Urban Analytics:

City Science, Prediction and Planning

Michael Batty
Centre for Advanced Spatial Analysis
CASA-UCL

12th April 2019

m.batty@ucl.ac.uk
@jmichaelbatty

Outline

- What Is Urban Analytics? Urban Analytics Defined
- A Fourfold Characterisation

  Data, Spatial Analysis, Simulation/Modelling, & Prediction, possibly Prescription or Design: &computers, the net etc.
- The High and Low Frequency City: Dynamics
- Four Example Applications

  Transit and ***Real-Timed Streamed Data***: Our Oyster Card Project

  New Spatial Data from Internet Activity: Airbnb

  Data-Driven Flow Models: Building Models Near the Data

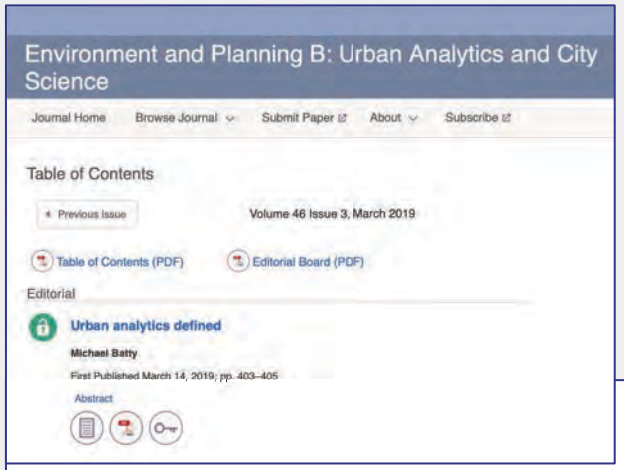
  The QUANT Model: For Very Large Areas, ***Models for Anyone Anytime, Anywhere***: Web-based Predictions
- Where To From Here: Just at the Beginning

What Is Urban Analytics? Urban Analytics Defined

I had a go at defining the scope because three years ago when the company Pion that ran the E&P family of journals we though we should get rid of the planning word which was in the main title and the subtitle and give the journal a slightly stronger edge to what we have done over the years. Hence the new sub-title Urban Analytics and City Science.

At much the same time but a bit before we set up our new Masters course in Smart Cities and Urban Analytics – largely using the term because it seemed to generalise spatial analysis and someone in the Bartlett objected to just the word Smart Cities. A couple of months ago I wrote an editorial on this in EPB and you can get it from here and if you look at what I have tweeted and I don't tweet much, then you can get it there as well

http://spatialcomplexity.blogweb.casa.ucl.ac.uk/files/2019/01/Editorial-EPB-46-1-2019.pdf



https://journals.sagepub.com/doi/pdf/10.1177/2399808319839494

Editorial

B Urban Analytics and City Science

Urban analytics defined

EPB: Urban Analytics and City Science
2019, Vol. 46(3) 403–405

DOI: 10.1177/2399808319839494
journals.sagepub.com/home/epb
SAGE

Suddenly you realise there is a new phrase or term that everyone is using and you wonder where it came from. Urban analytics is one of those clichés that seems to effortlessly roll off the tongue as though we have used it all our lives. It strictly originates from 'urban analysis',

A Fourfold Characterisation

Data, Spatial Analysis, Simulation/Modelling, & Prediction, possibly Prescription or Design

These categories are fairly obvious – they are what we have done in this field for a long time. But data has taken a new twist because we have so much more which is streamed in real time.

And dynamics is so much more on the agenda

While model styles are much more disaggregate with CA, ABM and micro-simulation all focusing on detail of the urban system much more than in the past.

But I should also flag computing itself, the move to graphics and the move to the web from the desktop – these are questions of how we produce what we do and how we communicate but also about how we use a new understanding from visual science and from crowdsourcing to improve what we do.

The High and Low Frequency City: Dynamics

To me, the real issue about urban analytics is the fact that real-time streaming of data has given us a strong momentum to move to what we can call the 'high frequency city'. This is the 24 hours city that in the past was managed in a low key way in terms of its routine functions but now with the advent of big data is providing us with a much richer sense of what cities re all about.

Much of our urban theory and data in the past has been about the low frequency city – how cities change over years and decades and longer but to this we now add changes in seconds, minutes, hours days and so on.

My examples will cover this mix and essentially we need new theories and new models to handle many of these questions.

And to elaborate my argument let me describe four applications that serve to define my interests but are simply the tip of the iceberg in terms of what urban analytics is now all about

Four Example Applications

Transit and *Real-Timed Streamed Data*: **Our Oyster Card Project**

Our first project is mining the real-time transit data, the tap-in and tap-outs for in our case, the London tube. And producing an O-D data set which is based on shortest routes; and then examining heterogeneous travel behaviour so we can mine the rich variety of the data set.

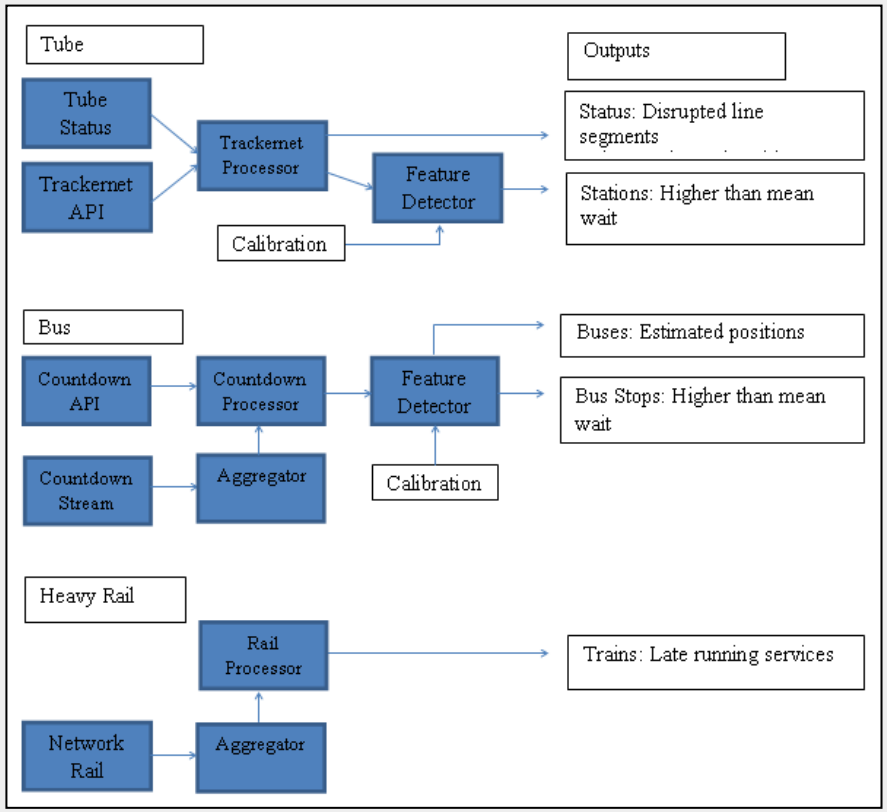
We also wish to tie up demand and supply for travel, examine urban structure that this reveals, and also explore disruption from the data set

We have extended this to Beijing and Singapore making some comparative analysis of what turn out to be quite similar subway systems

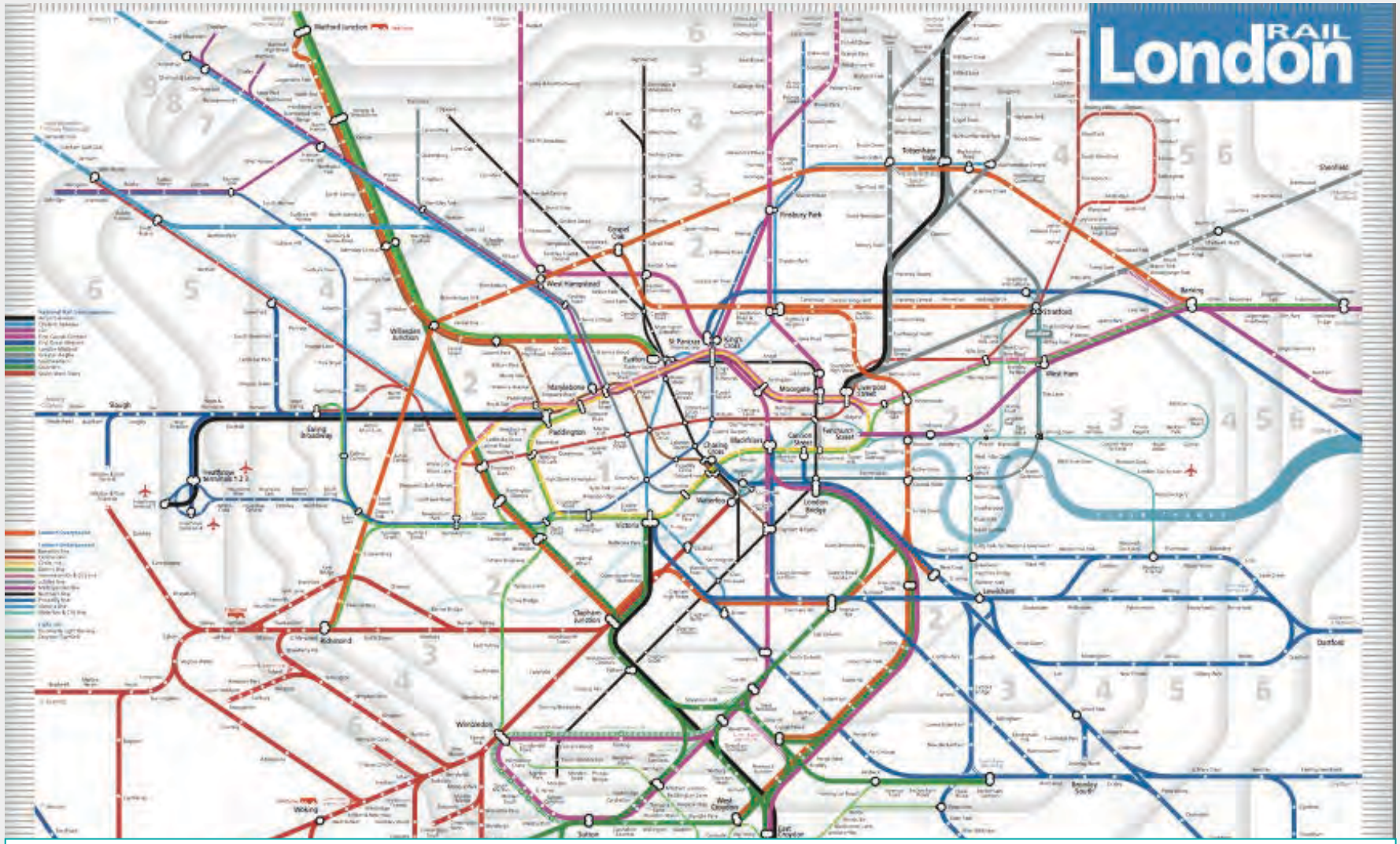
Real-Time Streaming: Oyster Card & Trackernet Data

Data Capture:

Demand by Travellers ***Supply of Trains***

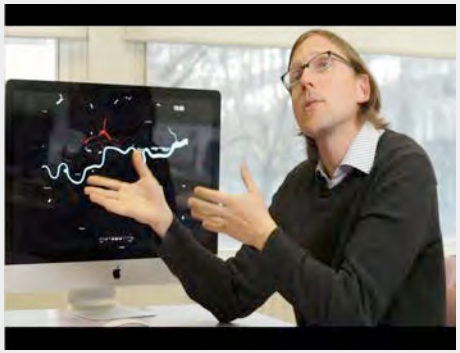


How do we match demand &supply without passenger tracking?



Tube, Overground and National Rail Networks in London where Oyster cards can be used

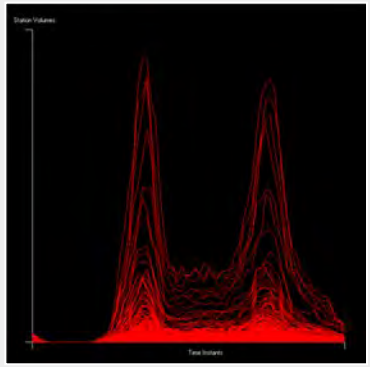
Synthesizing the O-D data, making assumptions about shortest paths

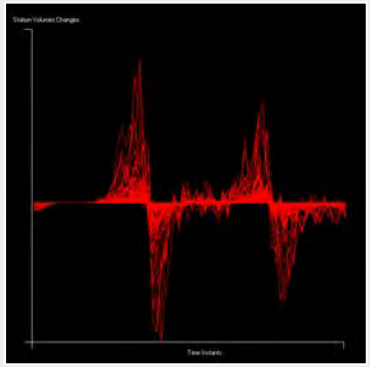


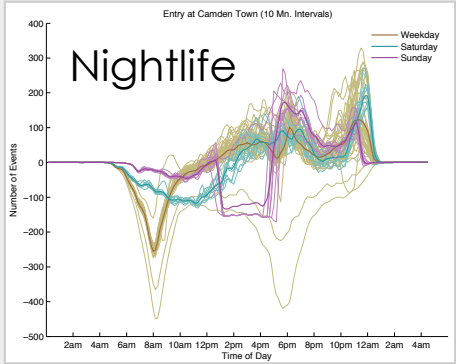
https://www.youtube.com/watch?v=9sAugcb2Qj4

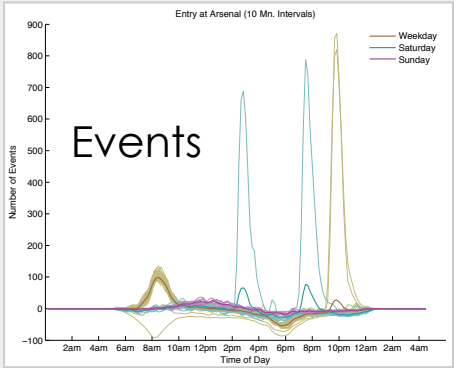
Variabilities – Heterogeneity and Travel Profiles

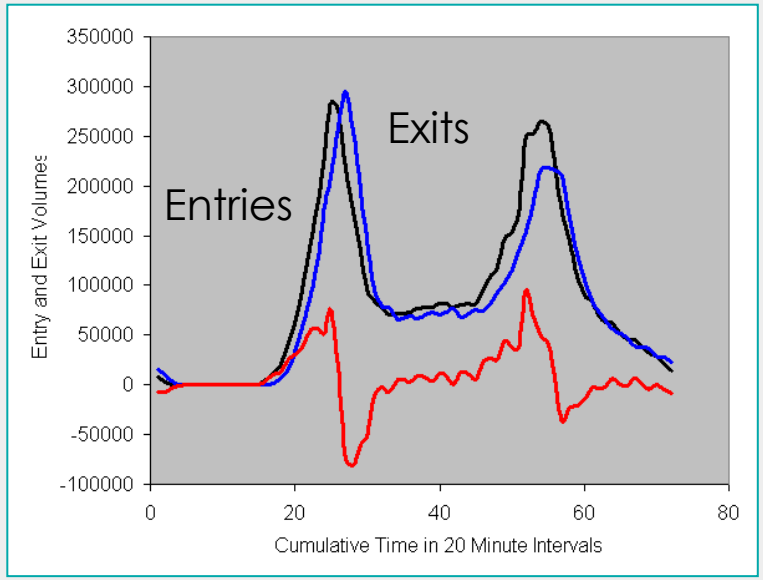
Diurnal flows morning & evening peaks, small blip (peak) around 10pm at night

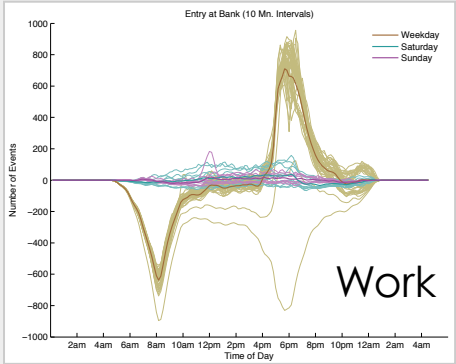


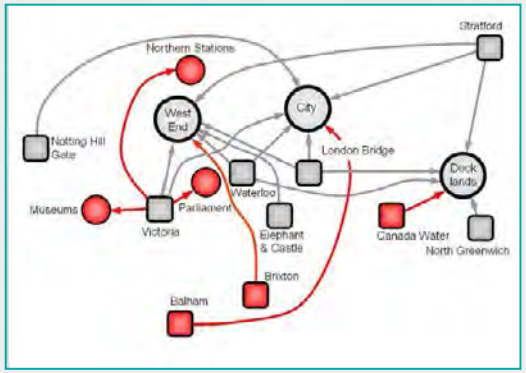




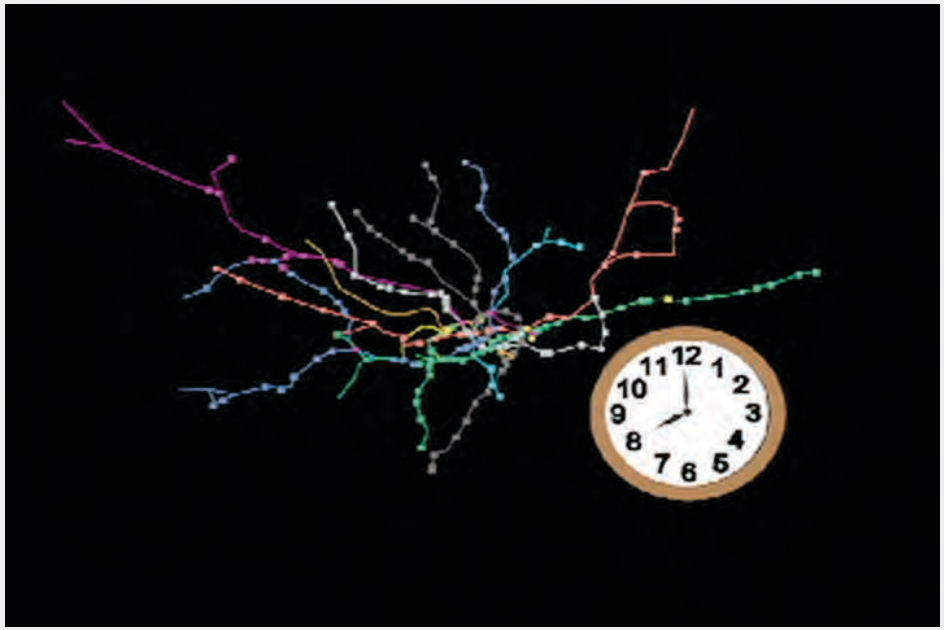








Roth C., Kang S. M., Batty, M., and Barthelemy, M. (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. **PLoS ONE 6(1):** e15923. doi: 10.1371/journal.pone.0015923



The biggest challenge is that the demand and supply data are not coordinated – we do not know what train a passenger gets on although we know where and when they enter the subway stations and also the precise position of the trains.

We cannot connect up the data as there is no tracking in the tube and even if there were it is unlikely we could get complete coverage.

Learning about Mobility from the Data Variabilities – Disruptions – Dynamics:

Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

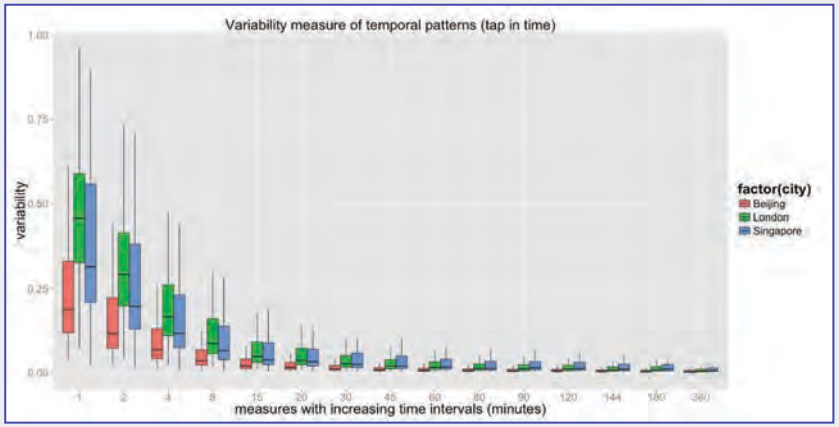
Table 1. Summary statistics of one-week of smart-card data (metro trips only)

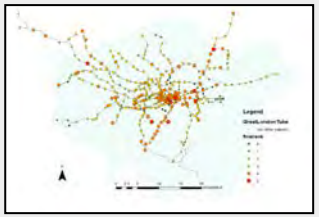
	London	Singapore	Beijing
Monday	3,457,234	2,208,173	4,577,500
Tuesday	3,621,983	2,250,597	4,421,737
Wednesday	3,677,807	2,277,850	4,564,335
Thursday	3,667,126	2,276,408	4,582,144
Friday	3,762,336	2,409,600	4,880,267
Number of stations (1)	400	130	233
Number of tube line	13	4	17
Area (2)	1,572 km^2	718.3 km^2	2267 km^2
Total population (3)	8.63 million	5.3 million	21.15 million
Ridership of Metro	20%	35%	21%
Length of metro lines	402km	182km (MRT+LRT)	465 km

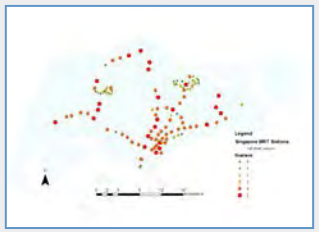
(1) Number of stations is the number of stations with smart-card records generated.

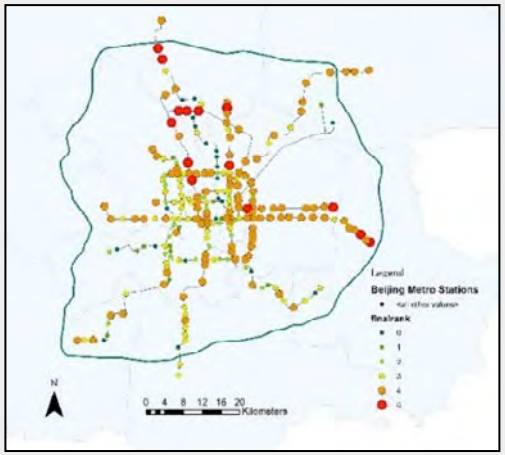
(2) The area of Beijing only counts the area enclosed by the 6th ring road for a fair comparison.

(3) From the World Population Review, http://worldpopulationreview.com/world-cities/ accessed 17 January 2016



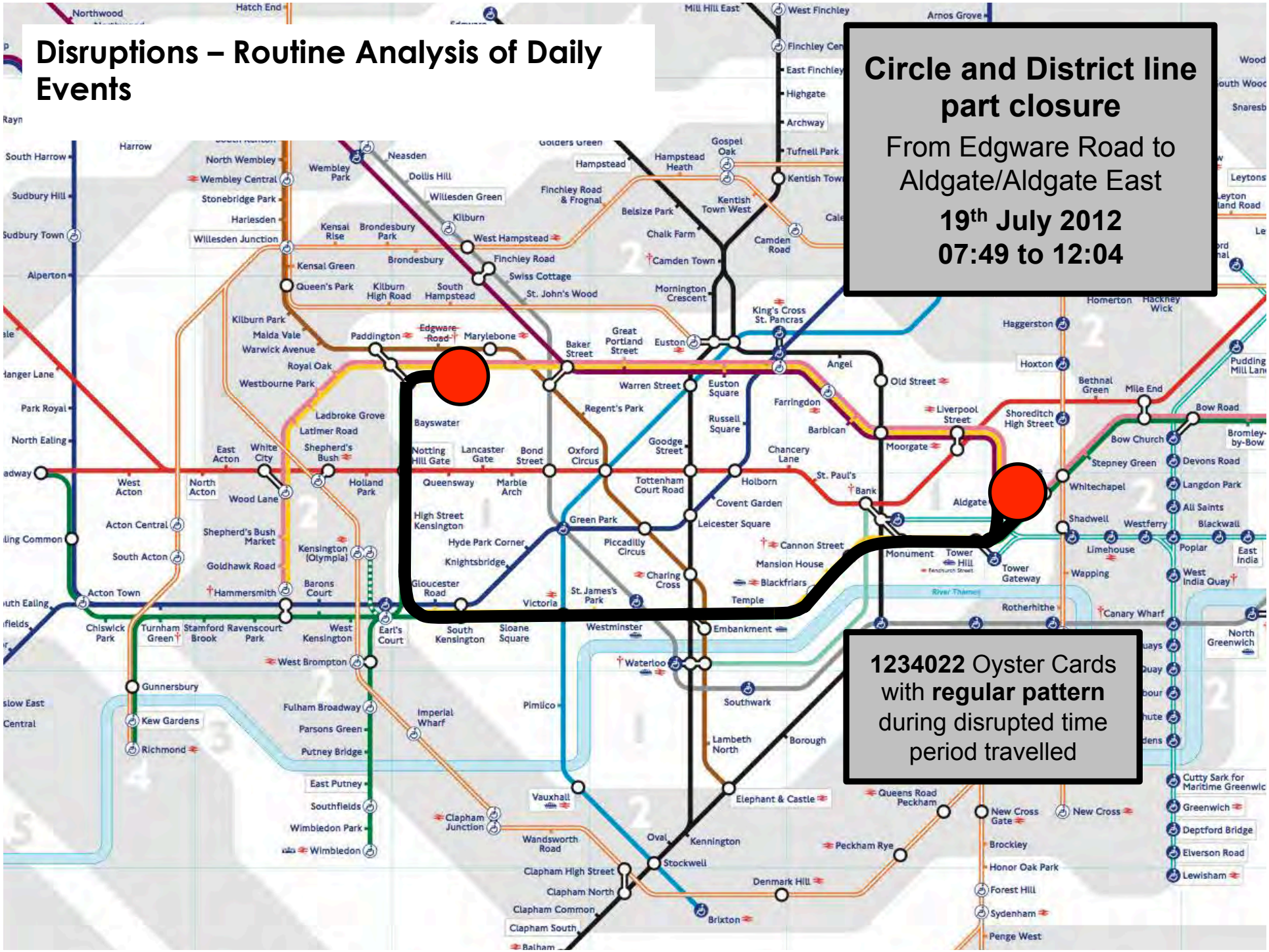






Zhong, C., Batty, M., Manley, E., Wan, J., Wang, Z., Che, F., and Schmitt, G. (2016) Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing using Smart-Card Data., **PLOS One**, http://dx.doi.org/10.1371/journal.pone.0149222

Disruptions – Routine Analysis of Daily Events



Circle and District line part closure

From Edgware Road to Aldgate/Aldgate East

19th July 2012

07:49 to 12:04

1234022 Oyster Cards with **regular pattern** during disrupted time period travelled

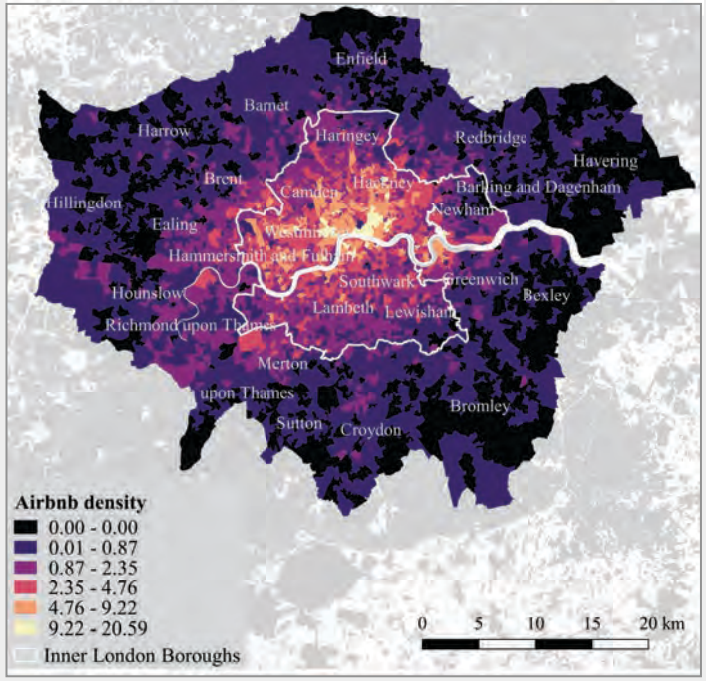
New Spatial Data **from Internet Activity: Airbnb**

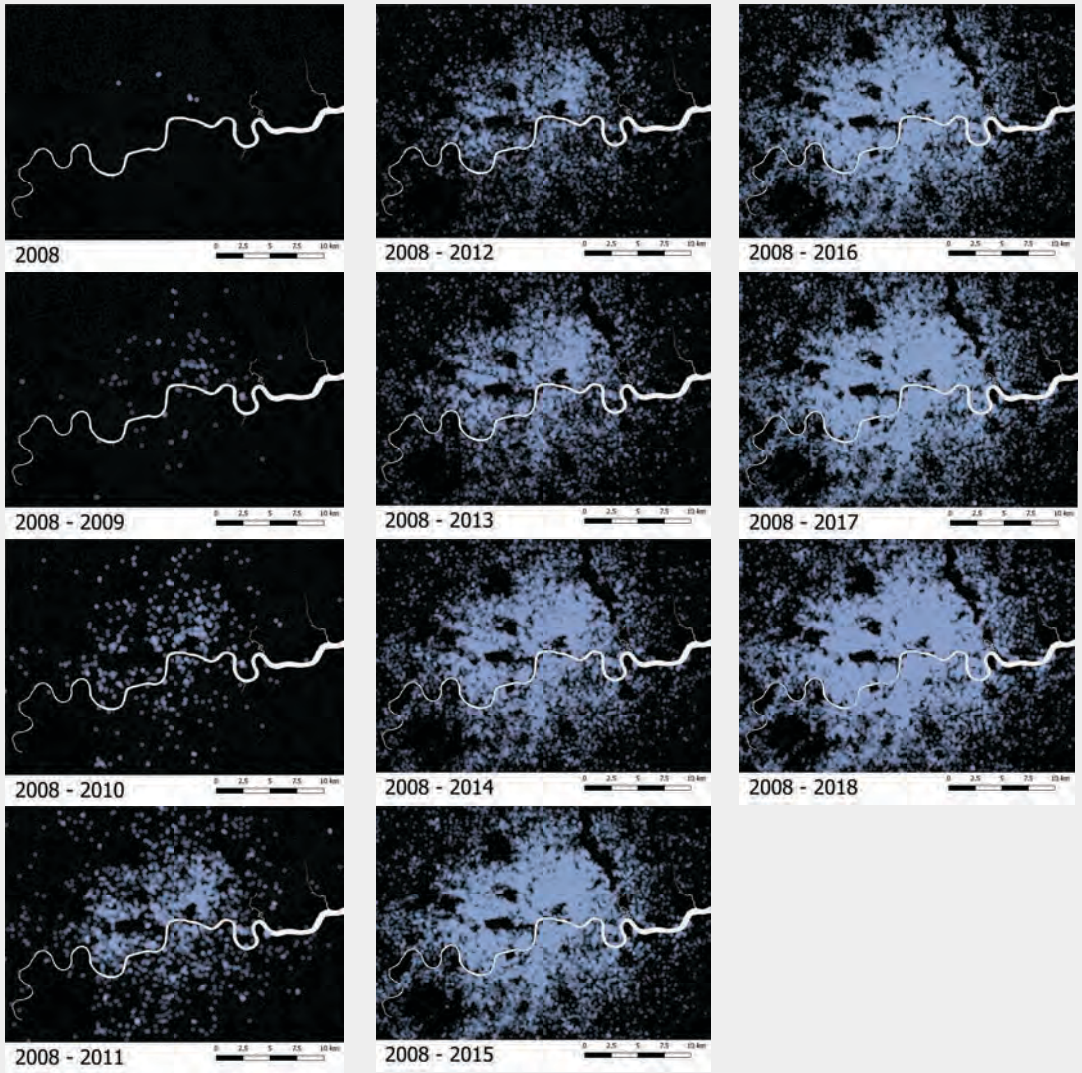
We are also working with Airbnb data for London which is available from 2008. The problem is that the data does not allow us to figure out those establishments that decide to de-register and thus we only ever see a snapshot of data at a cross section. The dynamics are tricky.

What we have discovered – which is obvious in hindsight – is that the conversion of dwelling space to Airbnb is probably reducing the overall supply of rented housing for longer term workers. This is likely to be the case in a very tight housing market like London which is transforming in any case due to inward migration.

We show through simple correlation and variance analysis that the spatial pattern of Airbnb is closest to pattern of rental flats which is also the most homogeneous areas of the city with respect to dwelling types. We use an entropy spread measure to show this.

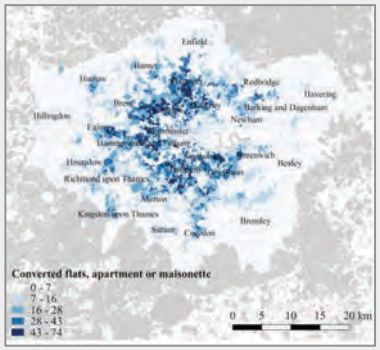
The Growth of Airbnb 2008-2018



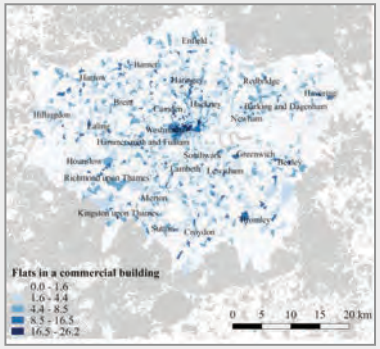


Flat Types

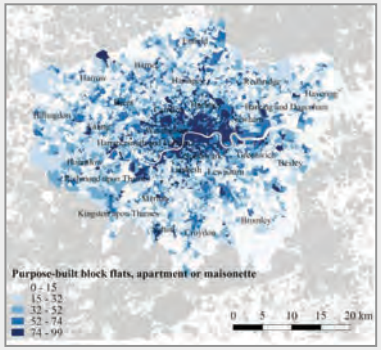
Conversions



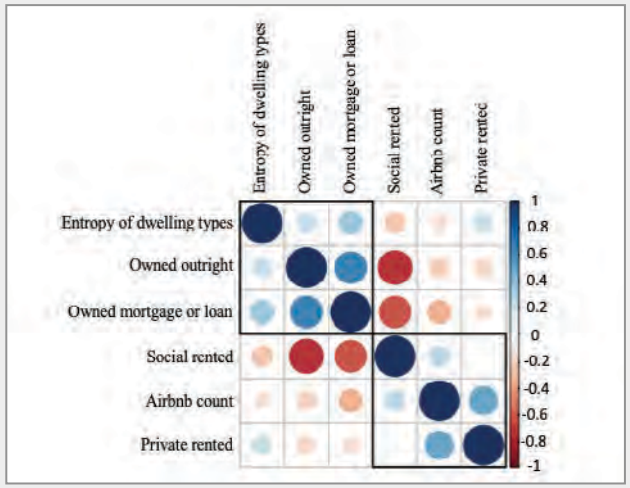
Commercial



Purpose Built

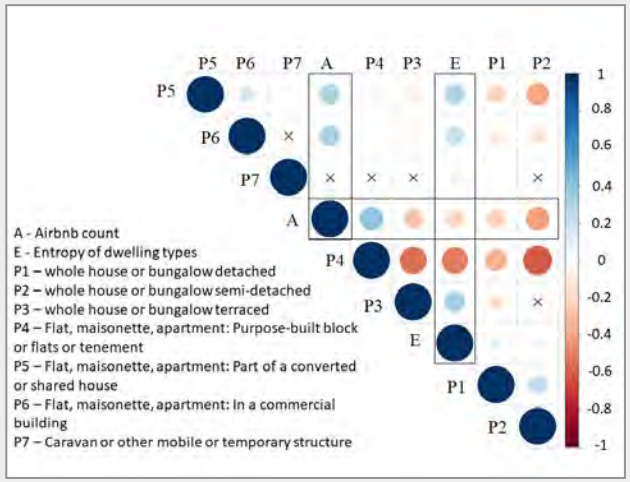


Correlations

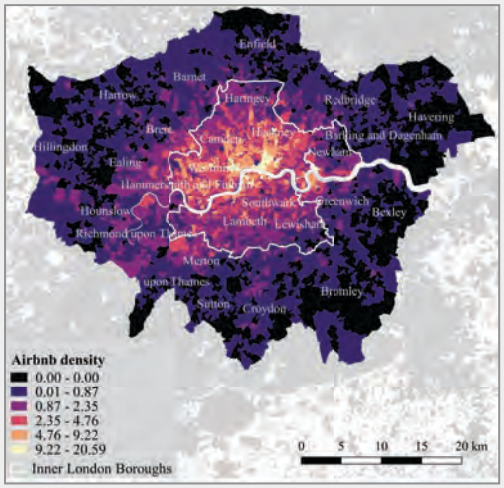


Entropy measures dwelling types k

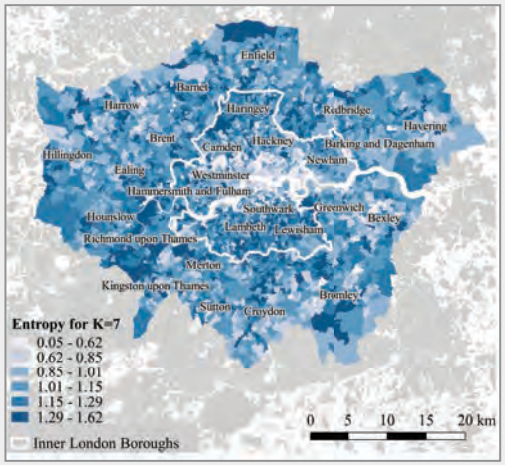
$$H_k = -\sum_{i=1}^{n} P_{ki} \log P_{ki} \quad where \quad \sum_{i=1}^{n} P_{ki}$$



Density of Airbnb



Entropy for k=7



Data-Driven Flows : **Building Models Near the Data**

The big question that was never really posed 50 years ago when spatial models were first developed was whether or not we could use the data as the model and then simulate changes from the data as the predictions.

The reason why we never did this was because we though the models were legion – the data didn't explain as much as our theories might be able to. In short why would we use the data as the model when we wanted a model to explain the data.

We are evolving a strategy for our urban models – mainly spatial interaction but any model –where we predict how the data is modified. This also enables us to handle dynamics much more effectively and we try to show how recent changes – marginal change – can be modelling on top of long term urban structure. To an extent this ideology is based on the notion that the best prediction for the future is the present ! Ok let me show you how.

Time t+1 ← Time t

$$T_{ij}(t+1) = \underbrace{KA_iB_jQ_{ij}}_{\text{Model}}T_{ij}(t)$$

Data Now

Data Then

$$T_{ij}(t+1) = KA_iB_j\cancel{Q_{ij}}T_{ij}(t)$$

Totals Origin Destination *i-j* Flow

…… **Model Parameters** ……

$$T_{ij}(t+1) = KA_iB_jT_{ij}(t)$$

The Fratar Model circa 1953

Conversion to Prior and Posterior Probability Form

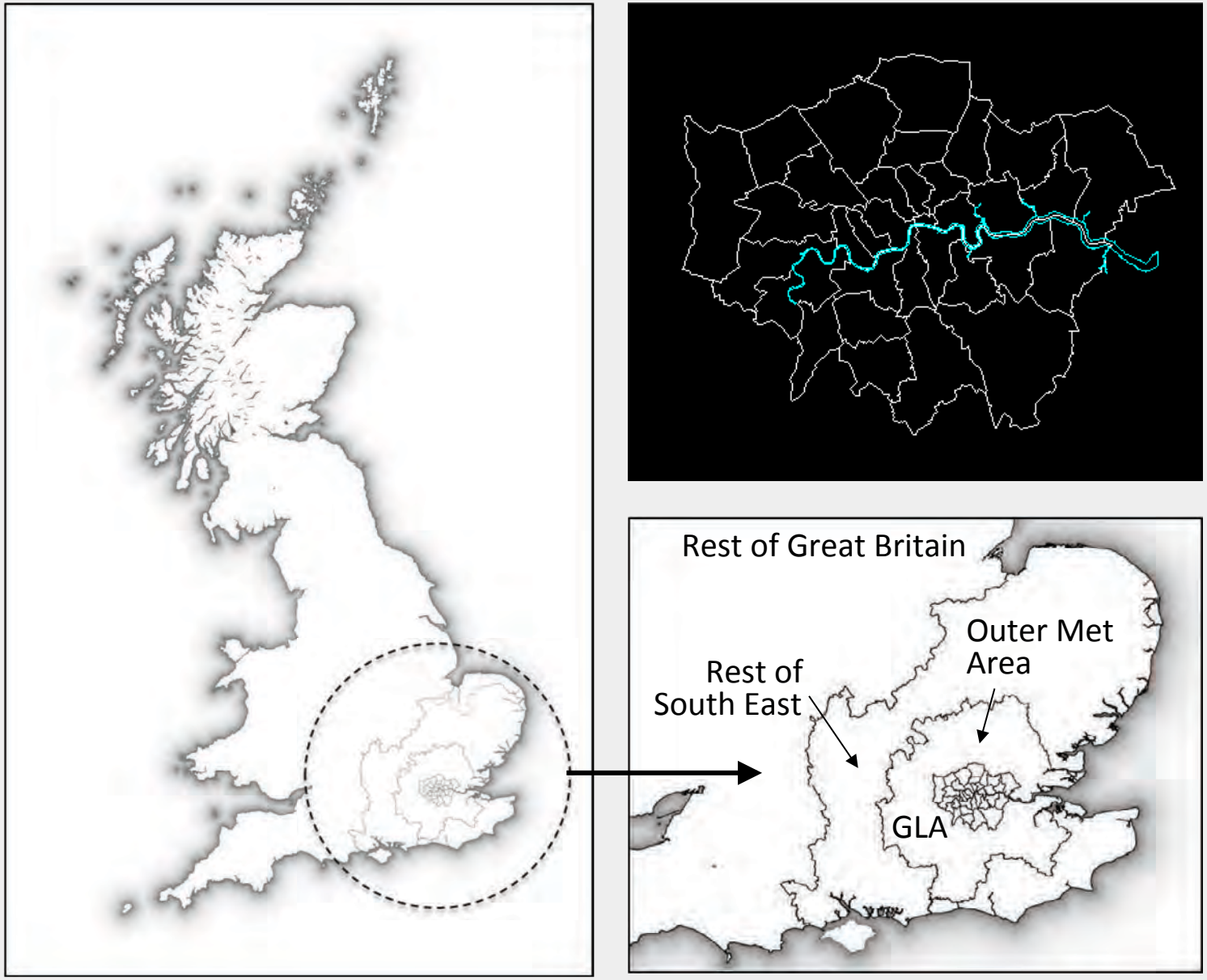
$$p_{ij}(t+1) = kA_iB_jQ_{ij}p_{ij}(t)$$

where we define these in terms of flow variables $T_{ij}(t+1),\ T_{ij}(t)$

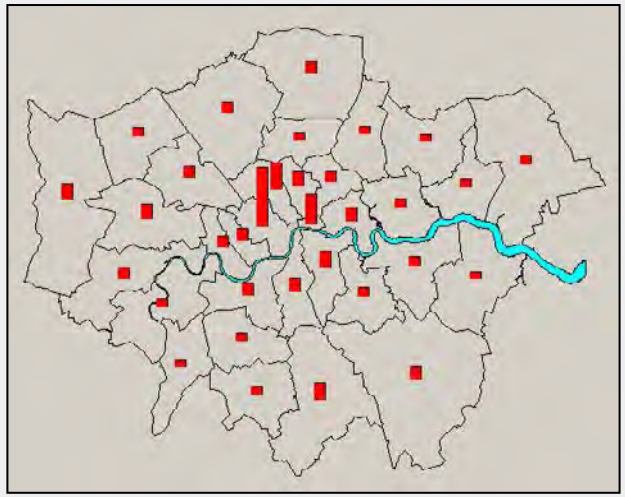
$$p_{ij}(t+1) = \frac{T_{ij}(t+1)}{T(t+1)} = kA_iB_jQ_{ij}\frac{T_{ij}(t)}{T(t)}$$

I could spend a long time talking about these models and over the years I have worked on them as ways of building dynamic models – using Bayesian updating.

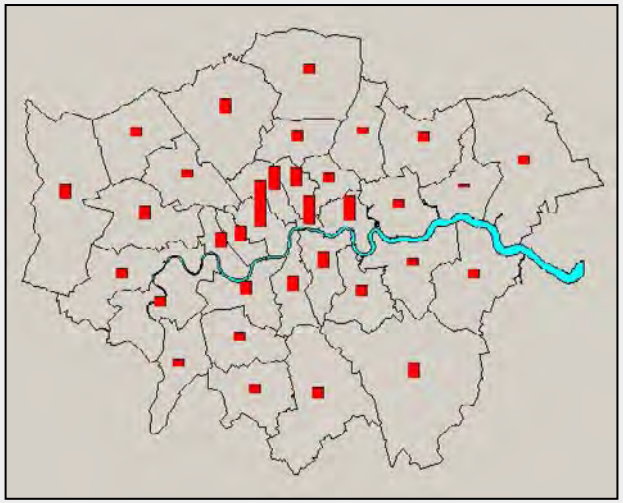
But we have never thought of them as data-driven models, always as ways of filling in missing information – it time to push them to the logical conclusions and begin to test these ideas – my example once again is London


Rest of Great Britain
Outer Met Area
Rest of South East
GLA

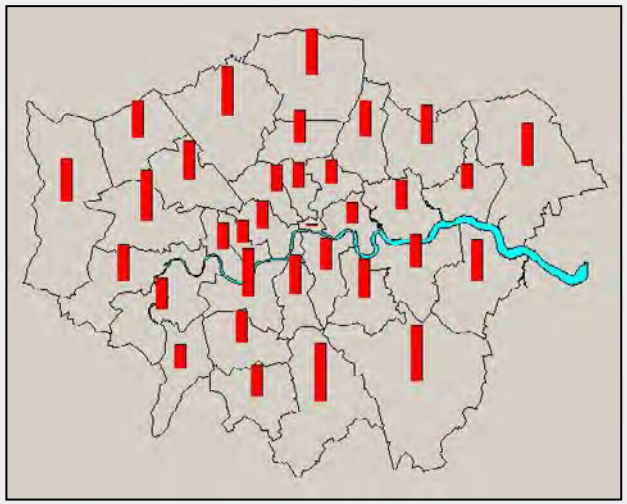
Employment 1991



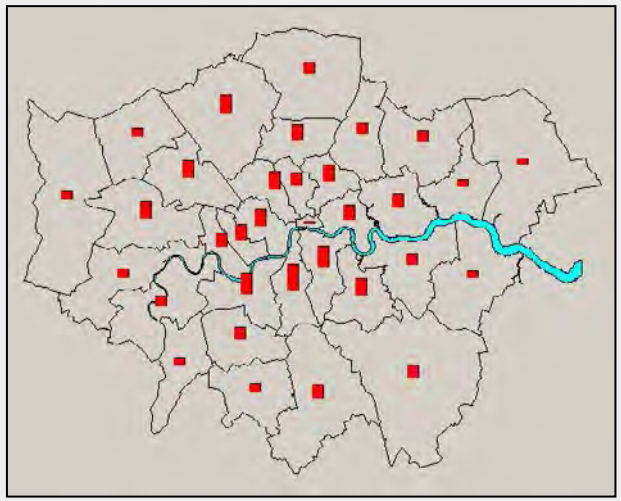
Emp Change 1991-2001



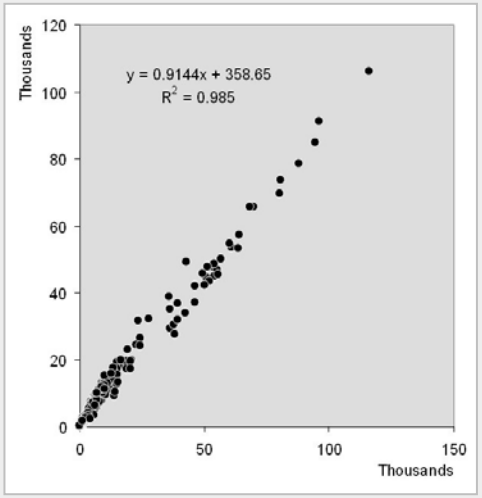
Population 1991



Pop Change 1991-2001

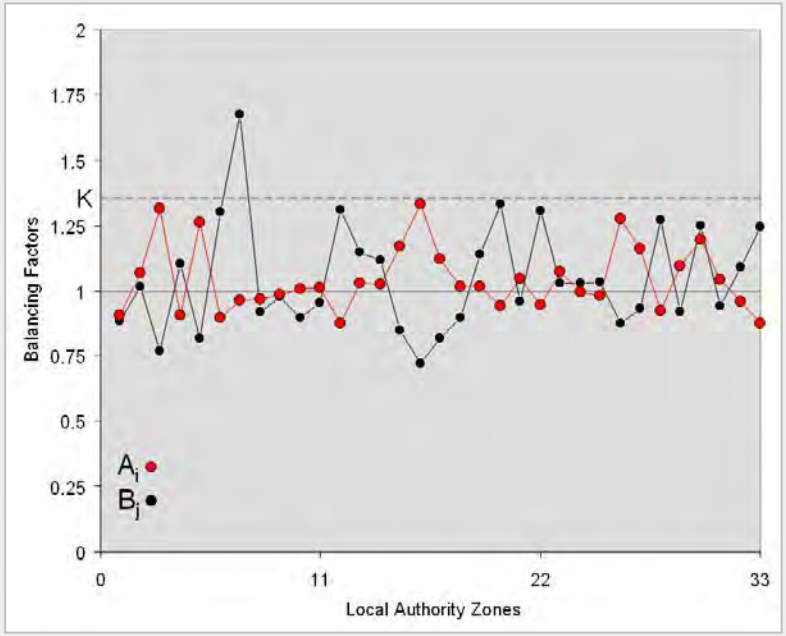


Prediction of Trips in 2001 is correlated with Trips in1991 at $r^2 = 0.985$ with a Sorenson index of 0.925



$$\hat{T}_{ij}(pred)\ v\ T_{ij}(2001)$$

We can predict 2001 from 1991 in terms of locational constraints, noting that the key model parameters now are k, A_i, B_j We need to explain these and in essence these are distortions of spatial structure, interpretable as housing and employment stress, economic surplus and so on. We can graph them as follows –noting that A_i, B_j are key model parameters & $k = 1.375$



Much much more but no time to take this further but simply to say that for our QUANT model, we intend to explore the future using the past & present in this data driven fashion

The QUANT Model: *Models for Anyone Anytime, Anywhere*: **Web-based Predictions**

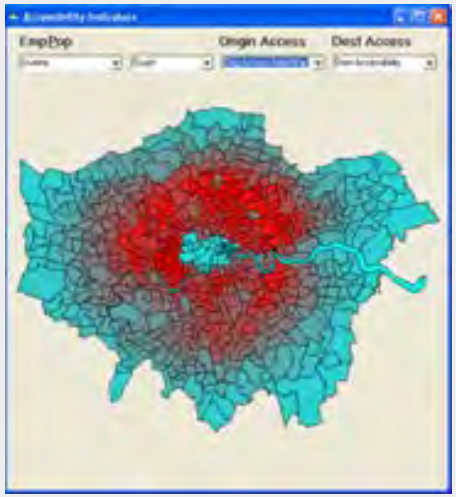
Alan Wilson yesterday talked about our QUANT model – we can now scale our models up to the nation and even beyond but working at small zones in the manner of past urban models.

Size in this sense is no longer a problem

We can also move out models into a new media – from the desktop to the web –making them available from any device with an internet connection. We can thus explore many problems that have impacts nationwide like new infrastructure – HS2 – migration – industrial strategy at the regional level etc.

There are still pretty massive computing problems but we can now build three mode spatial interaction models that run quite fast with 10,000 or so zones in about 30 secs. We are working to get the models ever faster and ever more available – e.g. DAFNI

We have a reasonable tradition with scaling our models up and in fact it was here in Newcastle that we were involved with our first one in the Tyndall cities project and then in Arcadia

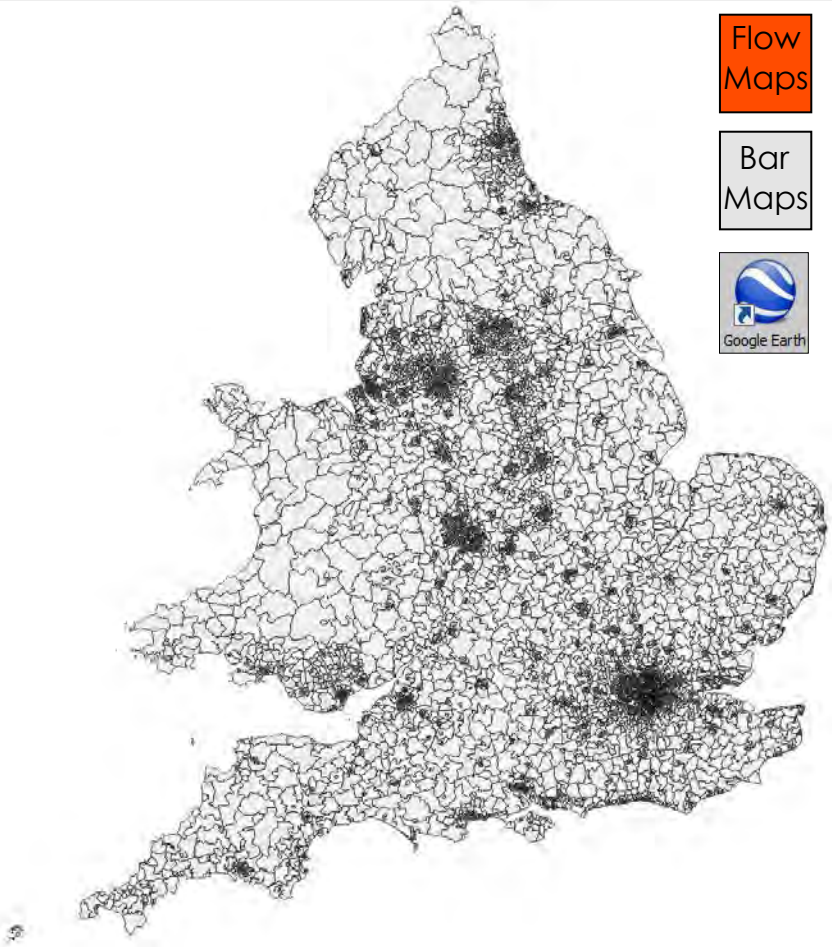


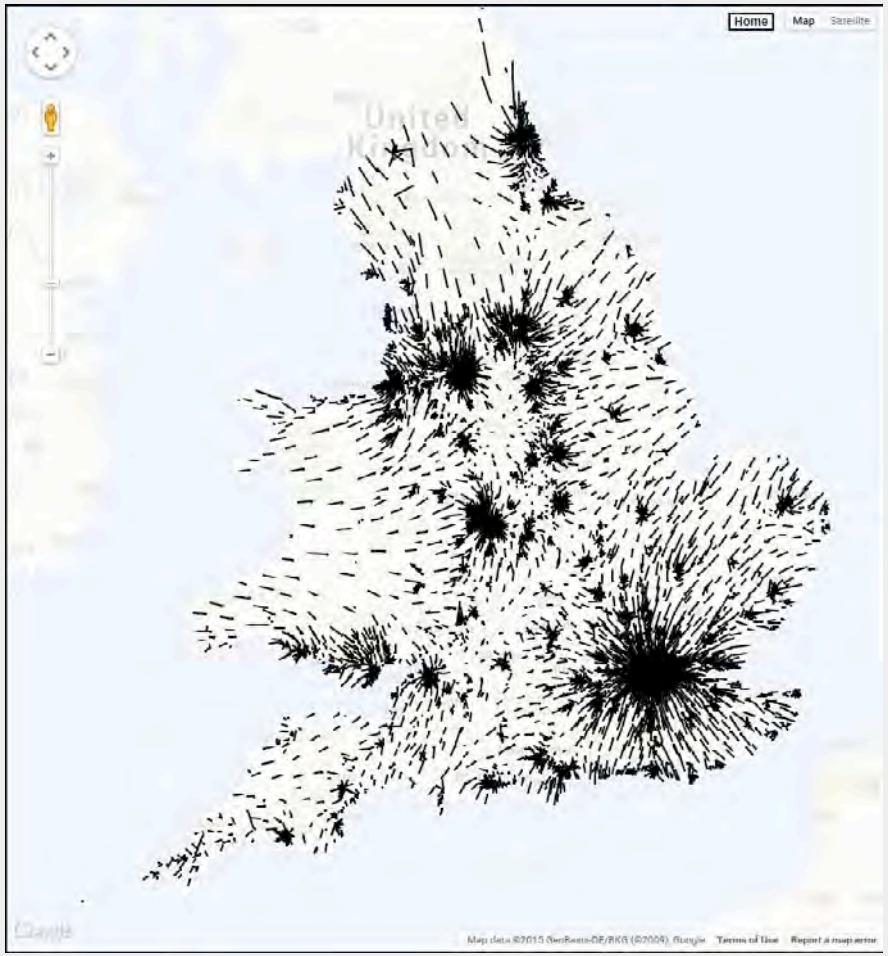
A couple more points. Our national model QUANT which means **Q**uantitative **U**rban **AN**alytics forecas**T**ing or some variant thereof is built at MSOA zonal level with some 7201 in E and W.

We have now added Scotland − the data there is a little different − hence the delay from the original model. And the total number of zones is now about 8417.

The model is a three-mode spatial interaction model whose networks are built from the fine scale nets for road, bus/ferry etc., and rail shown by Alan Wilson yesterday. The networks are scaled up to zonal level but shortest routes and changes are accomplished at the fine scale level and this is a major issue in exploring future scenarios involving transport other than rail where the problem is manageable and can be solved at speed.

We will now show some examples of the model − with Scotland − and then examine some scenarios such as the impact of new rail lines.

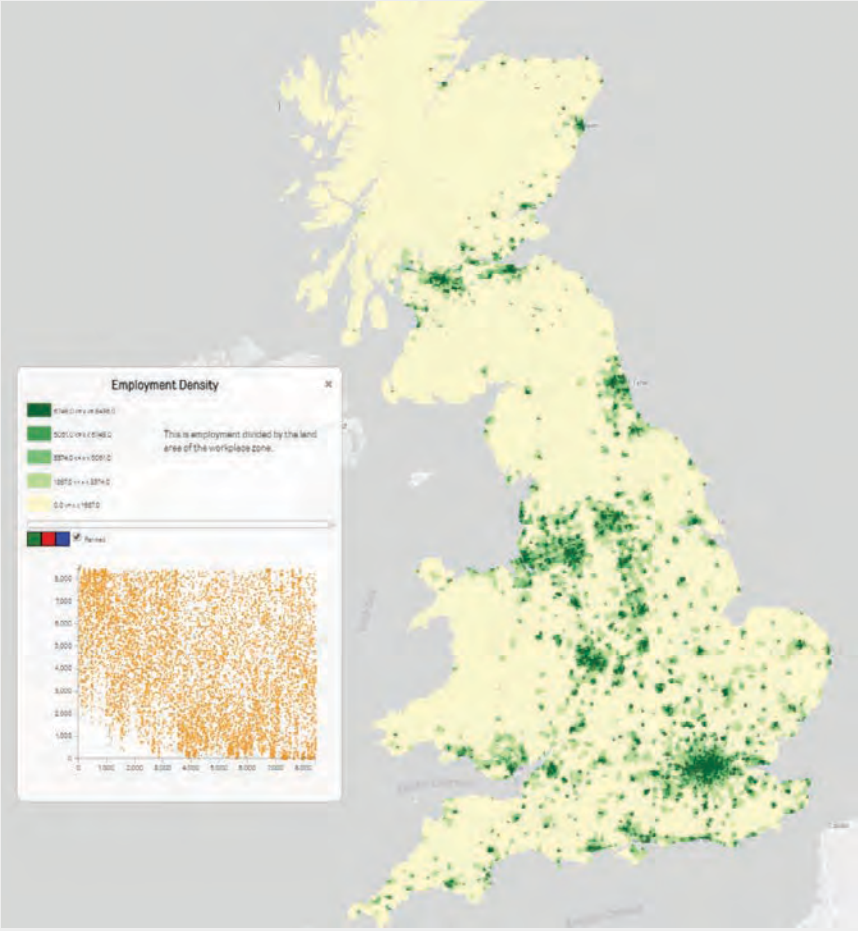




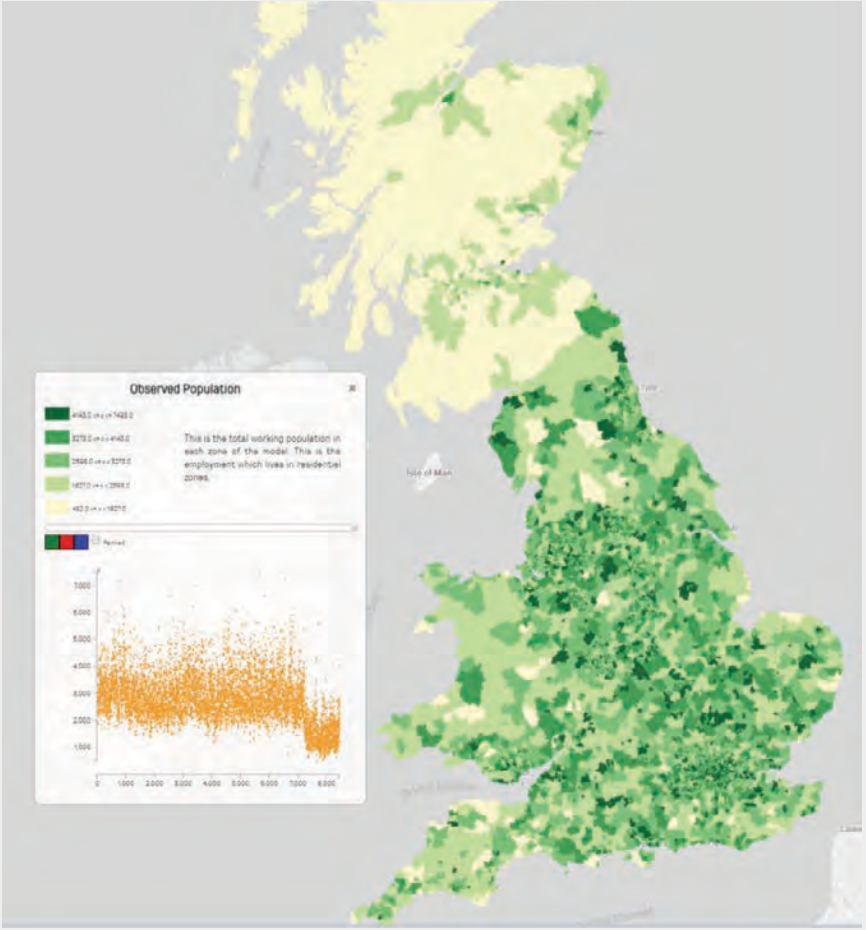
$$[x_i, y_i] = [[x_i, y_i], \left[\left[x_i + \frac{\sum_j T_{ij}\,[x_i - x_j]}{n}\right], \left[y_i + \frac{\sum_j T_{ij}\,[y_i - yy_j]}{n}\right]\right]]$$

Urban Analytics: City Science, Prediction and Planning

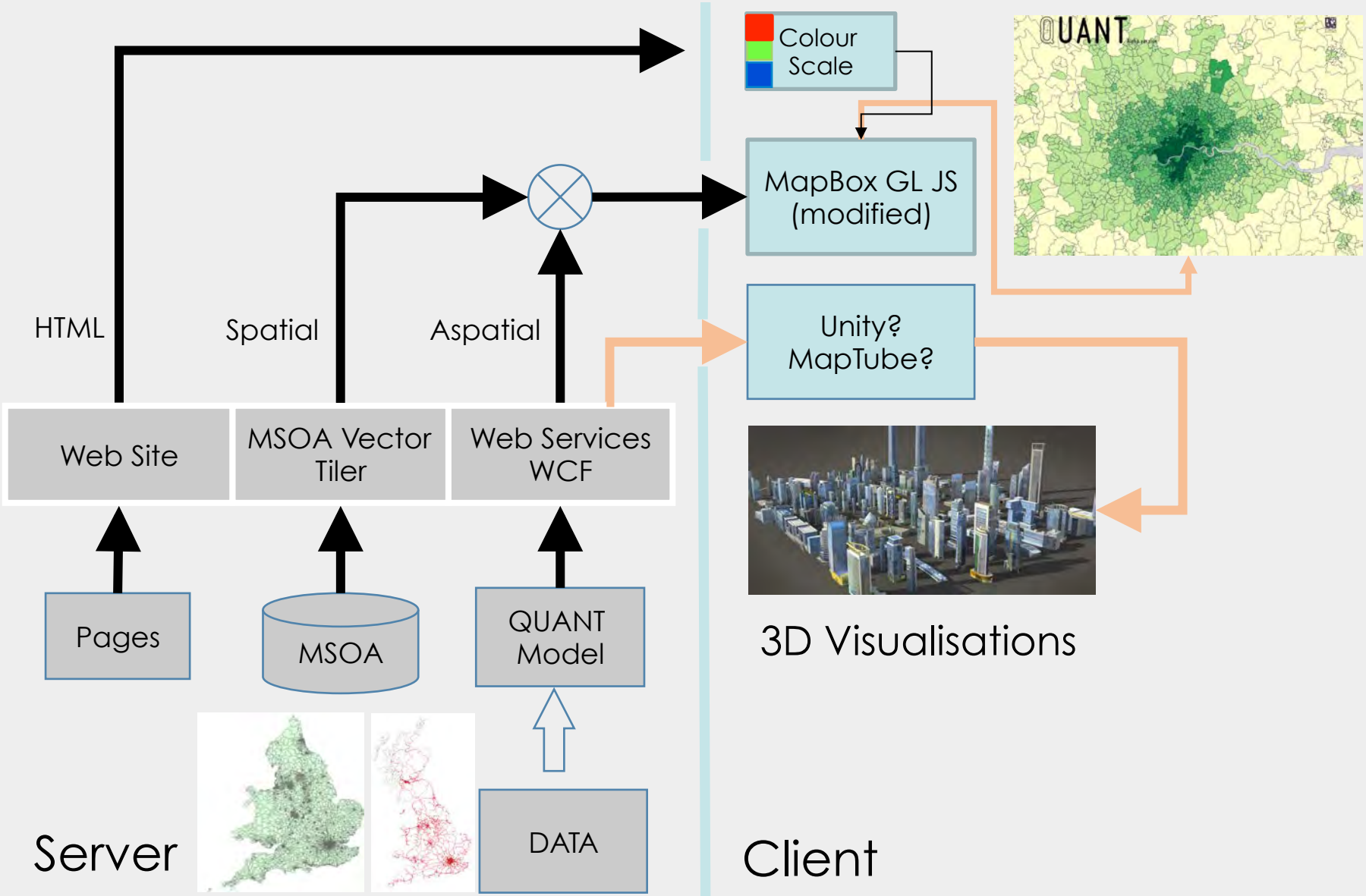
Employment Density

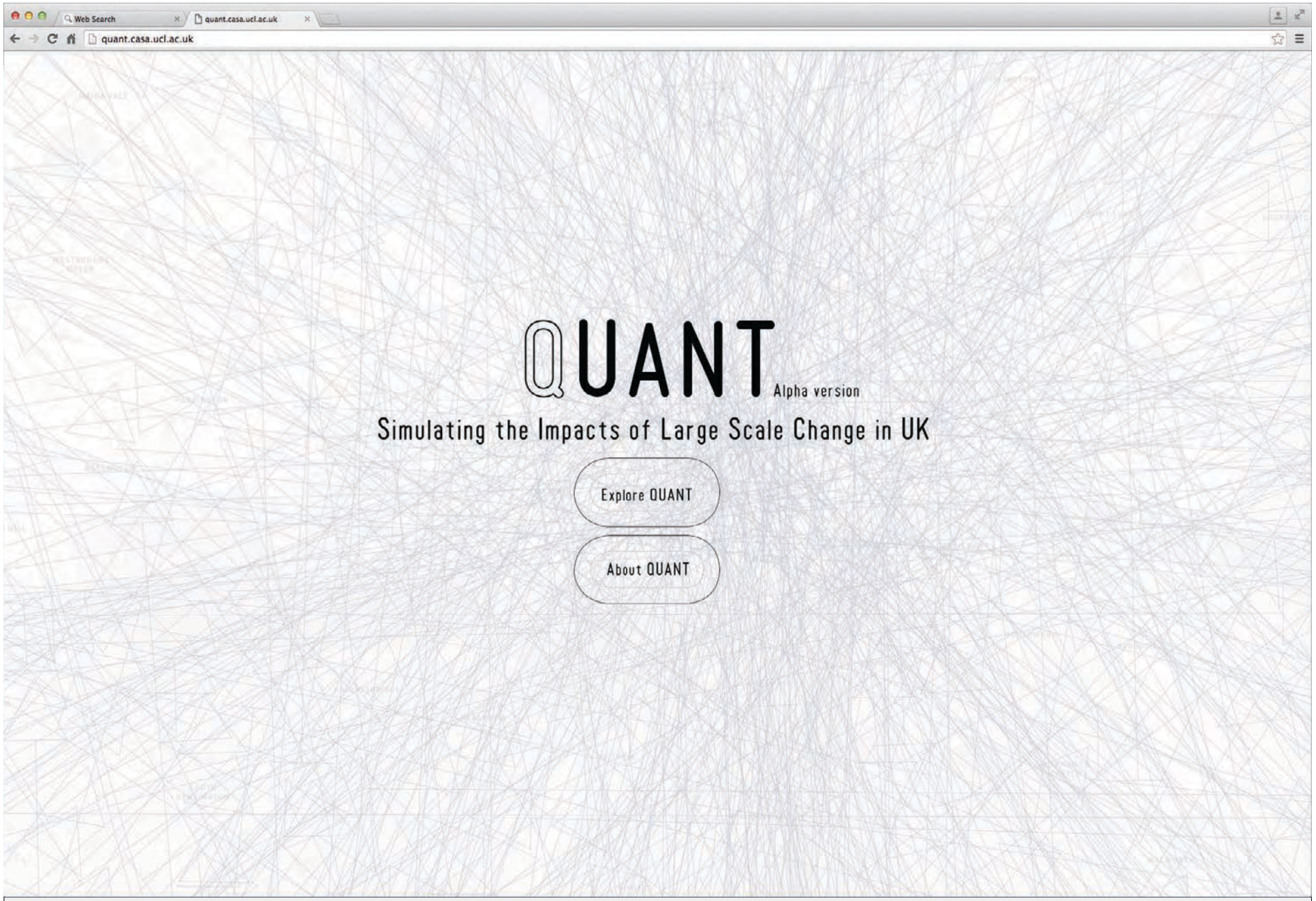


Population Counts



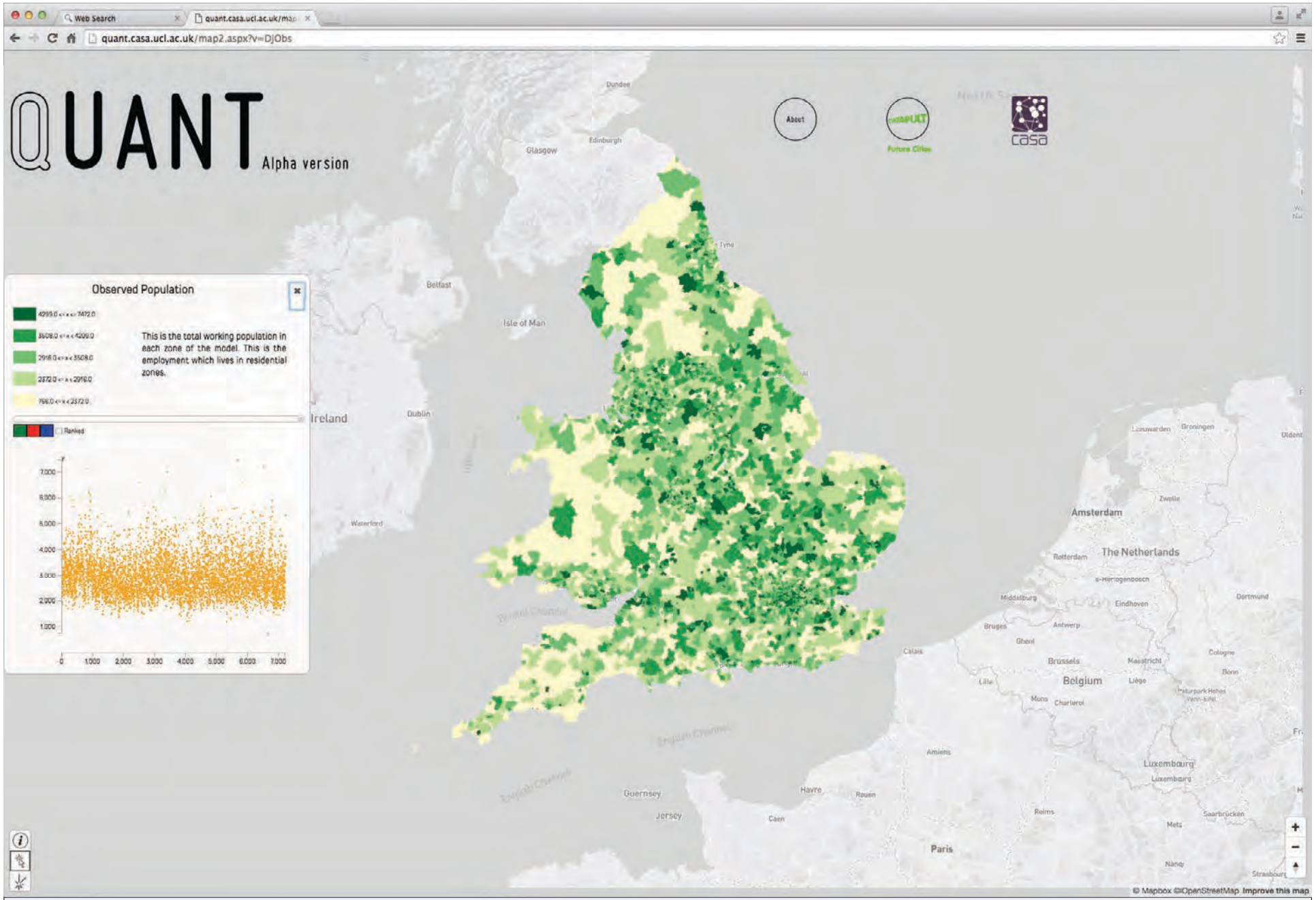
Client & Server Architecture





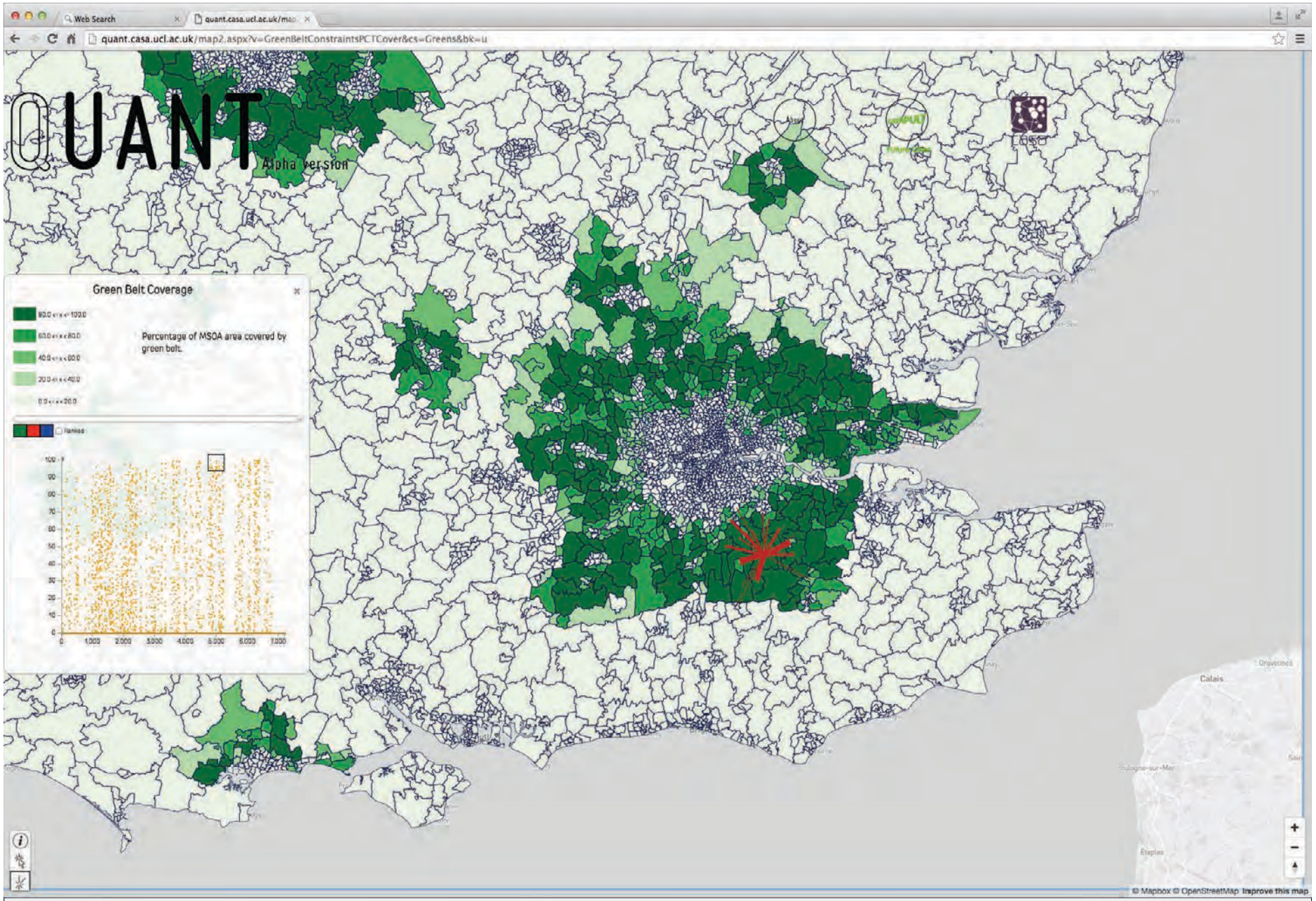
Urban Analytics: City Science, Prediction and Planning

Urban Analytics: City Science, Prediction and Planning

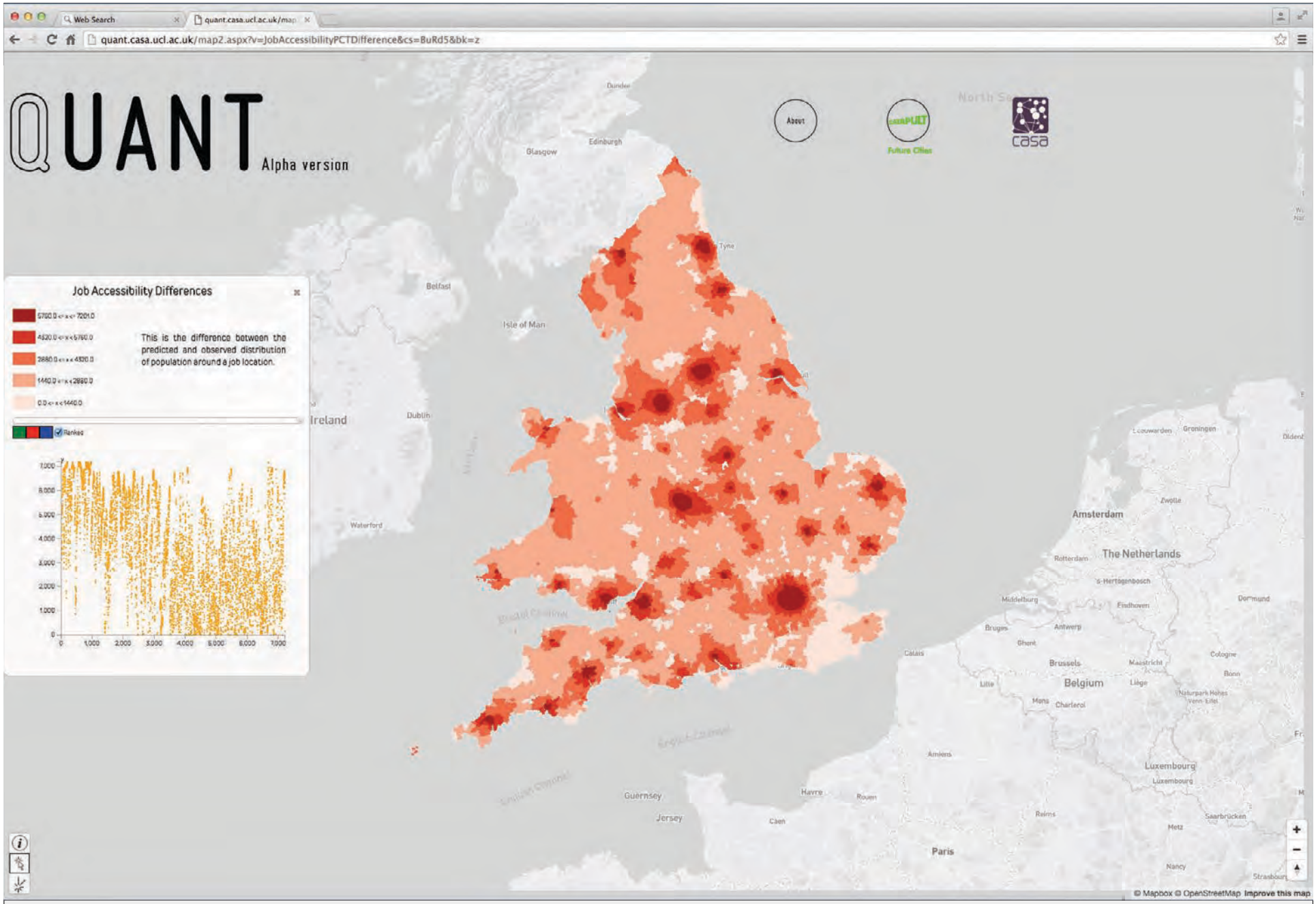


Urban Analytics: City Science, Prediction and Planning

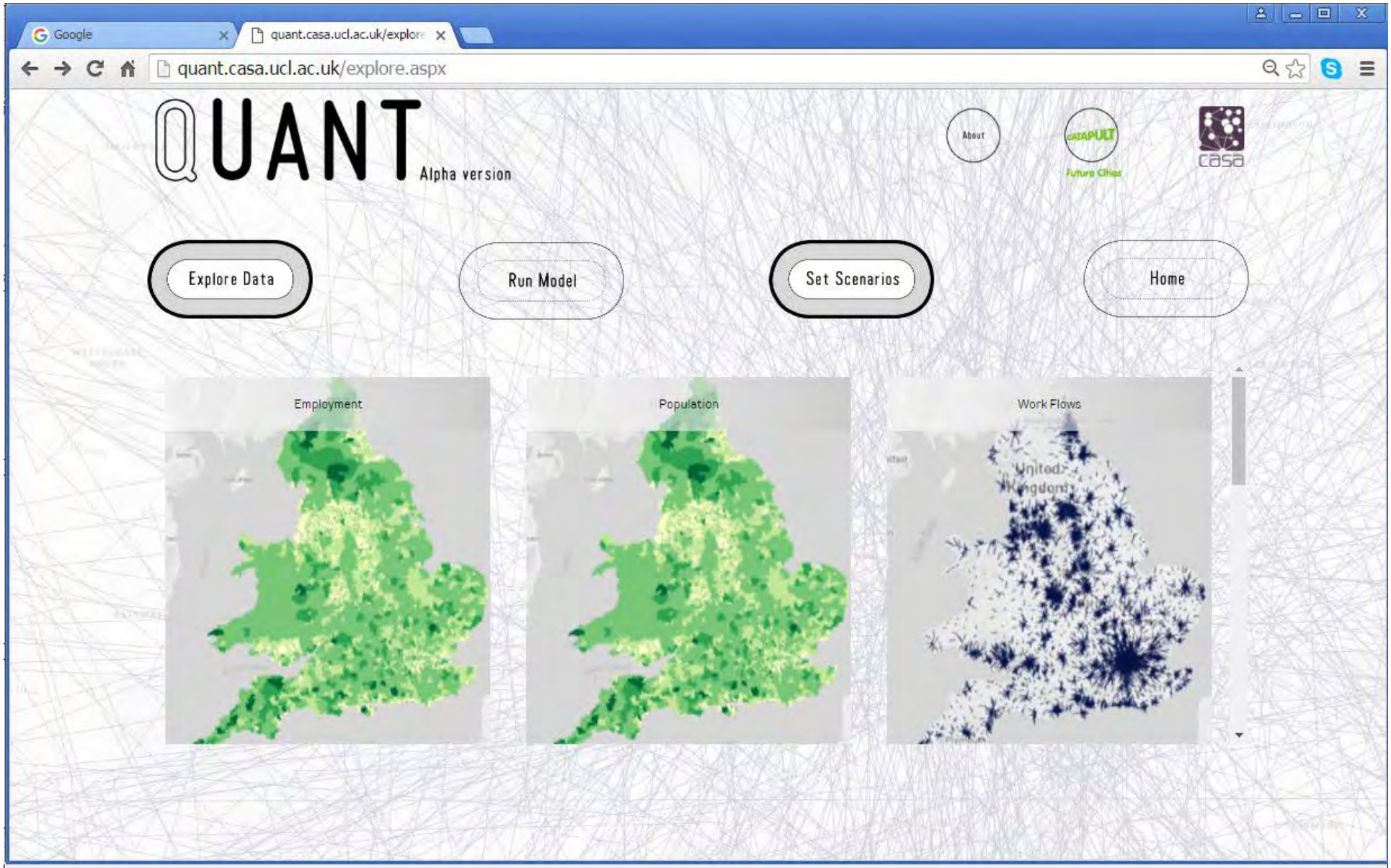
Urban Analytics: City Science, Prediction and Planning



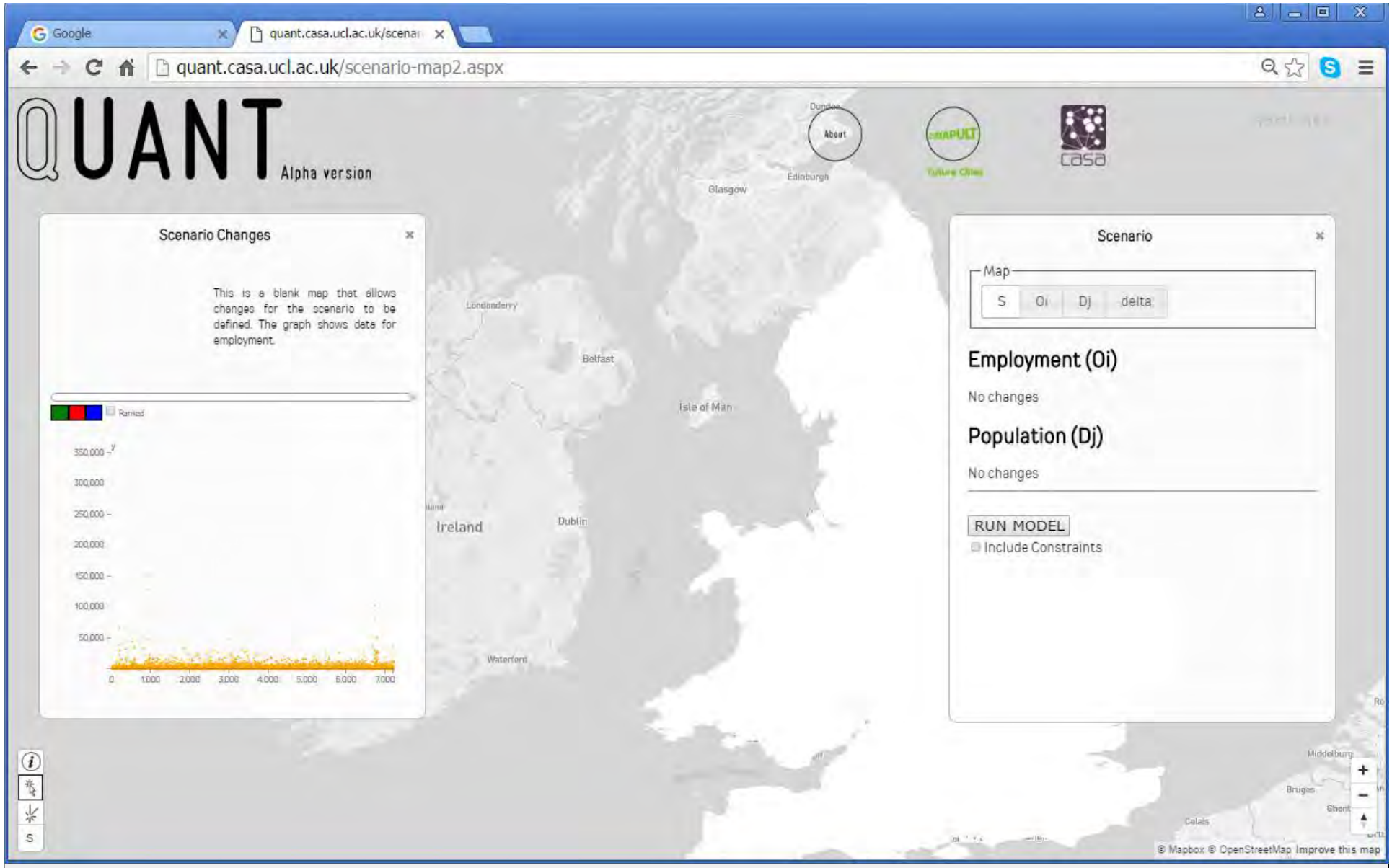
Urban Analytics: City Science, Prediction and Planning



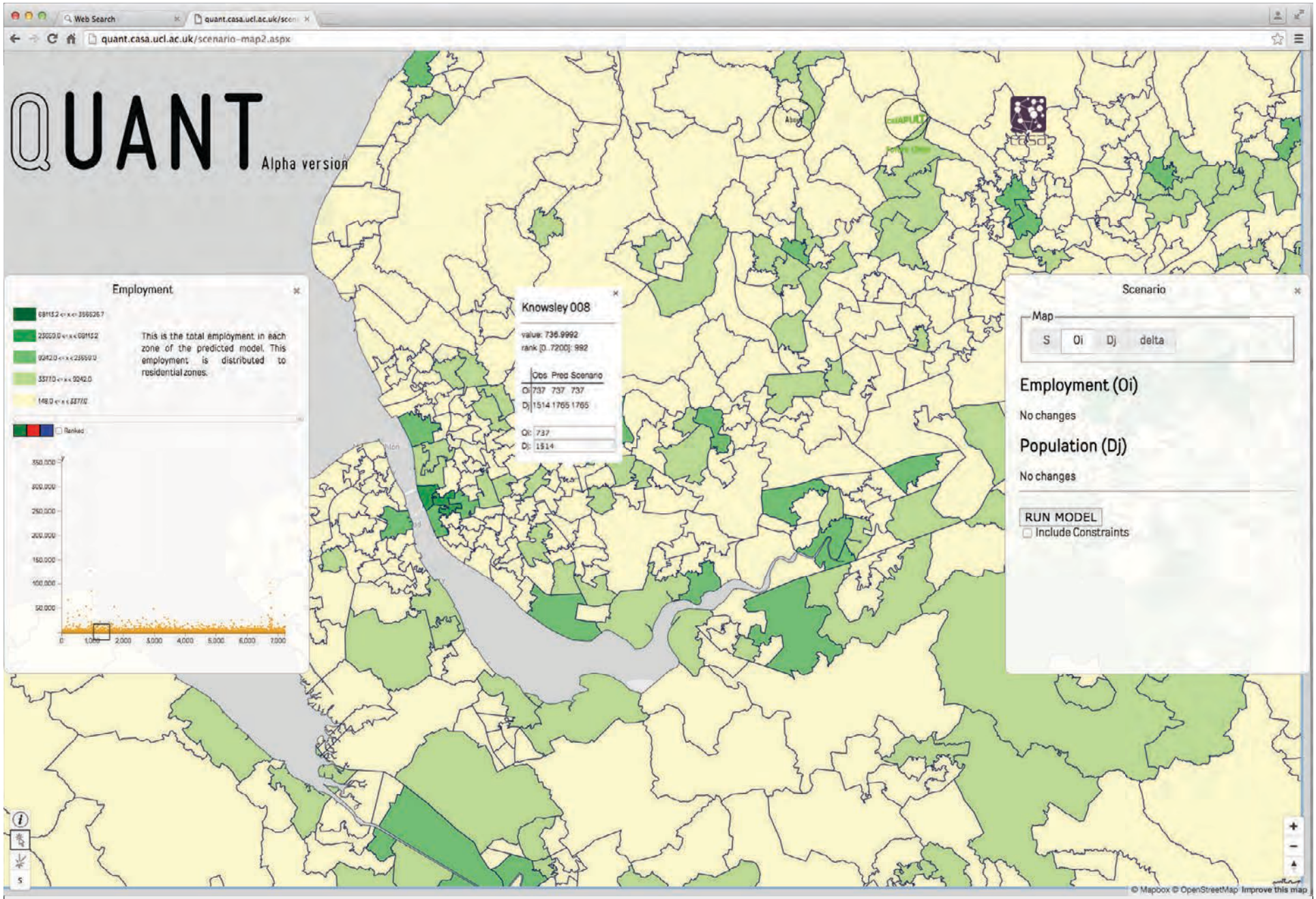
Urban Analytics: City Science, Prediction and Planning



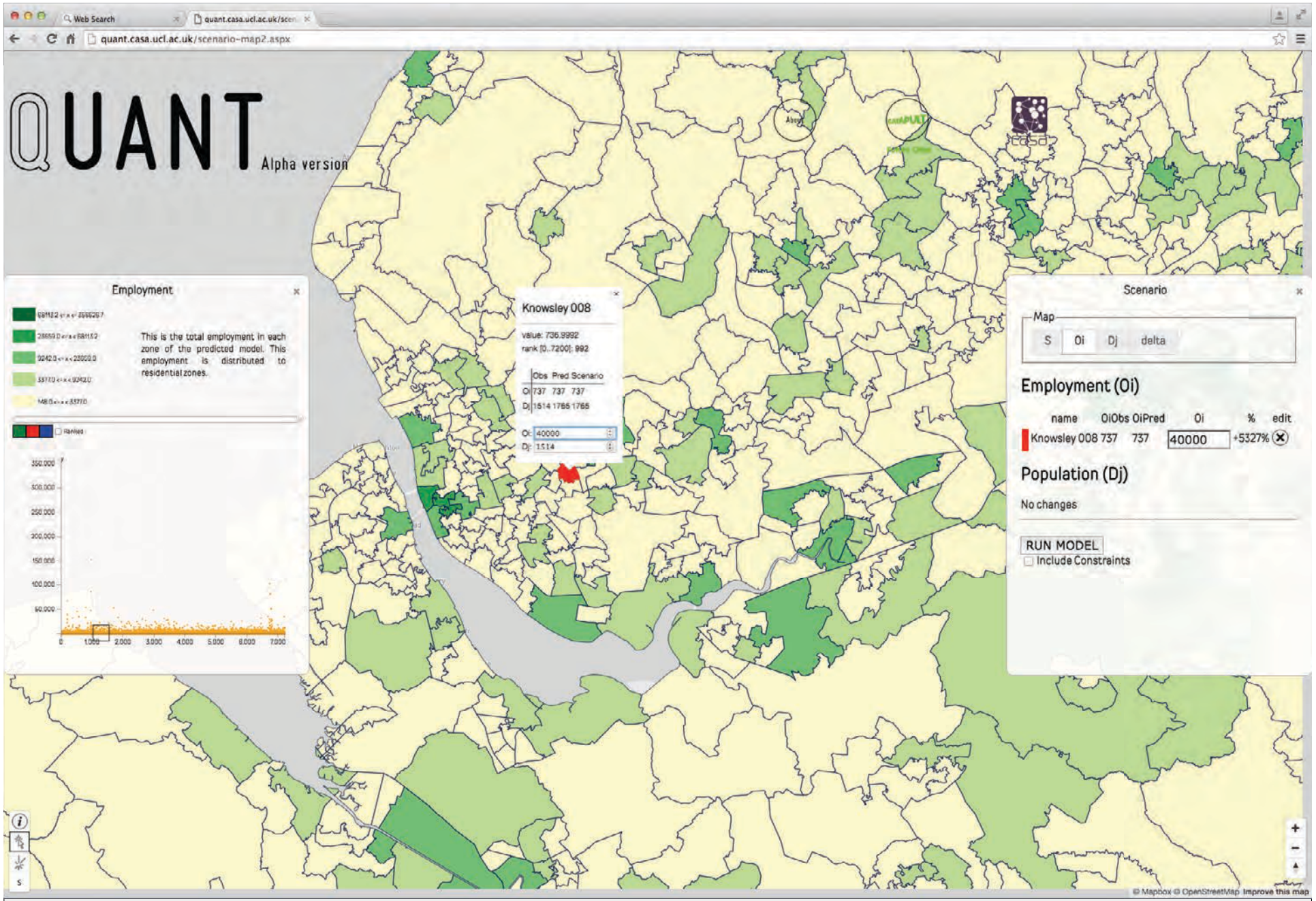
Urban Analytics: City Science, Prediction and Planning



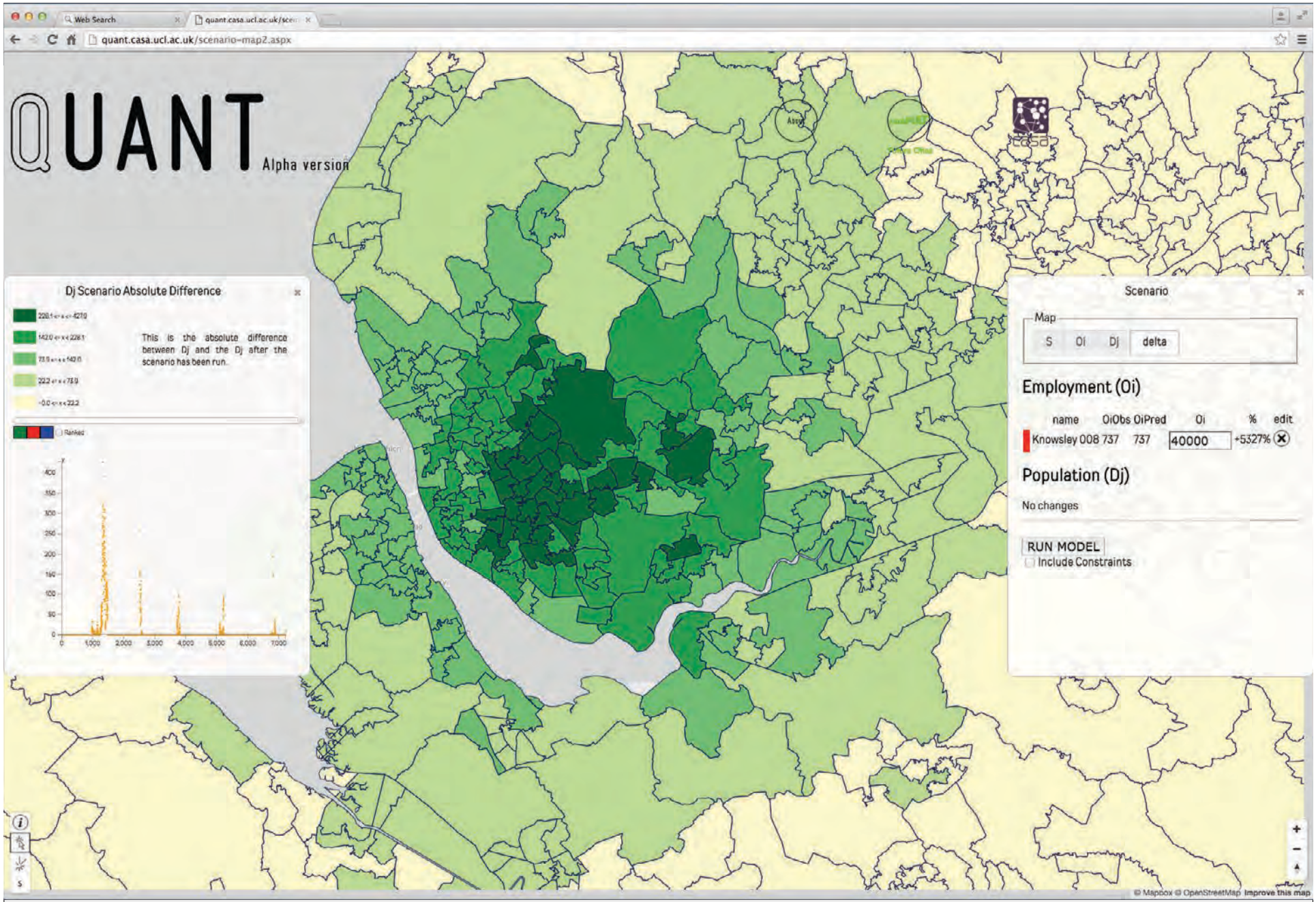
Urban Analytics: City Science, Prediction and Planning



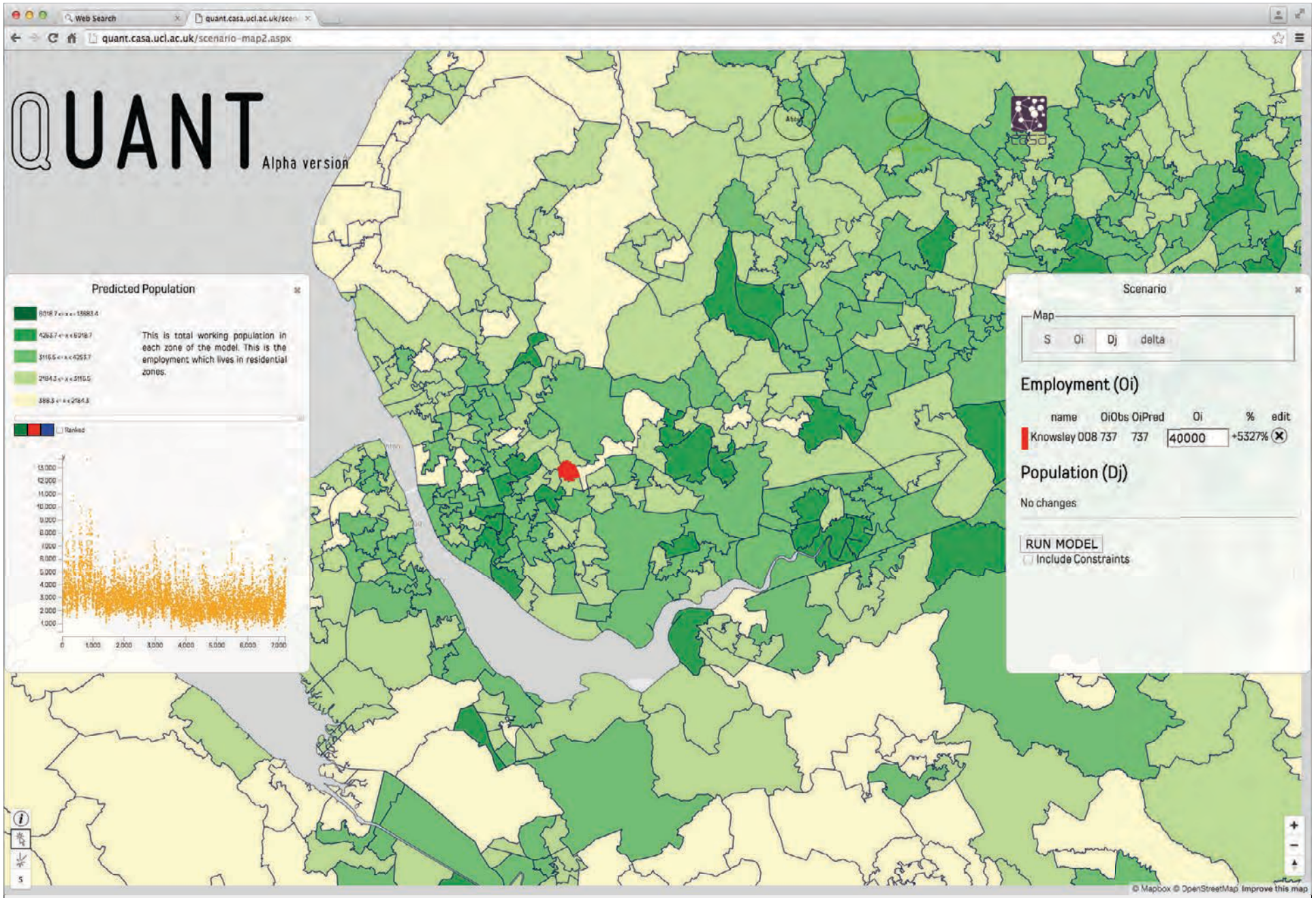
Urban Analytics: City Science, Prediction and Planning



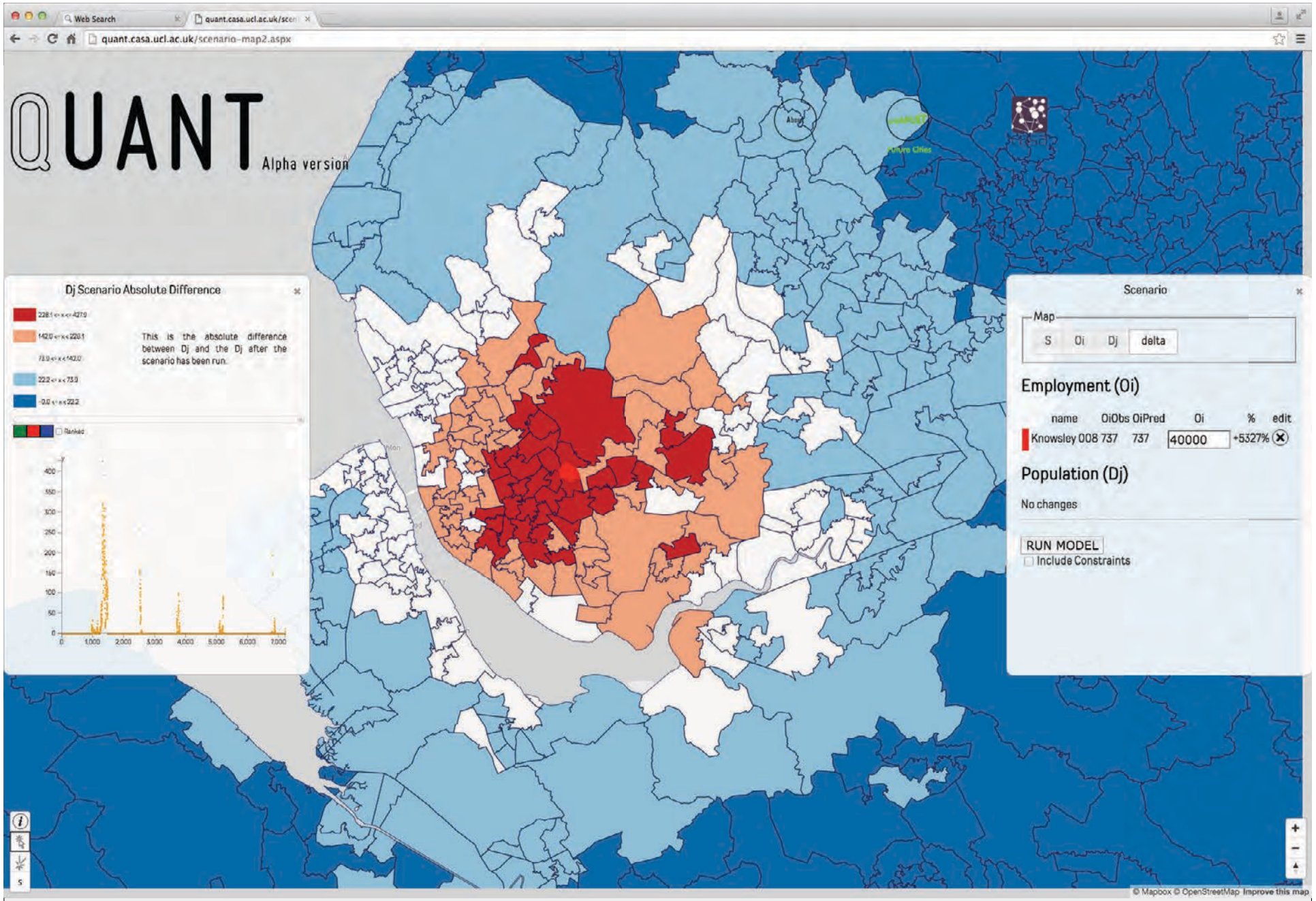
Urban Analytics: City Science, Prediction and Planning



Urban Analytics: City Science, Prediction and Planning



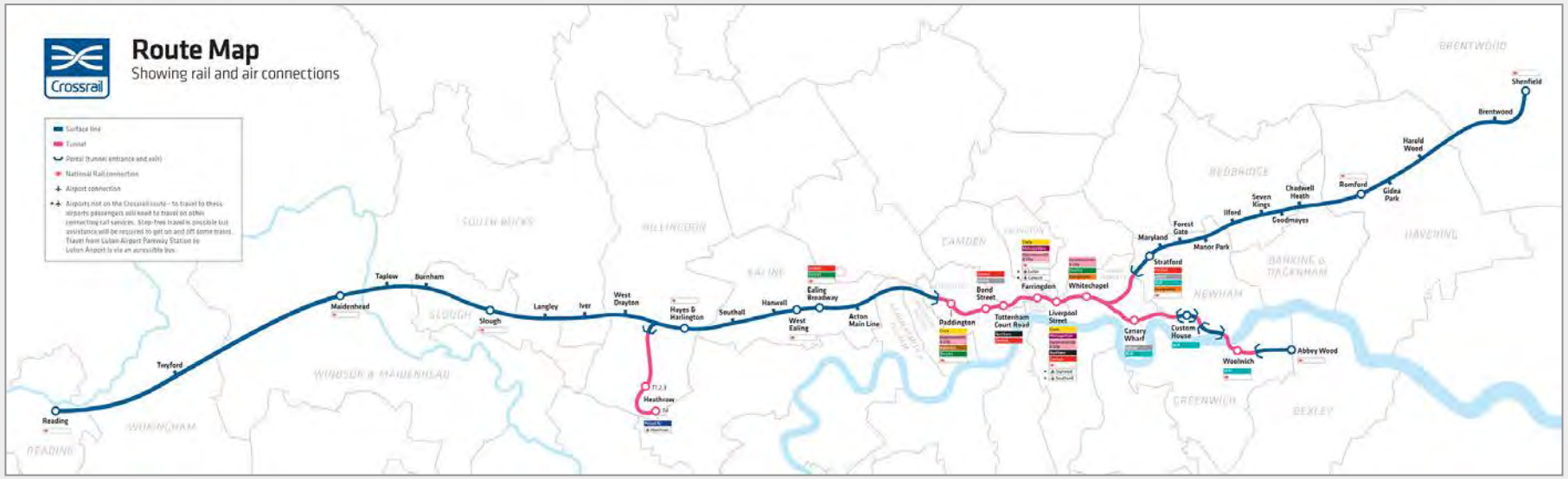
Urban Analytics: City Science, Prediction and Planning



Urban Analytics: City Science, Prediction and Planning

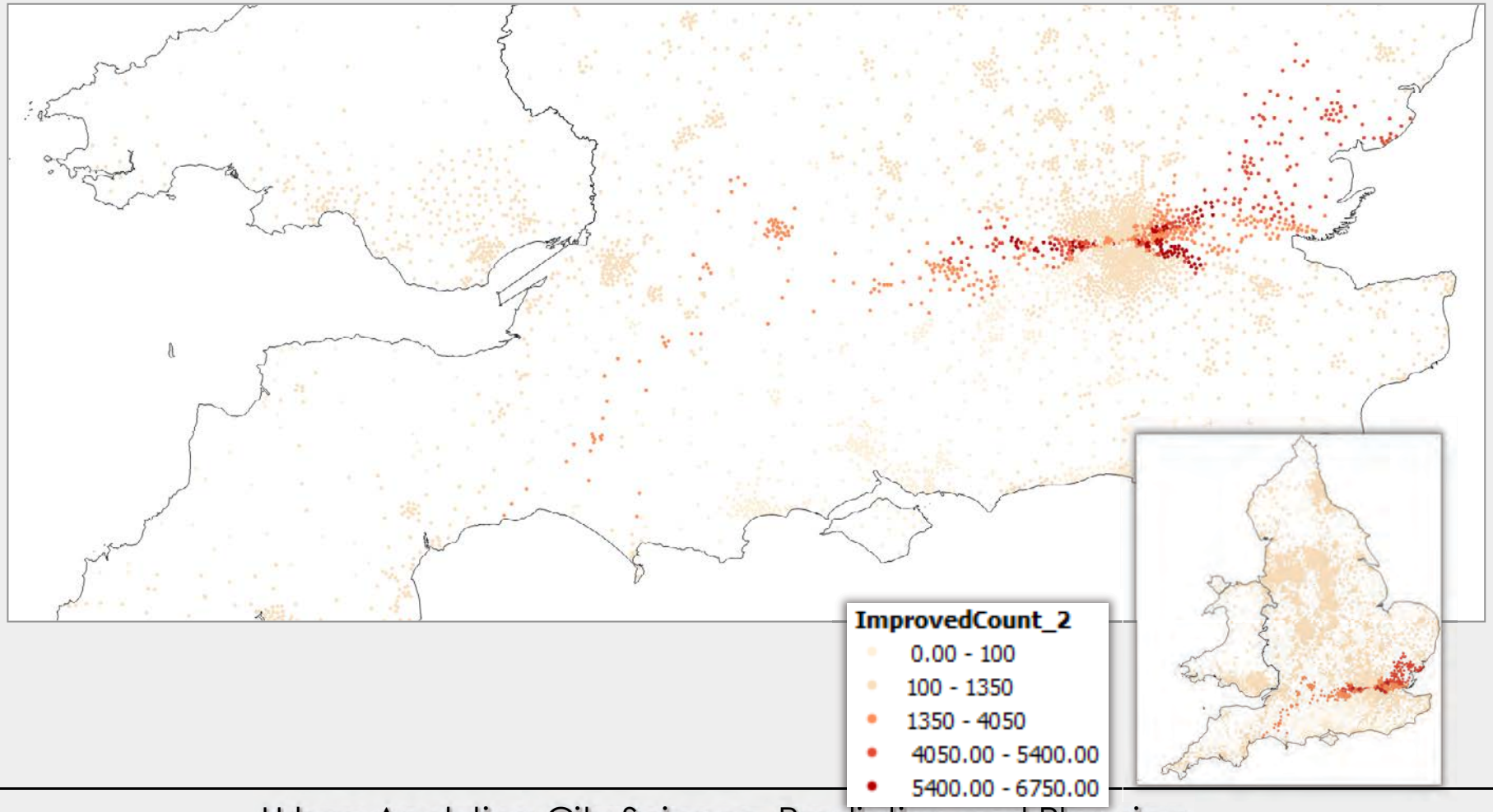
Crossrail

Reading, Heathrow, Shenfield, Abbey Wood



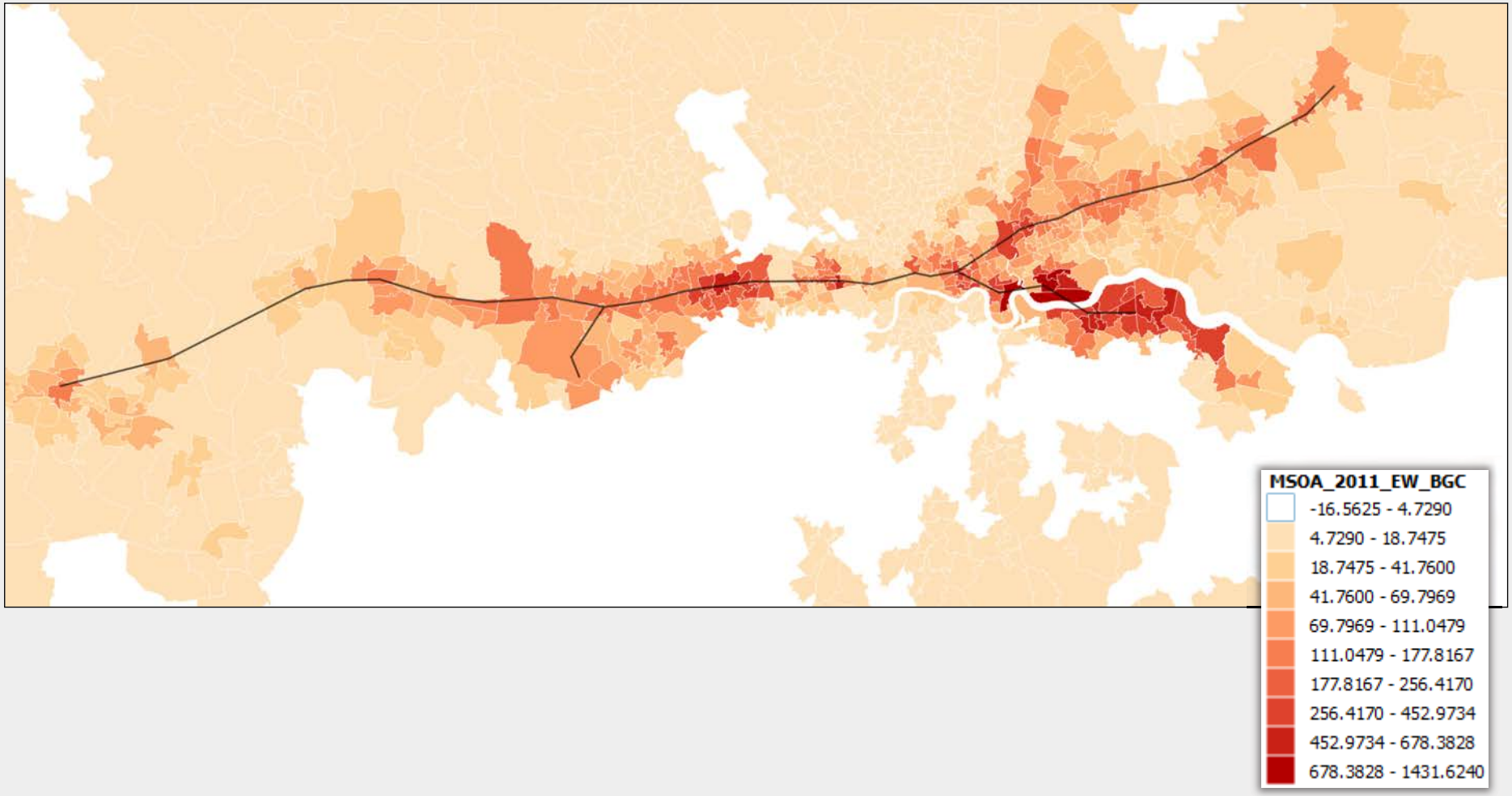
Crossrail

Number of Improved Journeys (ni)



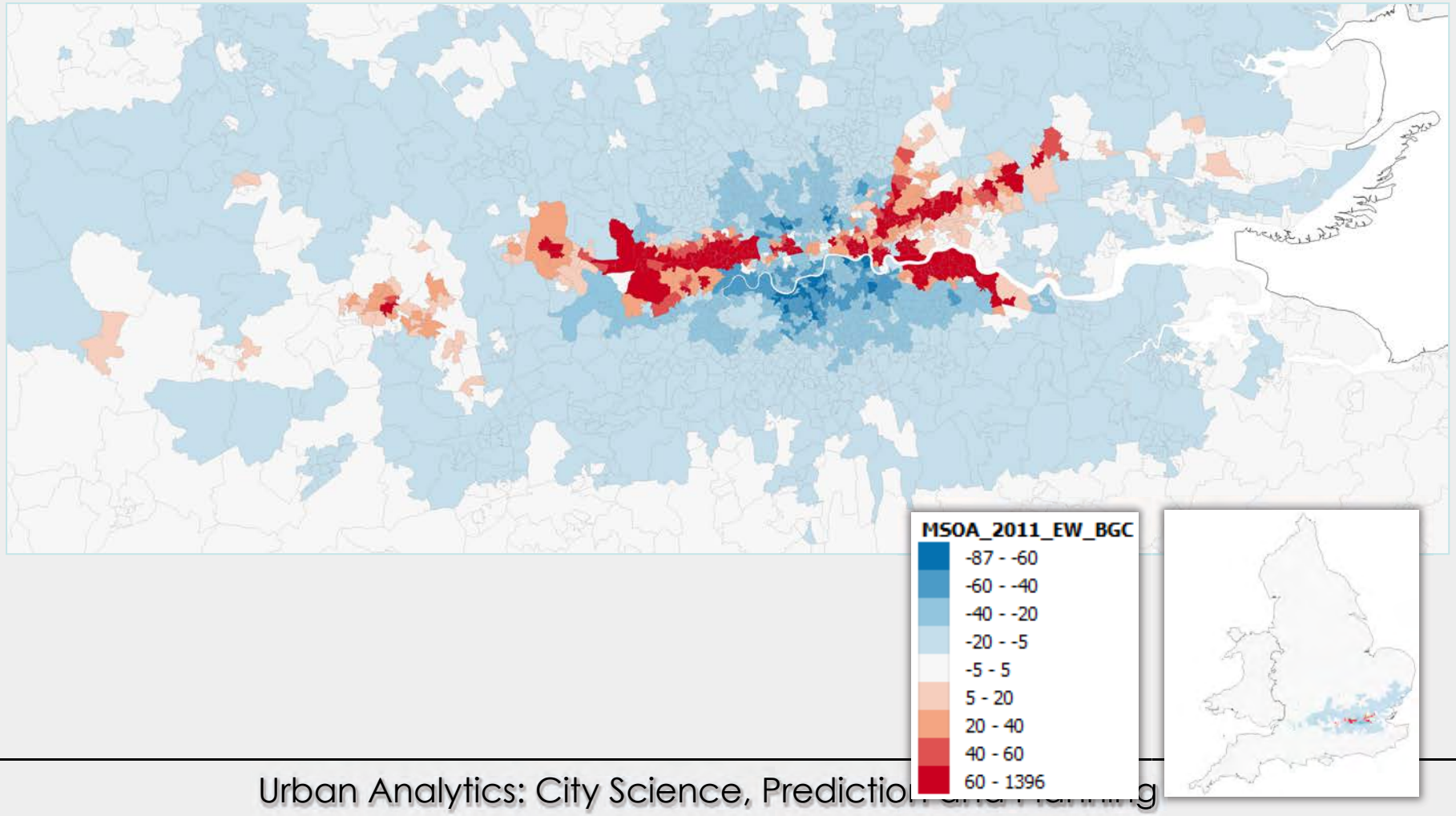
Crossrail

Population change (rail mode only)

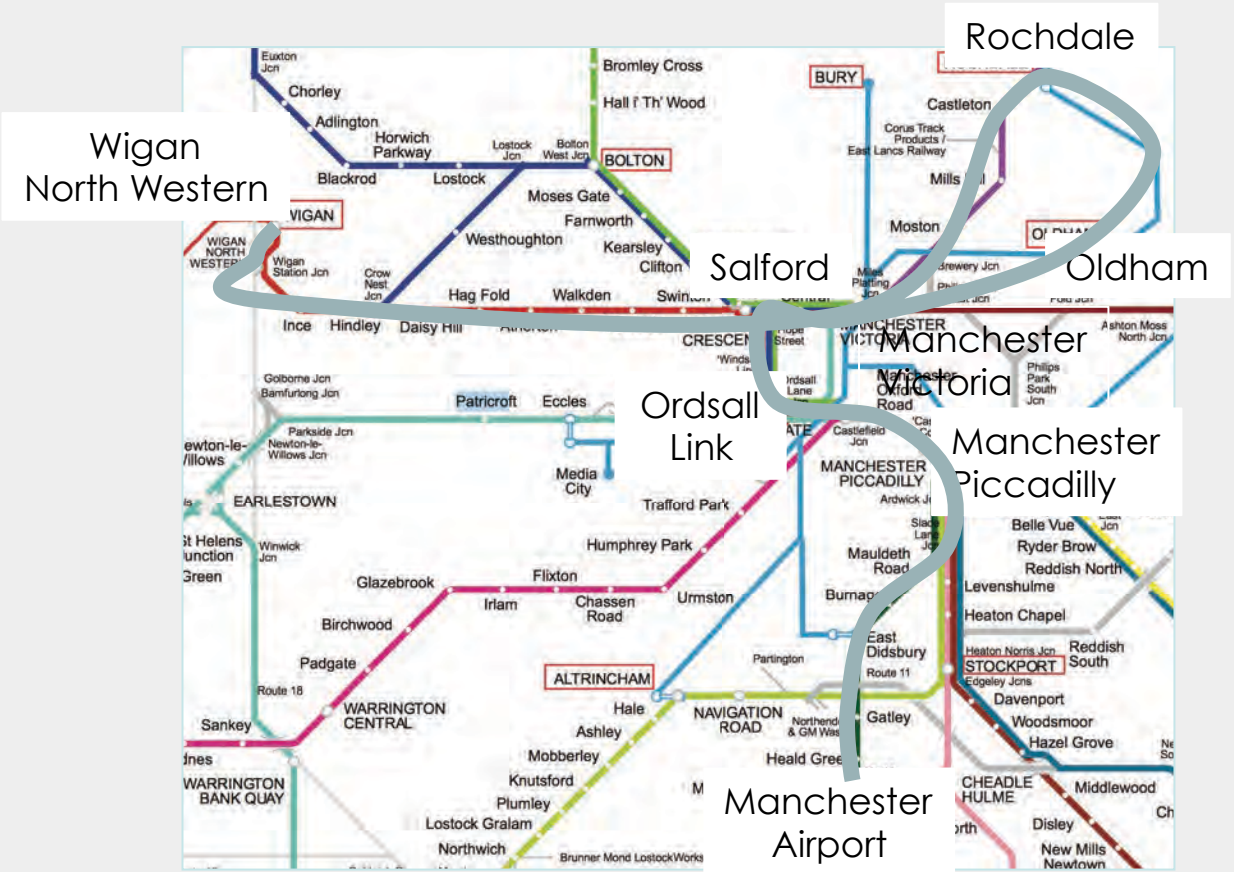


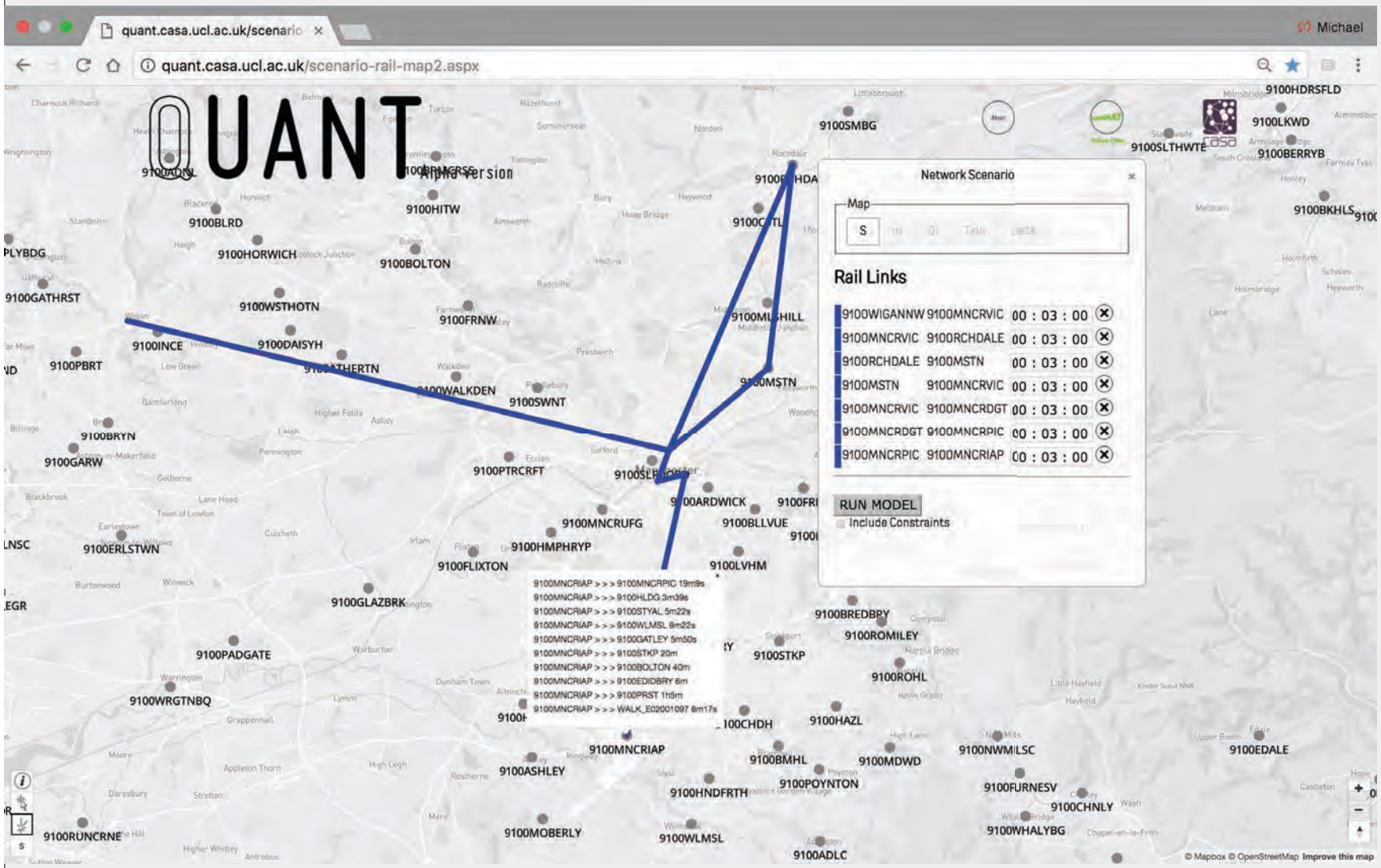
Crossrail

Population Change (all modes)

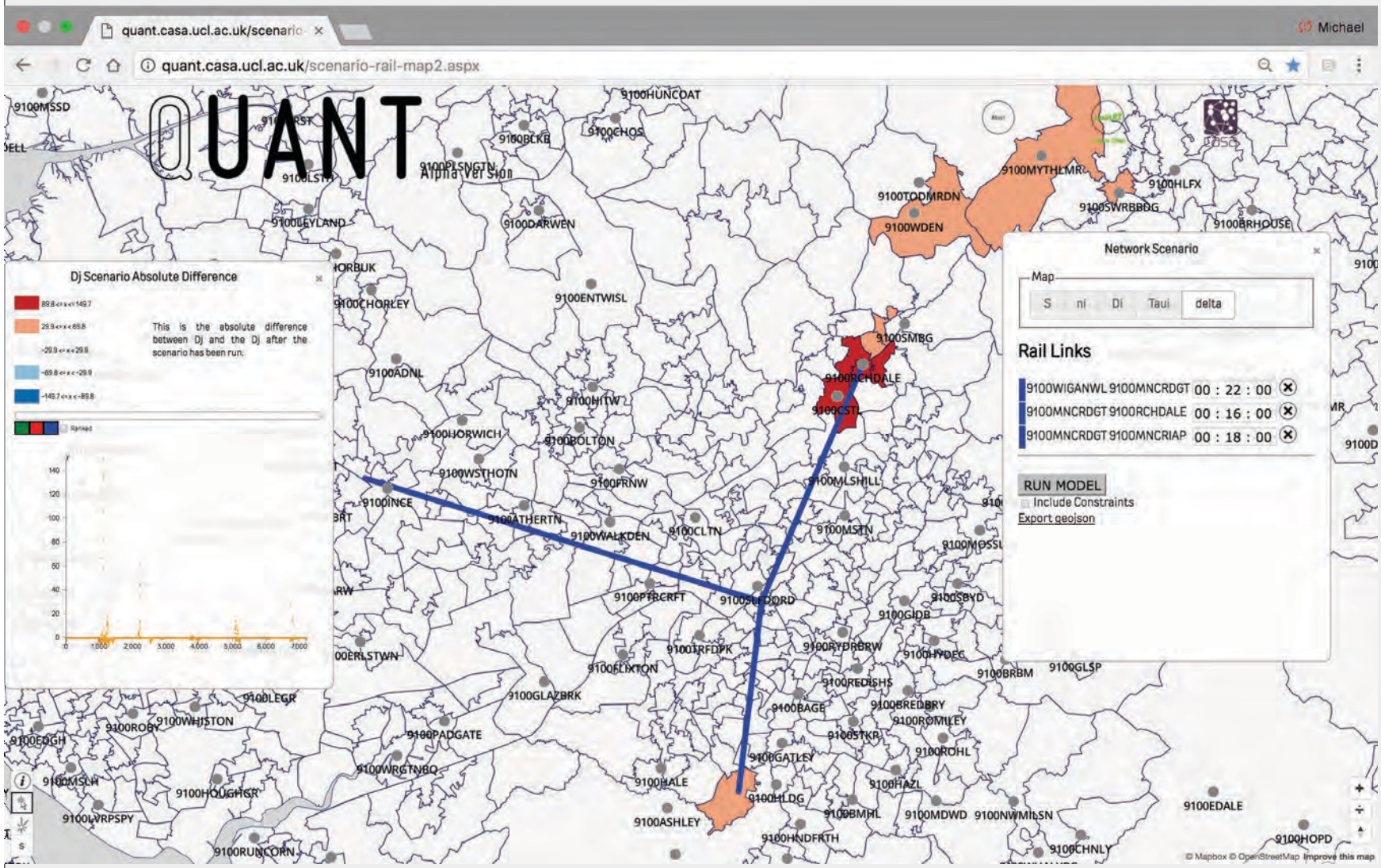



Rochdale
Wigan
North Western
Salford
Oldham
Manchester
Victoria
Ordsall
Link
Manchester
Piccadilly
Manchester
Airport

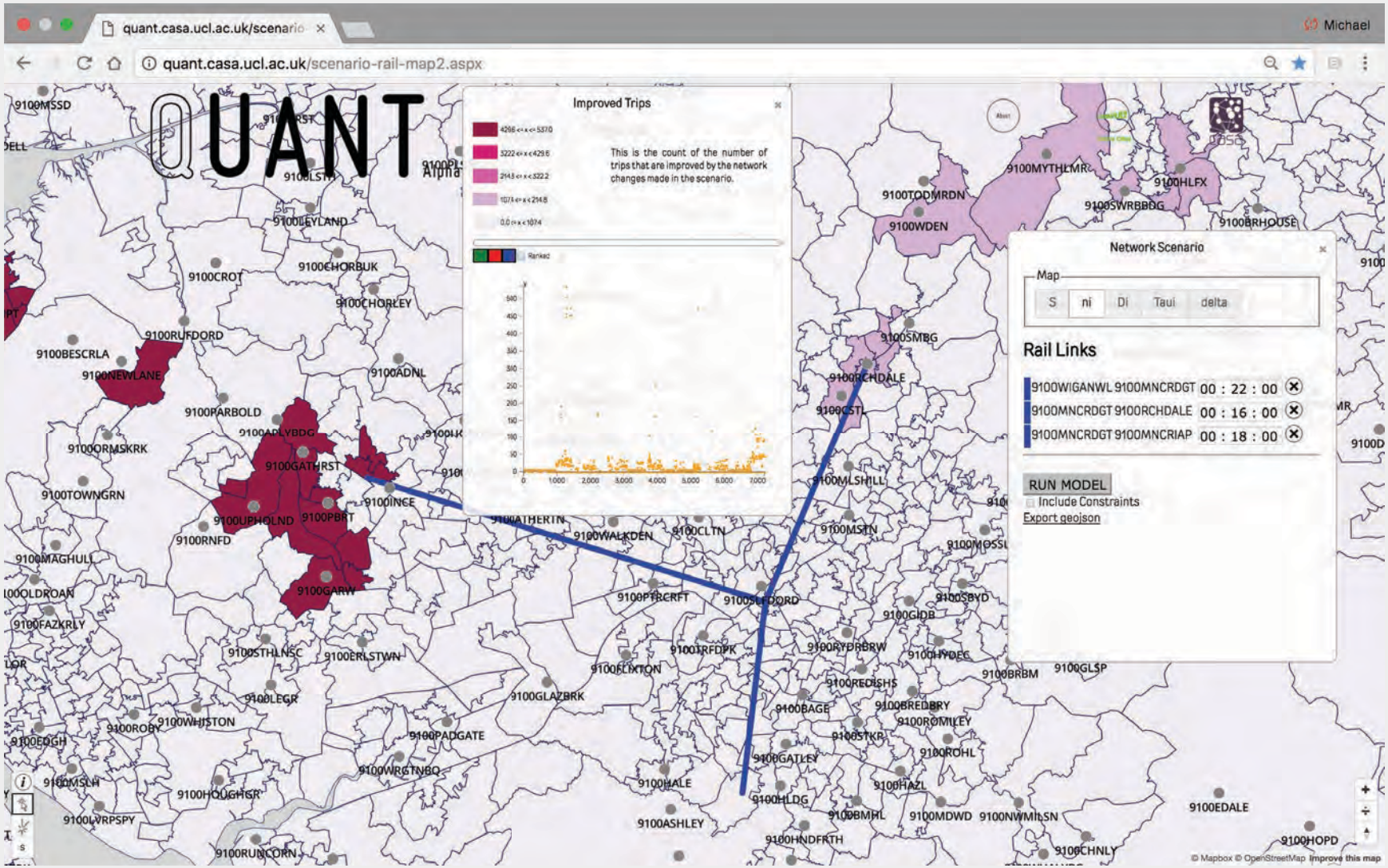

Rochdale
Wigan
North Western
Salford
Oldham
Manchester
Victoria
Ordsall
Link
Manchester
Piccadilly
Manchester
Airport
Euxton Jcn
Chorley
Adlington
Horwich Parkway
Blackrod
Lostock
Lostock Jcn
Bolton West Jcn
BOLTON
Bromley Cross
Hall i' Th' Wood
Moses Gate
Farnworth
Kearsley
Clifton
Westhoughton
Hag Fold
Walkden
Ince
Hindley
Daisy Hill
WIGAN
Wigan Station Jcn
Crow Nest Jcn
BURY
Castleton
Corus Track Products / East Lancs Railway
Mills
Moston
Brewery Jcn
Ashton Moss North Jcn
Philips Park South Jcn
Irlsall Lane
Castlefield Jcn
MANCHESTER PICCADILLY
Ardwick Jcn
Mauldeth Road
Burnage
East Didsbury
Route 11
Partington
Belle Vue
Ryder Brow
Reddish North
Levenshulme
Heaton Chapel
Heaton Norris Jcn
Reddish South
STOCKPORT
Edgeley Jcns
Davenport
Woodsmoor
Hazel Grove
Middlewood
Disley
New Mills Newtown
CHEADLE HULME
NAVIGATION ROAD
Gatley
Heald Green
Golborne Jcn
Bamfurlong Jcn
Patricroft
Eccles
Parkside Jcn
Newton-le-Willows Jcn
Media City
Trafford Park
Humphrey Park
Urmston
Chassen Road
Flixton
Irlam
Glazebrook
Birchwood
Padgate
Route 18
WARRINGTON CENTRAL
Sankey
EARLESTOWN
Winwick Jcn
St Helens Junction
Green
WARRINGTON BANK QUAY
ALTRINCHAM
Hale
Ashley
Mobberley
Knutsford
Plumley
Lostock Gralam
Northwich
Brunner Mond LostockWorks



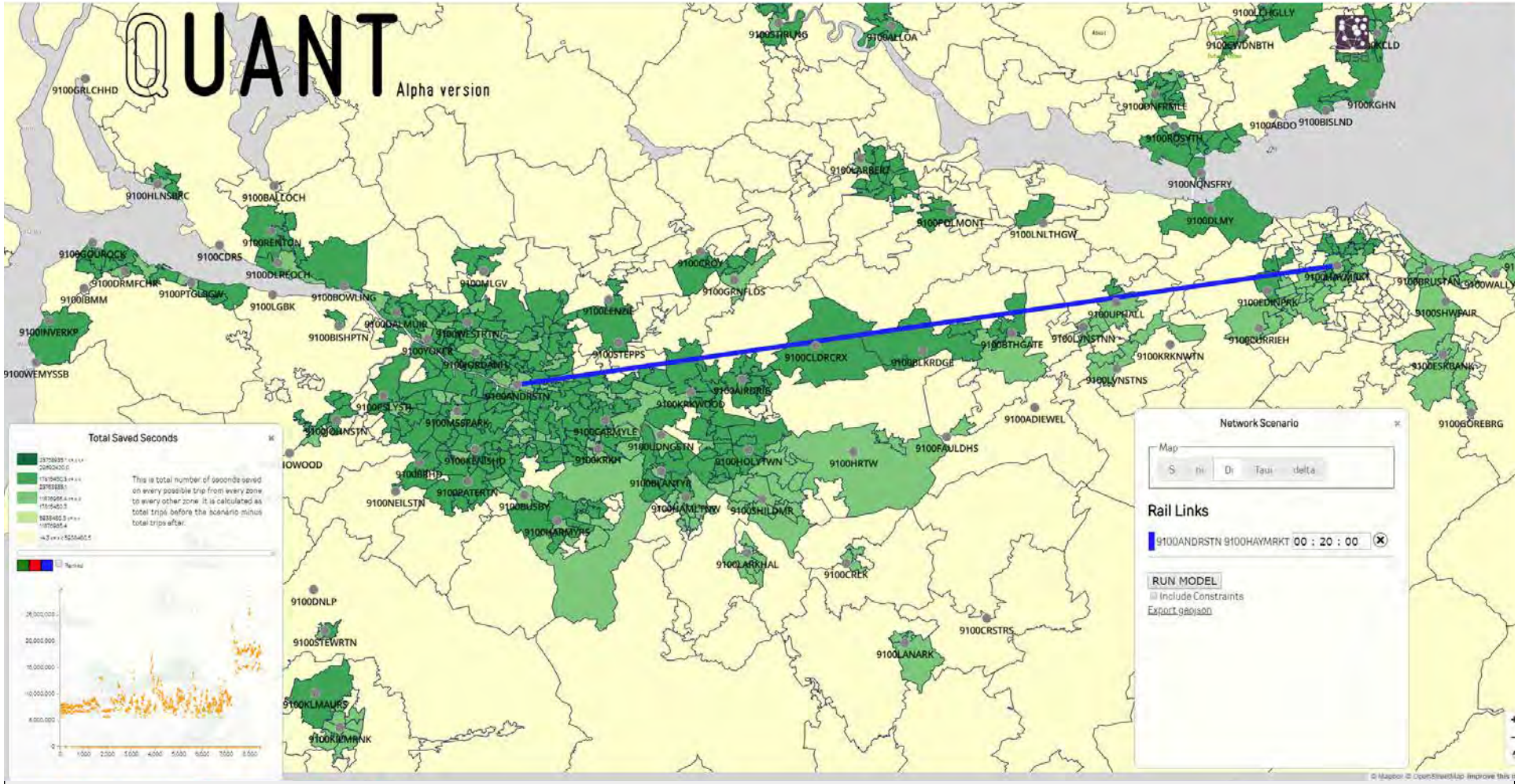
Urban Analytics: City Science, Prediction and Planning



Urban Analytics: City Science, Prediction and Planning



Urban Analytics: City Science, Prediction and Planning

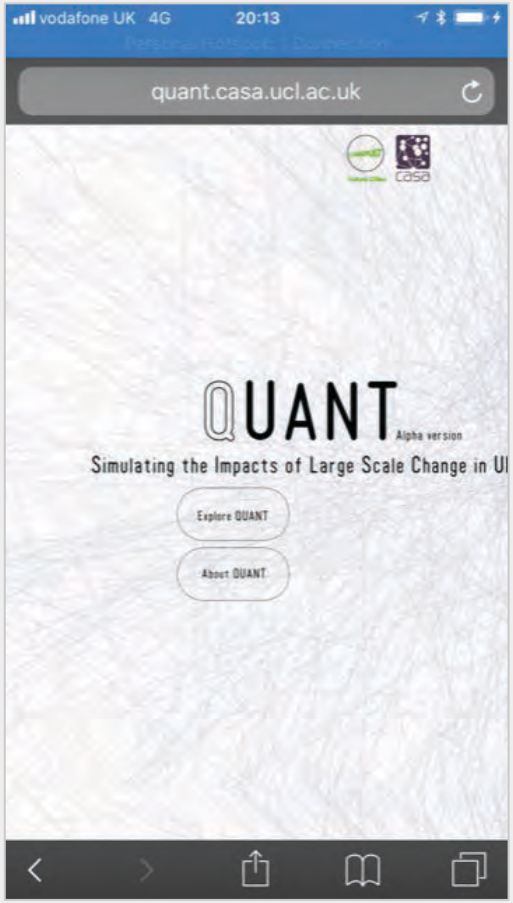


The High Speed Line from Glasgow to Edinburgh

Time between Glasgow-Edinburgh by train set to 20min.
The map shows the total number of seconds saved on every possible trip from any origin to any destination with the new line in place

Urban Analytics: City Science, Prediction and Planning

http://quant.casa.ucl.ac.uk/



m.batty@ucl.ac.uk
richard.milton@ucl.ac.uk

CATAPULT Future Cities

UKCRIC

UK Collaboratorium for Research in Infrastructure & Cities

Infrastructure and Projects Authority

National Infrastructure Commission

Where To From Here: Just at the Beginning

I cannot really summarise for there are countless kinds of urban analytics but let me make a few points to conclude

1. We need to use data much more, we need to build models that are close to data but are theories about the data and changes in data reflecting temporal dynamics
2. We need to build more than one model of the same phenomena – this is a longstanding notion and is intrinsic to many modelling applications but not yet to urban dynamics
3. We need to enquire into the representativeness of big data
4. I feel that much of what we are doing with such data is a fruitless quest because we need good theory to figure out when we have found something new in big data and to focus on where to look
5. We need to move towards crowdsourcing of model building, prediction and scenario setting as well as prescription
6. This means moving from desktop to web, employing more and more graphics and developing much better user interfaces

Draft

Data-Driven Urban Models

Michael Batty

Centre for Advanced Spatial Analysis (CASA)
University College London (UCL),
90 Tottenham Court Road, London W1N 6TR, UK

m.batty@ucl.ac.uk

28 July 2014

Cornell University

arXiv.org > cs > arXiv:1903.11205

Computer Science > Computers and Society

Airbnb's disruption of the housing structure in London

Zahratu Shabrina, Elsa Arcaute, Michael Batty

(Submitted on 27 Mar 2019 (v1), last revised 29 Mar 2019 (this version, v2))

This paper explores Airbnb, a peer-to-peer platform for short-term rental of housing accommodation, examining the geographical pattern of those establishments using data from London. Our purpose is to analyse whether or not the diversity of dwelling types correlate with the distribution of listings. We use a measure of spread based on entropy to indicate the diversity of dwelling types and look at its relationship with the distribution of Airbnb establishments, as well as the type of home ownership using correlation analysis. It is important to note that our study only considers domestic building types, and excludes any information on the diversity of land uses. Two important findings emerge from our analysis. Firstly, the spatial location of Airbnb rentals is negatively correlated with the diversity of dwelling types, and positively correlated with a single dwelling type, which corresponds in general to purpose built flats, conversions and flats in commercial buildings. Secondly, Airbnb is associated with areas that have a high proportion of privately rented properties, detracting more than 1.4% of the housing supply into short-term rentals. Such a phenomenon can reach up to 20% in some neighbourhoods, further exacerbating the process of gentrification. Finally, we discuss the implications of these findings as instruments to inform policies associated with the 'sharing' economy in relation to the disruption of the housing structure.

Subjects: Computers and Society (cs.CY); Physics and Society (physics.soc-ph)
Cite as: arXiv:1903.11205 [cs.CY]
(or arXiv:1903.11205v2 [cs.CY] for this version)

Download:
- PDF
- Other formats

Current browse context: cs.CY
< prev | next >
new | recent | 1903

Change to browse by:
cs
physics
physics.soc-ph

References & Citations
- NASA ADS

Google Scholar

Bookmark

Built Environment

Volume 42, Number 3, 2016

ISSN 0263-7960

Published by Alexandrine Press

Alexandrine Press
1 The Farthings
Marcham
Oxon OX13 6QD
Telephone: 01865 391518
Fax: 01865 391687
E-mail:
Alexandrine@tudkinassociates.co.uk

Built Environment is

BIG DATA AND THE CITY
Guest Editor: MICHAEL BATTY

PLOS ONE

RESEARCH ARTICLE

Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data

Chen Zhong[1]*, Michael Batty[1], Ed Manley[1], Jiaqiu Wang[1], Zijia Wang[2,3], Feng Chen[2,3], Gerhard Schmitt[4]

1 Centre for Advanced Spatial Analysis, University College London, London, United Kingdom, 2 School of Civil and Architectural Engineering, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing, P. R. China, 3 Beijing Engineering and Technology Research Centre of Rail Transit Line Safety and Disaster Prevention, No.3 Shangyuancun, Haidian District, Beijing, P. R. China, 4 Future Cities Laboratory, Department of Architecture, ETH Zürich, Zurich, Switzerland

* c.zhong@ucl.ac.uk

CrossMark

Environment and Planning A: Economy and Space

Journal Home | Browse Journal | Submit Paper | About | Subscribe

Table of Contents

Previous Issue | Volume 45 Issue 3, March 2013 | Next Issue

Visually-Driven Urban Simulation: Exploring Fast and Slow Change in Residential Location

Michael Batty

First Published January 1, 2013: pp. 532–552

Abstract
> Preview

http://quant.casa.ucl.ac.uk/

m.batty@ucl.ac.uk