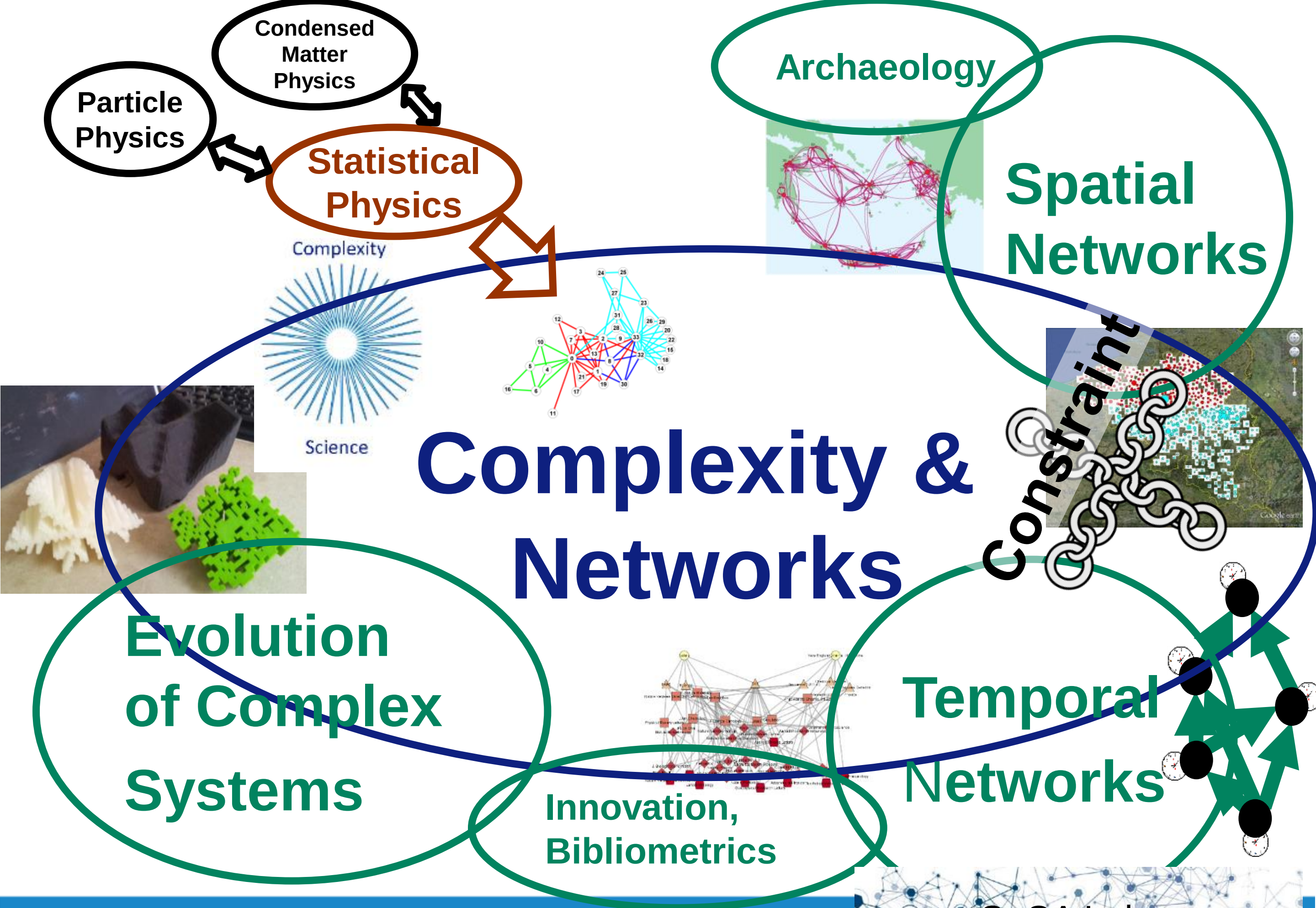
Imperial College London

The MacArthur Workshop on Urban Modelling & Complexity Science, UCL
20/10/2019

Complexity Science

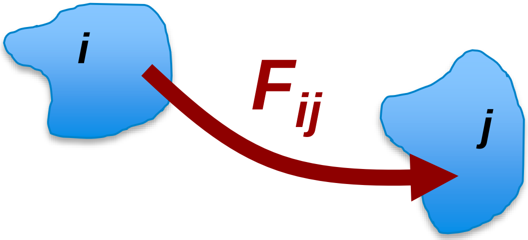
Properties of Simple Models of Spatial Interaction

Tim Evans



Condensed Matter Physics
Particle Physics
Statistical Physics
Complexity
Science
Archaeology
Spatial Networks
Constraint
Complexity & Networks
Evolution of Complex Systems
Innovation, Bibliometrics
Temporal Networks

SIM = Spatial Interaction Models



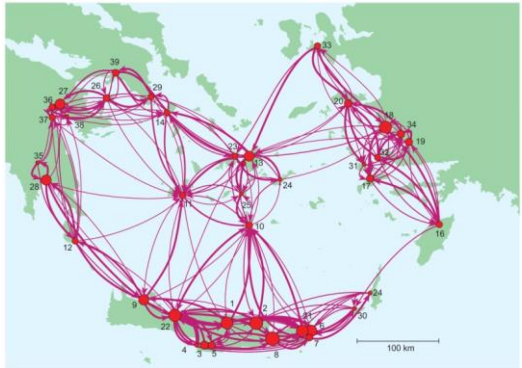
Model flows between nodes of different sizes embedded in space

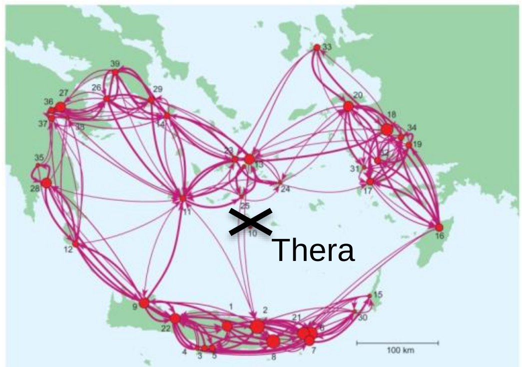
What did Simple Spatial Interaction Models ever do for us?

Look at general properties

HOW DOES SPACE INFLUENCE THE SYSTEM?

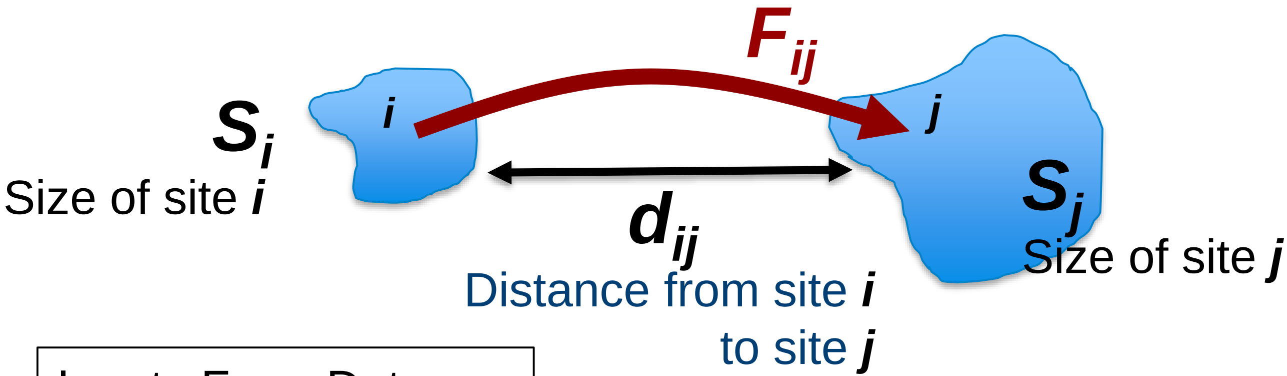
- Test Basic Principles
  - e.g. is site attractiveness a non-linear function of size? [Rihll & Wilson 1979, 81; Bettencourt at al. 2007]
  - e.g. CO_2 [Verbavatz & Barthelemy 2019]
  - Error propagation
- Comparisons, Null Models
  - e.g. Spatial Clustering [Expert at al. 2011]
  - “What if” studies
- General Predictions
  - e.g. destruction of Thera only weakens Aegean networks





[Knappett, TSE, Rivers, Antiquity 2012]

Simple Models of Spatial Interaction



Inputs From Data

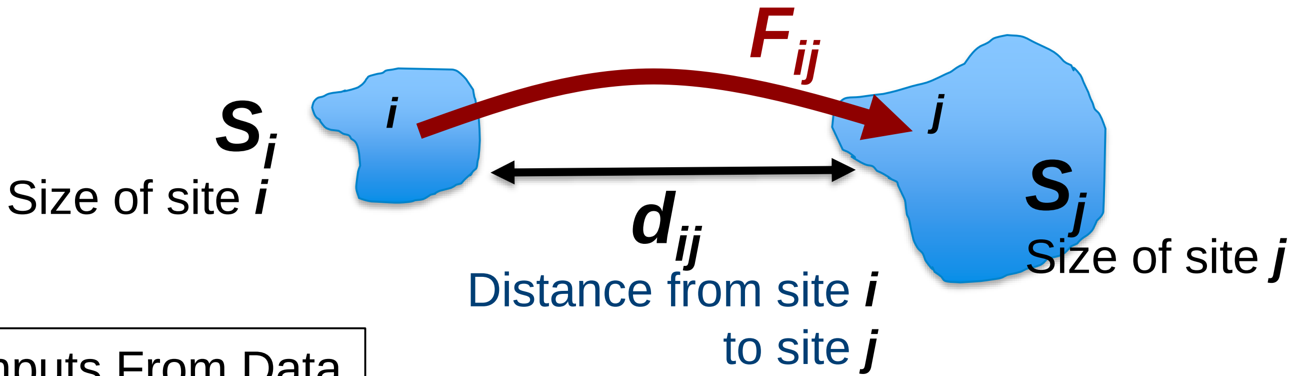
- Site sizes S_i
  - P_i Population
  - O_i Output
  - I_i Input
- Distances d_{ij}

Results

- Flows between sites F_{ij}

Tim Evans, Imperial College

Simple Models of Spatial Interaction



Inputs From Data

- Site sizes S_{ij}
  - P_i Population
  - O_i Output
  - I_i Input
- Distances d_{ij}

Modelling Choices

- Site parameters t_i, m_i, n_i
- Global Parameters α, β, …

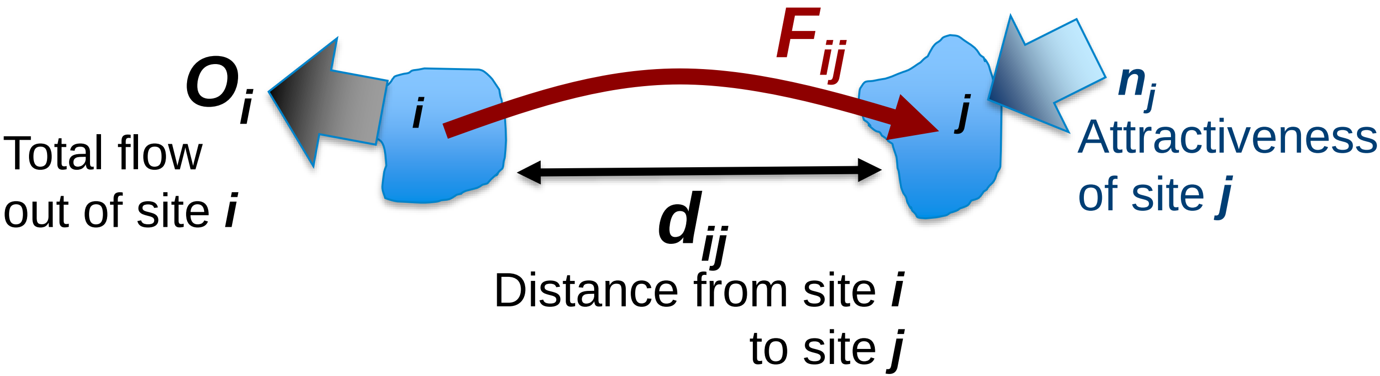
Results

- Flows between sites F_{ij}

Site Parameters: Data and Model

Data Values	Model Parameters
P_i Total Population	s_i generic size parameter
O_i Output e.g. # commuters leaving	t_i repulsive parameter $\Rightarrow$ controls output
I_i Input e.g. # commuters arriving	n_i attractiveness parameter $\Rightarrow$ controls input
	m_i aspiration parameter $\Rightarrow$ controls range

Production Constrained Gravity Model



$$F_{ij} = \frac{O_i \, n_j \, \exp(-\beta d_{ij})}{\underbrace{\sum_k n_k \, \exp(-\beta d_{ij})}_{\text{Normalisation}}}$$

Global Parameter β

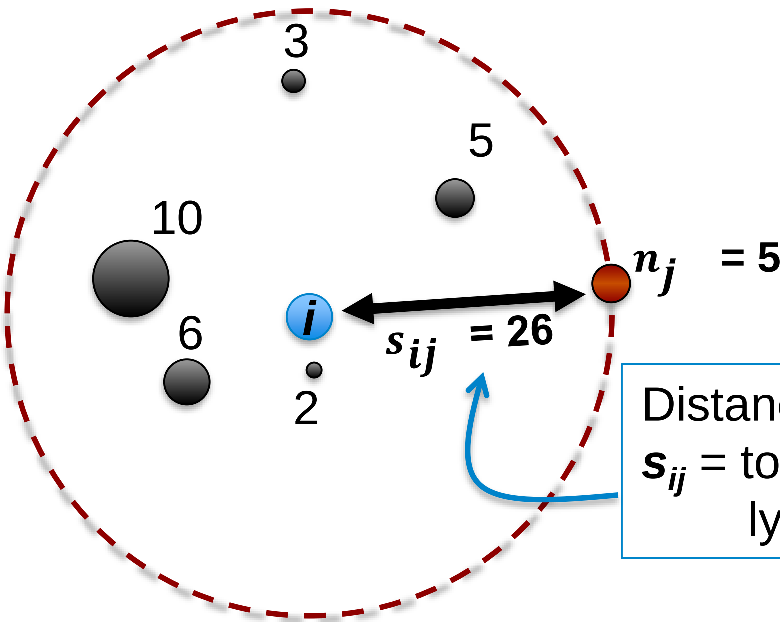
F_{ij} = Flow from site *i* to site *j*

Normalisation
Guarantees total flow out = O_i

Radiation Model [Simini et al, 2012]

- Uses **Intervening Opportunities** measure of distance

[Stouffer, 1940]



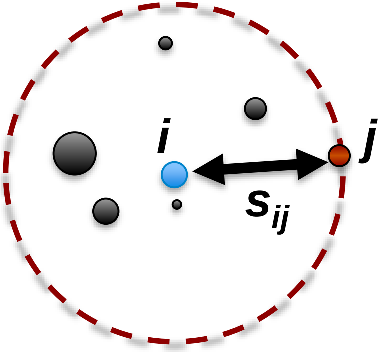
Distance from ***i*** and ***j***
$\boldsymbol{s_{ij}}$ = total opportunities lying between ***i*** and ***j***

Stouffer' Intervening Opportunities Model

[Stouffer, 1940]

For comparison with Radiation model, Stouffer suggested using the Intervening Opportunities measure of distance as follows

$$F_{ij} = t_i \, n_j \frac{1}{(s_{ij} + n_j)}$$



"… the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities." [p.846 Stouffer, 1940]

Radiation Model [Simini et al, 2012]

Interpretation

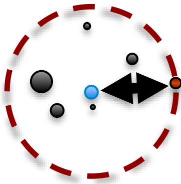
- t_i commuters leave site i to search for job
- Salary aspiration for each site set by m_i parameter
  - largest of m_i values from distribution $p(z)$
- Closest opportunities first
- Each of n_j opportunities at site j makes an offer
  - Offer z with probability $p(z)$
- Accept first offer meeting aspiration

= O_i Output of site i

?

NOT equal to I_j input of site j
BUT n_j controls input

Radiation Model



Intervening Opportunities + Record Statistics =

Production Constrained

t_i = total flow out of site ***i*** in model

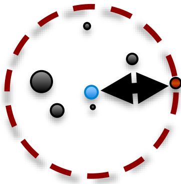
n_i controls input but not equal to input

$$F_{ij} = t_i \, n_j \underbrace{\frac{m_j}{(m_j + s_{ij})(m_j + s_{ij} + n_j)}}$$

F_{ij} = Flow from site ***i*** to site ***j***

Falls off on scale m_i
"aspiration" of commuters

Radiation Model Parameters



Intervening Opportunities + Record Statistics =

Three model site parameters:
t_i = total flow out of site i in model
n_i = controls input
m_i = controls range

Parameter-data map can introduce **global parameters**

Site Data Values

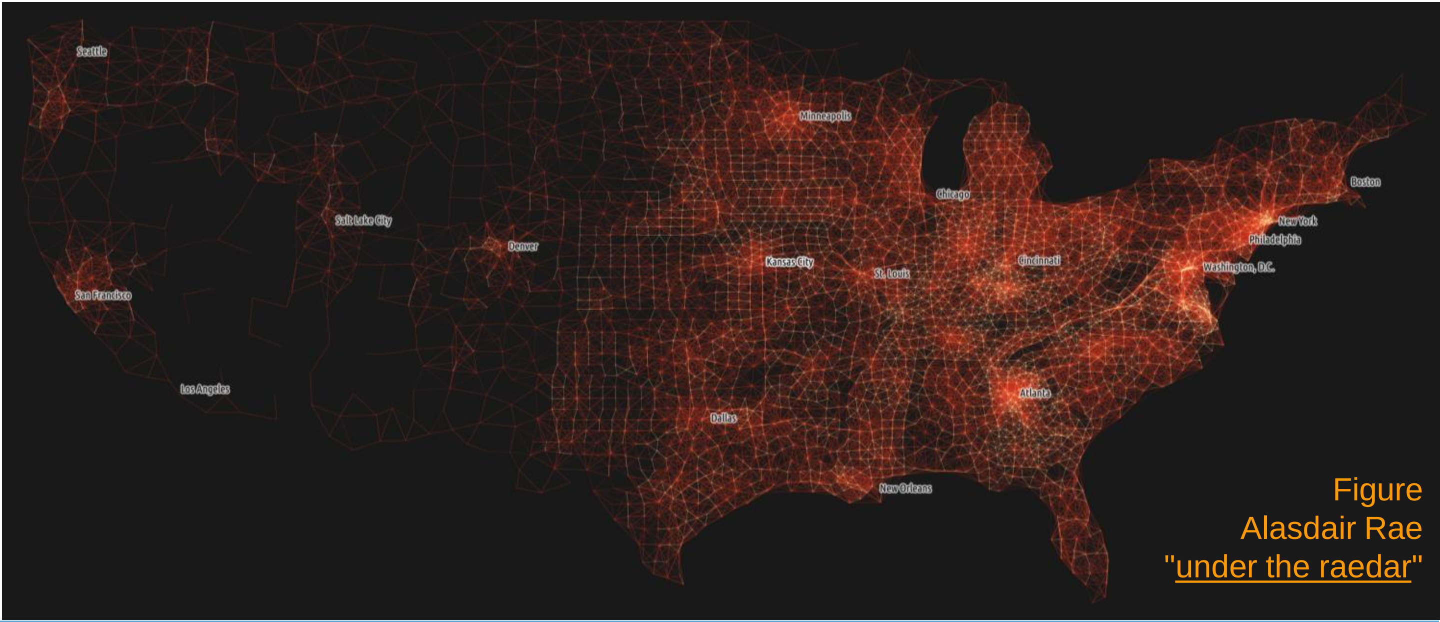
$$F_{ij} = \underbrace{t_i \, n_j \frac{m_j}{(m_j + s_{ij})(m_j + s_{ij} + n_j)}}$$

F_{ij} = Flow from site ***i*** to site ***j***

Fixed form,
no **global** model parameters

Matching Data and Spatial Interaction Models

Hilton, Sood, TSE, http://arXiv.org/abs/1909.07194
"Predictive limitations of spatial interaction models: a non-Gaussian analysis"



How to test Spatial Interaction Models

MODEL

F_{ij} = Flow
from site i
to site j

How do we compare data and model?

- What are appropriate statistics?
- What is a fair comparison?
- What are the best models?

Hilton, Sood & TSE, http://arXiv.org/abs/1909.07194

Statistics

- Sørensen-Dice coefficient
  [Gargiulo et al ’12; Lenormand et al ’12, ‘16; Masucci et al ‘13, Yang et al ‘14]

  Lacks a rigorous statistical basis.

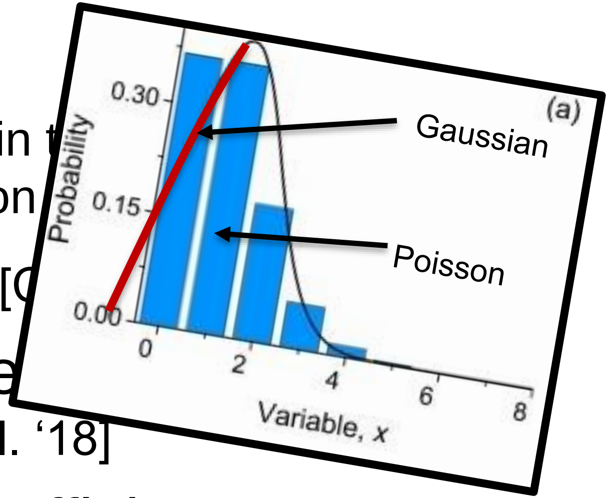
- The Kolmogorov-Smirnov test
  [Kang et al ‘15]

  Requires the two input functions to be independent, so invalid when estimate model parameters from data [Steinskog et al ‘07]

Gaussian Statistics

Statistics assuming the variation in t
pair follows a Gaussian distribution

- Mean Squared Errors [C
- The Coefficient of Dete
  [Masucci et al. ‘13, Curiel et al. ‘18]
- Pearson correlation coefficients [Liu et al. ‘15].



Problem:

- real data sets have no negative flows
- Data has a high proportion of very small flows

⇒ flows between each pair of sites cannot be assumed to be Gaussian

Suggested Protocol

Poisson Regression

- Assume models give average flow
- Assume fluctuations around mean are Poisson distributed
  - e.g. Radiation model has Binomial fluctuations, a good approximation to Poisson for large flows $\boldsymbol{t_i}$
- Look at log-likelihood $\boldsymbol{L}$
  - Measure of probability that data produced by model

$$\ln\big(L(F_{\min})\big) = \sum_{(i,j)} \Big(-\hat{F}_{ij} + F_{ij}\ln\big(\hat{F}_{ij}\big) - \ln\big(F_{ij}!\big)\Big)\,\Theta(F_{ij} - F_{\min})$$

Model value

Data value

Analysis Cutoff

Statistical Measures

- LOG LIKELIHOOD ***L***
  - Measure of probability that data produced by model
  - Does not compensate for number of parameters

- BAYESIAN INFORMATION CRITERION (BIC)
  - Harsh penalty for increasing number of parameters

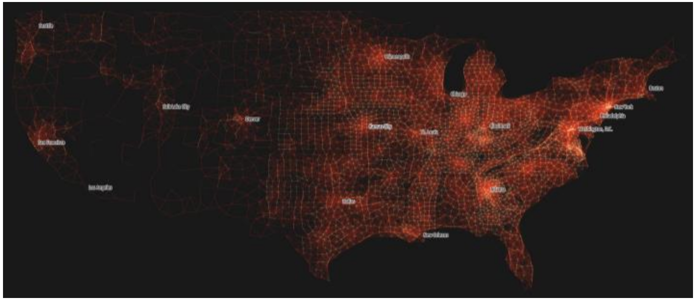
$$\mathrm{BIC}(F_{\min}) = k \ln(n(F_{\min})) - 2\ln\left(L(F_{\min})\right)$$

k = # parameters
n= # data points

- DEVIANCE $D = -2\ln\left({}^{L_s}/_{L}\right)$ where $\boldsymbol{L_s}$ is the saturated log likelihood, value of ***L*** if predictions matched data.

$$\mathrm{D}(F_{\min}) = 2\sum_{(i,j)}\left(F_{ij} - \hat{F}_{ij} + F_{ij}\ln\left(F_{ij}/\hat{F}_{ij}\right)\right)\Theta(F_{ij} - F_{\min})$$

Modern Commuting Data



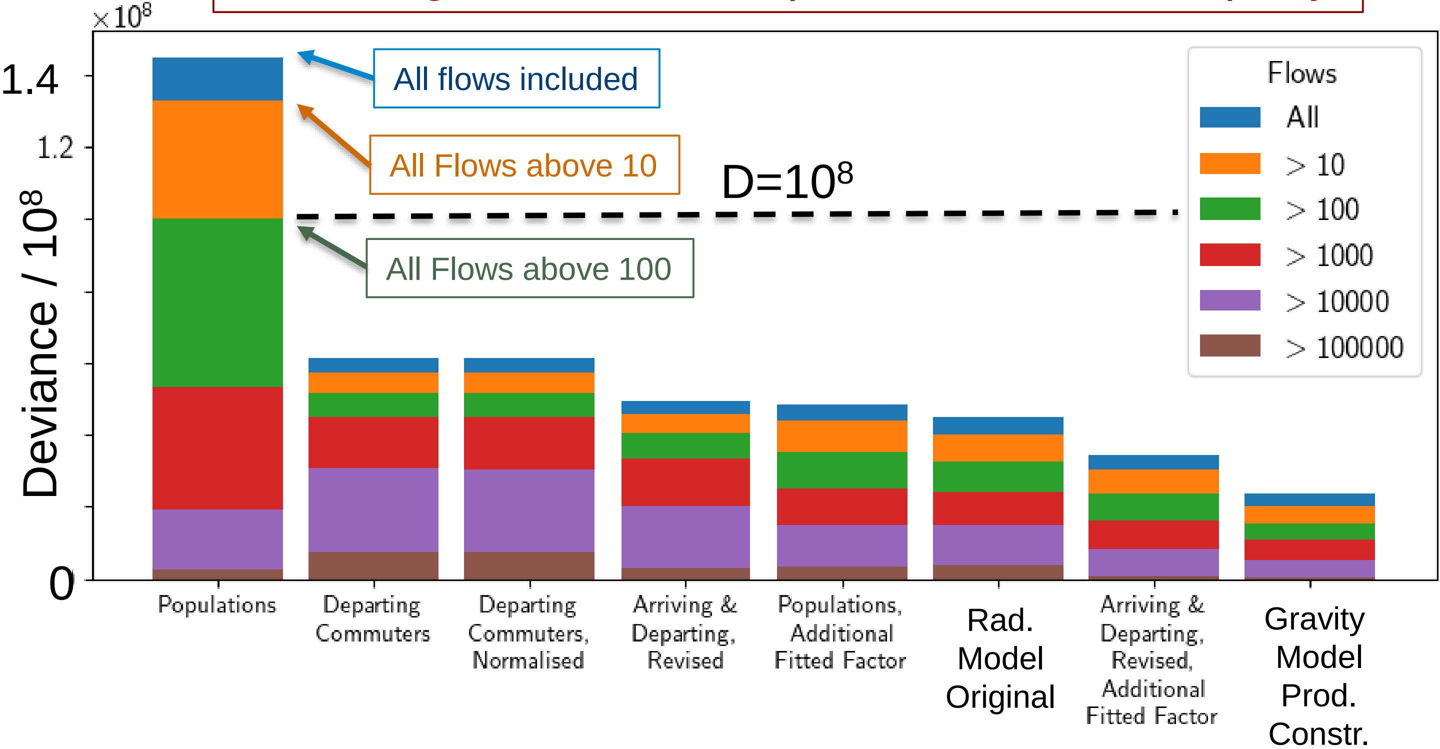
Flow	Number
All	9,665,881
>0	164,764
>10	77,432
>100	21,237
>1,000	7,058
>10,000	1,814
>100,000	212

- US Census 2000 asked "at what location did this person work last week?"
- 3109 counties within the 48 contiguous States
- 98.3% of county pairs have no flow

https://www.census.gov/data/tables/2000/dec/county-to-county-worker-flow-files.html

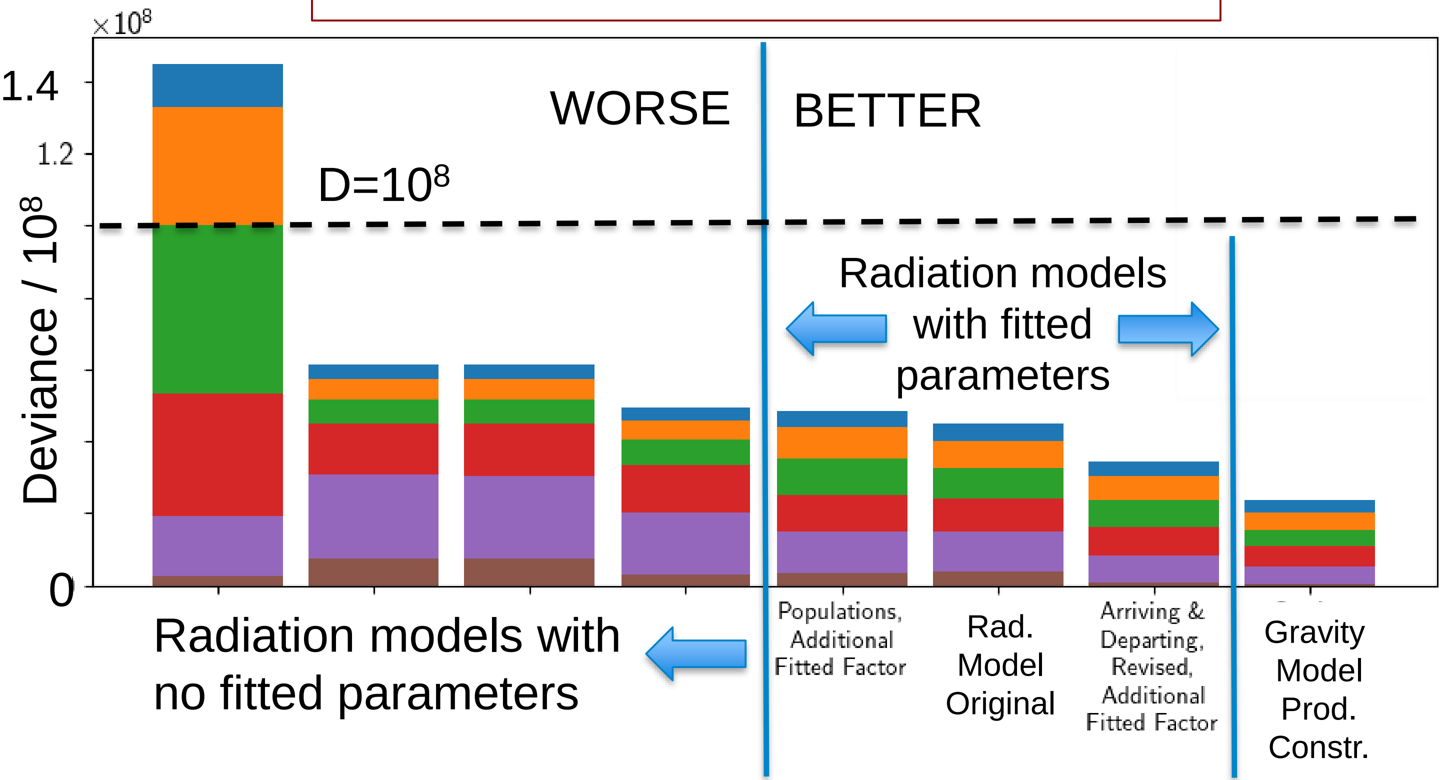
Deviance Results: US Census 2000

Removing Small Flows Improves All Models Equally



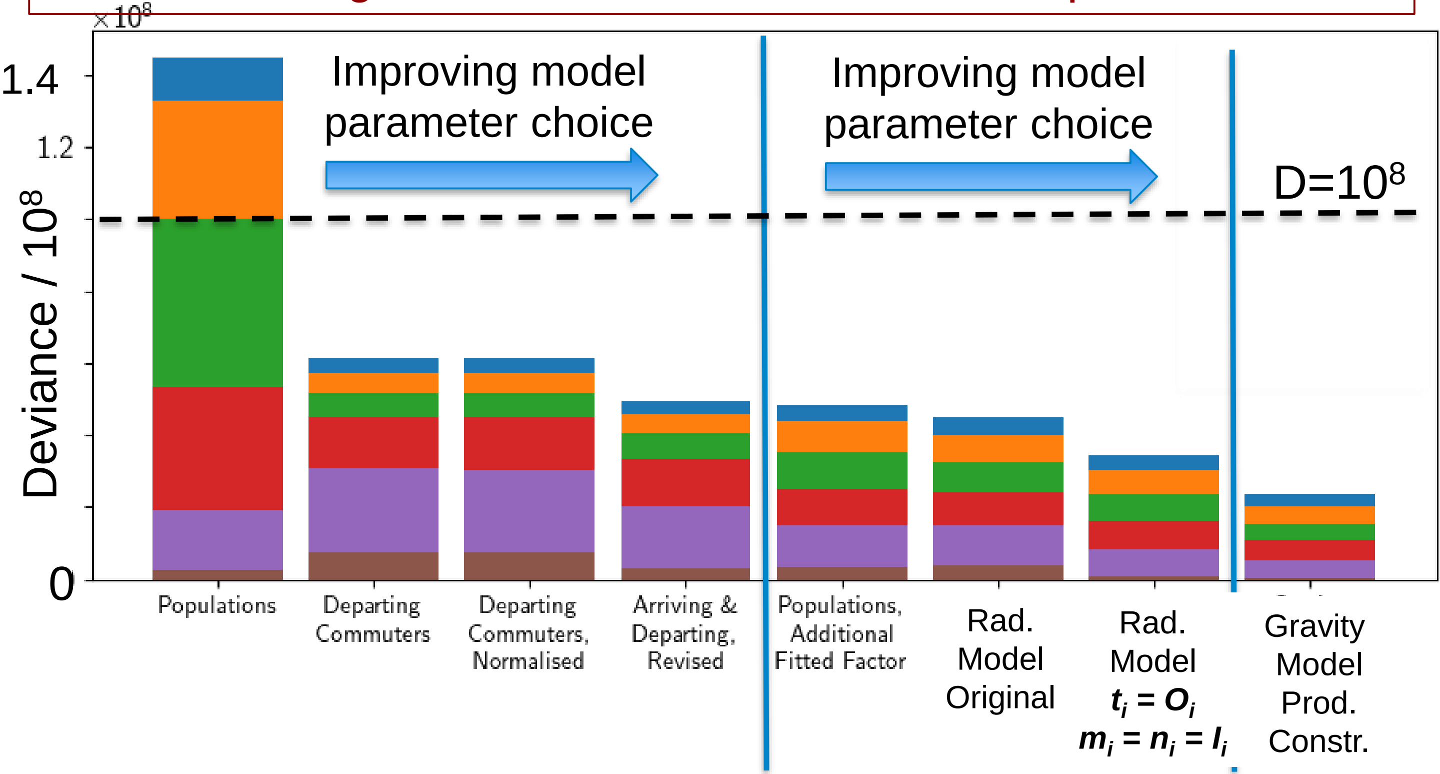
Deviance Results: US Census 2000

Models with Fitted Parameters are better



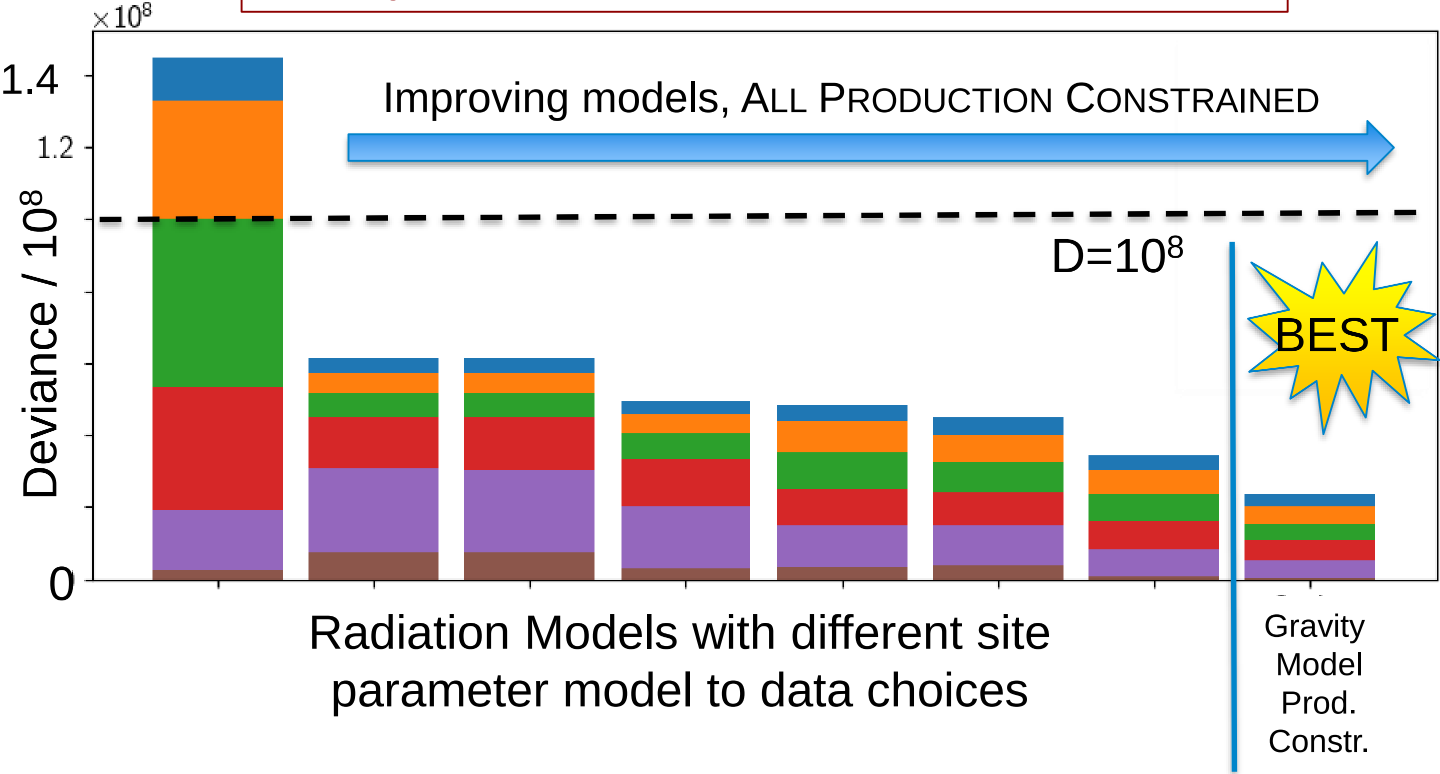
Deviance Results: US Census 2000

Good Matching of Model to Data Parameters Improves Results



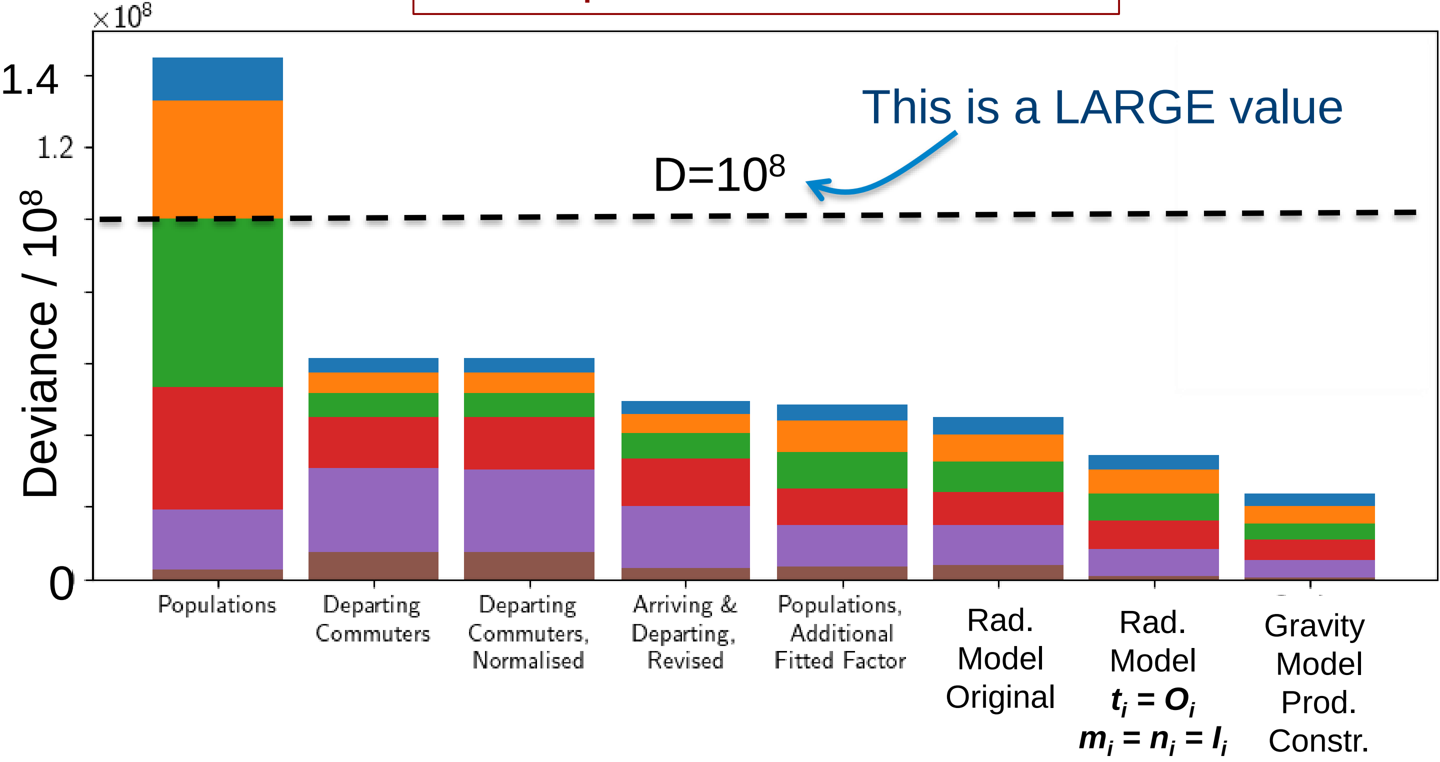
Deviance Results: US Census 2000

Gravity Model Better Than All Radiation Models



Deviance Results: US Census 2000

All Simple Models are Rubbish



[Hilton, Sood & TSE, http://arXiv.org/abs/1909.07194]

Summary

- Removing small flows left results unchanged
- Found all three statistics give same results
  - So many data points
- Using global fitted parameters improves results
  - So many data points, no penalty to add more parameters to simple models
- Match model parameters to data to improve results
  - Output parameters $t_i = O_i$
  - Input parameters $n_i = I_i$
  - Aspiration Parameter $m_i = n_i$ best (set by axioms)
- Gravity model is significantly better than Radiation model
  - Important to compare models with similar constraints
- All simple models give poor fits to real data
  - This is not why we use a simple spatial interaction model

[Also draws on work of TSE with Bamis & Gastner; Bamis MSc thesis 2014]

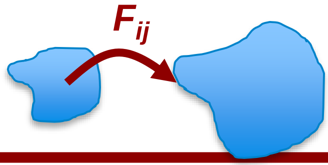
Properties of Spatial Interaction Models

What properties should a Spatial Interaction Model have?

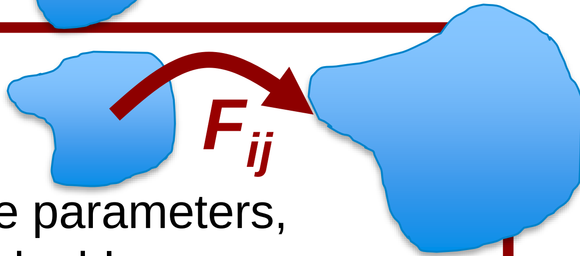
Understanding this can help

- Classify models
- Help us choose an appropriate model for a given problem
- Improve model parameter choices

Flow Scaling Axiom



FLOW SCALING AXIOM



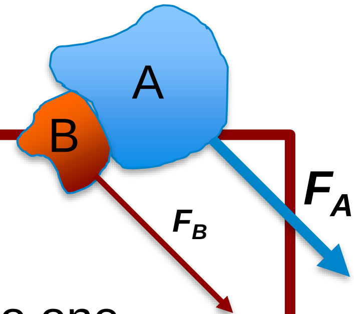
If we double all the site size parameters,
the flows should all double

$$t_i \to \lambda t_i, n_i \to \lambda n_i \quad \Rightarrow \quad F_{ij} \to \lambda F_{ij}$$

- Ensures only relative sizes matter
  - Useful in archaeological context
  - Units don't matter
- Simple (unconstrained) gravity model fails this
- Why enforce linearity? [Bettencourt et al 2007; Arcaute et al 2014]

Coarse Graining Axioms

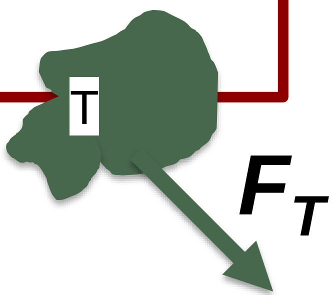
T



OUTPUT COARSE GRAINING AXIOM

If we combine two neighbouring sites into one,
the flows remain unchanged

$$t_T = t_A + t_B \quad \Rightarrow \quad F_T = F_A + F_B,$$



- Ensures spatial units don't matter
  - Should we use wards or boroughs in UK Census?

Radiation Model & Output Coarse Graining

To ensure output flows are consistent in Radiation model when you coarse grain requires **`aspiration=opportunities`** $m_i = n_i$

$$F_{ij} = t_i \, n_j \frac{n_j}{(n_j + s_{ij})(n_j + s_{ij} + n_j)}$$

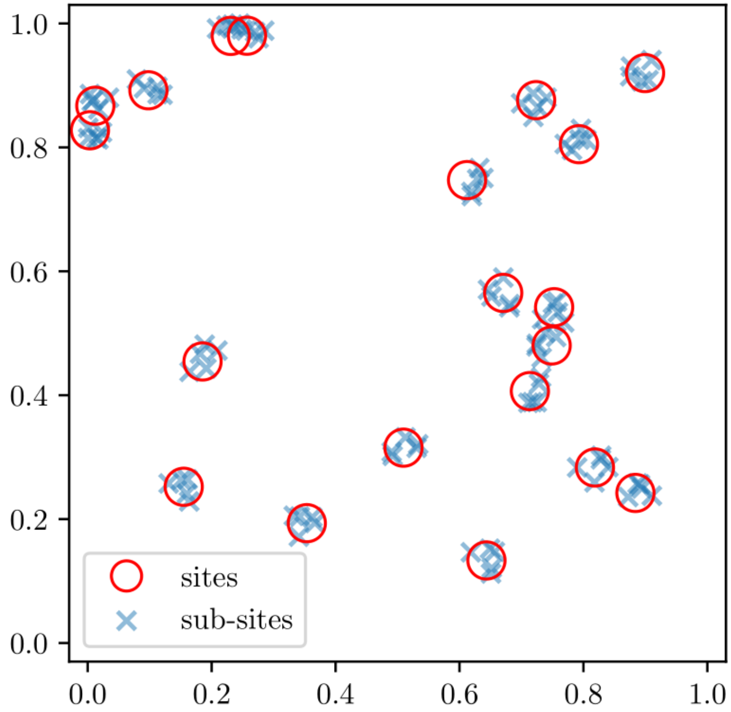
- One less parameter per site
  - Just output $\boldsymbol{t_i}$ and opportunities $\boldsymbol{n_i}$
- More consistent narrative
- Produced better results for US Census 2000 [Hilton, Sood, TSE 2019]

Coarse Graining in Practice

[TSE with Louf 2019]

- Theoretical analysis of coarse graining hard
  only produced simple results

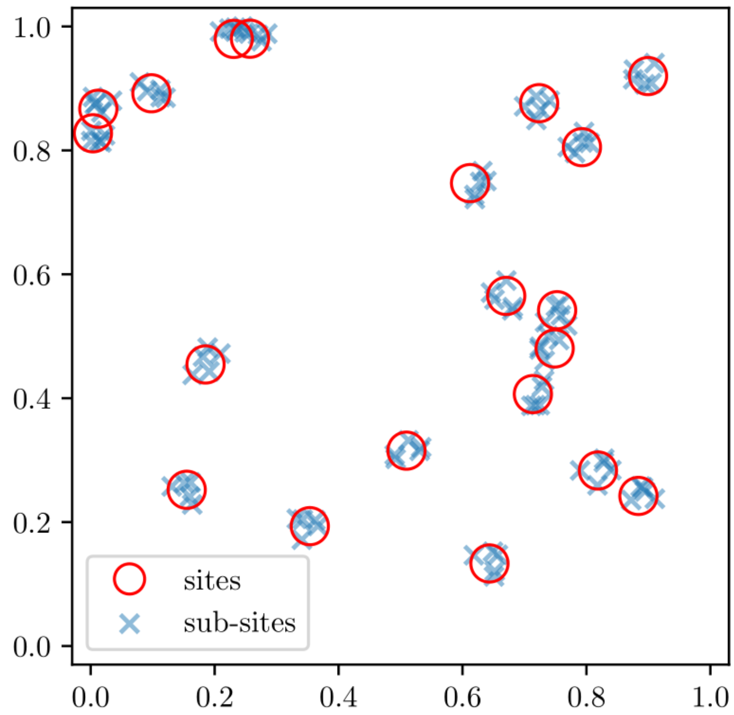
- Need to test effects of coarse graining on more realistic examples



Clustered PPP Model

- PPP=
  Poisson Point Process
  Here 20 points thrown down uniformly at random in box

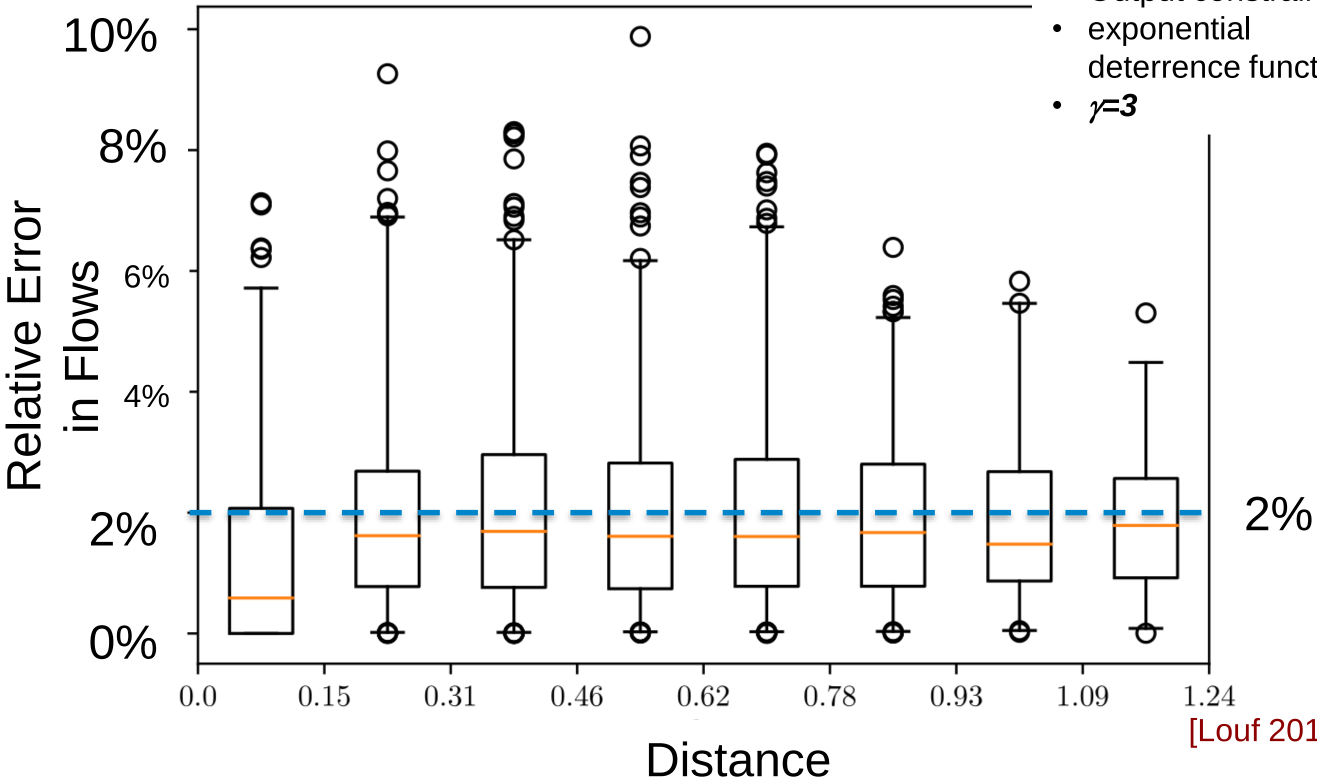
- Clustered
  Each random site has 5 subsites placed within radius ***r***



[Louf 2019]

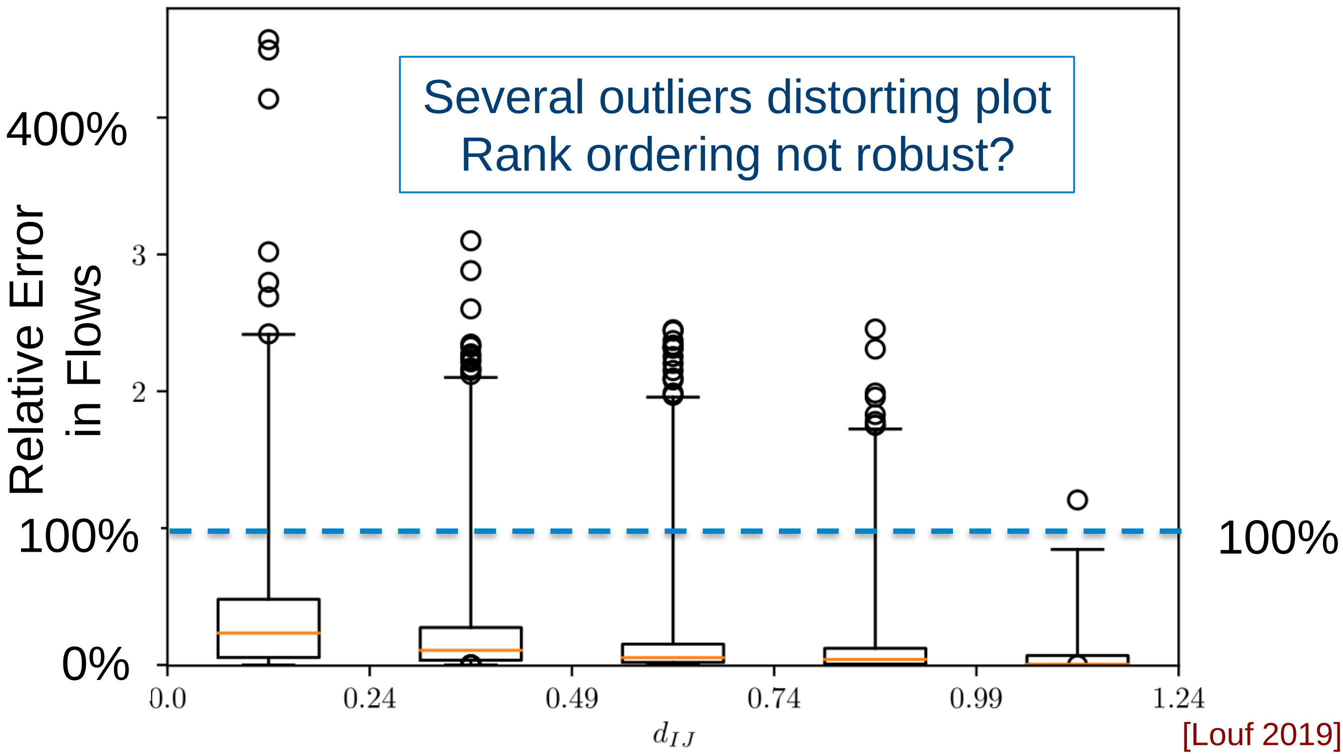
Clustered PPP & Gravity Model

- Output constrained
- exponential deterrence function,
- γ=3



[Louf 2019]

Clustered PPP & Radiation Model



[Louf 2019]

Coarse Graining and Robustness of Simple Models

[TSE with Louf 2019]

- Analytic estimates of effects of coarse graining on predicted flows are possible
- Simple unclustered PPP show larger differences than analytic results
- Adding clustered sites increases differences
- Intervening Opportunities measure more sensitive to coarse graining variations

Summary: Basic Properties

- Use simple properties to test and classify models
- Imposing property can constrain models
  - e.g. $\boldsymbol{m_i = n_i}$ in radiation model
- Models suggest coarse graining, the scale used for settlement units, is important

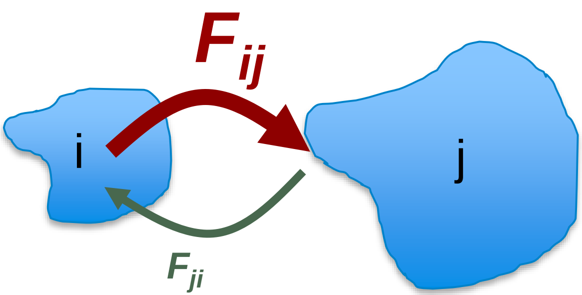
[Louf, MSc Thesis 2019]

[Also draws on work of TSE with Rivers; Bamis & Gastner; Bamis MSc thesis 2014]

Dynamical Spatial Interaction Models

What can we learn if we use flows to drive dynamics?

- Consider permanent movement of people between settlements as predicted by a Spatial Interaction Model
  - not daily commuting, longer time scale
- Difference in flow gives net migration



$$\underbrace{\Delta P_j}_{\text{Net migration}} = \overbrace{F_{ij}}^{\text{Flow in}} - \overbrace{F_{ji}}^{\text{Flow out}}$$

Migration from Simple Model

[Wilkinson, Emms, TSE 2018]

1. At time ***t*** we have the population of each site $P_i(t)$
2. Your chosen SIM give the flows between each pair of sites $F_{ij}(t)$ given the site populations $P_i(t)$
3. Dynamical equation gives population at next time step $P_i(t+1)$

$$\frac{dP_j}{dt} = \lambda(F_{ij} - F_{ji})$$

4. Repeat from 1.

Time scale for migration set by λ

City Distribution

Which Spatial Interaction Model can produce a reasonable city size distribution?

- Use modern US city distribution for comparison
- Sites uniformly scattered on square
- Stop simulation when have best fit to distribution
- Over 80 gravity and radiation models used

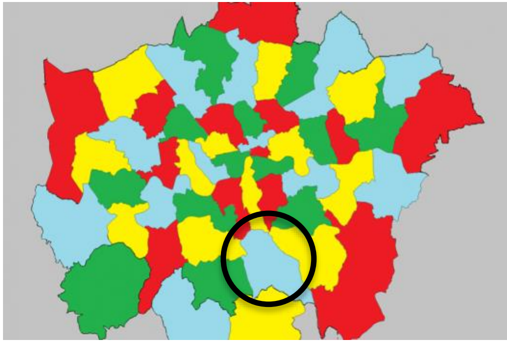
- Do large cities emerge?
- Do settlements cluster sensibly?

[Wilkinson, Emms, TSE 2018]

Results

- A few models get Zipf-like distributions
- No models produce the correct clustering

London Boroughs



Suggests that the attractiveness of a site for long distance (long time) migrants is

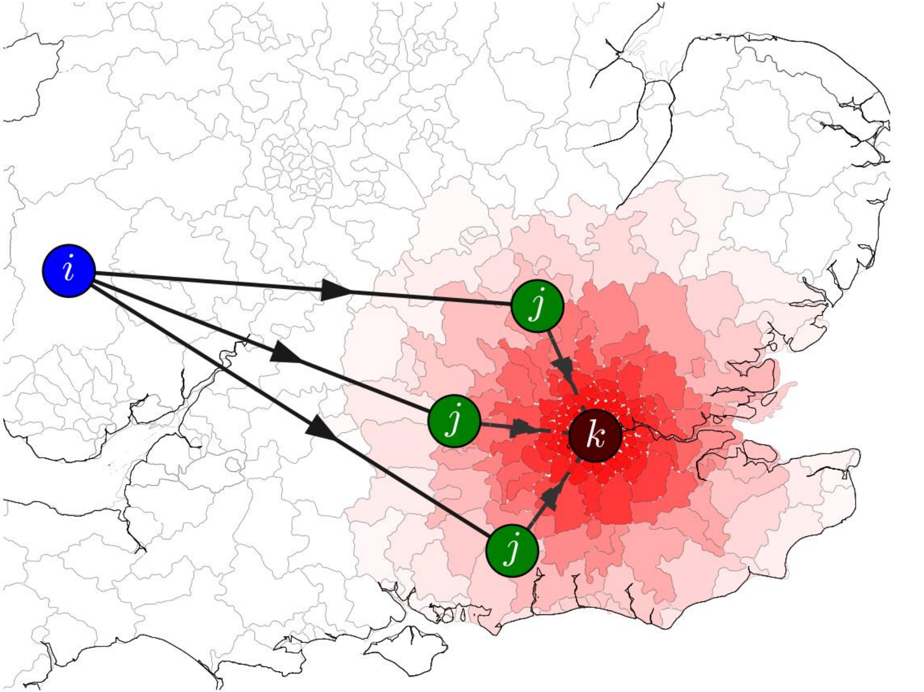
- **not** just a function of the site size
- It is a function of the local **neighbourhood**,
  the local ecology surrounding a site
  - Croydon is not very attractive on its own, but many people live there as it is 30 min to the City of London, IKEA is based there

Croydon

Another victim of 1960's British urban planners?

Best Solution: Two-Trip Model

- Migrate from ***i*** to ***j***
- Commute from ***j*** to ***k***
- Attractiveness of ***i*** depends on neighbours of ***j***



Best Solution: Two-Trip Model

- Adapt existing Spatial Interaction Model (output constrained gravity model)
- Make site attractiveness a function of neighbours using a second simple gravity model with shorter (commuting) distance scale

$$F_{ij} = \frac{P_i \, n_j \, \exp(-d_{ij}/D_m)}{\sum_k n_k \, \exp(-d_{ij}/D_m)}$$

Migration
- Output Constrained
- D_m Long Distance Scale

$$n_j = \sum_k P_k e^{-d_{jk}/D_c}$$

Commuting
- D_c Long Distance Scale

Summary

- Taking flows seriously leads to simle dynamic model of migration
- Only way to get a sensible settlement distribution emerging is if a short distance scale makes attractiveness of a site depend on the local region.

Conclusions

- Enforce all basic properties
  e.g. output and input constraints
- Use additional free parameters to improve fit
- All simple models are rubbish.
  - Do test basic principles
  - Do look for general features
  - Don’t plan your next journey based on them
-

Work with Benjamin HILTON and Abhijay SOOD

Thanks

Specific work reported here:-

- Theo Emms and James Wilkinson (2016-2017)
- Benjamin Hilton and Abhijay Sood (2018-2019)
  http://arXiv.org/abs/1909.07194
- Thomas Louf (2019)

Also drawing on earlier work and discussions with

- Michael Gastner (Yale-NUS College, Singapore) and Elias Bamis (2011-2012)
- Pierfrancesco Bosco (2016)
- M. Dolores Garcia Marti (2018)

and continuing collaboration with Ray Rivers (Imperial)

References

- Arcaute, E.; Hatna, E.; Ferguson, P.; Youn, H.; Johansson, A. & Batty, M., “City boundaries and the universality of scaling laws”, Journal of The Royal Society Interface, 2014 , 12 , 20140745-20140745.
- I. Bamis, “Constrained Gravity Models for Network Flows”, MSc Thesis, Imperial College London, 2012.
- Bettencourt, L. M. A.; Lobo, J.; Helbing, D.; Kühnert, C. & West, G. B., “Growth, innovation, scaling, and the pace of life in cities”, PNAS, 2007 , 104 , 7301-7306.
- T. Emms, “A Dynamical Analysis of Spatial Interaction Models”, MSci Thesis, Imperial College London, 2017.
- P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, “Uncovering space-independent communities in spatial networks,” PNAS 108, 7663–7668, 2011. DOI: http://doi.org/10.1073/pnas.1018962108
- Hilton, B., Sood, A. P. and Evans, T. S., “Predictive limitations of spatial interaction models: a non-Gaussian analysis”, 2019 http://arXiv.org/abs/1909.07194
- C. Knappett, T. S. Evans, and R. J. Rivers, “Modelling Maritime Interaction In The Aegean Bronze Age,” Antiquity , 82, 1009–1024, 2008. DOI: http://doi.org/10.1017/S0003598X0009774X
- C. Knappett, T. S. Evans, and R. J. Rivers, “The Theran eruption and Minoan palatial collapse: new interpretations gained from modelling the maritime network,” Antiquity , 85, 1008–1023, 2011. http://doi.org/10.1017/S0003598X00068459
- T. Louf, “An axiomatic study of spatial interaction models”, MSc Thesis, Imperial College London, 2019.
- T. E. Rihll and A. G. Wilson, “Spatial Interaction and Structural Models in Historical Analysis: Some Possibilities and an Example,” Histoire & Mesure , 2, 5–32, 1987.
- T. E. Rihll and A. G. Wilson, “Modelling settlement structures in Ancient Greece: new approaches to the polis,” in City And Country in the Ancient World , J. Rich and A. Wallace-Hadrill, Eds. Routledge, 1991, pp. 58–95.
- Simini, F.; Gonzalez, M. C.; Maritan, A. & Barabási, A.-L., “A universal model for mobility and migration patterns”, Nature, 2012, 484, 96-100.
- Stouffer, S. A., “Intervening Opportunities: A Theory Relating to Mobility and Distance”, American Sociological Review, 1940, 5, 845-867.
- United States Census Bureau. County-to-county worker flow files (2001). URL https://www.census.gov/data/tables/2000/dec/county-to-county-worker-flow-files.html
- US Census Bureau. US 2010 Census Data - demographic profile; 2010. URL https://www2.census.gov/census_2010/03-Demographic_Profile/
- V. Verbavatz & M. Barthelemy PLOS ONE 114, e0219559, 2019.
- J. Wilkinson, “A Dynamical Analysis of Spatial Interaction Models”, MSci Thesis, Imperial College London, 2017.