School of Geography
FACULTY OF ENVIRONMENT

UNIVERSITY OF LEEDS

The potential of simulation for administrative data and social research

Mark Birkin,

Professor of Spatial Analysis and Policy,

School of Geography, University of Leeds.

Content

UNIVERSITY OF LEEDS

Background to microsimulation

Review of methods

Microsimulation and the LS

Conclusions and Next Steps

Background

UNIVERSITY OF LEEDS

Technique of microsimulation modelling (MSM) introduced into economics (Guy Orcutt) in 1950s

- Recognise the distributional consequences of economic and financial policies
- e.g. change in the **rules** for housing benefit or personal taxation

Background

UNIVERSITY OF LEEDS

Very early development of MSM in geography also – attributable to Torsten Hagerstrand

- migration (1958)
- innovation-diffusion (1967)

Introduction to the mainstream in 1970s e.g. Wilson and Pownall (1976)

Background

UNIVERSITY OF LEEDS

MSM also popular in other social science and policy contexts

- e.g. Mike Murphy (sociology/ anthropology)
- e.g. Jane Falkingham (health care)
- e.g. McFadden, Williams (transport)

Strong links to CA and especially agent-based modelling

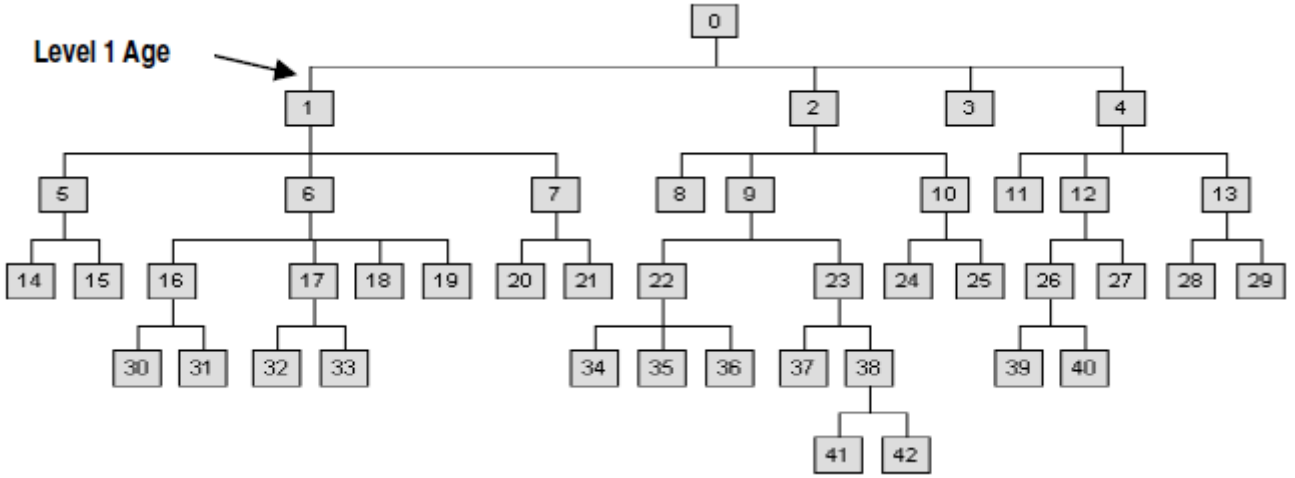
- e.g. Jim Doran (political science)
- e.g. Land use models etc

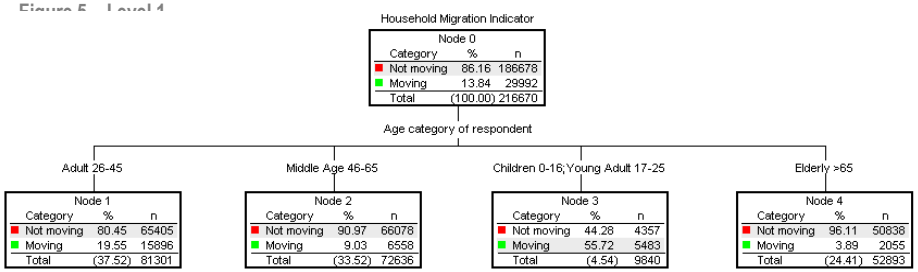
Example – Regeneration in Leeds

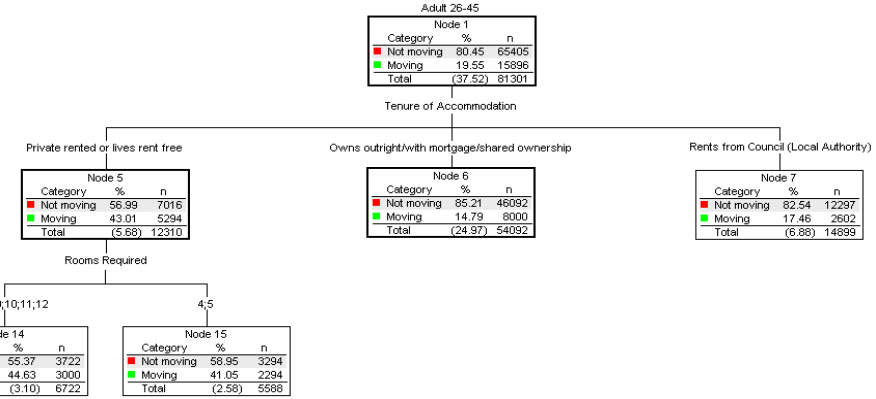
UNIVERSITY OF LEEDS

MicroSimulation Model

UNIVERSITY OF LEEDS







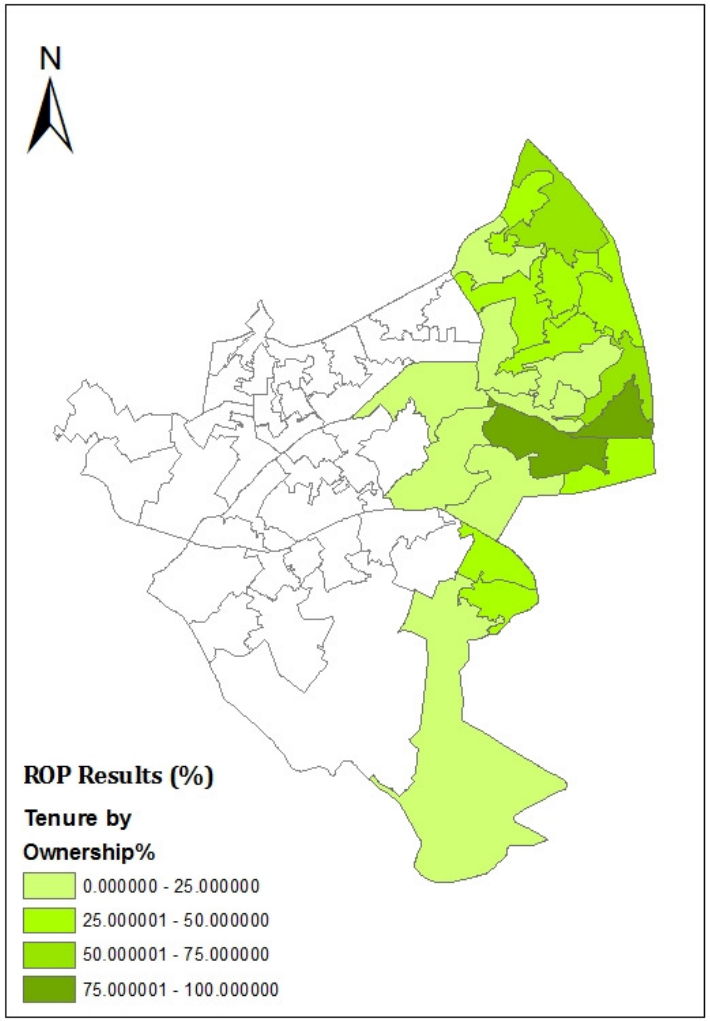
Model Rules

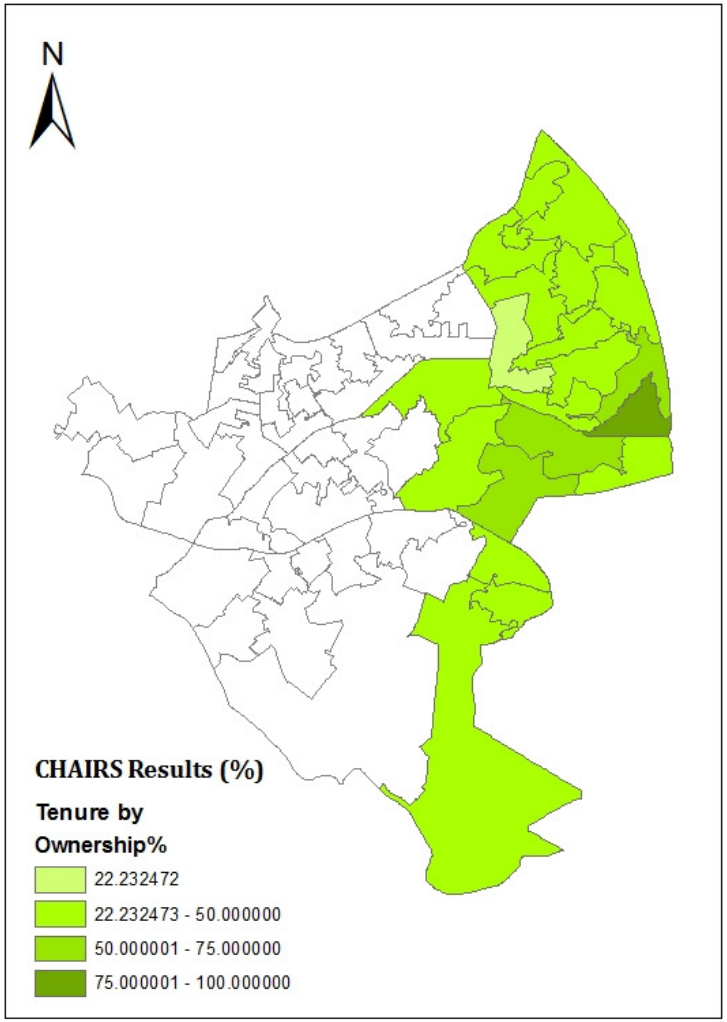
UNIVERSITY OF LEEDS

Rules	Definitions
Known areas	Households moves to area to which they are familiar
Index of Multiple Deprivation	Households move to better quality areas
Number of rooms requested	Households move to houses where the number of room is satisfactory
Ethnicity preference	Households move to areas where the ethnic make-up is tolerable
Transport routes available	Households move to areas where transport routes are readily available.
Socioeconomic status	Households move to houses they can afford.
Schools in proximity	Households containing school-aged children move to areas where schools are accessible.

Model Validation

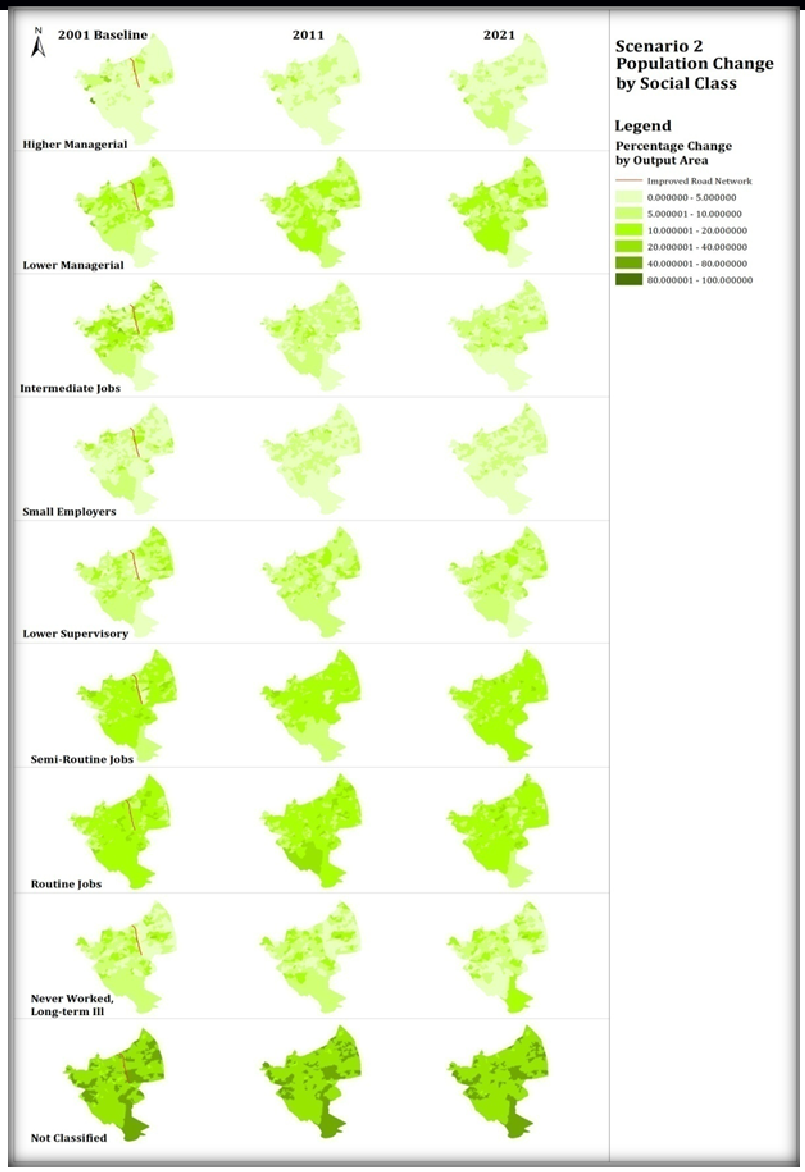
UNIVERSITY OF LEEDS





Scenario Analysis

UNIVERSITY OF LEEDS



Microsimulation Methods

UNIVERSITY OF LEEDS

Four fundamental components

- Population Reconstruction
- Projection
- Data Linkage
- Secondary models

Population Reconstruction

UNIVERSITY OF LEEDS

Population Reconstruction

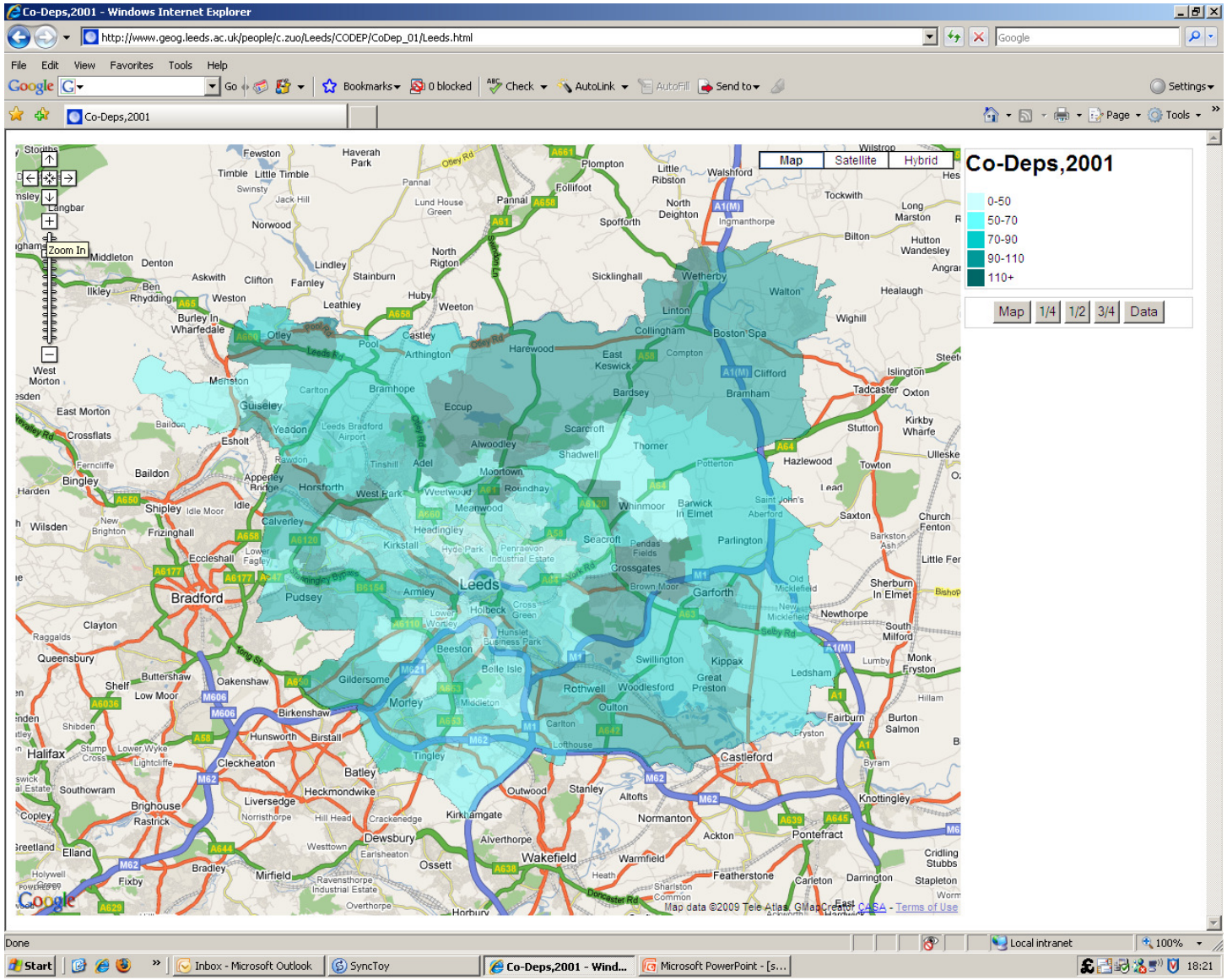
UNIVERSITY OF LEEDS

Methods are well-established

- Conditional probabilities (iterative proportional fitting)
  - (Birkin & Clarke,Beckmann et al)
- Deterministic reweighting (Smith, Procter)
- Regression-based methods (Tanton)
- Simulated annealing (Williamson, Ballas)
- Genetic algorithms (Birkin et al)

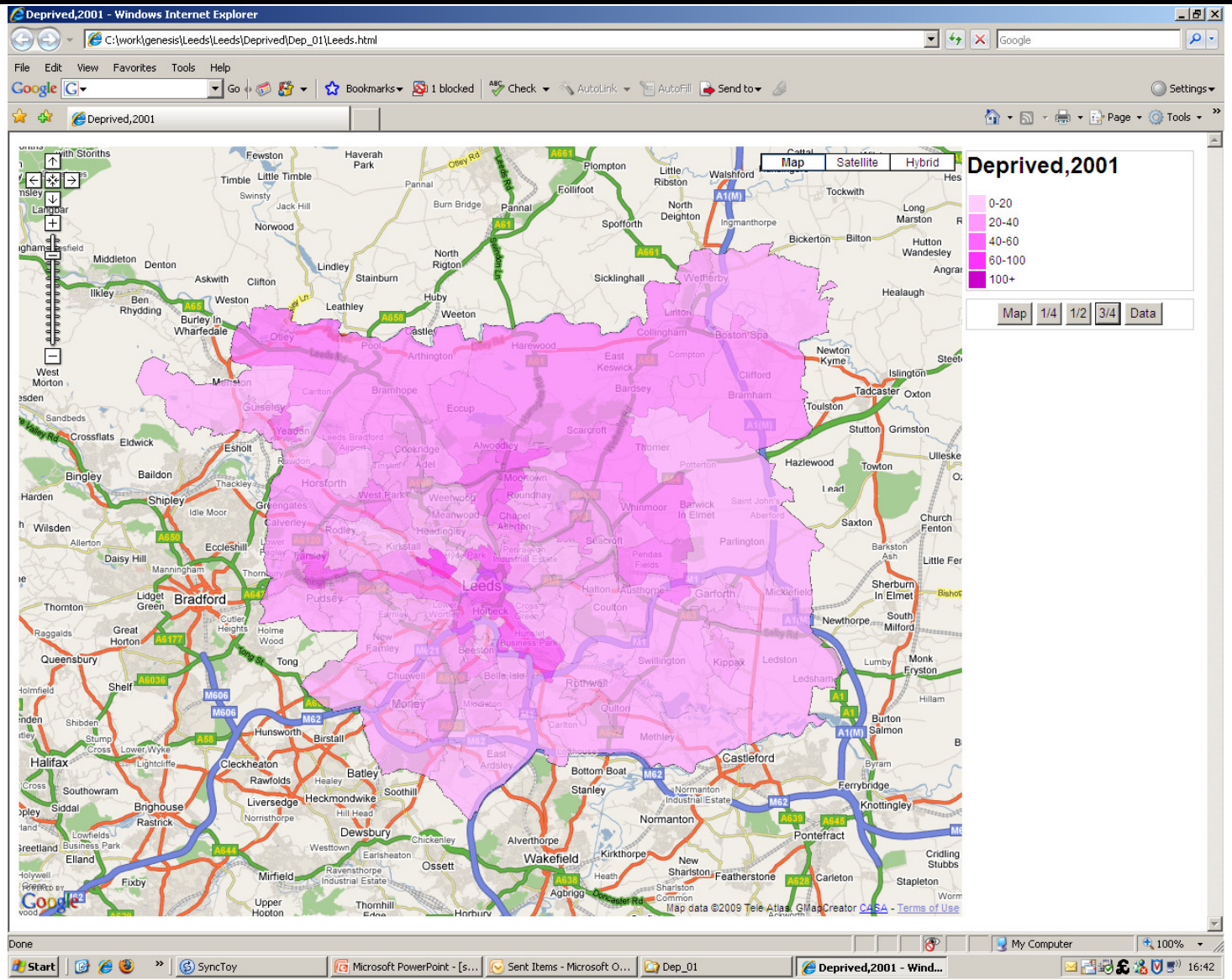
Population Reconstruction - Example

UNIVERSITY OF LEEDS



Population Reconstruction - Example

UNIVERSITY OF LEEDS



Moses Dynamic Model

UNIVERSITY OF LEEDS

- Ageing/mortality
- Fertility
- Inmigration
- Emigration
- Household formation
- Marital status
- Health status
- Local migration

Transition rates for fertility, mortality and migration are spatially disaggregated

- E.g. fertility: rates by age, marital status and location
- Event is simulated as a Monte Carlo process
  - Example: married woman, aged 28, living in Aireborough
  - Probability of maternity is 0.127
  - Pull a probability from a distribution of random numbers; if <= 0.127 then the event occurs

All events in discrete intervals of one year

Each module draws extensively on administrative sources

MoSeS Dynamic Model

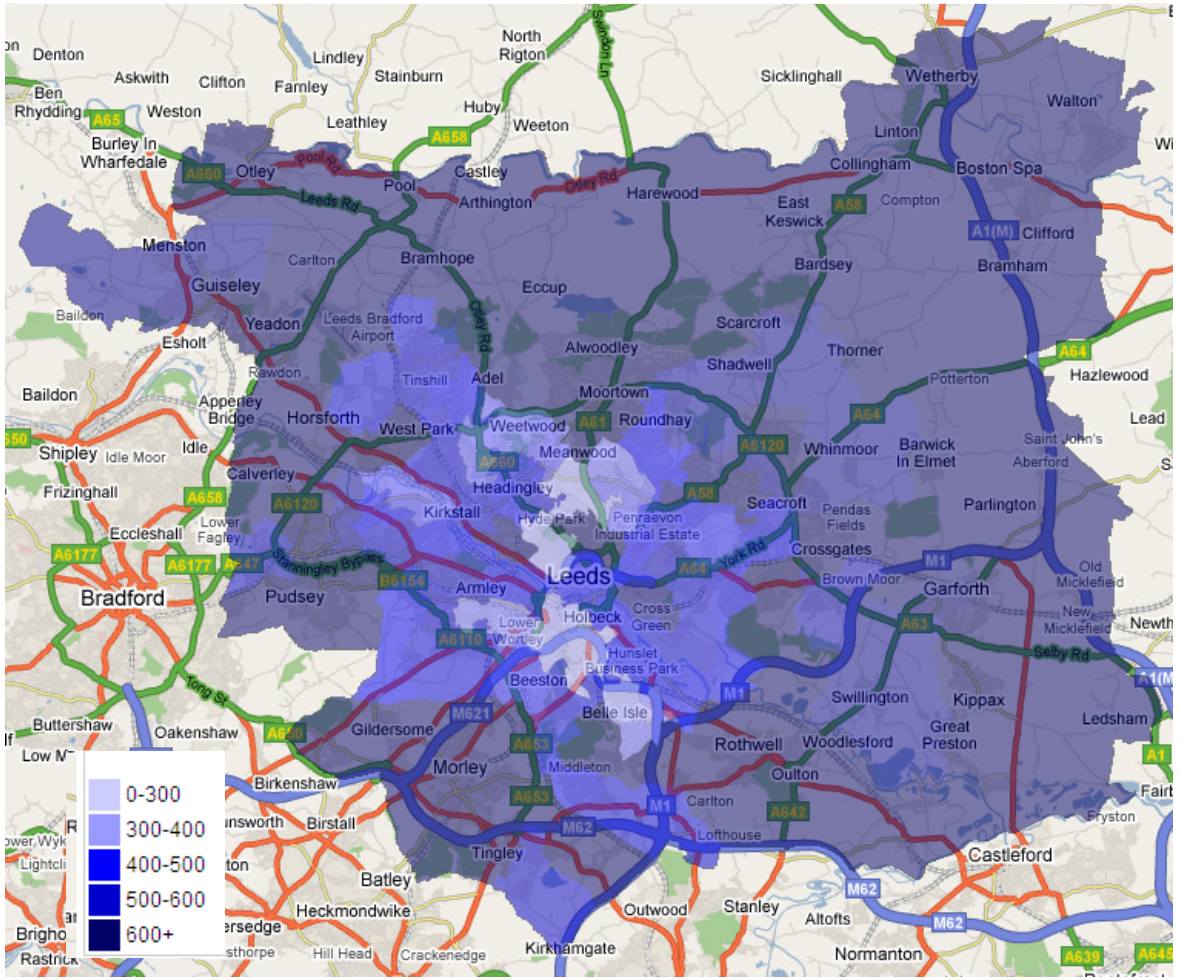
UNIVERSITY OF LEEDS

Simulation Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Gender	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Age	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Location	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Marital status	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
Household status	L	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Health	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
Gender	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Age	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Location	A	B	B	B	C	C	C	C	C	C	C	C	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
Marital status	S	S	S	S	C	C	C	C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Household status	L	S	S	S	C	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Health	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
Gender					M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Age					24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Location					C	C	C	C	C	C	C	C	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
Marital status					C	C	C	C	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Household status					C	C	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Health					G	G	G	G	G	G	G	G	G	G	G	P	P	M	M	G	G	G	G	G	G	G	G	G	G
Gender							M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Age							0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Location							C	C	C	C	C	C	D	D	D	D	D	D	D	D	D	D	D	D	D	E	E	E	E
Marital status							S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Household status							F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	S	S	S
Health							G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
Gender										F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Age										0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Location										C	C	C	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
Marital status										S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Household status										F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Health										G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G

MoSeS Dynamic Model - Example

UNIVERSITY OF LEEDS

LLTI in 2031



Data Linkage

UNIVERSITY OF LEEDS

Usually conceived in MSM as a synthetic procedure

- Example linkage between HSAR and BHPS
- Linkage process based on shared attributes
- Types of match
  - Exact match
    - One-to-many (Monte Carlo)
    - One-to-one
      - Danger of accidental matches?
        - Perturb input data?
        - Restrict attribute set (dynamically?)
  - Fuzzy match

Data Linkage (Example)

UNIVERSITY OF LEEDS

SAR Data

- Age
- Gender
- Ethnicity
- Marital status
- Occupation
- Additional variables (x25)

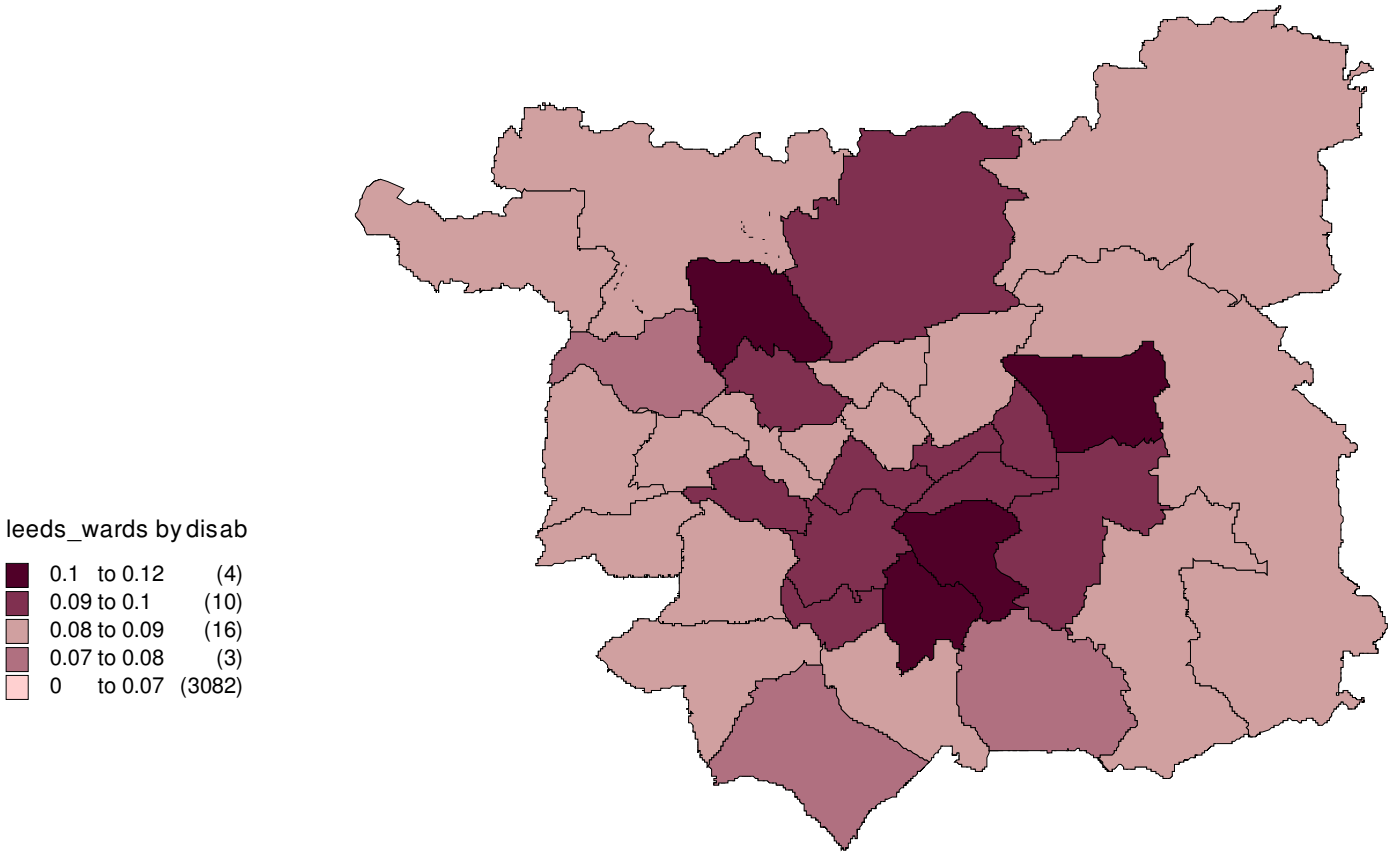
BHPS Data

- Age
- Gender
- Ethnicity
- Marital status
- Occupation
- Additional variables (x1000+)

Data Linkage (Example)

UNIVERSITY OF LEEDS

Disability in Leeds, 2006



Spatial Modelling

UNIVERSITY OF LEEDS

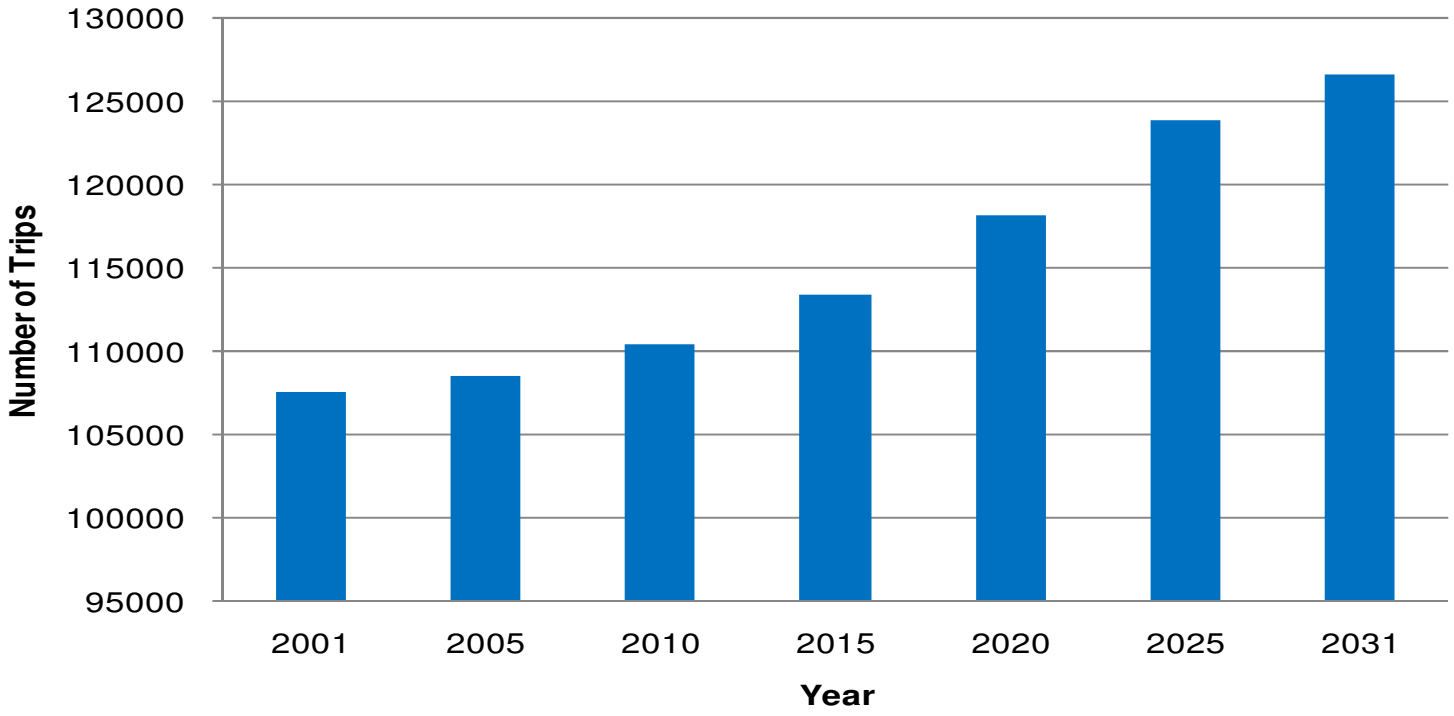
Use MSM data as inputs to

- Agent-based models
- Spatial interaction models
- Multi-level models?
- Econometric models?
- Pattern detection, geocomputation, network models, complex systems, AI, etc etc

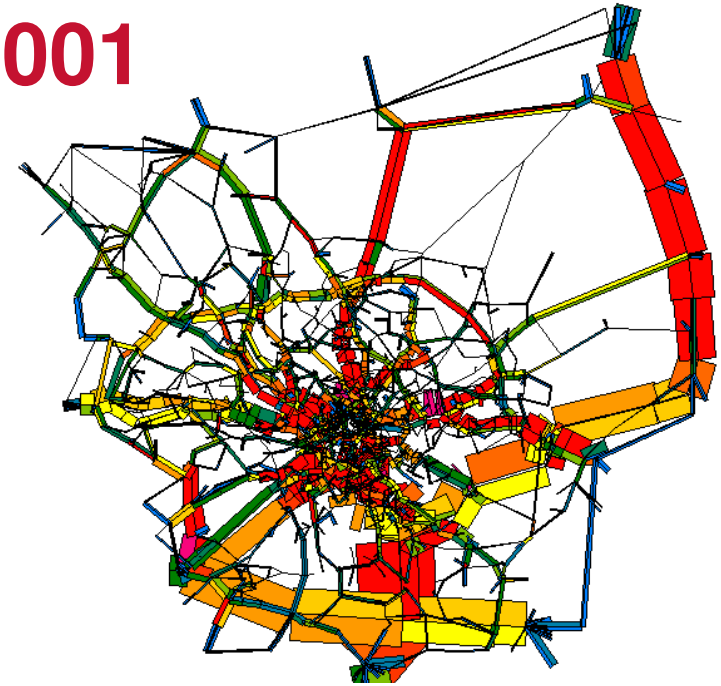
Modelling - Example

UNIVERSITY OF LEEDS

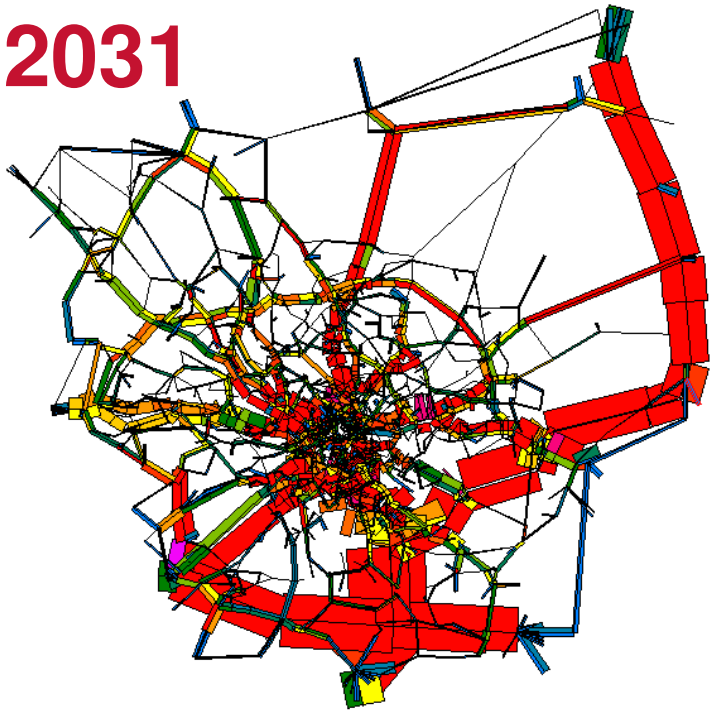
Number of trips during morning peak in Leeds from 2001 to 2030



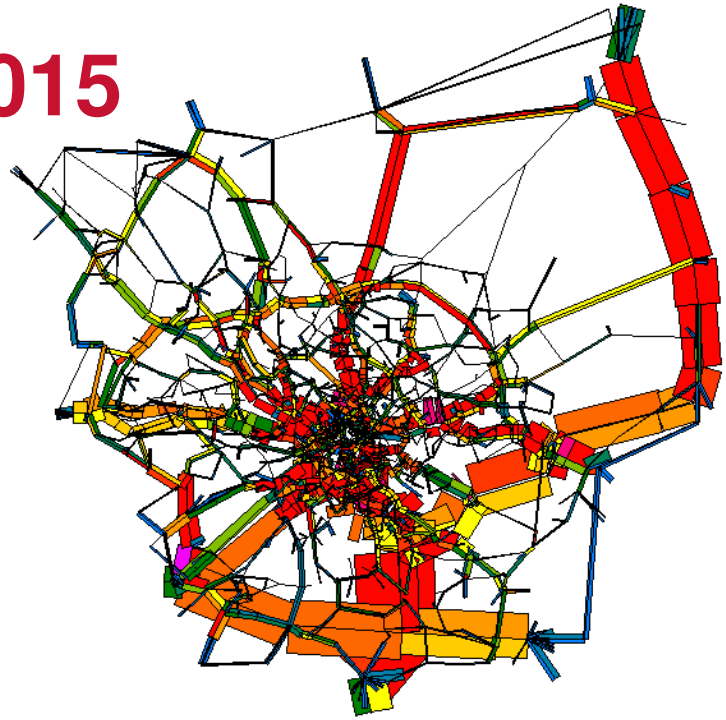
2001



2031



2015



Traffic Intensity *

	0		0.6
	0.1		0.7
	0.2		0.8
	0.3		0.9
	0.4		1.0
	0.5		

* Traffic Intensity=Traffic load/Road capacity

MSM and the NILS

UNIVERSITY OF LEEDS

Synthetic modelling/ Population Reconstruction

- Synthetic NILS dataset with modelled events (births, deaths, migration)
- Improve remote access for users (dummy analyses for implementation within a safe setting)
- Linkage to new data sources – education, transport,...?

Enhanced spatial models

- Models of morbidity, life expectancy which can be validated against NILS/ NIMS
- Representation of alcoholism, obesity, smoking...

Table 2: Estimates of birth statistics for a selection of wards in Leeds: first model run (source: Tomintz 2009)

Ward	Actual 2000-6	Predicted	Difference
Cookridge	294	298	4
Halton	187	195	8
Roundhay	229	232	3
Harehills	471	375	-96
Richmond Hill	272	214	-58
Seacroft	272	219	-53
Headingley	103	880	778

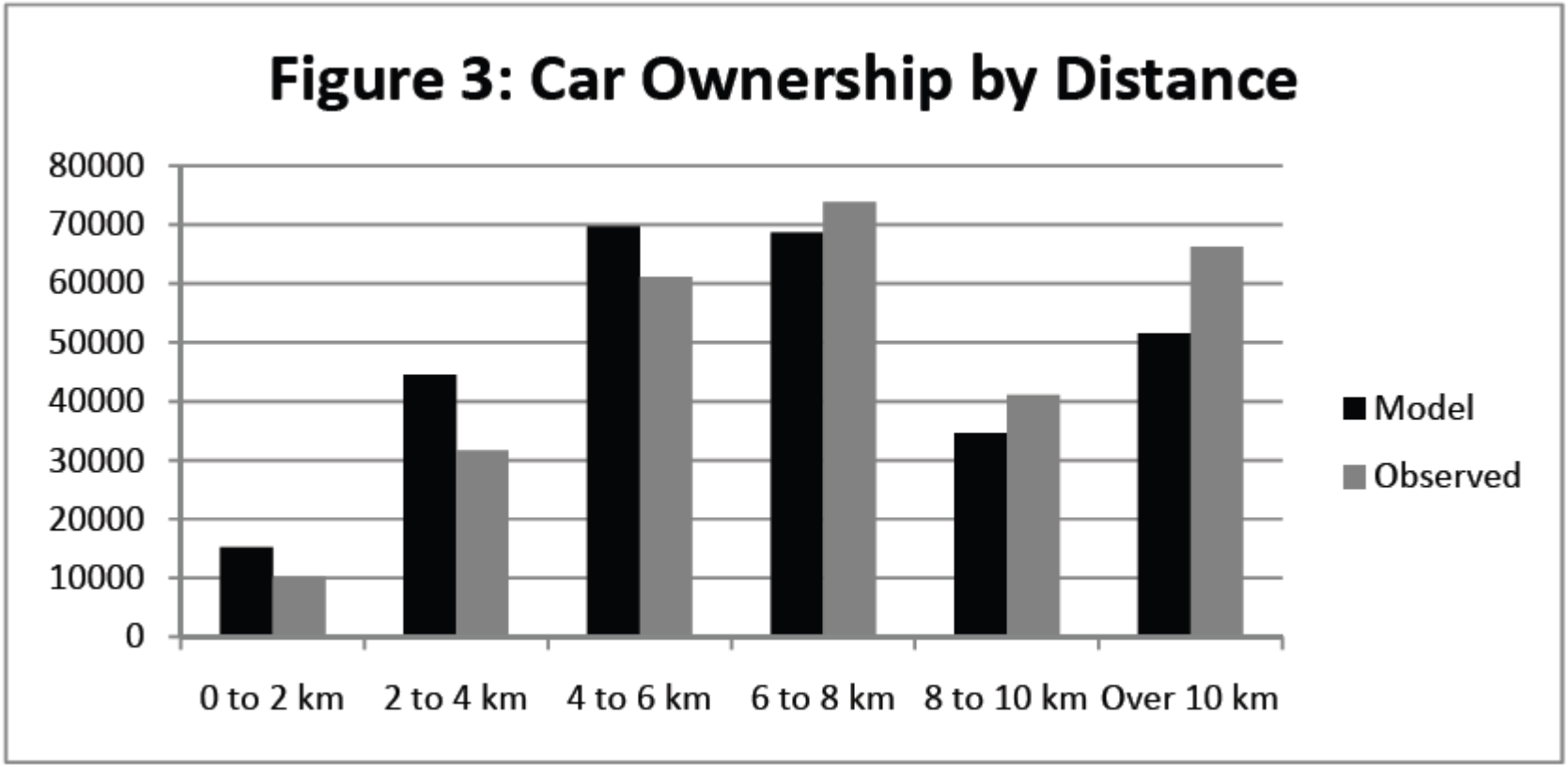
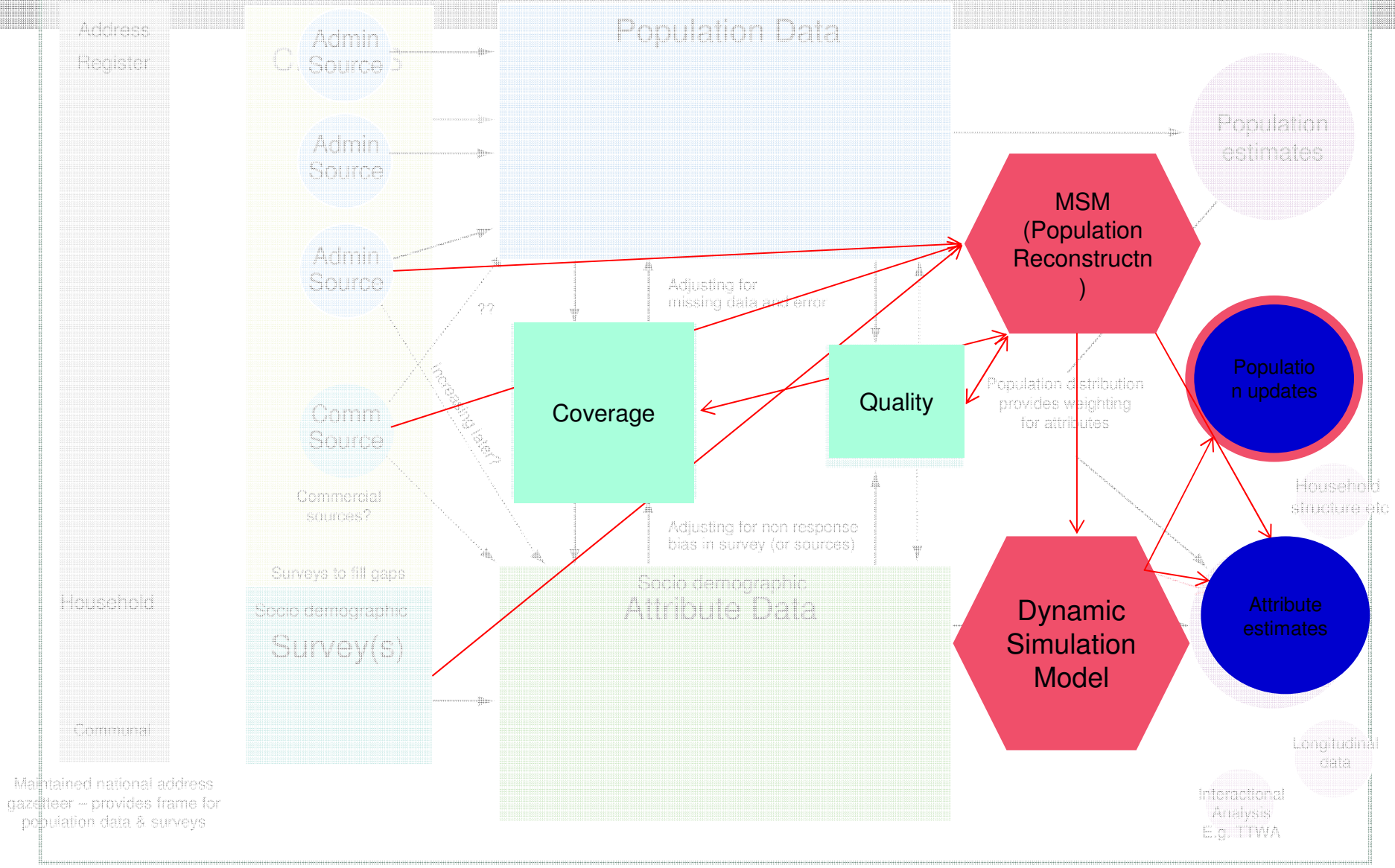


Figure 3: Car Ownership by Distance
80000
70000
60000
50000
40000
30000
20000
10000
0
0 to 2 km
2 to 4 km
4 to 6 km
6 to 8 km
8 to 10 km
Over 10 km
Model
Observed

MSM & NILS – Beyond 2011



Conclusions and Next Steps

UNIVERSITY OF LEEDS

Integration of MSM with the NILS:

- Synthetic data sets for research use and testing
- Substantive applications for data linkage, simulation and validation

Integration of MSM with GB LS

- Purposes as above
- Interaction with other datasets – Understanding Society??

Contribution of MSM/ data linkage to the Beyond 2011 programme?

Parallel development of linked simulations – MSM + SIM + ABM (Talisman) + other spatial and statistical techniques (NeISS)