Formal Modelling
(of social phenomena)

A Specialist Method

MRes, MMUBS

Me – Bruce Edmonds

- *Senior Research Fellow* and *Director* of the Centre for Policy Modelling (CFPM)
- Since 1994 developed the CFPM with Scott Moss as a research centre specialising in agent-based social simulation (http://cfpm.org)
- Now one of the leading such teams in this area in the world, e.g. major UK and EU projects
- One of the few centres in complexity science in the UK for a long time
- Editing a handbook: "*Simulating Social Complexity*" for Springer due out in 2009

What is a model?

Something, A, that is used to understand or answer questions about something else, B

- e.g: A scale model to test in a wind tunnel
- e.g: The official accounts of a business
- e.g: The minutes of a meeting
- e.g: A flow chart of a legal process
- e.g: A memory of a past event
- e.g: A computer simulation of the weather
- e.g: The analogy of fashion as a virus

Models usually abstract certain features and have other features that are irrelevant to what is modelled

What is a *formal* model?

Something that (in theory) can be written down precisely, whose content is specified without ambiguity

- e.g: mathematical/statistical relations, computer programs, sets of written rules

Can make exact copies of it

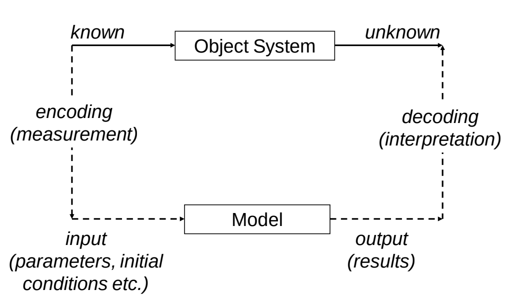
Agreed rules for interpreting/using them

Can make *certain* inferences from them

- *Not*: an analogy, a memory, a physical thing

There ***are grey areas****, degrees of formality*

The Modelling relation



Modelling Purposes

All modelling has a purpose (or several)

Including:

- Description
- Prediction
- Establishing/suggesting explanations
- Illustration/communication
- Exploration
- Analogy

These are frequently conflated!

The Modelling Context

All modelling has a context

- The background or situation in which the modelling occurs and should be interpreted
- Whether explicit or (more normally) implicit
- Usually can be identified reliably but not described precisely and completely
- The context inevitably hides many implicit assumptions, facts and processes

Modelling only works if there is a reliably identifiable context to model *within*

Descriptive formal models

Describes in precise terms the state(s) of what is observed

- e.g. the average height of a group of people
- e.g. The words that an individual said
- e.g. the correlation of height with arm span

A sequence of descriptive "snap-shots" can describe aspects of a process

- e.g. A Time series of average wages in UK

Evidence is often recorded as descriptive formal models

All sets of "data" are descriptive models

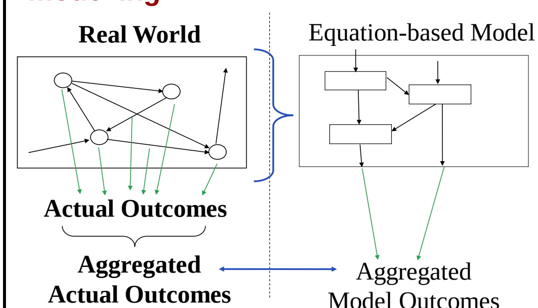
Analytic formal models

Where the model is expressed in terms that allow for formal inferences about its general properties to be made

- e.g. Mathematical formulae
- Where you don't have to compute the consequences but can *derive* them logically
- Usually requires numerical representation of what is observed (but not always)

Only fairly "simple" mathematical models can be treated analytically – the rest have to be simulated/calculated

Equation-based or statistical modelling



Statistical formal models

Where the collective properties of a group are modelled, eliminating some assumed randomness between individuals

- *Descriptive statistics* just summarise aspects of a group that are assumed to be representative of that group
- *Generative statistics* are a model of some process done using the combination of a target trend plus additional randomness

*Statistical models often rely on the "*Law of Large Numbers*" – that certain aspects are irrelevant and can be treated as random*

An analogy: An Ideal Gas

- ***The idea***: although the motion of each particle in the gas is not predictable, *taken together* the gas obeys regular laws and is predictable
- This is an idea that has seeped into the social sciences
- (Asimov 1962, page 7): "*Psycho-history dealt not with man, but with man-masses. It was the science of mobs; mobs in their billions ... The reaction of one man could be forecast by no known mathematics; the reaction of a billion is something else again*"

Problems with this idea…

- This only “works” if there is a signal that is separable from noise and…
  - …the “noise” is essentially random (Law of Large Numbers)…
  - …or can be safely ignored.
- But it is almost impossible to know either of these for sure!
- e.g. in stock markets, what seems to be random noise is rather the result of subtly linked social processes
- In other words, the context of modelling is inadequate and “leaky”

Computational formal models

Where a* process *is modelled in a series of precise instructions (*the program*) that can be “run” on a computer

- The same program always produces the same results (essentially) but...
- ...may use a “random seed” to randomise certain aspects
- Can be simple or very complex
- Often tries to capture more “qualitative” aspects of social phenomena

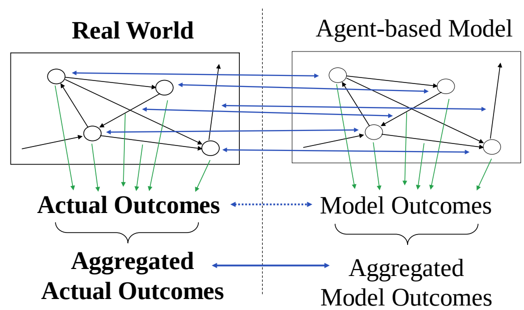
Example of Computational Model:
Schelling’s Segregation Model

Schelling, Thomas C. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* **1**:143-186.

Rule: each iteration, each dot looks at its 8 neighbours and if less than 30% are the same colour as itself, it moves to a random empty square

Segregation can result from wanting only a few neighbours of a like colour

Agent-based simulation



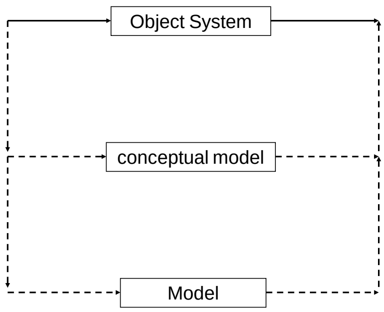
Characteristics of agent-based modelling

- Computational descriptions of processes
- Not analytically tractable
- More context-dependent…
- … but assumptions are much less drastic
- Detail of unfolding prcesses accessible
  – more criticisable (including by non-experts)
- Used to explore inherent possibilities
- Validatable by expert opinion *and* data
- Often very complex themselves

A trouble with such simulations

- Is that they are highly suggestive
- Once you play with them a lot, you start to “see” the world in terms of you model – a strong version of Kuhn’s *theoretical spectacles*
- They can help persuade beyond the limit of their reliability
- They may well not be directly related to any observations of social phenomena

Modelling a concept of something



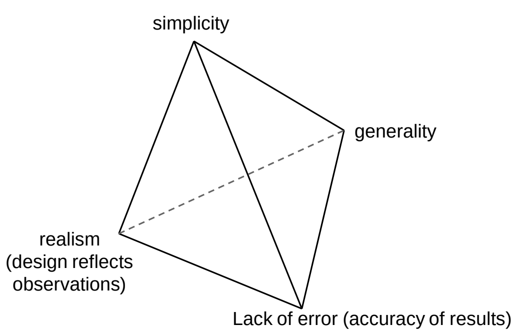
Sepecial Methods – formal modelling, http://cfpm.org/mres slide-19

Some Criteria for Judging a Model

- Soundness of design
  - w.r.t. knowledge of how the object works
  - w.r.t. tradition in a field
- Accuracy (lack of error)
- Simplicity (ease in communication, construction, comprehension etc.)
- Generality (when you can safely use it)
- Sensitivity (relates to goals and object)
- Plausibility (of design, process and results)
- Cost (time, effort, etc.)

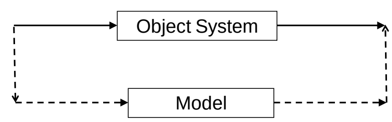
Sepecial Methods – formal modelling, http://cfpm.org/mres slide-20

Some modelling trade-offs



Sepecial Methods – formal modelling, http://cfpm.org/mres slide-21

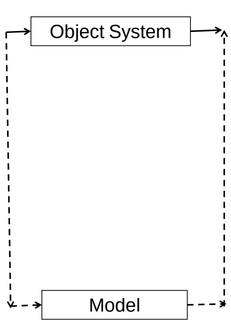
Complex Descriptive Model



Complex but directly relevant model – strong mapping to model, weak inference within model

Sepecial Methods – formal modelling, http://cfpm.org/mres slide-22

Abstract Theoretical Model



Simple model but abstract – strong inference within model, but weak mappings to and from the model

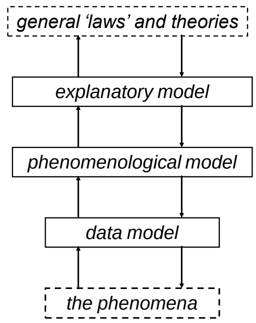
Sepecial Methods – formal modelling, http://cfpm.org/mres slide-23

Semantic complexity

- The difficulty of interpreting a rich meaningful domain and descriptions into an impoverished formal model
- Establishment of symbol meaning by:
  - Importing symbols from natural language
  - Use of symbols in context
  - Cycle of interaction and learning about symbols
  - Imputation by stakeholders and domain experts
- It is *very difficult* to go from models that strongly relate to data and those that give meaningful explanations
- But good science is when you have both

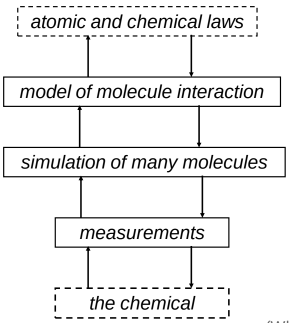
Simulation and Complexity - how they might relate, Oxford 2003, http://cfpm.org/~bruce slide-24

A possible layering of models (by abstraction)



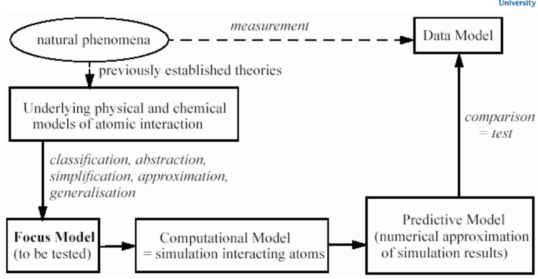
(What really happens)

A possible layering of models (by granularity and abstraction)



(What really happens)

An example from chemistry



Multiple models

- Parallel models
  - e.g. different models gained by different approaches and simplifications, whose results are compared (e.g. Lasers)
- Context-specific models
  - e.g. quantum models in micro-world and relativistic models in macro-world
- Clusters of models
  - e.g. use of analogical models alongside formal models in atomic physics

(What really happens)

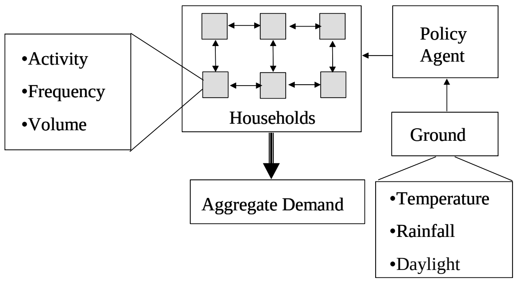
An Example

- Type: A complex agent-based descriptive simulation
- Context: statistical and other models of domestic water demand under different climate change scenarios
- Purposes:
  - to critique the assumptions that may be implicit in the other models
  - to demonstrate an alternative

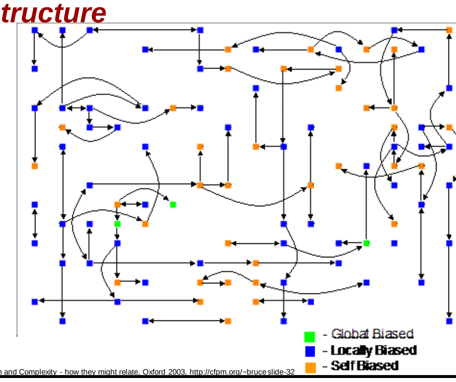
A *model of social influence and water demand*

- Investigate the possible impact of social influence between households on patterns of water consumption
- Design and detailed behaviour from simulation validated against expert and stakeholder opinion at each stage
- Some of the inputs are real data
- Characteristics of resulting aggregate time series validated against similar real data

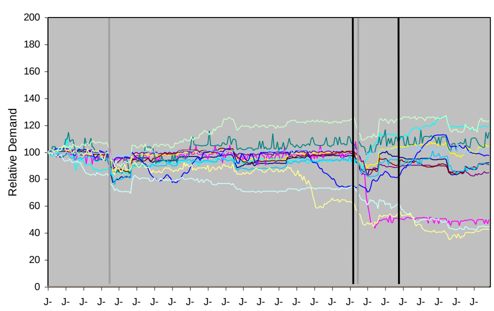
Simulation structure



Some of the household influence structure



Example results



Conclusions from Example

- The use of a concrete descriptive simulation model allowed the detailed criticism and, hence, improvement of the model
- The inclusion of social influence resulted in aggregate water demand patterns with many of the characteristics of observed demand patterns
- The model established how it *was possible that* processes of mutual social influence could result in widely differing patterns of consumption that were self-reinforcing

Useful?

- It does show some possible weaknesses and limitations in traditional statistical models
- The model has been imitated by researchers in Spain
- The local authority uses it to assess new residential developments to see some of the possible effects on water demand that could result
- Is this a good idea?

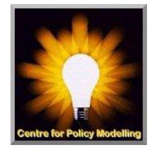
Conclusion – *advantages of formal modelling (for the social sciences)*

- Impressive ☹
- Little confusion about model
- Formal model can be copied and tried by others –a social “evolutionary” process
- Relatively easy to confront with evidence
- Strong inference step
- Helps unearth assumptions
- Suggests new questions to investigate
- Can be shown to be wrong (Popper) or *better* how it is wrong

Conclusion – *disadvantages of formal modelling*

- Impressive ☹
- Poor in terms of meaning
- Requires expertise
- Easy to fool oneself into thinking the world is like your model
- Tempting to take short-cuts
- Difficult to validate completely
- Difficult to list all assumptions
- Needs **lots** of evidence

The End



Bruce Edmonds
bruce.edmonds.name
Centre for Policy Modelling
cfpm.org
Slides
cfpm.org/mres