Census migration and commuting data: why look at who moves where?

Mike Coombes

Centre for Urban & Regional Development Studies Newcastle University

ACKNOWLEDGEMENTS

CURDS colleagues Tony Champion and Colin Wymer

NHPAU project colleagues at Heriot-Watt & Manchester Universities

Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Access was courtesy of ESRC/JISC Census of Population Program

Content of the presentation

1] **Summary of this rich resource**

- some datasets offer types of information never available previously

Content of the presentation

1] **Summary of this rich resource**

- some datasets offer types of information never available previously

2] **Features of flow datasets, including aspects that make them complex**

- they are not readily analysed by standard software

Content of the presentation

1] **Summary of this rich resource**

- some datasets offer types of information never available previously

2] **Features of flow datasets, including aspects that make them complex**

- they are not readily analysed by standard software

3] **Contrasting example uses of the datasets**

- differing approaches to analyses of potential commercial interest

Content of the presentation

1] **Summary of this rich resource**

- some datasets offer types of information never available previously

2] **Features of flow datasets, including aspects that make them complex**

- they are not readily analysed by standard software

3] **Contrasting example uses of the datasets**

- differing approaches to analyses of potential commercial interest

4] **Results from one recent relevant research project**

- a case study of using the datasets to define housing market areas

Content of the presentation

1] **Summary of this rich resource**

- some datasets offer types of information never available previously

2] **Features of flow datasets, including aspects that make them complex**

- they are not readily analysed by standard software

3] **Contrasting example uses of the datasets**

- differing approaches to analyses of potential commercial interest

4] **Results from one recent relevant research project**

- a case study of using the datasets to define housing market areas

5] **Forward look at other possible applications of interest to business users**

1] Today's focus is open data: mine 'flow' data

CURDS

Newcastle University

ARE THE CENSUS MIGRATION AND COMMUTING FLOW DATASETS 'OPEN'?

ONS are required to protect the confidentiality of citizens and also employers
- they classify datasets according to the risk they pose, as they explain:

> Public tables available under Open Government Licence — OPEN? yes

> …Safeguarded…accessed … under terms and conditions — 'ish'

- **…Secure…only accessible in the ONS Virtual Microdata Laboratory — no**

[nb. Commissioned tables can be specified to not be confidential — yes]

1] Today's focus is open data: mine 'flow' data

CURDS Newcastle University

ARE THE CENSUS MIGRATION AND COMMUTING FLOW DATASETS 'OPEN'?

ONS are required to protect the confidentiality of citizens and also employers
- they classify datasets according to the risk they pose, as they explain:

> Public tables available under Open Government Licence — OPEN? yes

> …Safeguarded…accessed … under terms and conditions — 'ish'

➢ …Secure…only accessible in the ONS Virtual Microdata Laboratory — no

[nb. Commissioned tables can be specified to not be confidential — yes]

Geographical detail is the key issue in deciding how much risk a dataset poses
- the finer the geographical detail, the greater the risk – BUT for flow data the issue is magnified because flow data has 2 geographies: FROM & TO

! No time here to detail how 'open' is each dataset for each level of geography !

ONS thankfully do not devalue 2011 flow data by 'rounding' 1s & 2s (as in 2001)
- but the commuting data 2011 has a new complexity in Workplace Zones

2011 commuting and migration flow datasets

SOURCES OF THE CENSUS MIGRATION AND COMMUTING FLOW DATASETS

Census 2O11 collected data on three types of location:

1 present address (+ usual address for visitors to derive 1* usual residence)

2 previous address if the person lived somewhere different 12 months ago

3 workplace# (# in Scotland only: this can be place of education)

2011 commuting and migration flow datasets

SOURCES OF THE CENSUS MIGRATION AND COMMUTING FLOW DATASETS

Census 2O11 collected data on three types of location:

1 present address (+ usual address for visitors to derive 1* usual residence)

2 previous address if the person lived somewhere different 12 months ago

3 workplace# (# in Scotland only: this can be place of education)

IN BRIEF:

commuting flows are FROM 1* usual residence TO 3 workplace#

migration flows are FROM 2 previous address TO 1* usual residence

2011 commuting and migration flow datasets

SOURCES OF THE CENSUS MIGRATION AND COMMUTING FLOW DATASETS

Census 2O11 collected data on three types of location:

1 present address (+ usual address for visitors to derive 1* usual residence)

2 previous address if the person lived somewhere different 12 months ago

3 workplace# (# in Scotland only: this can be place of education)

IN BRIEF:

commuting flows are FROM 1* usual residence TO 3 workplace#

migration flows are FROM 2 previous address TO 1* usual residence

…but there are also other datasets including:

- **movement of people with a second address**
- **flows of students from non-term address to 1 present address (term-time)**
- **commuting data in which FROM = second address if this said to be for work**
- **migration data highlighting international migrants**

2] Features of commuting and migration data

CURDS

Newcastle University

SOME WAYS IN WHICH MIGRATION AND COMMUTING DATASETS DIFFER

non-flow datasets	commuting datasets	migration datasets
base: individual OR household	ONLY individual	MOSTLY individual
coverage: all people	all employed and so full labour market coverage (+ students in Scotland)	only people who moved – biased sample of all people/households
variables: huge range of variables, and many cross-tabulated	fewer options, but include: social grade, mode of travel, age by sex	limited options include: social grade, current housing tenure, age by sex

2] Features of commuting and migration data

CURDS Newcastle University

SOME WAYS IN WHICH MIGRATION AND COMMUTING DATASETS DIFFER

non-flow datasets	commuting datasets	migration datasets
base: individual OR household	ONLY individual	MOSTLY individual
coverage: all people	all employed and so full labour market coverage (+ students in Scotland)	only people who moved – biased sample of all people/households
variables: huge range of variables, and many cross-tabulated	fewer options, but include: social grade, mode of travel, age by sex	limited options include: social grade, current housing tenure, age by sex

BUT the most critical difference is that flow datasets have two geographies:

spreadsheets are unwieldy if 2 variables have many thousand alternatives

…specialist software can be required, depending on the analyses needed

Recent trends continue established patterns

COMMUTING

NON-Census data shows the average worker commutes fewer times per week

… relates to more flexible working including part-time (home-)working

The distance commuted: the contrasts are north:south as well as urban:rural

…'driven' mainly by how well off local people are (*then* nearness of jobs)

SO areas where people increasingly commute further tend to have seen inward migration of better off people prone to commute further

Recent trends continue established patterns

COMMUTING

NON-Census data shows the average worker commutes fewer times per week
… relates to more flexible working including part-time (home-)working

The distance commuted: the contrasts are north:south as well as urban:rural
…'driven' mainly by how well off local people are (*then* nearness of jobs)

SO areas where people increasingly commute further tend to have seen inward migration of better off people prone to commute further

MIGRATION

Taking into account the increases among types of people more likely to move (eg well-educated, private renting) there is a long-term decline in migration

Most migration is short distance (eg. <20km), and these are the declining flows
SO declining move rates are due to housing and neighbourhood reasons

Recent trends continue established patterns

CURDS Newcastle University

COMMUTING

NON-Census data shows the average worker commutes fewer times per week
… relates to more flexible working including part-time (home-)working

The distance commuted: the contrasts are north:south as well as urban:rural
…'driven' mainly by how well off local people are (*then* nearness of jobs)

SO areas where people increasingly commute further tend to have seen inward migration of better off people prone to commute further

MIGRATION

Taking into account the increases among types of people more likely to move (eg well-educated, private renting) there is a long-term decline in migration

Most migration is short distance (eg. <20km), and these are the declining flows
SO declining move rates are due to housing and neighbourhood reasons

MIGRATION AND COMMUTING

The more complex working practices, with fewer but longer journeys to work, are related to rising reluctance to migrate, with the reasons including the multi-earner households' complex travel patterns

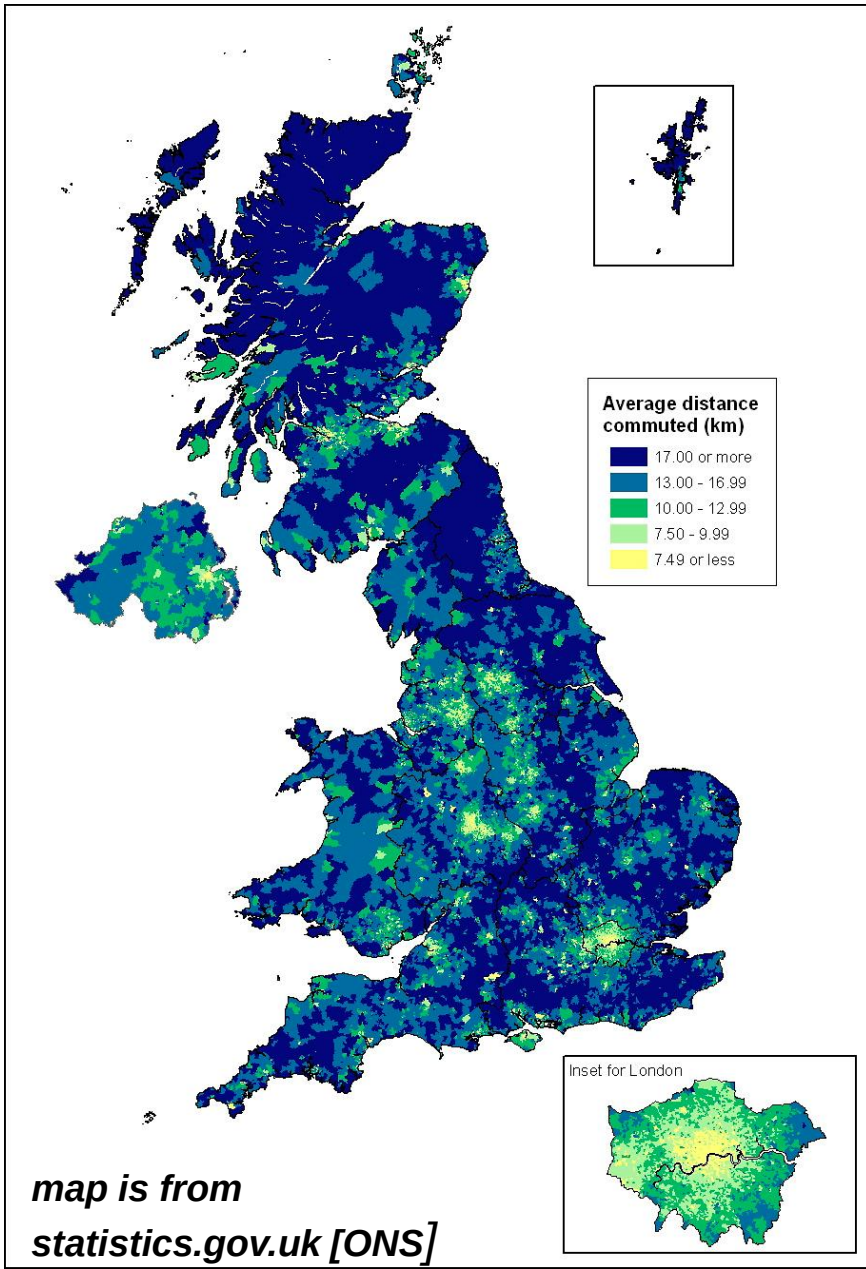
The pattern of longer distance commuting

The trends in survey data show a continued increase in the minority of workers who have longer distance journeys to work
(nb. as yet no Census local data for all UK)

This map of 2001 shows the areas (in blue) where the bulk of full-time workers travel more than 13km [*c.* 8 miles] to their work

Whilst most of Britain's area is blue by far the bulk of people live in the green/yellow areas with relatively short commuting trips: these are mostly towns or cities

This contrast is unchanged in the long-term

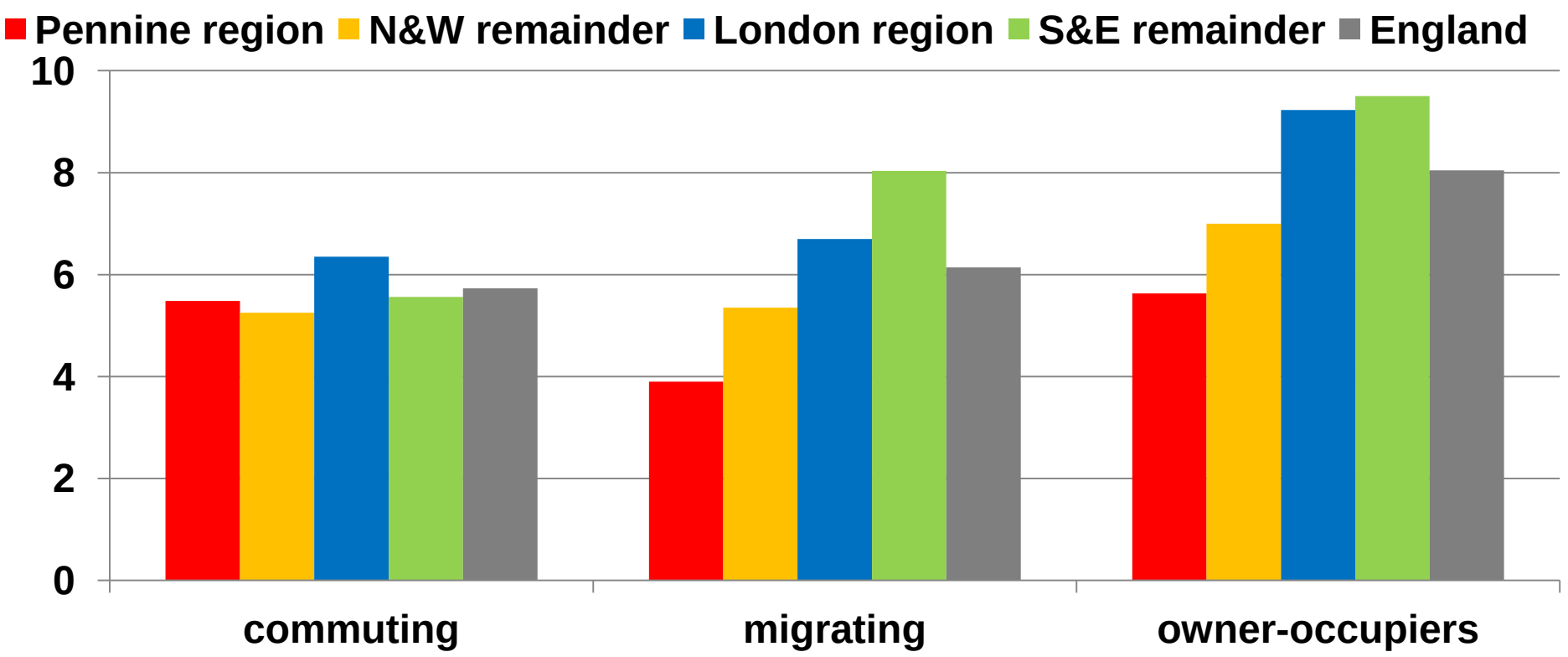


map is from statistics.gov.uk [ONS]

Are commuting and migration distances similar?

Distance commuted varies between urban and rural areas, and less by region, but migration distances vary markedly by broad region of the country

Distance commuted/migrated: mean of median flow (in/out) to constituent wards



3] Contrasting example uses of the datasets

Flow datasets can measure the relative strength of relationships between areas

… which is relevant to marketing because access is vital for local markets, and the arbitrage process that sets prices is shaped by geography

3] Contrasting example uses of the datasets

Flow datasets can measure the relative strength of relationships between areas

… which is relevant to marketing because access is vital for local markets, and the arbitrage process that sets prices is shaped by geography

The most familiar analysis uses flow data – perhaps from a form of transaction – to indicate the 'catchment area' of a particular location: this has limitations…

? how to draw the catchment area boundary (which flows are 'significant') ?

? what about competing locations, or nearby areas that are under-served ?

3] Contrasting example uses of the datasets

Flow datasets can measure the relative strength of relationships between areas
… which is relevant to marketing because access is vital for local markets, and the arbitrage process that sets prices is shaped by geography

The most familiar analysis uses flow data – perhaps from a form of transaction – to indicate the 'catchment area' of a particular location: this has limitations…

? how to draw the catchment area boundary (which flows are 'significant') ?

? what about competing locations, or nearby areas that are under-served ?

An alternative perspective requires the national territory to be divided into areas that are similar in their level of 'coherence' of the pattern of flows analysed: such analyses are not straightforward and most methods rather 'black-box'

?perhaps partly due to this difficulty, such analyses are mostly academic, although the partition of territory in this way also meets policy needs
eg. Census data used since 1970s to define Travel-to-Work Areas (TTWAs)

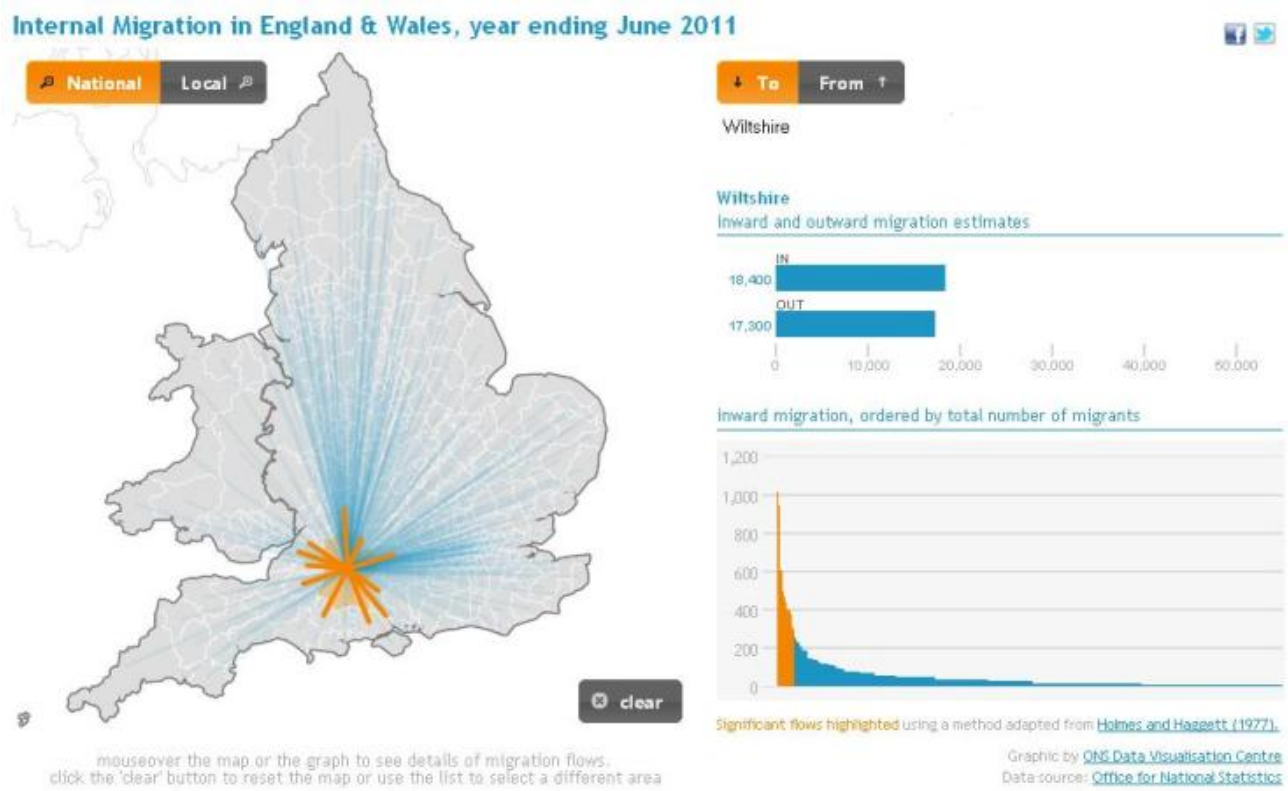
Patterns of flows around a selected location

The previous speaker covered catchment area modelling, and the potential role of commuting data from the Census to enhance existing approaches

Here this approach to analysing flow data is simply illustrated with a sample from one of the increasingly available visualisation tools provided by ONS, with this dataset migration flows rather than the more familiar commuting

www.ons.gov.uk/ons/dcp171778_280054.pdf

(Figure 6: Screenshot of Interactive Map of Internal Migration Flows in England and Wales -Year ending June 2011)



Dividing territories into areas: the TTWA method

This is an internationally recognised model for labour market area definitions

The key features of the TTWA method are instrumental for defining areas

- **all 'building block' zones to be in (only) 1 TTWA: no omissions or overlaps**
- **group the zones 1-by-1: not a top down process**
- **no 'centres' so no catchment areas: polycentric TTWAs emerge if they exist**
- **no contiguity constraint (but non-contiguous groupings are in fact rare)**

Dividing territories into areas: the TTWA method

This is an internationally recognised model for labour market area definitions

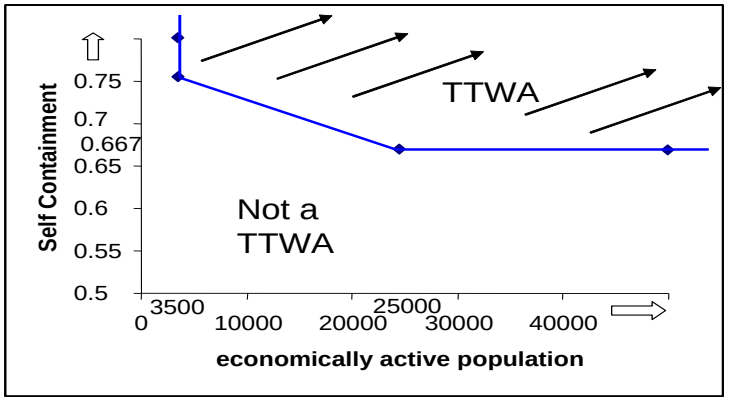
The key features of the TTWA method are instrumental for defining areas

- **all 'building block' zones to be in (only) 1 TTWA: no omissions or overlaps**
- **group the zones 1-by-1: not a top down process**
- **no 'centres' so no catchment areas: polycentric TTWAs emerge if they exist**
- **no contiguity constraint (but non-contiguous groupings are in fact rare)**

The most important measure is area "self-containment" in terms of flows…

eg. in commuting analyses: % of jobs in the area held by local residents AND % of employed residents working in the area

Can set different self-containment criteria, but the method combines these with some concern to avoid very small areas because they can have unreliable data, with some 'trade off' between criteria:



Brief visualisation excerpt from an analysis

Maps show state-of-play at selected steps in the grouping process as it gradually reduces the number of 'proto'TTWAs, sequentially identifying and 'disbanding' the one that is furthest from meeting the set criteria, then reassigning its zones

Here the 'proto'TTWAs are shown at stages in the process when in the whole UK there were:
8,000 4,000 2,000 1,000 500 & 250 (by when nearly all meet the final criteria)

Brief visualisation excerpt from an analysis

Maps show state-of-play at selected steps in the grouping process as it gradually reduces the number of 'proto'TTWAs, sequentially identifying and 'disbanding' the one that is furthest from meeting the set criteria, then reassigning its zones

Here the 'proto'TTWAs are shown at stages in the process when in the whole UK there were:
8,000 4,000 2,000 1,000 500 & 250 (by when nearly all meet the final criteria)

On these maps (showing central southern England, from an analysis of the UK):

the blue boundaries are the final 2001-based TTWAs

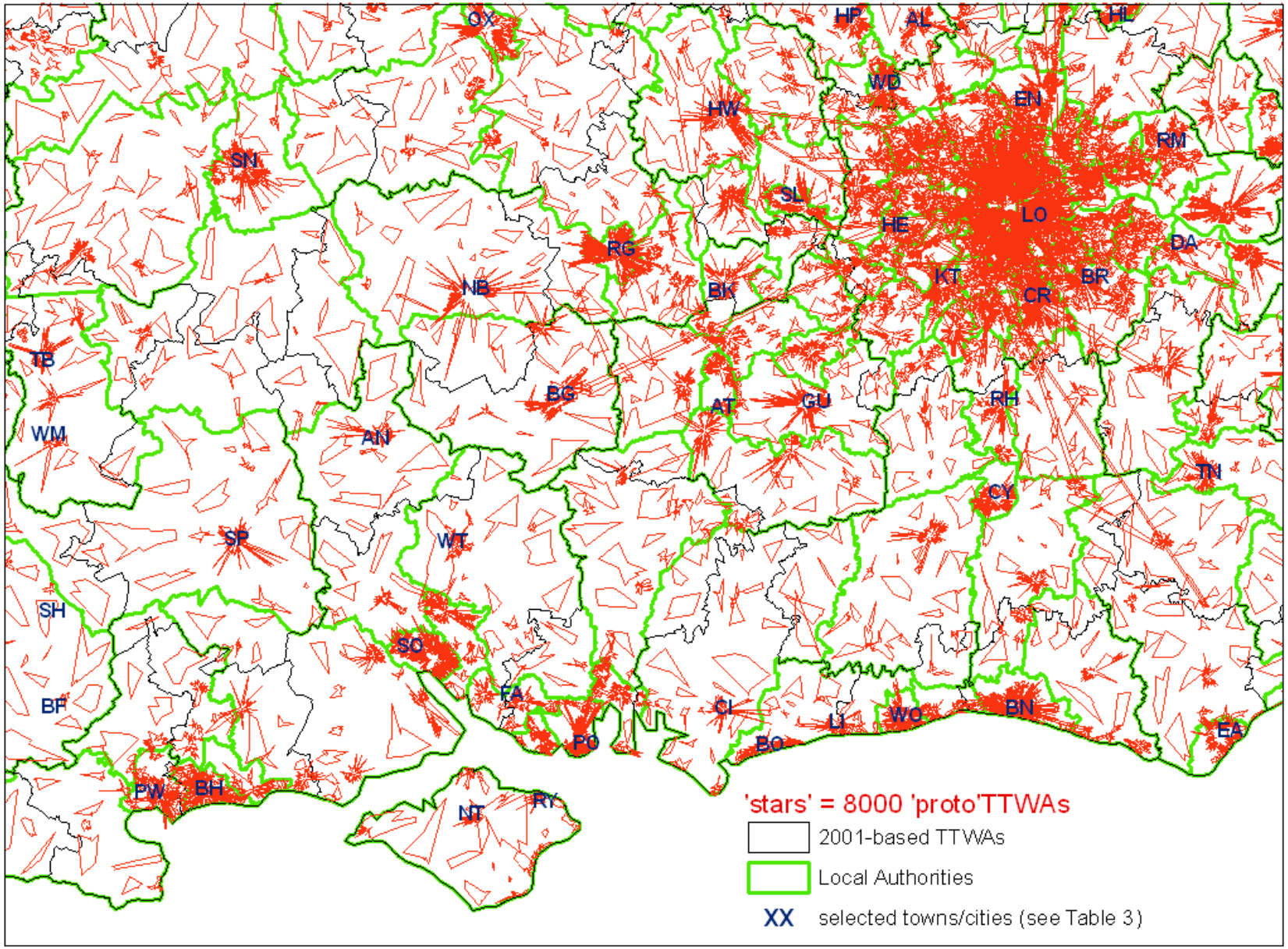
the green boundaries are Local Authority areas

the red 'stars' connect the LSOAs grouped together at that step of the process

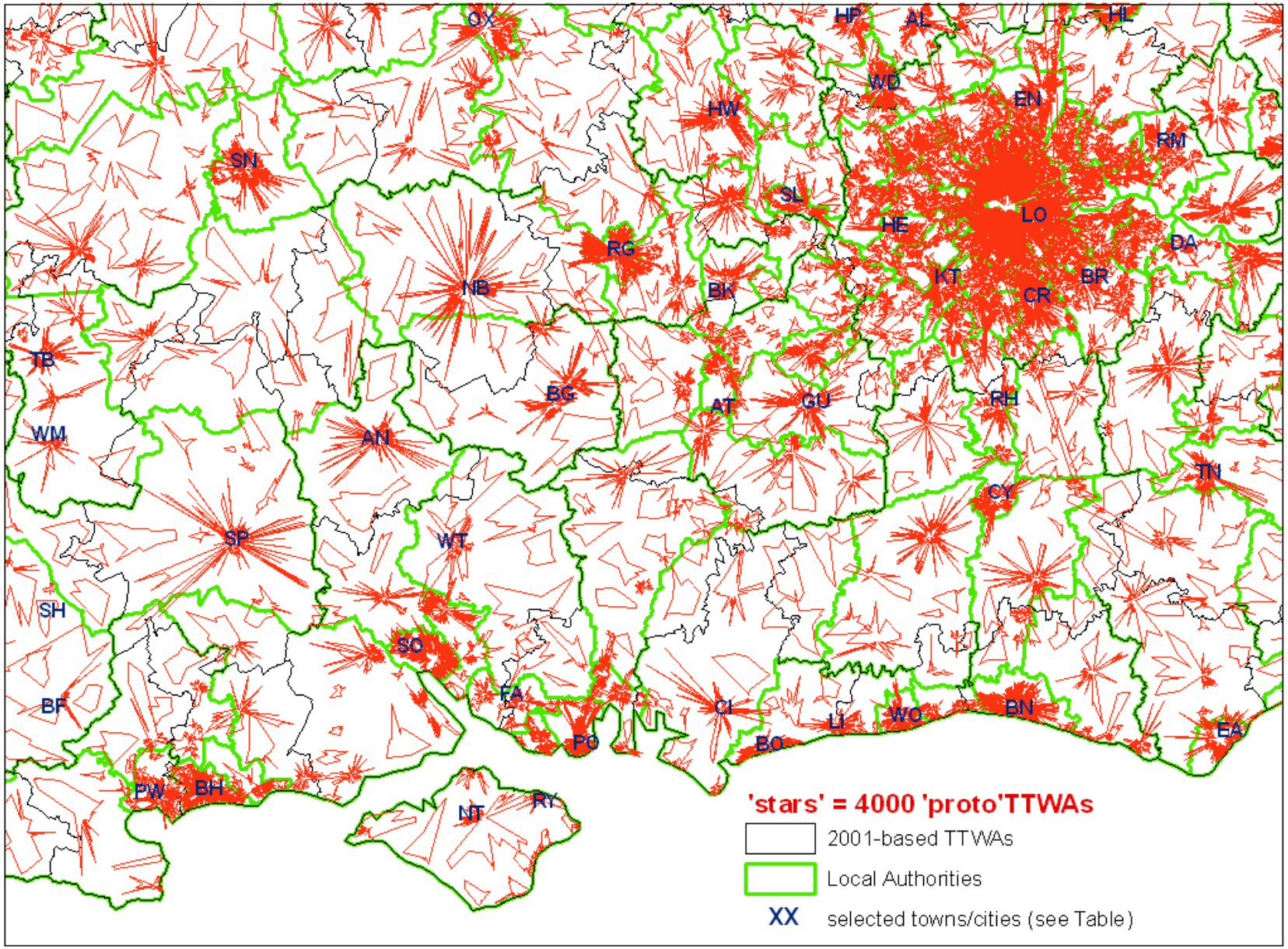
NB. the larger the 'star' the more it looks like a centre-&-hinterland <u>but it is not</u>

[+ selected towns/cities shown with 2 letter codes]

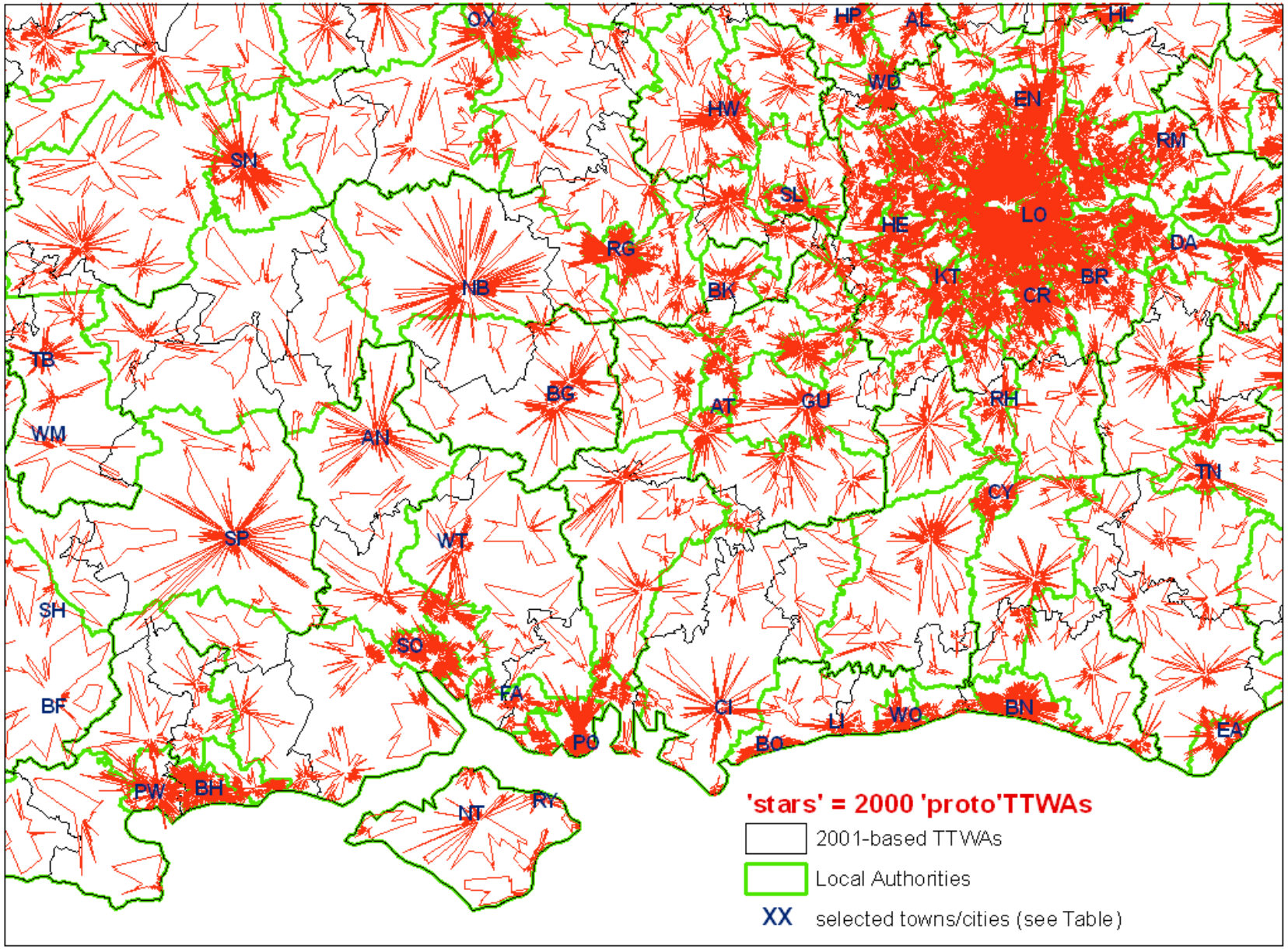
The sequence of maps gives a brief 'movie' of the process…



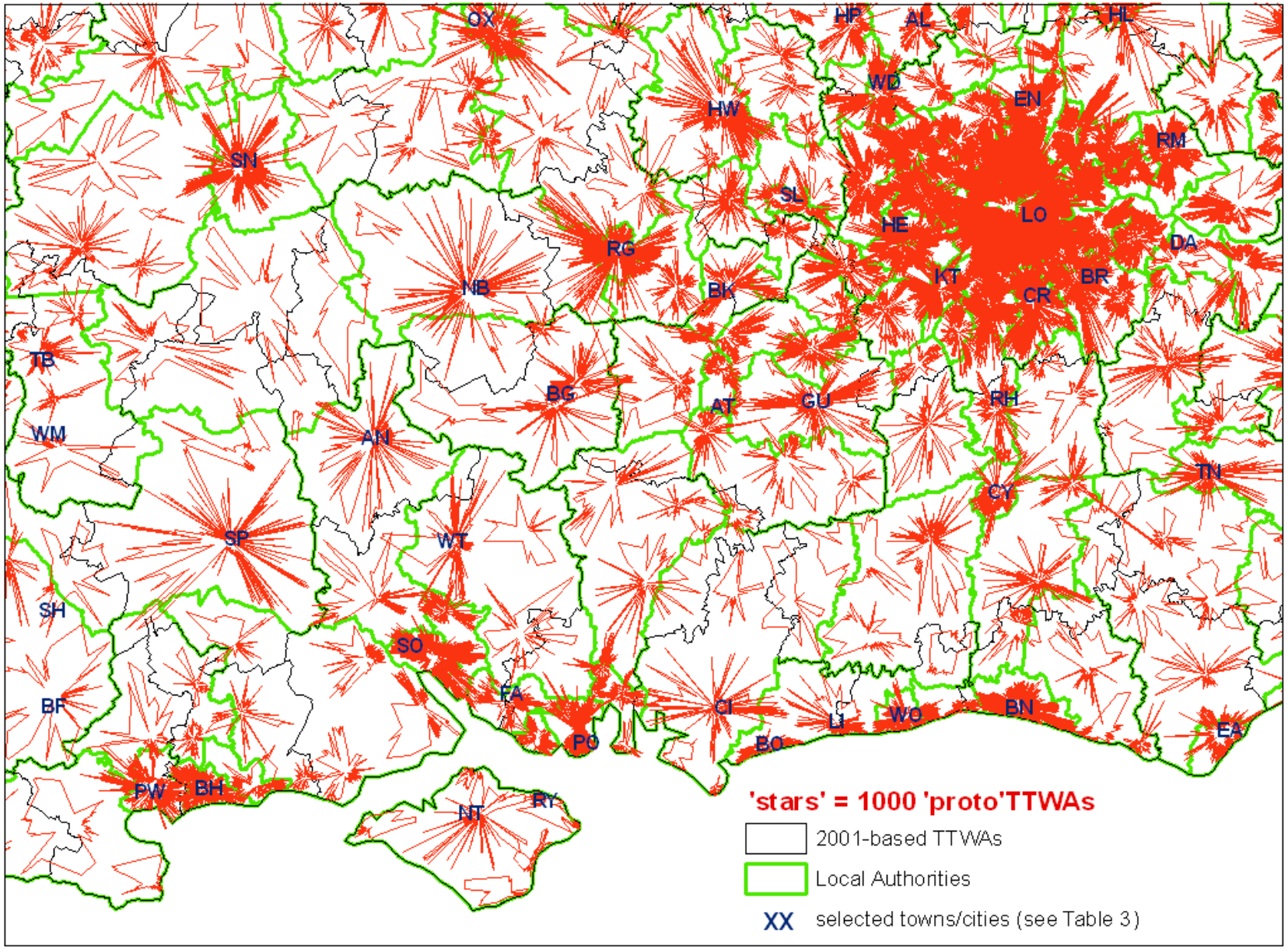
8000 ‘proto’TTWAs



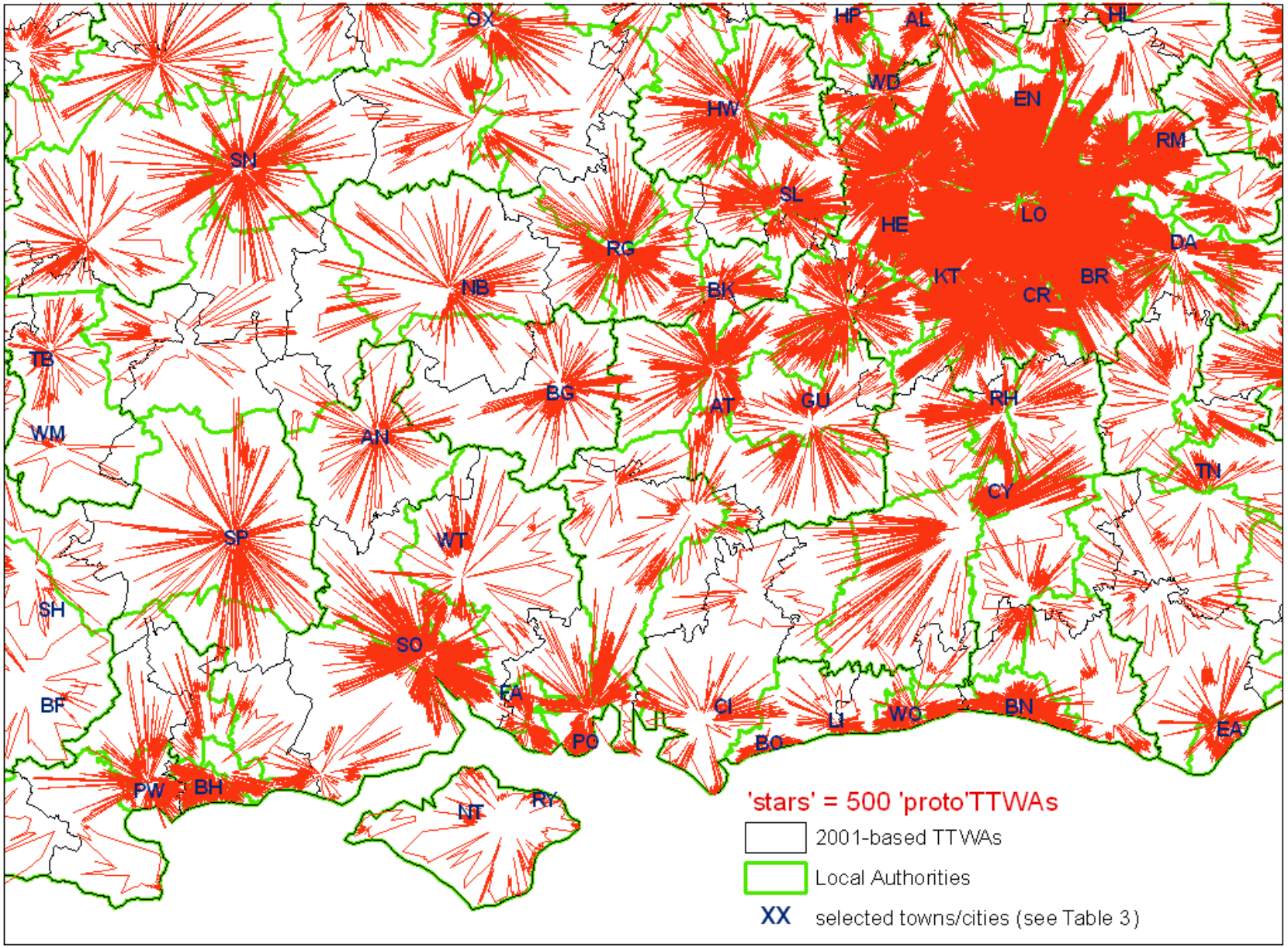
4000 ‘proto’TTWAs



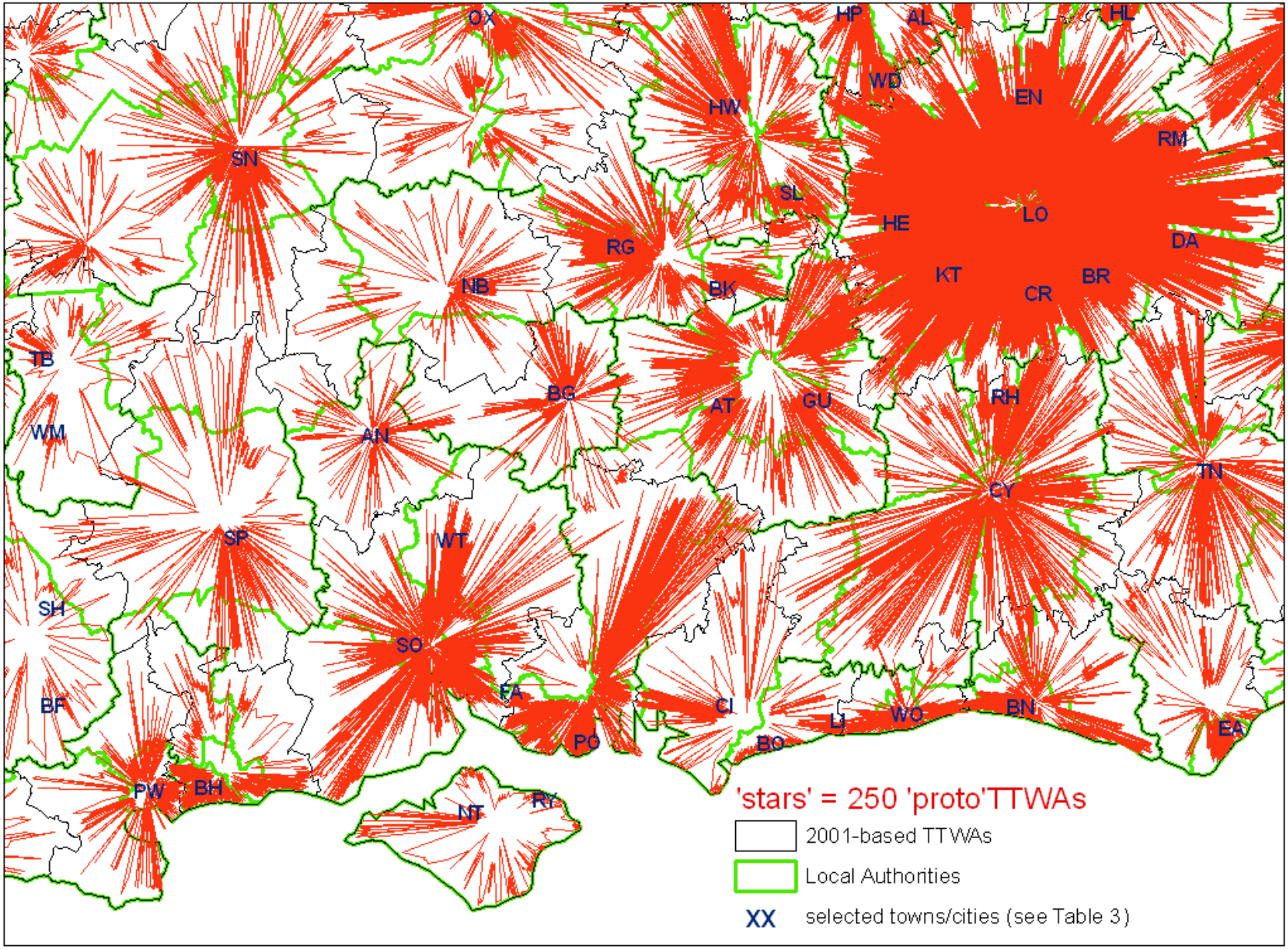
2000 ‘proto’TTWAs



1000 ‘proto’TTWAs



500 ‘proto’TTWAs



250 ‘proto’TTWAs

4] Results from one recent research study

The study was commissioned in 2009 by the previous government's NHPAU (National Housing & Planning Advisory Unit) to produce robustly defined Housing Market Areas (HMAs) that partitioned the territory of England

Consultation with interested parties approved an approach based on analysing migration patterns (reflecting the arbitrage process in the housing market) and also commuting patterns (because most moves are by people who are not at the same time changing their place of work)…there were also linked analyses of the geography of house price levels

4] Results from one recent research study

The study was commissioned in 2009 by the previous government's NHPAU (National Housing & Planning Advisory Unit) to produce robustly defined Housing Market Areas (HMAs) that partitioned the territory of England

Consultation with interested parties approved an approach based on analysing migration patterns (reflecting the arbitrage process in the housing market) and also commuting patterns (because most moves are by people who are not at the same time changing their place of work)…there were also linked analyses of the geography of house price levels

The requirement was for a set of sub-regional HMAs that were thus ≥ TTWAs

1 TTWAs were taken as 'building blocks' – ensuring that the resulting definition of HMAs would have respected major patterns of commuting

2 TTWAs were then grouped (by the TTWA method) using migration data, thus ensuring that each HMA internalised clustered flows of migrants*

**** Moving Group Reference Persons (MGRPs) aged on Census day at least 25 and who lived at a different specified UK address 12 months previously***

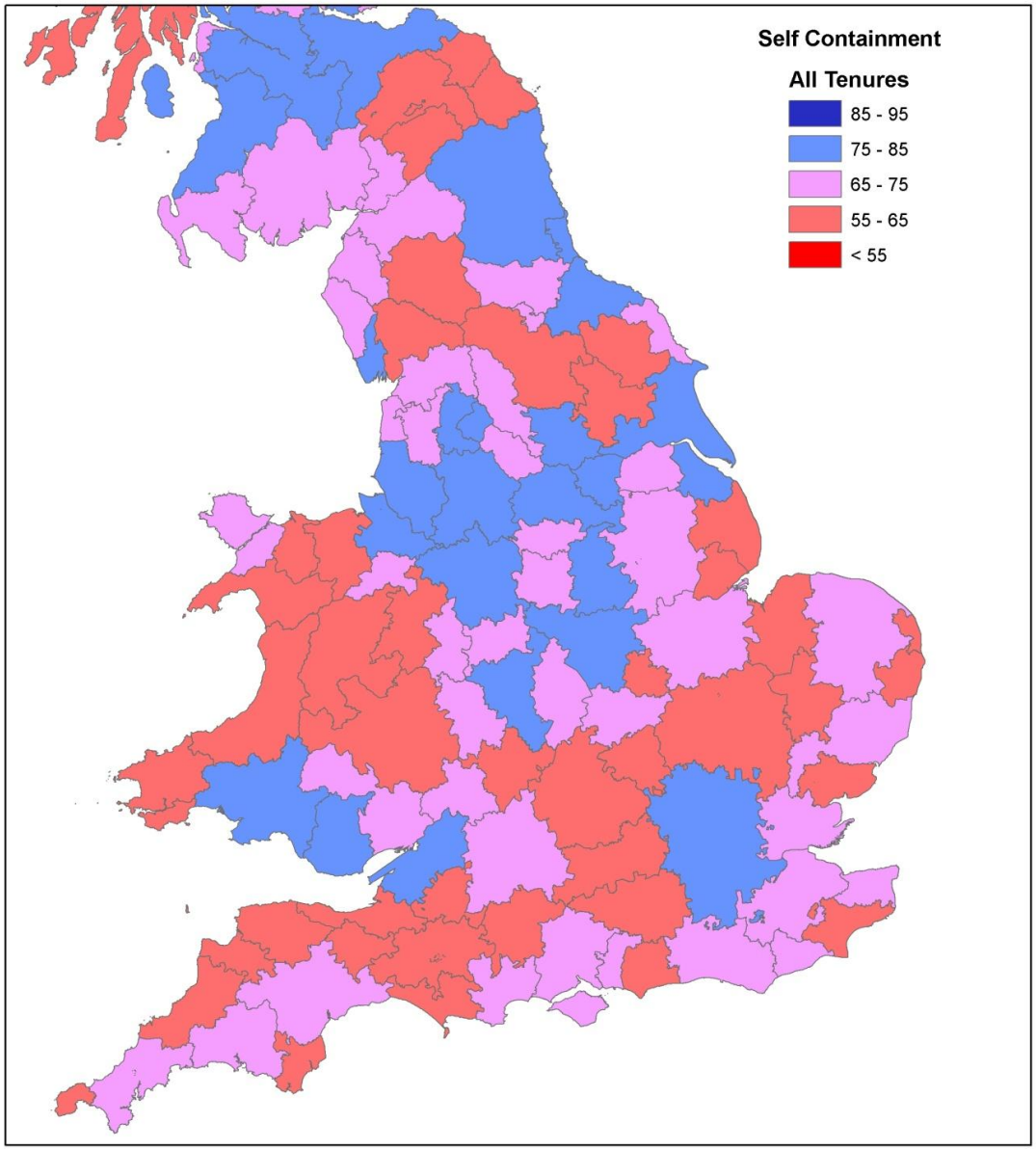
Key results: data on all MGRPs

All these *non*-tenure-specific HMAs are required by the definition analysis to have self-containment values that are at least 55%

HMAs' self-containment %s indicate how far all MGRPs see the boundaries as their effective 'mobility space'

blues = higher self-containment

reds = lower self-containment
(ie. any additional new housing in such areas faces the 'risk that people from other areas would moving there)

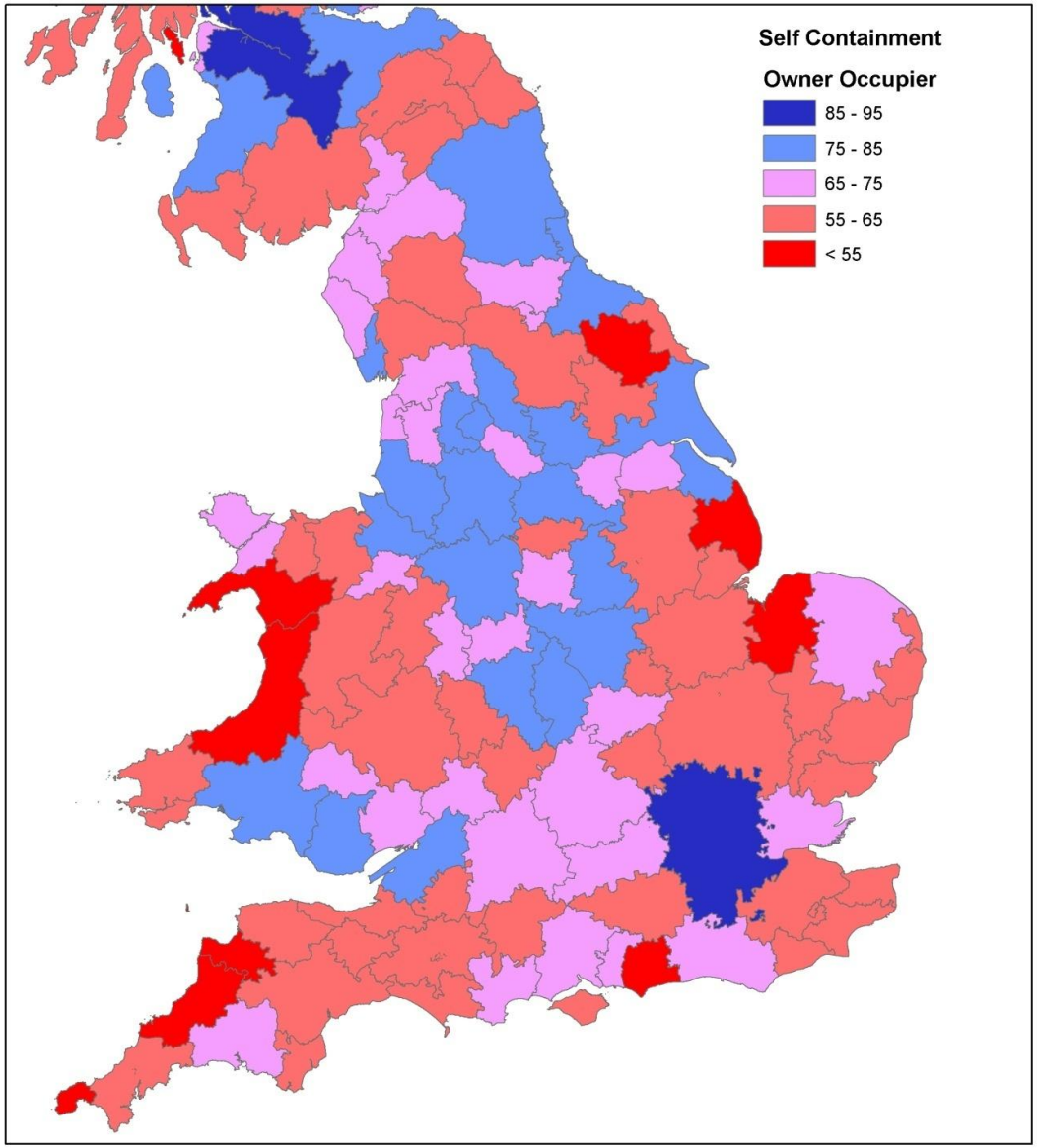


Owner Occupiers

Average migration distance for owner occupiers is very similar to that of all MGRPs, partly because this is the largest tenure sub-group

As a result this tenure's MGRPs' self-containment values are very similar to those for all MGRPs (the previous map)

So far as there are differences, they show the most urban HMAs *might* be subdivisible, while some remoter rural HMAs aren't self-contained

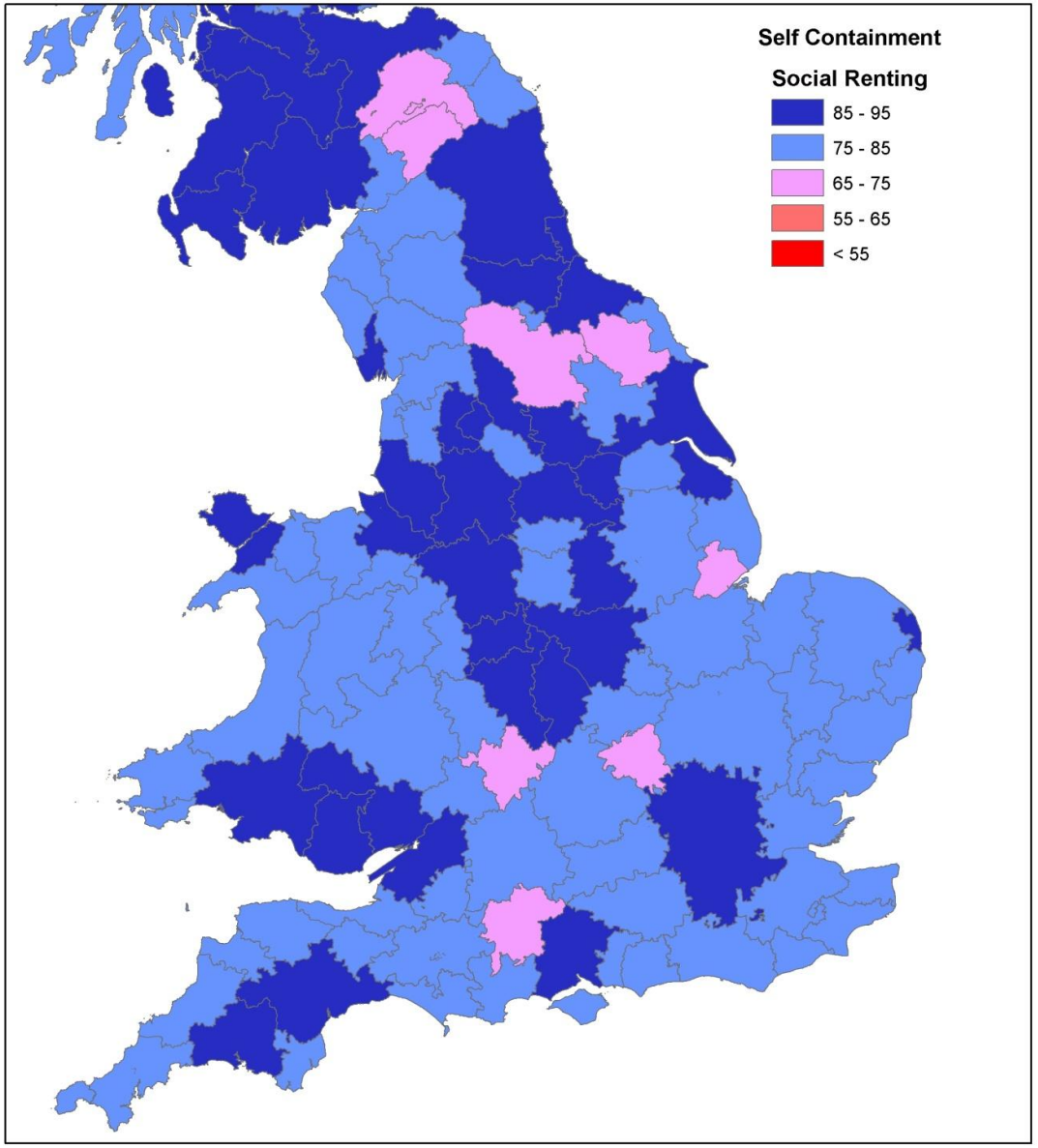


Social Renters

Average migration distance for social renters is very low, so these HMAs that were defined with All MGRP data are highly self-contained for this tenure group

Among the few areas with lower self-containment values the majority are very rural areas (some with military bases)

There are more social renters in the most urban HMAs, where the self-containment values tend to be very high (ie. these areas can be split into many separate HMAs for social renters)

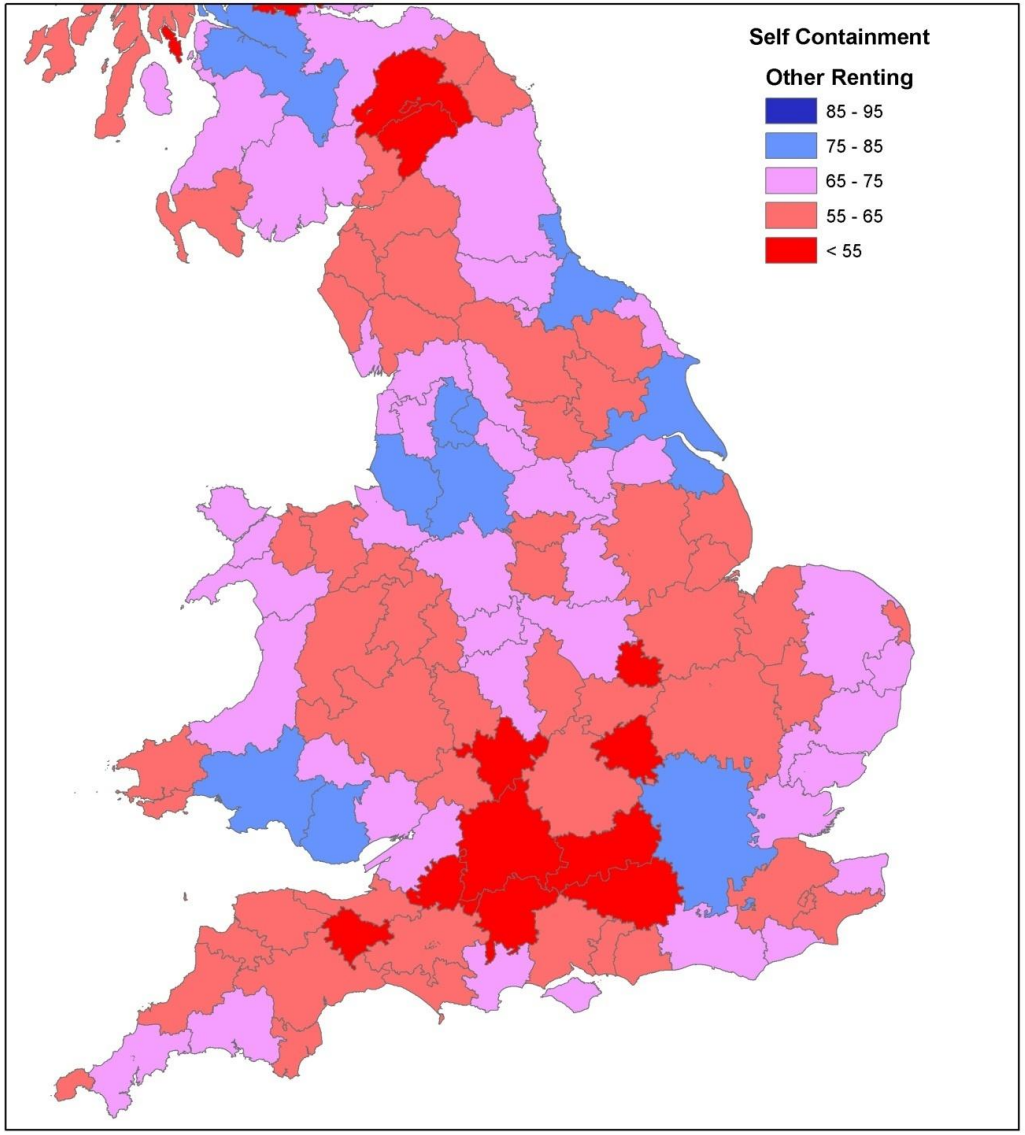


Other Renters

Average migration distance in this tenure is higher than for either of the other tenures, so the self-containment %s here are the lowest

As a result many of the HMAs (defined with All MGRP data) have very low % values here

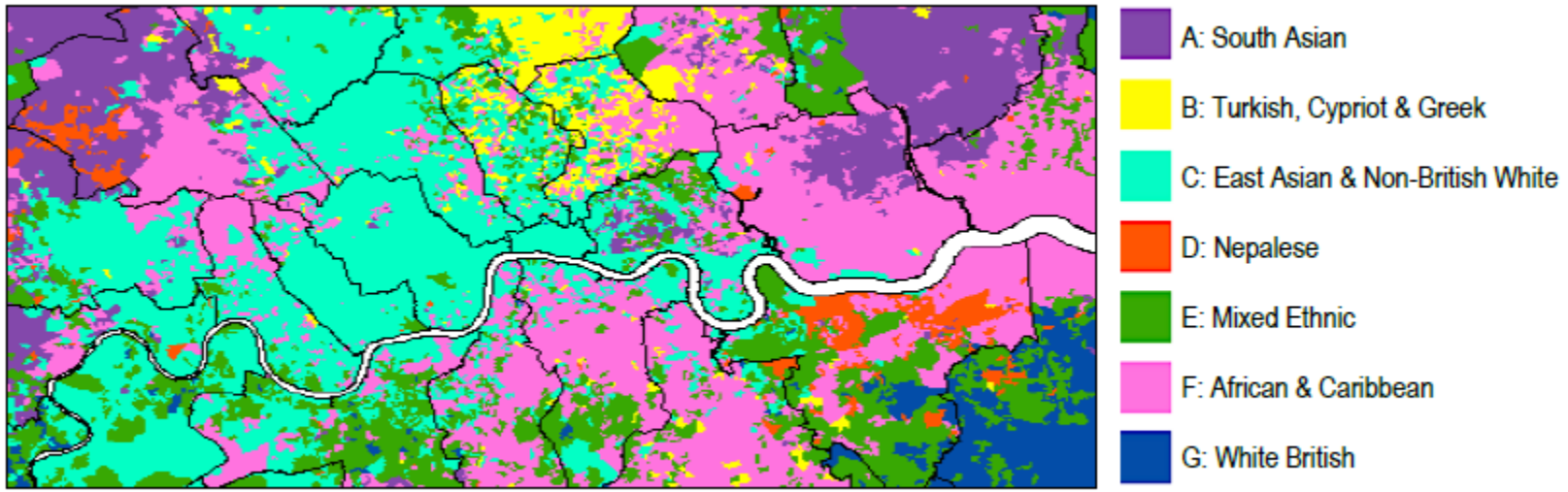
In low self-containment areas any new supply for 'other renters' is likely to have little impact on the local demand levels due to the high 'risk' of many people coming into the area to rent the new properties



5] Possible new future applications

There are increasing examples of the linking Census data with new 'Big Data' including from social media (eg. Twitter) to give a time of day dimension

Wei, Lansley & Rains (students at UCL) in recent work* with J Sainsbury
"How Open Data Resources can be used to define a composite measure of cultural identity and heritage for England & Wales"
produced a map of London by using standard tables from the Census:



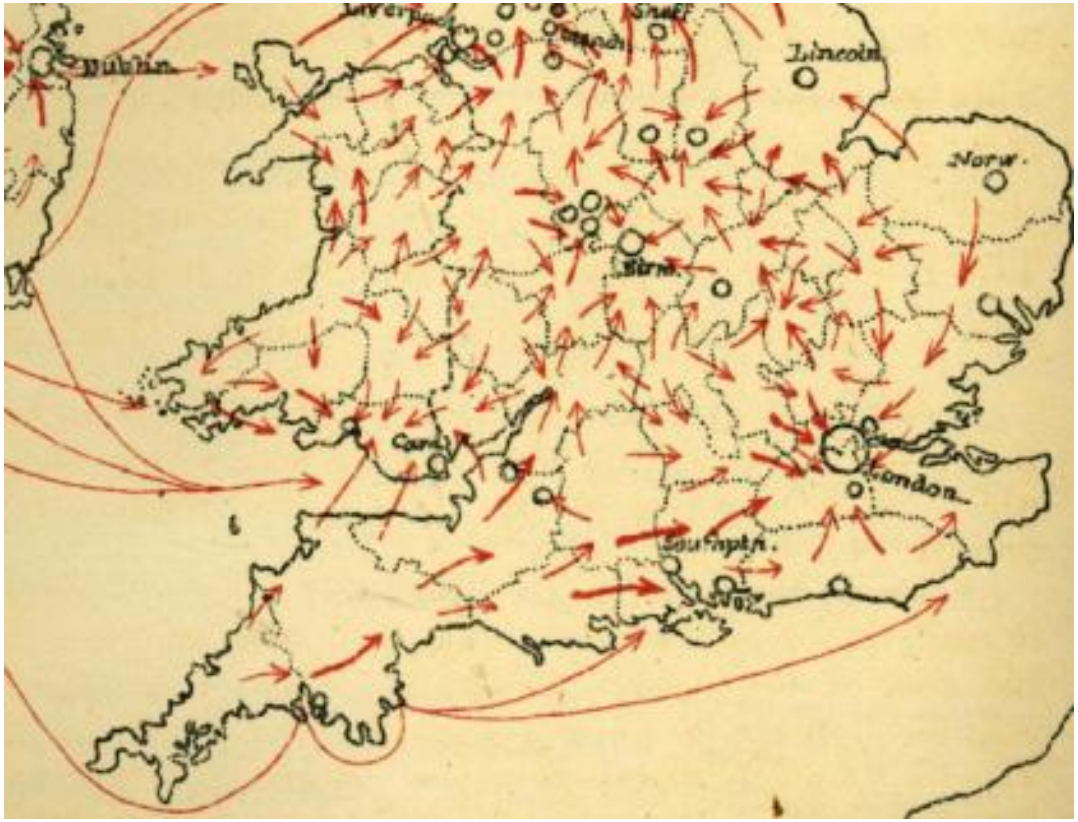
2011 Cultural, Ethnic and Linguistic Output Area Classification for England and Wales – Central London

…similar analyses will be possible with the data on the new population base "workday" which adds workers at their workplace to residents not in work

*** http://cdrc.ac.uk/wordpress/wp-content/uploads/2014/10/Report-Yiran-Wei.pdf**

…and a less novel potential application!

Ravenstein (1885) Currents of Migration (excerpt)*



- acknowledgement to Mike Batty (UCL)

  www.complexcity.info/chapter-2-figures

Flow data has a long history of being under-used

NOW the key issue must be:

? what do you need the data to do for you?

? data on movements that 'daily' OR 'occasional'

? flow analyses that show in/out patterns to/from a single location OR bounded 'clusters'

? new perspectives on the populations of areas, such as students out of term, or a 'workday' potential market