Introduction to Urban Modelling and Simulation

Topic 1: Models and Definitions

Michael Batty

m.batty@ucl.ac.uk

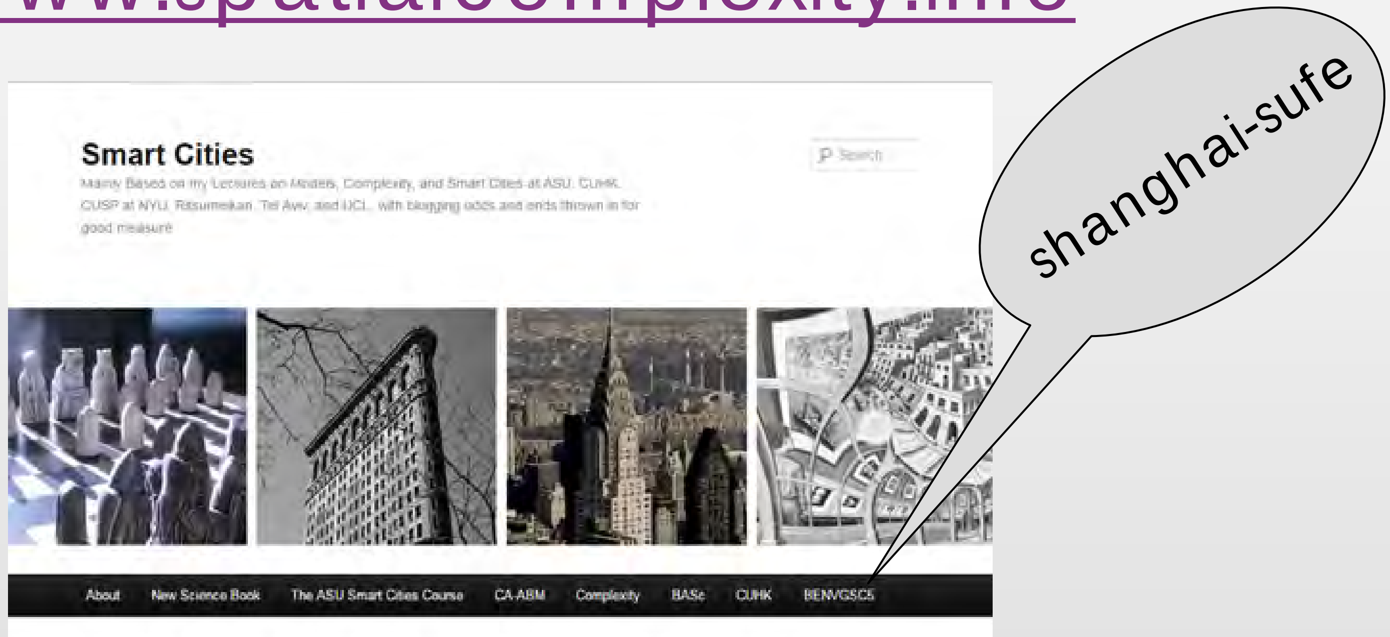
@jmichaelbatty

8 July, 2018

http://www.spatialcomplexity.info/

I will post the lectures on my blog

www.spatialcomplexity.info



You will find the link on this blog at

http://www.spatialcomplexity.info/shanghai-sufe

NOTHING THERE YET!

Outline: Topic 1: Models & Definitions

- What are Models? Relationships to Theory
- Definitions of Models
- A Classification: Icons, Analogs, Symbols,
- Aggregate viz Disaggregate Modelling
- Statics viz Dynamics
- The Paradigm Shift: Aggregates to Agents
- The Model-Building Process
- Facts and Theories, Factoids and Stylized Facts
- Verification, Validation, Goodness of Fit
- Calibration and Estimation

What are Models? Relationships to Theory

A theory is an *abstraction* of some phenomena, usually '*real*' but sometimes imagined in a form that makes the *simplification* or abstraction clear. A model is a simplification of *reality* which takes the theoretical abstractions and puts it into a form that we can manipulate. Simulation is often used to characterise this process of implementation.

In everything we do, we theorise, and more and more frequently we build models to demonstrate theory.

This is all fairly obvious – but the focus on theory is important because theory can be implicit as well as explicit. In fact in our growing quest to describe the world through models, theory is tending to become part and parcel of models.

The main reason for beginning with theory is that the conventional wisdom of science begins with theory and then *tests* theory against observations – data. It is impossible to approach the world without prior theory and without getting involved in where theory comes from, let us assume that whenever we model a phenomena we have in mind some theory no matter how implicit. Its sometimes hard to even extract the theory we hold but it always there.

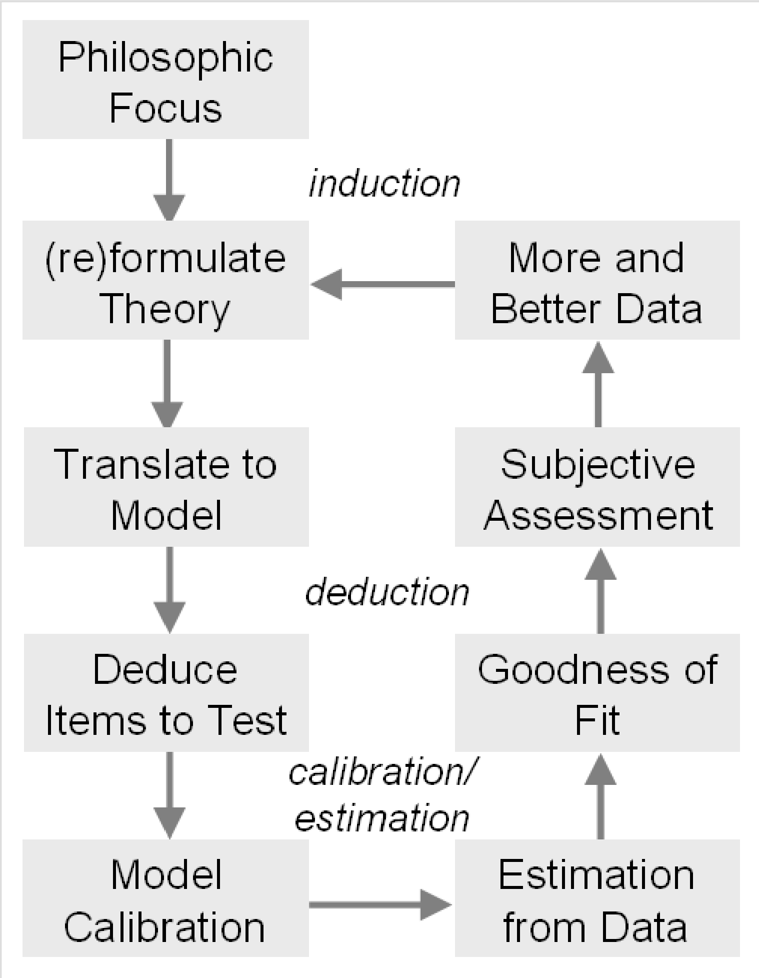
Thus the model- building process is really part and parcel of the scientific process – the scientific method where the current wisdom is that science tests theory by assembling data about reality which is designed to '*falsify*' the theory. This is scientific method *a la Popper* and it suggests that data or observations is the ultimate arbiter of good theory.

The method implies that this process of testing takes place in systems which are controllable in some science, are not volatile, as in experimental lab contexts. In fact as science has progressed, these conditions appear to be increasingly unlikely. Moreover as soon as we take powerful physical theory out of the lab, it becomes subject to volatile influences and can rarely achieve the predictive success it has in the lab

Hence the need for models – for theories in a form other than in the laboratory, where we can perform good testing.

The new form of the laboratory is the computer and instead of experimentation there is simulation. We could and perhaps we should spend time talking about this issue – for by no means all models are simulation models and all science is not based on computers. But increasingly science is intrinsically about computation and this is changing science itself. I also use the term ‘ science’ advisedly, in its most catholic sense ….. another debate perhaps later

Let me get some more terms out of the way – and to do this here is a simple picture of the scientific method.



Let me also say something about

induction

and

deduction

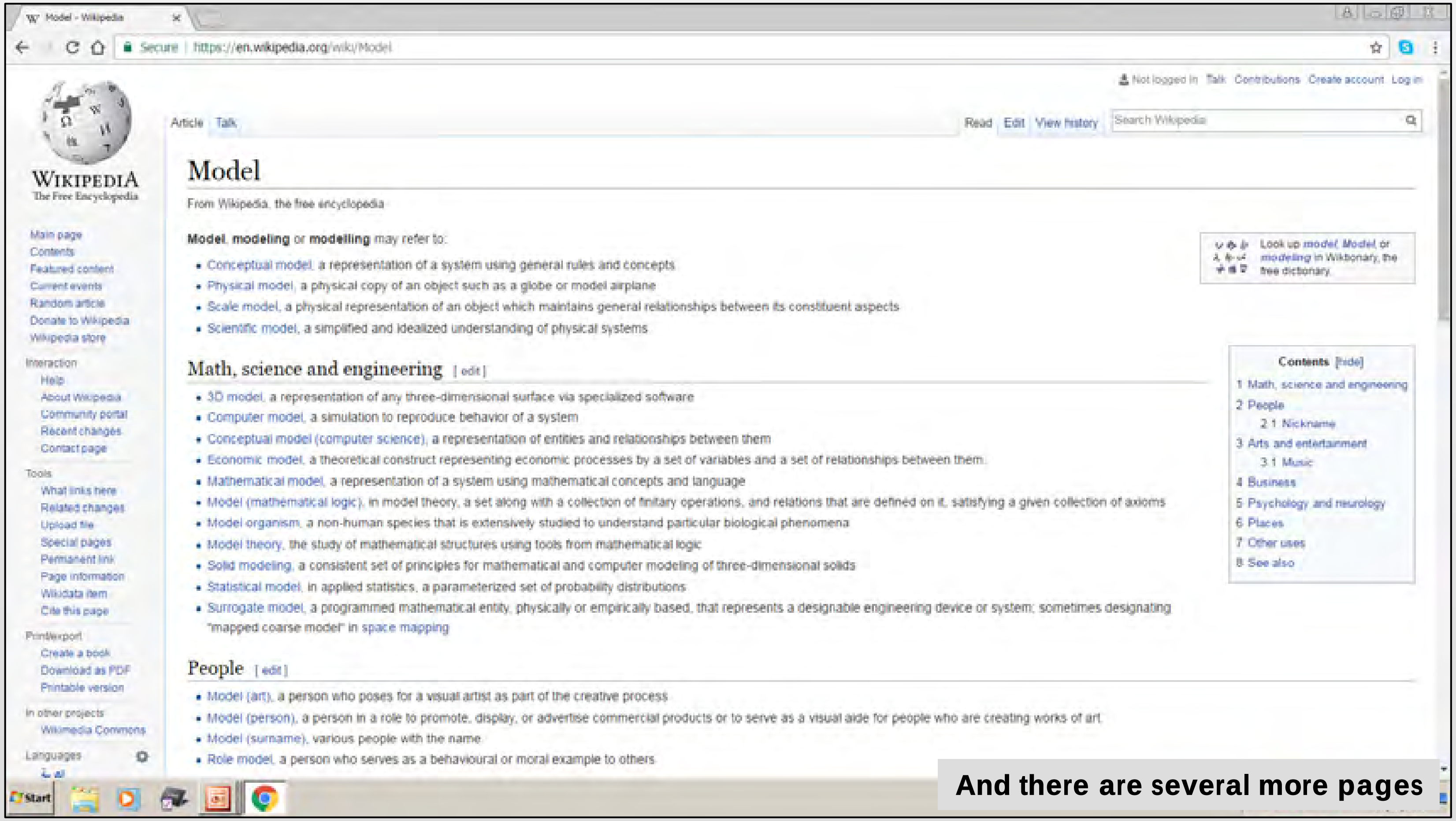
Definitions of Models

There are of course many types of models and although you may think that here we are only going to deal with mathematical or symbolic models, nothing could be further from the truth. Lowry's 1965 paper – "A Short Course in Model Design" paper that I recommend you read, classifies models, and we will draw loosely on his scheme.

Lowry, I. S. (1965) A Short Course in Model Design, **Journal of the American Institute of Planners**, **31**, 158-165.

There seem to be four different but generic ways of abstraction – **iconic**, **analog**, **symbolic** and **logic** but these categories are not mutually exclusive.

Don’t let me tell you what this range of ideas is about. Let me turn to Google and see what that says about the term ‘model’



A Classification: Icons, Analogs, Symbols

Iconic models are representations that visually convey what the real things looks like – maps are the classic example – these are largely representations – they may have some symbology but they are scaled down versions of the real thing.

Symbolic models represents systems in terms of the way they function, often through time and over space – these models are invariably mathematical.

Analog models are a half way house between iconic and symbolic. The key issue is that they take a representational and/or functional form of one system and apply it to another.

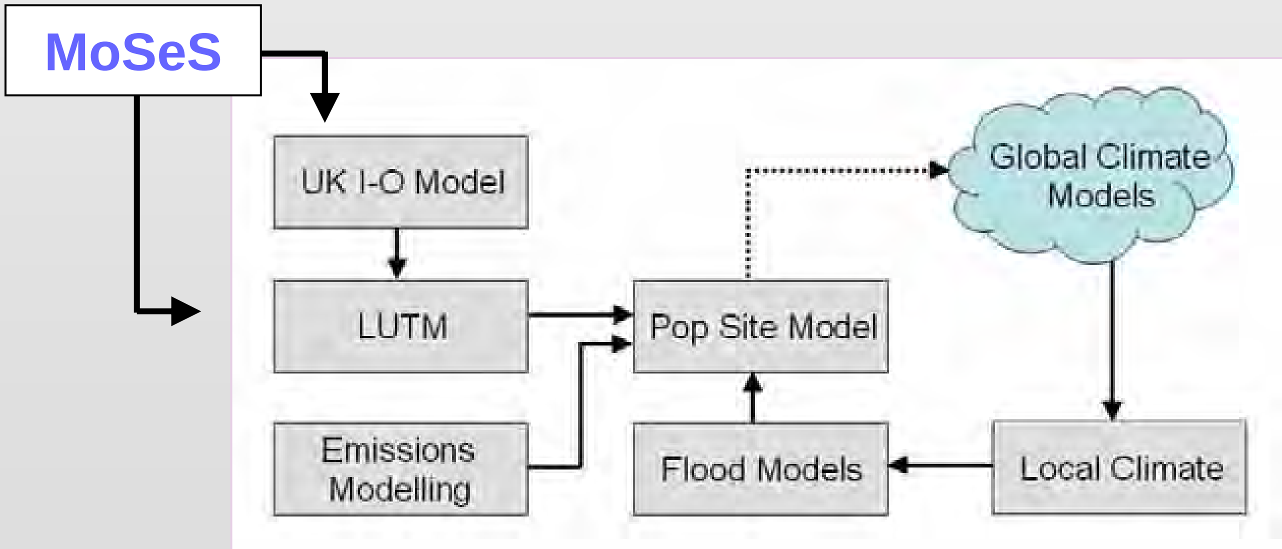
e.g. analogies between physical and human systems – the flow of blood in analogy to hydrodynamics developed for models of the atmosphere, traffic flow as an analog of an electrical network, and so on.

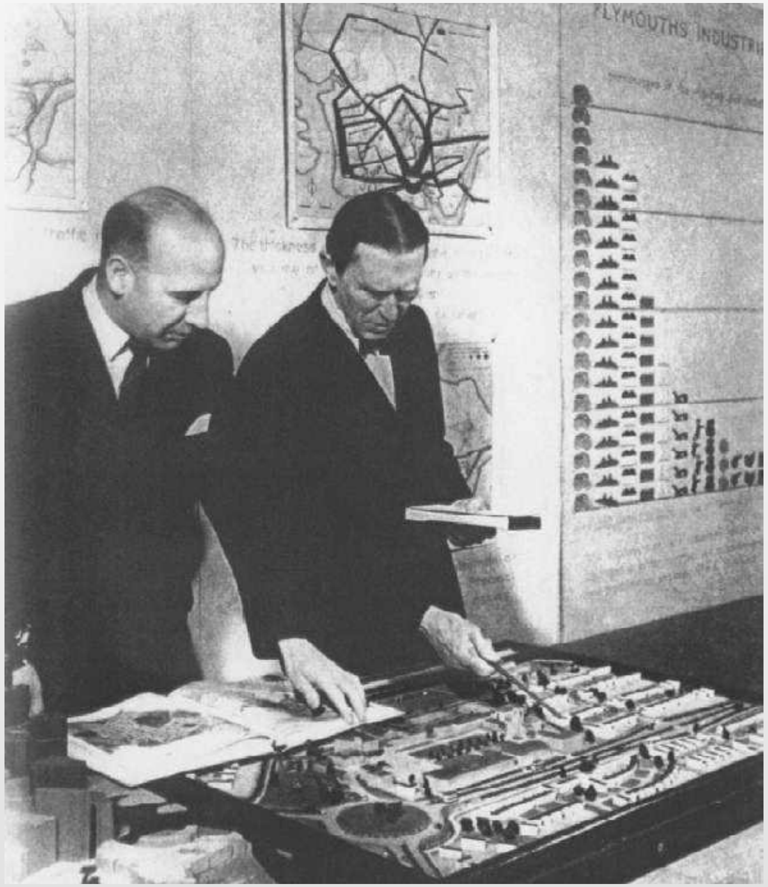
Logical models are symbolic in a sense but are based on causal connections composed of rules. We can mix, of course, any of these four types.

To this I am going to add Data-driven models. We can look at models existing on a spectrum from the data we collected about a situation that can contain elements of prediction within to fully predictive models that attempt to forecast the future or rather forecast events that have not yet happened or we have not yet observed

And I now want to introduce the idea of integrated models – different sectors and scales with different types of model – from iconic to symbolic.

Here is an example of how symbolic models are chained together – I don't want to spell out the maths and at this level the chains are like loops of spaghetti but flow diagrams are better






PIPERS


30 QEOP 3MINS
File Edit View Window Help
QUEEN ELIZABETH OLYMPIC PARK, LONDON E20 2ST, UK
1:11 PM
22°
VISUALISATIONS
FA
BU
LOD
BS
CITYGML
FOOTPRINT AREA
vilo
VIRTUAL LONDON
00:00:56

The reason why the term model has become so significant is that computers are increasingly being used as the ‘container’ or ‘media’ for many models as our world becomes digital.

Computers mean that iconic, analog, symbolic and logical models merge into one another, so for example we can have iconic models but built of mathematical structures as in GIS

And computer models are being generalised to all sorts of other things that we never used to call models – to plans, to processes of participation etc. Its is one of the most overworked words of the last half of the 20^{th} century and in this one it is widespread.

Aggregate viz Disaggregate Modelling

50 years ago when models first became identifiable as a distinct activity in science, and as the social sciences embraced them, they were usually statistical summaries or aggregations of elemental units.

Good examples were economic models based on macro economics, e.g. Keynesian models, econometric models

Population models, models based on social physics

There has always been a quest however to disaggregate – meaning that the model needs to be specified in more detail. Let me take an example – models of retail systems, called shopping models

Shopping trips = f (Population, Floorspace, Distance)

Shopping trips	Population	Floorspace	Distance
from zone i where people live to zone j where they shop	*zone i where people live*	*zone j where people shop*	*from zone i to zone j*

We might want to disaggregate the data into detailed types of population and detailed types of shopping, different transport networks and so on.

As computers have become ever faster and larger in terms of processing power, such models have become more and more disaggregate – in principle although data remains a constraint. We will show this a lot in the next two lectures

In fact as disaggregation has proceeded, models have changed in focus and a new stream of model where the fundamental elements themselves can be represented have become popular.

These are based on objects – or <u>agents</u> – where every element can be simulated – and we will say a lot more about these in later lectures.

The contrast is thus between <u>aggregates</u> and <u>agents</u>

Statics viz Dynamics

In passing, it would be remiss not to make the distinction between statics and dynamics. Models in social systems have tended to be static – comparative static or cross sectional as they are called in economics – with assumptions about that systems tend towards equilibrium.

In the last 20 years, all this has been thrown up in the air and dynamics has come onto the agenda in a big way. This has important implications for spatial systems where time has not been a popular feature of representations and models. The smart cities movement has introduced time into the agenda rather than just space.

The Paradigm Shift: Aggregates to Agents

I am not going to talk about this paradigm shift to other kinds of models but just to flag these ideas, and we will need to note models that I won't be introducing, but Sarah Wise will namely

- temporally dynamic models on fine scale spaces called cellular automata or CA models – what I introduced briefly in urban systems theory when I talked about fractals last term
- temporally dynamic models where individuals or objects move in space – agent-based models ABM or multi-agent models MAS; Sarah will spend three lectures on this

The Model-Building Process

In later discussions, we will return to the model building process and examine processes for defining a problem, theorising about the problem, formulating a model, operationalising the model, confronting the model with data, calibrating the model to the data, testing the model’s fit, taking the model elsewhere to truly test it, improving the model by extending the theory, and reiterating the process in this way.

But here we need to say something about facts and how we fit models to facts which come from data and we can distinguish the following.

Facts and Theories, Factoids & Stylized Facts: sometimes we call all this DATA

- Generally observations/data of the system being modelled or simulated are assembled and the model's predictions are compared against these 'facts'
- Facts are publicly agreed sets of observations over which there is 'no' disagreement
- Facts can range in quality from well defined observations to highly speculative pieces of data.
- Factoids and stylized facts are two types of observation that are sometimes also used in testing a model's predictive abilities

Factoids

1. A piece of unverified or sometimes inaccurate information that is presented in the press as factual, often as part of a publicity effort, and that is then accepted as true because of frequent repetition.
2. combining the word "fact" and the ending "-oid" to mean "like a fact".
3. Factoid has since developed a second meaning, that of a brief, somewhat interesting fact, that might better have been called a ‘factette’.
4. A 'factlet' is a fact that is tiny and trivial, and also correct.

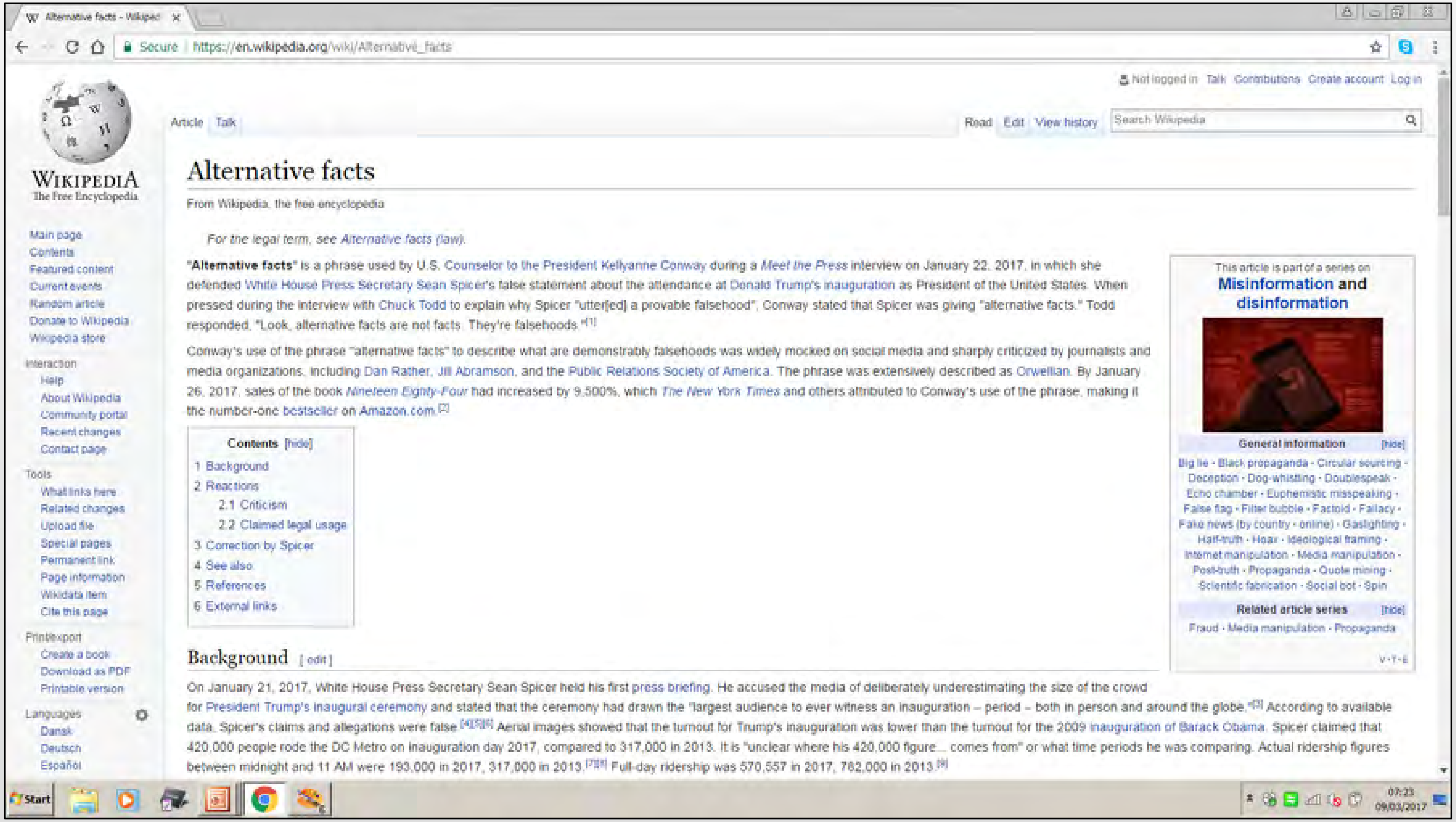
Stylized Fact

In social sciences, especially economics, a stylized fact is a simplified presentation of an empirical finding. While results in statistics can only be shown to be highly probable, in a stylized fact, they are presented as true.

A stylized fact is often a broad generalisation, which although essentially true may have inaccuracies in the detail.

Highly applicable to the assumptions of agent-based models which may not be verifiable but plausible

A New Definition Since Trump:
Alternative Facts: Fake News



Alternative facts

From Wikipedia, the free encyclopedia

For the legal term, see Alternative facts (law).

"**Alternative facts**" is a phrase used by U.S. Counselor to the President Kellyanne Conway during a *Meet the Press* interview on January 22, 2017, in which she defended White House Press Secretary Sean Spicer's false statement about the attendance at Donald Trump's inauguration as President of the United States. When pressed during the interview with Chuck Todd to explain why Spicer "utter[ed] a provable falsehood", Conway stated that Spicer was giving "alternative facts." Todd responded, "Look, alternative facts are not facts. They're falsehoods."[1]

Conway's use of the phrase "alternative facts" to describe what are demonstrably falsehoods was widely mocked on social media and sharply criticized by journalists and media organizations, including Dan Rather, Jill Abramson, and the Public Relations Society of America. The phrase was extensively described as Orwellian. By January 26, 2017, sales of the book *Nineteen Eighty-Four* had increased by 9,500%, which *The New York Times* and others attributed to Conway's use of the phrase, making it the number-one bestseller on Amazon.com.[2]

Contents [hide]

This article is part of a series on
Misinformation and disinformation

General information [hide]
Big lie · Black propaganda · Circular sourcing · Deception · Dog-whistling · Doublespeak · Echo chamber · Euphemistic misspeaking · False flag · Filter bubble · Factoid · Fallacy · Fake news (by country · online) · Gaslighting · Half-truth · Hoax · Ideological framing · Internet manipulation · Media manipulation · Post-truth · Propaganda · Quote mining · Scientific fabrication · Social bot · Spin

Related article series [hide]
Fraud · Media manipulation · Propaganda

Background [edit]

On January 21, 2017, White House Press Secretary Sean Spicer held his first press briefing. He accused the media of deliberately underestimating the size of the crowd for President Trump's inaugural ceremony and stated that the ceremony had drawn the "largest audience to ever witness an inauguration – period – both in person and around the globe."[3] According to available data, Spicer's claims and allegations were false.[4][5][6] Aerial images showed that the turnout for Trump's inauguration was lower than the turnout for the 2009 inauguration of Barack Obama. Spicer claimed that 420,000 people rode the DC Metro on inauguration day 2017, compared to 317,000 in 2013. It is "unclear where his 420,000 figure... comes from" or what time periods he was comparing. Actual ridership figures between midnight and 11 AM were 193,000 in 2017, 317,000 in 2013.[7][8] Full-day ridership was 570,557 in 2017, 782,000 in 2013.[9]

Verification (a model matches its design)

To check, confirm or prove the truth of something. To establish, prove, substantiate, attest, corroborate, support, confirm. In modelling it is usually used to see if the model is working properly

Validation (a model matches the data)

To meet some criterion/criteria associated with the model and or the data/observations. In general, validation is the process of checking if something satisfies a certain criterion. Examples would be: checking if a statement is true, if an appliance works as intended, if a computer system is secure, or if computer data is compliant with an open standard. This should not be confused with verification.

Goodness of Fit

A well defined measure of how the model's predictions match the known observations of facts, typically some measure of difference between predictions and observations.

Predictions are any outcome of the model, past, present or future

Calibration and Estimation

Calibration is the generic process of validation and verification. Estimation is the process or method of generating a precise estimate of some parameter characterising the model.

We are running out if time and that's enough really

Other Issues

consistency and reliability –with reliability is the consistency of your measurement,

I don't think there is a coherent discussion of all these issues *per se* as they are pieced together from multiple sources.

Sensitivity Testing
Process Modelling
Parsimony v richness
Scale, aggregation …. Space and time

And so on …………..

Background Reading

I will put this material up on my web tomorrow and some are on Moodle. There are five papers worth looking at

Batty, M. (2009) Urban Modeling, in R. Kitchin and N. Thrift (Eds) **International Encyclopedia of Human Geography**, Volume 12, Elsevier, Oxford, 51–58.

Batty, M. (2008) Spatial Interaction, in K. K. Kemp (Editor) **International Encyclopedia of Geographic Information Science**, Sage. Los Angeles, CA, 416-418.

Batty, M. and Torrens, P. (2005) Modelling and Prediction in a Complex World, **Futures**, **37** (7), 745-766.

Lowry, I. S. (1965) A Short Course in Model Design, **Journal of the American Institute of Planners**, **31**, 158-165.

Vanderleeuw, S. E. (2004) Why Model? **Cybernetics and Systems: An International Journal**, **35**, 117-128

And if you want some old, old background you can look at my old book **Urban Modelling** (1976). This is in the MSc room I think and if you need a copy there are some in the CASA library – if not come and ask me. You can download it anyway from my blogs

*This is the web direct web links for downloading the **Urban Modelling** book*

http://www.casa.ucl.ac.uk/urbanmodelling/

Introduction to Urban Modelling and Simulation

Topic 2: Basic Gravitation and Spatial Interaction

Michael Batty
m.batty@ucl.ac.uk
@jmichaelbatty

8 July, 2018

http://www.spatialcomplexity.info/

Outline: Basic Gravitation and Spatial Interaction

- Some Definitions
- Gravitation: The Basic Models
- Potential and Accessibility

We will continue this after lunch .....

Some Definitions

Again for I know you have had some of this in your quant methods class

Let me begin with the algebra of spatial interaction – we denote origins by the subscript i and destinations with the subscript j. Then the flow from i to j is T_{ij} . Ok now we usually arrange these flows in a matrix

$$T_{ij} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} & T_{15} & ..... & T_{1n} \\ T_{21} & T_{22} & T_{23} & T_{24} & T_{25} & ..... & T_{2n} \\ T_{31} & T_{32} & T_{33} & T_{34} & T_{35} & ..... & T_{13n} \\ . & & & & & & \\ & & & T_{ij}. & & & \\ . & & & & & & \\ T_{n1} & T_{n2} & T_{n3} & T_{n4} & T_{n5} & ..... & T_{nn} \end{bmatrix}$$

$$\sum_j T_{ij} = O_j$$

$$\sum_i T_{ij} = D_j$$

Now the usual operations are adding up over the rows or the columns as we show above using

Gravitation: The Basic Models

Let me begin with spatial interaction models once again and first define key terms. We are going to divide our spatial systems into small zones like Census Tracts which can either be called **origins** or **destinations**.

Origins are notated using the subscript i and destinations the subscript j. The original gravity model can be stated as

$$T_{ij} \propto \frac{P_i P_j}{d_{ij}^2} = K \frac{P_i P_j}{d_{ij}^2}$$

where we define $T_{ij}, P_i, P_j, d_{ij}^2, \text{and } K$ as trips, populations, distance squared and a scaling constant

In fact we can generalise the model first by noting that distance is like in the von Thunen model a measure of generalised travel cost c_{ij} and the populations are defined as measures of mass or activity as origin and destination activities O_i , D_j

Then

$$T_{ij} \propto \frac{O_i D_j}{c_{ij}^{\beta}} = K \frac{O_i D_j}{c_{ij}^{\beta}} = K O_i D_j c_{ij}^{-\beta}$$

Where β is the so-called friction of distance parameter controlling the effect of generalised travel cost. When β is large, the effect of distance is great and when it is small it is much less. This gives more trips when it is small than when it is big.

In all our models, we need to estimate these parameters and this is the process of calibration. We need to choose K and β so that the predicted trips T_{ij} are as close as possible to the observed trips T_{ij}^{obs}

We can do this in this simplest of models by fitting a linear regression to the logarithmic version of the model and when we take logs we get

$$\log \frac{T_{ij}}{O_i D_j} = \log K - \beta \log c_{ij}$$

We find the parameters by minimising the sum of the squares (squared deviations) between the predicted and observed trips, that is $\min \Phi = \min \sum_{ij} (T_{ij} - T_{ij}^{obs})^2$

Potential and Accessibility

In the 1940, the astronomer John Stewart suggested that a measure of potential could be produced from the gravity model that was an overall measure of the force of an object on all others. He defined this from the basic GM equation as potential V_i or potential per capita v_i

$$V_i = \sum_j T_{ij} \propto P_i \sum_j \frac{P_j}{d_{ij}^2}$$

$$v_i \propto \frac{V_i}{P_i} = \sum_j \frac{P_j}{d_{ij}^2}$$

This is essentially accessibility or nearness and it was first used as the basis for a simple urban model by

Walter Hanson in the late 1950s in a paper called "How Accessibility Shapes Land Use". There he said that the residential development in a place was a simple function of accessibility i.e.

$$R_i \propto \frac{V_i}{P_i} = \sum_j \frac{P_j}{d_{ij}^2}$$

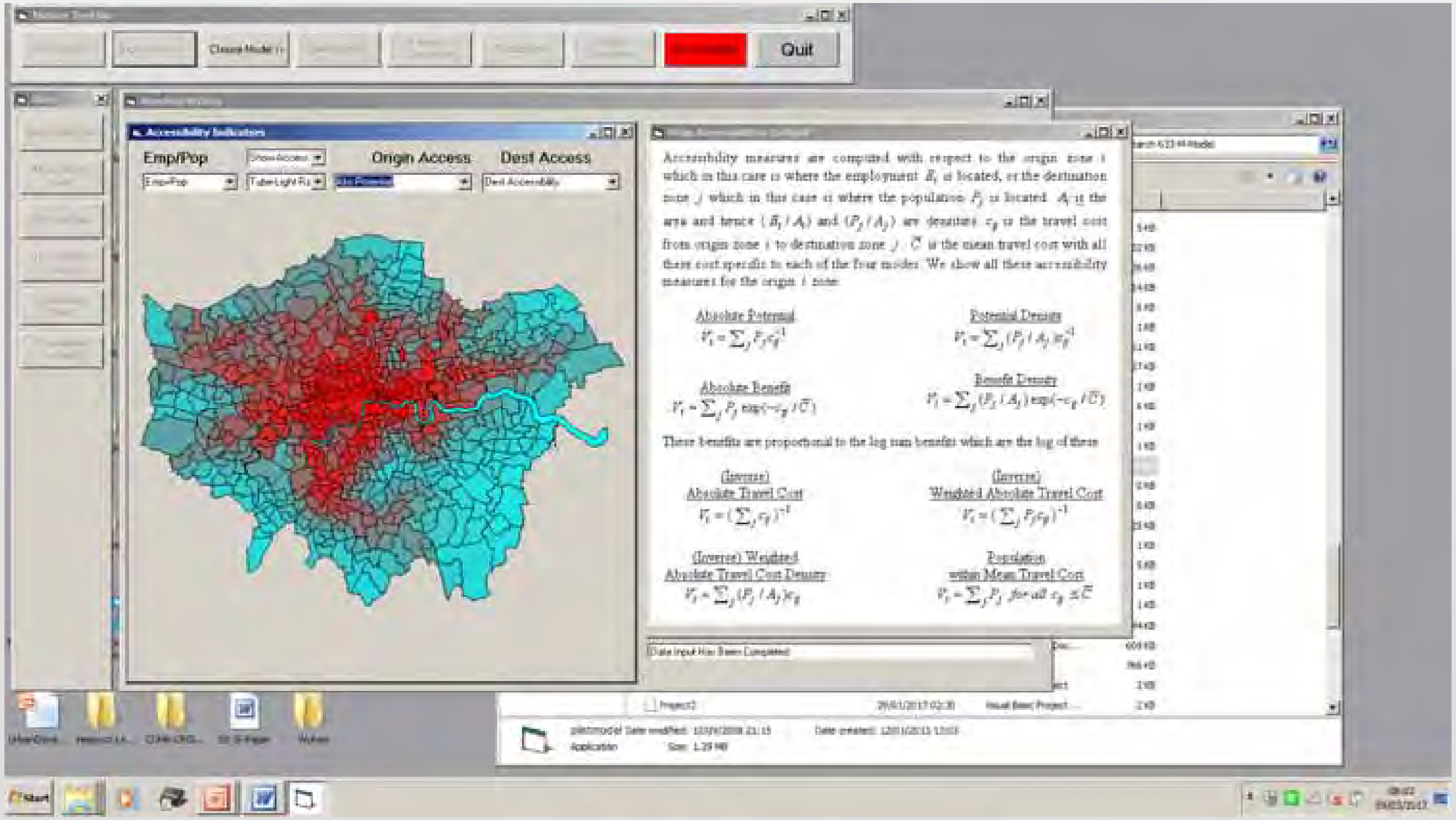
In fact if total residential development is R, then the equation can be written as

$$R_i = R \frac{(V_i / P_i)}{\sum_k (V_k / P_k)}$$

And this is our first operational land use model, the simplest

Walter G. Hansen, W. G. (1959) How Accessibility Shapes Land Use, **Journal of the American Institute of Planners**, Volume **25**, Pages 73-76 http://dx.doi.org/10.1080/01944365908978307

There are many definitions of accessibility but essentially they are all composite measures of attraction and deterrence from any one place to all others. The model I will show later has many of these



Accessibility measures are computed with respect to the origin zone i which in this case is where the employment E_i is located, or the destination zone j which in this case is where the population P_j is located. A_i is the area and hence (E_i / A_i) and (P_j / A_j) are densities. c_{ij} is the travel cost from origin zone i to destination zone j. $\overline{C}$ is the mean travel cost with all these cost specific to each of the four modes. We show all these accessibility measures for the origin i zone.

Absolute Potential

$$V_i = \sum_j P_j c_{ij}^{-1}$$

Potential Density

$$V_i = \sum_j (P_j / A_j) c_{ij}^{-1}$$

Absolute Benefit

$$V_i = \sum_j P_j \exp(-c_{ij} / \overline{C})$$

Benefit Density

$$V_i = \sum_j (P_j / A_j) \exp(-c_{ij} / \overline{C})$$

These benefits are proportional to the log sum benefits which are the log of these

(Inverse)
Absolute Travel Cost

$$V_i = (\sum_j c_{ij})^{-1}$$

(Inverse)
Weighted Absolute Travel Cost

$$V_i = (\sum_j P_j c_{ij})^{-1}$$

(Inverse) Weighted
Absolute Travel Cost Density

$$V_i = \sum_j (P_j / A_j) c_{ij}$$

Population
within Mean Travel Cost

$$V_i = \sum_j P_j \ \text{for all } c_{ij} \le \overline{C}$$

The original gravity model has been used for years but in the 1960s and 1970s various researchers cast it in a wider framework – deriving the model by setting up a series of constraints on its form which showed how it might be solved generating consistent models.

The *constraints logic* led to **consistent accounting.** The *generative logic* lead to analogies between **utility and entropy maximising** and opened a door that has not been much exploited to date between entropy, energy, urban form, physical morphology and economic structure. In particular the economic logic is called choice theory, specifically **discrete choice theory**

The key idea is to introduce constraints on the form that the model can take, and these relate to specifying what the model is able to predict. The more constraints we introduce on the model, the more we reduce the model's predictive power, but the idea of constraints also relates to what we know about the system in comparison with what we want to predict.

The idea of a framework for consistent generation of a model is that we can then handle the constraints systematically as we will now show.