Exploring the spatial scale of population variables in England and Wales in 2001 and 2011

Chris Lloyd

Email: c.d.lloyd@liverpool.ac.uk
Twitter: @lloydcd

Census Research User Conference 27th September 2013, London

UNIVERSITY OF LIVERPOOL

School of Environmental Sciences

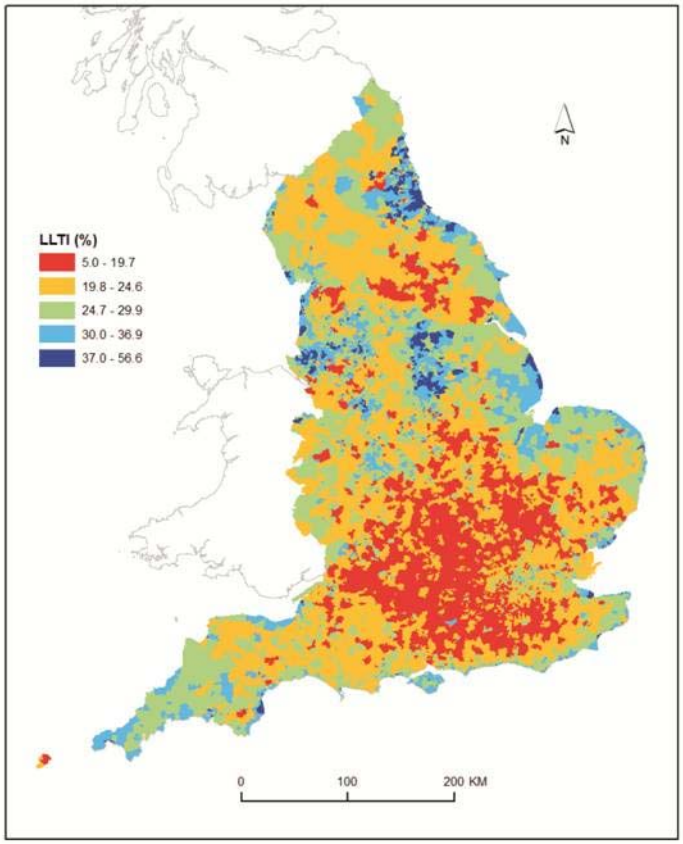
Outline

1. Defining spatial structure
   How a population sub-group is geographically distributed

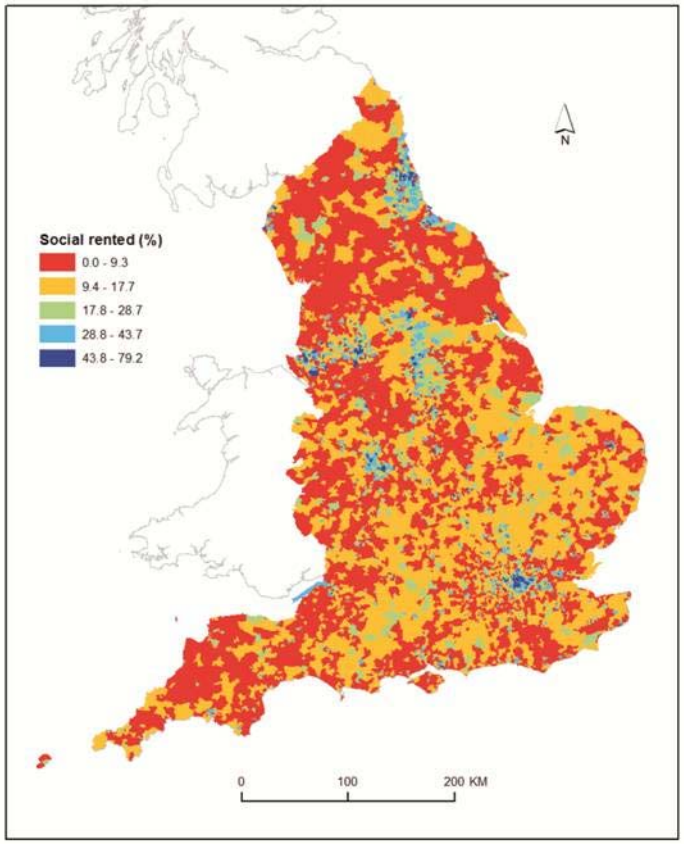
2. England and Wales analysis: 2001 and 2011
   What we can do with small area data

3. The need for small area Census statistics
   What we can't do without small area data

Spatial structure: LLTI and social rented HH in 2001, in England by CAS wards



(a)



(b)

Measures of population distribution

Index of dissimilarity, *D*, and Moran's *I* (autocorrelation coefficient)

D: measure of unevenness – how unevenly distributed is one group in relation to another? *D* = 0 indicates the groups are in the same proportion in all areas (e.g., 75, 25 in all areas); *D* = 1 indicates all areas are 100% of one group or the other; so, larger values indicate greater unevenness.

I: measure of clustering – negative values indicate neighbouring values tend to be dissimilar; values close to zero indicate no structure; positive values indicate spatial dependence or clustering.

For *I*, neighbours are given greater weight if they are nearby; so, *I* measures clustering between zones and their nearest neighbouring zones.

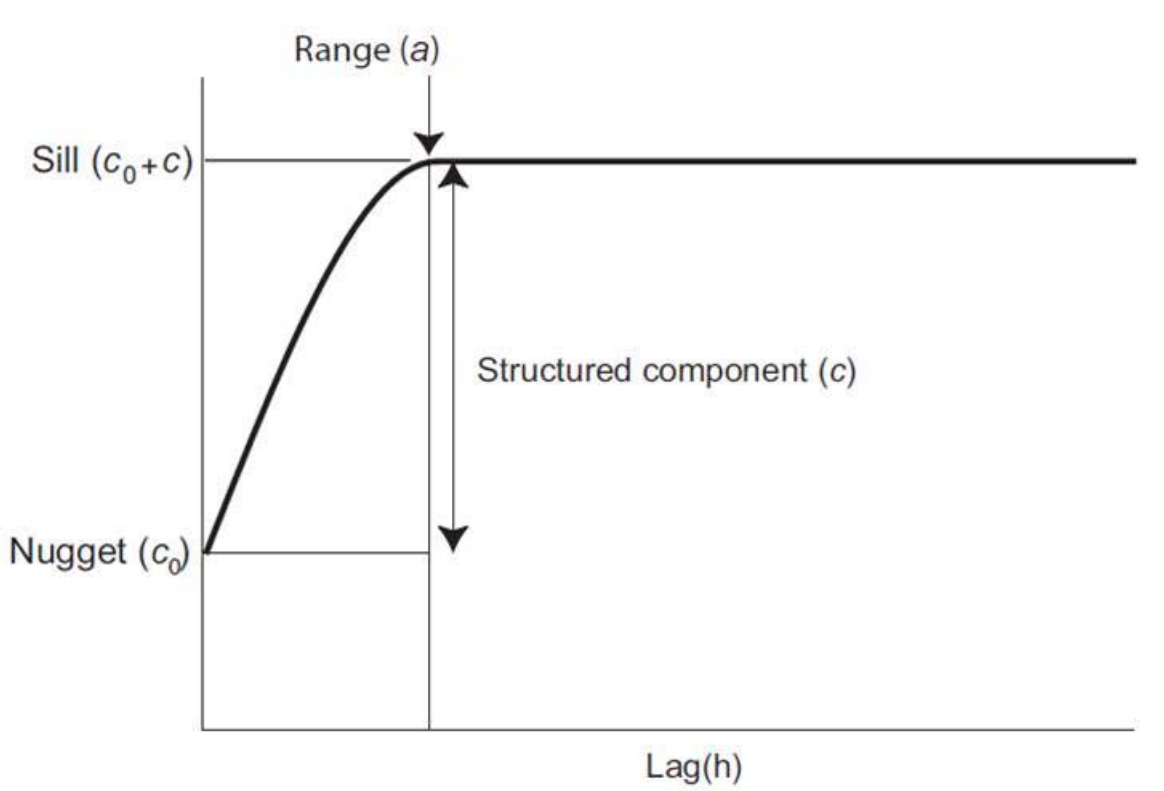
Measures of population distribution

Variogram: spatial dependence at different spatial scales

1. Take each data value in turn and compute its squared difference from each of the other values in the data set and store the distances between them
2. Group these differences into distance bins – e.g., all squared differences for pairs separated by 1 to 2 km and compute half of the average of these differences
3. Plot these (half) average differences against distances
4. The plot shows how difference between values changes as a function of distance

We can fit a model to the variogram which summarises its shape; the coefficients of this model provide a summary of the spatial structure of a variable

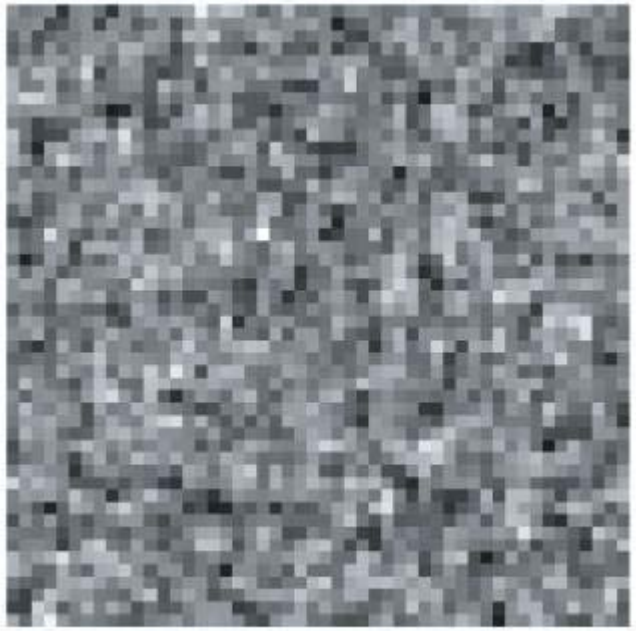
Variogram model



Bounded variogram model: nugget and effect and spherical component.

Provides a composite measure of clustering and polarisation: small nugget indicates localised clustering – with a large sill this indicates polarisation

Variograms



$a = 2$

$a = 40$

Simulated surfaces: spherical model with *a = 2 and 40.*

Variables

Variable	Description
A0to15	Persons aged 0 to 15 years
A16to29	Persons aged 16 to 29 years
A30to64	Persons aged 30 to 64 years
A65plus	Persons 65 years plus
WhiteNW	(Non) White persons
OwnOcc	Owner occupied households
SocRent	Social rented households
PrivRent	Private rented households
NoCarsVans	Households with (no) cars or vans
NoQualQual	Persons with (no) qualifications
EAEmployUnemp	(Un)employed persons
NSSEC12	Persons in NSSEC 1-2
NSSEC37	Persons in NSSEC 3-7
NSSEC8	Persons in NSSEC 8
LLTI	Persons with (no) LLTI

Data are of 2001 and 2011 Output Areas for England and Wales

Variables: log-ratios

Number of parts
2: Ethnicity (White/NonWhite), CarsVans (NoCars/Cars), Qual (Non qual/qual), Employ (employ/unemployEA), LLTI (LLTI/non LLTI)
3: Tenure (OwnOcc/PrivRent/SocialRent), NSSEC (NSSEC1and2/3to7/8)
4: Age (A0to15/A16to29/A30to64/A65plus)

Tenure (**Denominator**)

x_1	x_2	x_3
OwnOcc	PrivRent	**SocRent**
OwnOcc	**PrivRent**	

NSSEC

x_1	x_2	x_3
NSSEC12	NSSEC37	**NSSEC8**
NSSEC12	**NSSEC37**	

Age

x_1	x_2	x_3	x_4
A0to15	A16to29	A30to64	**A65plus**
A0to15	A16to29	**A30to64**	
A0to15	**A16to29**		

Index of dissimilarity, *D*

Variable	2001	2011	2011-2001
A0to15	0.159	0.161	0.002
A16to29	0.197	0.208	0.011
A30to64	0.110	0.102	-0.008
A65plus	0.258	0.274	0.016
WhiteNW	0.623	0.592	-0.031
OwnOcc	0.491	0.446	-0.045
SocRent	0.613	0.592	-0.021
PrivRent	0.384	0.371	-0.013
NoCarsVans	0.391	0.402	0.011
NoQualQual	0.223	0.255	0.032
EAEmployUnemp	0.329	0.300	-0.029
NSSEC12	0.271	0.265	-0.006
NSSEC37	0.239	0.207	-0.032
NSSEC8	0.464	0.374	-0.090
LLTI	0.197	0.199	0.002

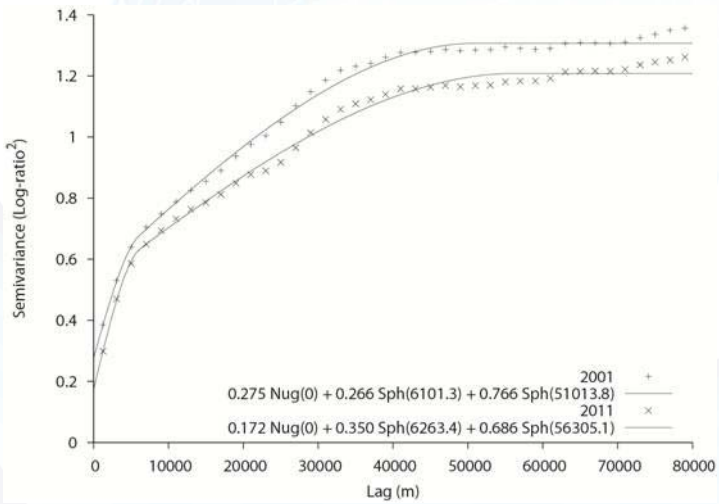
Data: Output Areas for England and Wales

Moran's *I* standard deviation

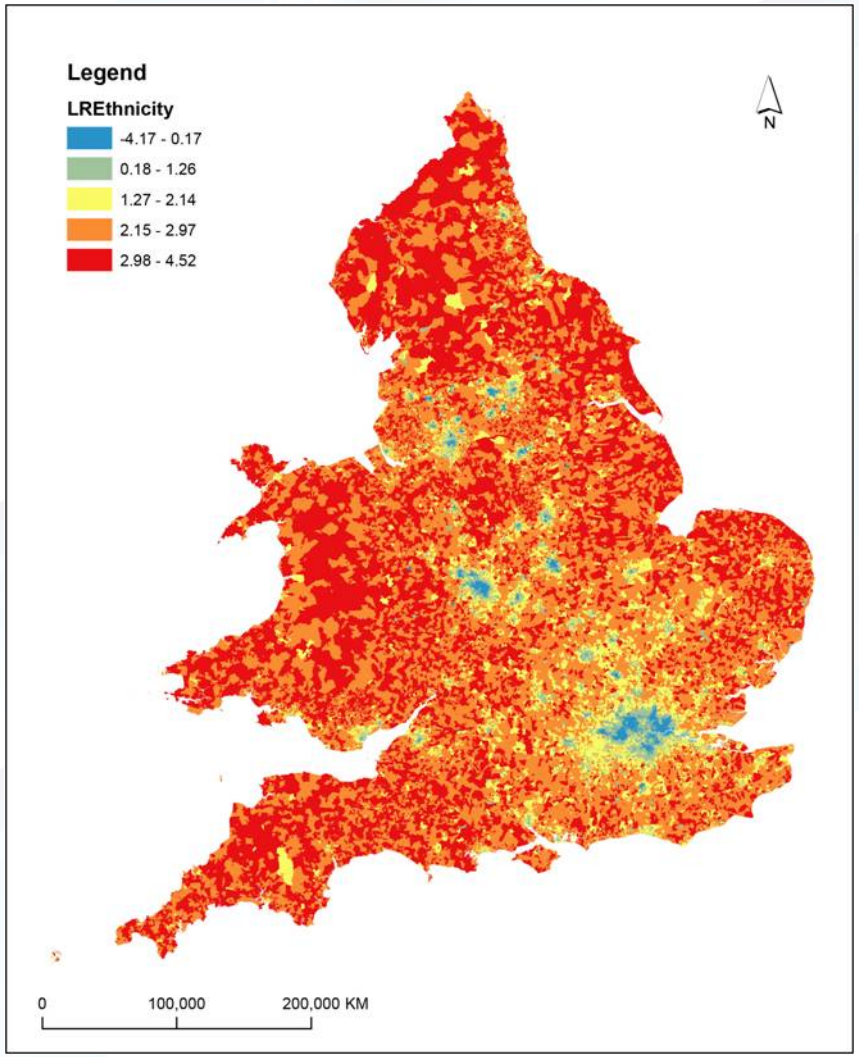
	I: 20 NN	*I*: 100 NN	*I*: 20 NN	*I*: 100 NN	SD	SD
	2001	**2001**	**2011**	**2011**	**2001**	**2011**
LRage1	0.345	0.216	0.448	0.321	0.692	0.731
LRage2	0.395	0.278	0.440	0.321	0.301	0.294
LRage3	0.431	0.335	0.524	0.417	0.411	0.418
LREthnicity	0.752	0.702	0.838	0.774	1.168	1.133
LRTenure1	0.388	0.255	0.400	0.248	1.409	1.325
LRTenure2	0.475	0.360	0.577	0.455	0.832	0.760
LRCarsVans	0.584	0.448	0.645	0.515	0.776	0.776
LRQual	0.598	0.468	0.564	0.437	0.493	0.507
LREmploy	0.445	0.359	0.475	0.367	0.602	0.522
LRNSSEC1	0.481	0.387	0.591	0.473	0.939	0.771
LRNSSEC2	0.664	0.530	0.697	0.561	0.499	0.479
LRLLTI	0.397	0.297	0.411	0.307	0.371	0.388

NN is nearest neighbours (smaller value = more local neighbourhood)
SD is standard deviation
Data: Output Areas for England and Wales

White/Non White



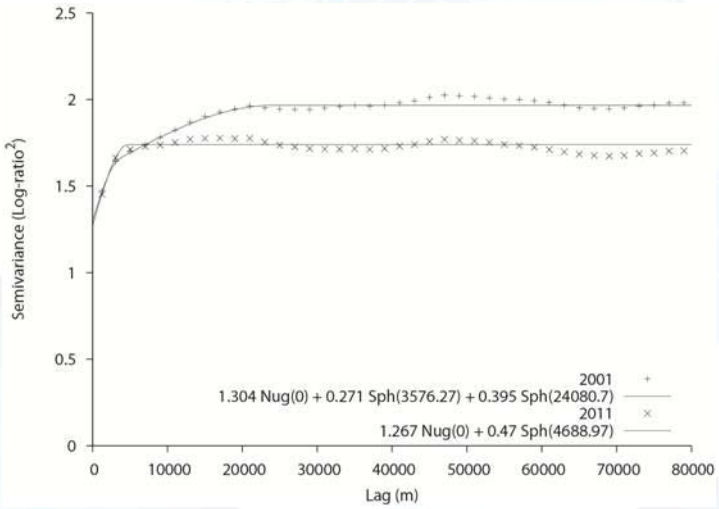
Clustering over small areas increased between 2001 and 2001, variance decreased – strong spatial structure, but regions are becoming more similar (consistent with standard deviation)

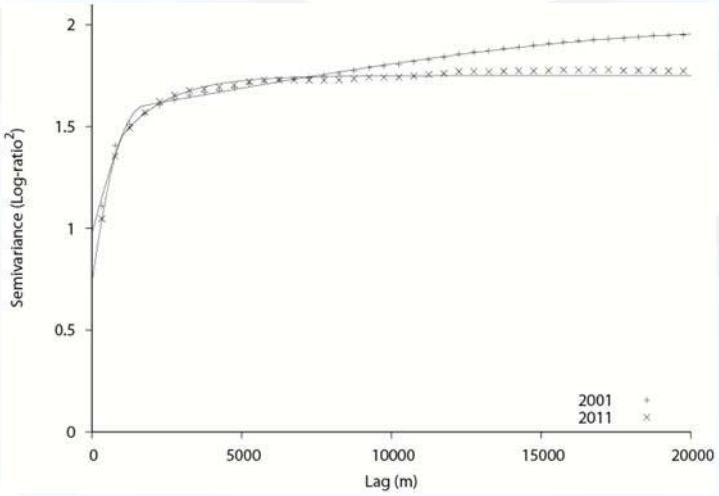


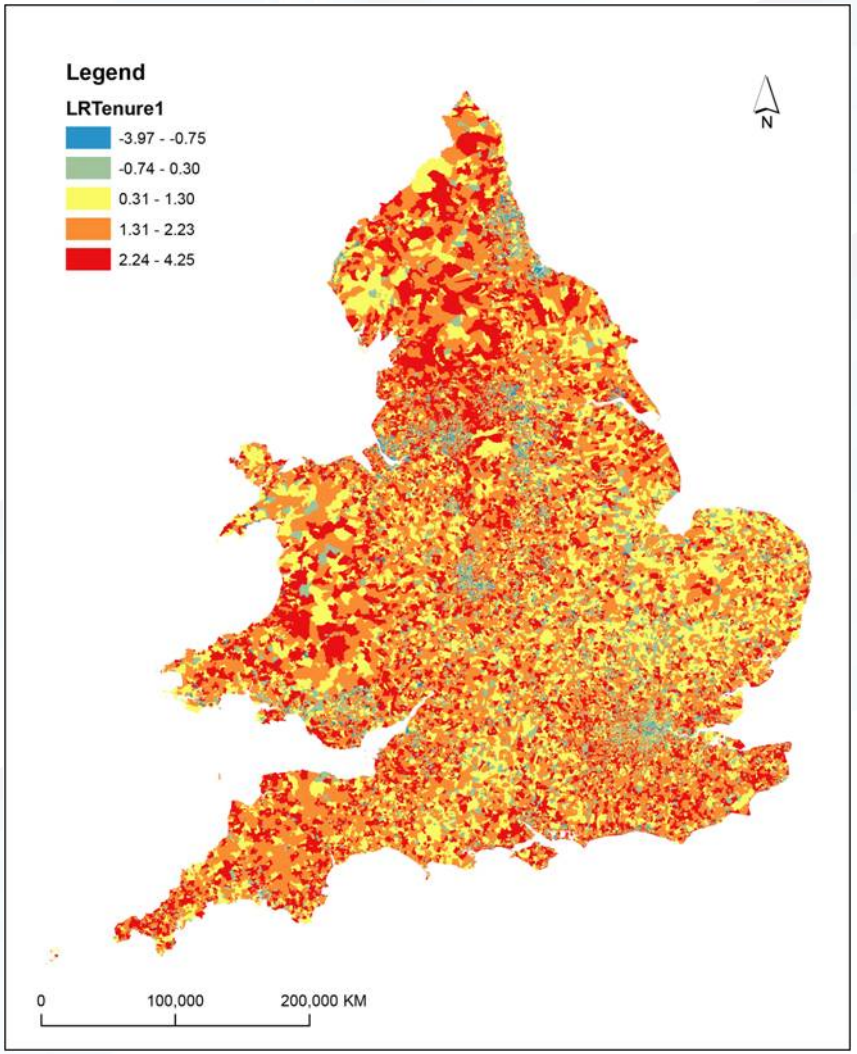
2011 OAs

Owner occupied HH and Private rented / Social rented HH

Clustering over small areas increased between 2001 and 2001, but variance decreased; much less spatial structure than for White/non White



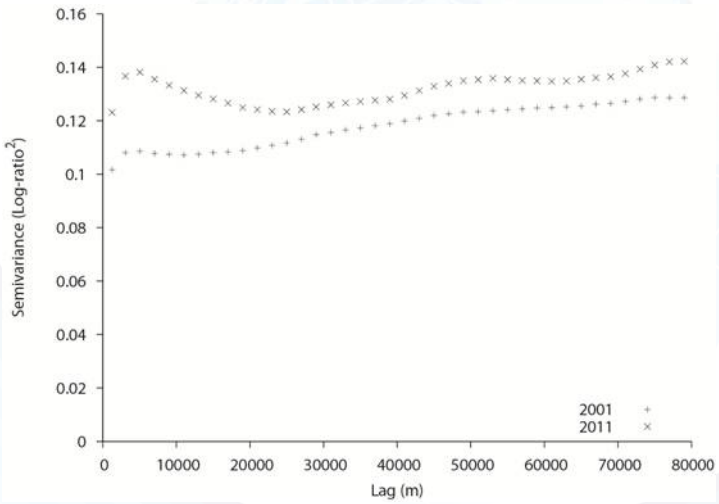


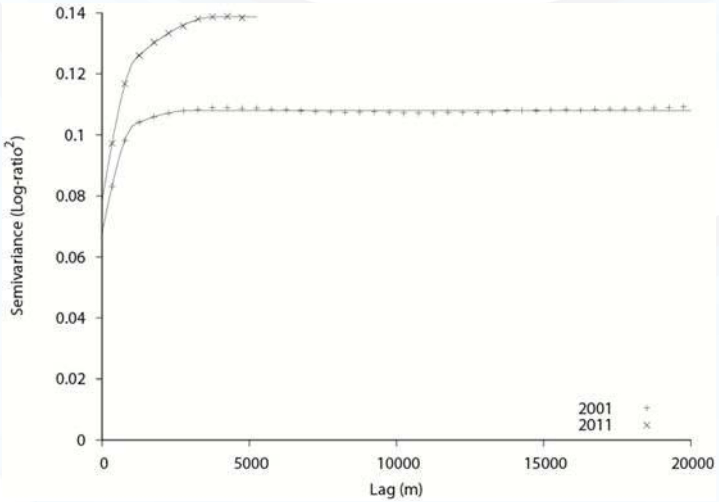


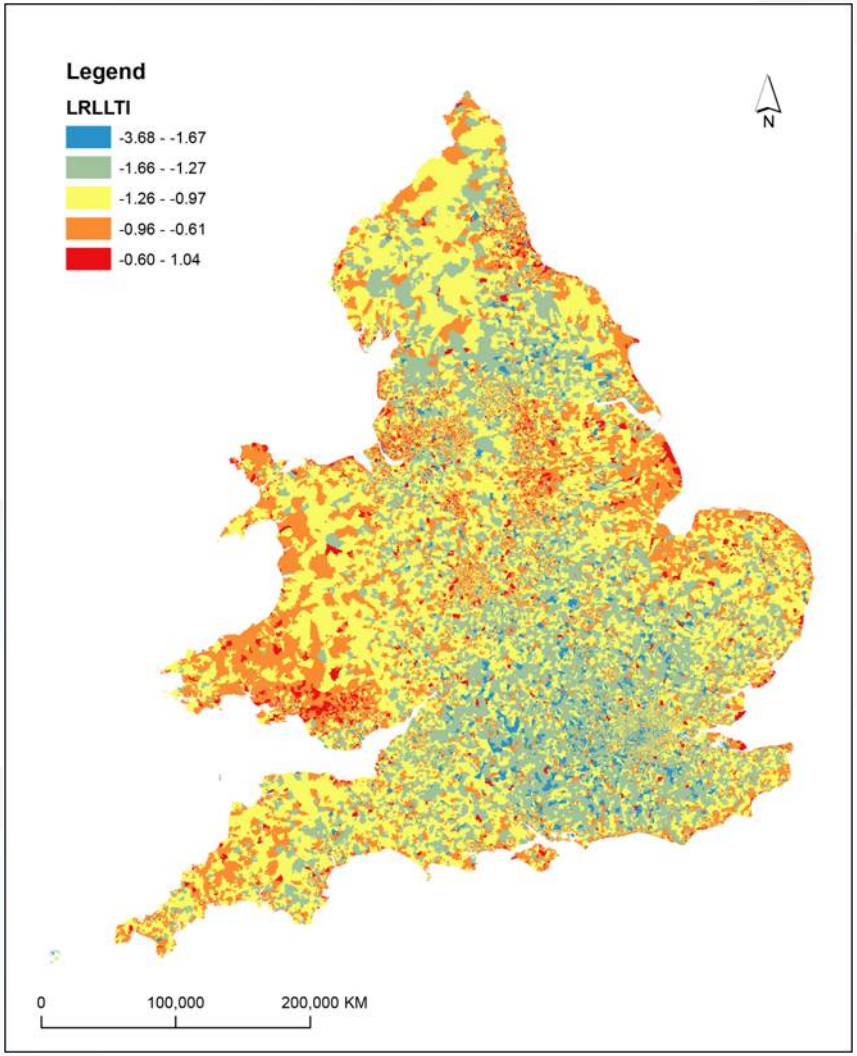
2011 OAs

LLTI/Non LLTI

Clustering over small areas, but no evidence of spatial structure > 10km; variance increased – regions are becoming more dissimilar







2011 OAs

Findings

- The age variables tend to be less uneven and less clustered than the other variables
- Small levels of clustering reflect, in some cases, high levels of clustering in some areas, but high variability elsewhere – social housing is a key example
- Between 2001 and 2011 unevenness in most population sub groups in England and Wales reduced
- Over the same period, there was an increase in localised clustering in the population by most of the demographic and socioeconomic variables assessed
- Taken together, the findings suggest that local areas have become more similar but, for many variables, this is against a background of reduced regional variation
- In simple terms, there is increased clustering *within* regions but decreased difference *between* regions for many population variables

The need for small area Census statistics

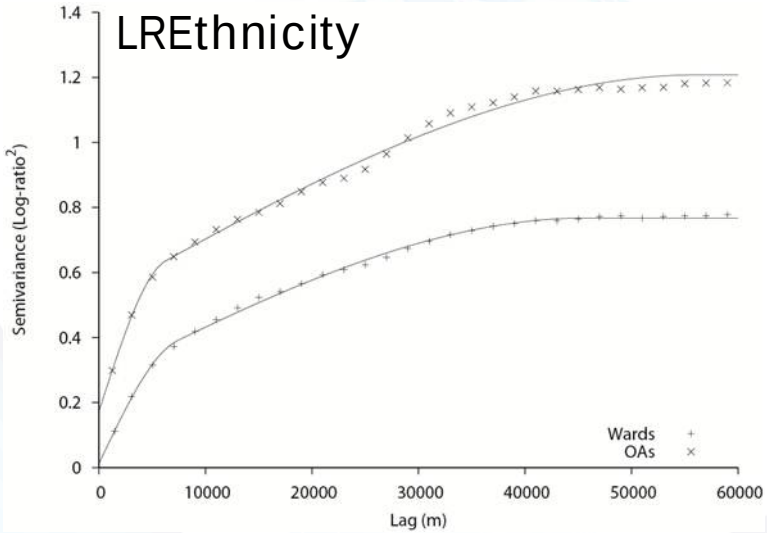
How much variation is there in population sub-groups at OA level?

How much information is lost if we move from, for example, OAs to wards?

As zone size increases, variation decreases (we lose information)

One approach is to estimate variograms for OAs and wards and assess the difference

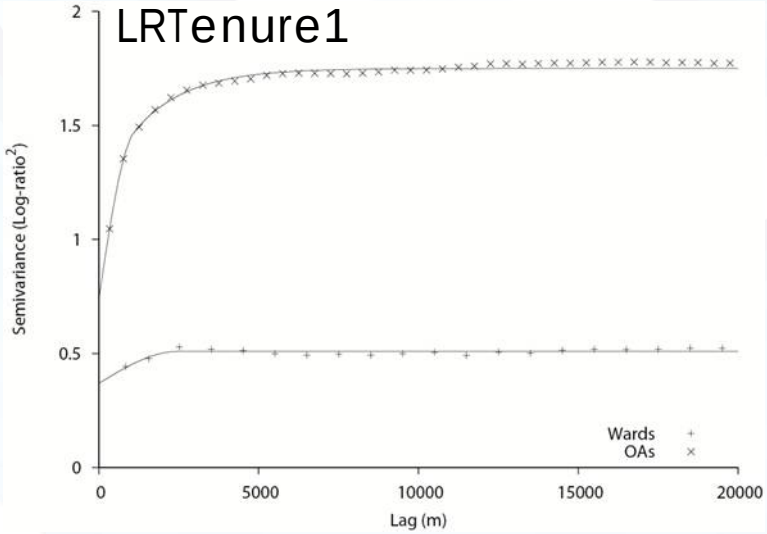
Variograms: OAs vs wards

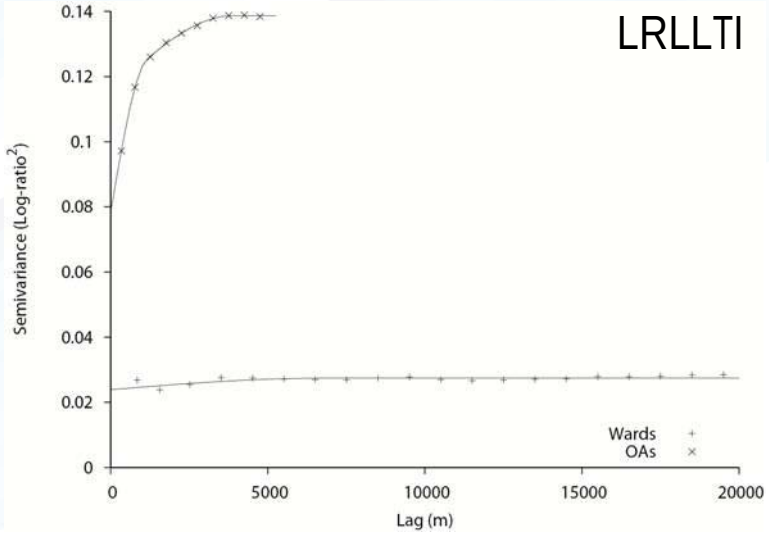


Results indicate that even moving from OAs to wards results in a considerable loss of information for most variables

Ratio of OA sill to ward sill:

LREthnicity	1.57
LRTenure1	3.44
LRLLTI	5.06





Summary

In simple terms, any application which considers area effects should make use of a zonal system which distinguishes effectively different areas – the zones must be small enough to resolve variation and if a variable changes over small distances (e.g., there are localised clusters) we need small zones

The results suggest that, judging by this case study, we need *very* small areas to meaningfully analyse population characteristics and even wards are too large in most cases

Without small area statistics we will be unable to say anything useful about where members of most population sub-groups live…

Acknowledgements

The Office for National Statistics are thanked for provision of the data for 2001 and 2011.

Office for National Statistics, 2001 Census: Digitised Boundary Data (England and Wales) [computer file]. ESRC/JISC Census Programme, Census Geography Data Unit (UKBORDERS), EDINA (University of Edinburgh)/Census Dissemination Unit. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.