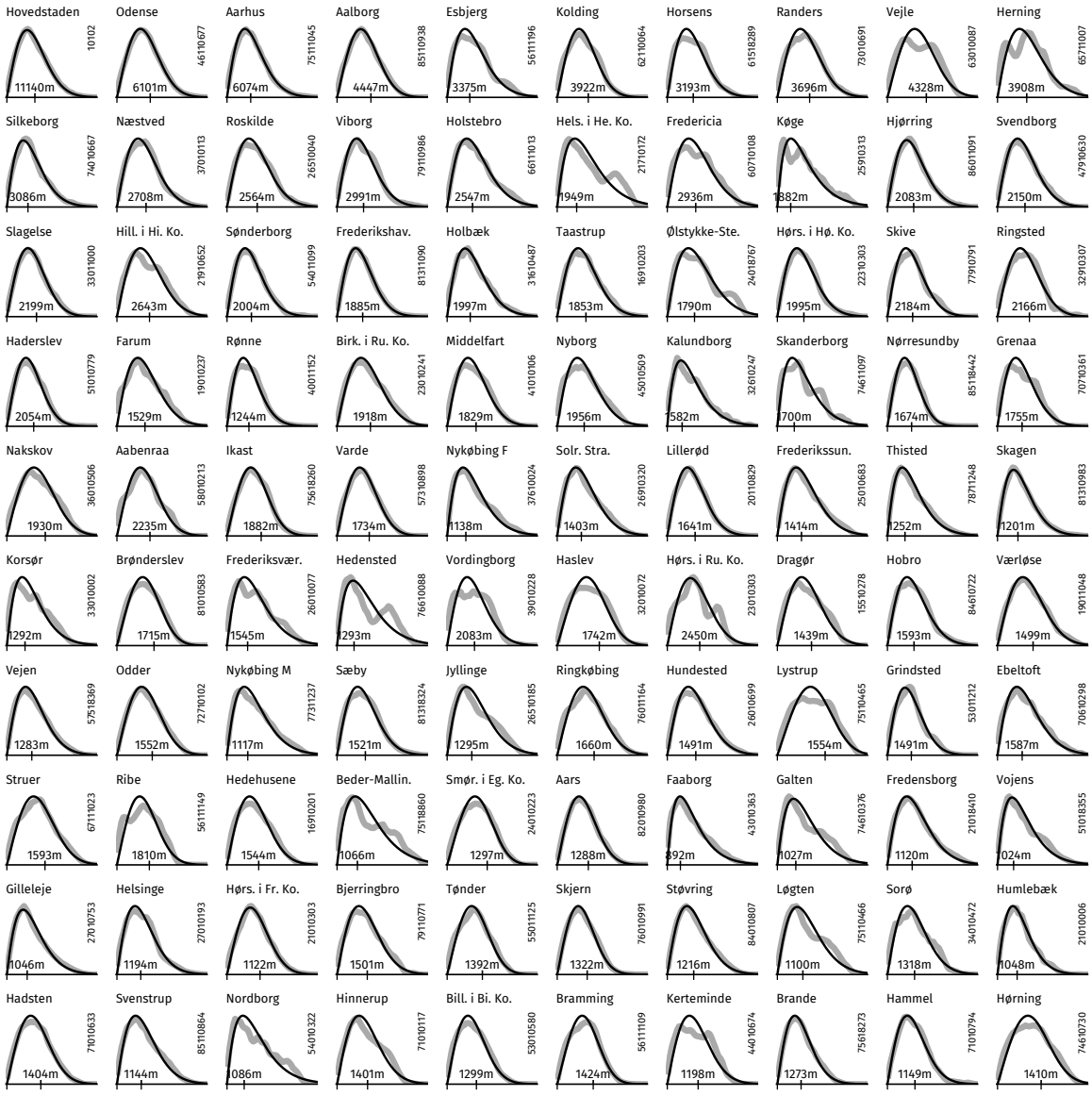

Figure 13: Empirical building pair distribution functions for the largest 100 cities in Denmark and fits of $p(r, R, m) = (r/R^2)\exp(-(r/R)^m)/\Gamma(2/m)$. We mark the inferred radius R.

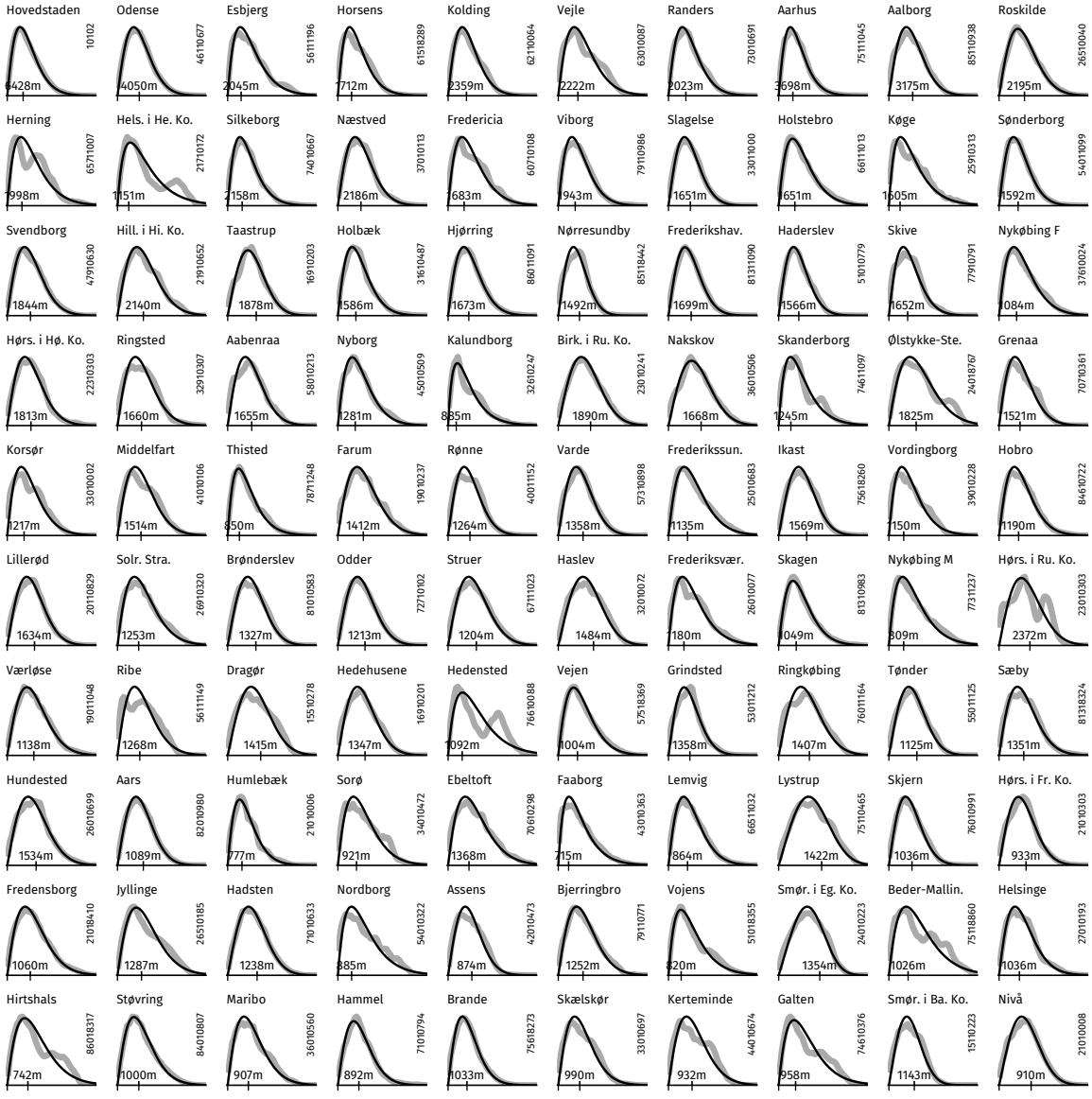

Figure 14: Empirical address pair distribution functions for the largest 100 cities in Denmark and fits of $p(r, R, m) = (r/R^2)\exp(-(r/R)^m)/\Gamma(2/m)$. We mark the inferred radius R.

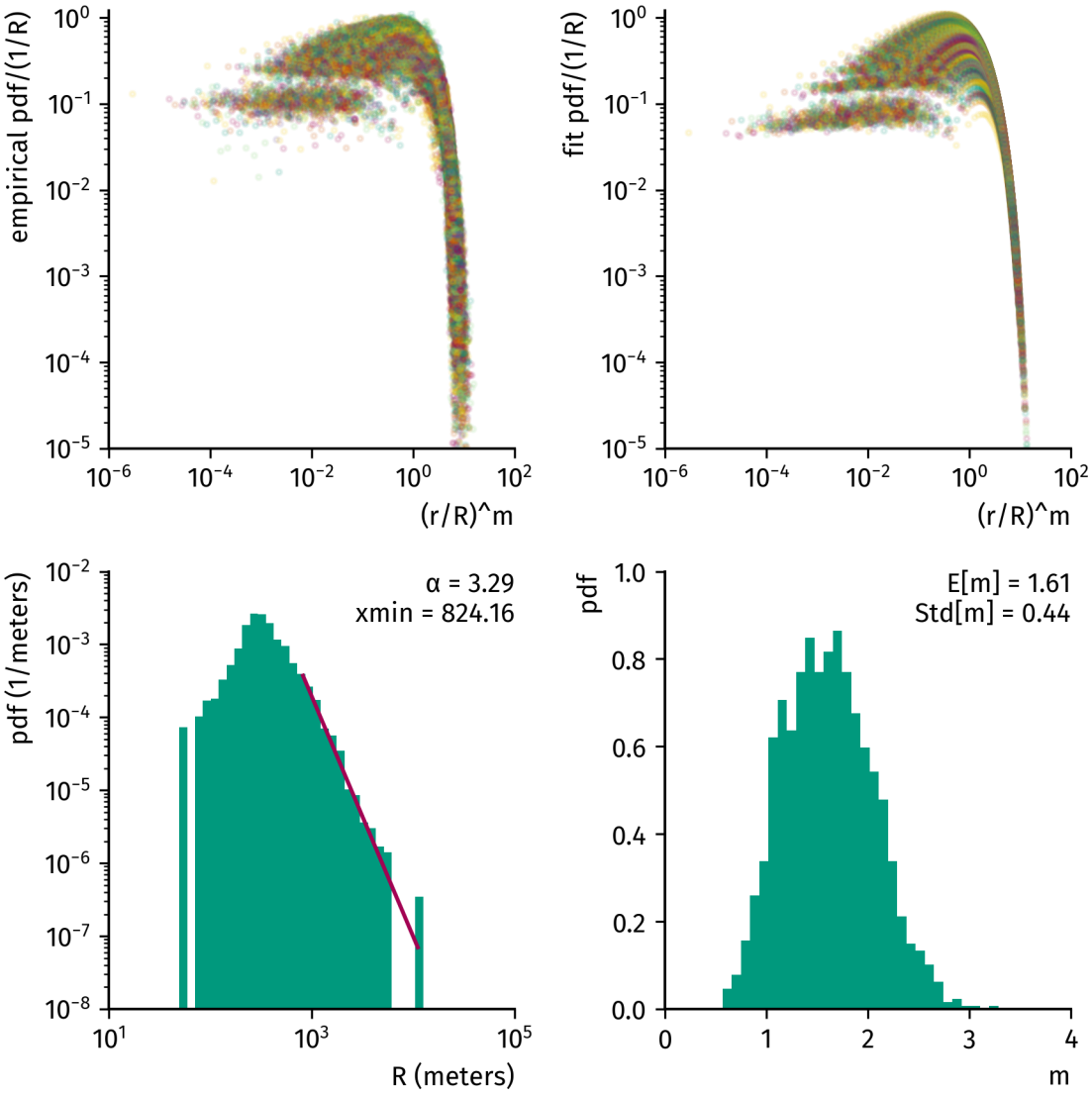

Figure 15: (Upper left) empirical pair distributions for all Danish towns with more than 30 buildings, rescaled by R and m obtained from (upper right) MLE fits of $p(r, R, m) = (r/R^2)\exp(-(r/R)^m)/\Gamma(2/m)$ to the empirical data. (Lower left) Distribution of inferred R values and power-law fit to its tail, giving $p(R) \propto R^{-\alpha}$ with $\alpha = 3.29$ and $x_{\min} = 824\,\mathrm{m}$. (Lower right) Distribution of inferred m values with $\langle m \rangle = 1.61$ and $\mathrm{Std}[m] = 0.44$.

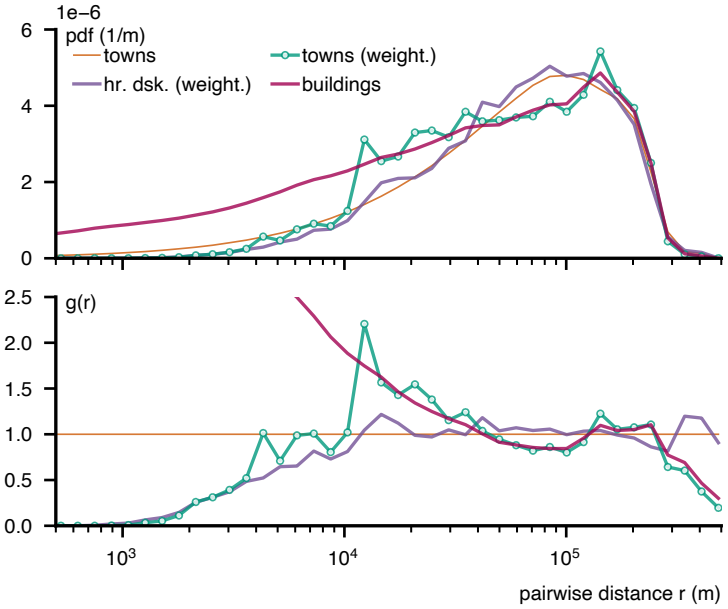

Figure 16: building-pair-weighted pair distribution function of cities, both for real city positions as well as a hard disk model without attractive force. Overlaid are the building pair distribution and the ideal gas pair distribution.

2.5.3 Between cities

To compute the pair distribution and $g(r)$ between cities, we define the 'city center' as the centroid of a city's (multi-) polygon, i.e. its geometric center: the center of mass of its shape.

Weighting each inter-city distance r_{ij} by the number of pairs of buildings $m_i m_j$ it contains, we find a pair distribution that approximates the empirical building pair distribution (see Fig. 16). This pair distribution has a clear peak in $g(r)$, suggesting an abstract attractive force between cities of a certain size. However, the onset of $g(r)$ is consistent with the weighted pair distribution of the hard-disk model configuration.

2.6 Pair distribution from independent patches of heterogeneous size

We consider a *multiverse* consisting of an infinite amount of universes indexed by i, each of which is inhabited by a patch of size R_i with N_i buildings inside, contributing to the multiverse with a pair distribution number density of $n(r, R_i) \approx N_i^2 p(r, R_i, m)$ (each universe contributes a number of building pairs amounting to N_i^2). Furthermore, we assume that there is a constant population density, which is proportional to the number of individuals per unit area, and that this density is inversely proportional to the square of the radius, $\rho_0 \propto N/R^2$. This implies that the number of individuals in a given area is proportional to

the area itself, $N \propto R^2$.

For each of these universes (or rather, each of these patches), we calculate the number of building pairs at distance r and assume that the patch sizes R are distributed according to some distribution $f(R)$, which is currently unspecified. The joint distribution of house pairs at distance r is then given by

$$n(r) \propto \int_0^\infty dR\, n(r,R) f(R) \tag{54}$$

$$\propto \int_0^\infty dR\, f(R) R^4 \frac{r}{R^2} \exp\left(-\left(\frac{r}{R}\right)^m\right). \tag{55}$$

To simplify the integral, we change variables to $\beta = 1/R$, such that $\mathrm{d}R = -\beta^2 \mathrm{d}\beta$ and

$$n(r) \propto \int_0^\infty d\beta\, f(\beta^{-1}) \frac{r}{\beta^4} \exp\left(-(\beta r)^m\right). \tag{56}$$

The integral in question has a solution in the case where $f = \beta^{4-\alpha}$ with $\alpha < 1$. We assume that this is the case, the integral now becomes,

$$n(r) = C \int_0^\infty \beta^{-\alpha} r \exp\left(-(r\beta)^m\right) d\beta \tag{57}$$

where C is a normalization constant.

We begin with the Gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt. \tag{58}$$

This integral converges for $z > 0$ if $z \in \mathbf{R}$. Substituting $t = (r\beta)^m$ yields $\mathrm{d}t = m r^m \beta^{m-1} \mathrm{d}\beta$ and therefore

$$\Gamma(z) = \int_0^\infty (r\beta)^{mz-m} \exp\left(-(r\beta)^m\right) m (r\beta)^m \beta^{-1} d\beta \tag{59}$$

$$= m r^{mz-1} \int_0^\infty \beta^{mz-1} r \exp\left(-(r\beta)^m\right) d\beta \tag{60}$$

$$= m r^{-\alpha} C^{-1} C \int_0^\infty \beta^{-\alpha} r \exp\left(-(r\beta)^m\right) d\beta. \tag{61}$$

Here, we introduce $z = (1-\alpha)/m$, which, due to the lower bound of $z > 0$, leads to an upper bound of $\alpha < 1$. We recognize the integral $n(r)$ on the right and therefore find

$$n(r) \propto r^\alpha C\, \Gamma\left(\frac{1-\alpha}{m}\right). \tag{62}$$

If the radius of patches in the multiverse were distributed according to the function $f(R) \propto R^{-(4-\alpha)}$, the initial growth of the joint pairwise distribution would follow a sublinear power law. Notably, the scaling exponent would not depend on the tail parameter m.

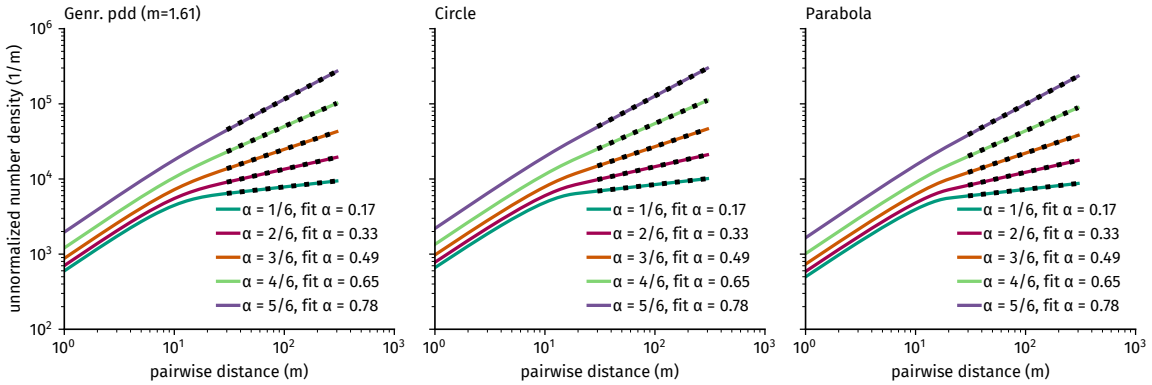

Figure 17: Integrating over the multiverse of patches of size R, summing up the contributions to the pairwise-distance distribution of all patches, for three assumptions of (i) generalized pair distribution, (ii) pair distribution of a uniform, circular patch of buildings, and (iii) a parabola pair distribution, with varying α. Here, we used a minimum patch size of $R = 10$m, which fixes a scale and leads to linear growth of the pairwise-distance distribution for small distances.

Let us consider the implications of this for the area of the patches. The area would scale as $A \propto R^2$ (or $R \propto \sqrt{A}$), indicating that the distribution of the area would follow

$$\tilde{f}(A) = \frac{\mathrm{d}R}{\mathrm{d}A} f(R = \sqrt{A}) \tag{63}$$

$$\propto \frac{1}{A^{(5-\alpha)/2}}. \tag{64}$$

In the data, we observe $\alpha = 0.67$. That means that the area of the patches would have to be distributed according to a power law with exponent $\mu = -(5-\alpha)/2 \approx -2.17$, which is well within what has been found empirically for cities [14]. Furthermore, our city radius inference analysis (cf. Fig. 15) indicates that $f(R) \propto 1/R^{3.29}$, which leads to $\mu = -2.15$. This demonstrates that the results of these two separate analyses are consistent. As previously demonstrated, the sublinear scaling exponent α is independent of the tail parameter m. We extend our analysis by using the parabola pair distribution model Eq. 23. We have

$$n(r) \propto \int_0^\infty dR\, n(r,R) f(R) \tag{65}$$

$$= \int_{r/2}^\infty dR\, f(R) R^4 \left(-\frac{3}{4R} \left(\frac{r}{R} \right)^2 + \frac{3r}{2R^2} \right). \tag{66}$$

Here, the lower bound in the integral comes from the condition that $r \leq 2R$. Now, as above,

we demand $f(R) \propto 1/R^{4-\alpha}$ to find

$$n(r) \propto \int_{r/2}^{\infty} dR\, R^{\alpha} \left(-\frac{3}{4R} \left(\frac{r}{R} \right)^2 + \frac{3r}{2R^2} \right) \tag{67}$$

$$= \int_{r/2}^{\infty} dR \left(-\frac{3r^2}{4R^{3-\alpha}} + \frac{3r}{2R^{2-\alpha}} \right) \tag{68}$$

$$\propto r^{\alpha}. \tag{69}$$

We solve the respective integrals $n(r)$ for the circle, generalized pair distribution, and parabola pair distribution models numerically and find that the above derivation holds for all three (see Fig. 17).

2.7 Pair distribution of a self-similar modular hierarchical model of building locations

One limitation of the multiverse approach is that it is only applicable if the patches are truly independent or sufficiently separated so that the scale of the pair distribution number density $n(r)$ does not affect the outcome. However, it is plausible that patches may be in close proximity to each other. This is supported by findings in [14]. Additionally, they discovered that a collection of these patches forms a fractal, or self-similar structure. While it is possible to demonstrate that a fractal dimension of building location does not necessarily result in sub-linear growth of the pair distribution function, it is certainly possible to investigate the consequences of such self-similar patch location on it.

We assume a self-similar modular hierarchical structure comprising patches of buildings. We start with a single unit of b buildings, each of which has a radius z_1. These are located within a patch of size R_1. Our objective is to regulate the building density in such a way that the number of buildings in a patch is given by $b = \theta(R_1^2/z_1^2)$. Here, θ represents the packing fraction. This implies that the patch radius is computed as follows: $R_1 = z_1\sqrt{b/\theta}$.

Now, consider that there are b of these patches of radius R_1, located in a larger patch of higher order, which is (self-)similar to the basic patch. We posit that each of the lower-order patches has a radius of $z_2 = R_1$, while the higher-order patch has a radius of $R_2 = z_2\sqrt{b/\theta}$.

Subsequently, we add $b-1$ similarly constructed patches to form an even larger patch. This entails constructing a self-similar structure of patches where the size of each patch of hierarchical order k is given by

$$R_k = z \left(\frac{b}{\theta} \right)^{k/2} \tag{70}$$

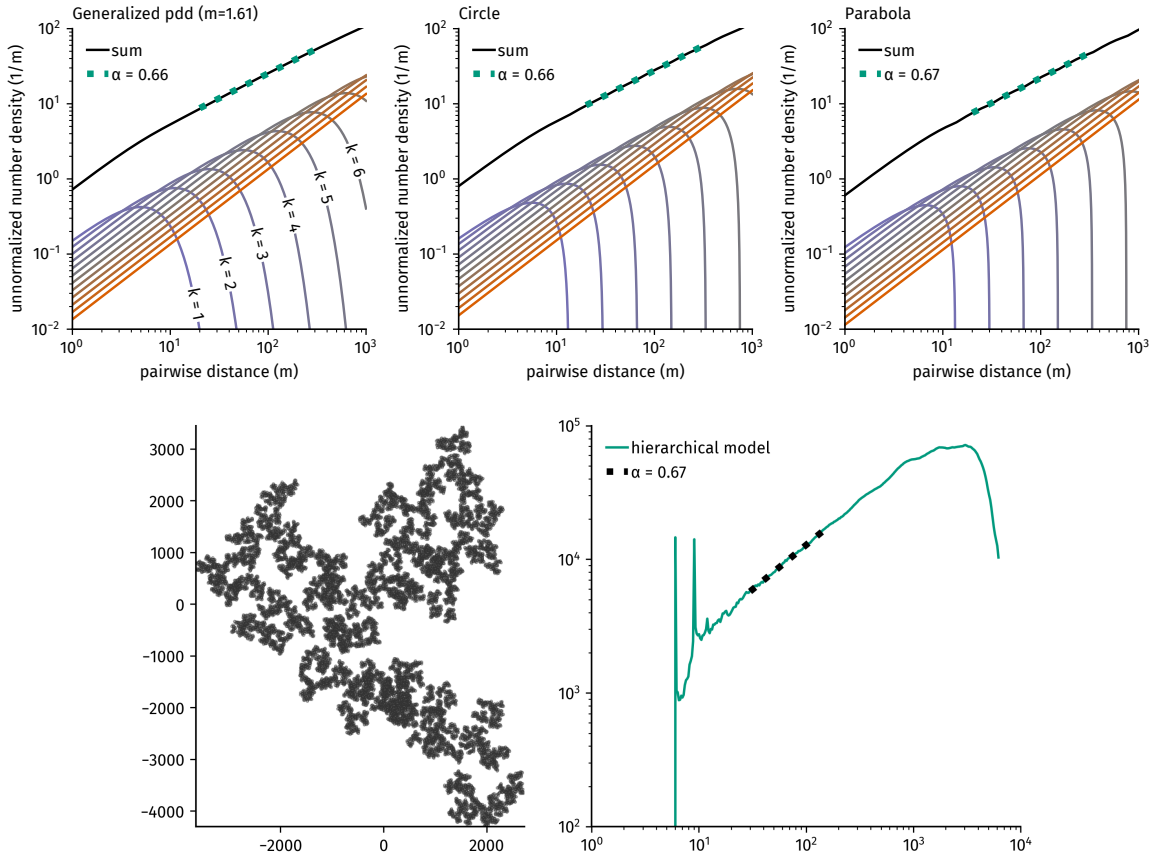

Figure 18: Upper two panels: Theoretical pairwise-distance distributions of the hierarchical building placement model with $N = 4$ and $\theta = 0.8$, in color the single contributions Eq. (73) of scale k, in black the total joint distribution function over all scales Eq. (74), including a power-law fit showing the resulting scaling exponent $\alpha \approx 2/3$. Lower two panels: A sample from the hierachical placement model with $L = 8$ hierarchy layers and the corresponding resulting pairwise-distance distribution of the whole structure, showing the desired scaling behavior on the mesoscale.

and the size of each sub-patch of a patch is given by

$$z_k = z_1 \left(\frac{b}{\theta}\right)^{(k-1)/2}. \tag{71}$$

Given a maximum number of L orders (layers), the total number of houses is eventually b^L. In each layer k, there are b^{L-k} patches of order k.

Now assume that for each patch i of order k and location $\boldsymbol{s}_{i,k}$, we distribute the location of its b sub-patches randomly within this patch such that the pair distribution function of sub-patches leads to a pair distribution $p(r, R_k$ of scale R_k. To prevent the patches from overlapping excessively, a collision algorithm is employed for the b sub-patches (preventing collisions between disks with radius z_k). We repeat this process recursively for each patch until the depth of the N-ary hierarchy tree reaches L. The leaves of the tree represent buildings. As these buildings will overlap, another collision algorithm is run until they no longer overlap.

We now estimate the joint pairwise-distance distribution of the entire structure. Consider a container patch that contains b sub-patches. If we assume that the total the B buildings within a sub-patch of scale $k-1$ are sufficiently concentrated within its center, the contribution of this patch of radius R_k will be proportional to the pair distribution $p(r, R_k)$, weighted with the total number of pairs of buildings within this container-patch, except the pairs of buildings within the same sub-patch. Consequently, the number of pairs therefore scales as $\propto B^2 b(b-1)$, or, neglecting the linear contribution, as $\propto (bB)^2$. For each patch in layer k, there are going to be $B = b^{k-1}$ buildings in a sub-patch. Therefore, the number of pairs of buildings in a patch of scale k grows as $\propto b^{2k}$.

Each of these patches contributes to approximately b^{2k} pairs to the joint pairwise-distance distribution. There are b^{L-k} patches of order k with radius R_k. Therefore, the total contribution to the joint pairwise-distance distribution of this scale is

$$\begin{aligned} n_k(r, R_k) &\propto b^k p(r, R_k) && (72)\\ &= b^k p(r, z(b/\theta)^{k/2}) && (73) \end{aligned}$$

where $p(r, R)$ depends on how sub-patches are distributed within patches.

The total pair distribution number density is given by

$$n(r) = \sum_{k=1}^{L} b^k p(r, z(N/\theta)^{k/2}). \tag{74}$$

An explicit form using the generalized pair distribution function model is

$$n(r) \propto \sum_{k=1}^{L} b^k \frac{r}{R_k^2} \exp\left(-\left(\frac{r}{R_k}\right)^m\right). \tag{75}$$

$$= \frac{r}{z^2} \sum_{k=1}^{L} \theta^k \exp\left(-\left(\frac{r}{z}\right)^m \left(\frac{\theta}{b}\right)^{km/2}\right). \tag{76}$$

We compute this equation numerically for varying θ, b, and different models for $p(r, R)$. Figure 18,illustrates examples for $b = 4$ and $\theta = 0.8$, which leads to a scaling of approximately $\propto r^{0.67}$.

To derive the dependence of the exponent of the observed scaling law on the parameters, we use the saddle-point approximation. First, we approximate the sum over hierarchy layers with an integral over a constant hierarchy layer density

$$n(r) \propto r \int_1^\infty dk \exp\left[\underbrace{k \ln\theta - \left(\frac{r}{z}\right)^m \exp\left(\frac{mk}{2} \ln\frac{\theta}{b}\right)}_{=-w(k)}\right]. \tag{77}$$

Relying on the saddle-point method, we can approximate the integral to find

$$n(r) \propto r \frac{\exp\left[-w(k_0)\right]}{\sqrt{w''(k_0)}} \tag{78}$$

where k_0 is the minimum of $w(k)$. We compute the derivatives

$$w'(k) = -\ln\theta + \left(\frac{r}{z}\right)^m \frac{m}{2} \ln\frac{\theta}{b} \exp\left(\frac{mk}{2} \ln\frac{\theta}{b}\right) \tag{79}$$

$$w''(k) = \left(\frac{r}{z}\right)^m \left(\frac{m}{2} \ln\frac{\theta}{N}\right)^2 \exp\left(\frac{mk}{2} \ln\frac{\theta}{b}\right) \tag{80}$$

and with $w'(k_0) = 0$ find the following equations for the minimum

$$\exp\left(\frac{mk_0}{2} \ln\frac{\theta}{b}\right) = \frac{\ln\theta}{\ln(\theta/b)} \left(\frac{z}{r}\right)^m \frac{2}{m} \tag{81}$$

$$k_0 = \frac{2}{m} \frac{1}{\ln\theta/b} \left[\ln\left(\frac{\ln\theta}{\ln(\theta/b)}\right) + m\left(\ln z - \ln r\right) + \ln\left(\frac{2}{m}\right)\right]. \tag{82}$$

Using the first of the two equations we note that for $w(k_0)$, the dependence on r cancels out in the second term and so

$$-w(k_0(r)) = k_0(r) \ln\theta - W \tag{83}$$

where W is an irrelevant constant. Second, we use the same equation to see that $w''(k_0)$ does not depend on r and therefore does not concern us any further either. As a step in between, this means that

$$n(r) \propto r \exp\left[k_0(r) \ln\theta\right]. \tag{84}$$

Looking at Eq. (82), we see that the only r-dependent term gives the minimum a structure of

$$k_0 = K - \frac{2}{\ln(\theta/b)} \ln r, \tag{85}$$

and therefore we find

$$n(r) \propto r \exp\left[-\frac{2\ln\theta}{\ln(\theta/b)} \ln r\right] \tag{86}$$

$$= r^{1-\frac{2\ln\theta}{\ln(\theta/b)}}, \tag{87}$$

i.e. the exponent of the sub-linear growth in the pair distribution function is given as

$$\alpha = 1 - \frac{2\ln\theta}{\ln(\theta/b)}. \tag{88}$$

We can compare this estimation with numerical results. Above we used $N = 4$ and $\theta = 0.8$ to find $\alpha \approx 0.67$. Eq. (88) yields $\alpha = 0.72$, an acceptable approximation. Notably, this result is also independent of the parameter m, which represents the explicit form of the pairwise distance distribution per patch. This explains why similar results are obtained despite the use of different geometries.

We can perform the same analysis for the parabola pair distribution model Eq. (23). We have

$$n(r) \propto \sum_{k=1}^{L} \frac{b^k}{R_k^2}\left(-\frac{3}{4}\frac{r^2}{R_k} + \frac{3r}{2}\right) \tag{89}$$

$$= \frac{r}{z^2}\sum_{k=1}^{L} \theta^k\left(-\frac{3}{4}\frac{r}{R_k} + \frac{3}{2}\right). \tag{90}$$

Again, we use the saddle point method to approximate

$$n(r) \propto r\frac{\exp\left[-w(k_0)\right]}{\sqrt{w''(k_0)}}, \tag{91}$$

with k_0 being the minimum of the function

$$w(k) = -k\ln\theta - \ln\left(-\frac{3}{4}\frac{r}{z}\frac{\theta^{k/2}}{b^{k/2}} + \frac{3}{2}\right). \tag{92}$$

We arrive at

$$n(r) \propto r\frac{\theta^{\frac{2\ln\left(\frac{4z\ln(\theta)}{r\left(2\ln(\theta)+\ln\left(\frac{\theta}{b}\right)\right)}\right)}{\ln\left(\frac{\theta}{b}\right)}} 3\sqrt{\pi}\ln\left(\frac{\theta}{b}\right)}{z^2\sqrt{\left(2\ln(\theta)+\ln\left(\frac{\theta}{b}\right)\right)\ln(\theta)}\left(2\ln(\theta)+\ln\left(\frac{\theta}{b}\right)\right)}. \tag{93}$$

Leaving aside the irrelevant prefactors, we have

$$\ln n(r) \propto \ln r + \ln\theta\left(K - \frac{2\ln r}{\ln(\theta/b)}\right) \tag{94}$$

$$\propto \ln r\left(1 - \frac{2\ln\theta}{\ln(\theta/b)}\right), \tag{95}$$

i.e. the same result as with the generalized pair distribution.

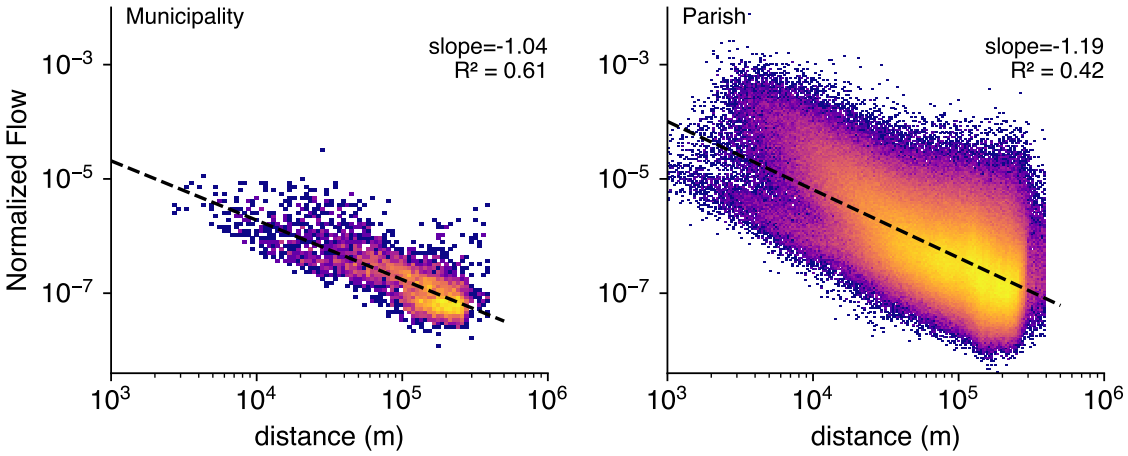

Figure 19: Gravity model as in [1] for migration between (left) municipalities in Denmark, (right) parishes in Denmark. The normalized flow represented on the y-axis is the number of moves from i to j divided by the product of population between i and j, $T_{i,j}/(N_iN_j)$.

3 Studying mobility with the pair distance function

3.1 Comparison with other Mobility models

3.1.1 Gravity

The gravity model as originally introduced in [1], states that the number of moves T from area i to j is proportional to,

$$T_{i\to j} \propto \frac{N_i N_j}{d_{i,j}} \tag{96}$$

with N_i the population at location i and $d_{i,j}$ the distance between locations i and j.

3.1.2 Radiation

In this section, we derive a continuous radiation model and demonstrate how it differs from our continuous gravity model. The radiation model of human mobility is usually represented by the following equation [15],

$$P_{ij} \propto \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})} \tag{97}$$

with P_{ij} the predicted number of people traveling from location i to location j, m_k the population at the origin location k, s_{ij} represents the total population in a circle centered at i with radius equal to the distance between i and j, excluding the populations of i and j.

Following [15], the probability of having a single particle emitted from location i to location j is,

$$P(1|m_i, m_j, s_{i,j}) = \int_0^\infty \mathrm{d}z P_{m_i}(z) P_{s_i}(< z) P_{m_j)(>z)} \tag{98}$$

with,

$$P_{m_i}(z) = \frac{\mathrm{d}P_{m_i}(< z)}{\mathrm{d}z} = m_i p(z)^{m_i - 1} \frac{\mathrm{d}p(< z)}{\mathrm{d}z} \tag{99}$$

if $m_i = 1$, then it simplifies to $P_1(z) = \mathrm{d}P(< z)/\mathrm{d}z$. Equation 98 simplifies to,

$$P(1|1, 1, s_{i,j}) = \int_0^\infty (1 - P(< z)) P(< z)^{s_{i,j}} \mathrm{d}P(< z) \tag{100}$$

$$= \frac{1}{(1 + s_{i,j})(2 + s_{i,j})} \tag{101}$$

In the general case when $s_{i,j} \gg 1$, that is, when the number of houses in the circle of center i and radius the distance i and j is greater than one, which is often the case when the distance is sufficiently large (see Fig. **??**),

$$P_{ij} = \frac{1}{(1 + s_{ij})(1 + 1 + s_{ij})} \sim s_{ij}^{-2}. \tag{102}$$

However, the number of addresses in the circle $s_{i,j}$ of center i and radius $R(i, j)$ is closely related to the primitive of the pair distribution function,

$$s_{x,y} = \int_\Omega dz p^d(z) \delta\{R(x, z) < R(x, y)\}, \tag{103}$$

$$s(r) = \int_x dx \int_y dy\, s_{x,y} p^d(x) p^d(y) \delta(r - R(x, y)), \tag{104}$$

$$s(r) = \int_x dx \int_y dy\, s_{x,y} p^d(x) p^d(y) \delta\{r > R(x, y)\} \tag{105}$$

In the meantime, the cumulative distribution of pair distribution is (with p the pair distribution function),

$$\Phi(r) = \int_0^r p(s) ds, \tag{106}$$

$$\Phi(r) = \int_0^r ds \int_\Omega dx \int_\Omega dy\, {}^d(x) p^d(y) \delta(s - R(x, y)), \tag{107}$$

$$\Phi(r) = \int_\Omega dx \int_\Omega dy\, p^d(x) p^d(y) \delta\{r > R(x, y)\}, \tag{108}$$

$$\Phi(r) = s(r). \tag{109}$$

If we come back to the derivation of the pair distribution,

$$f(x,y)\mathrm{d}^2x\mathrm{d}^2y = \pi(x,y)\varrho^d(x)\varrho^d(y)\mathrm{d}^2x\mathrm{d}^2y. \tag{110}$$

In our continuous gravity law model, we assume that $\pi(x,y) = \pi(R(x,y)) = \pi(r)$. On the other hand, the radiation model assumes that,

$$\pi(x,y) = s(x,y)^{-2} = \Phi(x,y)^{-2} \tag{111}$$

If we assume that on average, $s(x,y) \simeq \int_0^{|x-y|} \mathrm{d}r p(r) = \Phi(r)$. Then the two models will agree on average if,

$$\Phi(r)^2 = r \tag{112}$$

The pair distribution function would be $p(r) = \pm r^{-0.5}$. While the empirical pair distribution (Fig.2a) exhibits more complex patterns than this, and the two models appear to be irreconcilable.

3.2 The intrinsic distance cost as the inverse of distance

Consider an individual situated at location i. The utility associated with moving to location j can be expressed as follows,

$$U(i,j) = V(j) - C(r(i,j))$$

where $V(j)$ represents the inherent values of location j and $C(r(i,j))$ is a function of the distance $r(i,j))$ between i and j.

The probability of selecting location j can be calculated using the multinomial logit model, as outlined in [16],

$$f(i,j) = \frac{e^{U(i,j)}}{\sum_k e^{U(i,k)}}$$

The probability to move of a distance r is the sum over all locations k for which $r(i,k) = r$. This probability becomes independent of i. Thus, the probability of moving of a distance r aggregates over all locations i as $f(r) = \sum_i f(i,r)f(i)$. For simplification, let us that $f(i)$ is constant, implying that the probability of movement from any given location is equal. Furthermore, let us assume that for all $i,j \in \Omega$, $f(i,r) = f(j,r)$. In this case $f(r) \propto Nf(i,r)$.

Moreover, the number of locations k satisfying $r(i,k) = r$ corresponds to the pair distribution function, $p(r)$, up to a constant, with this constant being the total number of locations.

$$f(r) = \frac{\sum_{i\in\Omega}\sum_{k|r(i,k)=r} e^{U(i,k)}}{\sum_k e^{U(i,k)}}$$

Assuming a uniform value $V(k)$ across all locations, we derive:

$$f(r) = \frac{\sum_{i\in\Omega}\sum_{k|r(i,k)=r} e^{V-C(r(i,k))}}{\sum_k e^{U(i,k)}} \sim p(r)e^{V-C(r)}$$

with $p(r)$ the pair distribution function evaluated at r.

Given the human tendency to perceive distances logarithmically, especially for larger magnitudes, we posit that the cost function $C(r) \propto \log(r)$. This hypothesis is supported by the observation that humans might perceive numerical differences in a logarithmic fashion, such as perceiving the difference between 1 and 2 similarly to the difference between 10 and 20 or even 100 and 200, as demonstrated in [17]. Consequently, upon normalization by the pair distribution function, we derive the equation

$$\frac{f(i,r)}{p(i,r)} \sim e^{V-\log(r)} = e^{\log(V_2)-\log(r)} = \frac{V_2}{r}$$

where $V = \log(V_2)$. This leads to

$$\frac{f(i,r)}{p(i,r)} \sim \frac{1}{r}$$

which is equivalent to equation (**??**) of the intrinsic distance cost, given $\pi(r) = \frac{1}{r}$.

3.3 Reconciling Opportunity and Distance-Based Mobility Paradigms

As highlighted in [18, 19, 20], in classical mobility studies, two divergent theories explain the movement of individuals. The first, inspired by Newton's law of gravity, suggests that mobility decreases as the physical distance between locations increases, often modeled as a power law of distance [1]. The second theory argues that mobility is not directly related to distance but to intervening opportunities. It suggests that individuals prioritize locations based on the availability of closer opportunities rather than distance itself, leading to movements driven more by the opportunity distribution than by travel constraints [21, 15].

The gravitational model describes how the flow of individuals between two locations decreases as the distance between them increases, analogous to the decrease in gravitational pull with distance in Newton's theory. This decline in mobility is modeled using functions such as exponential or power laws [22, 23, 24]. In this framework, the population sizes of the start and end locations act like masses, attracting travel in direct proportion to their size and inversely to the square of the distance between them. Known for its simplicity and ease of computation, the gravity model is widely used in various fields such as migration,

intercity communication, and traffic flow analysis. It can also be extended by integrating socio-economic characteristics of locations to improve its accuracy and applicability [25].

Conversely, the intervening opportunity model emphasizes that mobility is mainly driven by the availability of intervening opportunities rather than distance [26]. It suggests that individuals are influenced more by the availability of nearby opportunities than by the distance to distant opportunities, with the spatial distribution of these opportunities dictating destination choices, making distance a less critical factor. The radiation model [15], an extension of this theory, streamlines the analysis by assuming that the chosen opportunity is the optimal one. This model correlates opportunity density with population density and provides a mathematical formula for predicting trip endpoints. The model has also been extended to a continuous spatial framework, providing a nuanced understanding of how social factors influence movement patterns [27]. **??**

Our framework provides a unifying approach to the intervening opportunity model and distance-based model In the equation **??**,

$$f(r) \propto \pi(r)p(r). \tag{113}$$

The pair distribution function $p(r)$ captures the available opportunities to move, when we divide the moving distance distribution by the pair distribution, we normalize the empirical mobility traces by their intervening opportunities. Consequently, the remaining quantity is found to be dependent on the distance, with the result that $pi(r) = f(r)/p(r) \simeq 1$. This can be interpreted as a distance term analogous to the gravity model. Furthermore, our framework explain why intra-city movements are better fitted by a gravity model with an exponential decay, which is similar to the exponential decay of population density [28].

3.4 Local piece-wise power law

In this section, we investigate whether the continuous gravity model still holds at the scale of cities. The observed moving distribution $f(r)$ can be decomposed into the observed moving distribution of each city c_i in the set $\mathcal{C}$ of cities in Denmark,

$$f(r) = \sum_{c_i \in \mathcal{C}} f_{c_i}(r), \tag{114}$$

where $f_{c_i}(r)$ only counts moves originating from city c_i, i.e. $f_c(r) = \int_{x \in c} \mathrm{d}x \int_0^\infty \mathrm{d}y f(x, y)$. For each city's observed moving distance the appropriate normalization is not the country-wide pair distribution but the relative pair distribution function, which counts pairs with at least one location in the city of interest. We define,

$$p_c(r) = \int_{x \in c} \mathrm{d}x \int_{y \in \Omega} \varrho^d(x)\varrho^d(y)\mathrm{d}x\mathrm{d}y. \tag{115}$$

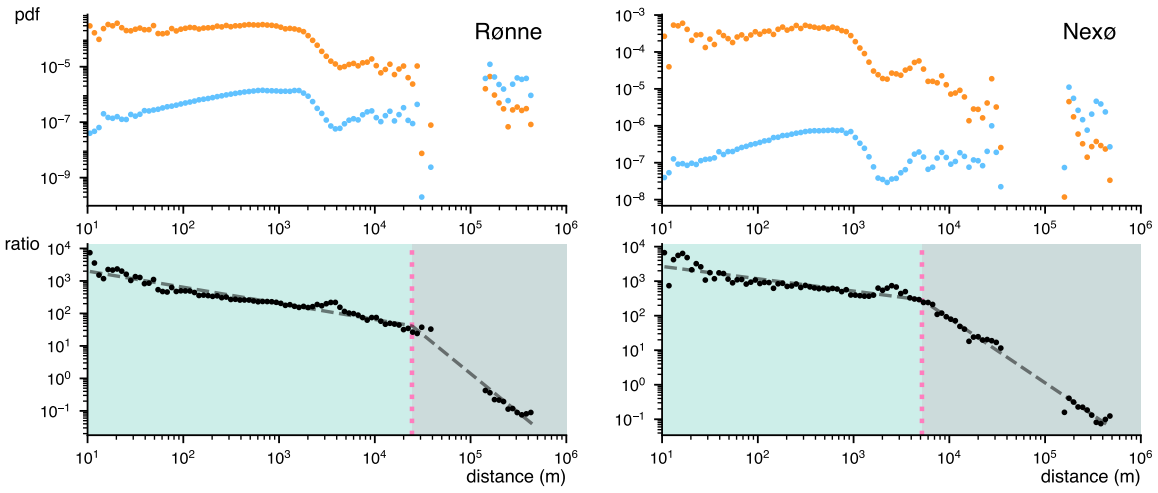

Figure 20: Piecewise intrinsic distance for the city of Rønne and Nexø both situated on an island.

We can now compute the relative intrinsic distance cost, $\pi_c(r) = f_c(r)/p_c(r)$. In the main manuscript Figures.3f-g show that the relative intrinsic distance cost does not follow the continuous gravity law and leads to a piece-wise intrinsic distance cost, Figures 22-26 show the same process for many more cities. The piece-wise intrinsic distance cost is a piece-wise power law distribution of the form:

$$\pi_c(r) = \begin{cases} C_1(R_c/r)^{\beta}, & 0 \leq r \leq R_c \\ C_2(R_c/r)^{\alpha}, & r > R_c \end{cases} \tag{116}$$

where R_c is the mobility inferred city radius, $\alpha > 1$, $\beta \in \mathbb{R}$, and $x \in \mathbb{R}^+$. The three parameters α, β, r_c are estimated by maximum-likelihood and compared to a log-normal, Pareto, and exponential Pareto distribution [**?**]. The likelihood ratios and p-values for all cities are reported in figures 27-29. We obtain that on average, $\beta = 0.60 \pm 0.20$, $\alpha = 2 \pm 0.21$, R_c follow a heavy-tailed distribution between a power law and a log-normal distribution.

3.4.1 The remarkable case of islands

In the section, we highlight the remarkable point that the piece-wise two-step intrinsic distance cost for cities still holds remarkably well for islands (Fig.20). In the case of islands, the inter-city part of the piece-wise process (when the exponent is on average 2), it mostly across the sea, and therefore the relative pair distance distribution is null. However, if the distance is large enough to reach the nearest landmass, the pair distance distribution is positive again and the 2 exponent can be interpolated between the island and the nearest landmass (see Fig.20).

3.5 Simulation of piece-wise intrinsic distance cost

In this section, we try to recover the intrinsic distance cost, $\pi(r) = 1/r$ from the aggregation of the piece-wise process of equation 116. We can decompose the intrinsic distance cost as follows,

$$\pi(r) = \frac{f(r)}{p(r)} = \sum_{c \in \mathcal{C}} \frac{f_c}{p(r)} = \sum_{c \in \mathcal{C}} \pi_c(r) \frac{p_c(r)}{p(r)} \tag{117}$$

The only term that is unknown in this equation is $p_c(r)$ the relative pair distance distribution function for each city. However, we have already shown that the pair distribution of cities follows a generalized gamma distribution within the city boundary. Beyond the city boundary, we have shown that the positions of cities follow an ideal gas (i.e. random position but no overlap), which implies that we can model the relative pair distribution function with a linear growth beyond the city radius. The relative pair distribution function of a city, c, is thus,

$$\hat{p}_c(r) = \begin{cases} C_1(r/R_c^2)\exp(-(r/R_c)^m), & 0 \leq r \leq R_c \\ C_2 r, & r > R_c \end{cases} \tag{118}$$

where C_1 and C_2 are two constants such that $\hat{p}_c$ is continuous. To obtain the global pair distribution we need to weight each relative pair distribution by its contribution, which is equal to the number of addresses. We assume that the number of addresses is proportional to the square of the radius and $p_c = R_c^2 \hat{p}_c$. For simplicity, we also assume that the general pair distribution function p follows,

$$p(r) = \begin{cases} C_3 r^{2/3}, & 0 \leq r \leq R_{\max} \\ C_4, & r > R_{\max} \end{cases} \tag{119}$$

where $R_{\max} = \max\{R_c, c \in \mathcal{C}\}$ the largest city radius. The form of the pair distribution function closely follows the empirical one of Figure 2a. Therefore, we obtain the intrinsic distance cost when aggregating over the city scale,

$$\pi_c(r)\frac{p_c(r)}{p(r)} = \begin{cases} C_5(r^{-1/6}/R_c^2)\exp(-(r/R_c)^m), & 0 \leq r \leq R_c \\ C_6 r^{-5/3}, & R_c < r \leq R_{\max} \\ C_7 r^{-1}, & r > R_{\max} \end{cases} \tag{120}$$

Remarkably, the inverse of distance scaling of the intrinsic distance cost for $r \geq R_{\max}$emerges as a sole consequence of the random distribution of city locations. The scaling $r \leq R_{\max}$ results from the aggregation over the scale of cities, akin to our patch model SI. 2.6 or [29]. We obtain that for $r \leq R_{\max}$,

$$\pi(r) = \int_0^r C_5(r^{-1/6}/R_c^2)\exp(-(r/R_c)^m)g(R)\mathrm{d}R + \int_r^\infty C_6 r^{-5/3} g(R)\mathrm{d}R \tag{121}$$

where g is the distribution of the city radius. From the main text Fig.3, g can be modeled as a log-normal or a power-Pareto distribution. We solve the equation 121 numerically by simulating over the city intrinsic distance cost over an artificial geography of Denmark (eq.118,119). We recover the intrinsic distance cost $\pi(r) = 1/r$ (Fig.21).

Figure 21: For each city in Denmark, the distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, and the dark teal to inter-city moves.

Figure 22: For each city in Denmark, the distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, and the dark teal to inter-city moves.

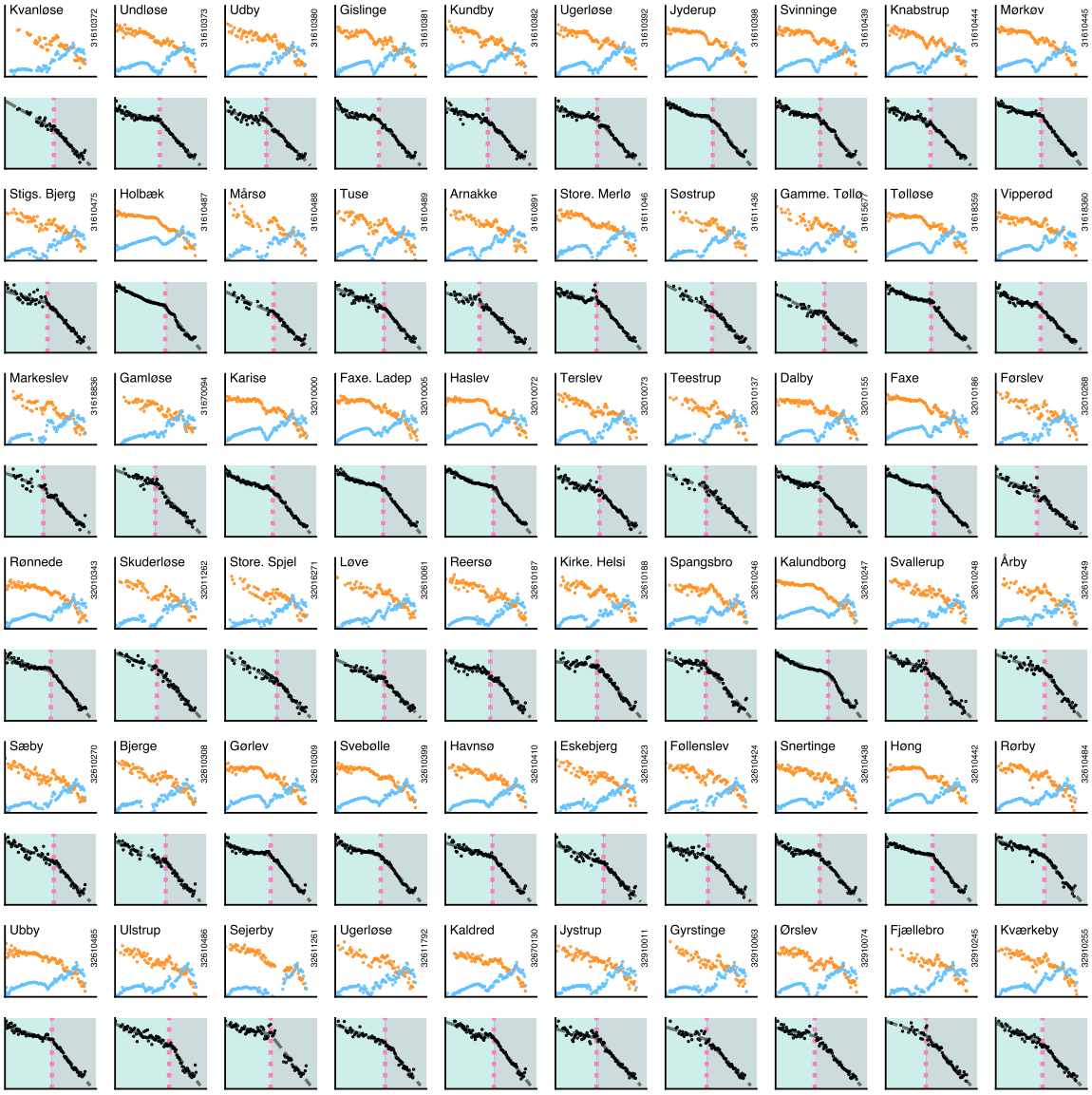

Figure 23: For each city in Denmark, the distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, and the dark teal to inter-city moves.

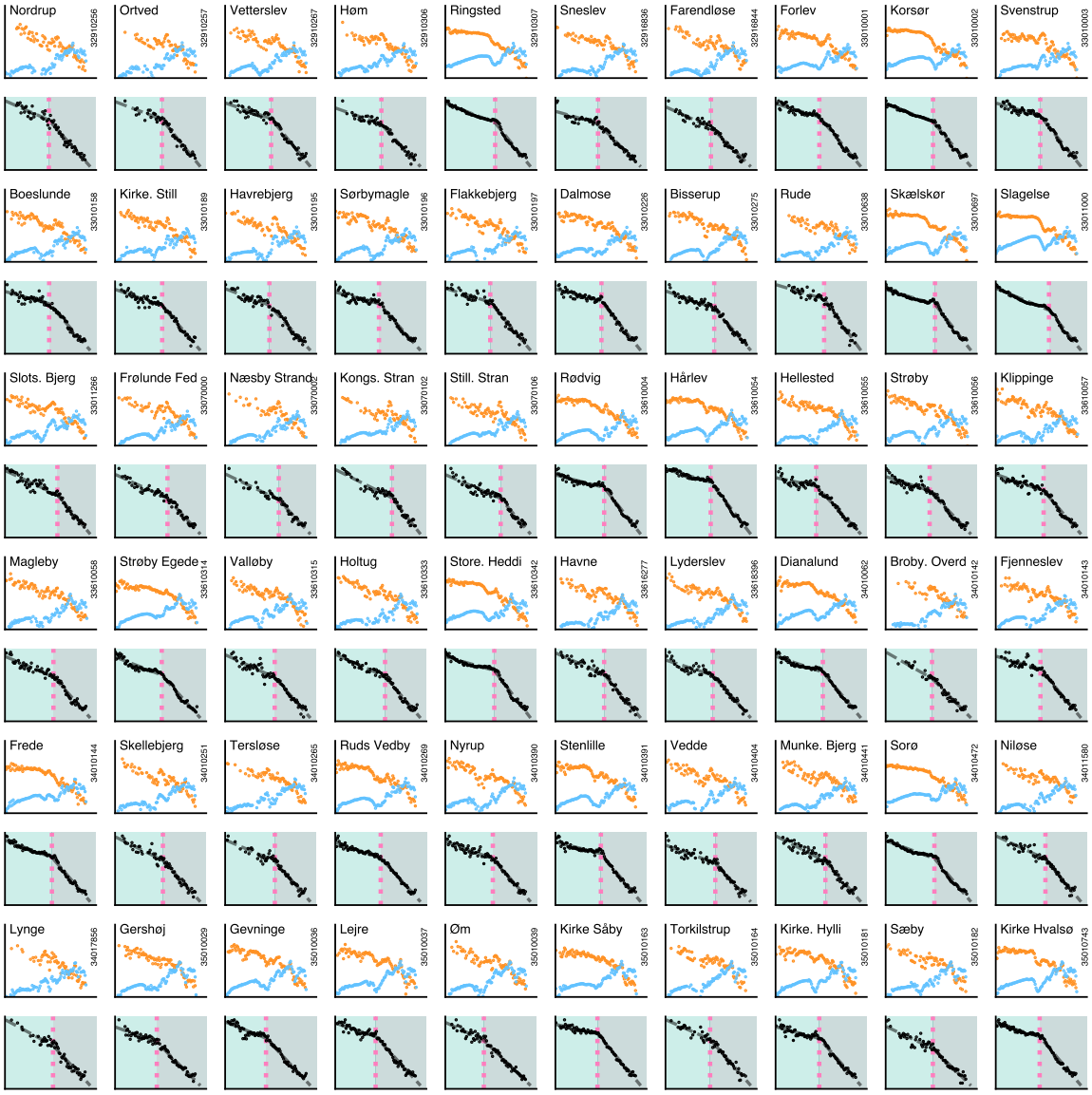

Figure 24: For each city in Denmark, the distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, and the dark teal to inter-city moves.

Figure 25: For each city in Denmark, the distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, and the dark teal to inter-city moves.

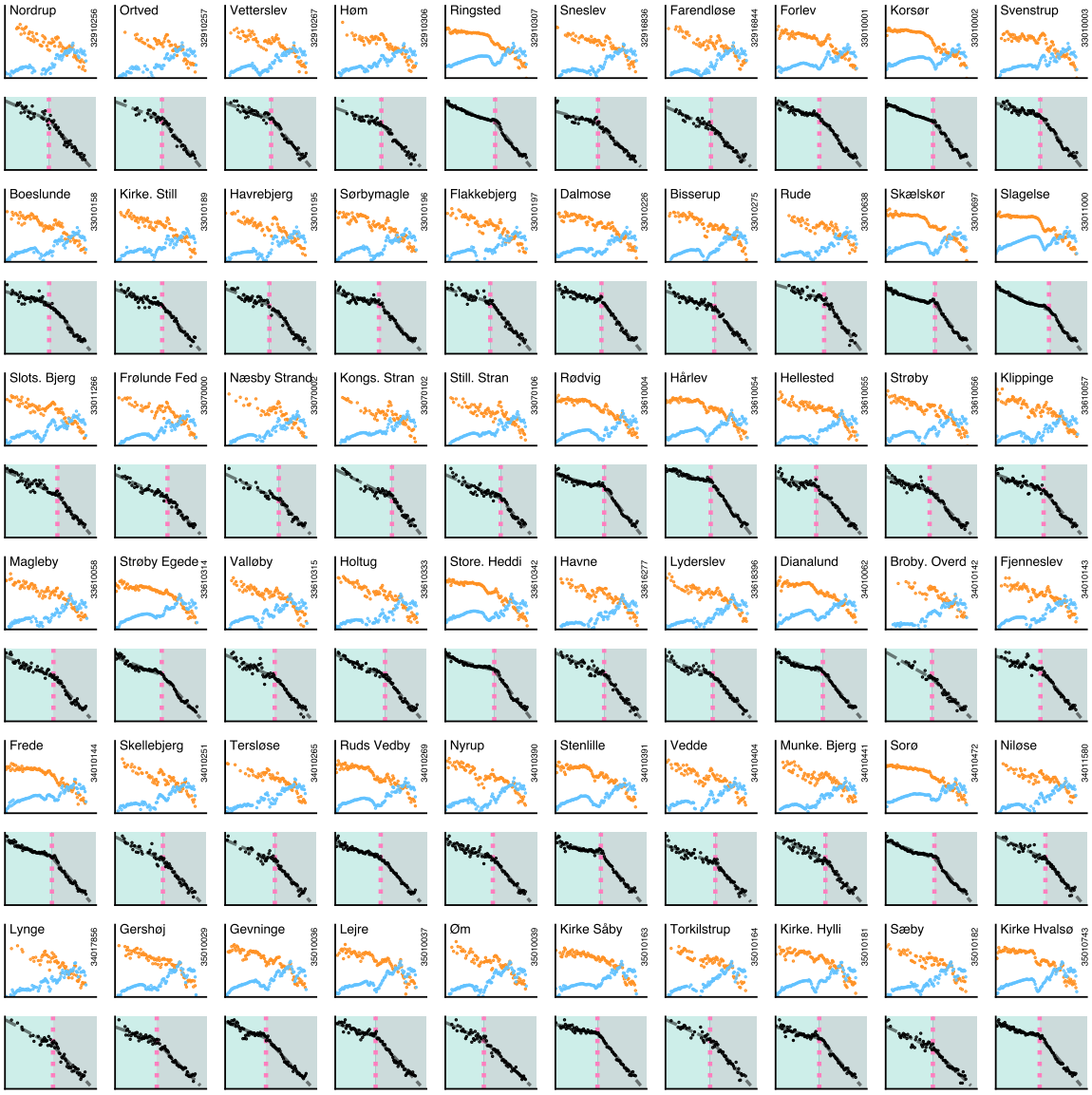

Figure 26: For each city in Denmark, distribution of observed moving distance (orange) and the relative pair distribution function. Pairs are restricted to those containing at least one address from the city of interest. The second plot shows the intrinsic distance cost (black), as the ratio between the observed moving distance and the relative pair distribution function. The pink dashed line is the inferred mobility city radius, the light teal corresponds to intra-city moves, the dark teal to inter-city moves.

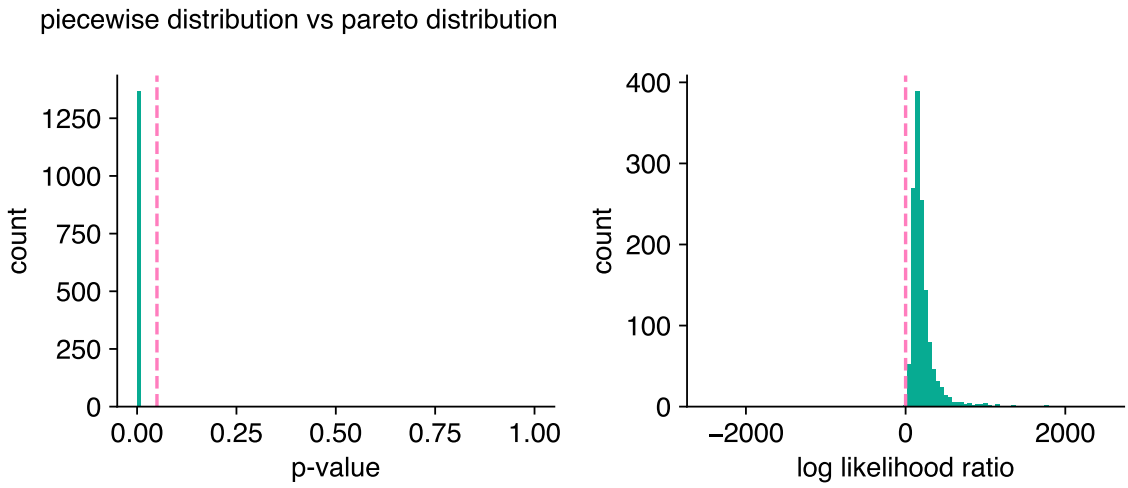

Figure 27: For each city of the piecewise fit, Figure.22-26, we compare the log-likelihood ratio between the piecewise power law distribution of equation 116 and a Pareto distribution. The left plot shows the distribution of p values with a pink line at $p = 0.05$. The right plot shows the distribution of the log-likelihood ratio, with a pink line at the decision boundary $R = 0$. A positive log-likelihood ratio indicates that the piecewise models describe the data better, the p-value give the significance of the result.

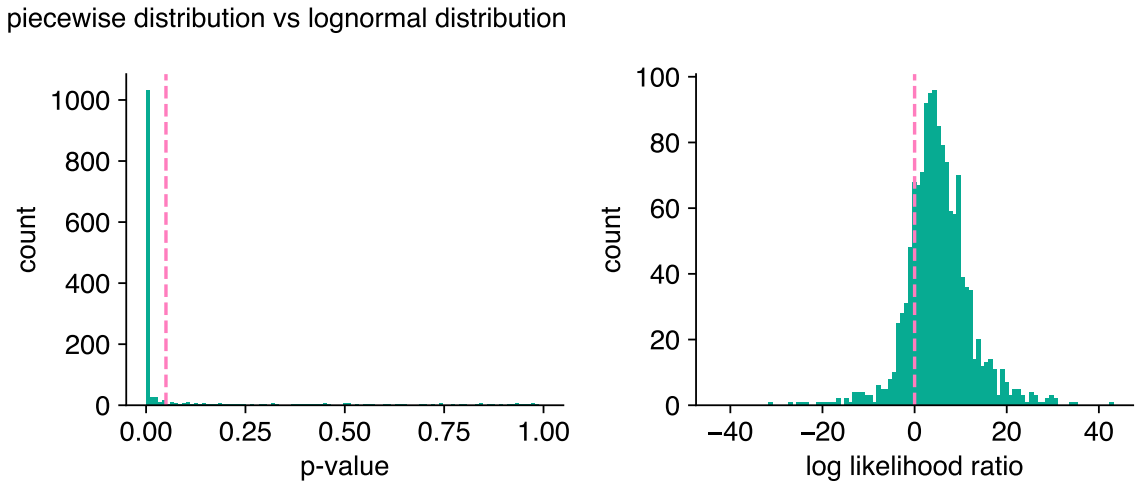

Figure 28: For each city of the piecewise fit, Figure.22-26, we compare the log-likelihood ratio between the piecewise power law distribution of equation 116 and a lognormal distribution. The left plot shows the distribution of p values with a pink line at $p = 0.05$. The right plot shows the distribution of the log-likelihood ratio, with a pink line at the decision boundary $R = 0$. A positive log-likelihood ratio indicates that the piecewise models describe the data better, the p-value give the significance of the result.

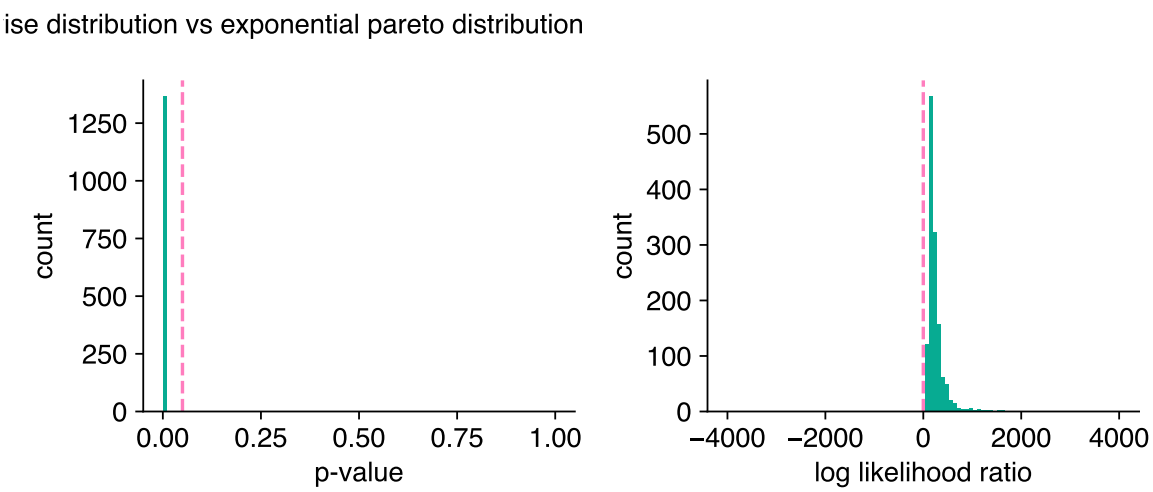

Figure 29: For each city of the piecewise fit, Figure.22-26, we compare the log-likelihood ratio between the piecewise power law distribution of equation 116 and a piecewise exponential-Pareto distribution. The left plot shows the distribution of p values with a pink line at $p = 0.05$. The right plot shows the distribution of the log-likelihood ratio, with a pink line at the decision boundary $R = 0$. A positive log-likelihood ratio indicates that the piecewise models describe the data better, the p-value give the significance of the result.

4 The definition of city

The precise demarcation of a city — where a city ends and its boundaries — is a pressing issue in academic literature [30]. Depending on the definition adopted, different results can be obtained [31]. To ensure that our results are not just a byproduct of our chosen city definition and that they could potentially be applied to other definitions, we evaluated the robustness of the city boundaries.

First, we adopted the city definition provided by Danmarks Statistik [32]. According to this definition, Denmark consists of 1,473 cities.

To ensure more robust results, we compare the city definition to cities defined by density-based clustering techniques. These methods distinguish densely clustered data points that represent urban or urbanized areas from sparse or noisy regions that typically correspond to rural or sparsely populated areas. DBSCAN is particularly well-suited for this task because of its innate ability to delineate clusters of different geometries, which is critical for accommodating the non-uniform shapes of urban regions. This algorithm sorts data points into core points, boundary points, and noise based on surrounding data density. However, one of the main challenges is to determine the optimal values for ϵ (neighborhood search radius) and minimum data points, as these parameters can determine the level of detail of the identified urban zones. [33] .

For data covering cities with different population densities, HDBSCAN stands out. Building

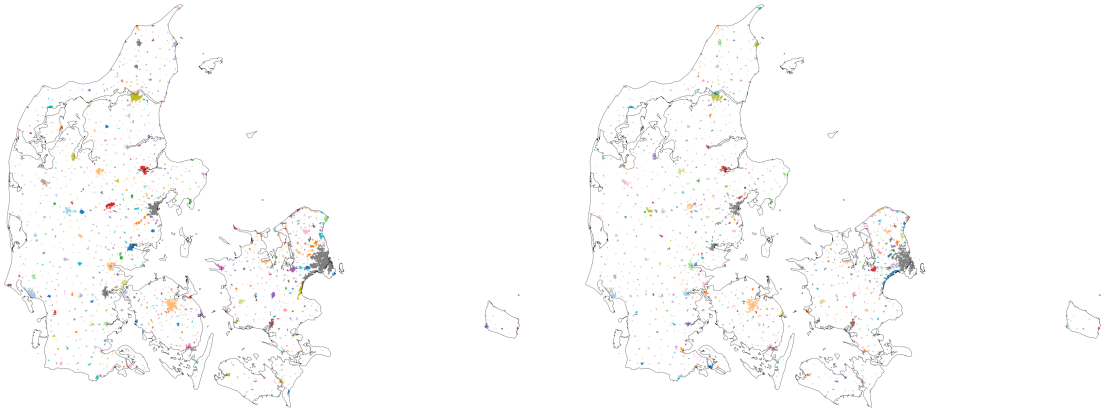

Figure 30: Comparison between the Danmarks Statistik definition of cities (on the left) and the cluster obtained from HDBSCAN (on the right). The colors are ordered according to the size of the cluster/city.

on the foundation laid by DBSCAN, HDBSCAN employs a hierarchical tactic that makes the fixed ϵ value redundant. By examining the density hierarchy, HDBSCAN can differentiate between densely populated cities and smaller towns within the same data collection. This granularity, coupled with the flexible clustering results, provides a detailed view of the boundaries of cities [34].

Copenhagen, however, presents a unique scenario. Its urban influence extends well beyond its administrative boundaries, as evidenced by phenomena such as commuting patterns. As a result, neighboring cities have been merged with Copenhagen to form the Copenhagen metropolitan area, or Hovedstadsområdet, as defined by DST. Notably, the clustering results from HDBSCAN reflected this merging, indicating the combined urban sprawl of the Copenhagen region.

Normalized Mutual Information (NMI) and Adjusted Mutual Information (AMI) quantify the similarity between two clusterings. Both metrics measure the information shared between the true labels and the labels assigned by a clustering algorithm. Therefore, we can use them to evaluate the performance of clustering algorithms in the absence of ground truth labels. The NMI is defined as the mutual information between two clusterings divided by the geometric mean of their entropies [35]. However, a limitation of the NMI is that it ignores the random grouping of clusters, i.e. random cluster assignments can produce a non-zero NMI value. The AMI, on the other hand, corrects for this limitation by adjusting the score to account for chance, ensuring that random cluster assignments result in an AMI score close to zero [36]. Therefore, the AMI provides a more accurate representation of the similarity between two clusterings in our case where the number of clusters is not fixed.

The table 2 shows the value of the indices for the two algorithms. An NMI/AMI score of

0.9 indicates that the two clusterings share a significant amount of information. Such a high score typically indicates that the two clusterings are almost identical, with only a few data points potentially clustered differently. Therefore, the city definition from Danmarks Statistics is substantially similar to the one we obtain using density-based clustering techniques. Furthermore, table 1 shows that merging Copenhagen into the Hovedstadsområdet is a better definition of a city according to the density-based clustering techniques.

Algorithm	NMI	AMI
DBSCAN	0.88	0.87
HDBSCAN	0.91	0.89

Table 1: NMI and AMI values between empirical city clustering (Copenhagen merged) and algorithmic clustering (DBSCAN and HDBSCAN).

Algorithm	NMI	AMI
DBSCAN	0.71	0.67
HDBSCAN	0.67	0.65

Table 2: NMI and AMI values between the empirical cities (Copenhagen not merged) clustering and the algorithmic clustering (DBSCAN and HDBSCAN)

The parameters for HDSCAN were chosen to optimize the NMI between the clustering and the real city labels. Figure 31 shows the values for different values of *minimum samples* and *minimum cluster size*, the rest of the parameters follow the default values of this implementation [34].

Figure **??** shows all addresses colored by the cluster they belong to. The background color is the official city border.

5 Maximum likelihood estimation of power laws

Fitting the statistical power-law model, to examine the distance distribution of human movement, we fit a truncated power law with the form,

$$p(x) \propto x^{-alpha} e^{-\lambda x}. \tag{122}$$

where α is a constant parameter of the distribution (known as the scaling parameter or exponent), x is the travel distance ($x > x_{min} > 0$), and λ is the parameter of the exponential distribution. x_{min} represents the shortest distance above which the power law scaling relationship

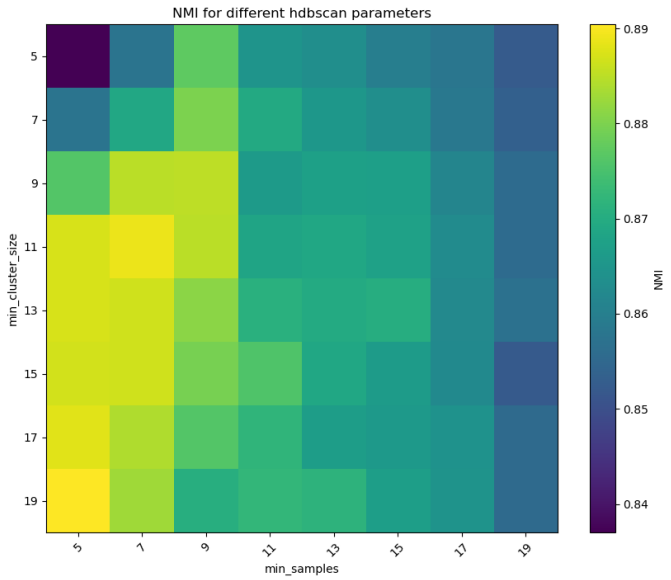

Figure 31: Heatmap of the NMI between the clustering of addresses locations and their official city and the values for different combinations of HDSCAN parameters.

begins. The scaling parameter α must be estimated before finding the optimal values of x_{min}. The methods of [37] find x_{min} by generating a power-law fit starting from each unique value in the data set, then selecting the one that gives the minimum Kolmogorov-Smirnov distance between the data and the fit. For any given value of x_{min}, we estimated the scaling parameter using maximum likelihood estimation. The goodness of fit for the truncated power law distribution was considered in comparison to the fit of other distributions (e.g., power law, exponential, and lognormal).

Although some of the power laws present a more complicated picture, we then used the methods developed in this [**?**], which carefully extend [37] procedures for a wider class of distributions.

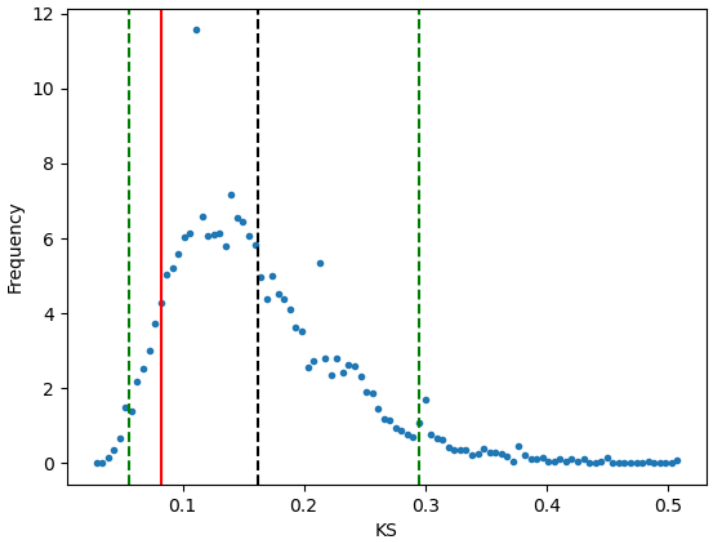

Figure 32: Kolmogorov-Smirnov (KS) test for Houston data, the blue dots represent the distribution of the test statistics over the sampled data. The mean is represented by a dashed black line. The 5th and 95th percentiles are represented by dashed green lines, indicating the range within which 90% of the observed values fall.

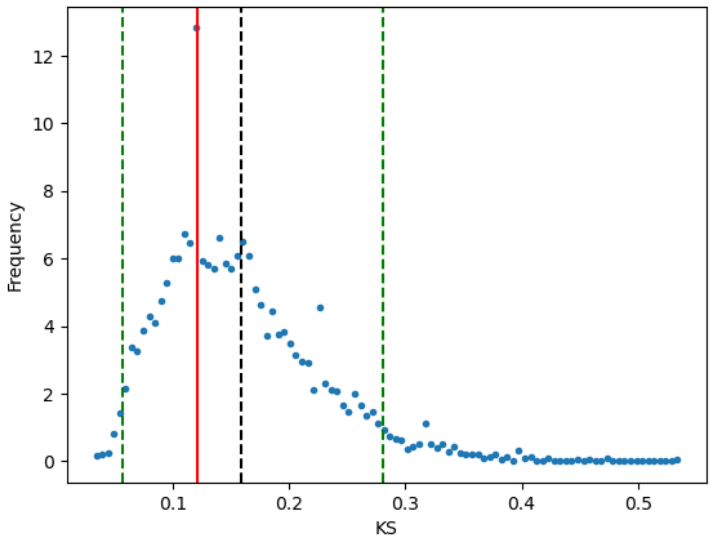

Figure 33: Kolmogorov-Smirnov test for Singapore data, the blue dots represent the distribution of the test statistics over the sampled data. The mean is represented by a dashed black line. The 5th and 95th percentiles are represented by dashed green lines, indicating the range within which 90% of the observed values fall.

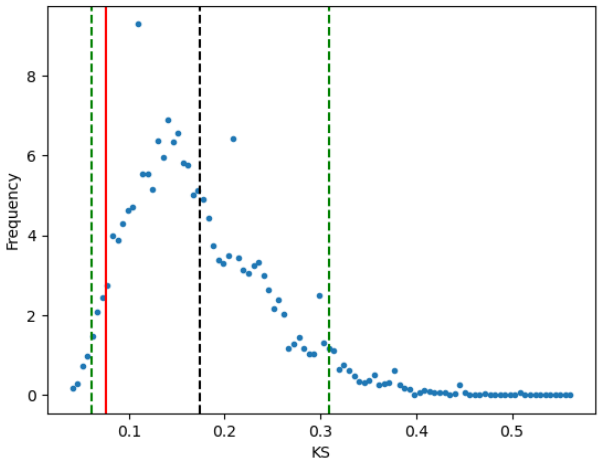

Figure 34: Kolmogorov-Smirnov test for San Francisco data, the blue dots represent the distribution of the test statistics over the sampled data. The mean is represented by a dashed black line. The 5th and 95th percentiles are represented by dashed green lines, indicating the range within which 90% of the observed values fall.

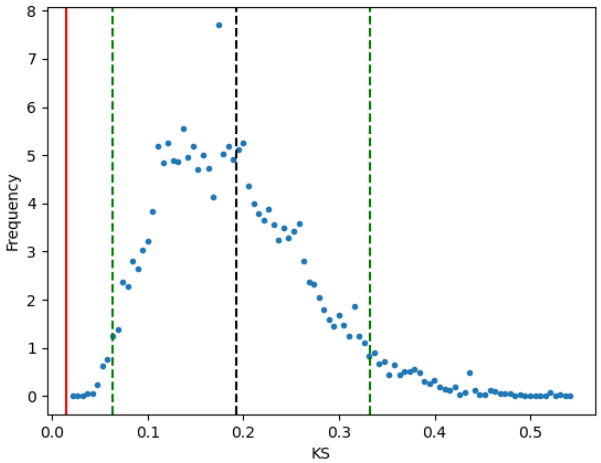

Figure 35: Kolmogorov-Smirnov test for France data, the blue dots represent the distribution of the test statistics over the sampled data. The mean is represented by a dashed black line. The 5th and 95th percentiles are represented by dashed green lines, indicating the range within which 90% of the observed values fall.

References

[1] George K. Zipf. The p1p2/d hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.

[2] Marc Wiedermann, Jonathan F. Donges, Jürgen Kurths, and Reik V. Donner. Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes. *Physical Review E*, 93(4):042308, April 2016.

[3] Benjamin F. Maier. Generalization of the small-world effect on a model approaching the Erdős–Rényi random graph. *Scientific Reports*, 9(1):9268, December 2019.

[4] Benjamin F Maier, Cristián Huepe, and Dirk Brockmann. Modular hierarchical and power-law small-world networks bear structural optima for minimal first passage times and cover time. *Journal of Complex Networks*, 7(6):865–895, December 2019.

[5] Benjamin F. Maier. *Spreading Processes in Human Systems*. PhD thesis, Humboldt University of Berlin, January 2020.

[6] Mireille Boutin and Gregor Kemper. On reconstructing n-point configurations from the distribution of distances or areas. *Advances in Applied Mathematics*, 32(4):709–735, 2004.

[7] Eric W Weisstein. Disk line picking. *MathWorld–A Wolfram Web Resource.*, 2023.

[8] J. E. Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43:461–482, 1931.

[9] Jean-Pierre Hansen and I. R. McDonald. *Theory of Simple Liquids: with Applications to Soft Matter*. Academic Press, 2nd ed edition, 1986.

[10] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[11] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[12] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, November 2009.

[13] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *PLoS ONE*, 9(1):e85777, January 2014.

[14] Hernán A. Makse, Shlomo Havlin, and H. Eugene Stanley. Modelling urban growth patterns. *Nature*, 377(6550):608–612, October 1995.

[15] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[16] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[17] Stanislas Dehaene. The number sense: How the mind creates mathematics. *Oxford University Press*, 1997.

[18] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[19] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: Universal patterns in human urban mobility. *PLOS ONE*, 2012.

[20] Mattia Mazzoli, Alex Molas, Aleix Bassolas, Maxime Lenormand, Pere Colet, and José J. Ramasco. Field theory for recurrent mobility. *Nature Communications*, 10:3895, 2019.

[21] S. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5:845–867, 1940.

[22] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant. A universal model of commuting networks. *PLoS ONE*, 7:e45985, 2012.

[23] M. Lenormand, A. Bassolas, and J. J. Ramasco. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51:158–169, 2016.

[24] S. Erlander and N. F. Stewart. *The Gravity model in transportation analysis: Theory and extensions*. VSP, Utrecht, The Netherlands, 1990.

[25] R. Patuelli, A. Reggiani, S. P. Gorman, P. Nijkamp, and F.-J. Bade. Network analysis of commuting flows: A comparative static approach to german data. *Networks and Spatial Economics*, 7:315–331, 2007.

[26] M. Schneider. Gravity models and trip distribution theory. *Papers in Regional Science*, 5:51–58, 1959.

[27] Filippo Simini, Amos Maritan, and Zoltán Néda. Human mobility in a continuum approach. *Journal Name Here*, Volume Here(Issue Number Here):Page Numbers Here, Year Here.

[28] Xiao Liang, Jichang Zhao, Li Dong, and Ke Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific reports*, 3(1):2983, 2013.

[29] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587:402–407, 2020.

[30] Elsa Arcaute, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface*, 12(102), 2015.

[31] Luís M. A. Bettencourt. The origins of scaling in cities. *Science*, 340(6139):1438–1441, 2013.

[32] Danmarks Statistik. Moduldata for befolkning og valg. `https://www.dst.dk/da/Statistik/dokumentation/Times/moduldata-for-befolkning-og-valg`, 2024. Accessed: 2024-02-26.

[33] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[34] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.

[35] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. In *Journal of machine learning research*, pages 583–617, 2002.

[36] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837–2854, 2010.

[37] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.