LIONEL MARCH MEMORIAL CONFERENCE

UNIVERSITY OF CAMBRIDGE 20-21 March 2019

casa UCL

What Do We Need to Know Before We Begin? Lionel March, Prior Probabilities, and Information

Michael Batty

Centre for Advanced Spatial Analysis, CASA-UCL

20th March 2019

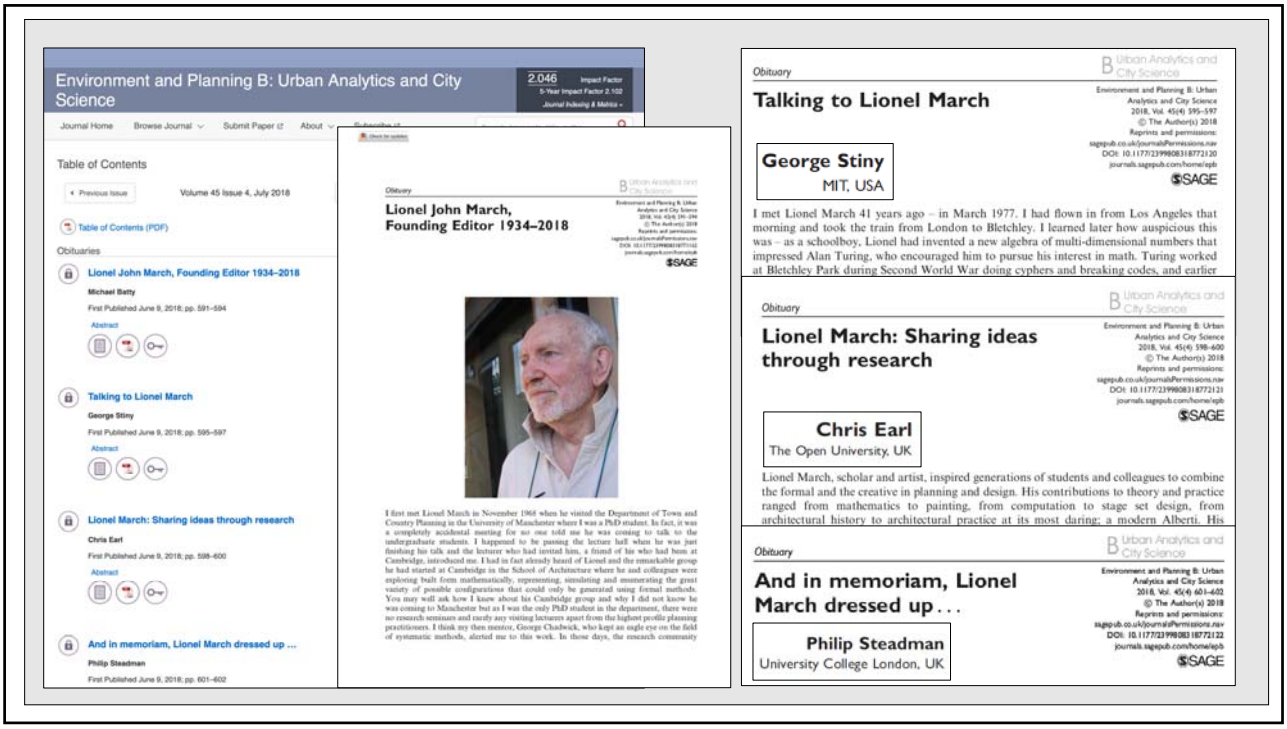
m.batty@ucl.ac.uk

@jmichaelbatty





Lionel March: 1934 – 2018



You can download the Obituaries from

http://spatialcomplexity.info/files/2018/06/Lionel-March-Obituaries-EPB.pdf

You can download this presentation as a PDF from

http://spatialcomplexity.info/

I want to talk about Lionel's contribution of course and I will do this in relation to our work together when we were both in the Faculty of Engineering at the University of Waterloo, me from 1974–1975 and Lionel from 1974–1976.

I am going to talk about our work on using Bayesian methods in thinking about spatial interaction. We published three papers on this idea and I will tell you about them. I will divide my talk into three parts

1 *the ideas, the methods;*
2 *then about what Lionel and myself did with them;*
3 *finally a suggestion that there is much unfinished business;*

where we can push these ideas into those about more generic data-driven modelling, about scientific explanation

The Ideas: Prior and Posterior Probabilities

In our world, the idea that we already have information about how we might explain any phenomenon before we set about actually explaining it, is generic.

We can encode this in the notion of *Prior Probabilities* – of some information explaining the location, say, of an activity – and then modifying this probability with some theory of how the activity locates in a place, possibly changing in this place. This is the *Posterior Probability* and new information is introduced to update the prior to the posterior.

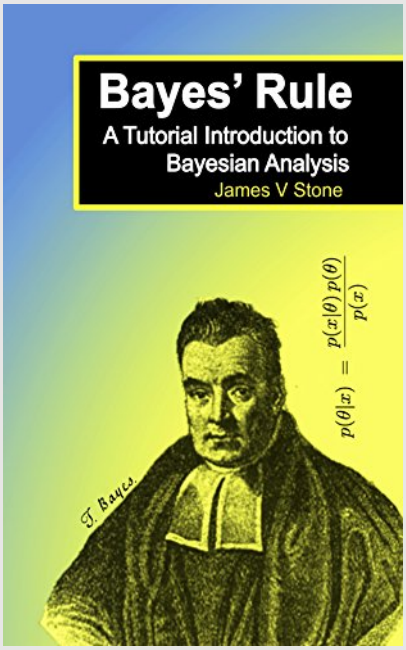
There is a well-defined theorem to represent this and it is called *Bayes theorem*, sometimes *rule* and it is an ever more popular.

Before I tell you about how Lionel and myself embarked on this problem, I will state Bayes' rule so that you have an idea of the formality of all this.

$$p(\theta \mid x) = \frac{p(x \mid \theta)}{p(x)}\, p(\theta)$$

$$\begin{pmatrix} \text{posterior} \\ \text{probability} \end{pmatrix} = \begin{pmatrix} \text{likelihood} \\ \text{ratio} \end{pmatrix} \begin{pmatrix} \text{prior} \\ \text{probability} \end{pmatrix}$$

The posterior observed event θ is and the prior observed event is x



In fact, we can derive Bayes equation quite naturally from a simple equality

$$p(\theta \mid x)\, p(x) = p(x \mid \theta)\, p(\theta)$$

θ is **this**; x is **that**;

This says: the probability of **this** given **that** occurring, times the probability of **that** occurring is equal to the probability of **that** given **this** occurring, times the probability of **this** occurring.

This is both intuitively obvious and confusing as much of probability theory is.

Let us now reformulate this simple relation in the algebra and notation that we understand here in spatial modelling

We can write the posterior as equal to the product of the likelihood ratio times the prior probability as

$$P_i = \Lambda_i Q_i$$

where the probabilities are normalised as $\sum_i P_i = 1 \quad \sum_i Q_i = 1$

Now we can write the equation where we identify the normalisation directly as

$$P_i = Z L_i Q_i$$

Where

$$Z = \frac{1}{\sum_i L_i Q_i} \quad ; \quad P_i = \frac{L_i Q_i}{\sum_i L_i Q_i}$$

Different Kinds of Prior Probabilities

Essentially the prior probability is turned into a posterior by adding new information which is essentially the model - - we can think of this as what we have already Q_i times the new information which is what we hypothesise.

Let us look at what we have already in terms of a cities example

- The simplest Q_i is say something like the capacity or area of land or the number of houses say in residential location – we need these before we can occupy them with residents
- We could also assume a null hypothesis

$$Q_i = 1/n \qquad and \qquad \sum_{i=1}^{n} Q_i = 1$$

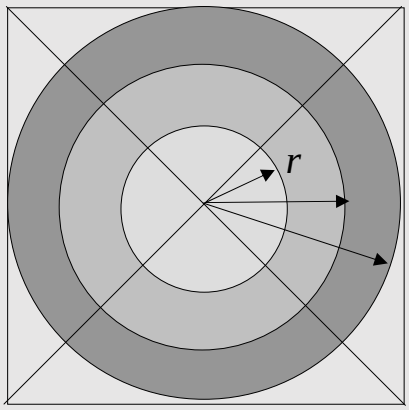
And this means that the prior has no effect

- We might also think of the prior as some sort of geometric default – in other words due to the effect of space. Let us now assume that if an individual travels from an origin to ever further destinations, the number of possible destinations increases in proportion to the distance travelled and this the probability of ending at one of those destinations declines inversely with distance r_i

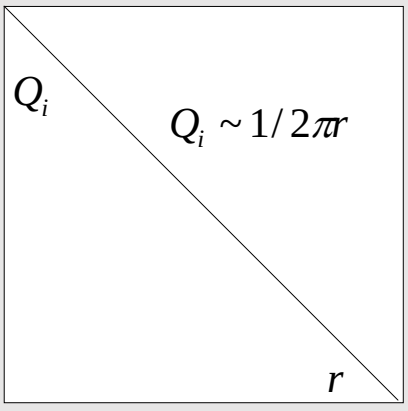
  We thus have to factor out this effect – in other words it already exists and this we might define the prior as

$$Q_i \approx 1/r_i = r_i^{-1} / \sum_i r_i^{-1}$$

We can articulate this effect in the following picture



$$A = 2\pi r^2$$
$$dA/dr = 2\pi r$$



This is related to James S Coleman's *Method of Residues* which I will recount below as one of our applications

Now here is a more controversial idea – the model is the data

- We might think of the prior as the data and thus the model becomes updating the data into the 'data' – in short the best model we have is the data. This sounds crazy but it helps us to think of the model as an update of what we know and we know the data. We can thus write the model as

$$Q_i = \frac{Q_i}{\sum_i Q_i} = P_i = \frac{L_i Q_i}{\sum_i L_i Q_i}$$

- In fact if we update the data by changing some aspect of it, we do what Fratar did in 1954 – and also what various people have done with respect to using biproportional factoring to update I-O and trip matrices

- We can now introduce our last idea of the prior – as an updating of the data through time – or using the model as the data to produce another model – through real time or even through 'model time'. We write the posterior-prior relation as

$$P_i(t+1) = Z(t)\,L_i(t)\,P_i(t) = \Lambda_i(t)P_i(t)$$

where the prior is now the model prediction at the previous time interval; and where the initial model is the prior data

$$P_i(t) = Q_i$$

Then by recursion we can express the posterior as a function of the original prior and the sequence of likelihoods

$$P_i(t+n) = \Lambda_i(t+n-1)P_i(t+n-1)$$

We can now simplify this as

$$P_i(t+n) = \Lambda_i(t+n-1) \ldots \Lambda_i(t+1)\Lambda_i(t)P_i(t) = P_i(t)\prod_{\tau=1}^{t}\Lambda_i(t)$$

Lionel in his inimitable way called this a *Bayes Chain.* I don't think anyone else has called it this but he did; and I think we only referred to it only once as this

- There is one last thing before I tell you what we did together. In fact what we did involves what I have just been talking about but the big unifying idea involves using information theory to derive these models.

  In short if you define information as Kullback & Leibler defined it in 1951, not long after Shannon, the idea is that information is the difference between the prior and the posterior i.e.

Now information is thus defined as

$$I = \sum_i P_i \log \frac{P_i}{Q_i}$$

And we can minimise this subject to all the information that we think is relevant to the model; but first let us look at the formula when $Q_i = 1/n$

$$I = \sum_i P_i \log \frac{P_i}{1/n} = \log n + \sum_i P_i \log P_i = \log n - H$$

Now let us minimise this and we form the Lagrangian

$$\min L = \sum_i P_i \log \frac{P_i}{Q_i} - \mu(\sum_i P_i - 1) - \lambda(\sum_i P_i\, r_i - R)$$

We get the standard model but it has a prior probability now

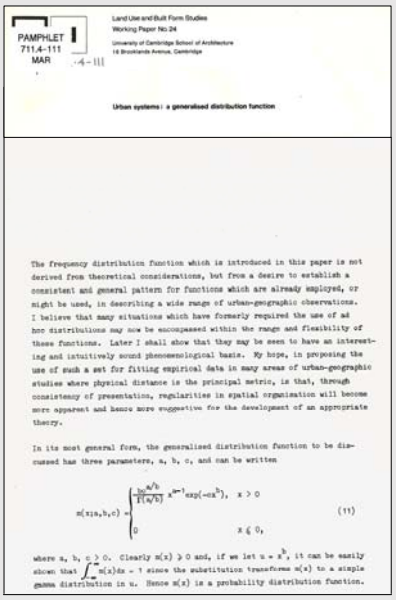
$$P_i = \exp(-\mu)\exp(-\lambda r_i)Q_i$$

and we can collect all the terms and simplify the model as

$$P_i = \frac{r_i^{-1}\exp(-\lambda r_i)}{\sum_i r_i^{-1}\exp(-\lambda r_i)} = \frac{Q_i\exp(-\lambda r_i)}{\sum_i Q_i\exp(-\lambda r_i)}$$

Now we have shown that the model uses in this case a constraint on distance travelled plus a prior based on inverse distance to – to produce the model – this in fact is like a gamma distribution.

And who looked at this when all of us began work in the area in 1967-8 – but Lionel March


PAMPHLET
711.4-111
MAR

Land Use and Built Form Studies
Working Paper No.24
University of Cambridge School of Architecture
18 Brooklands Avenue, Cambridge

Urban systems: a generalised distribution function

The frequency distribution function which is introduced in this paper is not derived from theoretical considerations, but from a desire to establish a consistent and general pattern for functions which are already employed, or might be used, in describing a wide range of urban-geographic observations. I believe that many situations which have formerly required the use of ad hoc distributions may now be encompassed within the range and flexibility of these functions. Later I shall show that they may be seen to have an interesting and intuitively sound phenomenological basis. My hope, in proposing the use of such a set for fitting empirical data in many areas of urban-geographic studies where physical distance is the principal metric, is that, through consistency of presentation, regularities in spatial organisation will become more apparent and hence more suggestive for the development of an appropriate theory.

In its most general form, the generalised distribution function to be discussed has three parameters, a, b, c, and can be written

$$m(x;a,b,c) = \begin{cases} \frac{bc^{a/b}}{\Gamma(a/b)} x^{a-1}\exp(-cx^b), & x > 0 \\ 0 & x \leqslant 0, \end{cases} \qquad (11)$$

where a, b, c > 0. Clearly m(x) > 0 and, if we let $u = x^b$, it can be easily shown that $\int_{-\infty}^{\infty} m(x)dx = 1$ since the substitution transforms m(x) to a simple gamma distribution in u. Hence m(x) is a probability distribution function.

Lionel March

1969

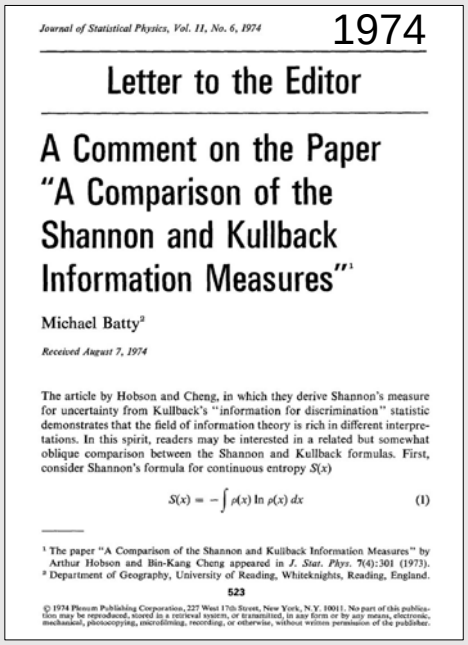
This is Working Paper 24 of the centre for Land Use and Built Form Studies in the University of Cambridge, prepared for presentation at the meeting of the British Section of the Regional Science Association, London, 21–23 August 1969.

August 1969
Second edition July 1971

$$m(x;a,b,c) = \begin{cases} \frac{bc^{a/b}}{\Gamma(a/b)} x^{a-1}\exp(-cx^b), & x > 0 \\ 0 & x \leqslant 0, \end{cases}$$

So We Began Our Joint Work in 1974: My Second Theme

Now when we joined each other at Waterloo, I had been working with ideas about information and I had worked on a measure called *Spatial Entropy* which can be seen as an information difference. I won't go into this here but in 1973 another paper by Hobson and Cheng appeared about information differences. Here is a letter I wrote to the journal back in 1973/4 before I went to Waterloo



Journal of Statistical Physics, Vol. 11, No. 6, 1974

1974

Letter to the Editor

A Comment on the Paper "A Comparison of the Shannon and Kullback Information Measures"[1]

Michael Batty[2]

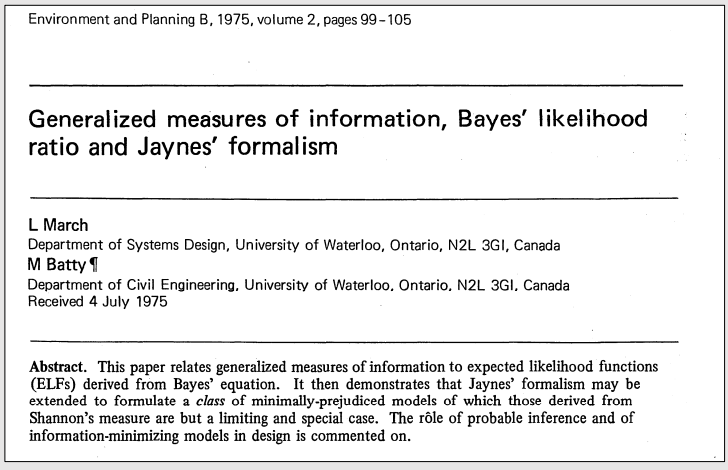
Received August 7, 1974

The article by Hobson and Cheng, in which they derive Shannon's measure for uncertainty from Kullback's "information for discrimination" statistic demonstrates that the field of information theory is rich in different interpretations. In this spirit, readers may be interested in a related but somewhat oblique comparison between the Shannon and Kullback formulas. First, consider Shannon's formula for continuous entropy $S(x)$

$$S(x) = -\int \rho(x) \ln \rho(x)\, dx \qquad (1)$$

[1] The paper "A Comparison of the Shannon and Kullback Information Measures" by Arthur Hobson and Bin-Kang Cheng appeared in *J. Stat. Phys.* 7(4):301 (1973).
[2] Department of Geography, University of Reading, Whiteknights, Reading, England.

© 1974 Plenum Publishing Corporation, 227 West 17th Street, New York, N.Y. 10011. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission of the publisher.



Environment and Planning B, 1975, volume 2, pages 99 – 105

Generalized measures of information, Bayes' likelihood ratio and Jaynes' formalism

L March
Department of Systems Design, University of Waterloo, Ontario, N2L 3GI, Canada
M Batty ¶
Department of Civil Engineering, University of Waterloo, Ontario, N2L 3GI, Canada
Received 4 July 1975

Abstract. This paper relates generalized measures of information to expected likelihood functions (ELFs) derived from Bayes' equation. It then demonstrates that Jaynes' formalism may be extended to formulate a *class* of minimally-prejudiced models of which those derived from Shannon's measure are but a limiting and special case. The rôle of probable inference and of information-minimizing models in design is commented on.

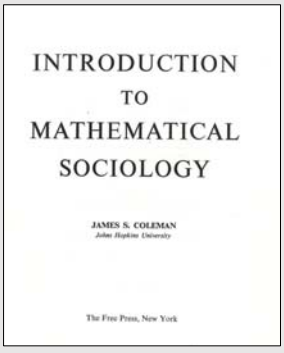
- Our first paper articulated the idea of Bayes theorem and information minimising, drawing on Aczel's ideas – Aczel was at Waterloo too in the same department as Bill Tutte: *Combinatorics and Optimisation*

Lionel knew all about this somehow: the paper by Jaynes on prior probabilities was important then and also Lionel went to see Myron Tribus who was at Xerox In Rochester NY, not so far away.

We then decided to look at what a good prior was and Lionel introduced me to James S. Coleman's method of residues – based on the idea that you filter out what is obvious and then what is left – the residues are to be explained.

In Coleman's book (1964), he shows this for the notion of geometric space that gets exponentially larger as you travel away from an origin, hence making the probability of travelling to those further spaces intrinsically decreasing.

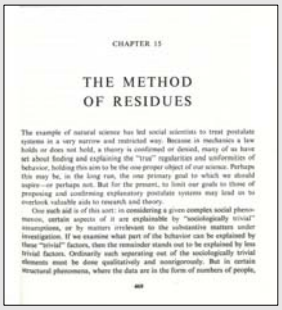
We both knew about the book but it was Lionel who knew about the method. I then applied it to a gravity model formulation in the Toronto region. I did buy the book in fact when I returned to UK in 1975 and it still floats around CASA



INTRODUCTION
TO
MATHEMATICAL
SOCIOLOGY

JAMES S. COLEMAN
John Hopkins University

The Free Press, New York



CHAPTER 15

THE METHOD
OF RESIDUES

The example of natural science has led social scientists to treat postulate systems in a very narrow and restricted way. Because in mechanics a law holds or does not hold, a theory is confirmed or denied, many of us have set about finding and explaining the "true" regularities and uniformities of behavior, holding this aim to be the one proper object of our science. Perhaps this may be, in the long run, the one primary goal to which we should aspire—or perhaps not. But for the present, to limit our goals to those of proposing and confirming explanatory postulate systems may lead us to overlook valuable aids to research and theory.

One such aid is of this sort: in considering a given complex social phenomenon, certain aspects of it are explainable by "sociologically trivial" assumptions, or by matters irrelevant to the substantive matters under investigation. If we examine what part of the behavior can be explained by these "trivial" factors, then the remainder stands out to be explained by less trivial factors. Ordinarily such separating out of the sociologically trivial elements must be done qualitatively and nonrigorously. But in certain structural phenomena, where the data are in the form of numbers of people,

Environment and Planning A, 1976, volume 8, pages 189 – 214

The method of residues in urban modelling

M Batty
Department of Geography, University of Reading, Whiteknights, Reading, RG6 2AF, Berkshire, England
L March
Department of Design Technology, The Open University, Walton Hall, Milton Keynes, MK7 6AA, England
Received 11 July 1975

Abstract. This paper seeks to extend the macrostatic approach to urban modelling by treating modelling problems as many-stage processes. Within such a process the early stages are concerned with explaining the relatively trivial characteristics of the phenomena of interest, and the later stages are devoted to explaining more important behavioural issues. Coleman (1964) calls this approach the 'method of residues', and its power is first demonstrated here by a reinterpretation of the well-known gravity model. An ad hoc test of the method on the Toronto-centred region serves to emphasise the need for a more formal approach, and thus an analogy between the method and the Bayesian viewpoint is introduced. A method of information minimising, more general but consistently and unambiguously related to the method of entropy maximising, is used to make the formal approach operational, and the method is used to generate an 'extended' family of spatial-interaction models. A number of spatial-interaction models are derived, and the paper is concluded by a test of two of these models on the Toronto-centred region.

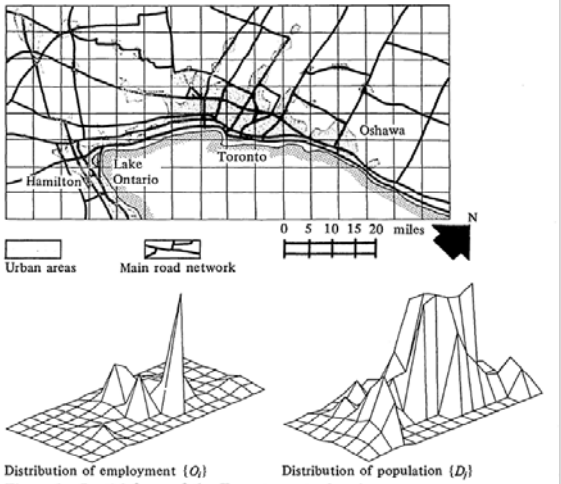


Figure 1. Spatial form of the Toronto-centred region.

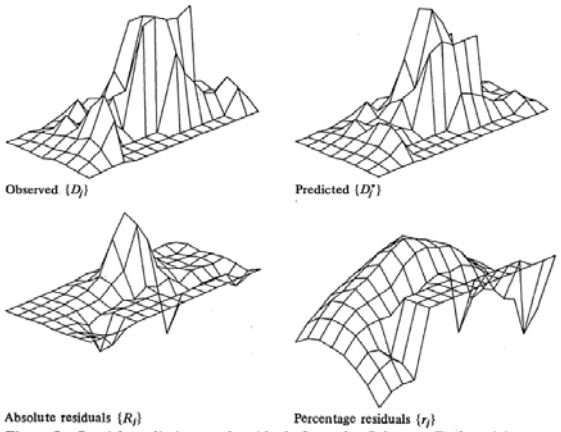


Figure 3. Spatial predictions and residuals from the Coleman–Zipf model.

Waterloo was a truly remarkable place – they gave you money for doing research and engineering gave us money to run a two day colloquium on all this in July 1975



COLLOQUIUM ON SOCIO-ECONOMIC SYSTEMS DESIGN IN URBAN AND REGIONAL PLANNING

A RESEARCH SEMINAR HELD ON JULY 28 - 29, 1975, IN THE UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA.

ORGANISERS

Michael Batty, Department of Civil Engineering, University of Waterloo, and Department of Geography, University of Reading.

Lionel March, Department of Systems Design, University of Waterloo, and Department of Architecture, University of Cambridge.



University of Waterloo

FINAL PROGRAMME

COLLOQUIUM ON SOCIO-ECONOMIC SYSTEMS DESIGN IN URBAN AND REGIONAL PLANNING: 28-29 JULY, 1975

in

ARTS LECTURE HALL, ROOM 105

MONDAY, 28th JULY, 1975:

9-30 a.m. Coffee

10-00 a.m. Lionel March, Department of Systems Design, University of Waterloo, and Department of Architecture, University of Cambridge: Introduction to the Seminar

10-30 a.m. Marcial Echenique, Director, the Martin Centre for Urban and Architectural Studies, University of Cambridge: Experience with Urban Models in Britain and South America

11-30 a.m. Bruce Hutchinson, Department of Civil Engineering, University of Waterloo, and A.C. Sarna, Central Road Research Laboratory, Delhi, India: A Land-Use Transport Model for Delhi, India

12-30 p.m. - 2-00 p.m. LUNCH

2-00 p.m. DISCUSSION GROUPS

(i) (i) in Room 105: Land Use Models in Developing Countries: Discussion to be lead by Bruce Hutchinson

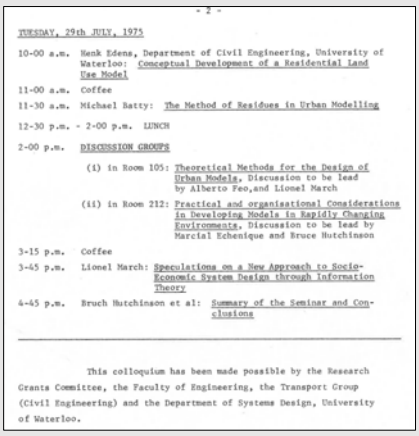
(ii) in Room 212: The Role of Model-Building in the Planning Process: Discussion to be lead by Michael Batty

(iii) in Room 213: The Theory of Design and Model-Building in Social and Economic Systems: Discussions to be lead by Lionel March.

3-15 p.m. Coffee

3-45 p.m. Alberto Feo, Instituto de Urbanismo, Faculty of Architecture, Central University of Venezuela, Caracas: The CARACAS Models: Problems, Performance and Planning Applications

4-45 p.m. Michael Batty, Department of Civil Engineering, University of Waterloo, and Department of Geography, University of Reading: Summary of the First Day's Proceedings



- 2 -

TUESDAY, 29th JULY, 1975

10-00 a.m. Henk Edens, Department of Civil Engineering, University of Waterloo: Conceptual Development of a Residential Land Use Model

11-00 a.m. Coffee

11-30 a.m. Michael Batty: The Method of Residues in Urban Modelling

12-30 p.m. - 2-00 p.m. LUNCH

2-00 p.m. DISCUSSION GROUPS

(i) in Room 105: Theoretical Methods for the Design of Urban Models, Discussion to be lead by Alberto Feo,and Lionel March

(ii) in Room 212: Practical and organisational Considerations in Developing Models in Rapidly Changing Environments, Discussion to be lead by Marcial Echenique and Bruce Hutchinson

3-15 p.m. Coffee

3-45 p.m. Lionel March: Speculations on a New Approach to Socio-Economic System Design through Information Theory

4-45 p.m. Bruch Hutchinson et al: Summary of the Seminar and Conclusions

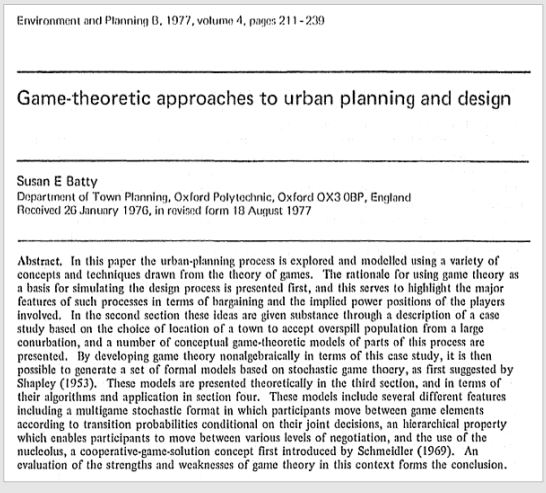
---

This colloquium has been made possible by the Research Grants Committee, the Faculty of Engineering, the Transport Group (Civil Engineering) and the Department of Systems Design, University of Waterloo.

I returned to the UK – to the University of Reading where I was a lecturer in Geography – and my wife finished her MASc degree in Systems Design – another pretty unique coincidence as she had applied to do this not knowing that Lionel was going to Waterloo as we were at that time. This was pretty amazing in any case – going to Waterloo at the same time as Lionel – which I only found out about at the Land Use Models conference here in Cambridge in July 1974 and we were destined to fly in early August.

My wife in fact wrote a Master's thesis on game-theory in design and Lionel was her advisor. All pretty incestuous and she published a paper on this in EPB in 1977

Anyway we didn't worry about things like that in those days



Environment and Planning B, 1977, volume 4, pages 211 - 239

Game-theoretic approaches to urban planning and design

Susan E Batty
Department of Town Planning, Oxford Polytechnic, Oxford OX3 0BP, England
Received 26 January 1976, in revised form 18 August 1977

Abstract. In this paper the urban-planning process is explored and modelled using a variety of concepts and techniques drawn from the theory of games. The rationale for using game theory as a basis for simulating the design process is presented first, and this serves to highlight the major features of such processes in terms of bargaining and the implied power positions of the players involved. In the second section these ideas are given substance through a description of a case study based on the choice of location of a town to accept overspill population from a large conurbation, and a number of conceptual game-theoretic models of parts of this process are presented. By developing game theory nonalgebraically in terms of this case study, it is then possible to generate a set of formal models based on stochastic game theory, as first suggested by Shapley (1953). These models are presented theoretically in the third section, and in terms of their algorithms and application in section four. These models include several different features including a multigame stochastic format in which participants move between game elements according to transition probabilities conditional on their joint decisions, an hierarchical property which enables participants to move between various levels of negotiation, and the use of the nucleolus, a cooperative-game-solution concept first introduced by Schmeidler (1969). An evaluation of the strengths and weaknesses of game theory in this context forms the conclusion.

My last piece of work with Lionel was after I returned. We applied information-minimising to a little model of the Reading region. I remember presenting this in Foster Court where Geography was located in UCL in late 1975, maybe 1976.

And a book came out with our paper with a somewhat amazing array of authors – let me show you



Contents

1 Introduction 1
R L Martin, N J Thrift, R J Bennett

Part I: Problems of method and measurement

2 Dynamics of patterns on a hard-data backcloth 19
R H Atkin

3 An introduction to the theory of fuzzy sets in the context of the construction of representational, spatial economic theory 37
W D Macmillan

4 Some implications of the use of spatially grouped data 53
I N Williams

7 Space–time processes and spatial interaction models
R P Haining

8 Kalman filter modelling of time-varying processes in urban and regional analysis
R L Martin

9 Dynamic urban models based on information-minimising
M Batty, L March

Part III: Applications

10 Adaptive parameter space–time models: an entropy-maximising application using the Kalman filter with unknown prior parameters
R J Bennett

13 Spatial forecasting: a view from the touchline 205
P Haggett

And so to my third and last theme

I am very attracted to the idea of data-driven models and I think that we can go a very long way with this – all our data in fact is lagged in the past and I believe we could well make a lot of progress in using that data directly as it contains all the information which is encoded about how the urban system works – in other words we make our models ever better by continually updating them – priors into posteriors and so on and we start with the data.

I had a go at this about 3 years ago and never completed it. I said I would have another go and present a paper last November at the NARSA meeting in San Antonio, TX, but I had bad flu and couldn't go.

I Intend to have a go in the next couple of weeks and build some of these data driven models and present the ideas at a meeting on 'Urban Analytics' in Newcastle.

The gist of the idea is that if we have some data about a place and time, by the time we build a model of this data it will have changed, so our model might be an updating – a prediction of the data.

So if we have data about spatial interaction in a matrix $T_{ij}(obs)$

we need to update this to $T_{ij}(pred)$ as

$$T_{ij}(pred) = f_{ij}[T_{ij}(obs)]$$

This is a version of the prior to posterior model where prior is data and posterior is a prediction – data observed after the prior

There are many things we can say about this and these represent ideas to test. I will list them to finish

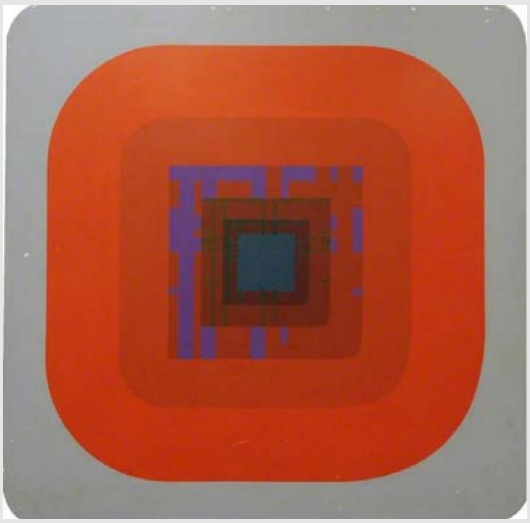
- The prior is data at time t, the posterior data at time t+1
- The functional transform is a method of residues-like model
- There may be constraints of various kinds on the form of the posterior that need to be reflected in the functional form
- There may be predictions for some of the posterior that we need to account for
- There may be missing data in the prior and the posterior that require modelling synthetically
- Thus the mix of prior and posterior can be messy

- This leads to what I consider an interesting idea.
- Let us replace all our urban models of this genus which we have built in the last 50 years with the data, embrace the implicit dynamics of data transformation
- In this way we might begin to see how good our predictions might be

Let me end with a nice picture not *of* but *from*

Lionel

I hope it reflects his interest in priors and posteriors



Quartet(1) The University of Hull Art Collection, 1968