



HAL
open science

Systematic comparison of trip distribution laws and models

Maxime Lenormand, Aleix Bassolas, José J. Ramasco

► **To cite this version:**

Maxime Lenormand, Aleix Bassolas, José J. Ramasco. Systematic comparison of trip distribution laws and models. 2015. hal-01175812v1

HAL Id: hal-01175812

<https://hal.science/hal-01175812v1>

Preprint submitted on 13 Jul 2015 (v1), last revised 19 Jan 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Systematic comparison of trip distribution laws and models

Maxime Lenormand,¹ Aleix Bassolas,¹ and José J. Ramasco¹

¹*Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB),
Campus UIB, 07122 Palma de Mallorca, Spain*

Trip distribution laws are basic for the travel demand characterization needed in transport and urban planning. Several approaches have been considered in the last years. One of them is the so-called gravity law, in which the number of trips is assumed to be related to the population at origin and destination and to decrease with the distance. The mathematical expression of this law resembles Newton's law of gravity, which explains its name. Another popular approach is inspired by the theory of intervening opportunities and it has been concreted into the so-called radiation models. Individuals are supposed to travel until they find a job opportunity, so the population and jobs spatial distributions naturally lead to a trip flow network. In this paper, we perform a thorough comparison between the gravity and the radiation approaches in their ability at estimating commuting flows. We test the gravity and the radiation laws against empirical trip data at different scales and coming from different countries. Different versions of the gravity and the radiation laws are used to estimate the probability that an individual has to commute from one unit to another, called trip distribution law. Based on these probability distribution the commuting networks are simulated with different trip distribution models. We show that the gravity law performs better than the radiation law to estimate the commuting flows, to preserve the structure of the network and to fit the commuting distance distribution although it fails at predicting commuting flows at large distances. Finally, we show that both approaches can be used in absence of detailed data for calibration since the only parameter of both the gravity and the radiation laws depends only on the scale of the geographic unit.

INTRODUCTION

Everyday, billions of individuals around the world travel. These movements form a socio-economic complex network, backbone for the transport of people, goods, money, information or even diseases at different spatial scales. The study of such spatial networks is consequently the subject of an intensive scientific activity [1–3]. Some examples include the estimation of population flows [4–11], transport planning and modeling [1, 2], spatial network analysis [12, 13], study of urban traffic [12] and modeling of the spreading of infectious diseases [14–16].

Trip distribution modeling is thus crucial for the prediction of population movements, but also for an explanatory purpose, in order to better understand the mechanisms of human mobility. There are two major approaches for the estimation of trip distribution at an aggregate level. The traditional gravity approach, in analogy with the Newton's law of gravitation, is based on the assumption that the amount of trips between two locations is related to their populations and decays with a function of the distance [3, 17–20]. This approach has been widely used in the past few decades to model flows of population [4–6, 8, 9, 14–16], volume of international trade [21, 22], traffic in transport networks [23, 24] and phone communications [25]. In contrast to the gravity law, the Stouffer's law of intervening opportunities [26] hinges on the assumption that the number of opportunities plays a more important role in the location choices than the distance, particularly in the case of migration choices. In the same vein, the recently proposed radiation approach [7, 10, 27, 28] is inspired by a simple diffusion model where the amount of trips between two locations depends on their populations and the

number of opportunities between them.

The gravity law and the radiation law have been compared several times during the last years giving the superiority to either of the approaches depending on the study [7, 8, 10, 29, 30]. Two main issues can be identified in these comparisons. First, the inputs used to simulate the flows are not always identical. For example, in the comparison proposed in [29], the gravity law tested takes as input the population, whereas the radiation law is based on the number of jobs. Second, in all these studies, the models used to generate the trips from the radiation and the gravity laws are not constrained in the same way. The radiation models are always production constrained, this means that the number of trips, or at least an estimation of the number of trips generated by census unit, is preserved. The models used to generate the trips with the gravity laws can be either, unconstrained [29, 30], only the total number of trips is preserved or doubly constrained [8, 10], both the trips produced and attracted by a census unit are preserved. To fairly compare the gravity and the radiation approaches the same input data must be used and, most importantly, we need to differentiate the law, gravity or radiation, and the modeling framework used to generate the trips from this law. Indeed, both the gravity and the radiation laws can be express as a probability to move from one place to another, called trip distribution law, and based on these probability distributions, the total number of trips can then be simulated using different trip distribution models including different level of constraints.

In this work, we test and compare, in a systematic and rigorous way, gravity and radiation laws against commuting census data coming from six different countries using four different constrained mod-

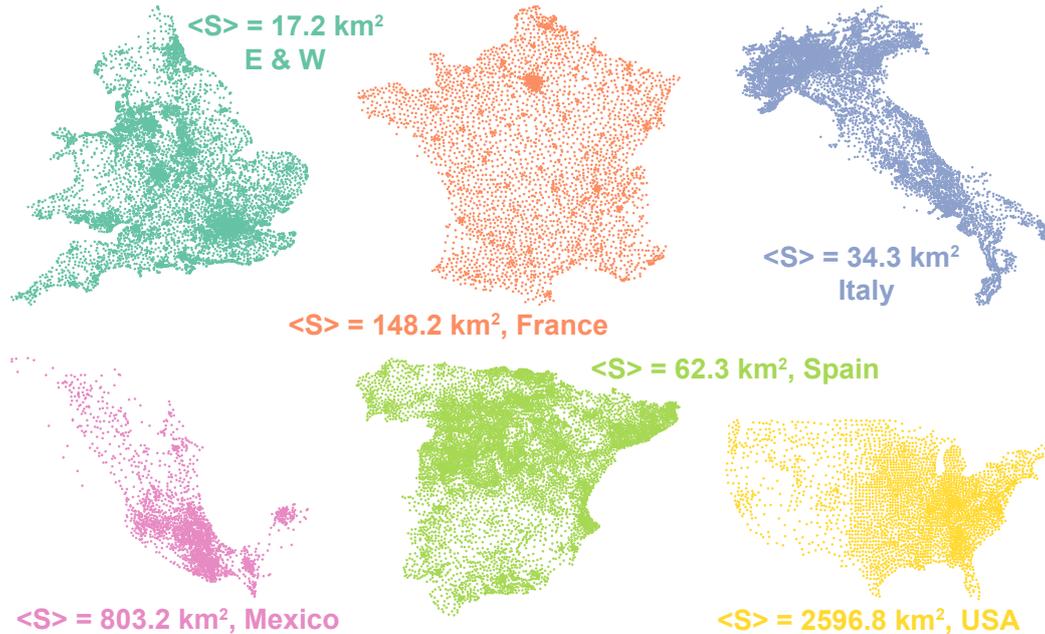


Figure 1: Position of the unit's centroids for the six countries. $\langle S \rangle$ represents the average surface of the census unit (i.e., municipalities, counties or wards).

els to generate the networks: unconstrained model, single constrained models (production or attraction) and the well-known doubly constrained model. For the gravity law, since the form of the distance decay functions may vary from one study to another [3, 14, 15, 31, 32] both the power and the exponential forms are tested to model the impact of the distance. For the radiation law, two versions are considered, the original free-parameter model [7] and the extended version proposed in [10]. The simulated networks are compared with the observed ones on different aspects showing that, globally, the gravity law with an exponential distance decay function outperforms the other laws in the estimation of commuting flows, the conservation of the commuting network structure and the fit of the commuting distance distribution even if it fails at predicting commuting flows at large distances. Finally, we show that both the gravity laws and the extended radiation law can be used in absence of detailed data for calibration since their only parameter depends only on the scale of the geographic census unit.

MATERIALS AND METHODS

Datasets

In this study, we test the gravity and the radiation laws against census commuting data of six countries: England and Wales, France, Italy, Mexico, Spain and the United States of America (hereafter called E&W, FRA, ITA, MEX, SPA and USA, respectively) and two cities: London and Paris (hereafter LON and

PAR, respectively). Each case study is divided into n census units of different spatial scale: from the Output Area in London with an average surface of 1.68 km^2 to the counties in the United States with an average surface of 2596.8 km^2 . Figures 1 and 2 display the centroids of the census units for the eight case studies. See the Appendix for a detailed description of the datasets. For each unit, the statistical offices provide the following information:

- T_{ij} , the number of trips between the census unit i and j (i.e. number of individuals living in i and working in j);
- d_{ij} , the great-circle distance between the unit i and the unit j computed with the Haversine formula;
- m_i , the number of inhabitants in unit i .

In this work we consider only inter-unit flows (i.e. $T_{ii} = 0$), mainly because it is not possible to estimate intra-units flows with the radiation laws. We note $N = \sum_{i,j=1}^n T_{ij}$ the total number of commuters, $O_i = \sum_{j=1}^n T_{ij}$ the number of out-commuters (i.e. number of individuals living in i and working in another census unit) and $D_j = \sum_{i=1}^n T_{ij}$ the number of in-commuters (i.e. number of individuals working in j and living in another census unit).

Comparison of trip distribution laws and models

The purpose of the trip distribution models is to split the total number of trips N in order to generate a

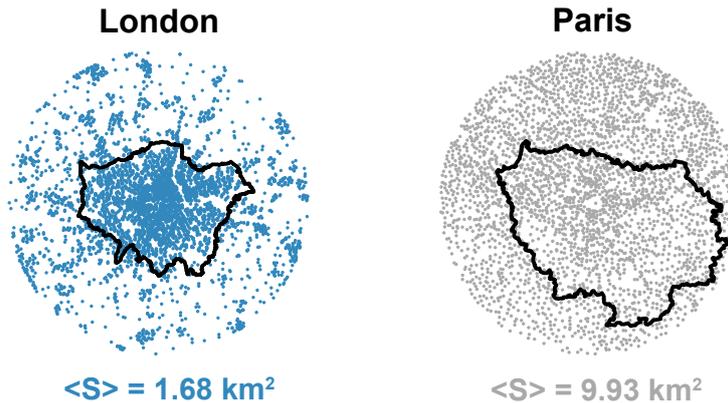


Figure 2: Position of the units' centroids around London (left) and Paris (right). The black contours represent the boundaries of the Greater London Authority (left) and the french *département* Ile de France (right). $\langle S \rangle$ represents the average unit surface.

trip table $\tilde{T} = (\tilde{T}_{ij})_{1 \leq i, j \leq n}$ of the estimated number of trips from each census area to every other. Note that by trip we are referring to commuting travels from home to work, there is a return trip not considered in \tilde{T} and N is also equivalent to the number of unique commuters. The trip distribution depends on, on one hand, the characteristics of the census units and the way they are spatially distributed, and, on the other hand, the level of constraints required by the model. Therefore, to fairly compare different trip distribution modeling approaches we have to consider separately the law used to calculate the probability to observe a trip between two census units, called trip distribution law, and the trip distribution model used to generate the trip allocation from this law.

Gravity and radiation laws

The purpose of this study is to test the capacity of both the gravity and the radiation approaches to estimate the probability p_{ij} that out of all the possible travels in the system we have one between the census area i and j . This probability is asymmetric in i and j as the flows themselves, and, by convention, the self-loops are excluded of the analysis $p_{ii} = 0$. This probability is normalized to all possible couples of origins and destinations, $\sum_{i, j=1}^n p_{ij} = 1$. Note that p_{ij} does not refer to the conditional probability of a trip starting in i finishes in j $\mathbb{P}(1|i, j)$. There exists a relation between both of them:

$$p_{ij} = \mathbb{P}(i) \mathbb{P}(1|i, j) \quad (1)$$

where $\mathbb{P}(i)$ stands for the probability of a trip starting in i . $\mathbb{P}(1|i, j)$ will appear later for the radiation law as a function of the populations of origin m_i , destination m_j and the jobs opportunities between them s_{ij} , $\mathbb{P}(1|m_i, m_j, s_{ij})$, but the basis of our analysis will be p_{ij} .

Gravity law In the simplest form of the gravity approach, the probability of commuting between two units i and j is proportional to the product of the origin population m_i and destination population m_j , and inversely proportional to the travel cost between the two units:

$$p_{ij} \propto m_i m_j f(d_{ij}), \quad i \neq j \quad (2)$$

The travel cost between i and j is usually modeled with an exponential distance decay function,

$$f(d_{ij}) = e^{-\beta d_{ij}} \quad (3)$$

or a power distance decay function,

$$f(d_{ij}) = d_{ij}^{-\beta} \quad (4)$$

As mentioned in [3], the form of the distance decay function can change according to the dataset, therefore, both the exponential and the power forms are considered in this study. In both cases, the importance of the distance in commuting choices is adjusted with a parameter $\beta > 0$.

Radiation law The radiation approach, proposed by [7], is inspired by a diffusion model where each individual living in an area i has a certain probability of being "absorbed" by another area j , which in turn produces the following trip distribution probability:

$$p_{ij} \propto m_i \frac{\mathbb{P}(1|m_i, m_j, s_{ij})}{\sum_{k=1}^n \mathbb{P}(1|m_i, m_k, s_{ik})}, \quad i \neq j \quad (5)$$

This probability is proportional to the origin population m_i and to $\mathbb{P}(1|m_i, m_j, s_{ij})$, which, as mentioned above, is the conditional probability that a commuter living in i with population m_i is attracted to j with population m_j , given that there are s_{ij} job opportunities in between. s_{ij} is the number of opportunities

(approximated by the population in this case) in a circle of radius d_{ij} centered in i (excluding the source and destination). This conditional probability needs to be normalized because the probability for an individual living in a census unit i of being absorbed by another census unit is not equal to 1 in case of finite system but equal to $1 - \frac{m_i}{M}$ where M is the total population [29].

The original radiation model is free of parameters and, therefore, it does not require calibration. The conditional probability $\mathbb{P}(1|m_i, m_j, s_{ij})$ is expressed as

$$\mathbb{P}(1|m_i, m_j, s_{ij}) = \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})} \quad (6)$$

However, some recent works have shown that the model fails to describe human mobility compared to more classic approaches particularly on a small scale [8, 29, 30, 33]. To circumvent these limitations, an extended radiation model have been proposed by [10]. In this extended version, the probability $\mathbb{P}(1|m_i, m_j, s_{ij})$ is derived under the survival analysis framework introducing a parameter α to adjust the effect of the scale,

$$\mathbb{P}(1|m_i, m_j, s_{ij}) = \frac{[(m_i + m_j + s_{ij})^\alpha - (m_i + s_{ij})^\alpha](m_i^\alpha + 1)}{[(m_i + s_{ij})^\alpha + 1][(m_i + m_j + s_{ij})^\alpha + 1]} \quad (7)$$

Constrained models

After the description of the probabilistic laws, the next step is to materialize the people commuting. The purpose is to generate the commuting network $\tilde{T} = (\tilde{T}_{ij})_{1 \leq i, j \leq n}$ by drawing at random N trips from the trip distribution law $(p_{ij})_{1 \leq i, j \leq n}$ respecting different level of constraints according to the model. We are going to consider four different types of models:

1. *Unconstrained model.* The only constraint of this model is to ensure that the total number of trips \tilde{N} generated by the model is equal to the total number of trips N observed in the data. In this model, the N trips are randomly sample from the multinomial distribution

$$\mathcal{M}\left(N, (p_{ij})_{1 \leq i, j \leq n}\right) \quad (8)$$

2. *Production constrained model.* This model ensures that the number of trips "produced" by a census unit is preserved. For each unit i , O_i trips are produced from the multinomial distribution

$$\mathcal{M}\left(O_i, \left(\frac{p_{ij}}{\sum_{k=1}^n p_{ik}}\right)_{1 \leq j \leq n}\right) \quad (9)$$

3. *Attraction constrained model.* This model ensures that the number of trips "attracted" by a unit is preserved. For each census unit j , D_j trips are attracted from the multinomial distribution

$$\mathcal{M}\left(D_j, \left(\frac{p_{ij}}{\sum_{k=1}^n p_{kj}}\right)_{1 \leq i \leq n}\right) \quad (10)$$

4. *Doubly constrained model.* This model, also called production-attraction constrained model ensures that both the trips attracted and generated by a census unit are preserved using two balancing factors K_i and K_j calibrated with the *Iterative Proportional Fitting* procedure [34]. The relation between K_i , K_j , p_{ij} and the trip flows is given by

$$\tilde{T}_{ij} = K_i K_j p_{ij} \quad (11)$$

Unlike the unconstrained and single constrained models, the doubly constrained model is a deterministic model. Therefore, the simulated network \tilde{T} is a fully connected network in which the flows are real numbers instead of integers. This can be problematic since we want to study the capacity of both the gravity and the radiation approaches to preserve the topological structure of the original network. To bypass this limitation N trips are randomly sample from the multinomial distribution

$$\mathcal{M}\left(N, \left(\frac{\tilde{T}_{ij}}{\sum_{k,l=1}^n \tilde{T}_{kl}}\right)_{1 \leq i, j \leq n}\right) \quad (12)$$

Goodness-of-fit measures

Common part of commuters We calibrate the parameters β and α using the common part of commuters (CPC) introduced in [6, 8]:

$$CPC(T, \tilde{T}) = \frac{2 \sum_{i,j=1}^n \min(T_{ij}, \tilde{T}_{ij})}{\sum_{i,j=1}^n T_{ij} + \sum_{i,j=1}^n \tilde{T}_{ij}} \quad (13)$$

This indicator is based on the Sørensen index [35]. It varies from 0, when no agreement is found, to 1, when the two networks are identical. In our case, the total number of commuters N is preserved, therefore the Equation (13) can be simplify to

$$CPC(T, \tilde{T}) = 1 - \frac{1}{2} \frac{\sum_{i,j=1}^n |T_{ij} - \tilde{T}_{ij}|}{N} \quad (14)$$

which represents the percentage of good prediction as defined in [36].

In order to assess the robustness of the results regarding the choice of goodness-of-fit measures, we also test the results obtained with the normalized root mean square error

$$NRMSE(T, \tilde{T}) = \frac{\sum_{i,j=1}^n (T_{ij} - \tilde{T}_{ij})^2}{\sum_{i,j=1}^n T_{ij}} \quad (15)$$

and the information gain statistic

$$I(T, \tilde{T}) = \sum_{i,j=1}^n \frac{T_{ij}}{N} \ln \left(\frac{T_{ij}}{\tilde{T}_{ij}} \right) \quad (16)$$

$I(T, \tilde{T})$ is zero if T and \tilde{T} are equal and grows as the difference between them increases.

Common part of links The ability of the models to recover the topological structure of the original network can be assessed with the common part of links (CPL) defined as

$$CPL(T, \tilde{T}) = \frac{2 \sum_{i,j=1}^n \mathbb{1}_{T_{ij}>0} \cdot \mathbb{1}_{\tilde{T}_{ij}>0}}{\sum_{i,j=1}^n \mathbb{1}_{T_{ij}>0} + \sum_{i,j=1}^n \mathbb{1}_{\tilde{T}_{ij}>0}} \quad (17)$$

where $\mathbb{1}_X$ is equal to one if the condition X is fulfilled and zero otherwise. The common part of links measures the proportion of links in common between the simulated and the observed networks (i.e. links such as $T_{ij} > 0$ and $\tilde{T}_{ij} > 0$). It is null if there is no link in common and one if both networks are topologically equivalent.

Common part of commuters according to the distance In order to measure the similarity between the observed commuting distance distribution and the ones simulated with the models, we introduce the common part of commuters according to the distance (CPC_d). Let us consider N_k the number of individuals having a commuting distance in the bin between $2k - 2$ and $2k$ kms. The CPC_d is equal to the CPC based on N_k instead of T_{ij}

$$CPC_d(T, \tilde{T}) = \frac{\sum_{k=1}^{\infty} \min(N_k, \tilde{N}_k)}{N} \quad (18)$$

RESULTS

In this section, we compare the four laws: gravity with an exponential or a power distance decay function, and the original and the extended radiation laws. We test these laws against empirical data coming from eight different case studies using four constrained models to estimate the flows. For each

constrained model, the parameters β and α are calibrated so as to maximize the CPC. Since the models are stochastic, we consider an average CPC value measured over 100 replications of the trip distribution. Similarly, all the goodness-of-fit measures are obtained by calculating the average measured over 100 network replications. It is important to note that the networks generated with the constrained models are very stable, the stochasticity of the models does not affect the statistical properties of the network. Therefore, the goodness-of-fit measures does not vary much with the different realizations of the multinomial sampling. For example, within the 100 network instances for all models and case studies, the CPC varies, at most, by 0.09% around the average.

Estimation of commuting flows

Figure 3 displays the common part of commuters obtained with the different laws and models for the eight case studies. Globally, the gravity laws give better results than the radiation laws. For the gravity laws, the results improve with the exponential rather than with the power distance decay function. For the radiation laws, the extended version outperforms the original one. In the top left panel, we observe the results for the unconstrained model. In this case, the radiation laws give better results than the gravity for most case studies. However, these better performances are due to the normalization factor used in Equation 5. Indeed, this normalization implies that the probability of having a trip originating in a census unit i is proportional to the population of i , which is not necessarily the case for the gravity laws. If we use the same type of normalization for the gravity p_{ij} :

$$p_{ij} \propto m_i \frac{m_j f(d_{ij})}{\sum_{k=1}^n m_k f(d_{ik})}, \quad i \neq j \quad (19)$$

we observe that the "normalized" gravity laws give better results than the radiation (dark red circles and blue squares in the top left panel of Figure 3).

To compare the constrained models performances, we plot in Figure 4a the CPC obtained with the four models according to the laws averaged over the eight case studies. As expected, more constrained the model is, higher the CPC becomes. Unconstrained models are able to reproduce on average around 50% of the observed commuting network against 70% for the doubly constrained model. It is interesting to note that, the attraction constrained model gives better results than the production constrained model. This can be explained by the fact that the job demand is easier to estimate than the job offer, which can be related to extra economic questions. This is in agreement with the results obtained with a uniform distribution ($p_{ij} \propto 1$) plotted in Figure 3d.

The results obtained with the normalized root mean square error, the information gain statistic and the

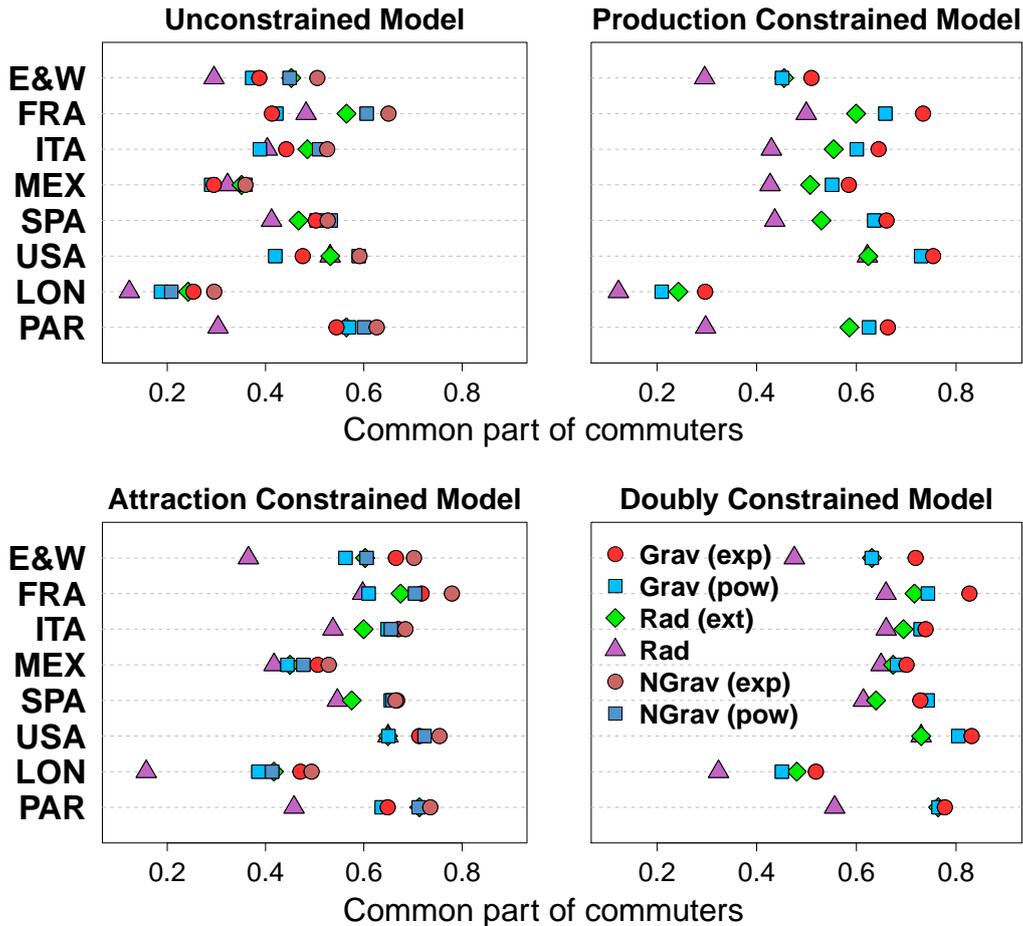


Figure 3: Common part of commuters according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum values observed in the 100 realizations but in most cases they are too close to the average to be seen.

average common part of in- and out-commuters are available in Appendix (Figures S1-S4). Although the results obtained with these goodness-of-fit measures are very similar to the ones obtained with the CPC, one question to note is that, except in England & Wales and in London, the radiation laws give smaller normalized root mean square error values than the gravity and normalized gravity laws with the unconstrained model.

Structure of the commuting network

We consider next the capacity of the gravity and the radiation laws to recover the structure of the empirical commuting networks. Figure 4b shows the average common part of links obtained with the different laws and models (the CPL according to the unconstrained models, the gravity and radiation laws for the eight case studies are available in Figure S5 in

the Appendix). We observe that the gravity law with an exponential distance decay function outperforms the other laws when the unconstrained and the single constrained models are used to generate the flows. However, when the doubly constrained models is used, very similar results are obtained except for the original version of the radiation law. In any case, the common part of links never exceed 0.55, this can be explained by the fact that, globally, the different laws fail at reproducing the number of links (as well as the average degree of the commuting networks).

Indeed, as it can be seen in Figure 5, which displays the ratio between the number of links generated with the models and the observed ones, the radiation law and the gravity law with the exponential distance decay function tend to underestimate the number of links whereas the two other laws overestimate it. The same results are found with the average degree (see Figure S6 in Appendix for more details).

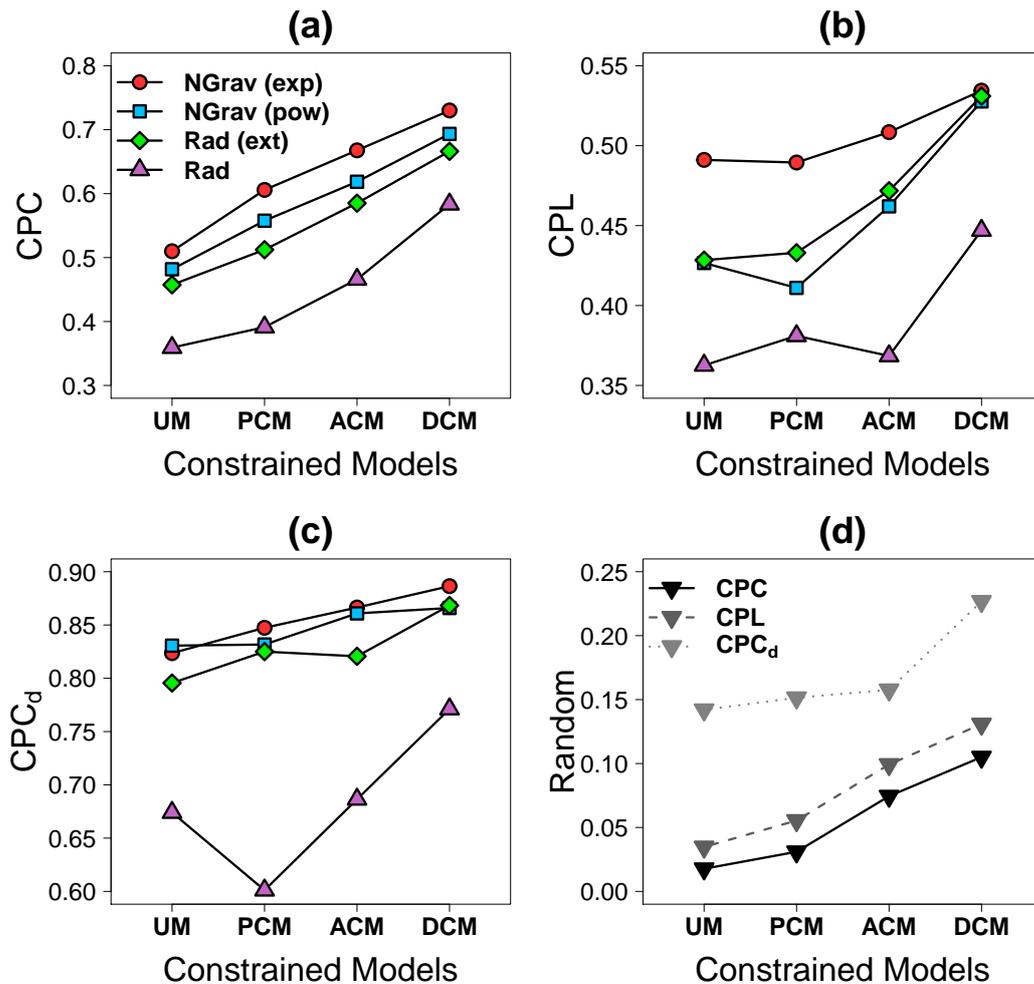


Figure 4: Performance of the unconstrained model (NC), the production constrained model (PCM), the attraction constrained model (ACM) and the doubly constrained model (DCM) according to the gravity and the radiation laws (a)-(c) and a uniform distribution (d). (a) Average CPC. (b) Average CPL. (c) Average CPC_d. The red circles represent the normalized gravity law with the exponential distance decay function; The cyan squares represent the normalized gravity law with the power distance decay function; The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. The point down triangles represent the uniform distribution, from dark to light grey, the CPC, the CPL and the CPC_d

Commuting distance distribution

Another important feature to study is the commuting distance distribution. Figure 4c shows the average common part of commuters according to the distance obtained with the different models and laws (the CPC_d for the eight case studies are available in Figure S7 in the Appendix). Here again, the best results are obtained with the gravity law with the exponential distance decay function. Although, the results are globally good, and, except the original radiation law, the gravity and radiation laws are able to reproduce more than 80% of the commuting distances.

To go further, we plot in Figure 6 the observed and the simulated commuting distance distributions obtained with the production constrained model in France and United States. We can clearly see that

the gravity law with an exponential distance decay function is better for estimating commuting distances which are below a certain threshold equal to 50 km in France and 150 km in United States. After this threshold, the gravity law with an exponential distance decay function fails at estimating the commuting flows. On the contrary, the radiation laws and the gravity law with a power distance decay function are able to estimate commuting flows at large distances. However, we have to keep in mind that the proportion of commuters traveling such long distances are less than 6% in France and 5% in United States. Besides, one can legitimately wonder whether these long travels are repeated twice per day or if they may be an artifact of the way in which the census information is collected.

In addition to inter-census units trip networks, we have also consider within-city mobility. In order to

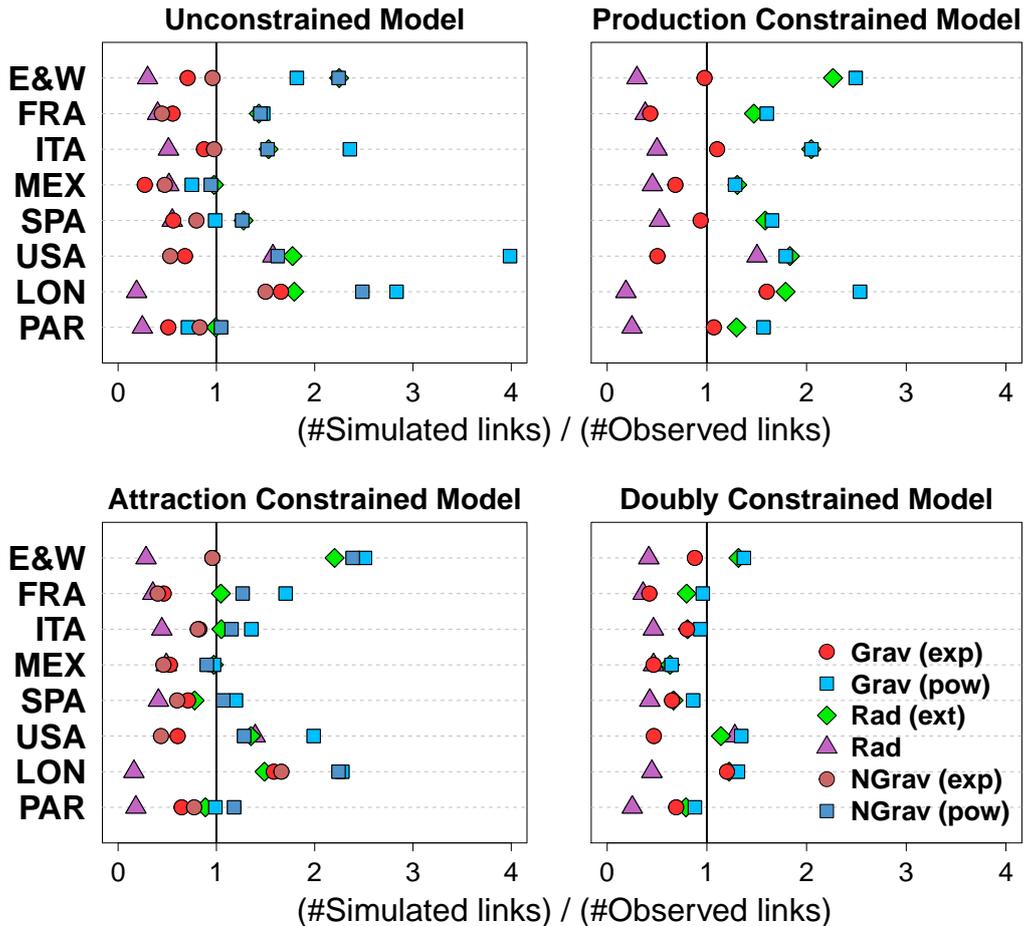


Figure 5: Ratio between the simulated and the observed number of links according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

study the ability of both approaches at predicting intra-city trips, we use Twitter data to extract a commuting networks between grid cells of 1 km² in the metropolitan area of Madrid. The results obtained are consistent with those shown here for the census (see Figure S8 in Appendix for more details).

Gravity and radiation laws based on the number of out- and in-commuters

In Equations 2 and 5, the population is used instead of the number of out-commuters O_i and the number of in-commuters D_j , which are usually preferred since they are a more faithful reflection of the job demand and offer. The job demand and offer are considered to be related to the population but the proportion is rarely direct (it needs to be adjusted with an exponent) and according to the case study, the fit can

be bad (see Figures S9-S10 in Appendix for more details). In order to assess the robustness of the results to changes in the input data, we consider the results obtained with the gravity law (Equation 20) and the radiation law (Equation 21) based on the number of out- and in-commuters. In the radiation case, s_{ij} is the number of in-commuters in a circle of radius d_{ij} centered in i (excluding the source and destination) and the role of the populations in the gravity law is taken by O_i and D_j . To be more specific, the gravity law becomes:

$$p_{ij} \propto O_i \frac{D_j f(d_{ij})}{\sum_{k=1}^n D_k f(d_{ik})}, \quad i \neq j \quad (20)$$

while the radiation law can be written as

$$p_{ij} \propto O_i \frac{\mathbb{P}(1|D_i, D_j, s_{ij})}{\sum_{k=1}^n \mathbb{P}(1|D_i, D_k, s_{ik})}, \quad i \neq j \quad (21)$$

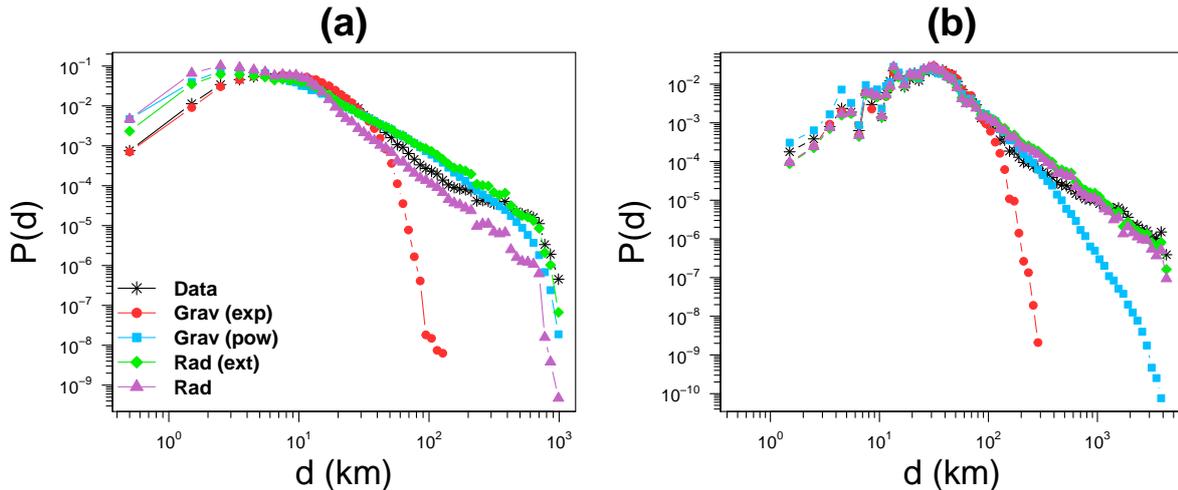


Figure 6: Commuting distance distribution observed in the data and simulated with the production constrained model. (a) France and (b) United States. The red circles represent the gravity law with the exponential distance decay function; The cyan squares represent the gravity law with the power distance decay function; The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. The black stars represent the census data.

Figure 7 displays the average CPC, CPL and CPC_d obtained with the four models according to the laws averaged over the eight case studies (results per case studies are available in Figures S11-S13 in Appendix). As it can be seen on these plots the results observed in Figure 4 are quite stable to changes in the input data.

Parameter calibration in the absence of detailed data

An important issue with the estimation of commuting flows is the model or law calibration. Indeed, how to calibrate the parameters β and α in the absence of detailed data? This problem has already been tackled in previous studies [8, 10, 15]. In [8], the authors have shown that, in the case of the exponential form of the gravity law, the value of β can be directly inferred from the average census unit surface with the relationship $\beta = 0.3 \langle S \rangle^{-0.18}$. Similarly, [10] proposed to estimate the value of α in the extended radiation law with the average spatial scale $l = \sqrt{\langle S \rangle}$ using the functional relationship $\alpha = 0.0085 l^{1.33}$.

In Figure 8a, 8b and 8c, we plot the calibrated value of β and α as a function of the average census unit surface $\langle S \rangle$ for the four constrained models. Figure 8a shows the relationship obtained with the gravity law with an exponential distance decay function. We observe that the coefficients of the relationship are the same than the one obtained in [8]. This is not surprising since three datasets out of the six used here coincide. In this case, the value of β decreases with larger spatial scales. This can be explained by the fact that β in the exponential form of the gravity law is proportional to the inverse of the characteristic distance and

such distance increases with the average unit surface since the shorter distance trips are excluded (Figure S13b in Appendix). Figure 8b displays the same relationship for the power form of the gravity law, in this case the value of β increases with the scale to fit the tail of the distribution. Larger (negative) exponents imply a faster decay of the power-law correcting factor. In fact, we observe in the data that, globally, the absolute value of the slope of the tail increases with the scale (Figure S13c in Appendix). Finally, we plot in Figure 8c the relationship between α and the average unit surface, the exponent obtained is similar to the one reported in [10].

As in [8], it is possible to assess the quality of the parameter estimation by measuring its impact on the CPC. For each law, model and case study, we have measured the relative error between the CPC obtained with the calibrated value of the parameter and the CPC obtained with the simulated one. The results are presented in Figure 8c. The relative errors are globally small and vary at most by 4% of the original CPC values for the gravity laws and 10% for the extended radiation law. This means that the parameter value can be directly inferred from the scale, and thus, commuting networks at different scales can be generated without requiring detailed data for calibration.

DISCUSSION

In summary, we have compared different versions of the gravity and the radiation laws. These two approaches have already been compared in the past but using different inputs, number of parameters and/or type of constraints. For this reason, the aim of this work has been to bring some light into the discus-

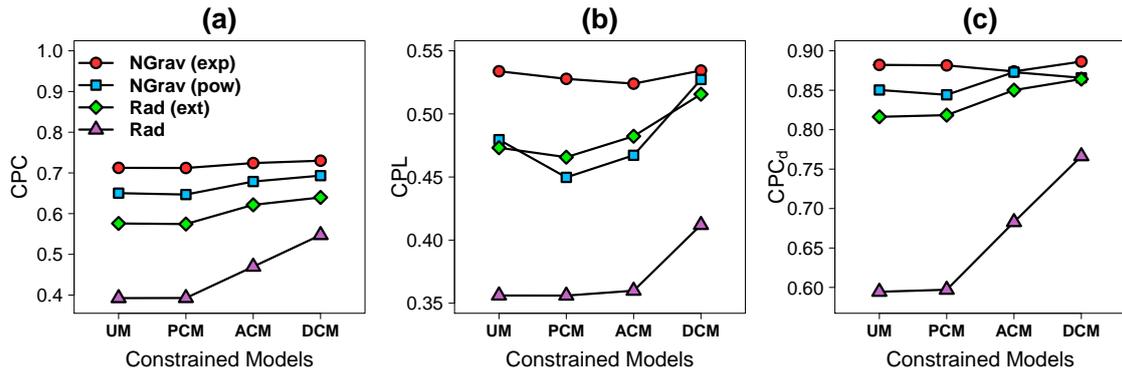


Figure 7: Performance of the unconstrained models according to the gravity and the radiation laws. (a) Average CPC. (b) Average CPL. (c) Average CPC_d . The red circles represent the gravity law with the exponential distance decay function; The cyan squares represent the gravity law with the power distance decay function; The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law.

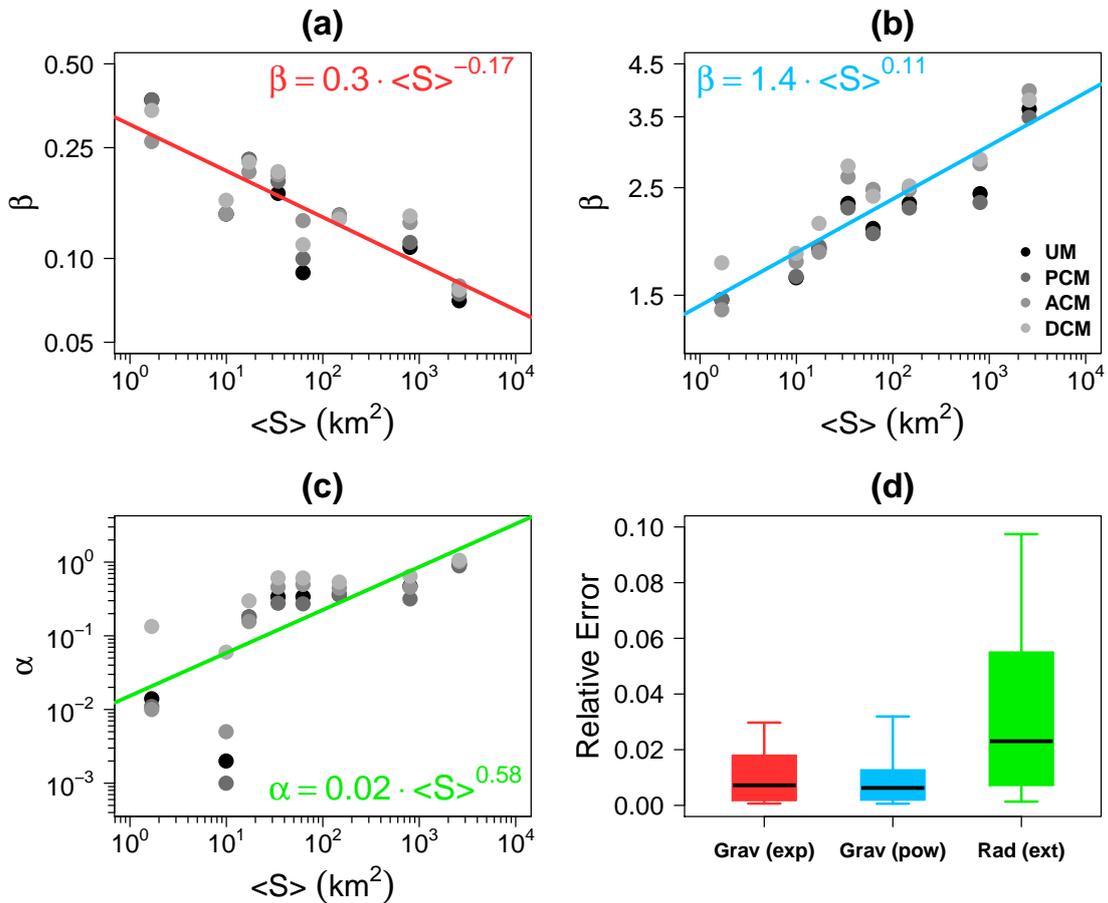


Figure 8: Parameter value as a function of the average unit surface. (a) Gravity laws with an exponential distance decay function. (b) Gravity laws with a power distance decay function. (c) Extended radiation law. (d) Boxplots of the relative error between the CPC obtained with a calibrated value of the parameters and the CPC obtained with values estimated with the regression models. The boxplot is composed of the first decile, the lower hinge, the median, the upper hinge and the last decile.

sion by systematically comparing the radiation and the gravity laws taking care of dissociating the probabilistic laws and the constrained models used to generate the trip networks. We show that, globally, the gravity approach outperforms the radiation laws to estimate the commuting flows but also to preserve the commuting network structure and to fit of the commuting distance distribution. More particularly the gravity law with the exponential distance decay function give better results than the other laws even if it fails at estimating commuting flows at large distances. The reason for this is that most of the travels are short-range, which are better capture by the gravity law with exponential decay in the distance. The large distance commuting trips are few and probably associated with weekly rather than daily commuting. To handle these different types of mobility, it may be necessary to investigate further the nature of the trips and to consider even mixed models for different displacement lengths. The superiority of the gravity law is very robust to the choice of goodness-of-fit measure and to the change of input. Regarding a more practical issue which is the calibration of the parameters without detailed data, we show that the parameter values can be estimated with the average unit area. This allows for a direct estimation of the commuting flows even in the absence of detailed data for calibration.

The objective of this work has been to establish the

basis for a fair and systematic comparison separating probabilistic laws and different degrees of constraint trip generation models. The use of these models in contexts such as urban and infrastructure planning, where large investments are at stake, imposes the need for the selection of the aptest model before taking decisions based on its results.

The software package to generate spatial networks using the approach described in the paper can be downloaded from <https://github.com/MaximeLenormand/RadGrav>.

ACKNOWLEDGEMENTS

Partial financial support has been received from the Spanish Ministry of Economy (MINECO) and FEDER (EU) under the project INTENSE@COSYP (FIS2012-30634), and from the EU Commission through projects LASAGNE and INSIGHT. The work of ML has been funded under the PD/004/2013 project, from the Conselleria de Educacin, Cultura y Universidades of the Government of the Balearic Islands and from the European Social Fund through the Balearic Islands ESF operational program for 2013-2017. JJR acknowledges funding from the Ramón y Cajal program of MINECO.

-
- [1] J. Rouwendal and P. Nijkamp. Living in two worlds: A review of home-to-work decisions. *Growth and Change*, 35(3):287–303, 2004.
 - [2] J.D. Ortúzar and L.G. Willumsen. *Modeling Transport*. John Wiley and Sons Ltd, New York, 2011.
 - [3] M. Barthelemy. Spatial Networks. *Physics Reports*, 499:1–101, 2011.
 - [4] J.-M. Choukroun. A general framework for the development of gravity-type trip distribution models. *Regional Science and Urban Economics*, 5(2):177–202, 1975.
 - [5] A. G. Wilson. Land-Use/Transport Interaction Models: Past and Future. *Journal of Transport Economics and Policy*, 32(1):pp. 3–26, 1998.
 - [6] F. Gargiulo, M. Lenormand, S. Huet, and O. Baqueiro Espinosa. Commuting network model: getting to the essentials. *Journal of Artificial Societies and Social Simulation*, 15(2):13, 2012.
 - [7] F. Simini, M. C. González, A. Maritan, and A.-L. Barabasi. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.
 - [8] M. Lenormand, S. Huet, F. Gargiulo, and G. Defuant. A Universal Model of Commuting Networks. *PLoS ONE*, 7:e45985, 2012.
 - [9] M. Lenormand, S. Huet, and F. Gargiulo. Generating French Virtual Commuting Network at Municipality Level. *Journal of Transport and Land Use*, 7(1):43–55, 2014.
 - [10] Y. Yang, C. Herrera, N. Eagle, and M. C. González. Limits of Predictability in Commuting Flows in the Absence of Data for Calibration. *Scientific Reports*, 4(5662), 2014.
 - [11] O. Sagarra, M. Szell, P. Santi, A. Diaz-Guilera, and C. Ratti. Supersampling and network reconstruction of urban mobility. *arXiv preprint arXiv:1504.01939*, 2015.
 - [12] A. De Montis, M. Barthelemy, A. Chessa, and A. Vespignani. The structure of interurban traffic: A weighted network analysis. *Environment and Planning B: Planning and Design*, 34(5):905–924, 2007.
 - [13] A. De Montis, A. Chessa, M. Campagna, S. Caschili, and G. Deplano. Modeling commuting systems through a complex network analysis: A study of the Italian islands of Sardinia and Sicily. *The Journal of Transport and Land Use*, 2(3):39–55, 2010.
 - [14] C. Viboud, O.N. Bjrnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. *Science*, 312(5772):447–451, 2006.
 - [15] D. Balcan, V. Colizza, B. Goncalves, H. Hud, J.J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21484–21489, 2009.
 - [16] M. Tizzoni, Bajardi P., Decuyper A., Kon Kam King G., Schneider C. M., Blondel V., Smoreda Z., González M. C., and Colizza Vittoria. On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Comput Biol*, 10(7):e1003716, 07 2014.
 - [17] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences*,

- pages 666–704, 1781.
- [18] H. C. Carey. *Principles of Social Science*. Lippincott, 1858.
- [19] G. K. Zipf. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, 11(6):677–686, 1946.
- [20] S. Erlander and N. F. Stewart. *The Gravity model in transportation analysis : theory and extensions*. Topics in transportation. VSP, Utrecht, The Netherlands, 1990.
- [21] J. Anderson. A Theoretical Foundation for the Gravity Equation. *American Economic Review*, 69(1):106–16, 1979.
- [22] J. Bergstrand. The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *The Review of Economics and Statistics*, 67(3):474–81, 1985.
- [23] W.-S. Jung, F. Wang, and H.E. Stanley. Gravity model in the korean highway. *Europhysics Letters*, 81(48005), 2008.
- [24] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of The Royal Society Interface*, 7(48):1097–1103, January 2010.
- [25] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):L07003+, July 2009.
- [26] S. A. Stouffer. Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, 5(6):845–867, 1940.
- [27] F. Simini, Maritan A., and Néda Z. Human Mobility in a Continuum Approach. *PLoS ONE*, 8(3):e60069, 03 2013.
- [28] Y. Ren, M. Ercsey-Ravasz, P. Wang, M. C. González, and Z. Toroczkai. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications*, 5(5347), 2014.
- [29] A. Masucci, J. Serras, A. Johansson, and M. Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Phys. Rev. E*, 88:022812, Aug 2013.
- [30] X. Liang, J. Zhao, L. Dong, and K. Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3(2983), 2013.
- [31] A.S. Fotheringham. Spatial structure and distance-decay parameters. *Annals, Association of American Geographers*, 71(3):425–436, 1981.
- [32] J.J. de Vries, P. Nijkamp, and P. Rietveld. Exponential or power distance-decay for commuting? An alternative specification. *Environment and Planning A*, 41(2):461–480, 2009.
- [33] H. Commenges. Modèle de radiation et modèle gravitaire : du formalisme à l’usage. In *SAGEO*, 2014.
- [34] W. E. Deming and F. F. Stephan. On a Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals Are Known. *Annals of Mathematical Statistics*, 11:427–444, 1940.
- [35] T. Sørensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, 5:1–34, 1948.
- [36] M. Lenormand and G. Deffuant. Generating a Synthetic Population of Individuals in Households: Sample-Free vs Sample-Based Methods. *Journal of Artificial Societies and Social Simulation*, 14(12):12, 2013.
- [37] A. and Colet P. Tugores. Big data and urban mobility. In *Proceedings of the 7th Iberian Grid infrastructure conference*, 2013.
- [38] M. Lenormand, M. Picornell, O. Garcia Cantú, A. Tugores, T. Louail, R. Herranz, M. Barthélemy, E. Frías-Martínez, and J. J. Ramasco. Cross-checking different source of mobility information. *PLoS ONE*, 9(8):e105184, 2014.

APPENDIX

Datasets

- The England & Wales dataset comes from the 2001 Census in England and Wales made available by the Office for National Statistics (data available online at <https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=124>).
- The French dataset was measured for the 1999 French Census by the French Statistical Institute (data available upon request at http://www.cmh.ens.fr/greco/adisp_eng.php).
- The Italian’s commuting network was extracted from the 2001 Italian Census by the National Institute for Statistics (data available upon request at <http://www.istat.it/it/archivio/139381>).
- Data on commuting trips between Mexican’s municipalities in 2011 are based on a microdata sample coming from the Mexican National Institute for Statistics (data available online at <http://www3.inegi.org.mx/sistemas/microdatos/default2010.aspx>).
- The Spanish dataset comes from the 2001 Spanish Census made available by the Spanish National Statistics Institute (data available upon request at http://www.ine.es/en/censo2001/index_en.html).
- Data on commuting trips between United States counties in 2000 comes from the United State Census Bureau (data available online at <https://www.census.gov/population/www/cen2000/commuting/index.html>).

Table SI: Presentation of the datasets

Case study	Number of units	Number of links	Number of Commuters
England & Wales	8,846 wards	1,269,396	18,374,407
France	3,645 cantons	462,838	12,193,058
Italy	7,319 municipalities	419,556	8,973,671
Mexico	2,456 municipalities	60,049	603,688
Spain	7,950 municipalities	261,084	5,102,359
United State	3,108 counties	161,522	34,097,929
London	4,664 Output Areas	750,943	4,373,442
Paris	3,185 municipalities	277,252	3,789,487

Normalized root mean square error

The normalized root mean square error weights large errors more heavily than small ones.

$$NRMSE(T, \tilde{T}) = \frac{\sum_{i,j=1}^n (T_{ij} - \tilde{T}_{ij})^2}{\sum_{i,j=1}^n T_{ij}} \quad (22)$$

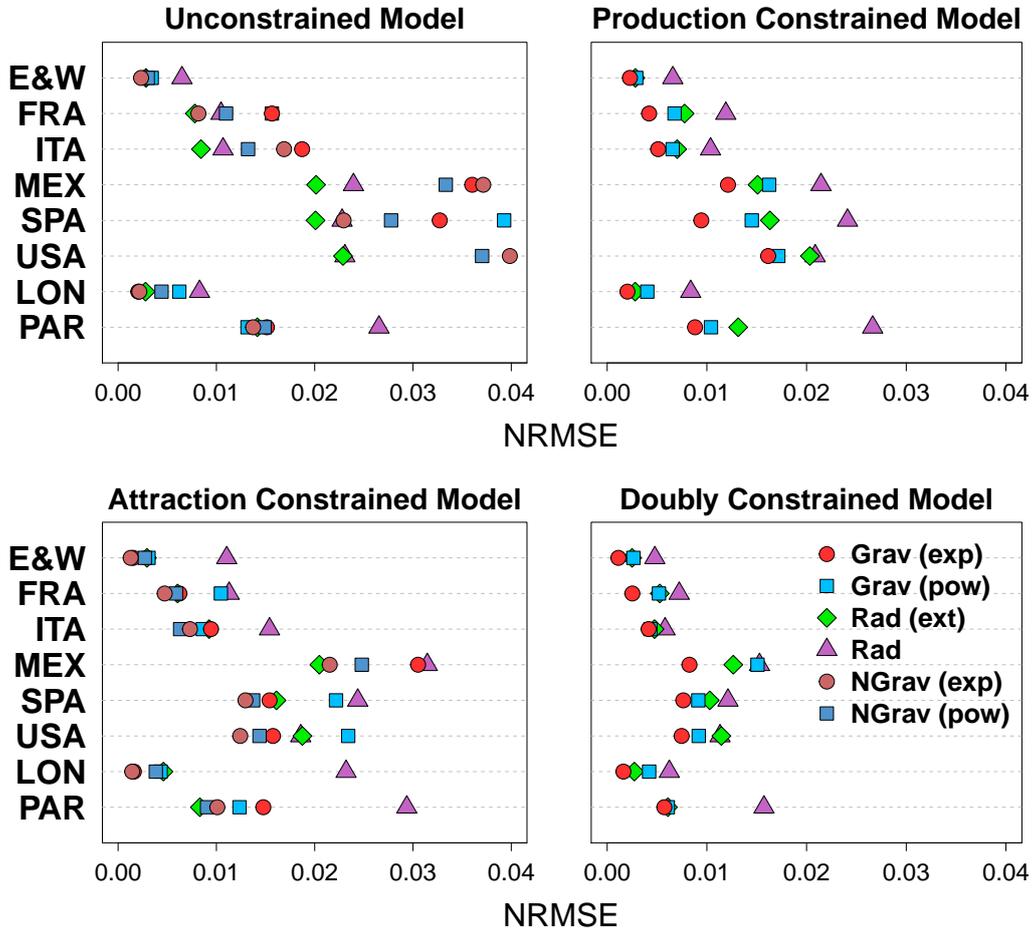


Figure S1: Normalized root mean square error according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Information gain statistic

This statistic measures the difference between two probability distributions, thus, lower values indicate more accurate goodness of fit.

$$I(T, \tilde{T}) = \sum_{i,j=1}^n \frac{T_{ij}}{N} \ln \left(\frac{T_{ij}}{\tilde{T}_{ij}} \right) \quad (23)$$

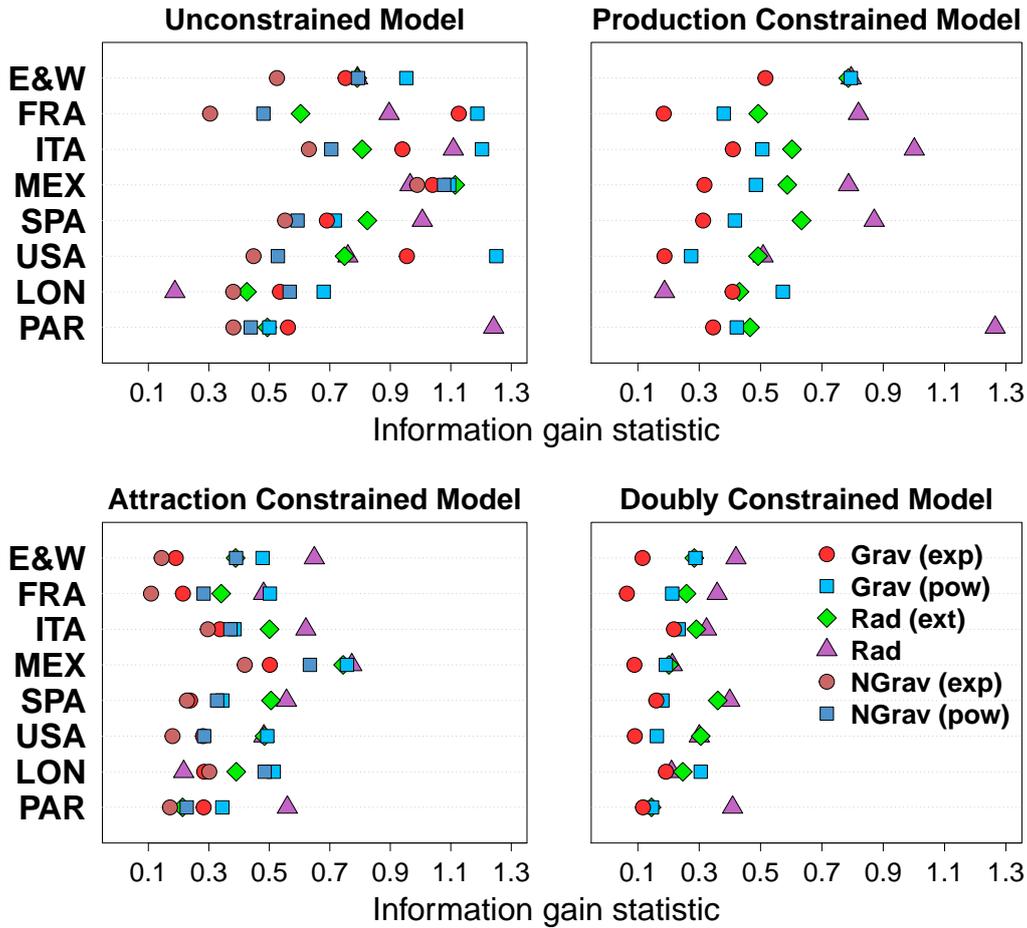


Figure S2: Information gain statistic according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Average common part of out- and in-commuters

To better understand how the CPC is distributed at a more granular level, it is possible to compute the CPC by unit with the common part of out-commuters (Equation 24) and the common part of in-commuters (Equation 25) that we can average over the n units.

$$CPC_i = \frac{2 \sum_{j=1}^n \min(T_{ij}, \tilde{T}_{ij})}{O_i + \tilde{O}_i} \quad (24)$$

$$CPC_j = \frac{2 \sum_{i=1}^n \min(T_{ij}, \tilde{T}_{ij})}{D_j + \tilde{D}_j} \quad (25)$$

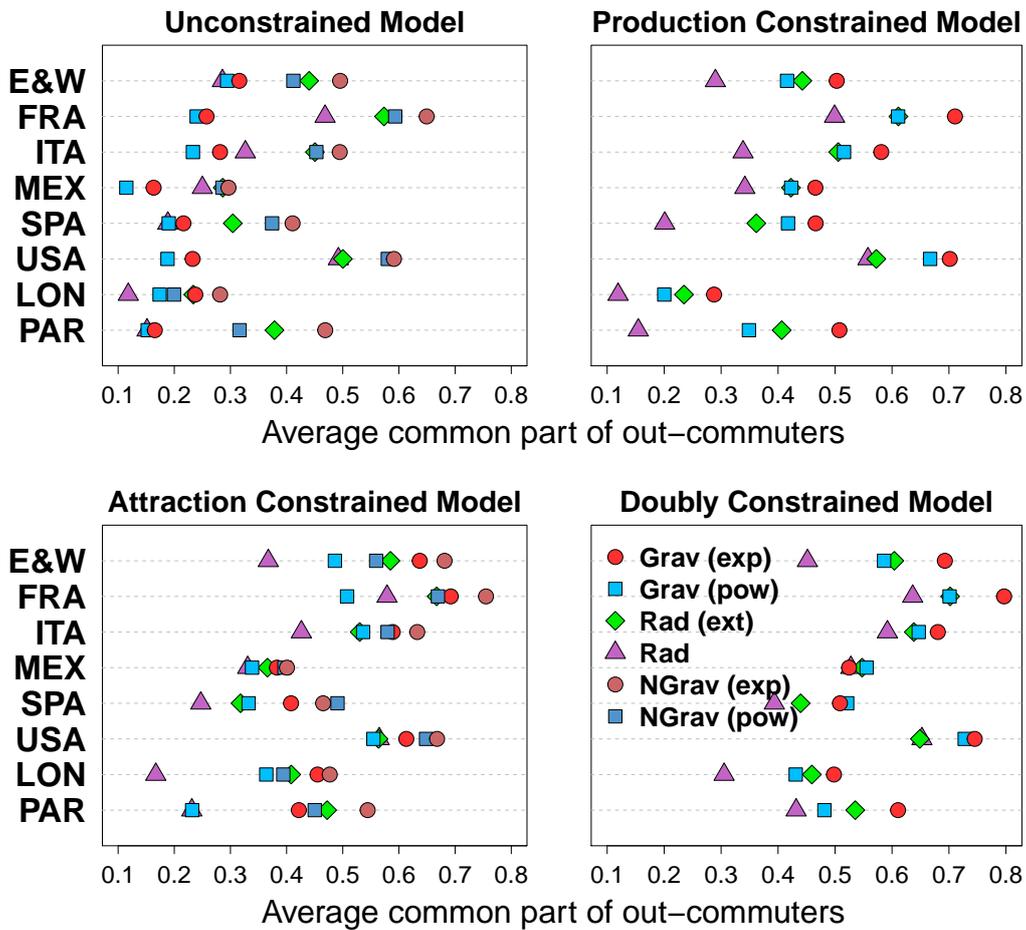


Figure S3: Average common part of out-commuters according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

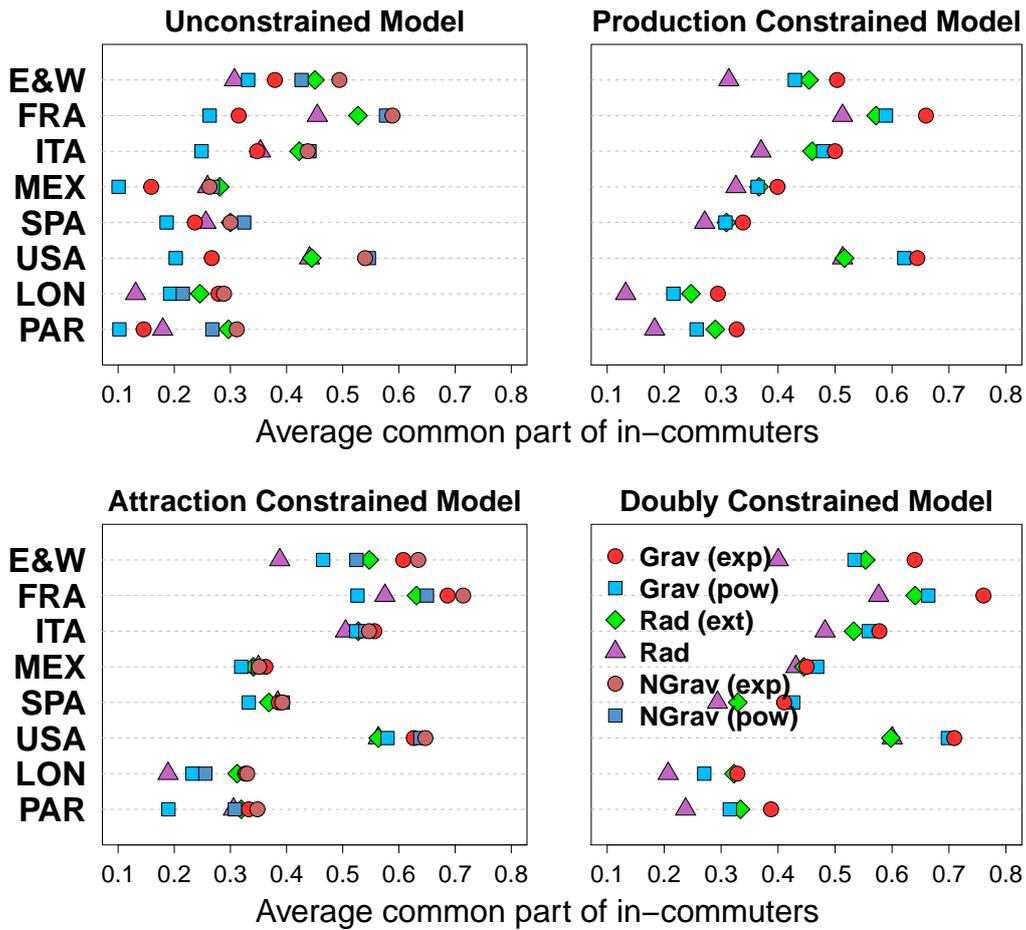


Figure S4: Average common part of in-commuters according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Common part of links

$$CPL(T, \tilde{T}) = \frac{2 \sum_{i,j=1}^n \mathbb{1}_{T_{ij}>0} \cdot \mathbb{1}_{\tilde{T}_{ij}>0}}{\sum_{i,j=1}^n \mathbb{1}_{T_{ij}>0} + \sum_{i,j=1}^n \mathbb{1}_{\tilde{T}_{ij}>0}} \quad (26)$$

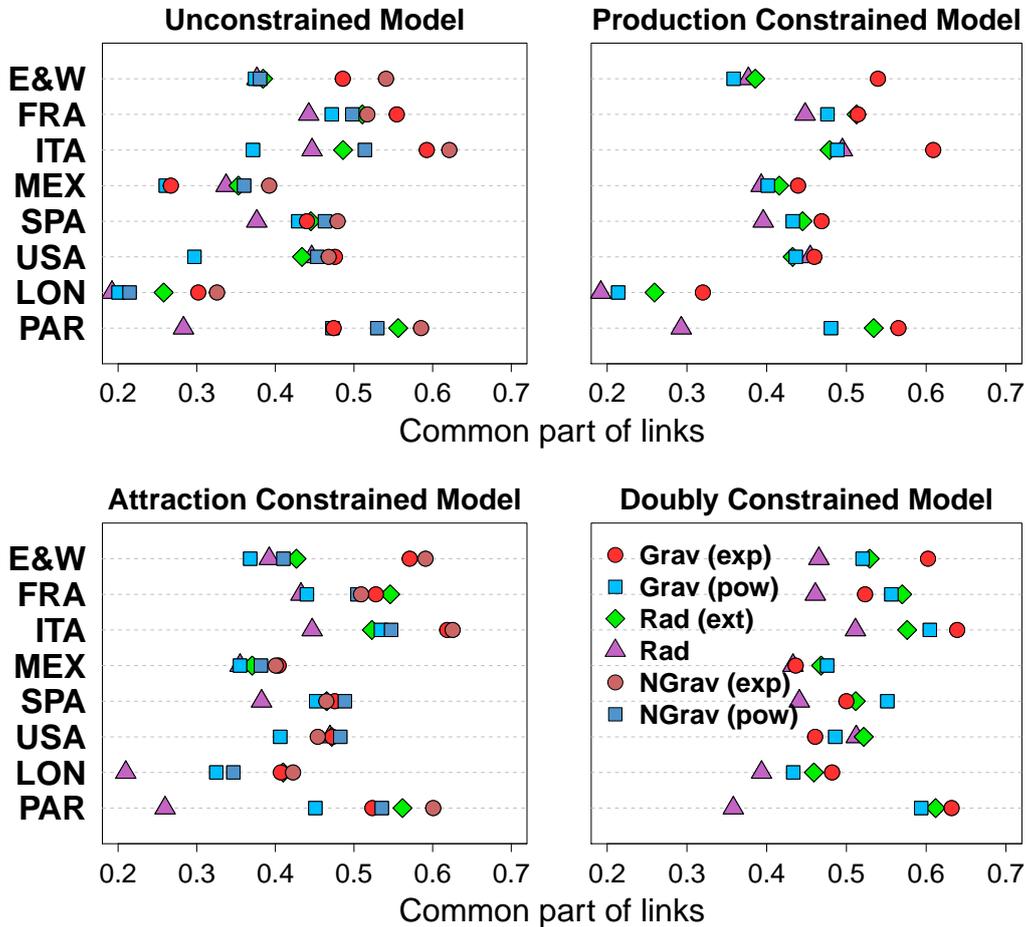


Figure S5: Common part of links according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Ratio between the simulated and the observed average degree

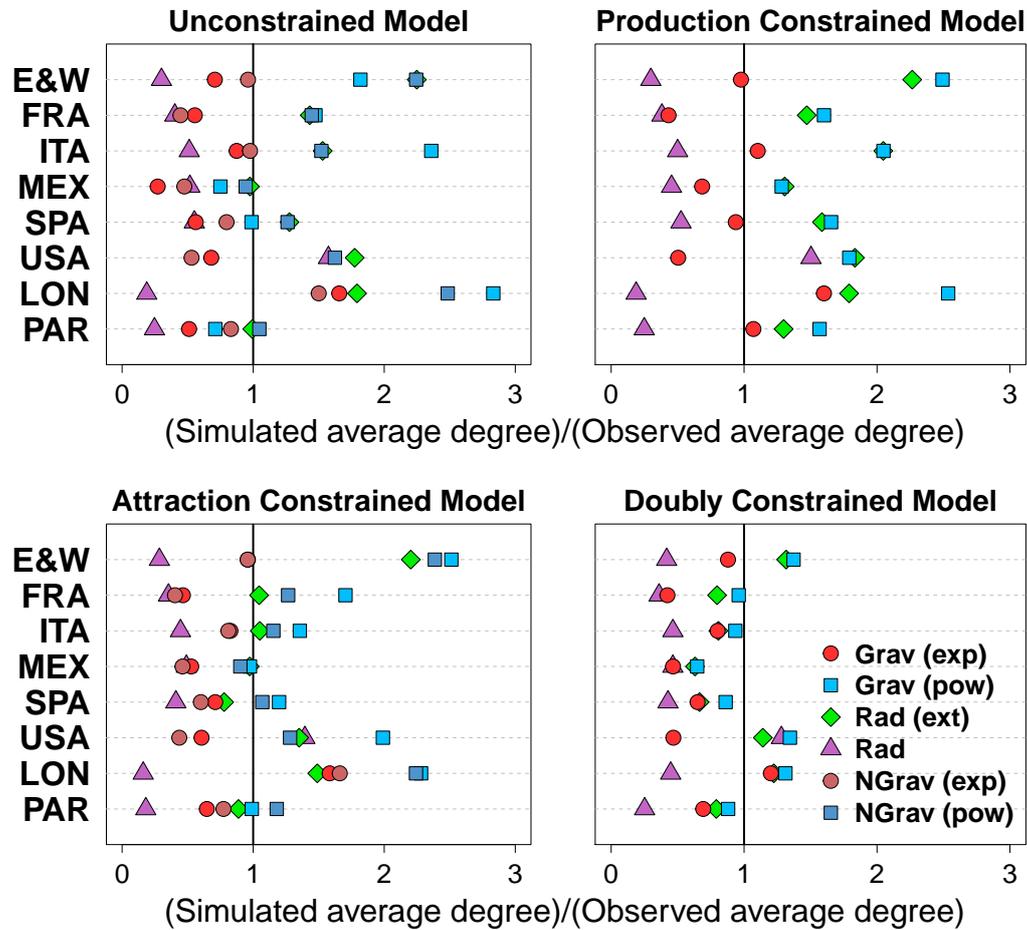


Figure S6: Ratio between the simulated and the observed average degree. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Common part of commuters according to the distance

$$CPC_d(T, \tilde{T}) = \frac{\sum_{k=1}^{\infty} \min(N_k, \tilde{N}_k)}{N} \quad (27)$$

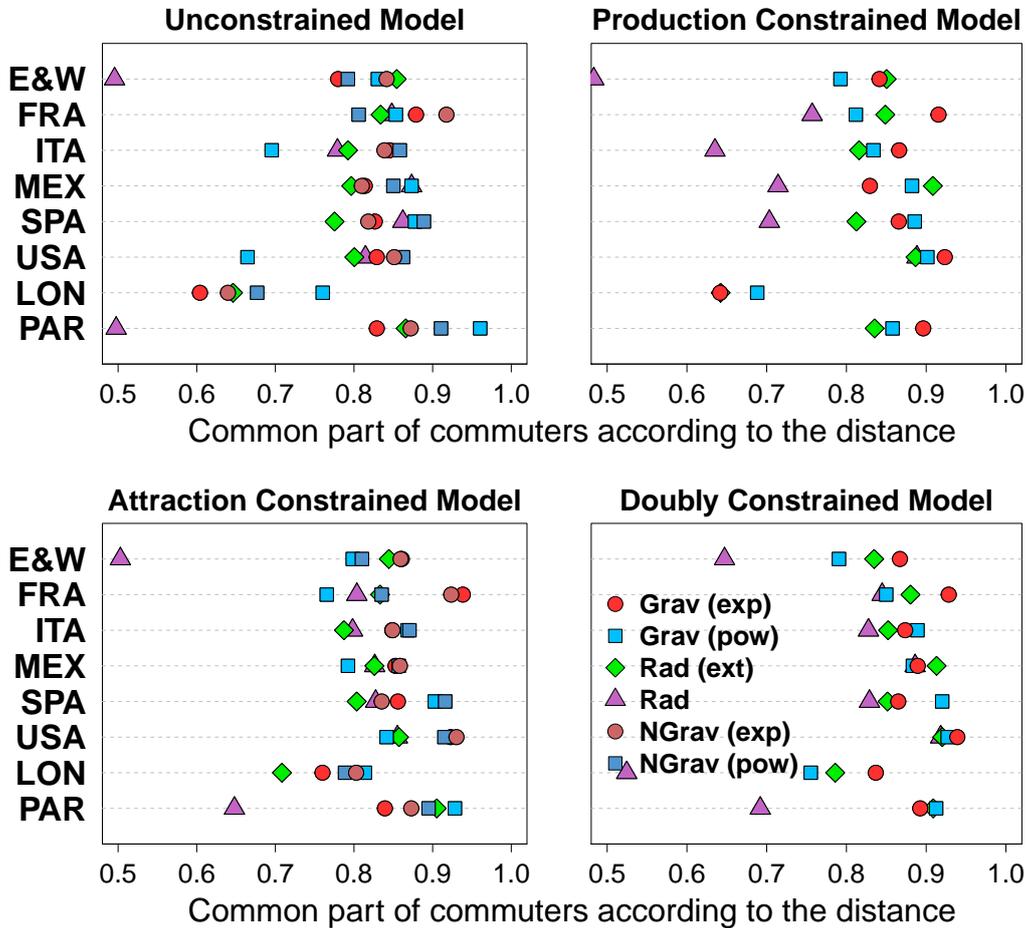


Figure S7: Common part of commuters according to the distance according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

Flows of commuters in Madrid

The metropolitan area of Madrid contains a population of 5,512,495 inhabitants (2009) within an area of 1,935 km^2 . The metropolitan area is divided into a regular grid of square cells of lateral size 1 km. The dataset comprehends geolocated tweets of 50,272 in Madrid in the time period going from September 2012 to December 2013. These users were selected because it was detected from the general data streaming with the Twitter API that they have emitted at least a geolocated tweet from one of the two cities. Later, as a way to increase the quality of our database, a specific search over their most recent tweets was carried out [37]. The identification of the OD commuting matrices using Twitter is the one explained in [38]. We first filter out the Twitter users for whom we do not enough information to extract the place of work and residence (less than 100 tweets on weekdays in all the dataset). Then, the users' home and work are identified as the cell most frequently visited on weekdays by each user between 8 pm and 7 am (home) and between 9 am and 5 pm (work).

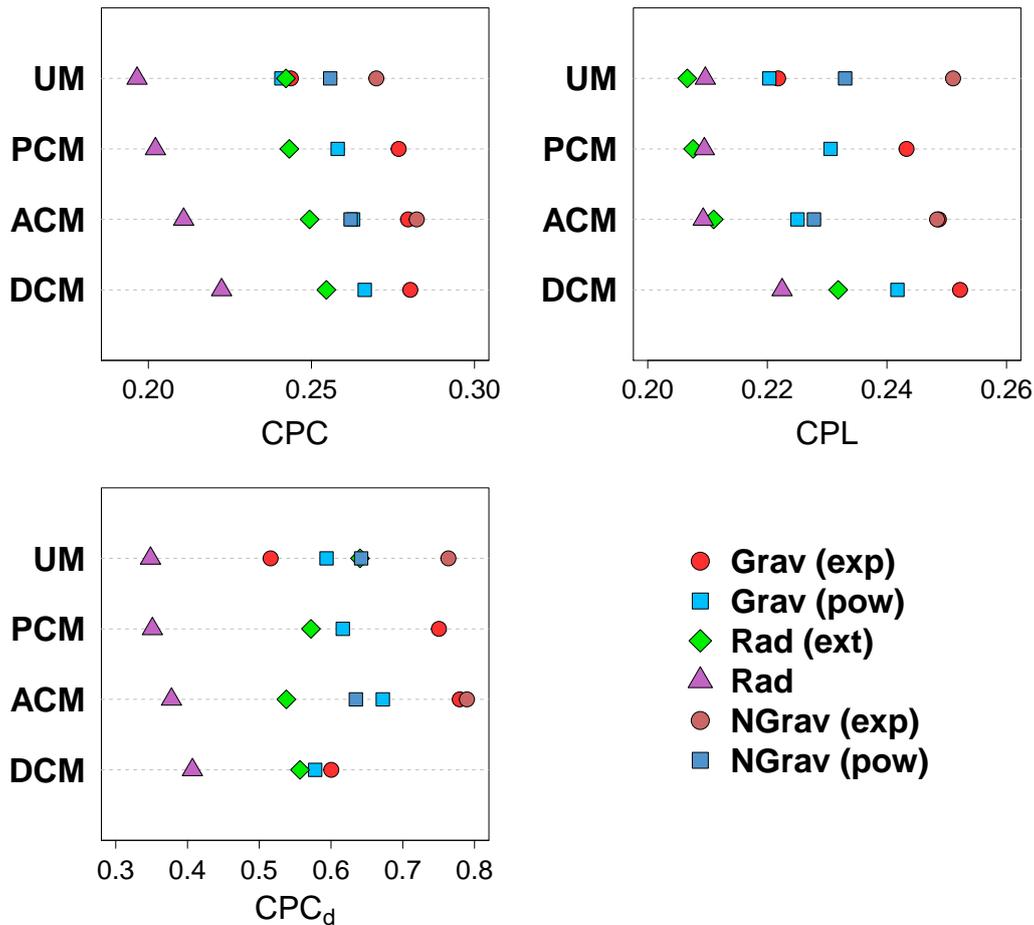


Figure S8: Common part of commuters according to the unconstrained models, the gravity and radiation laws. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

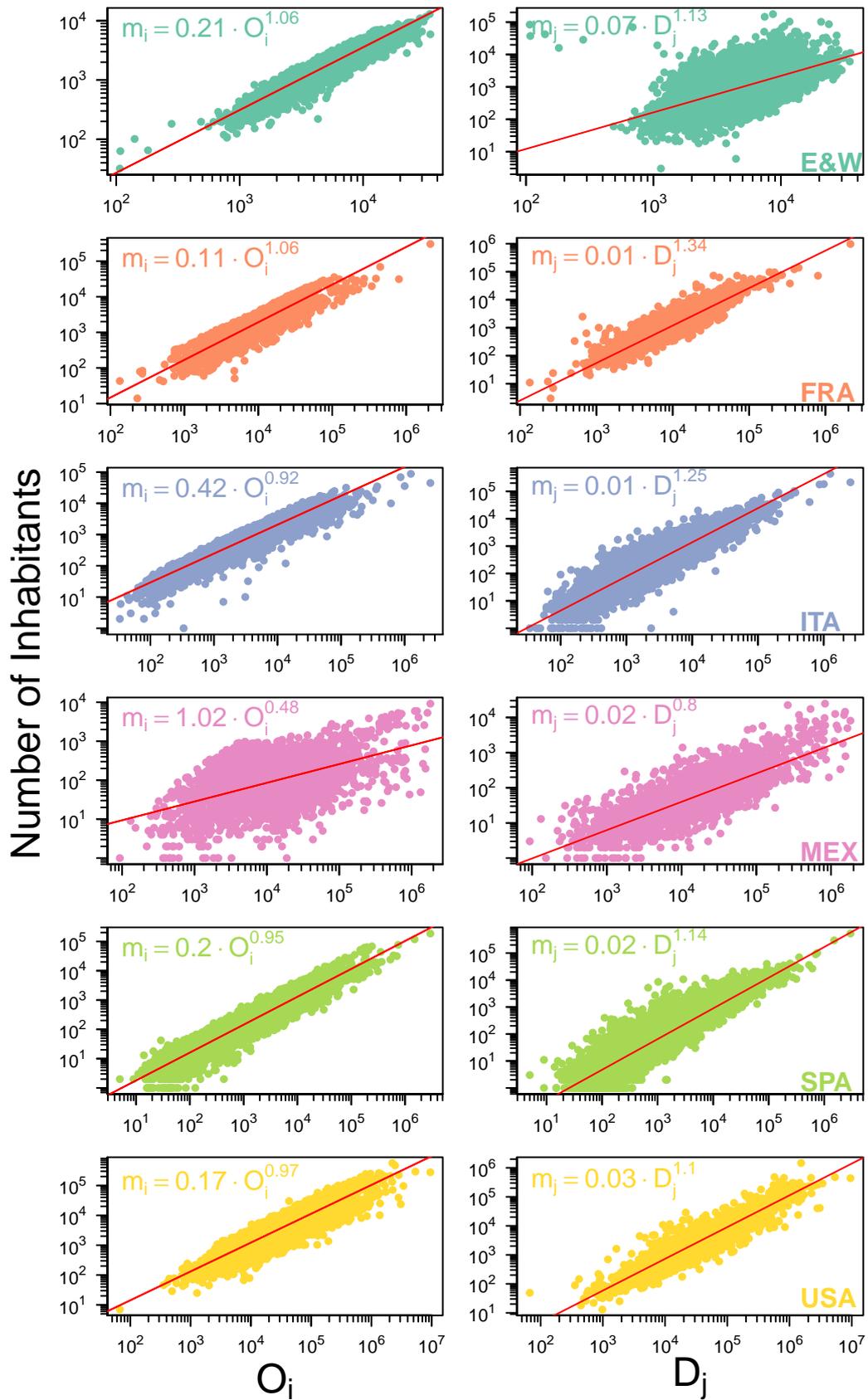


Figure S9: Number of out-commuters O_i (left) and in-commuters D_i (right) as a function of the number of inhabitants for each of the eight case studies. From the top to the bottom: England and Wales, France, Italy, Mexico, Spain and United States. The red line represents the regression line.

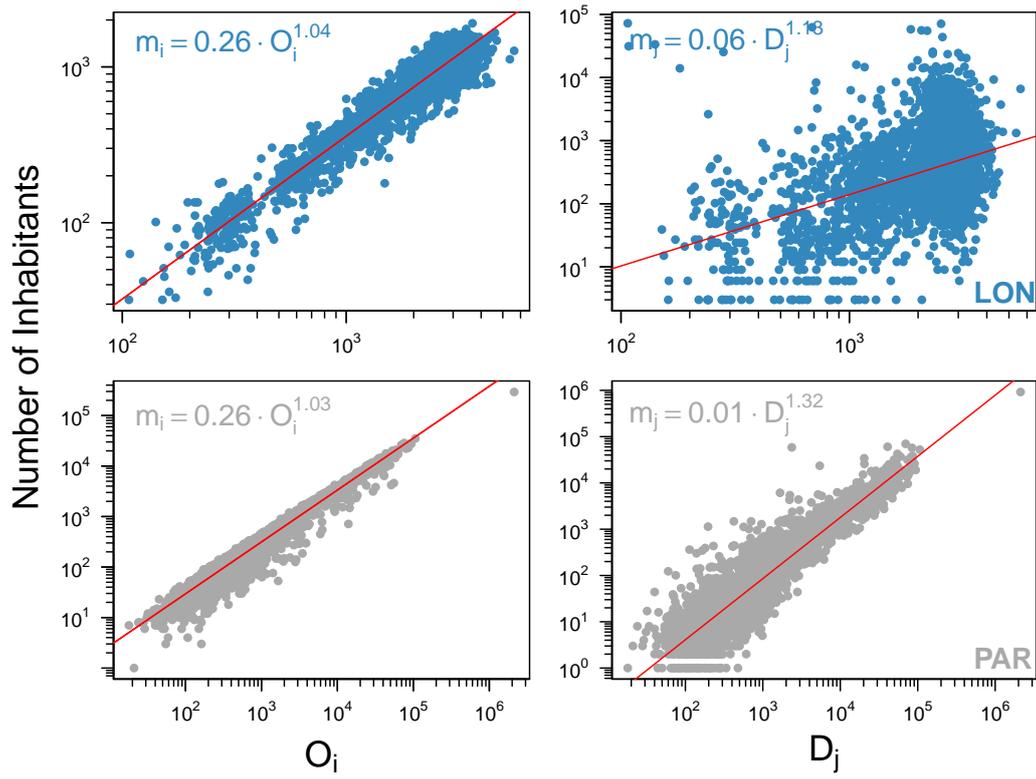


Figure S10: Number of out-commuters O_i (left) and in-commuters D_i (right) as a function of the number of inhabitants for each of the London (top) and Paris (bottom). The red line represents the regression line.

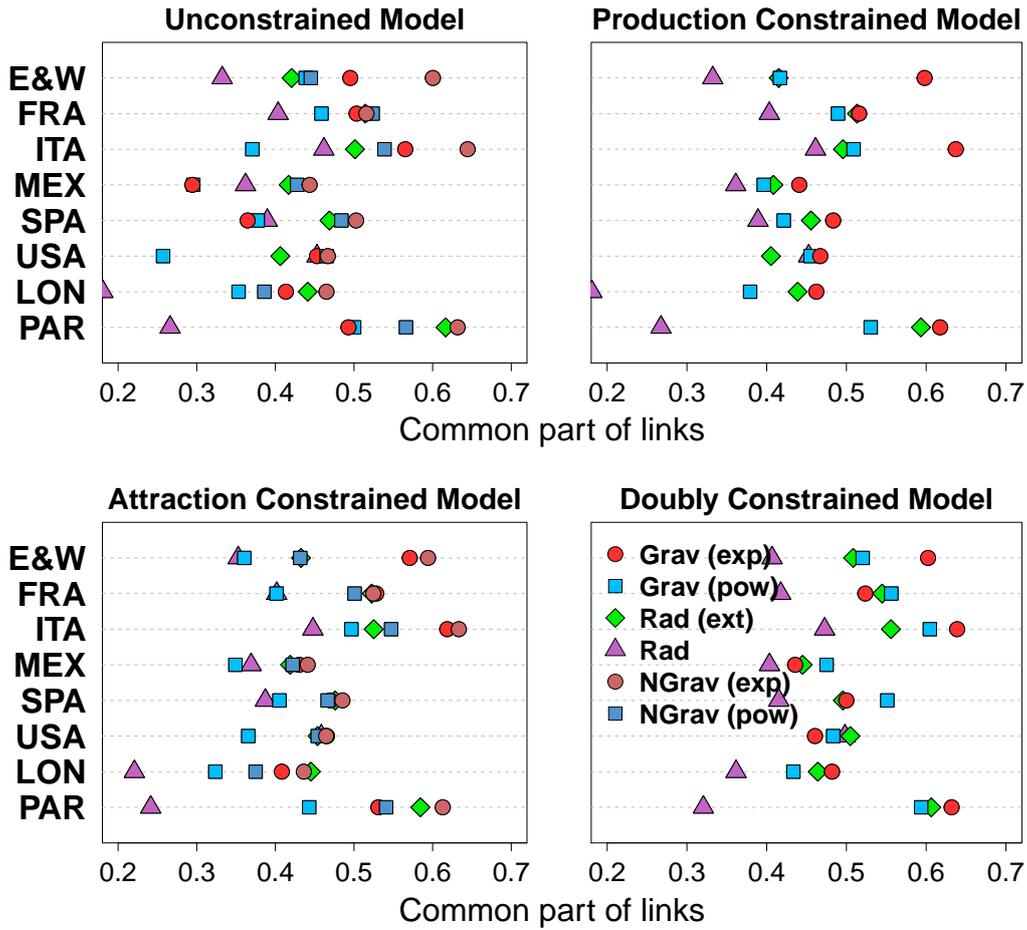


Figure S12: Common part of links according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

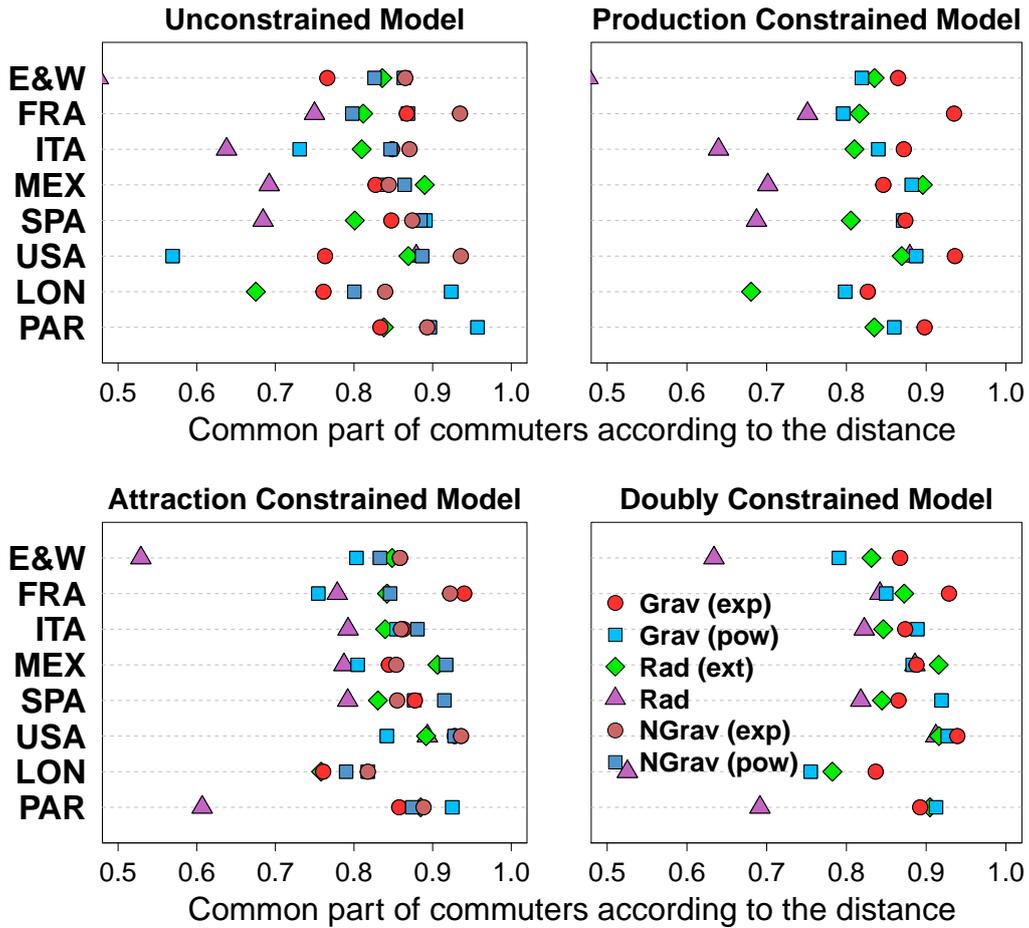


Figure S13: Common part of commuters according to the distance according to the unconstrained models, the gravity and radiation laws for the eight case studies. The circles represent the gravity law with the exponential distance decay function (in red the original version and in dark red the normalized one); The squares represent the gravity law with the power distance decay function (in cyan the original version and in blue the normalized one); The green diamonds represent the extended radiation law; The purple triangles represent the original radiation law. Error bars represent the minimum and the maximum but in most cases they are too close to the average to be seen.

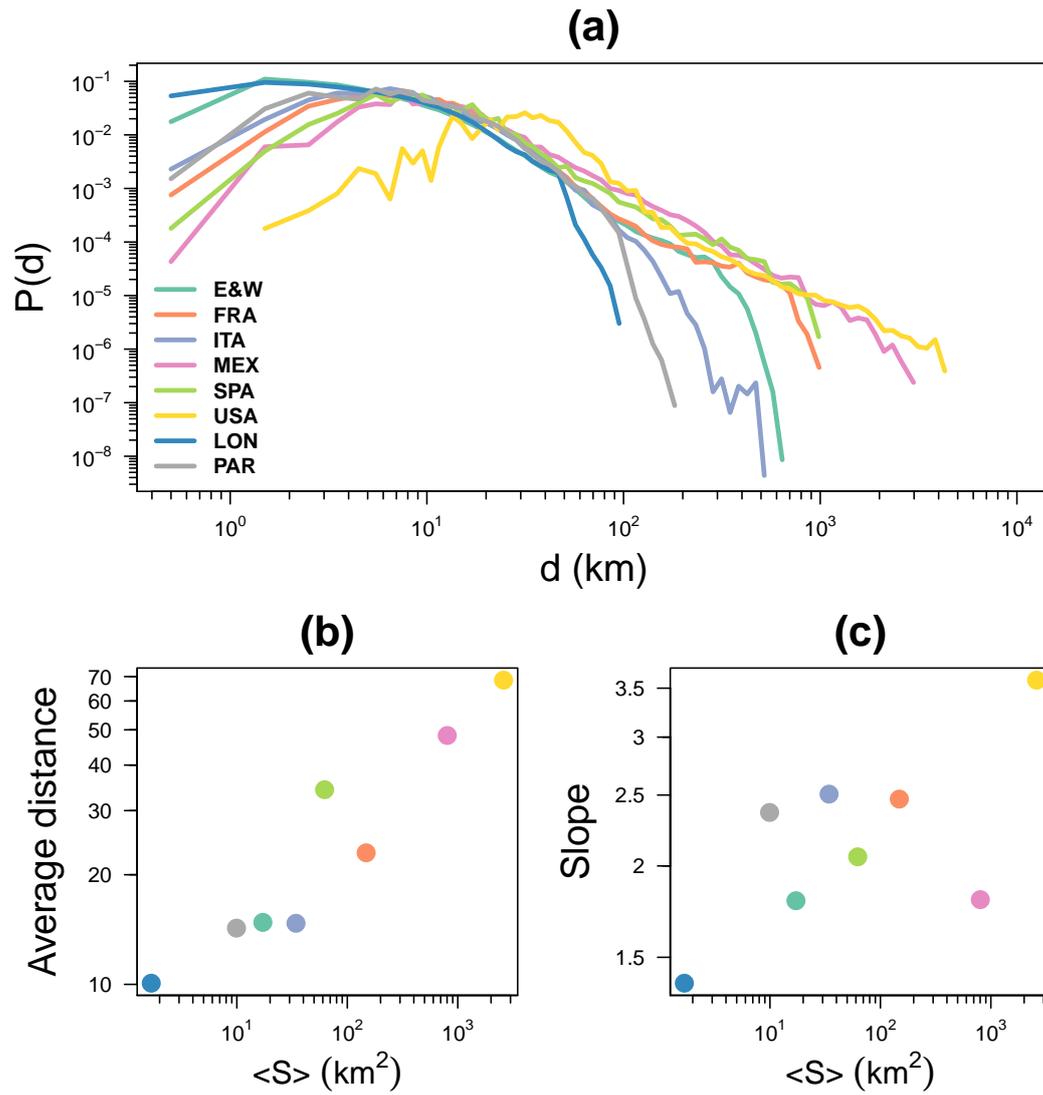


Figure S14: Observed commuting distance distributions. (a) Commuting distance distribution according to the case study. (b) Average commuting distance as a function of the average unit surface. (c) Slope of the tail according to the average unit surface.

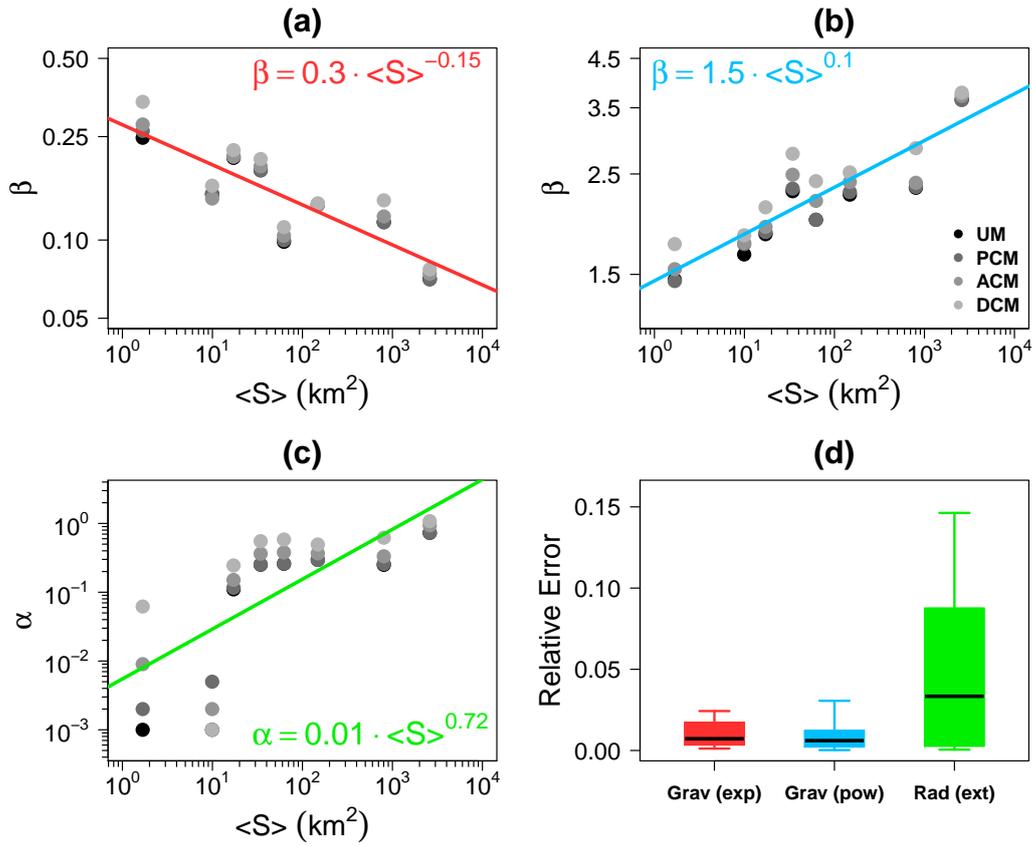


Figure S15: Parameter value as a function of the average unit surface. (a) Gravity laws with an exponential distance decay function. (b) Gravity laws with a power distance decay function. (c) Extended radiation law. (d) Boxplots of the relative error between the CPC obtained with a calibrated value of the parameters and the CPC obtained with values estimated with the regression models. The boxplot is composed of the first decile, the lower hinge, the median, the upper hinge and the last decile.