

Rich Harris, Spatial Modelling Research Group, Geographical Sciences

Towards an understanding of research ethics for computational, e-social science

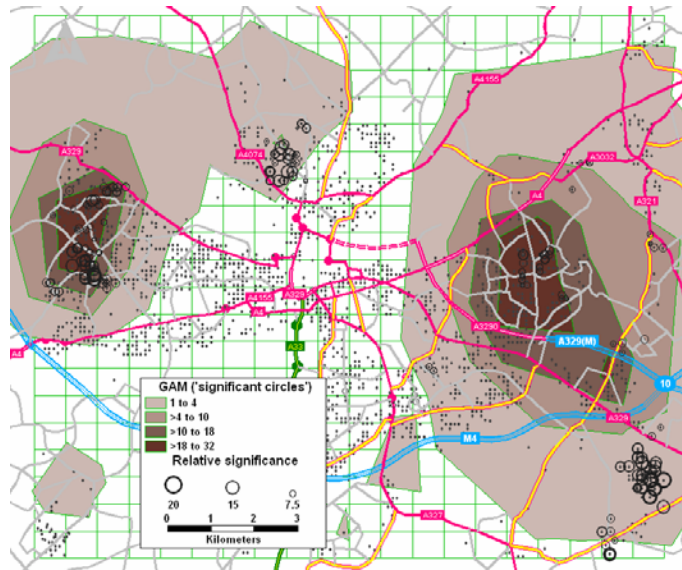
A geographical perspective



Context

- Grid computing as a practice of e-social science
- No one clear definition of what e-social science is or does but
 - “reasonable consensus that, at minimum, e-social science includes High Performance Computing, Grid computing and associated technologies...” (Fielding, nd, p.8)
 - Computational Grid; Access Grid; Data Grid
- Focusing here on e-social science as a computational project “characterized by massive distributed datasets, distributed computing and the need for complex visualizations” (Martin, 2005 p.2).

Why the interest?



$$\text{GCSESCORE}_{ijk} = \beta_{0jk} + -2.833(0.468)\text{FSM}_{ijk} + -3.534(0.327)\text{Boy}_{ijk} + \\ 0.146(0.076)\text{KS2MN}_{ijk} + 2.205(0.053)\text{KS3MN}_{ijk} + \\ 0.227(0.056)\text{Sch_GCSE}_{jk} + \beta_{6k}\text{ethnic_diff}_{jk} + e_{ijk}$$

$$\beta_{0jk} = -46.018(2.136) + v_{0k} + u_{0jk}$$

$$\beta_{6k} = -0.788(0.987) + v_{6k}$$

$$\begin{bmatrix} v_{0k} \\ v_{6k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 2.246(0.751) & \\ -1.970(2.362) & 27.779(12.090) \end{bmatrix}$$

$$u_{0jk} \sim N(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 2.505(1.182)$$

$$e_{ijk} \sim N(0, \sigma_e^2) \quad \sigma_e^2 = 85.625(2.177)$$

$$\text{Deviance(MCMC)} = 24974.350(3427 \text{ of } 3849 \text{ cases in use})$$

← abduction

induction

→ deduction

- Both are based on subsets of extremely large datasets
- Both involve iterative repeat testing to construct confidence intervals (to judge significance)

The spatial dimension

- The data I use are geographically referenced
- Grid computing as an evolution of:
 - GIS (geographical information systems)
 - Geocomputation
 - Geosimulation
 - Geovisualization
- 13 September 2005 Agenda Setting Workshop: GIS and the Grid (Leeds)
- “Some of the enduring fundamentals of socioeconomic GIS research agendas remain to be addressed, but to these should be added the development of appropriate spatial methodologies for e-social research” (Martin, *ibid.*, p.3).

The rub

- GIS and its associated technologies have been sharply debated in (human) geography where their place in social research have been strongly contested and resisted.
(see proceedings for brief history and explanation)
- Similar concerns and hostilities will be encountered by e-social science?

The 13 lines of critique against GIS in 'Ground Truth' (source: Openshaw, 1997)

- relevance to a computationally-based e-social science?
 - a GIS geography can be condemned as a handmaiden of the state;
 - GIS is a revival of discredited positivist epistemologies; GIS emphasizes maps and spaces;
 - GIS is an ordered representation of space within a Cartesian framework;
 - GIS is a discipline dominated by facts rather than knowledge and understanding;
 - a GIS geography implies a neglect of themes not included in the data;
 - GIS continues and reinforces a structural distortion of knowledge and uneven data access;
 - GIS is not a value-free technology;
 - GIS has a propensity for evil intent;
 - GIS is the technology of surveillance;
 - GIS software used for direct marketing are an apotheosis of instrumental reason;
 - grandiose claims and hype;
 - the inevitability of ethical inconsistency.

Three broad lines of critique

- Constructivist
 - Scientific enquiry as socially constituted by the networks in which they operate as opposed to any objective and teleological progress
- Critical social theorist
 - Socially constituted and socially situated technologies have social consequences and effects
- The humanistic
 - GIS as ‘thin description’
 - “Capable of identifying certain characteristics and patterns of data, but incapable of describing or explicating the meaningful nature of social life” (Cloke et al., 2004 p.283)
 - Contrasts with ‘thick description’
 - “information about the context of an act, the intentions and meanings that organise actions and its subsequent evolution” (*op. Cit.*)
 - Social science is an oxymoron?
 - Computational social science more so?

Are spatial scientists uninterested in ethics?

- “[Spatial scientists] tend to reduce methodology to technique, being bothered about the correct running of an appropriate statistical test but less about anything entailed in the deriving of the data on which the test is conducted (unless relevant to deciding on which particular test is suitable), *not anything following conceptually, politically, ethically or otherwise from choosing to tackle the data statistically rather than some other way*” (emphasis added)
 - A powerful and easily reproduced critique
 - ‘I am not a number...’
 - Although not actually true...

- Esnard's (1998) 'Portable, Provisional Codes' for GIS use in local governance:
 - "[Focus] is on process, data quality, community needs, liability, and data access as the framework for *reflective* use of GIS technology" (p.37, emphasis added).
- URISA's & GIS Certification Institute's Code of Ethics
 - "based on the ethical principle of always treating others with respect and never merely as means to an end... It requires us to consider the impact of our actions on other persons and to modify our actions to reflect the respect and concern we have for them. It emphasizes our obligations to other persons, to our colleagues and the profession, to our employers, and to society as a whole."

Towards an understanding of research ethics for computational, e-social science

- Concerned with data protection (and IPR) and data privacy
- But more than this because:
 - Data protection precisely is that – the protection of data not the data subject (Cho, 2005)
 - Informational privacy is not synonymous with personal privacy
 - Personal privacy may routinely be infringed by ‘not personal’ data

Towards an understanding of research ethics for computational, e-social science

- Concerned with openness, accessibility and accountability, e.g. by using open source programming to permit examination of ‘software sorting’
- But more than this because:
 - “code does not simply exist, created and working in easily defined and examined ways... is not... deterministic (i.e., code determines in absolute, nonnegotiable means, everyday practices) or universal (i.e., such determinations occur in all places and all times in a simple cause-and-effect manner). ... [T]here is no neat marriage between coded objects, infrastructures, processes and assemblages, and [particular effects of code. Instead, technicity varies as function of the nature of code, people, and context” (Dodge & Kitchin, 2005 p. 170).

Towards an understanding of research ethics for computational, e-social science

- Concerned with social shaping and knowledge (re-) production
 - “the likely effects of the widespread adoption of grid technologies on [the construction of] (social) scientific knowledge” (Woolgar, nd pp. 1-2).
- But more than this:
 - “to include all social aspects of the genesis, use, implementation and effects of the new e technologies” (*op. Cit.*).
 - The likely effects on those about whom social scientific knowledge is being generated.

Towards an understanding of research ethics for computational, e-social science

- Concerned with the socially situated nature, limits and possible consequences of a computational, e-social science
 - “Grid technology does not only provide speed. It gives us the ability to store and process virtually unlimited volumes of data at a lower cost. Now we can keep customer data over the whole lifetime of interaction with a company. And, as the more data you put into a model the better it is, we will be able to build models which have the depth of data and learning that enables them to be truly predictive... I firmly believe that grid technology will be one of the biggest influences in driving direct marketing into a truly data-focused discipline” (letters page, *Database Marketing*).
- But more than that because:
 - I can see no sense in becoming moribund by over-reaction or disengagement

A footnote from recent history

- The timing of publication of ‘Ground Truth’ obscures the meetings that were taking place between ‘the protagonists’ and ‘the antagonists’ of GIS (Schuurman, 2000)
 - Critical GIS
- Recognise the ESRC’s recent call for proposals to investigate the social shaping and the social impact of e-(social/) science.
- Time for NCeSS’ own ‘Friday Harbor’?
 - To debate the epistemological, ontological and ethical framework for computational, e-social science.