Scaling laws of cities

Markus Schläpfer

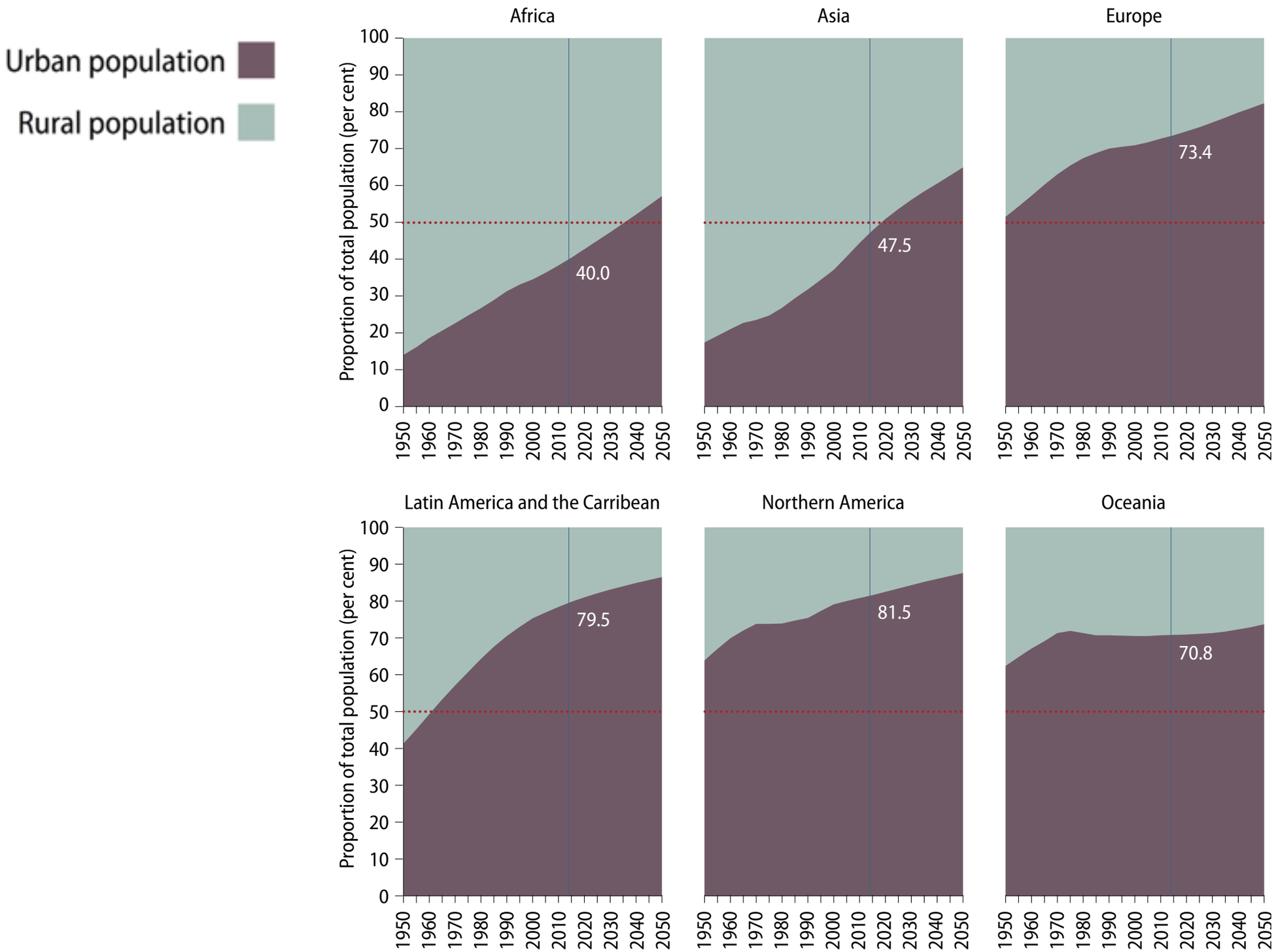
Future Cities Laboratory, ETH-Singapore Centre
School of Computer Science and Engineering, NTU

NTU Winter School on Complexity Science
March 14, 2017

Mortula / shutterstock.com

Ever-increasing complexity of cities

A. Population growth and mass urbanization



UN Population Division, 2015

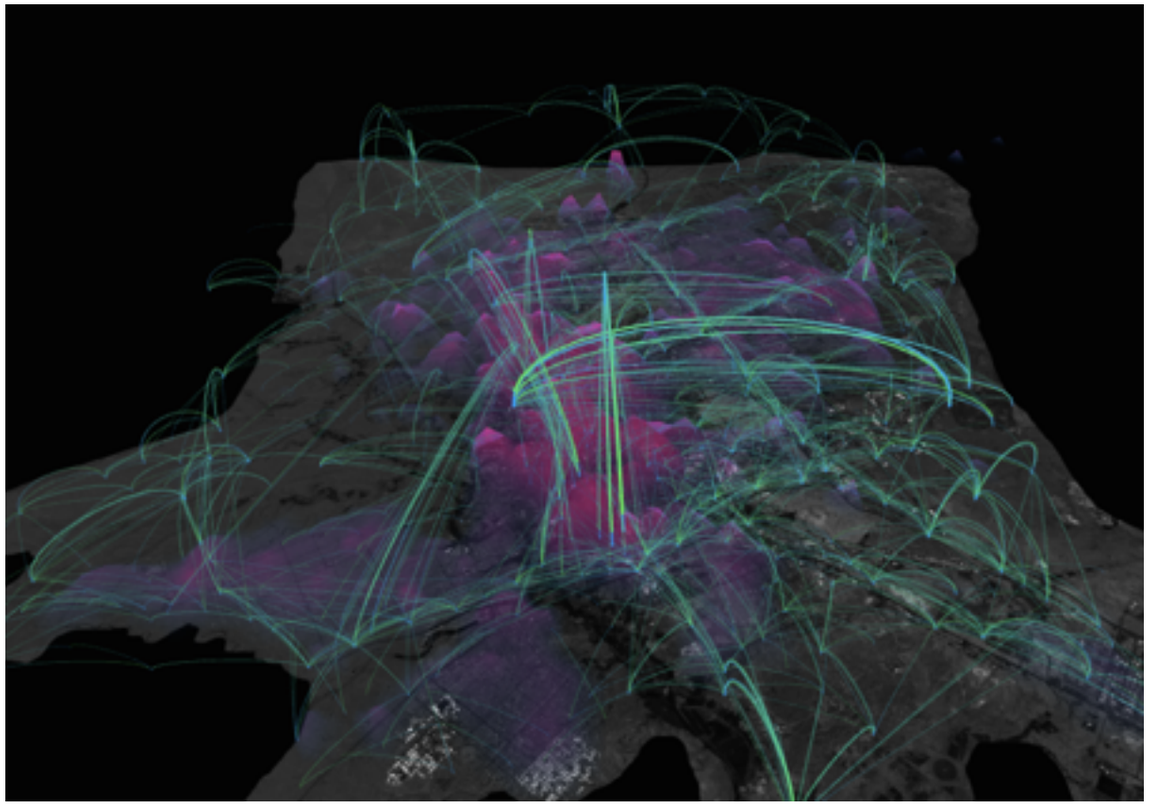
Ever-increasing complexity of cities

B. New forms of urban organization

Monocentricity

Univ. Munster

Polycentricity



MIT/Senseable City Lab, Kael Greco

Increasing uncertainties in urban planning and design

- Urban mobility
- Infrastructure design
- Social sustainability
  (social segregation, job accessibility)
- …

Urgent need for a quantitative *understanding* of cities

Growing availability of human activity data

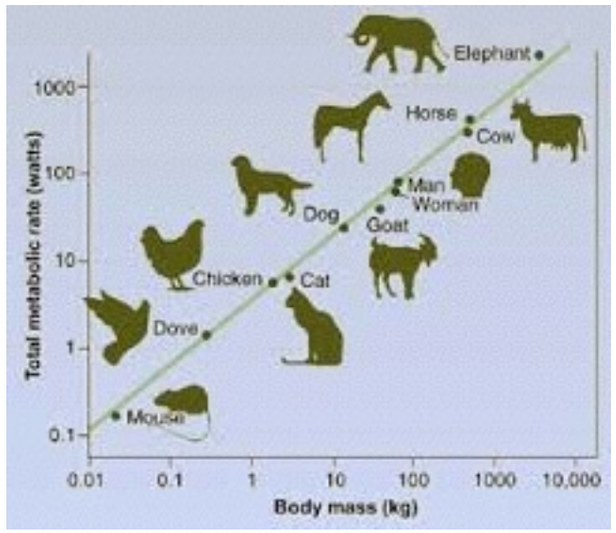
- Mobile phone data
- Smart card data from public transportation
- GPS traces from vehicular devices
- Location-based social networks
  (Foursquare, Twitter, Flickr, Running Apps, etc.)
- User-generated mapping projects (OpenStreetMap)
- Open data provided by city governments
- ….

Content

1. Urban scaling laws
2. Urban structure: building heights and shapes
3. Urban dynamics: the movement of people in cities
4. Application: infrastructure design

1. Urban Scaling Laws

The scaling of socio-economic quantities with city size



Luís M.A. Bettencourt

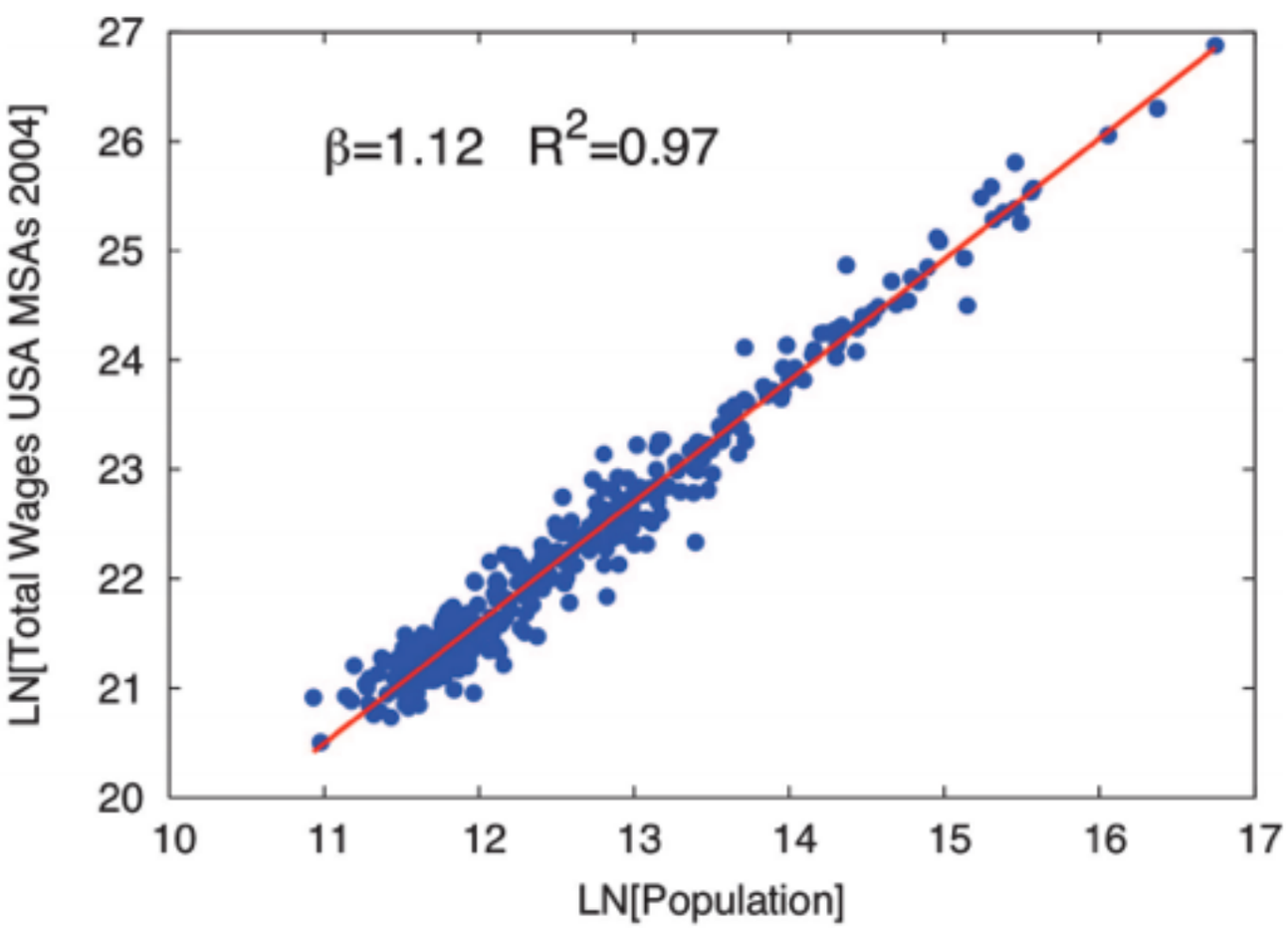
Geoffrey B. West

$$Y \propto N^{\beta}$$

Exponent

City population size

Socio-economic quantity
(wages, patents, crime,
AIDS cases etc.)



$$\beta \approx 1.15 > 1$$

L.M.A. Bettencourt et al.
Proc. Natl. Acad. Sci. USA (2007)

Greater population —
„faster life and greater dividends“

Y	β	95% CI	Adj-R^2	Observations	Country–year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
″Supercreative″ employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003

≈15% per capita increase in wages, GDP, patents etc.
for each doubling of city size

L.M.A. Bettencourt et al.
Proc. Natl. Acad. Sci. USA (2007)

Network of human interactions as a unifying mechanism?

PHYSICAL REVIEW E **79**, 016115 (2009)

Superlinear scaling for innovation in cities

Samuel Arbesman*
Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, Massachusetts 02115, USA

Jon M. Kleinberg†
Computer Science, Cornell University, Ithaca, New York 14853, USA

Steven H. Strogatz‡
Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA
(Received 29 September 2008; published 30 January 2009)

Superlinear scaling in cities, which appears in sociological quantities such as economic productivity and creative output relative to urban population size, has been observed, but not been given a satisfactory theoretical explanation. Here we provide a network model for the superlinear relationship between population size and innovation found in cities, with a reasonable range for the exponent.

nature COMMUNICATIONS

ARTICLE

Received 4 Dec 2012 | Accepted 30 Apr 2013 | Published 4 Jun 2013

DOI: 10.1038/ncomms2961

Urban characteristics attributable to density-driven tie formation

Wei Pan[1], Gourab Ghoshal[1,†], Coco Krumme[1], Manuel Cebrian[1,2,3] & Alex Pentland[1]

The Origins of Scaling in Cities

Luís M. A. Bettencourt

Despite the increasing importance of cities in human societies, our ability to understand them scientifically and manage them in practice has remained limited. The greatest difficulties to any scientific approach to cities have resulted from their many interdependent facets, as social, economic, infrastructural, and spatial complex systems that exist in similar but changing forms over a huge range of scales. Here, I show how all cities may evolve according to a small set of basic principles that operate locally. A theoretical framework was developed to predict the

Growing availability of human activity data

- Mobile phone data

Mobile phone data - exemplary data sources

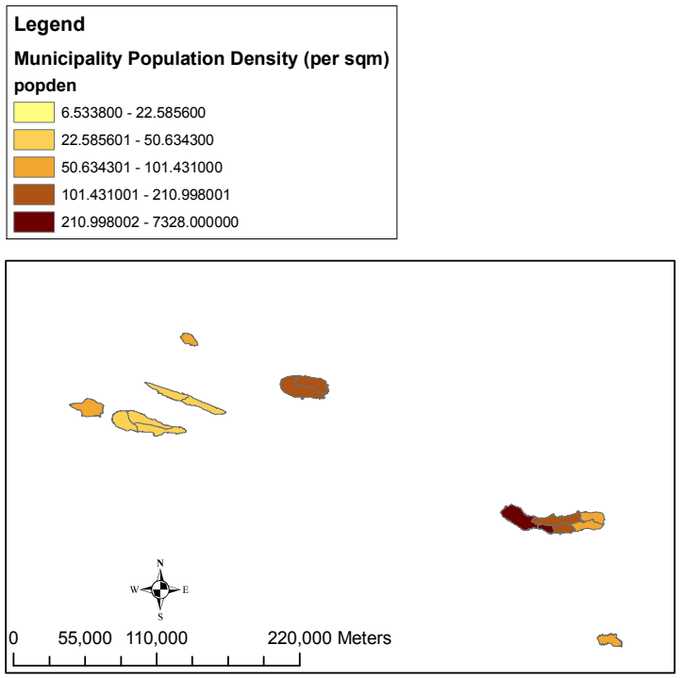
- Open data
  - *Italy* - Telecom Italia Open BigData Initiative
    http://theodi.fbk.eu/openbigdata

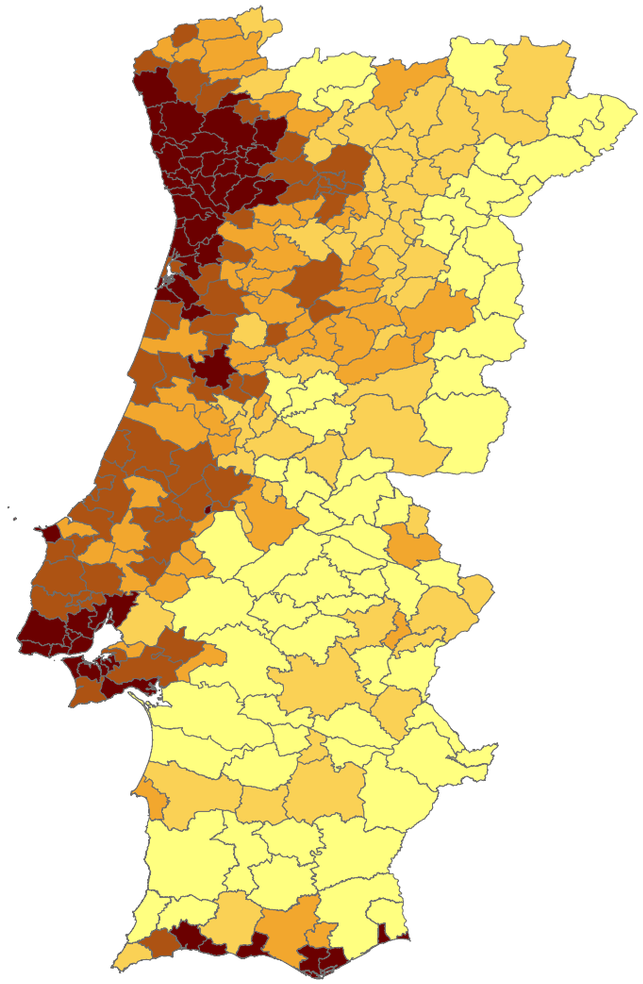
- Big data research competitions
  - *Ivory Coast* - Orange D4D Challenge 2013
    http://www.d4d.orange.com/en/Accueil
  - *Senegal* - Orange D4D Challenge 2015
    http://www.d4d.orange.com/en/Accueil

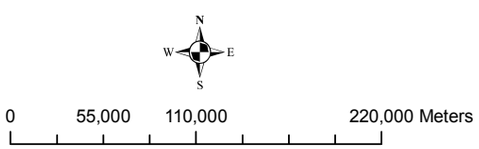
  - *Italy* - Telecom Italia BigData Challenge 2015
    http://www.telecomitalia.com/tit/en/bigdatachallenge.html

- (Telco providers and data analytics companies)

Lets look into the data!







Several millions of anonymized call detail records (CDRs) from Portugal for a period of ≈15 months

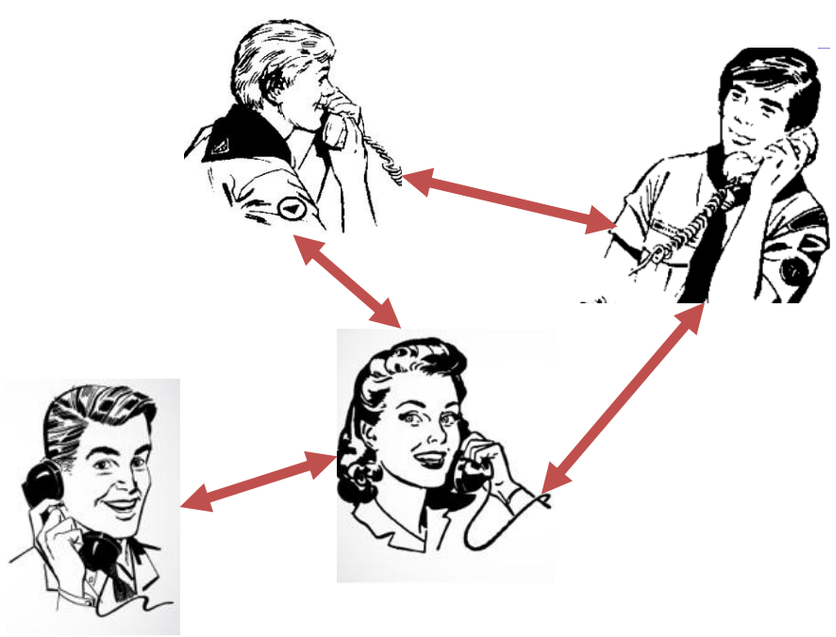
Call detail records (CDRs)

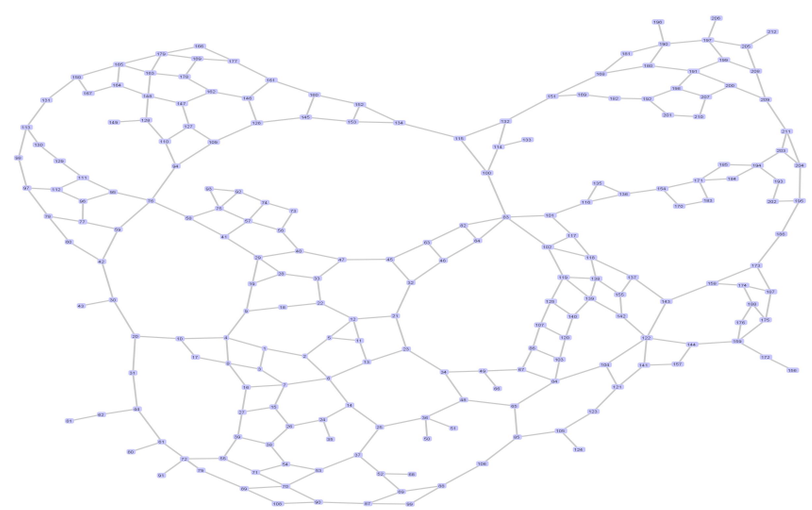
- Anonymized ID (surrogate number) of the caller
- Anonymized ID of the callee
- Start time of the call
- Duration of the call
- The locations of the antennas routing the call

Inferring the interaction network

Mobile phone user → Node

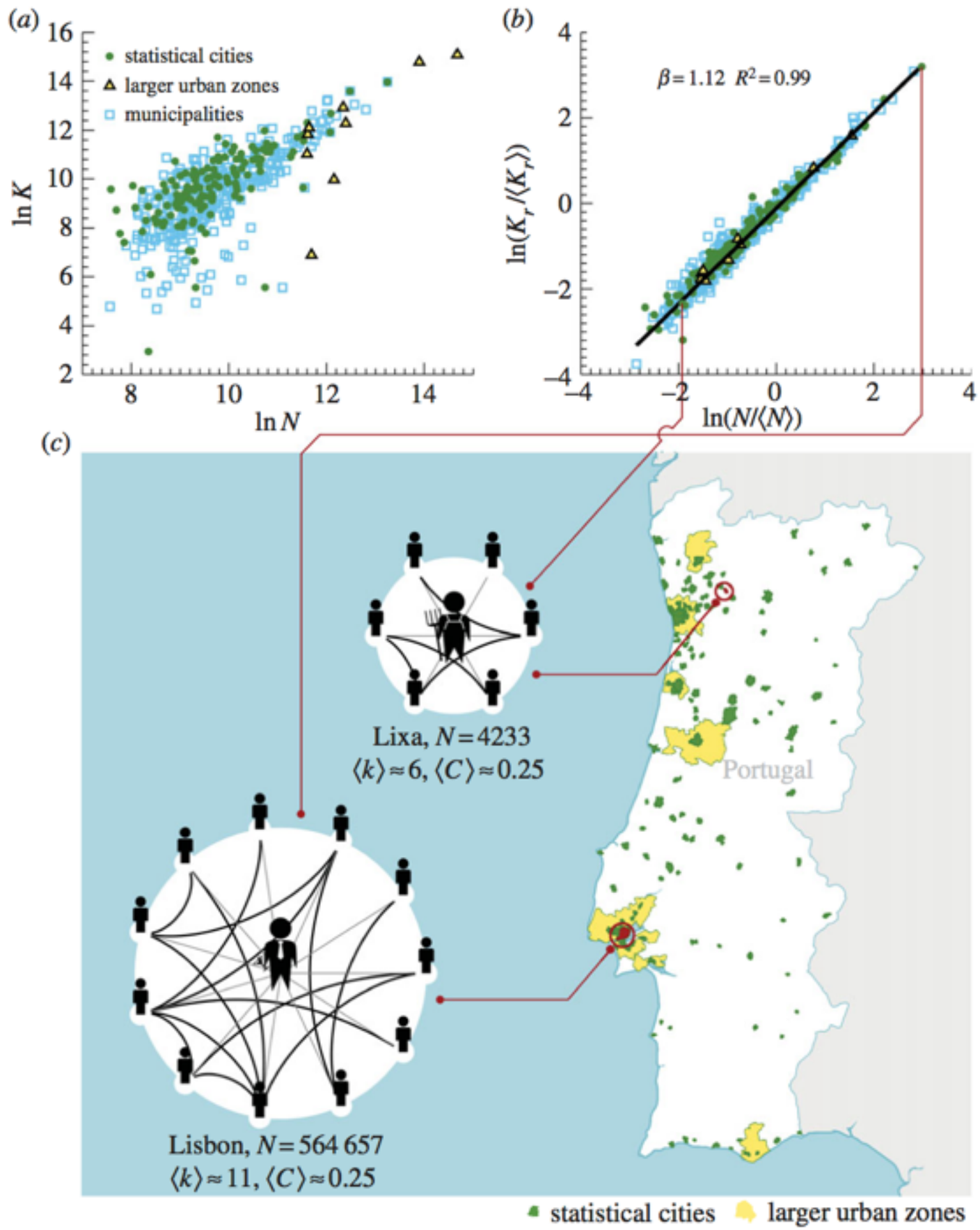
Reciprocal call between two users → Link



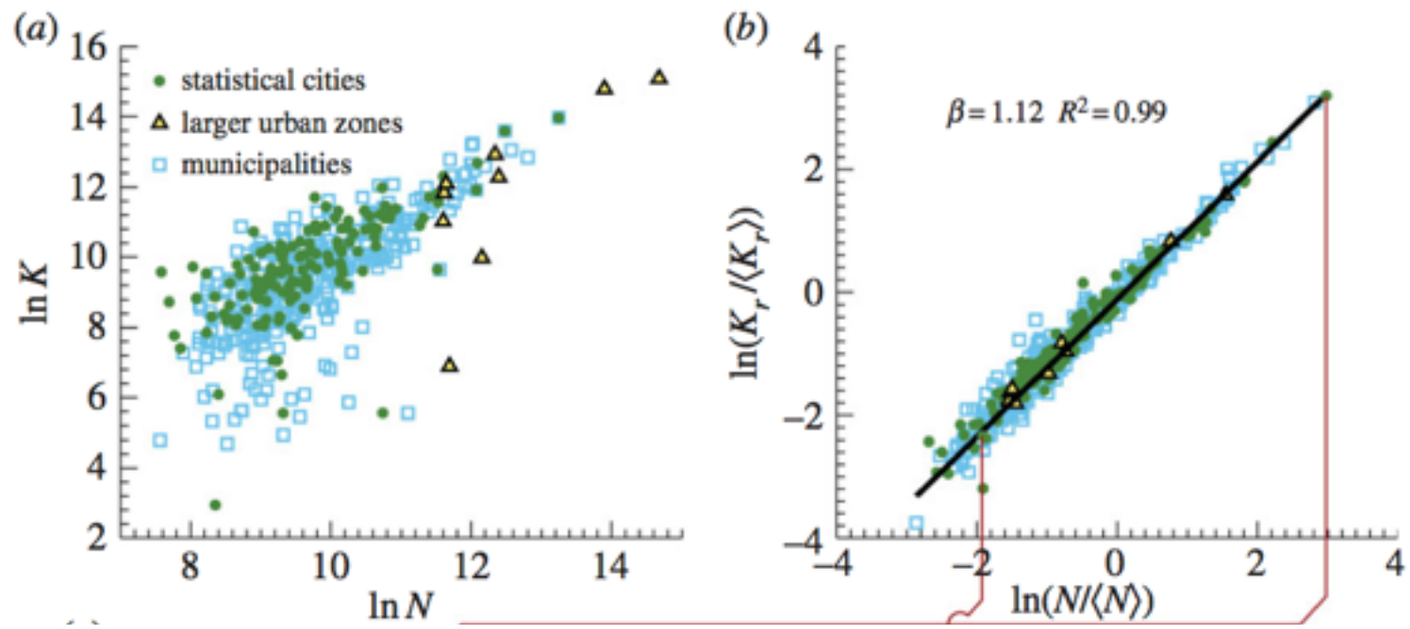


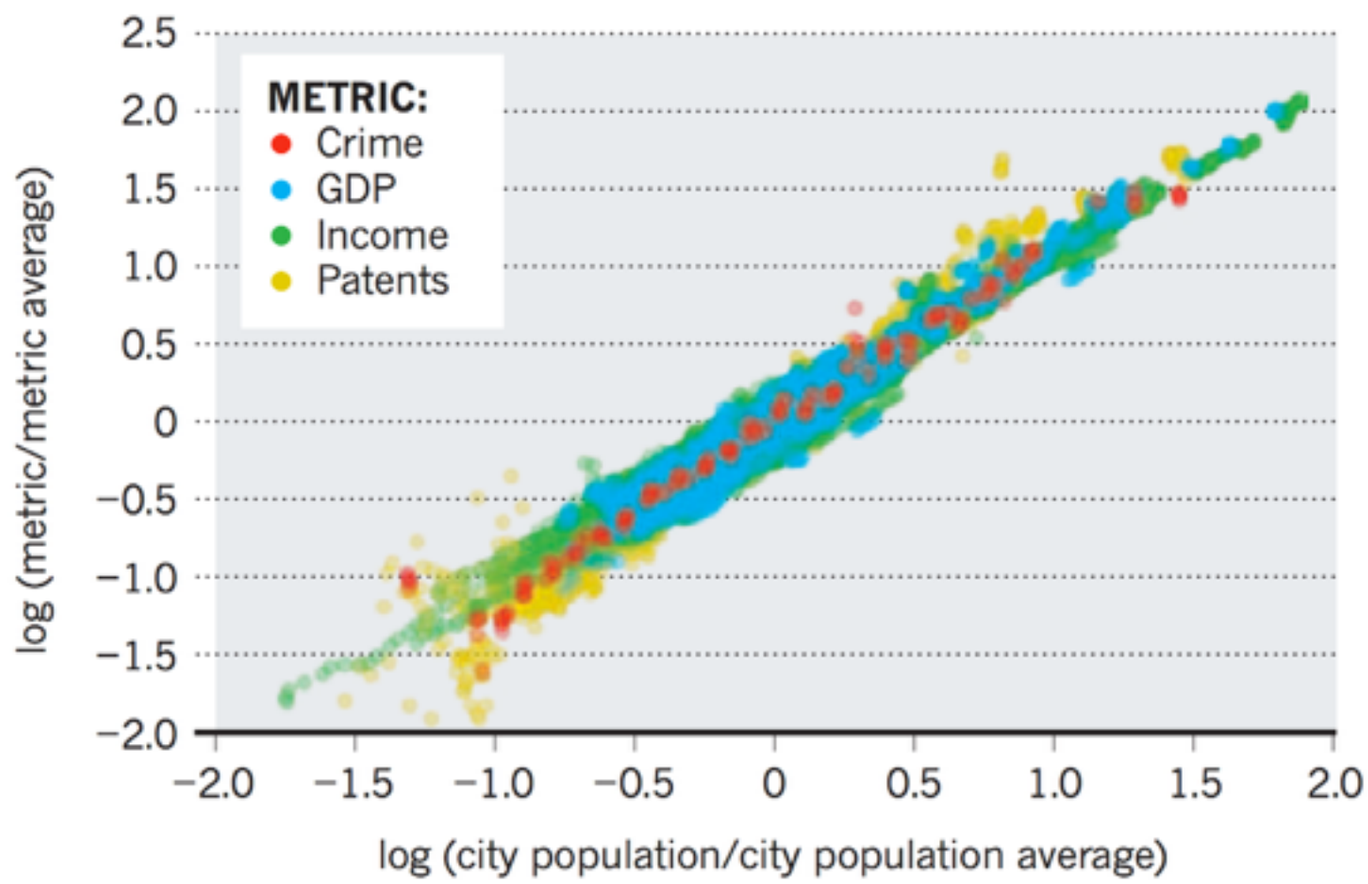
Portugal data:
1.6 Mio nodes
6.8 Mio links

Human interactions



M. Schläpfer et al. Royal Society Interface, 2014


(a)
statistical cities
larger urban zones
municipalities
ln K
ln N
(b)
β = 1.12 R² = 0.99
ln(K_r/⟨K_r⟩)
ln(N/⟨N⟩)


METRIC:
Crime
GDP
Income
Patents
log (metric/metric average)
log (city population/city population average)

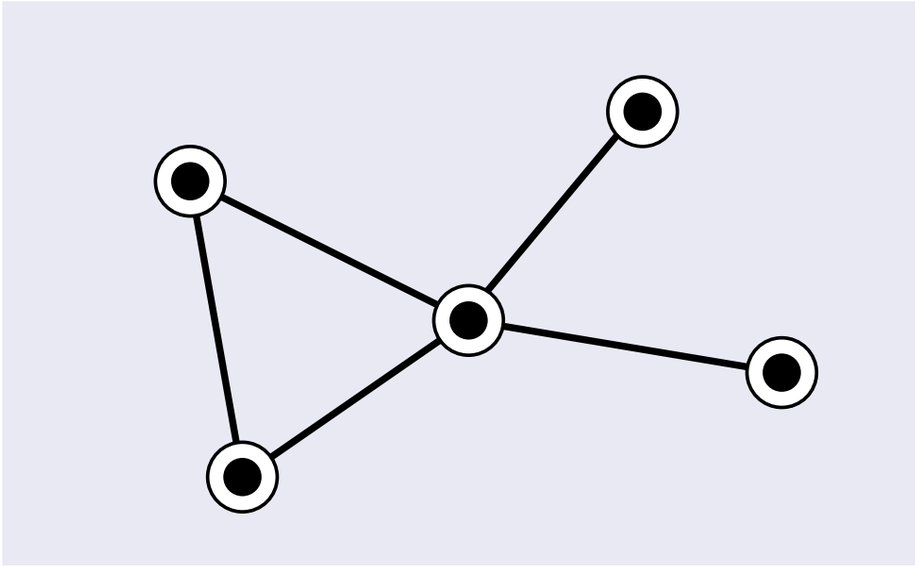
city definition	number	network type	ΔT (days)	Y	β	95% CI
Portugal						
statistical city	140	reciprocal	409	degree (K_r)	1.12	[1.11, 1.14]
				call volume (V_r)	1.11	[1.09, 1.12]
				number of calls (W_r)	1.10	[1.09, 1.11]
			92	degree (K_r)	1.10	[1.09, 1.11]
				call volume (V_r)	1.10	[1.08, 1.11]
				number of calls (W_r)	1.08	[1.07, 1.10]
		non-reciprocal	409	degree (K_r)	1.24	[1.22, 1.25]
				call volume (V_r)	1.14	[1.12, 1.15]
				number of calls (W_r)	1.13	[1.12, 1.14]
larger urban zone	9(8)	reciprocal	409	degree (K_r)	1.05	[1.00, 1.11]
				call volume (V_r)	1.11	[1.02, 1.20]
				number of calls (W_r)	1.10	[1.05, 1.15]
		non-reciprocal	409	degree (K_r)	1.13	[1.08, 1.18]
				call volume (V_r)	1.14	[1.05, 1.23]
				number of calls (W_r)	1.13	[1.08, 1.18]
municipality	293	reciprocal	409	degree (K_r)	1.13	[1.11, 1.14]
				call volume (V_r)	1.15	[1.13, 1.17]
				number of calls (W_r)	1.13	[1.11, 1.14]
UK						
urban audit city	24	reciprocal	31	degree (K)	1.08	[1.05, 1.12]
				degree, land-mobile (K_{lm})	1.14	[1.11, 1.17]
				call volume (V)	1.10	[1.07, 1.14]
				number of calls (W)	1.08	[1.05, 1.11]

Nodal clustering

Clustering coefficient:
Probability that one‘s contacts are also connected with each other.

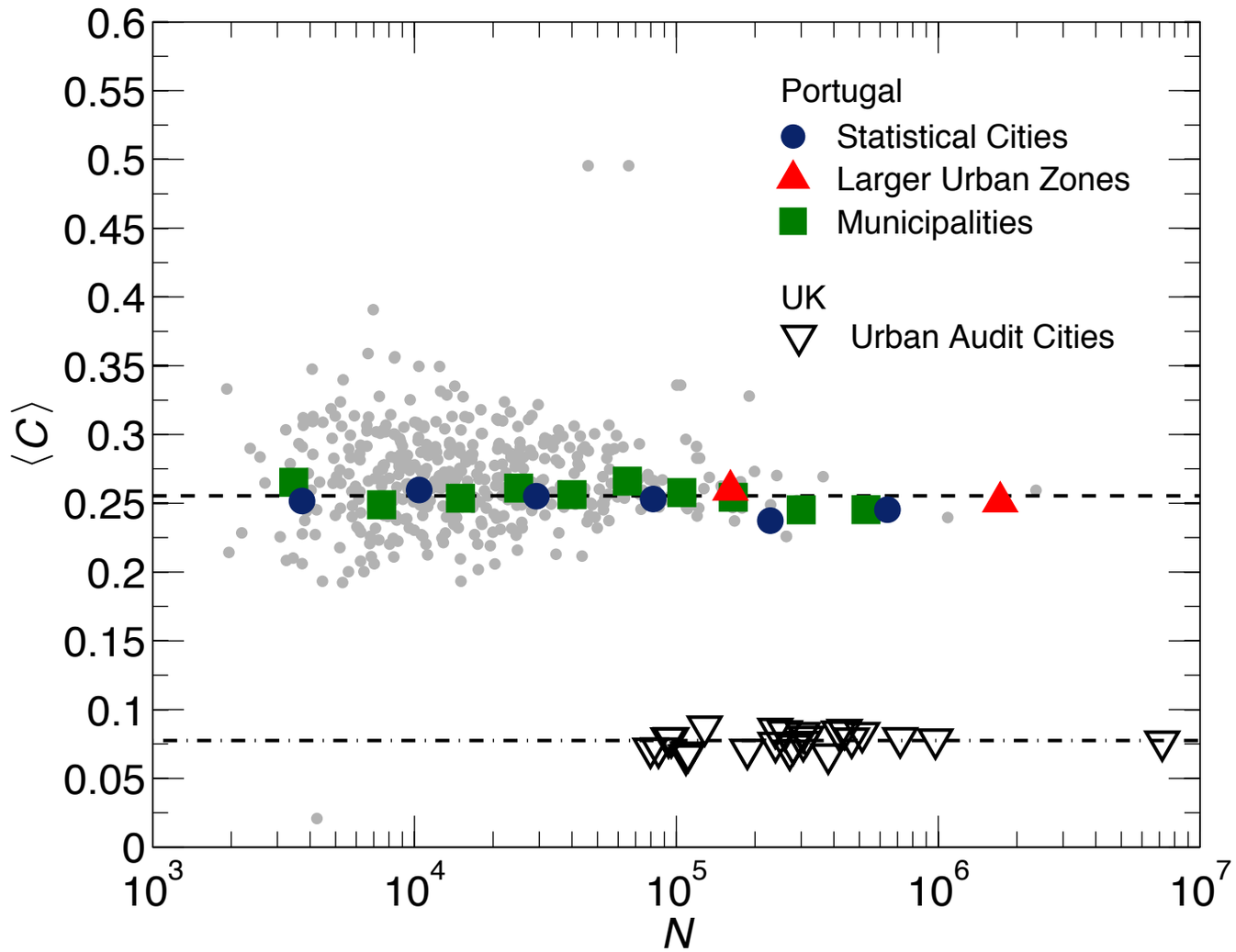
$$C_i \equiv 2z_i / \left[k_i(k_i - 1)\right]$$

z_i Number of links between the k_i neighbours
k_i Degree of node i



As larger cities provide a larger pool of people, the clustering coefficient should decrease if contacts were established at random.

Nodal clustering



- Average clustering is an invariant of city size.
- Even in large cities we live in groups that are as tightly knit as those in small towns or ‘villages’.

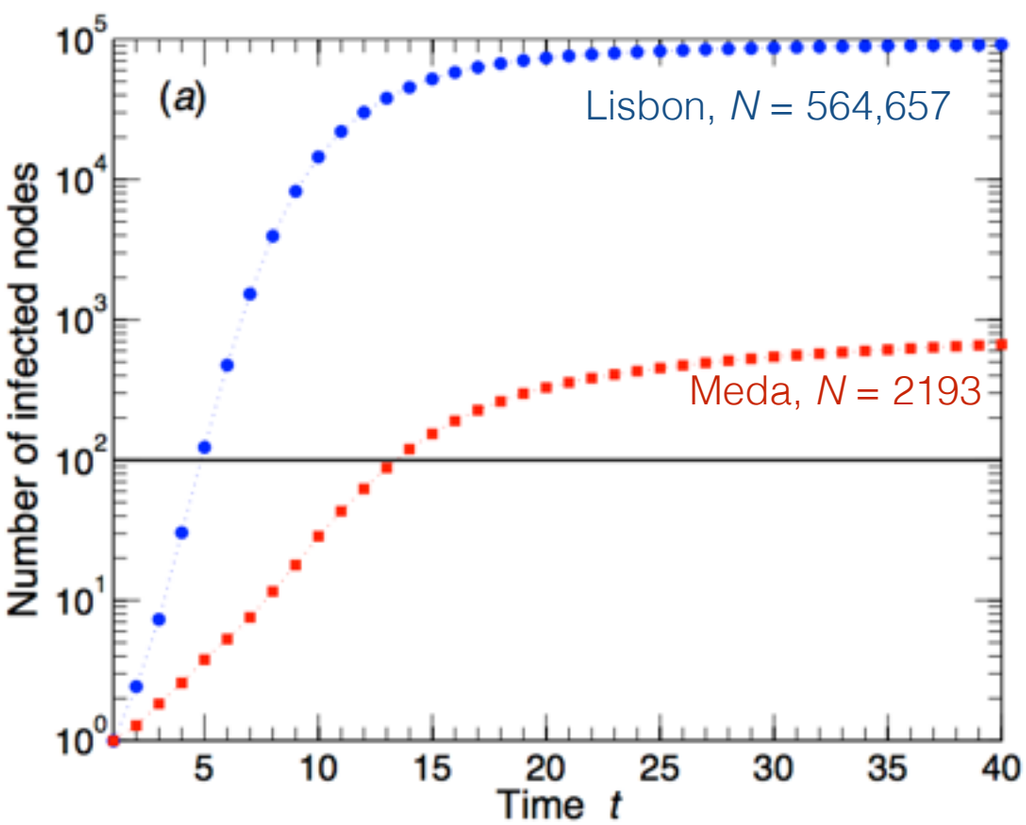
Schläpfer M. et al., J Royal Soc Interface, 11(98):20130789, 2014

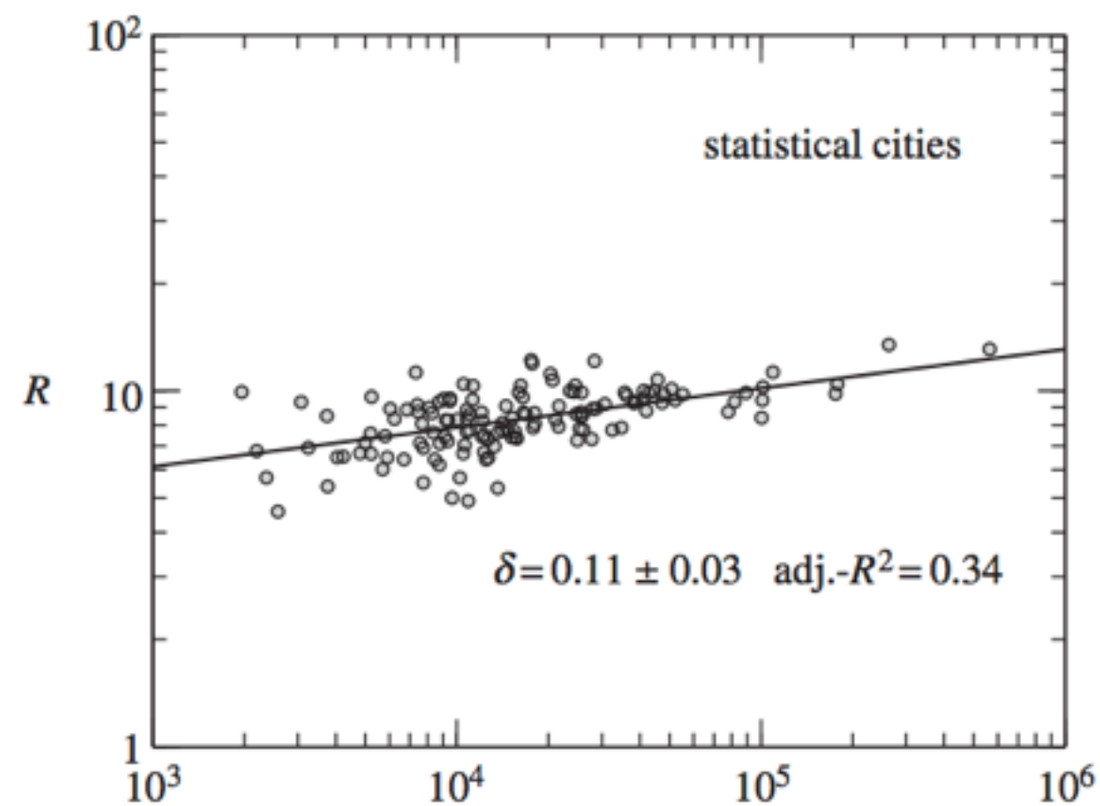
Acceleration of spreading processes

Susceptible-infected (SI) model

$$P_{ij} \propto v_{ij}$$

P_{ij} Transmission probability
v_{ij} Call volume between user i and user j





Potential ‚hidden‘ biases

Test on different data sets

- UK, mobile phones and landlines (Schläpfer et al. 2014)
- Ivory Coast, mobile phones (Andris and Bettencourt, 2014)
- „Unnamed“ European Country, mobile phones (Llorente, 2015)
- US and Europe, Twitter data (Tizzoni, 2015)
- Switzerland, mobile phones (Büchel and von Ehrlich, 2016)

2. Urban Structure: Building Heights and Shapes

PERSPECTIVE

Building functional cities

J. Vernon Henderson,[1*] Anthony J. Venables,[1,2] Tanner Regan,[1] Ilia Samsonov[1]

The literature views many African cities as dysfunctional with a hodgepodge of land uses and poor "connectivity." One driver of inefficient land uses is construction decisions for highly durable buildings made under weak institutions. In a novel approach, we model the dynamics of urban land use with both formal and slum dwellings and ongoing urban redevelopment to higher building heights in the formal sector as a city grows. We analyze the evolution of Nairobi using a unique high–spatial resolution data set. The analysis suggests insufficient building volume through most of the city and large slum areas with low housing volumes near the center, where corrupted institutions deter conversion to formal sector usage.

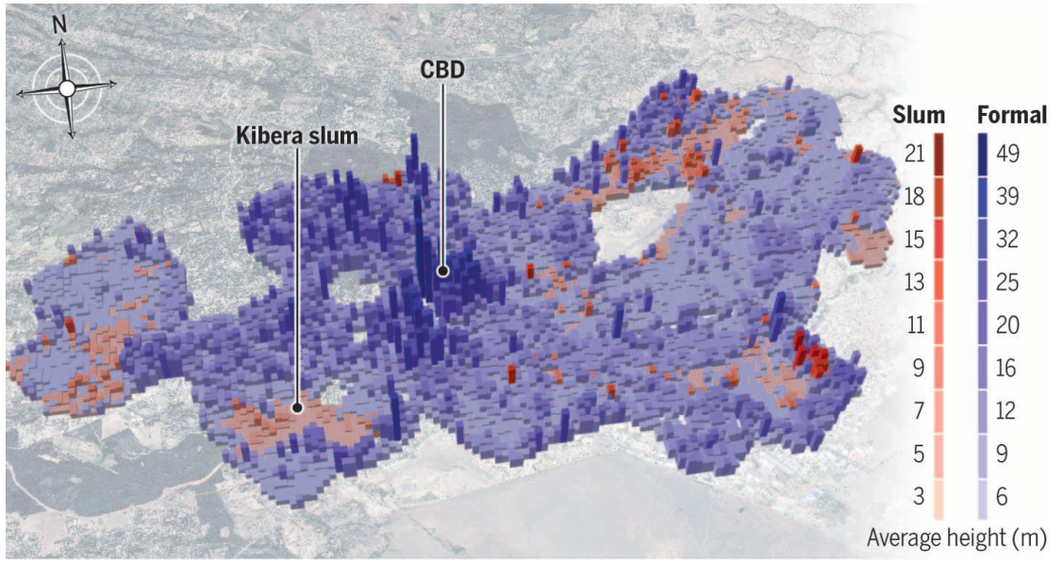
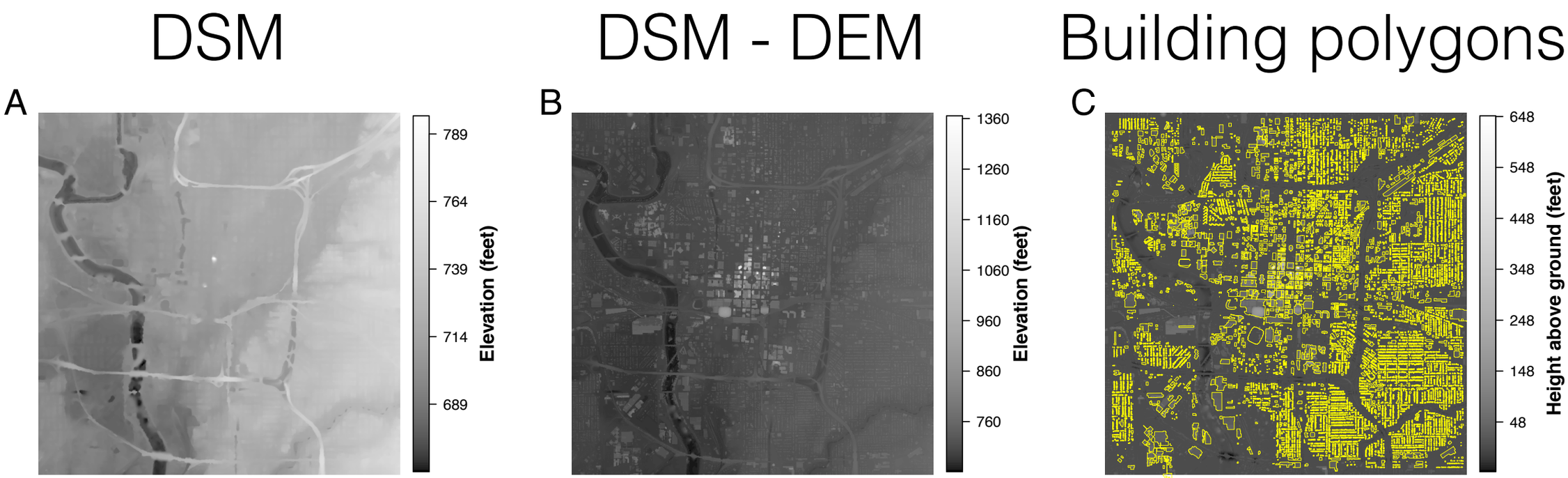


Fig. 1. City of Nairobi building height and distribution. Nairobi shows average built height in 2015 as 150-m by 150-m cells split across the formal and slum sectors. The compass (top left) points north. The location of the Kibera slum and the CBD are marked. The boundary of the city spans about 22 km east to west and 11 km north to south; the map tilt may distort the appearance of distances. Modified from HRV. [Background imagery Airbus Defense and Space 2016, taken from the SPOT5 satellite 20 September 2004].

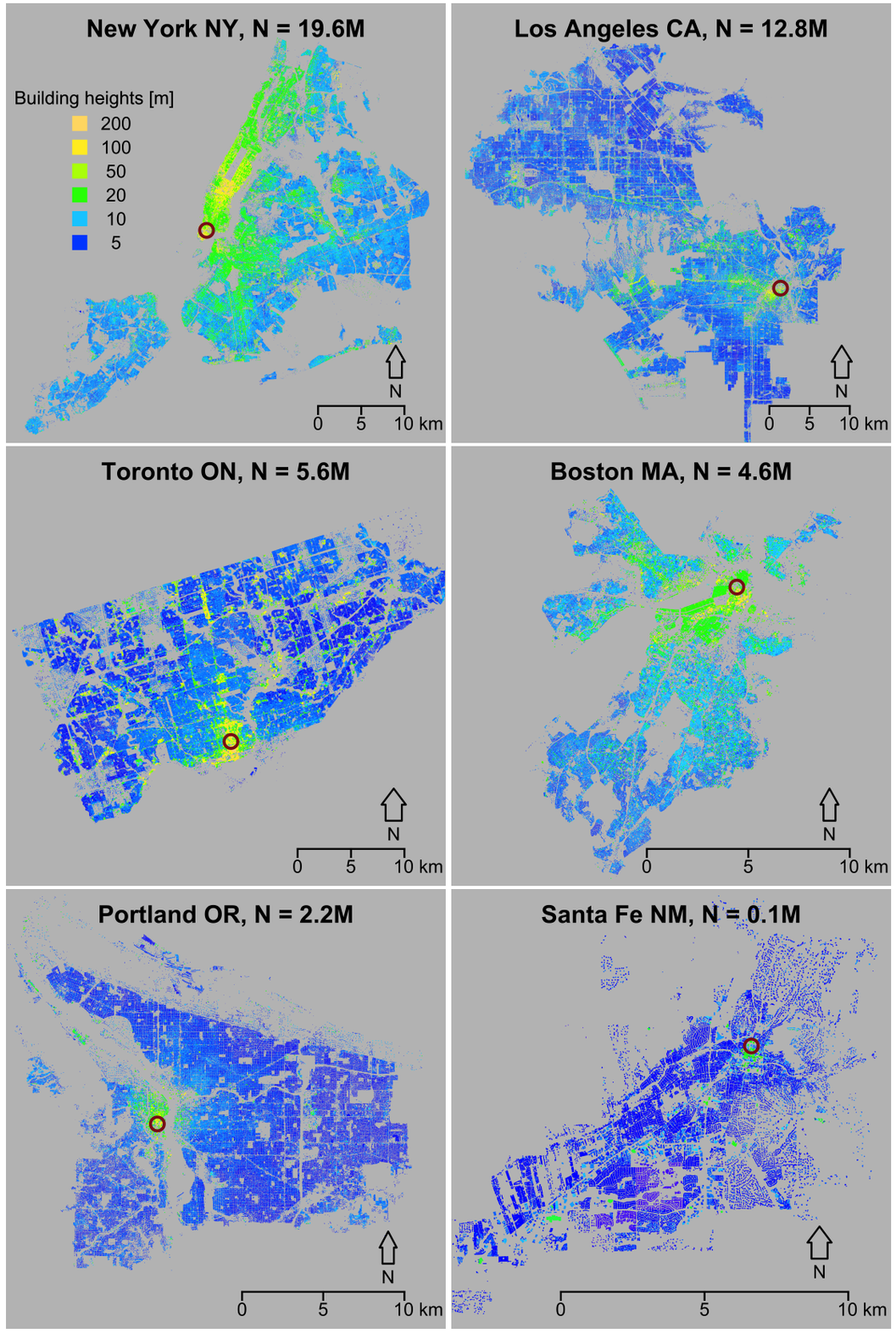
Generating simple 3D city models



DSM: Digital surface model
DEM: Digital elevation model

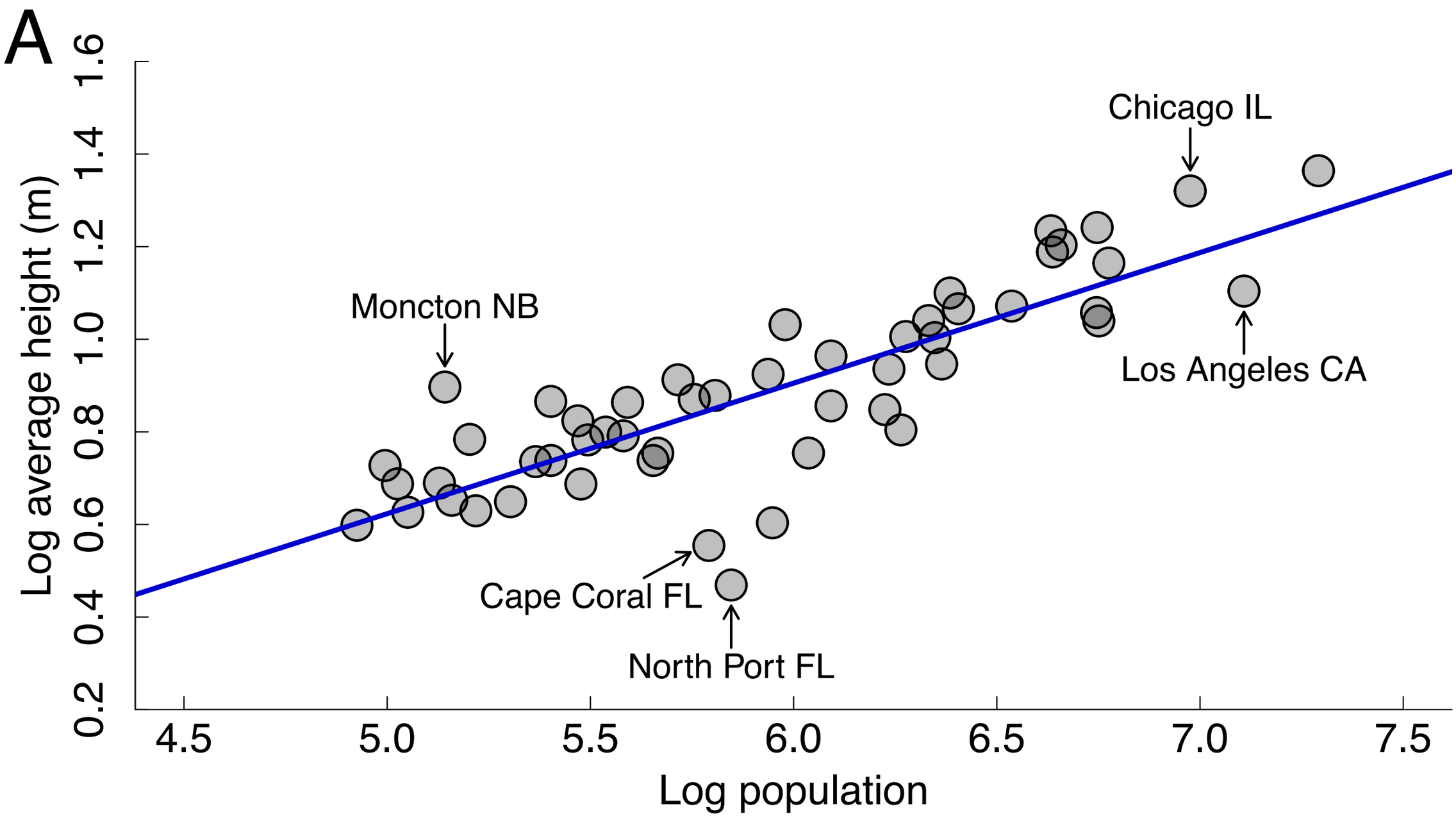
Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

Building heights



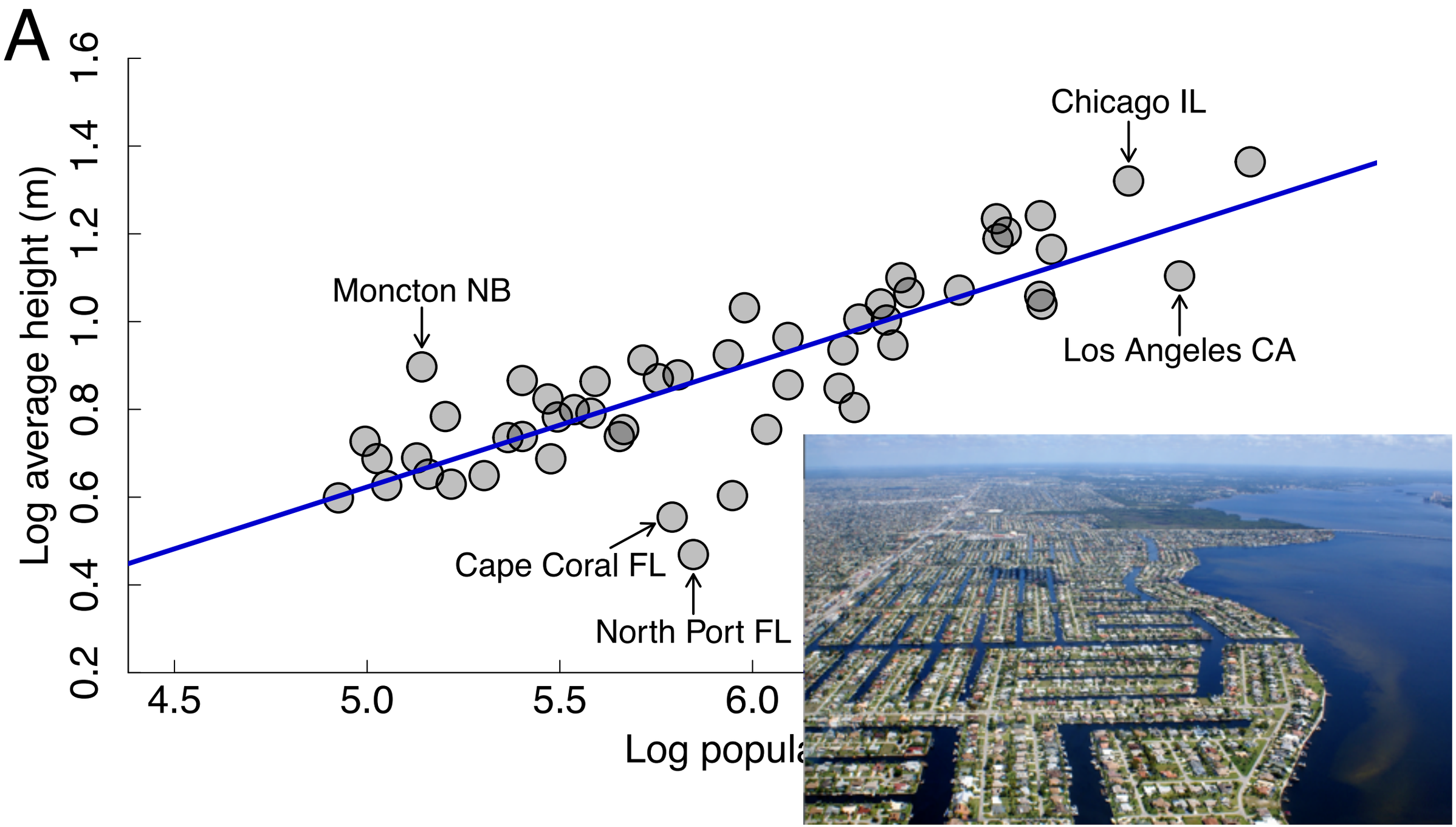
Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

Building heights



Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

Building heights



Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

Height prediction from urban scaling theory

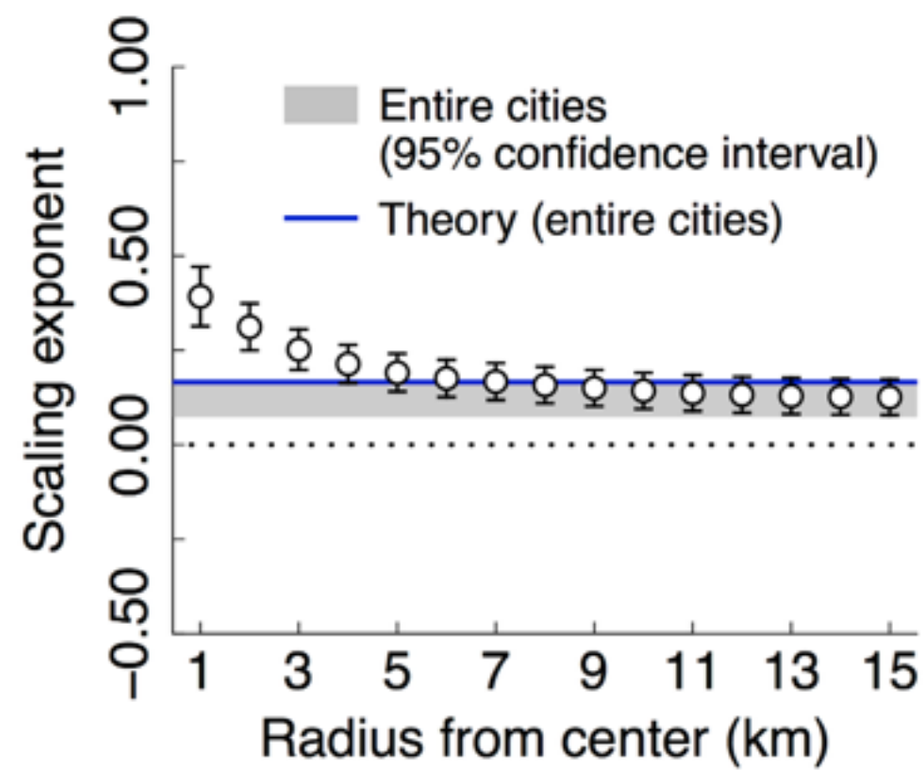
For cities to be **functional:**

$$h \propto N^{1/6}$$

Building height

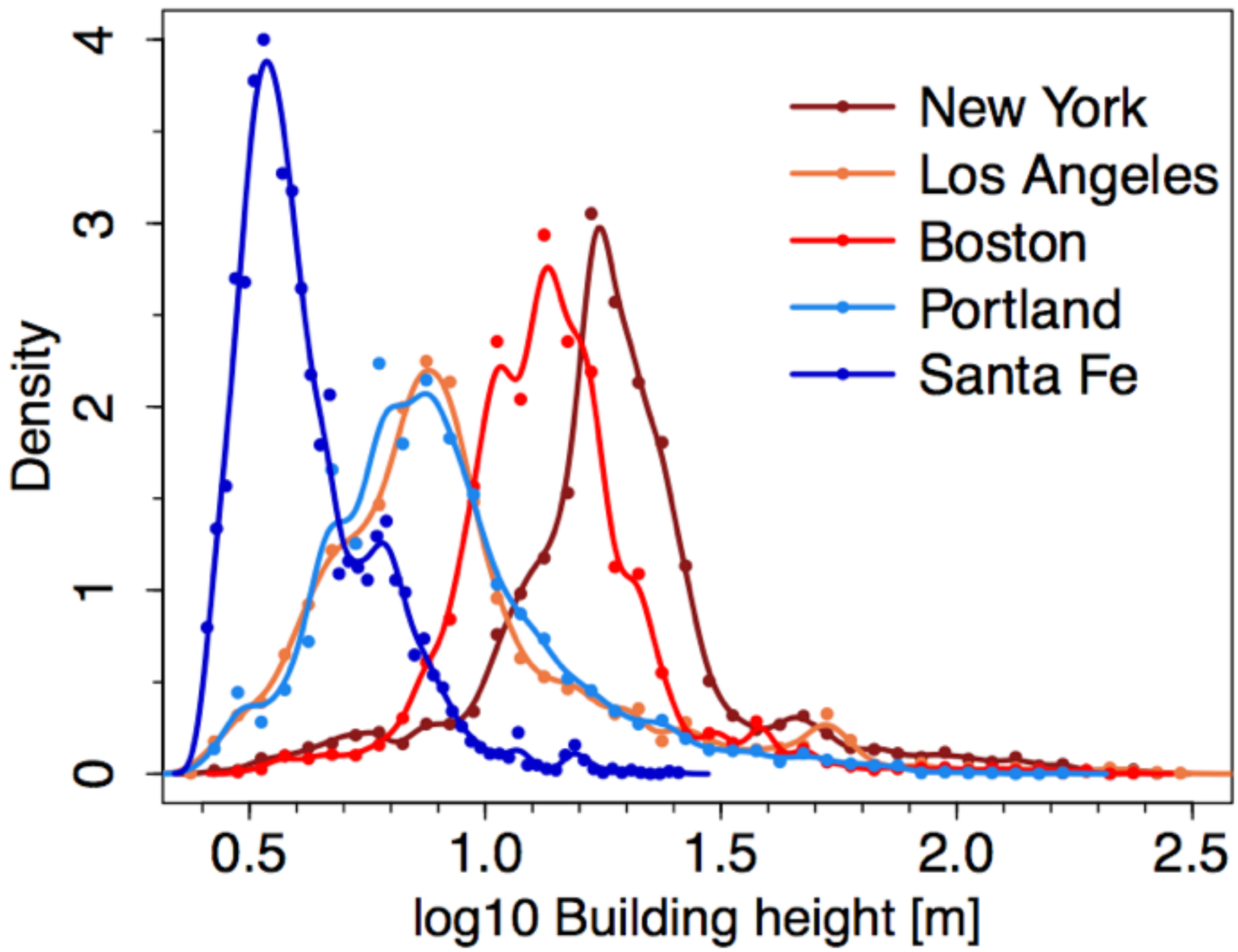
Population size

Scaling exponent



Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

Building heights distribution

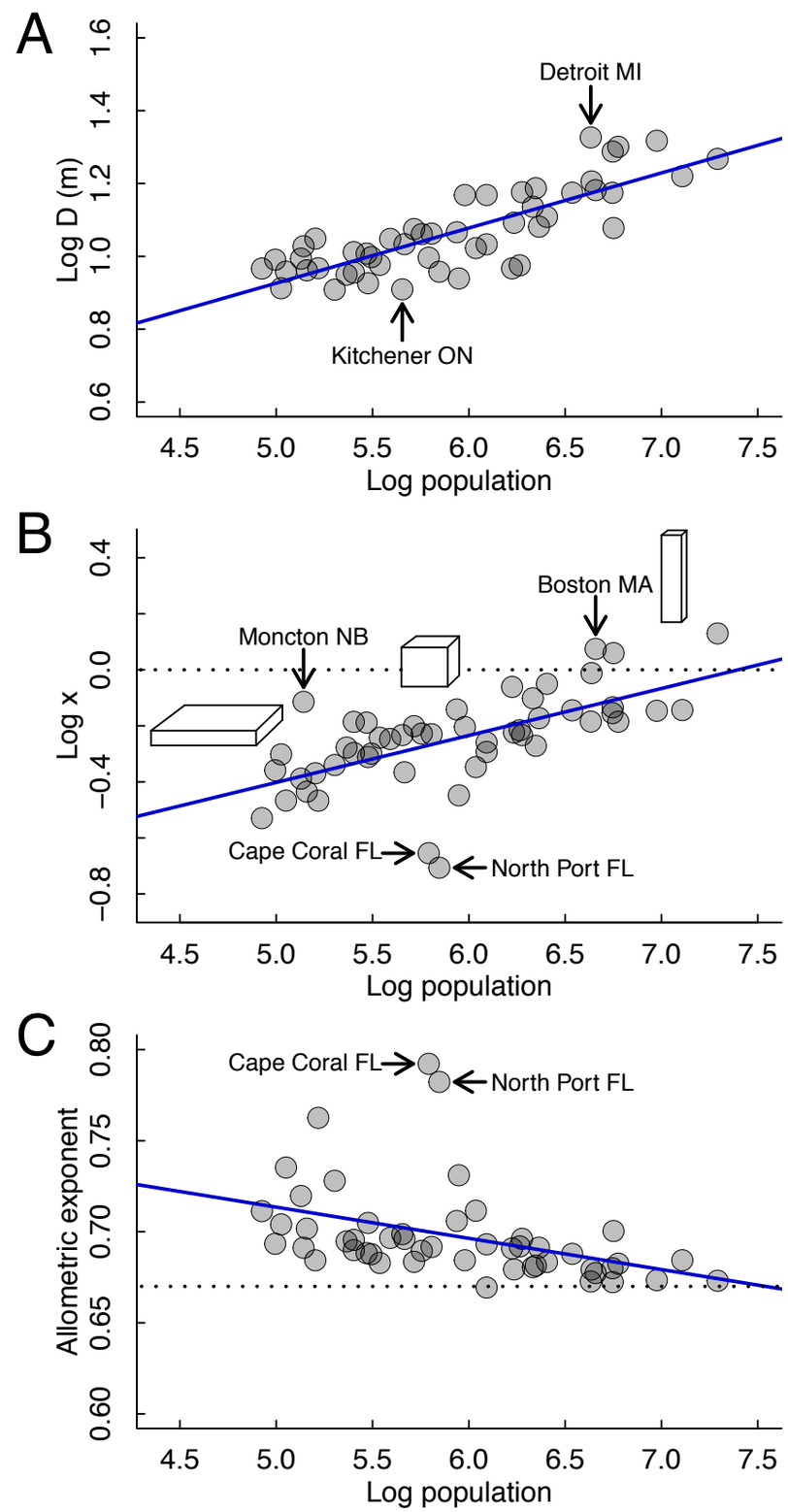


Building shapes

Building volume

Building shape

Allometric scaling



Schläpfer, Lee, Bettencourt, arXiv:1512.00946, 2015

3. Urban Dynamics: Movement of People in Cities

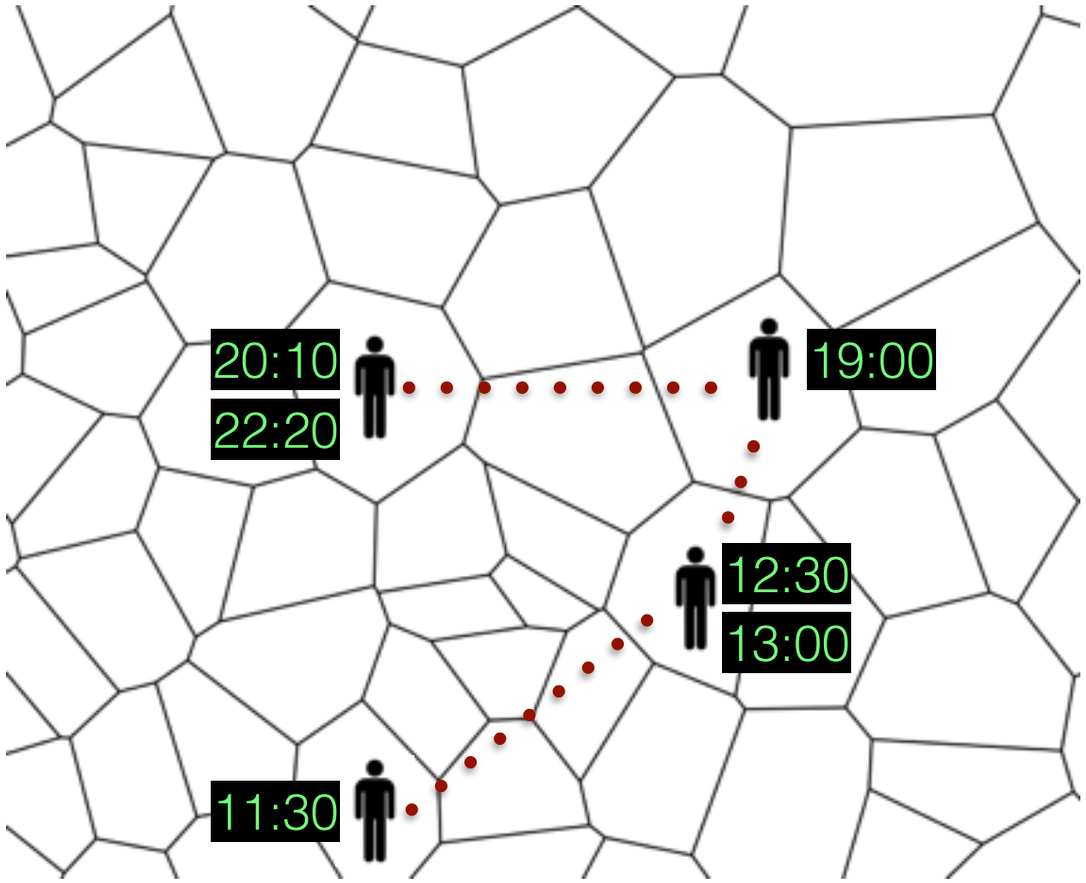
‚Collective‘ movements in cities

Source: New York Time Lapse by Dimid Vazhnik, 2015

Individual trajectories from mobile phone data

User ID, Timestamp, Cell tower ID

```
1,2013-01-24 11:30:00,599
1,2013-01-24 12:30:00,608
1,2013-01-24 13:00:00,608
1,2013-01-24 19:00:00,446
1,2013-01-24 20:10:00,323
1,2013-01-24 20:30:00,323
1,2013-01-24 22:00:00,323
1,2013-01-24 22:20:00,323
```



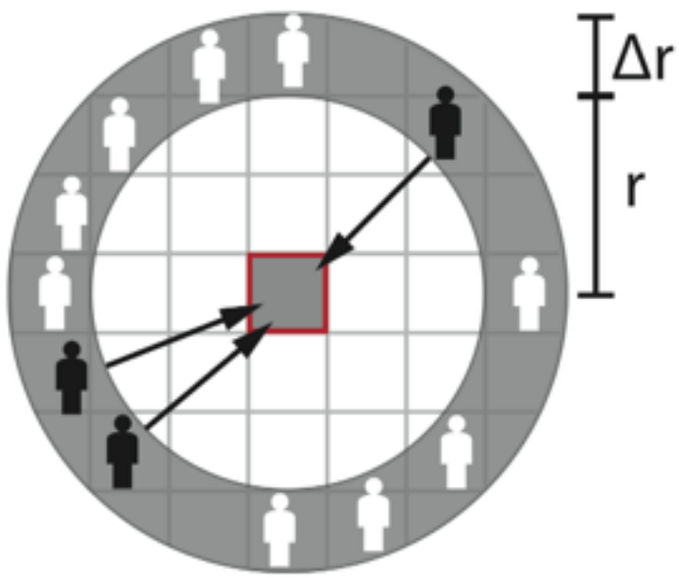
1. ***How many*** people visit a given location?
2. From ***how far*** do they come?
3. ***How often*** do they visit?

Lets look into the data!

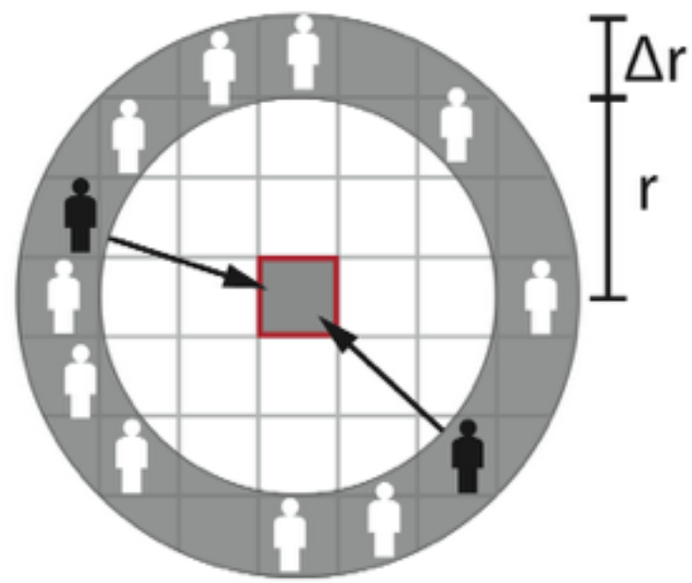
- Greater Boston area
- $\approx$2 Mio. mobile phone users over 4 months
- $\approx 10^9$ location based records per month (triangulation)
- 46,210 locations (500m x 500m grid cells)

Quantifying the attractiveness of locations

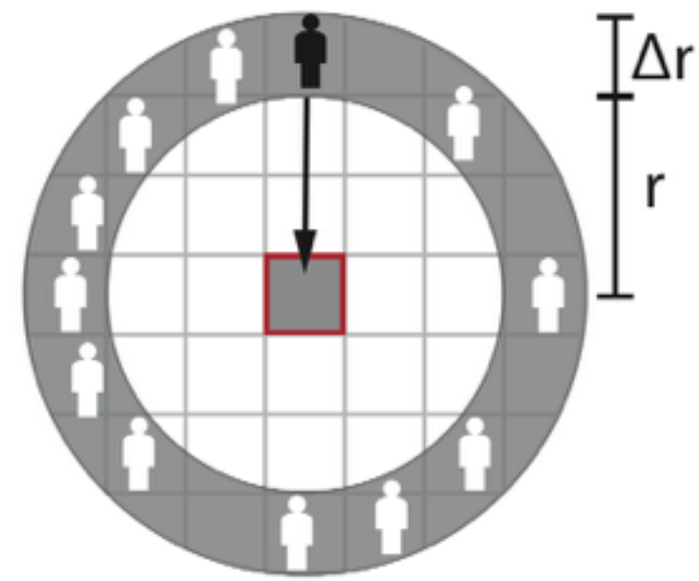
1 visit per month

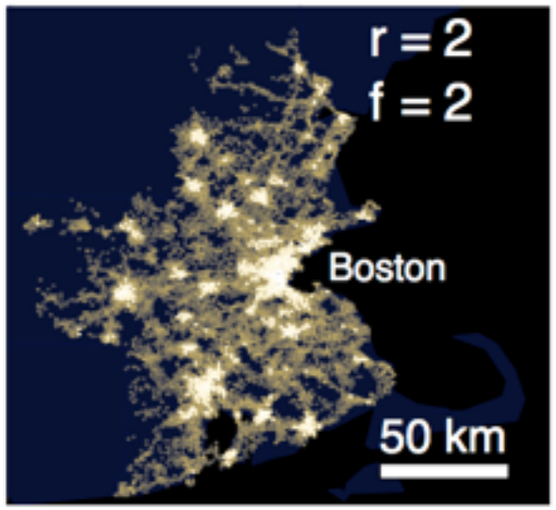


2 visits per month



3 visits per month



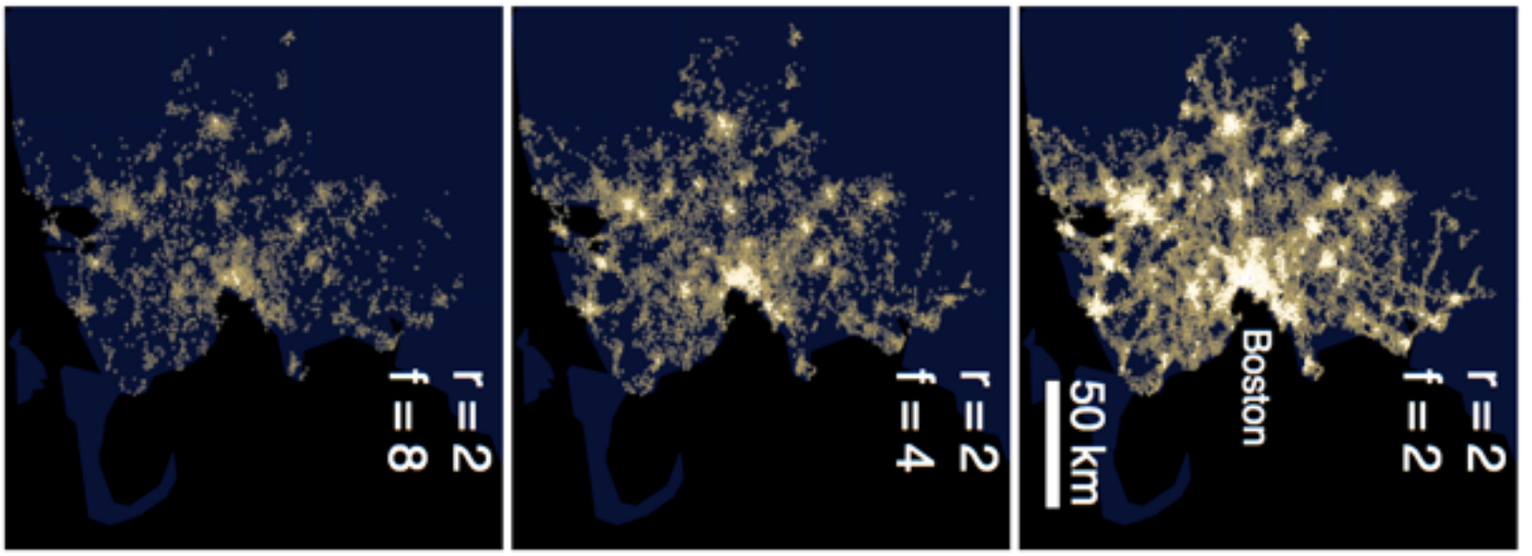


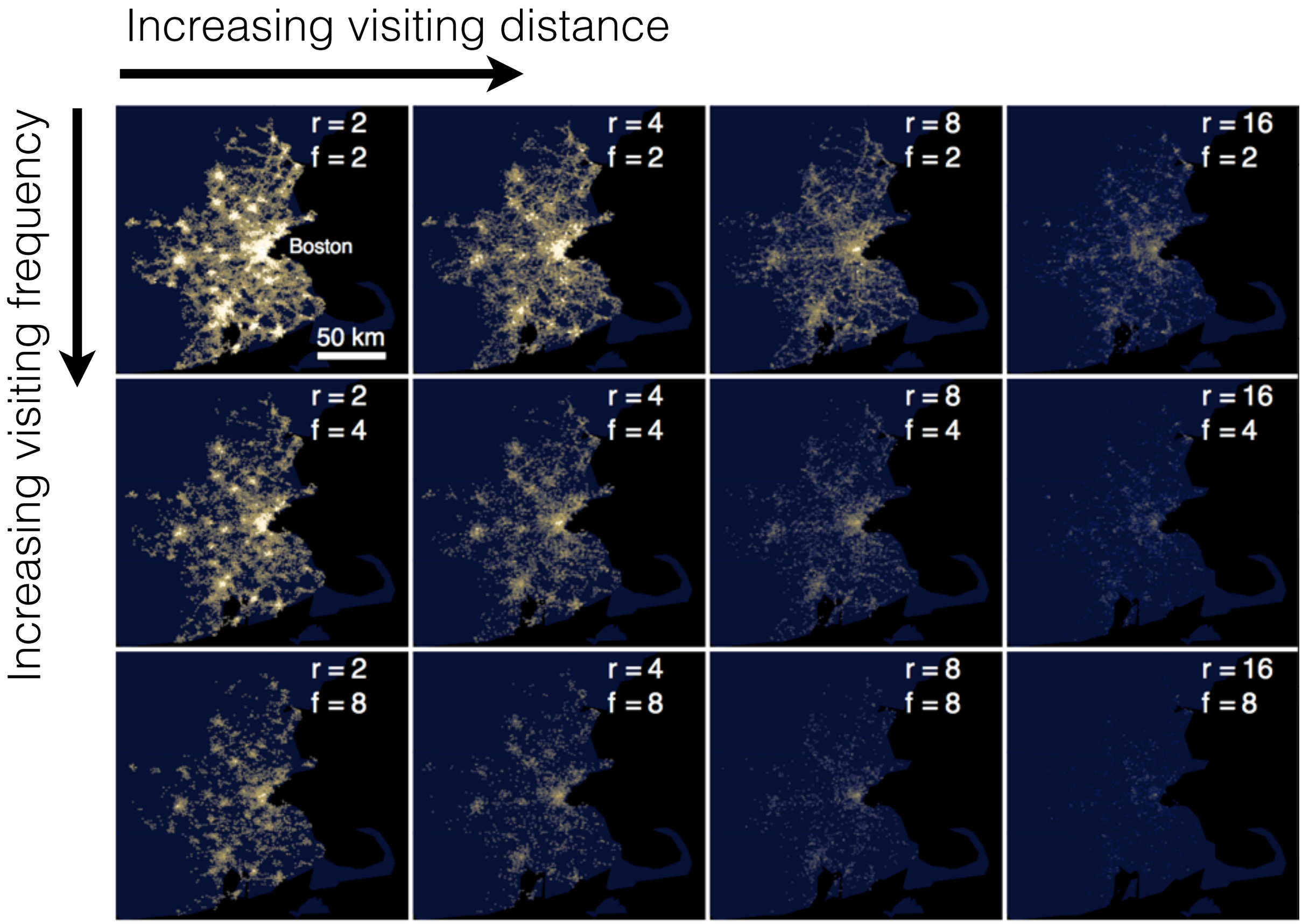
r visiting distance (km)
f visiting frequency (visits per month)

Brightness of pixel: number of visitors, ***q*(*r*,*f*)**

Increasing visiting distance

Increasing visiting frequency




Increasing visiting distance
Increasing visiting frequency
r = 2
f = 2
Boston
50 km
r = 4
f = 2
r = 8
f = 2
r = 16
f = 2
r = 2
f = 4
r = 4
f = 4
r = 8
f = 4
r = 16
f = 4
r = 2
f = 8
r = 4
f = 8
r = 8
f = 8
r = 16
f = 8

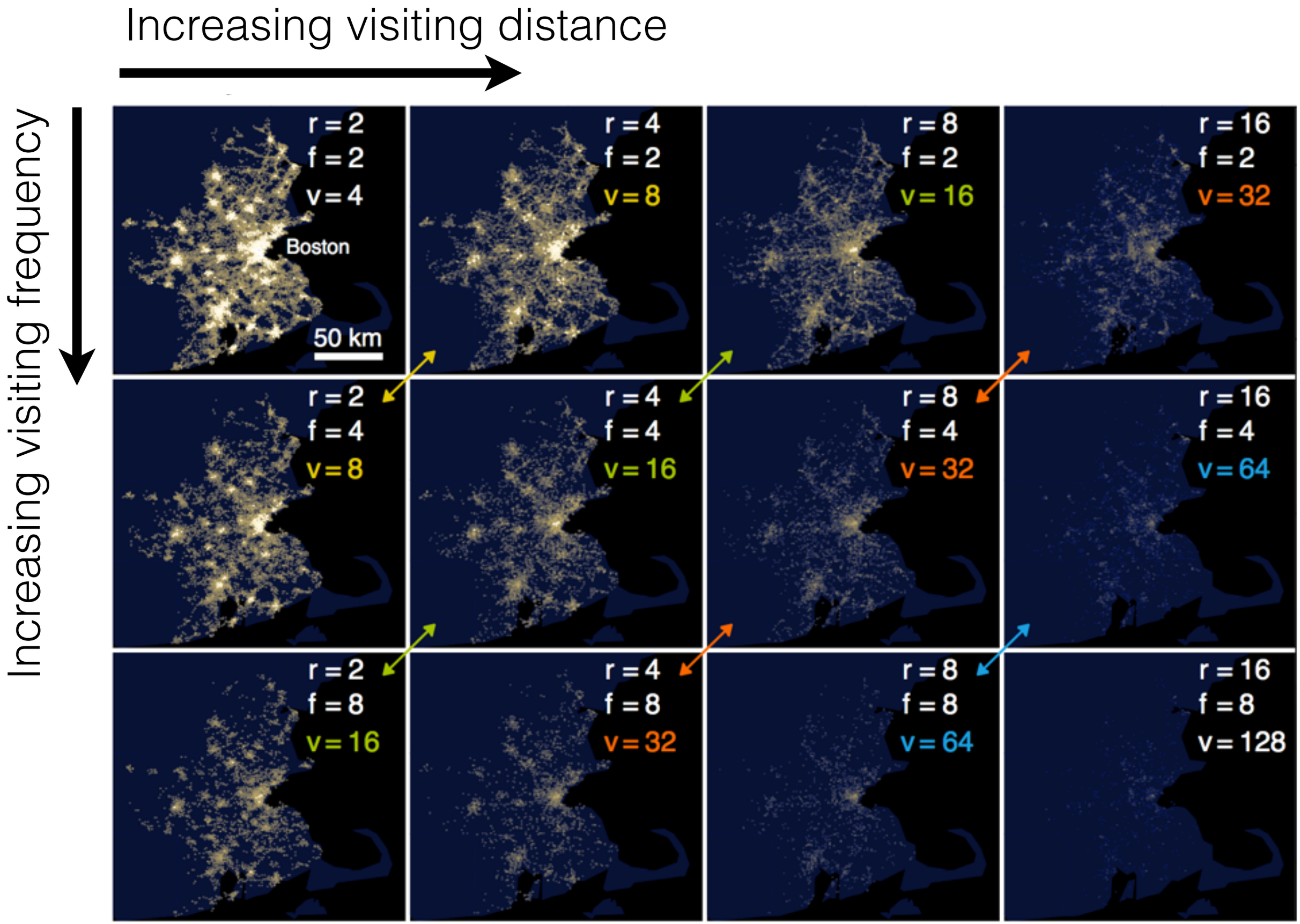
Dimensional analysis

$$q = q\left[r, f, u, Y(N)\right]$$

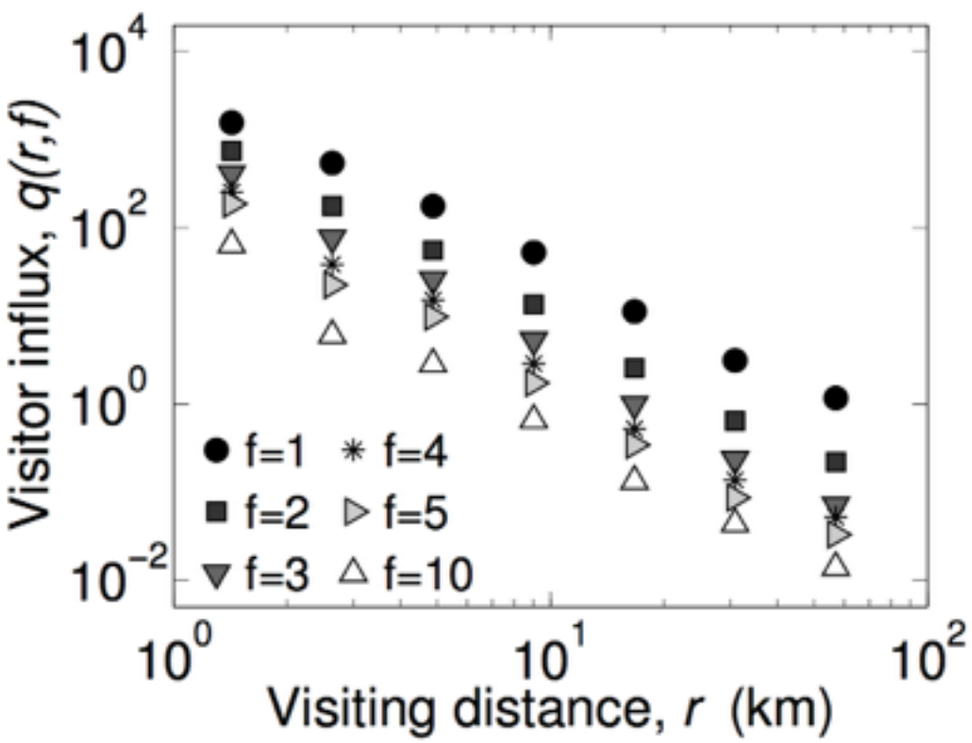
Travel speed

Socioeconomic features

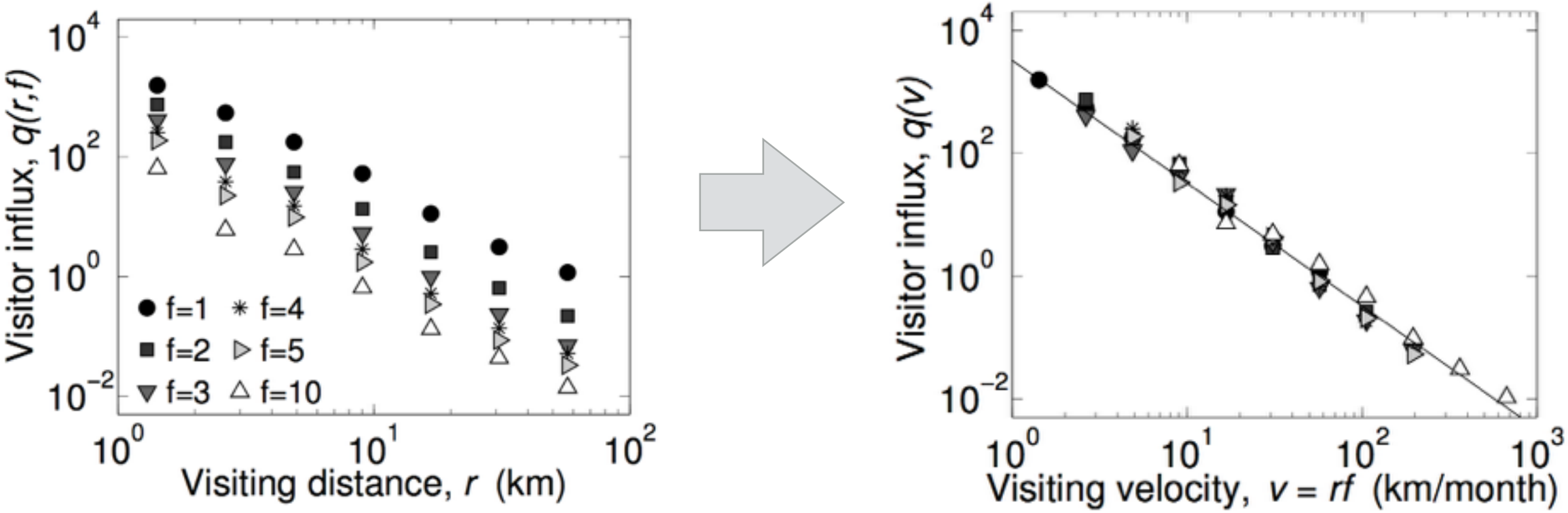
$$\Rightarrow q(r, f) = G\left(\frac{rf}{u}\right) = F(rf)$$


Increasing visiting distance
Increasing visiting frequency
r = 2
f = 2
v = 4
Boston
50 km
r = 4
f = 2
v = 8
r = 8
f = 2
v = 16
r = 16
f = 2
v = 32
r = 2
f = 4
v = 8
r = 4
f = 4
v = 16
r = 8
f = 4
v = 32
r = 16
f = 4
v = 64
r = 2
f = 8
v = 16
r = 4
f = 8
v = 32
r = 8
f = 8
v = 64
r = 16
f = 8
v = 128

Newbury Street, Boston



Newbury Street, Boston



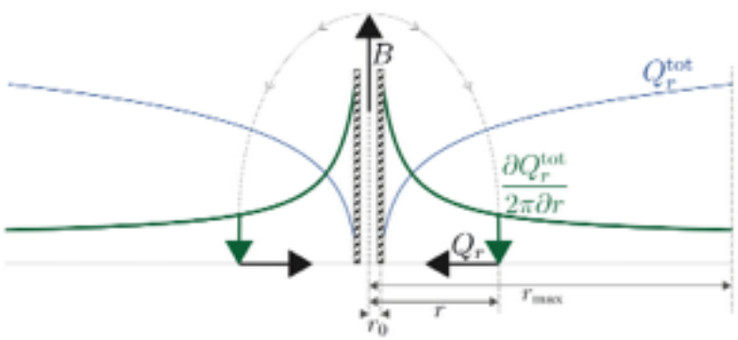
Example (v = 20km/month):

number of visitors coming from 5 km and 4 times a month
=
number of visitors coming from 10 km and 2 times a month
=
number of visitors coming from 20 km and once a month

What is the **functional relation** between:

- number of visitors,
- their travel distance from home,
- their visiting frequency?

Fluid dynamic model



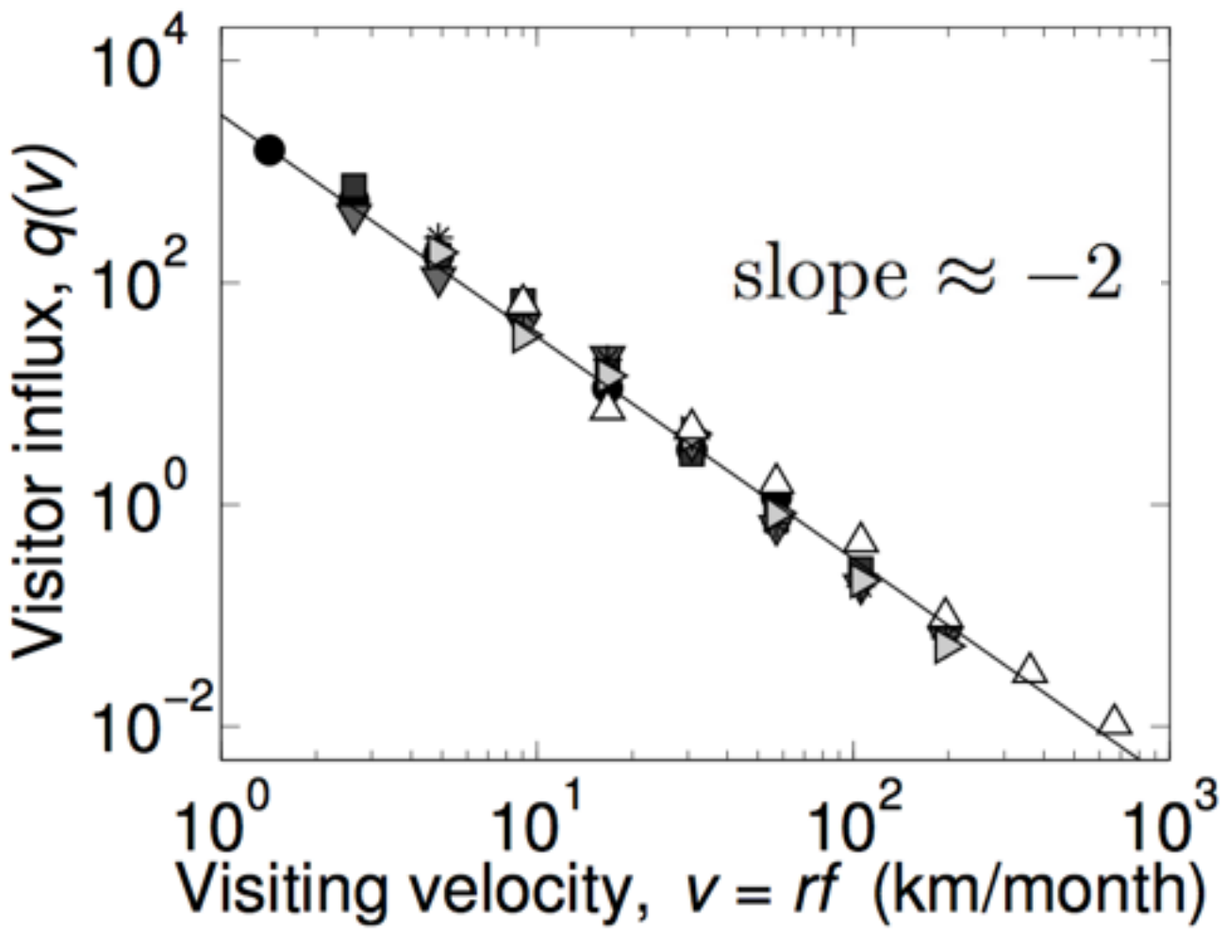
Dimensional argument

$$\Rightarrow q(r, f) = G\left(\frac{rf}{u}\right) = F(rf)$$

Theoretical expectation

Number of visitors
$\propto$
[travel distance x visiting frequency]$^{-2}$

Number of visitors
∝
[travel distance x visiting frequency]$^{-2}$



Schläpfer, Szell, Ratti, West (in preparation)

1. ***How many*** people visit a given location?
2. From ***how far*** do they come?
3. ***How often*** do they visit?

1. ***How many***
2. From
3. ***How often***

1. ***How many*** people visit a given location?

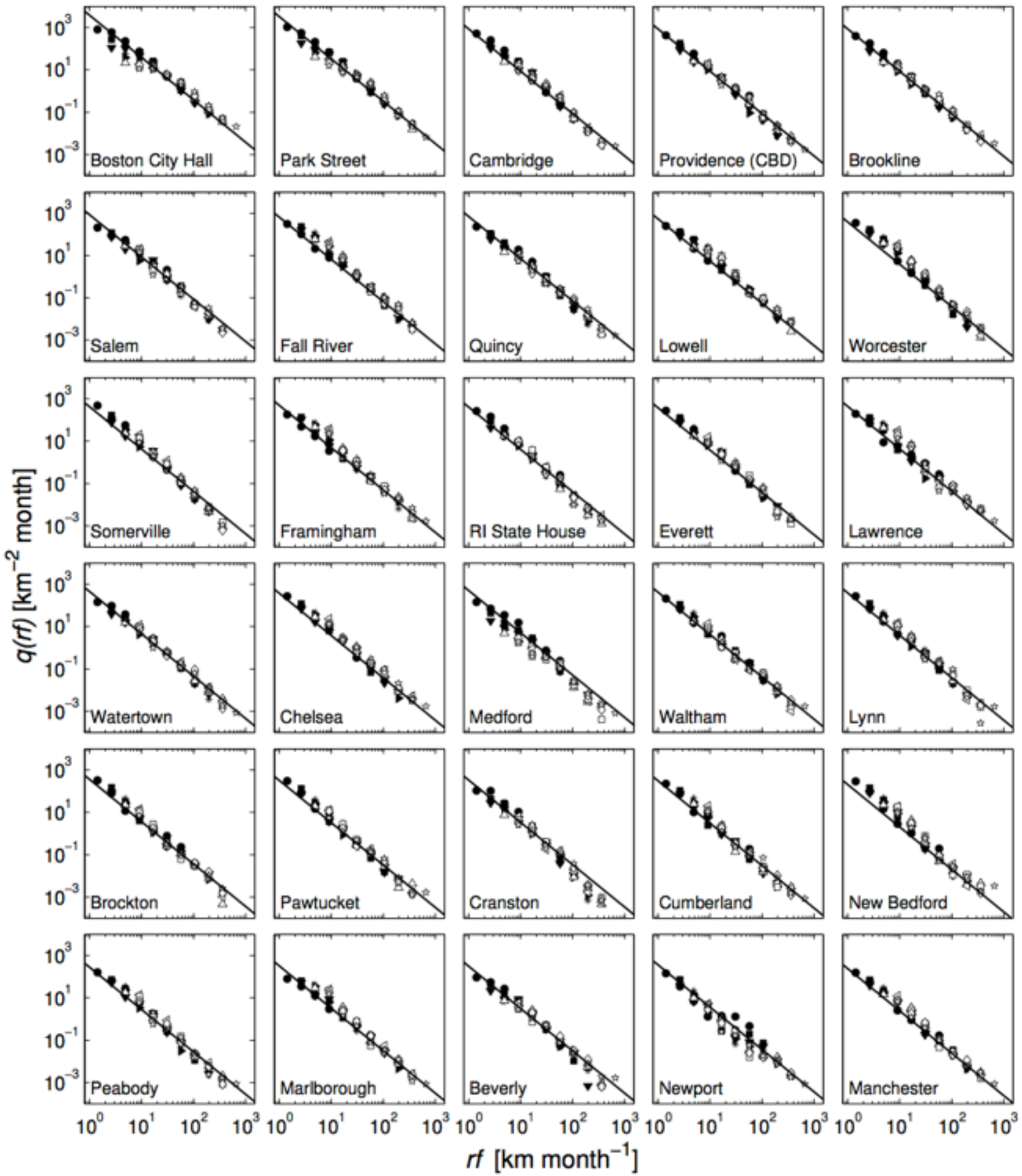
predicts

predicts

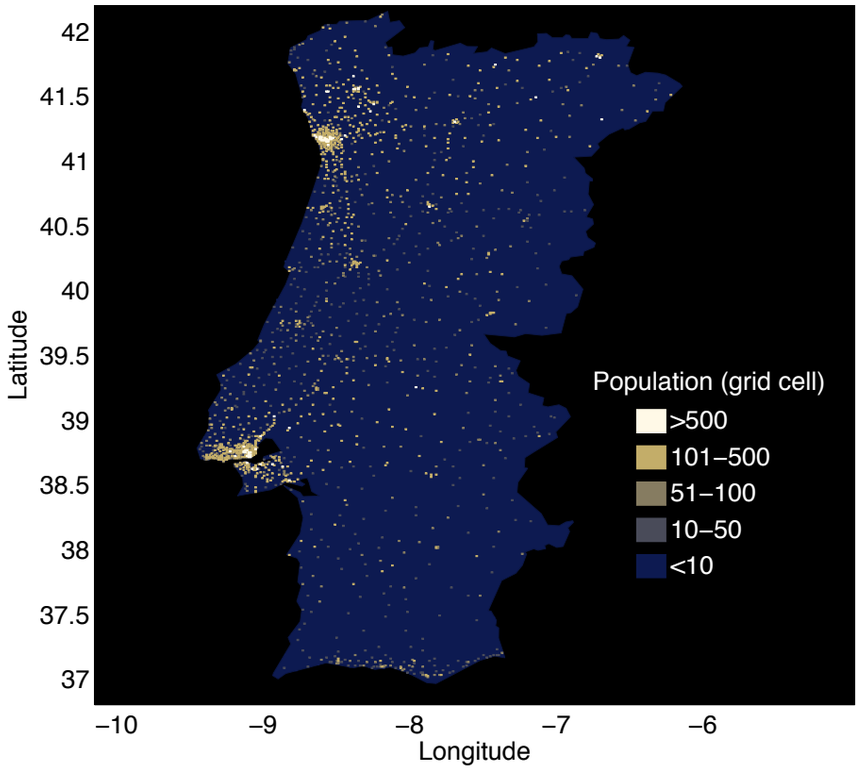
From ***how far*** do they come.

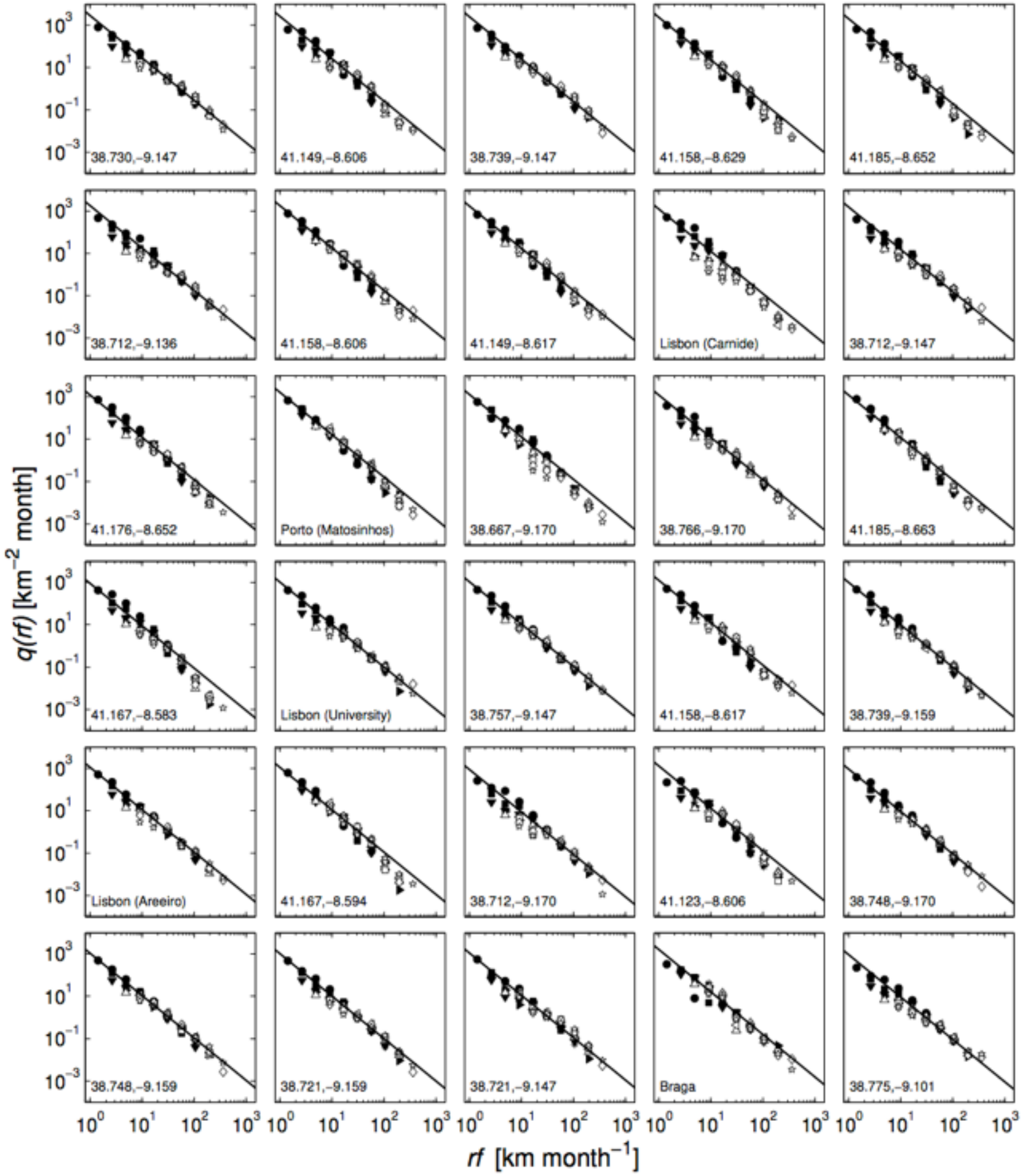
How often do they visit.

Greater Boston

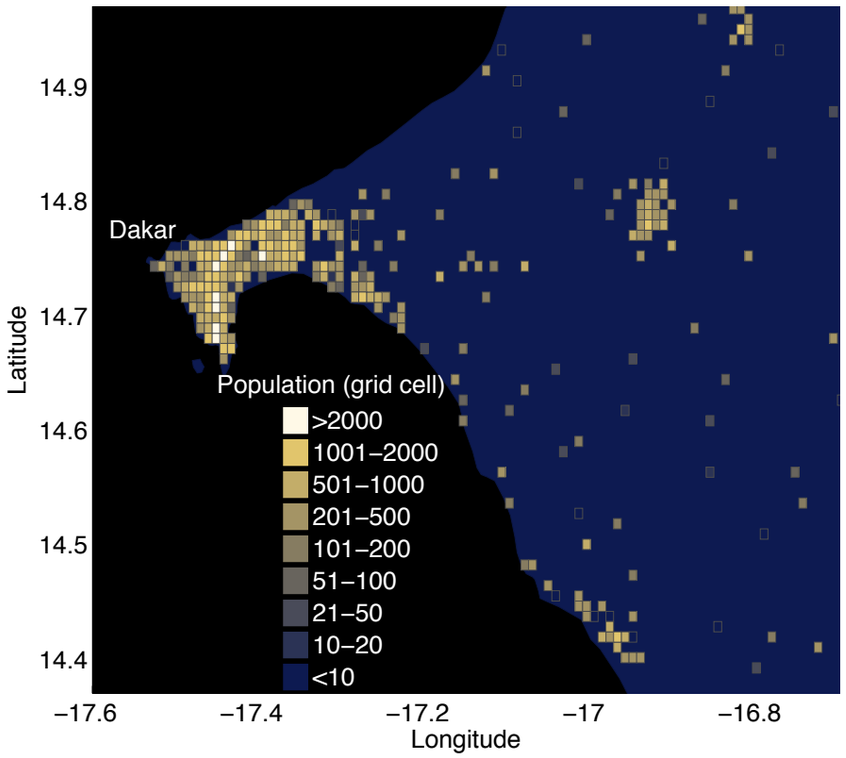


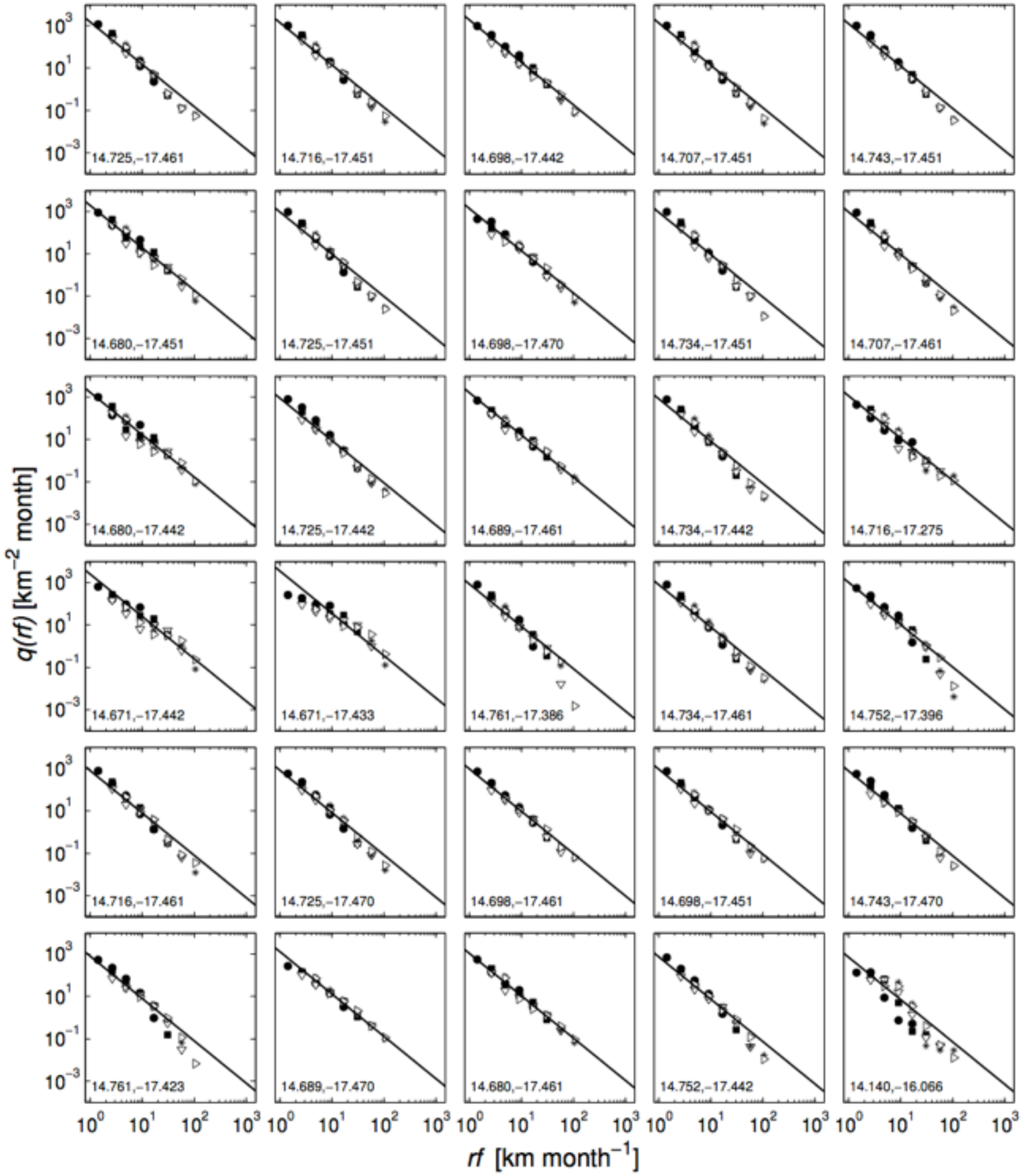
Portugal


Population (grid cell)
>500
101–500
51–100
10–50
<10
Latitude
Longitude
42
41.5
41
40.5
40
39.5
39
38.5
38
37.5
37
−10
−9
−8
−7
−6

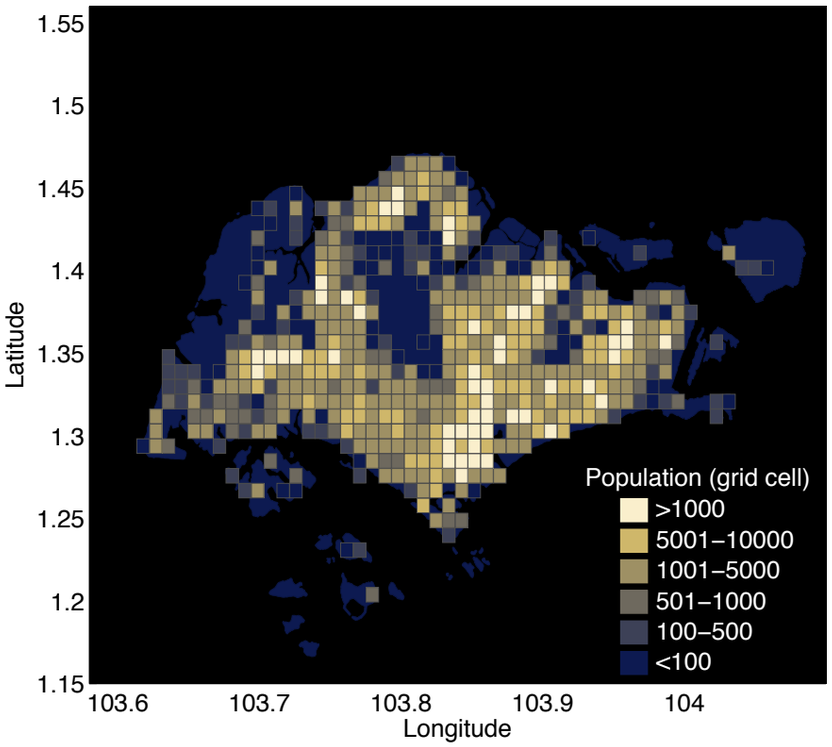

q(rf) [km−2 month]
rf [km month−1]
38.730,−9.147
41.149,−8.606
38.739,−9.147
41.158,−8.629
41.185,−8.652
38.712,−9.136
41.158,−8.606
41.149,−8.617
Lisbon (Carnide)
38.712,−9.147
41.176,−8.652
Porto (Matosinhos)
38.667,−9.170
38.766,−9.170
41.185,−8.663
41.167,−8.583
Lisbon (University)
38.757,−9.147
41.158,−8.617
38.739,−9.159
Lisbon (Areeiro)
41.167,−8.594
38.712,−9.170
41.123,−8.606
38.748,−9.170
38.748,−9.159
38.721,−9.159
38.721,−9.147
Braga
38.775,−9.101

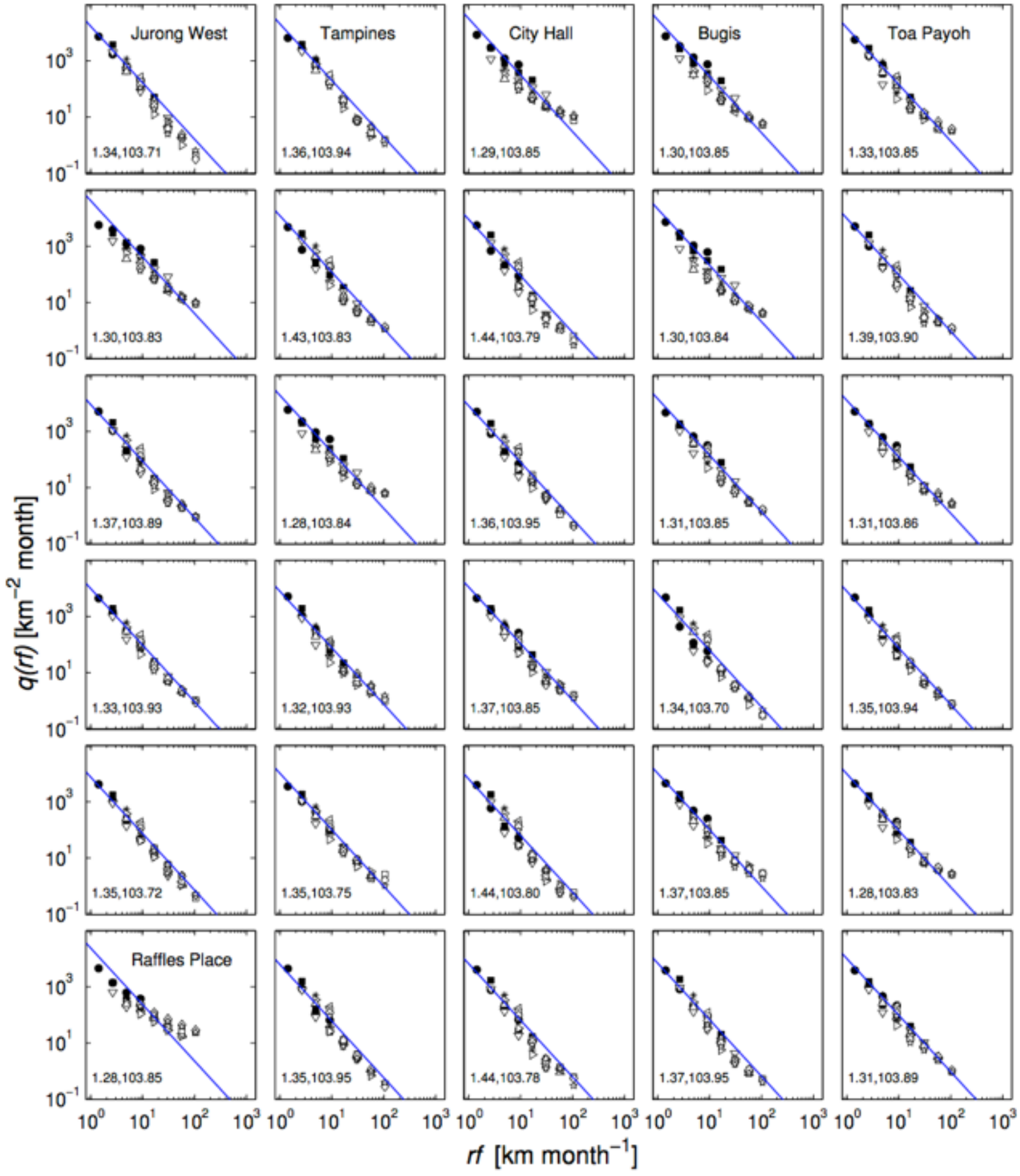
Senegal

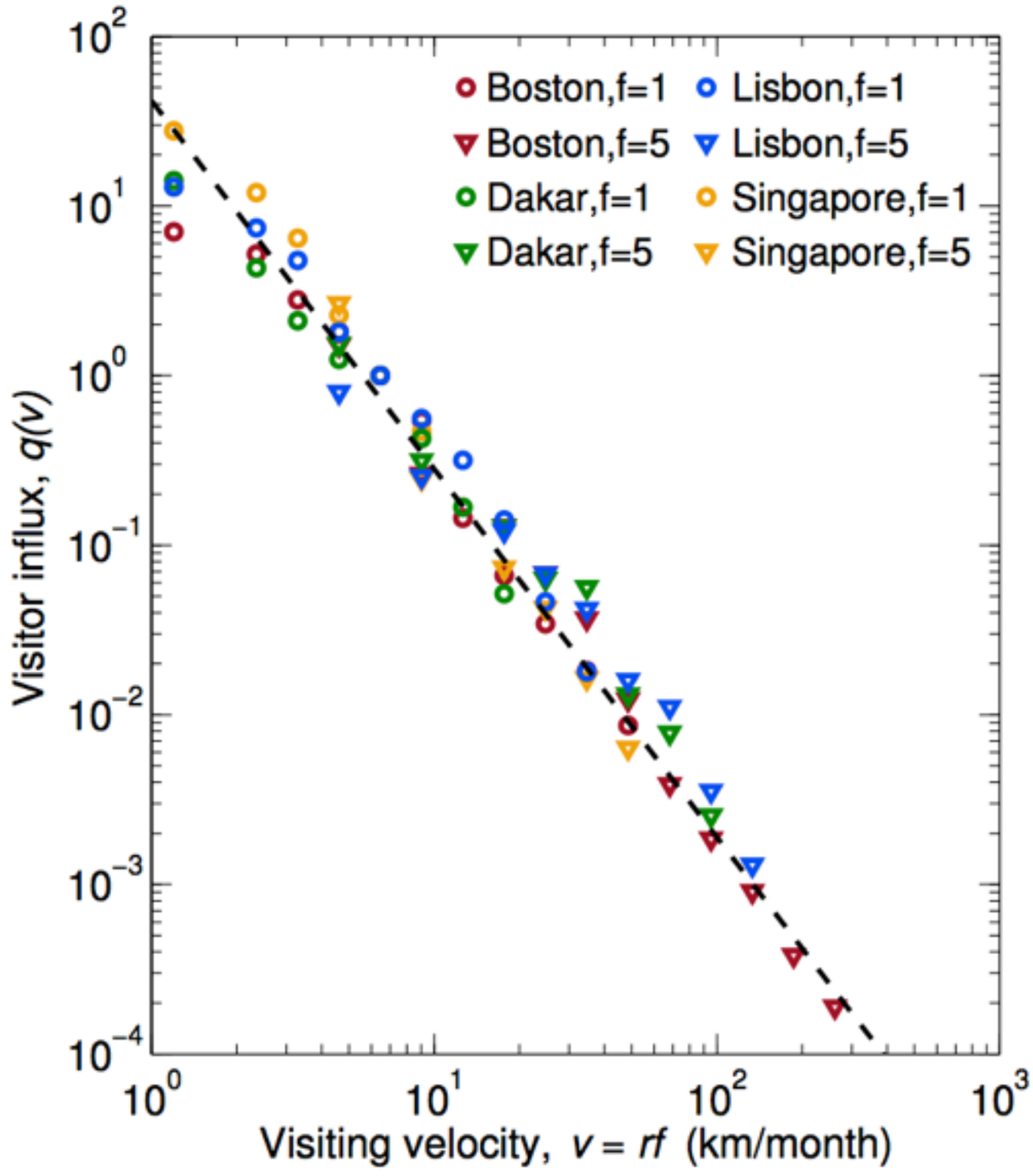

Dakar
Population (grid cell)
>2000
1001–2000
501–1000
201–500
101–200
51–100
21–50
10–20
<10
Latitude
14.9
14.8
14.7
14.6
14.5
14.4
−17.6
−17.4
−17.2
−17
−16.8
Longitude


q(rf) [km−2 month]
rf [km month−1]
14.725,−17.461
14.716,−17.451
14.698,−17.442
14.707,−17.451
14.743,−17.451
14.680,−17.451
14.725,−17.451
14.698,−17.470
14.734,−17.451
14.707,−17.461
14.680,−17.442
14.725,−17.442
14.689,−17.461
14.734,−17.442
14.716,−17.275
14.671,−17.442
14.671,−17.433
14.761,−17.386
14.734,−17.461
14.752,−17.396
14.716,−17.461
14.725,−17.470
14.698,−17.461
14.698,−17.451
14.743,−17.470
14.761,−17.423
14.689,−17.470
14.680,−17.461
14.752,−17.442
14.140,−16.066

Singapore

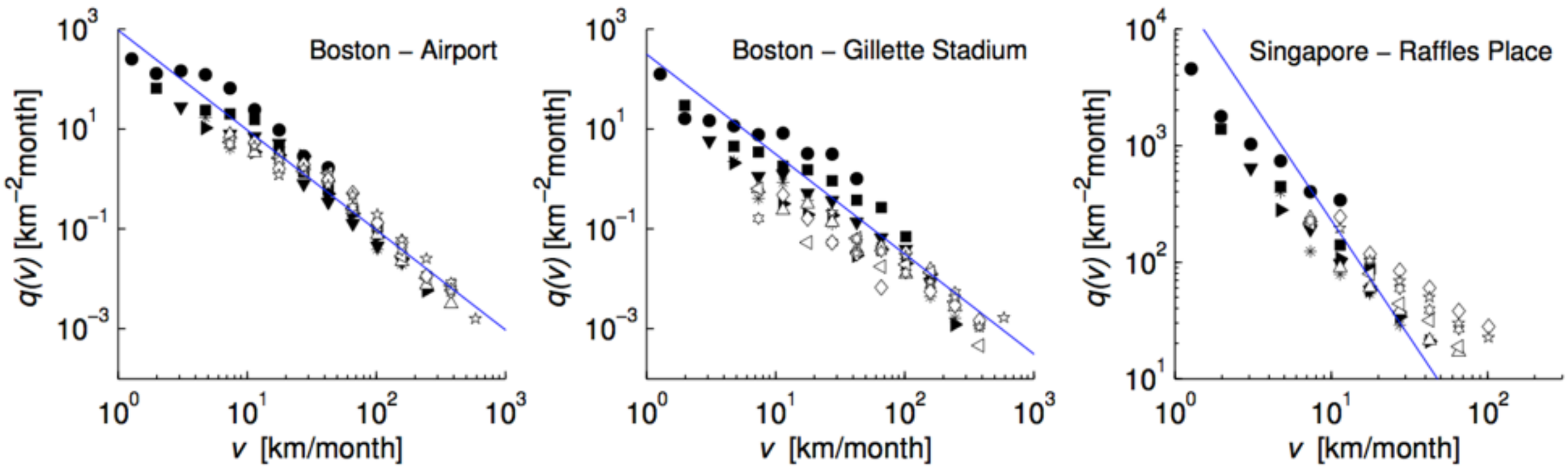

Latitude
1.55
1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
103.6
103.7
103.8
103.9
104
Longitude
Population (grid cell)
>1000
5001–10000
1001–5000
501–1000
100–500
<100

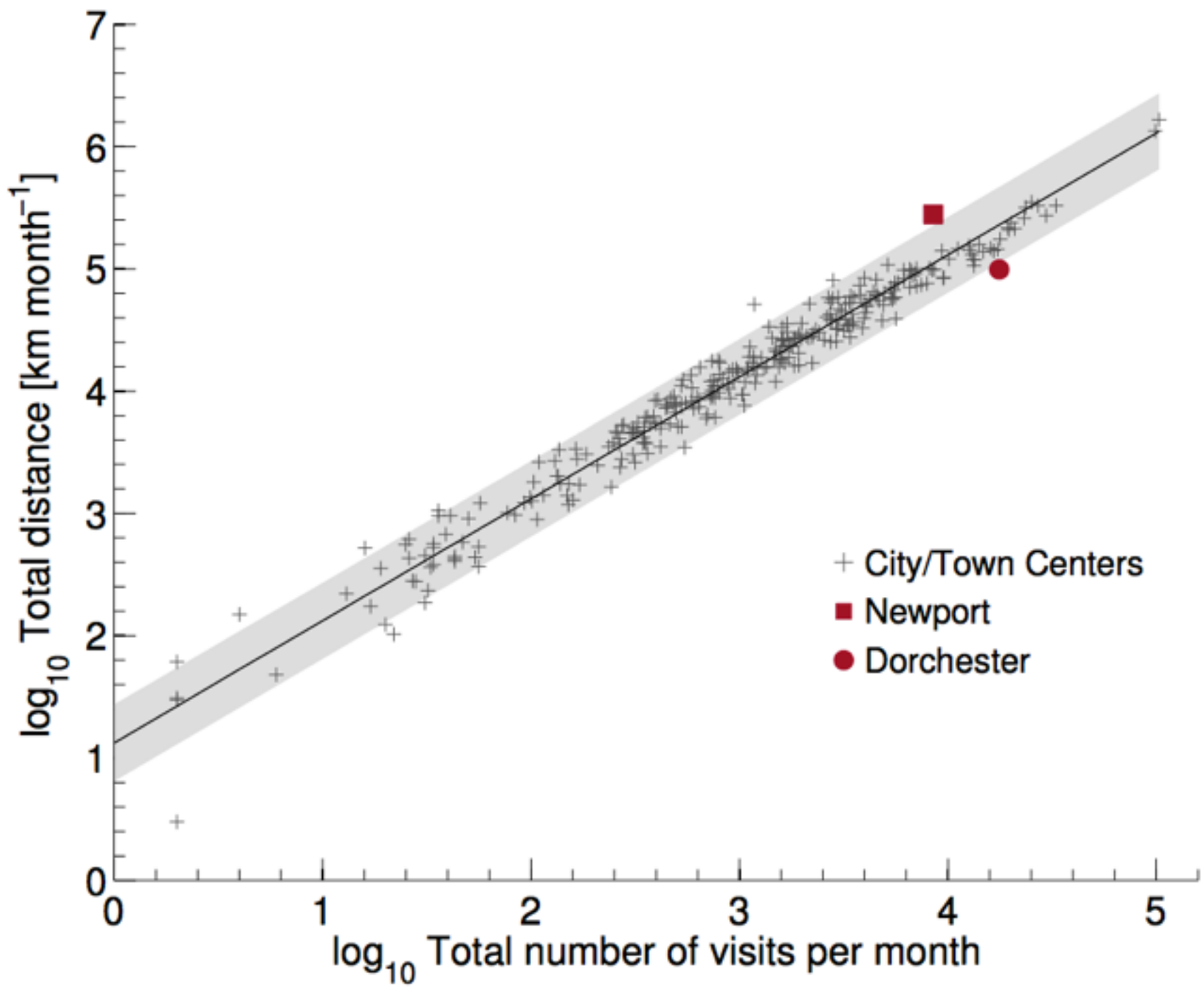

Jurong West
Tampines
City Hall
Bugis
Toa Payoh
1.34,103.71
1.36,103.94
1.29,103.85
1.30,103.85
1.33,103.85
1.30,103.83
1.43,103.83
1.44,103.79
1.30,103.84
1.39,103.90
1.37,103.89
1.28,103.84
1.36,103.95
1.31,103.85
1.31,103.86
1.33,103.93
1.32,103.93
1.37,103.85
1.34,103.70
1.35,103.94
1.35,103.72
1.35,103.75
1.44,103.80
1.37,103.85
1.28,103.83
Raffles Place
1.28,103.85
1.35,103.95
1.44,103.78
1.37,103.95
1.31,103.89
q(rf) [km^-2 month]
rf [km month^-1]



Schläpfer, Szell, Ratti, West (in preparation)

Locations with ‚anomalous‘ behavior



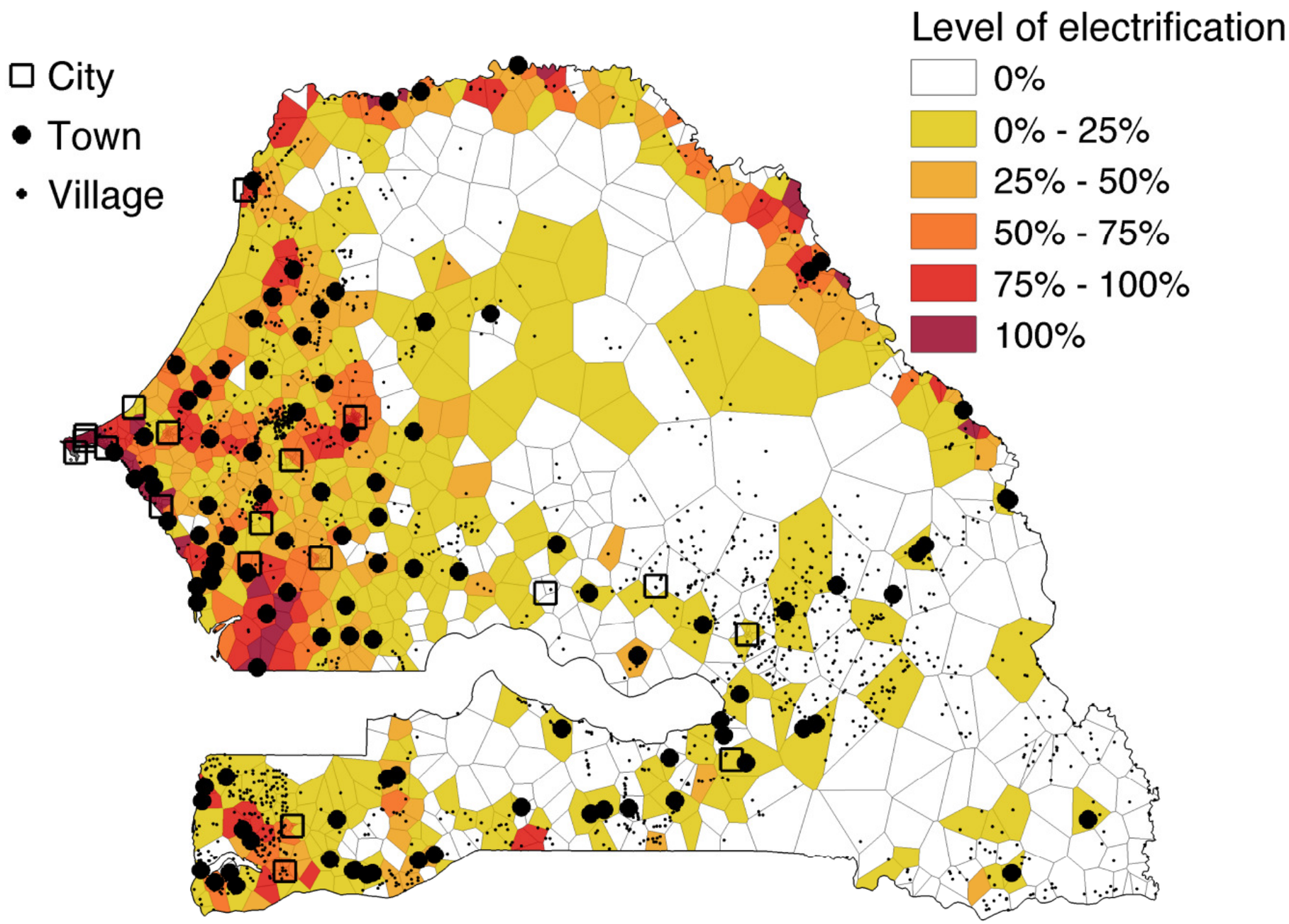

7
6
5
4
3
2
1
0
0
1
2
3
4
5
log10 Total distance [km month−1]
log10 Total number of visits per month
City/Town Centers
Newport
Dorchester

4. Application: Infrastructure design

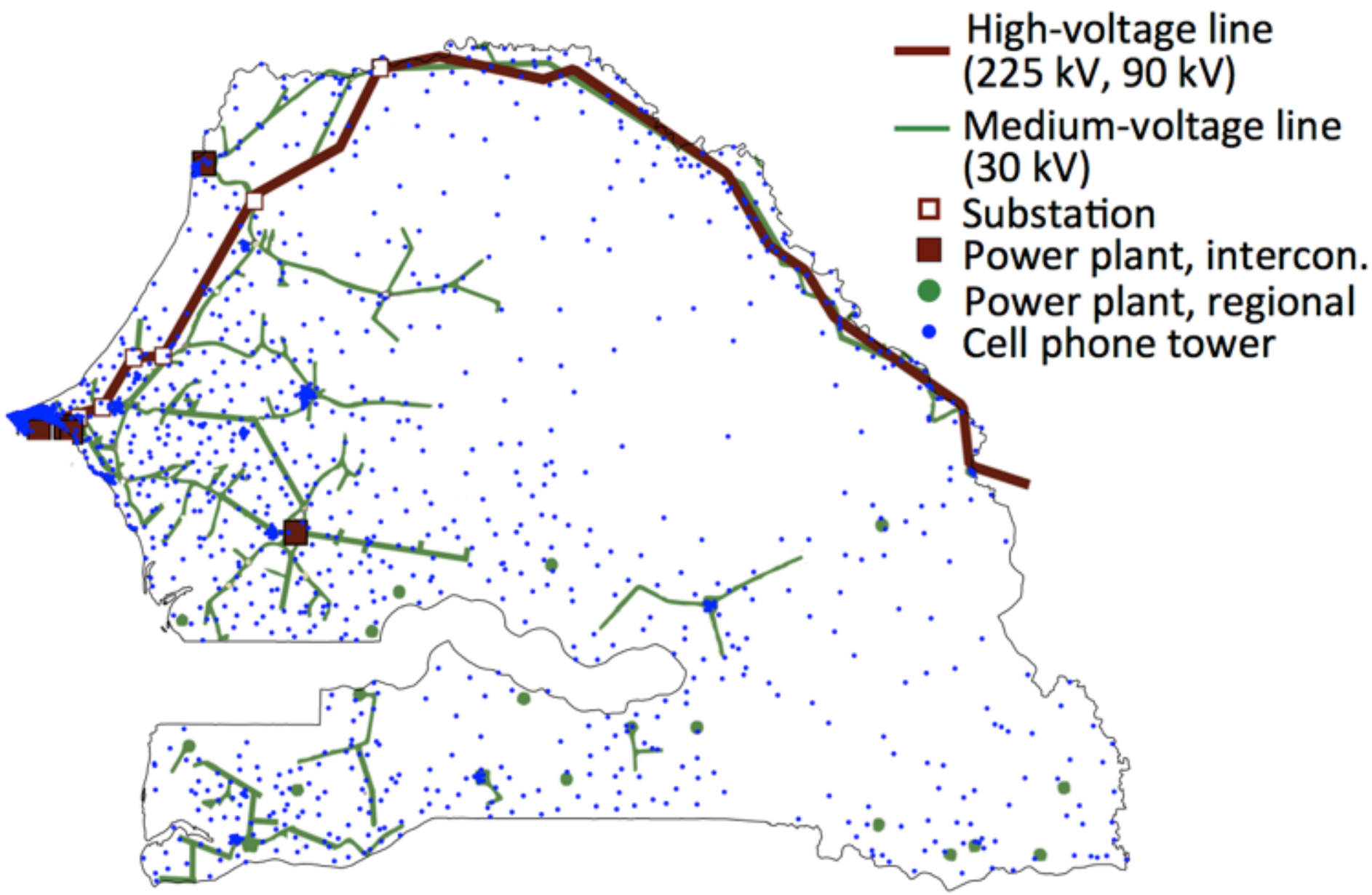
Electrification planning in developing countries

D4D challenge

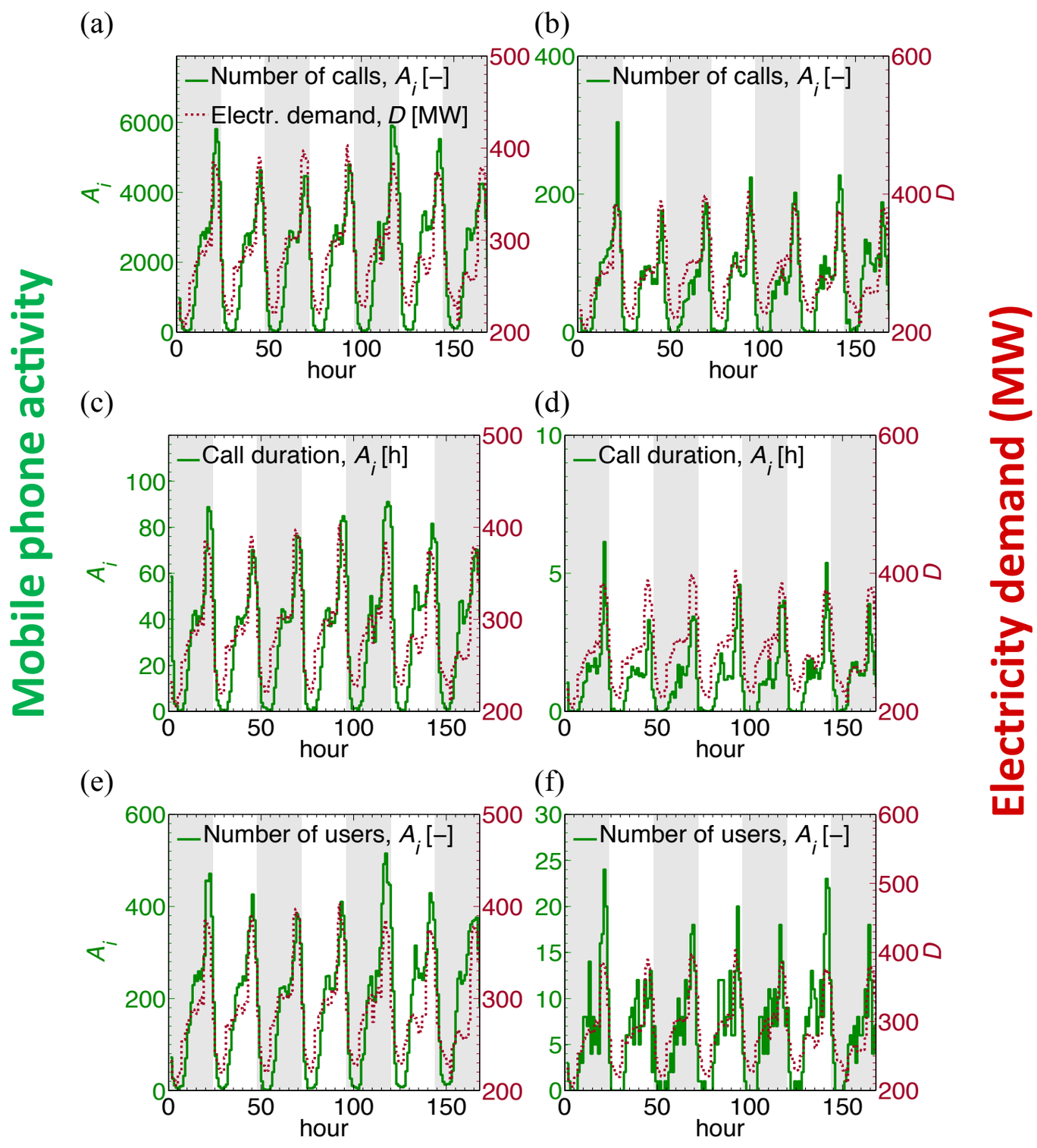
Electrification rates in Senegal



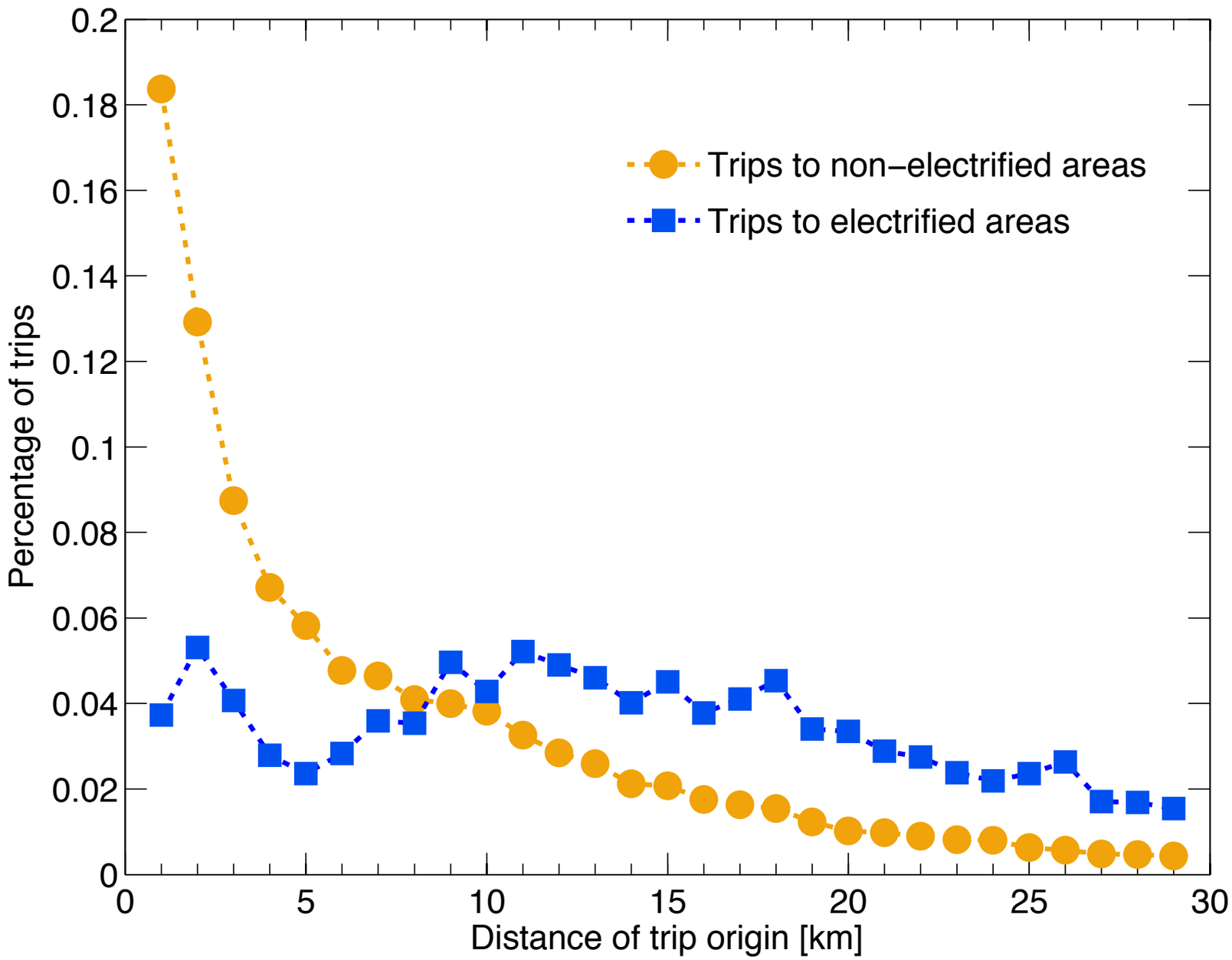
Using information from mobile phone infrastructure to facilitate electrification



Mobile phone data as a proxy for electricity demand



But not only…


Trips to non−electrified areas
Trips to electrified areas
Percentage of trips
Distance of trip origin [km]
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
0
5
10
15
20
25
30

Electrification technology optioneering: techno-economic analysis

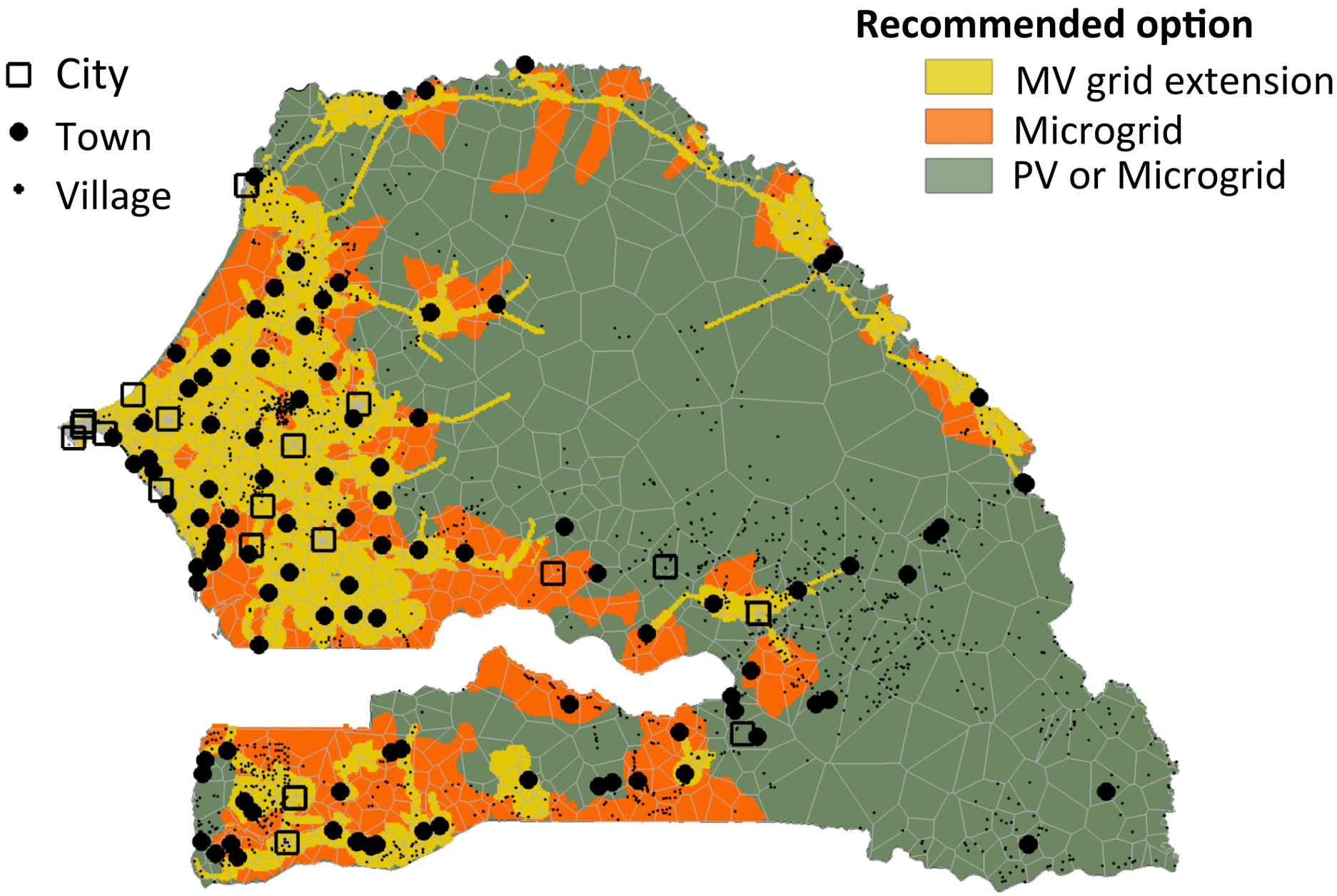
- Medium voltage electricity grid extension

- Diesel-based microgrid installation

- Traditional and minimalistic solar photovoltaic system

The model and equations are in the paper ;)

Electrification recommendations



Martinez-Cesena, Mancarella, Ndiaye, Schläpfer
D4D Challenge 2015, First Prize (best overall) and Energy Prize

Take home:
urban ‚big‘ data

„Don’t“s	„Do“s
give new insights per se	
are free of ‚hidden‘ biases	

Take home:
urban ‚big‘ data

„Don’t“s	„Do“s
give new insights per se	allow testing „old“ ideas
are free of ‚hidden‘ biases	cover large parts of the population
	objective measurements

Thank you!

Markus Schläpfer — schlaepfer@arch.ethz.ch