# *Small Areas for the Republic of Ireland Implications for All-Island Datasets*

**Martin Charlton**

National Centre for Geocomputation
National University of Ireland Maynooth
Maynooth, Co Kildare


38. HOTEL ADELPHIA, PHILADELPHIA, PA

# Outline

- Spatial Data and Spatial Units
- Atomic Small Areas
- The show so far…

# Spatial Data and Spatial Units

- Spatial data is awkward stuff
- It tends to behave in unexpected ways
- This has implications for the analyses that we might wish to make

- It should be possible to obtain a reliable answer to the question “what was the average county male unemployment rate on April 28 2002?”

# Unemployment

- I can download the counts of unemployed males and economically active males for each county, calculate the rate for each county and take the average.
- It's 8.51%... or is it?
- If I use Electoral Divisions instead of counties and repeat the calculations it's 7.59%

# It gets worse...

- Taking the percentage of residents in social classes 6 and 7 we obtain
  - Counties: 23.15%
  - EDs: 22.62%
- And the correlation between the two indicators is different depending on whether I use Counties or EDs...
  - Counties: 0.67
  - EDs: 0.62

# ... and worse

- If I regress unemployment on social class I get two entirely different models for the differing areal units...

- $Unemp_{ed} = 0.46*SocCl_{ed} - 2.24$
- $Unemp_{cy} = 0.42*SocCl_{cy} - 1.82$

# Friday, December 29th 1933

2:30 *P. M.*—SECTION I (Rose Room, Adelphia Hotel)
(JOINT MEETING WITH THE AMERICAN SOCIOLOGICAL SOCIETY)

CHAIRMAN:
C. E. Gehlke, Western Reserve University

TOPIC:
**The Use of Census Tracts** [2]

PAPERS:

*Significance of Infant Mortality Rates for Small Geographic Areas*
Robert E. Chaddock, Columbia University [3]

*Sampling Errors and Interpretations of Social Data Ordered in Space and Time*
Frederick F. Stephan, Council of Social Agencies, Pittsburgh

*Some Problems in the Correlation of Spatially Distributed Variables*
Jerry A. Neprash, Franklin and Marshall College

*Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material*
C. E. Gehlke and Katherine Biehl, Western Reserve University

*Demonstration of Map Spotting Machine*
Earle E. Young [4]

DISCUSSION:
Howard Whipple Green, Health Council, Cleveland, Ohio

# A problem is described…

CERTAIN EFFECTS OF GROUPING UPON THE SIZE OF THE CORRELATION COEFFICIENT IN CENSUS TRACT MATERIAL[1]

By C. E. Gehlke and Katherine Biehl

Variations in the size of the correlation coefficient seem conditioned upon changes in the size of the unit used, with a smaller value of $r$ associated with the smallest unit rather than with the largest. Various ways

Journal of the American Statistical Association, 1934

# Modifiable Areal Units

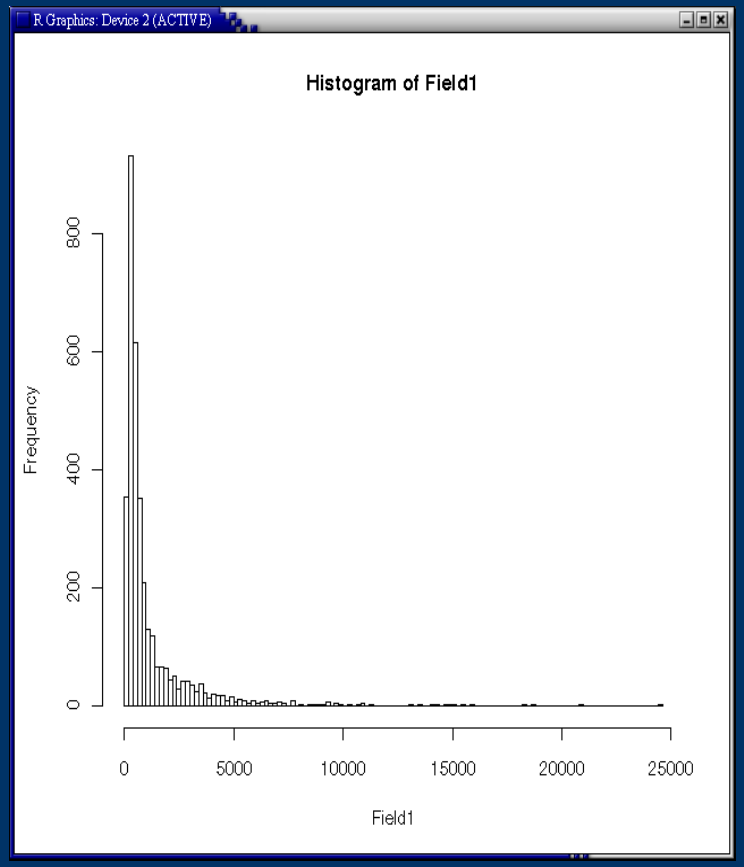
- Geographers have coined the term 'modifiable areal unit problem' for this phenomenon
- It affects all analyses of spatially aggregated data
- The size of the areal units affects the analysis
- The arrangement of the areal units affects the analysis
  - There are lots of ways of aggregating small units into the same number of larger ones

# Aggregated data

- Data from Censuses is usually released in some spatially aggregated form
- A trade-off between over-aggregation and confidentiality is required
  - County level: 8.51%
  - ED level: 7.56%
  - ? level: ?%
  - Ah, Mrs O'Leary, the census says your husband is out of work

# Electoral Divisions

- EDs have wildly varying populations
  - In 2002 the range was from 55 to 24405
  - The mean was 1145, and the median 526.5
  - The distribution is highly positively skewed
  - **EDs are the smallest areal units for which census data are available in the RoI**



# RoI/NI MAUP…

- **Electoral Divisions** in RoI
  - Average Population: 1144
  - Minimum: 55
  - Maximum: 24404
  - Variance: 1760.8
- **Output Areas** in NI
  - Average Population: 336
  - Minimum: 109
  - Maximum: 2582
  - Standard deviation: 83.1

# Atomic Small Areas

- One solution is to create a new geography in the RoI – areas smaller than EDs, but with a large enough population to avoid problems of confidentiality
- We can't easily follow the example of NI and create the equivalent of Output Areas
- We have no postcodes...

# Desiderata for Small Areas

- They should nest into EDs – ED boundary is sacrosanct
- They should cover the whole country – any location is in one and only one SA
- Minimum size must be 65 households
- Should be relatively homogeneous
- Should use street centrelines as a unifying feature
- Should be constrained by natural boundaries (rivers, railways, major roads)
- Should be ‘consistently small’

# Refinements

- Experiments indicate that in more rural locations a geography built on townlands was preferable
- Doughnut polygons (Kilkenny Rural is but one of many) have proved to be awkward to handle – there are 56 of these!
- Various technical issues
  - the procedure uses a mixture of vector and raster based approaches – some spline smoothing has been applied in the raster-vector conversion

# Projects

- Stage 1: The DOEHLG funded a pilot project to determine the feasibility of an automated approach – report submitted April 2005
- Stage 2: The Information Society funded a second study to test and develop the approach developed in Stage 1 – report submitted in January 2006

- Roll out: Ordnance Survey Ireland funded a project to roll out small areas for the whole of the RoI – work is in progress

# Building Blocks

- An address gazetteer, GeoDirectory, updated quarterly (An Post/OSi)
  - Residential delivery point counts
  - Commercial delivery point locations
  - Grid references available at building level
- Road centrelines from the 1:50000 Discovery data (OSi)
- Townland boundaries (OSi)
- Other boundaries – railway lines, watercourses, paths

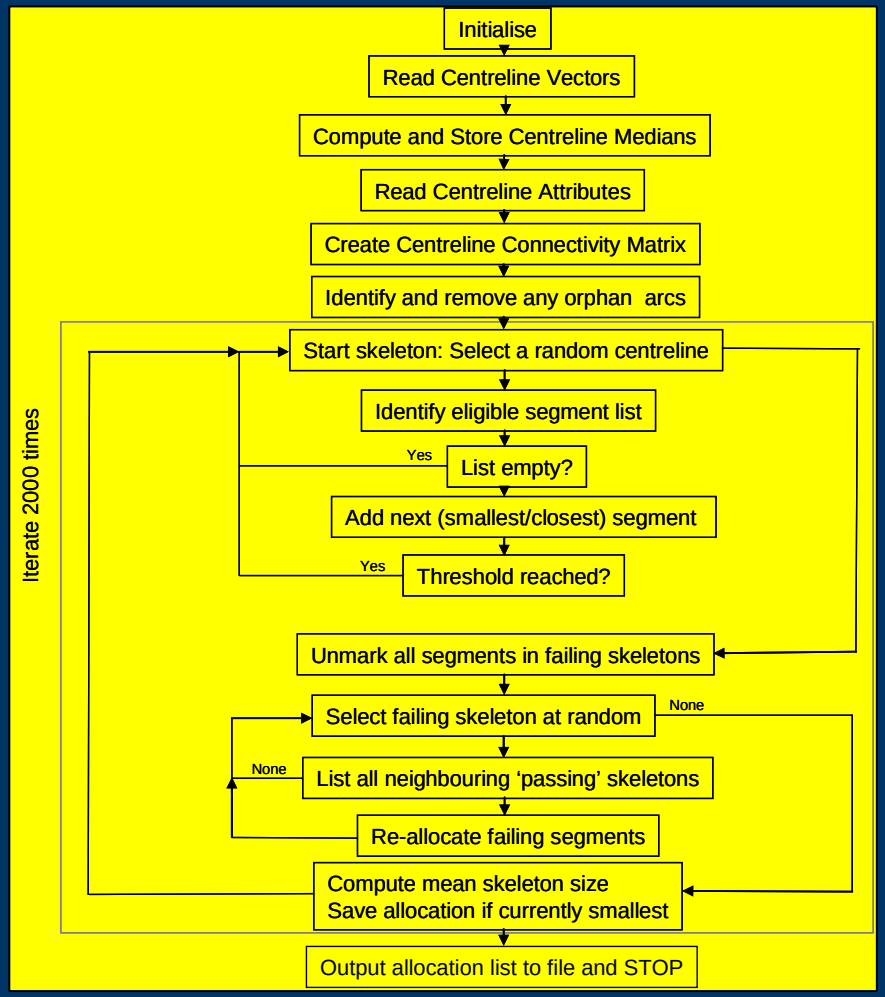
# Development

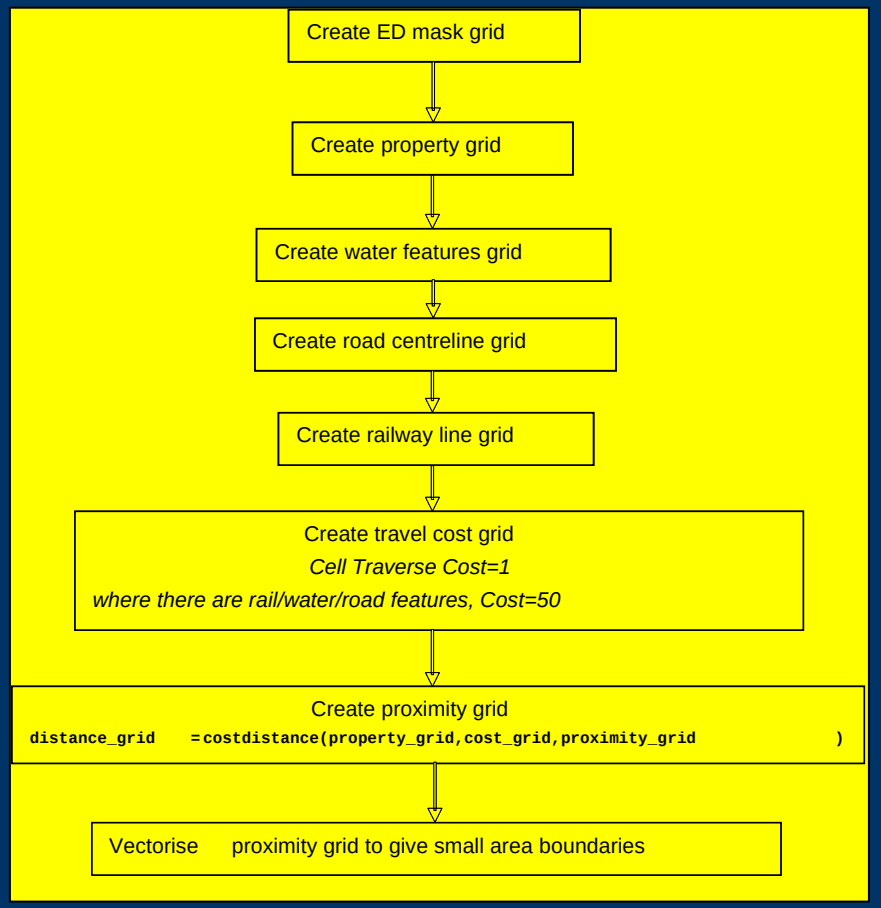
- Stage 1: algorithm implemented in ArcINFO and tested on Maynooth and Leixlip
- Stage 2: algorithm extended and further developed on 10 other EDs

- Roll out:
  - algorithm extensively modified, external code modified
  - Data obtained for whole of RoI
  - Move from R & D to Production phase – “tooling up”

# Algorithm

- The ‘algorithm’ is implemented in ArcINFO
  - It has a rich command language
  - It has a flexible and easily modified macro language
  - It runs on a unix workstation
- Certain stages are implemented by bespoke programs coded in FORTRAN and C++


Initialise
Read Centreline Vectors
Compute and Store Centreline Medians
Read Centreline Attributes
Create Centreline Connectivity Matrix
Identify and remove any orphan arcs
Iterate 2000 times
Start skeleton: Select a random centreline
Identify eligible segment list
Yes
List empty?
Add next (smallest/closest) segment
Yes
Threshold reached?
Unmark all segments in failing skeletons
Select failing skeleton at random
None
None
List all neighbouring 'passing' skeletons
Re-allocate failing segments
Compute mean skeleton size
Save allocation if currently smallest
Output allocation list to file and STOP


Create ED mask grid
Create property grid
Create water features grid
Create road centreline grid
Create railway line grid
Create travel cost grid
Cell Traverse Cost=1
where there are rail/water/road features, Cost=50
Create proximity grid
distance_grid =costdistance(property_grid,cost_grid,proximity_grid )
Vectorise proximity grid to give small area boundaries

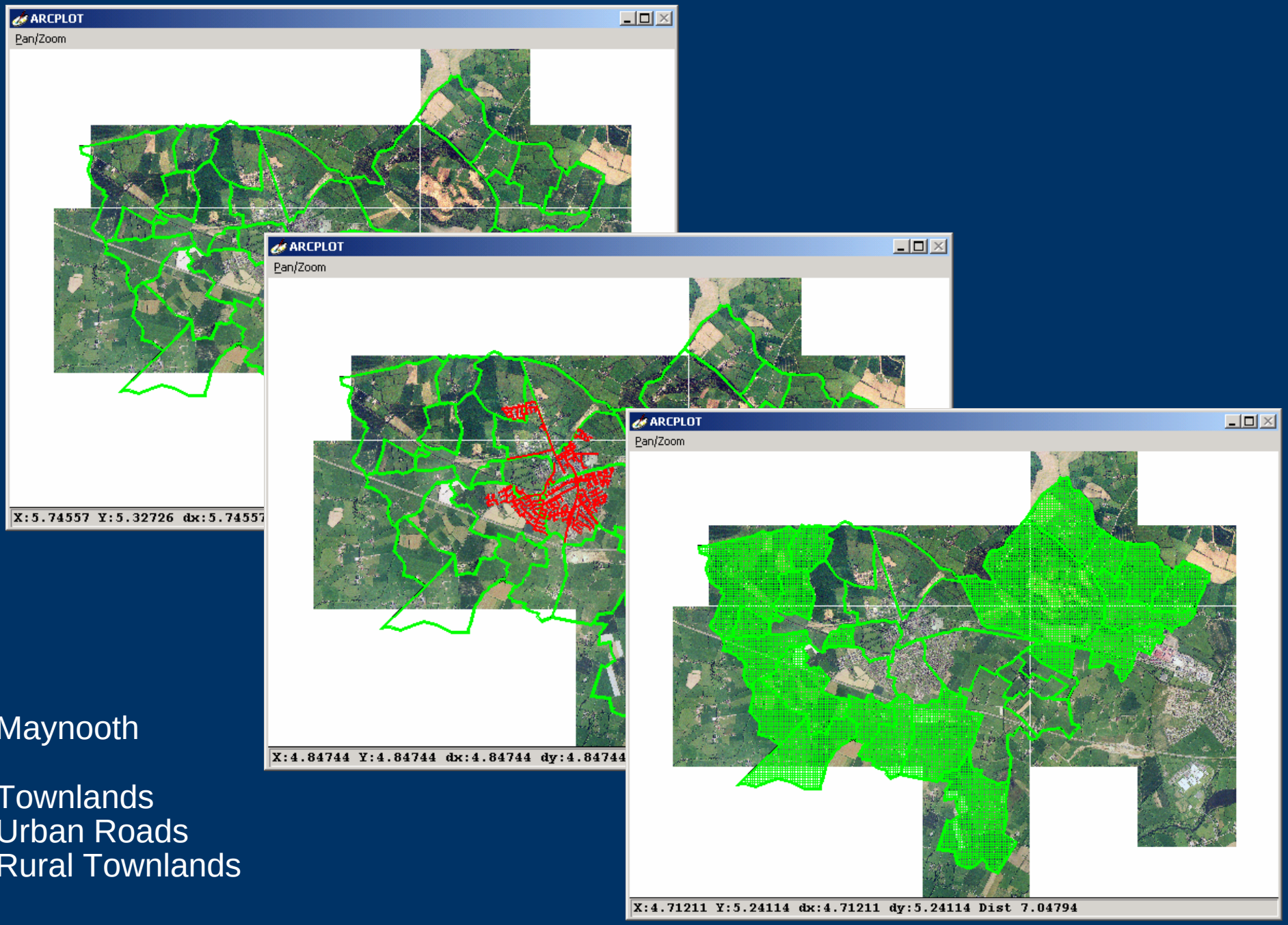
# Townland processing...

```
/*
/* Join the small area allocation to the PAT
/*
&if [iteminfo rural.pat -info small_area -exists] &then arc dropitem rural.pat
   rural.pat small_area
arc joinitem rural.pat rural.sa rural.pat rural# rural-id
&system echo Rural coverage and small area allocation table merged >>report.txt

&if [exists rural.totals -info] &then &sv x [delete rural.totals -info]
arc statistics rural.pat rural.totals small_area
sum sum-res_boxes
end
&system echo Small area residential delivery points tallied >>report.txt

&if [exists rural_proto -cover] &then arc kill rural_proto all
arc dissolve rural rural_proto small_area
&system echo Internal townland boundaries in each SA dissolved >>report.txt

arc joinitem rural_proto.pat rural.totals rural_proto.pat small_area small_area
&system echo Small area residential delivery point totals merged with proto
   coverage >>report.txt

&if [exists rural_sa -cover] &then arc kill rural_sa all
arc reselect rural_proto rural_sa
reselect sum-sum-res_boxes ge 65
[unquote ' ']
n
n
&system echo Rural small areas over 65 households created >>report.txt
```

# Algorithm

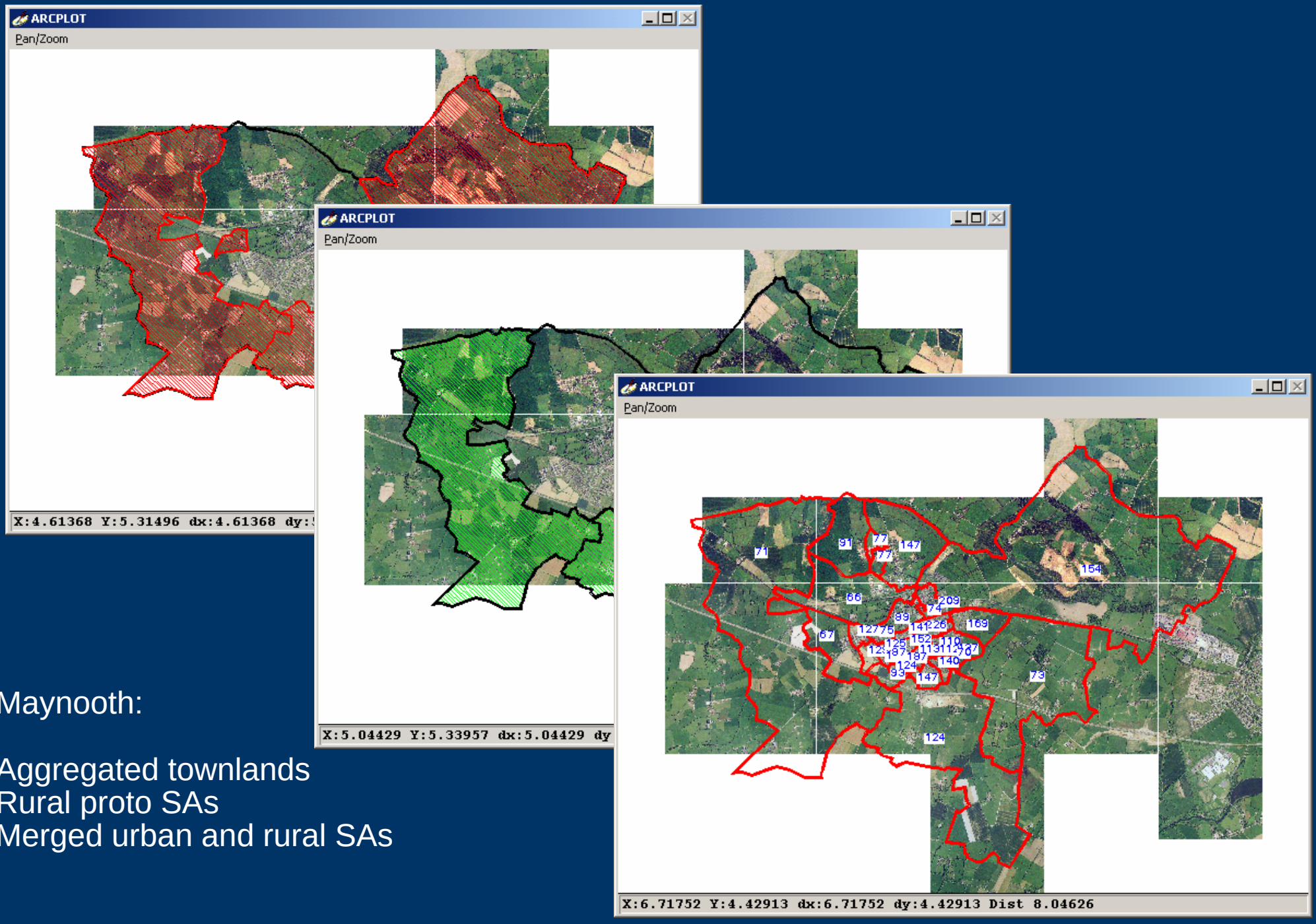
- The algorithm runs to some 1300 lines of code
- The external programs form another 3000 lines of code

- Moving from R & D to production is not a trivial task…
  - System random number generator replaced by Mersenne Twister
    - 623-dimensionally equidistributed uniform pseudorandom number generator
    - It will produce $2^{19937}$-1 numbers before repeating
    - Seed uses the system clock



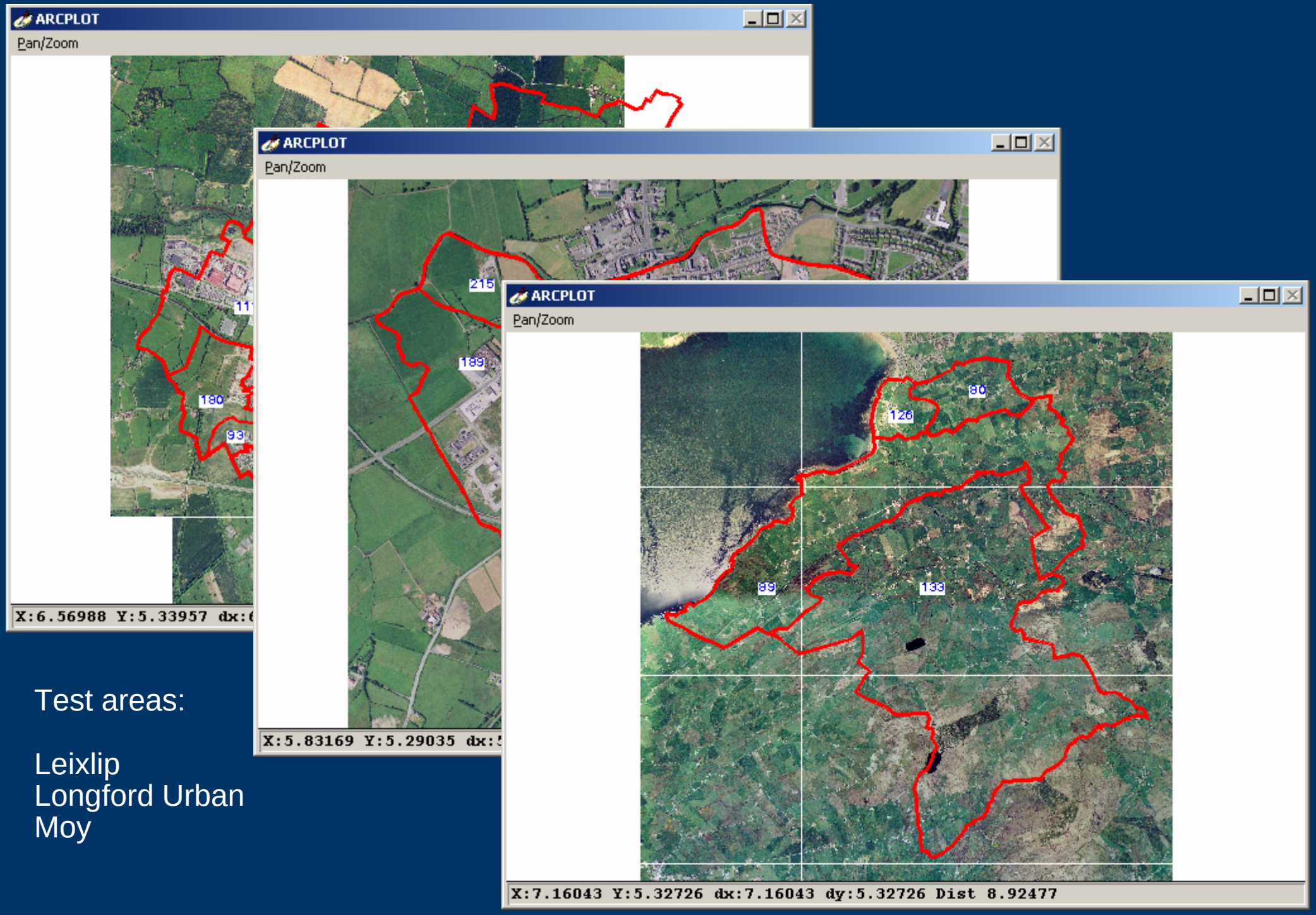

Maynooth

Townlands
Urban Roads
Rural Townlands



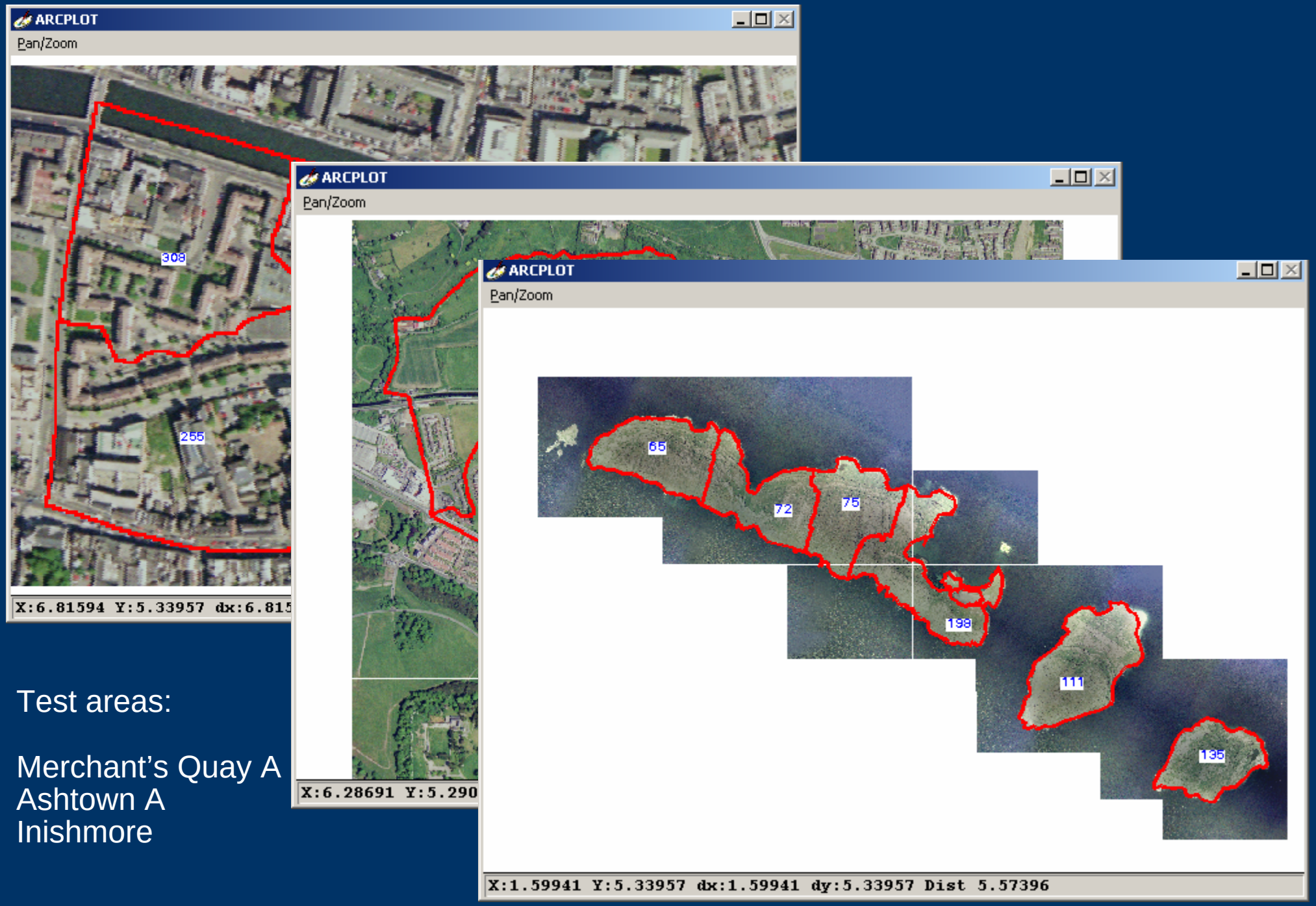

Maynooth:

Aggregated townlands
Rural proto SAs
Merged urban and rural SAs



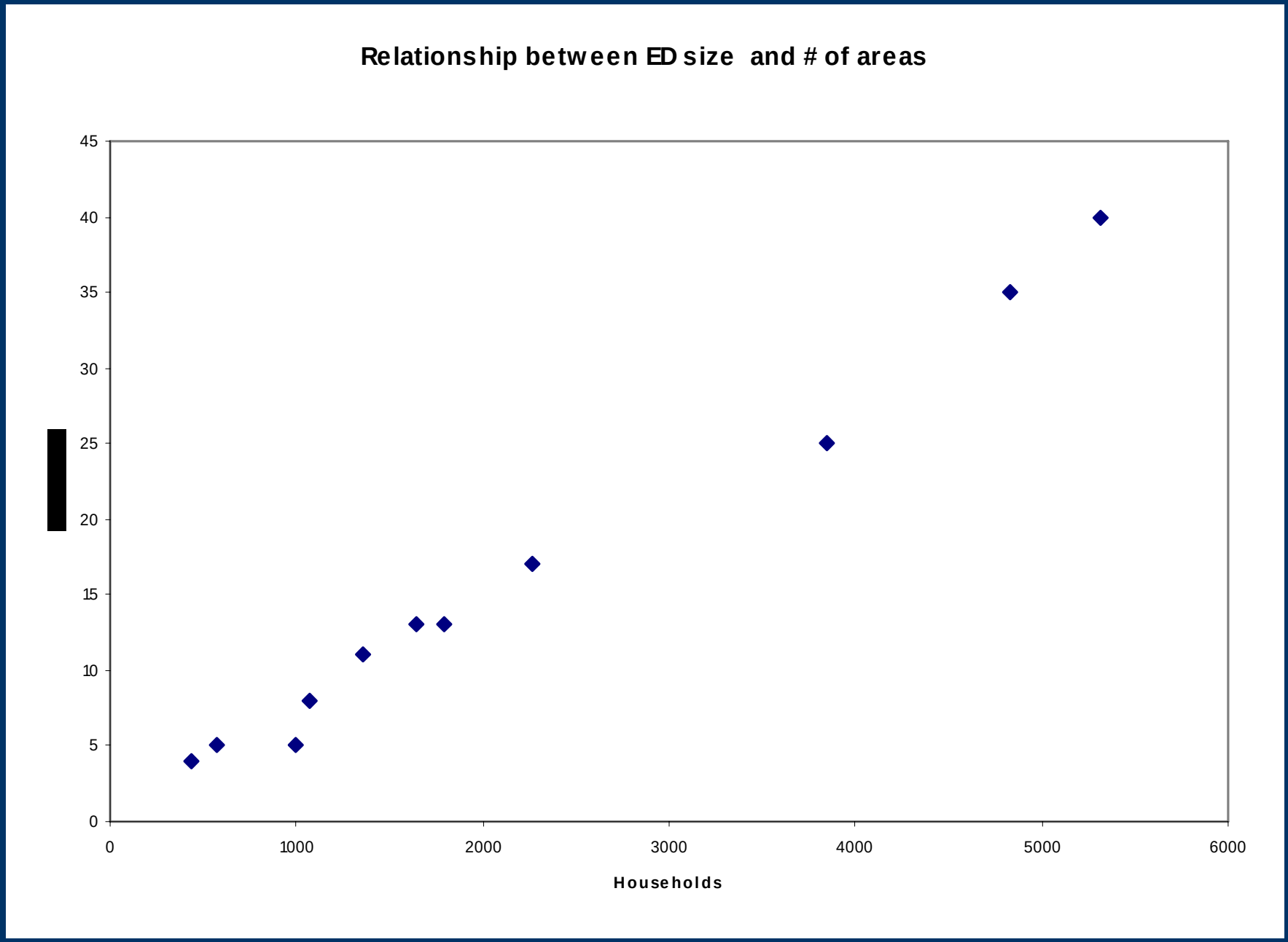

Test areas:

Leixlip
Longford Urban
Moy

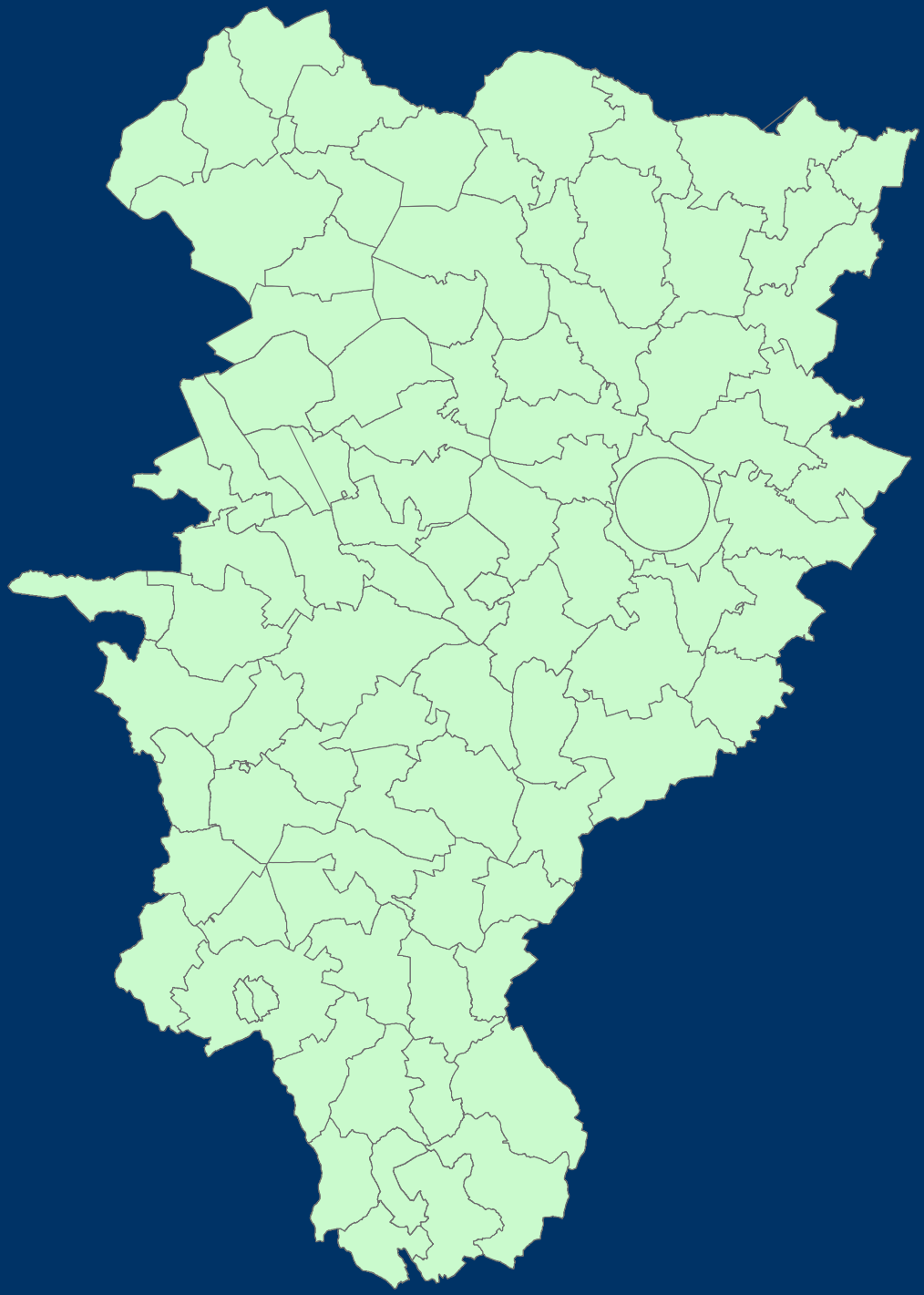



Test areas:

Merchant's Quay A
Ashtown A
Inishmore


Relationship between ED size and # of areas
45
40
35
30
25
20
15
10
5
0
0
1000
2000
3000
4000
5000
6000
Households

# Kildare



| County | Eds | Pop'n | H'holds | SAs | Pop/SA | HH/SA |
|---|---|---|---|---|---|---|
| **Carlow** | 54 | 46014 | 17195 | 194 | 237 | 89 |
| **Dublin** | 322 | 1122821 | 420429 | 4092 | 274 | 103 |
| **Kildare** | 89 | 163944 | 60957 | 635 | 258 | 96 |
| **Kilkenny** | 113 | 80339 | 29651 | 341 | 236 | 87 |
| **Laois** | 98 | 58774 | 22591 | 249 | 236 | 91 |
| **Longford** | 55 | 31068 | 12111 | 170 | 183 | 71 |
| **Louth** | 42 | 101821 | 38703 | 423 | 241 | 91 |
| **Meath** | 92 | 134005 | 53938 | 604 | 222 | 89 |
| **Offaly** | 87 | 63663 | 23769 | 264 | 241 | 90 |
| **Westmeath** | 106 | 71858 | 27064 | 322 | 223 | 84 |
| **Wexford** | 124 | 116596 | 45566 | 621 | 188 | 73 |
| **Wicklow** | 82 | 114676 | 42870 | 467 | 246 | 92 |
| **Clare** | 155 | 103277 | 38210 | 485 | 213 | 79 |
| **Cork** | 398 | 447829 | 167234 | 1900 | 236 | 88 |
| **Kerry** | 166 | 132527 | 48110 | 674 | 197 | 71 |
| **Limerick** | 173 | 175304 | 64225 | 749 | 234 | 86 |
| **Tipperary** | 175 | 140131 | 52367 | 645 | 217 | 81 |
| **Waterford** | 130 | 101546 | 38580 | 448 | 227 | 86 |
| **Galway** | 238 | 209077 | 78661 | 1011 | 207 | 78 |
| **Leitrim** | 78 | 25799 | 10646 | 164 | 157 | 65 |
| **Mayo** | 154 | 117446 | 43431 | 636 | 185 | 68 |
| **Roscommon** | 112 | 53774 | 20734 | 289 | 186 | 72 |
| **Sligo** | 82 | 58200 | 21480 | 302 | 193 | 71 |
| **Cavan** | 93 | 56546 | 21929 | 297 | 190 | 74 |
| **Donegal** | 149 | 137575 | 50415 | 719 | 191 | 70 |
| **Monaghan** | 70 | 52593 | 18655 | 255 | 206 | 73 |
| | **3437** | **3917203** | **1469521** | **16956** | **231** | **87** |

# Output

- Output will be a series of shapefiles with a small area codes
  - CSO code * 1000 + sequence #
  - OSi code * 1000 + sequence #
- Will also provide another table for GeoDirectory
  - Building ID from BUILDINGS table
  - Small Area codes

# Comparison

- Interim results for RoI Small Areas
  - Average Population: 231
  - Average Households: 87
- Output Areas in NI
  - Average Population: 336
  - Average Households: 125

# Outcomes

- By creating compatible geographies for the reporting units across the Island we can minimise the effects of the MAUP
- This means that comparisons across the border are not affected by differential scaling
- Should we wish to create higher level geographies which are not tied to administrative units, we have a consistent set of basic spatial units to work from

# Almost there…

- Inevitably the output from an automated process may lead to some oddities
  - Heron's beaks (caused when two natural boundaries run a few metres apart in parallel)
  - Bow-ties
- Will require some manual editing
- 18 different measures of complexity to identify problem polygons…