Lectures on Urban Modelling January 2017

Gravitation and Spatial Interaction

Michael Batty

m.batty@ucl.ac.uk

@jmichaelbatty

16 January, 2017

http://www.spatialcomplexity.info/

Outline

- Gravitation: The Basic Models
- Potential and Accessibility
- Spatial Interaction and Trip Distribution: Constraints on Volume & Location
- Derivation Methods: Entropy-Maximising
- Residential Location, Modal Split
- The London Tyndall Model: Applications
- Transportation Modelling: The Four Stage Process
- Next Time: Modular Modelling: Coupled Spatial Interaction

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. Now the original gravity model can be stated as

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

where we define $T_{ij}, P_i, P_j, d_{ij}^2$, 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

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

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 traipse 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

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

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

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.

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

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

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.

Trip Distribution: Constraints on Volume & Location

We must move quite quickly now so let me introduce the basic constraints on spatial interaction and then state various models

The constraints are usually specified as origin constraints and destination constraints as

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

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

And we can take our basic gravity model and make it subject to either or both of these constraints or not at all

So what we get are four possible models

<u>Unconstrained</u> $T_{ij} = KO_iD_jc_{ij}^{-\beta}$

<u>Singly (Origin) Constrained</u> $T_{ij} = A_iO_iD_jc_{ij}^{-\beta}$

so that the volume of trips at the origins is conserved

<u>Singly (Destination) Constrained</u> $T_{ij} = B_jO_iD_jc_{ij}^{-\beta}$

so that the volume of trips at the destinations is conserved

<u>Doubly Constrained</u> $T_{ij} = A_iB_jO_iD_jc_{ij}^{-\beta}$

trip volumes at origins + destinations are conserved

The first three are location models, the last is the transportation model

Now the simplest way to work out what the constants mean is to note the constraints equations and then add and factor the model subject to the constraints. Let us begin with the simplest gravity model which is

$$T_{ij} = KO_i D_j c_{ij}^{-\beta}$$

Then as K pertains to the whole system if we add this model up over i and j we can factor our K as

$$\sum_{ij} T_{ij} = K \sum_{ij} O_i D_j c_{ij}^{-\beta} = T \quad \text{and then}$$

$$K = T \Big/ \sum_{ij} O_i D_j c_{ij}^{-\beta} \quad \text{and the model becomes}$$

$$T_{ij} = T \frac{O_i D_j c_{ij}^{-\beta}}{\sum_{ij} O_i D_j c_{ij}^{-\beta}}$$

Now the singly constrained – origin and then destination and the doubly constrained models follow directly and we will simply state there full forms noting that we need to find

$$(1) \quad T_{ij} = A_i O_i D_j c_{ij}^{-\beta} = O_i \frac{D_j c_{ij}^{-\beta}}{\sum_j D_j c_{ij}^{-\beta}} \quad \text{origin - constrained}$$

$$(2) \quad T_{ij} = B_j O_i D_j c_{ij}^{-\beta} = D_j \frac{O_i c_{ij}^{-\beta}}{\sum_j O_i c_{ij}^{-\beta}} \quad \text{destination constrained}$$

$$T_{ij} = A_i B_j O_i D_j c_{ij}^{-\beta}$$

$$(3) \quad A_i = 1/\sum_j B_j D_j c_{ij}^{-\beta} \quad \text{origin - destination constrained}$$

$$B_j = 1/\sum_i A_i O_j c_{ij}^{-\beta}$$

Entropy-Maximising and Related Measures

Now we have only dealt with constraints through consistent accounting – we now need to deal with generative methods that lead to the same sort of accounting– entropy maximising, information minimising, utility maximising and random utility maximising, and also various forms of nonlinear optimisation – in fact all these methods may be seen as a kind of optimisation of an objective function – entropy utility and so on – subject to constraints.

In essence what we do is define a function to optimise which is some measure of the variation in the model and we then optimise it using calculus subject to some basic constraints of the kind that we have been using

First we define entropy as Shannon information and we convert all our equations and constraints to probabilities.

$$p_{ij} = \frac{T_{ij}}{T} = \frac{T_{ij}}{\sum_{ij} T_{ij}}$$

Shannon entropy is a measure of spread or compactness in spatial systems

$$H = -\sum_i \sum_j p_{ij} \log p_{ij}$$

We maximise this entropy subject to origin and destination constraints or some combination of these but noting now that we need another constraint on travel cost which is equivalent to energy so that we can derive a model

$$\sum_i \sum_j p_{ij} c_{ij} = \hat{C}$$

We thus set up the problem as

$$\max \quad H = -\sum_i \sum_j p_{ij} \log p_{ij}$$

subject to

$$\sum_j p_{ij} = p_i$$

$$\sum_i p_{ij} = p_j$$

$$\sum_i \sum_j p_{ij} c_{ij} = \hat{C}$$

But note that the probabilities always add to 1, that is

$$\sum_i \sum_j p_{ij} = \sum_i p_i = \sum_j p_j = 1$$

From this we get the Boltzmann-Gibbs distribution for the probabilities

By setting up a Lagrangian which is the method of maximisation, then we get

$$p_{ij} = \exp(-\lambda_i - \lambda_j - \lambda c_{ij})$$

or

$$T_{ij} = Tp_{ij} = \overbrace{A_i O_i}\,\overbrace{B_j D_j}\,\overbrace{\exp(-\lambda c_{ij})}$$

Now we can generate any model in the family of four models – unconstrained, singly-constrained (origin or destination) and doubly constrained by setting the redundant constraint parameters equal to zero and simplifying the model

To derive a residential location model which is origin constrained – we know the information at the origin but want to predict the flows to the destination and add up these flows to predict activity at the destination, we

We thus set up the problem as

$$\begin{cases} \max \quad H = -\sum_i \sum_j p_{ij} \log p_{ij} \\ \textit{subject to} \\ \sum_j p_{ij} = p_i \\ \cancel{\sum_i p_{ij} = p_j} \\ \sum_i \sum_j p_{ij} c_{ij} = \hat{C} \end{cases}$$

And we get

$$\begin{cases} p_{ij} = \exp(-\lambda_i - \lambda c_{ij}) \\ \textit{or} \\ T_{ij} = Tp_{ij} = A_i O_i \exp(-\lambda c_{ij}) = O_i \dfrac{\exp(-\lambda c_{ij})}{\sum_j \exp(-\lambda c_{ij})} \\ \textit{where} \\ D_j'' = \sum_i T_{ij} \end{cases}$$

Several things to note:

There is no attractor value at the destination – we would need to put this in as a constraint – i.e. a piece of information to be incorporated by the model

This is a location model – we predict activity at the destination – in the case of a model that predicts how many people working in zone i O_i live in zone j, this is D'_j where the prime ' is the notation for predicted

Now let us put this model back into the entropy equation and see what we get – let us put the model back in in its exponential form

$$p_{ij} = \exp(-\lambda_i - \lambda c_{ij})$$

Then what we get is

$$H = -\sum_i \sum_j p_{ij} \log p_{ij} = -\sum_i \sum_j p_{ij} (-\lambda_i - \lambda c_{ij})$$

$$= \sum_i p_i \lambda_i + \lambda \hat{C} = \sum_i p_i \log \sum_j \exp(-\lambda c_{ij}) / p_i + \lambda \hat{C}$$

What we need to note is that entropy is partitioned into a fixed energy and free energy – the fixed is the second term and the free is the first – a series of weighted log-sums and it is often thought of a kind of accessibility.

In this case it is the sum of accessibilities, one for each origin zone. It has strong relations to utility in the random utility maximising version of this kind of model which is central to discrete choice theory

Residential Location, Modal Split

Let me illustrate in two ways how we can build models using this framework

If we say that residential location depends on not only travel cost but also on money available for housing we argue as before that

The model is singly constrained – we know where people work and we want to find out where they live – so origins are workplaces and destinations are housing areas
The model then lets us predict people in housing
We argue that people will trade-off money for housing against transport cost

And we then set up the model as follows

This time using not the probability form but the trip activity-volume form, we get

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

$$\sum_i \sum_j T_{ij} c_{ij} = C$$

$$\sum_i \sum_j T_{ij} R_j = R$$

leads to

$$T_{ij} = A_i O_i \exp(\vartheta R_j) \exp(-\lambda c_{ij})$$

Note that we now add a constraint on money available for housing (like rent) R_j. We can of course find out from this location model how many people live in destination housing zones, so again it is a distribution as well as a location model

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

We can extend this model in lots of ways and we will show some of these later. We also can think about disaggregating the model into different transport modes – let us call each mode k and then set up the model so that we can predict T_{ij}^k as follows

The model is singly (origin) constrained because we want to predicts how many people travel from work to home. Given we know how many people work at origins, and we want to predict what mode of transport k they travel on. Then

$$\sum_j \sum_k T_{ij}^k = O_i$$

$$\sum_i \sum_j \sum_k T_{ij}^k F_j = F$$

$$\sum_i \sum_j T_{ij}^k c_{ij}^k = C^k$$

And the model can be specified as

$$T_{ij}^{k} = O_i \frac{F_j \exp(-\lambda^k c_{ij}^k)}{\sum_j \sum_k F_j \exp(-\lambda^k c_{ij}^k)} = O_i \frac{F_j \exp(-\lambda^k c_{ij}^k)}{\sum_j F_j \sum_k \exp(-\lambda^k c_{ij}^k)}$$

Note that the mode split is a ratio of the competitive effects of each travel cost, that is

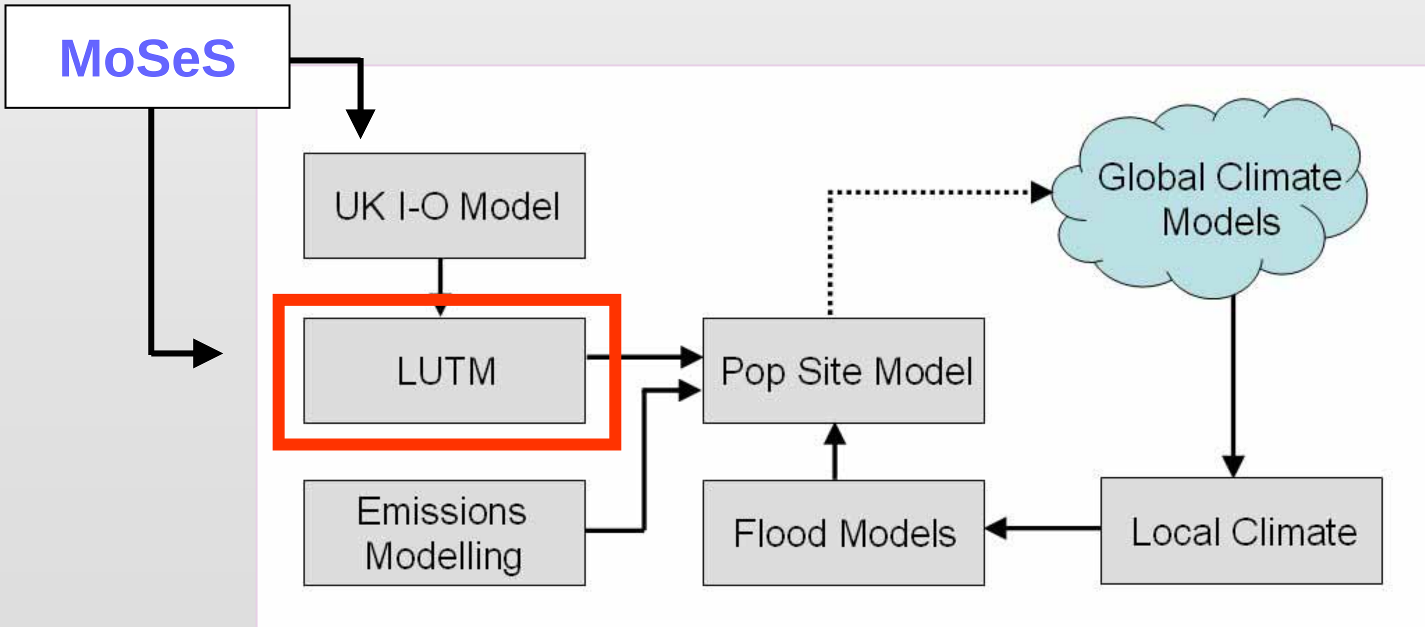
$$\frac{T_{ij}^{k}}{T_{ij}^{\ell}} = \frac{\exp(-\lambda^k c_{ij}^k)}{\exp(-\lambda^\ell c_{ij}^\ell)}$$

In short the model is not only distributing trips so that locations compete but also that modes compete BUT modes do not compete per se with locations

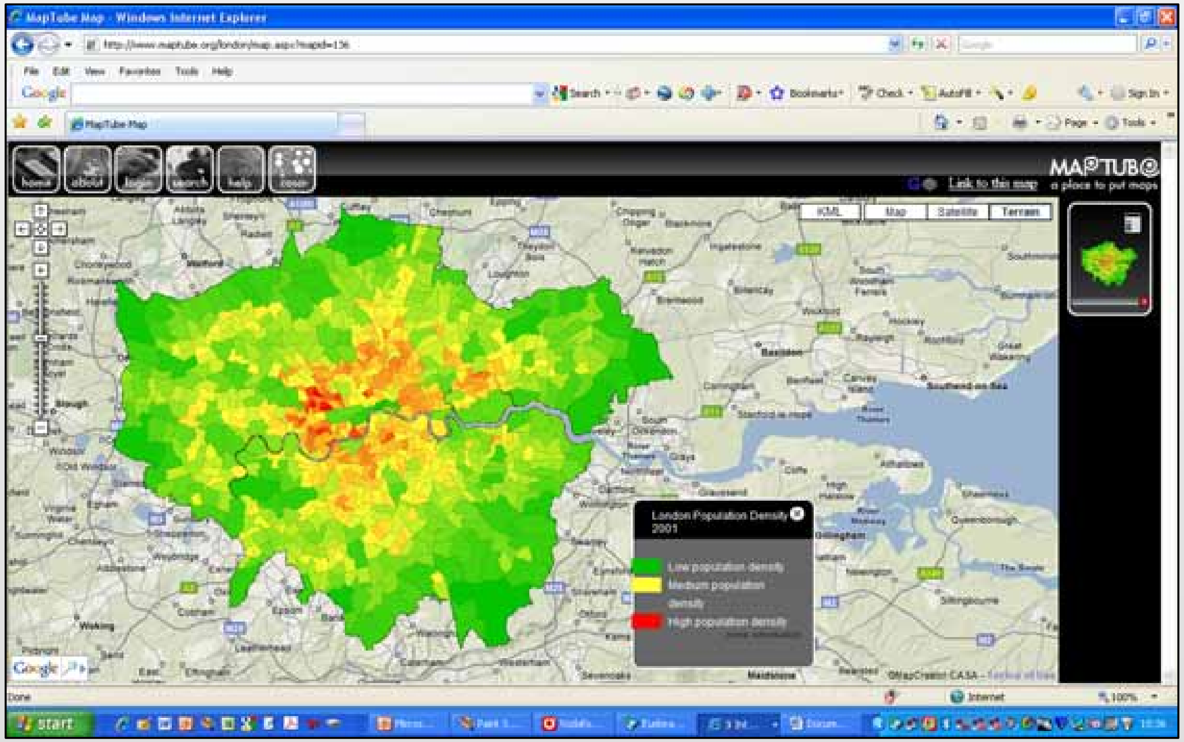
Now let us see how we can build this model for real

The London Tyndall Model: Applications

I already introduced the model last time as an example of an integrated model – with several components – the one we will dwell on here is the residential location model which is essentially a version of the singly constrained model that we have been outlining. Here is the block diagram of the model stages to remind you of what it is



Essentially we have built this model for Greater London which is divided into 633 zones – the area has 7.7m population and about 4.3m jobs – we have four modes – road (car), heavy rail, light rail and tube, and bus – walk/bike is a residual mode.
To fix ideas let me show the extent of the area first



Splash Screen – for the desktop version

London and the Thames Gateway Land Use Transportation Model

Cities Research Programme

Tyndall°Centre for Climate Change Research

CASA UCL

Newcastle

This program is a rudimentary land-use transportation model built along classical lines which allocates population and employment to small zones of the urban system. It uses spatial interaction principles which bind the population sector (residential or housing) to employment sector (work or industrial and commercial) through the journey to work (work trips) and the demand from services (which loosely translate into trips made to the retail and commercial sector).

The model is being built for Greater London and the Thames Gateway at ward level - 633 in all - so that it can be used in a wider process of integrated assessment focussed on assessing the impact of climate change on small areas in this metropolitan region. In particular rises in sea level and pollution are key issues, and as such the model sits between aggregate assessments of environmental changes associated with global and regional climate change models and environmental input output models, and much more disaggregate models related to the detailed hydrological implication of long term climate change.

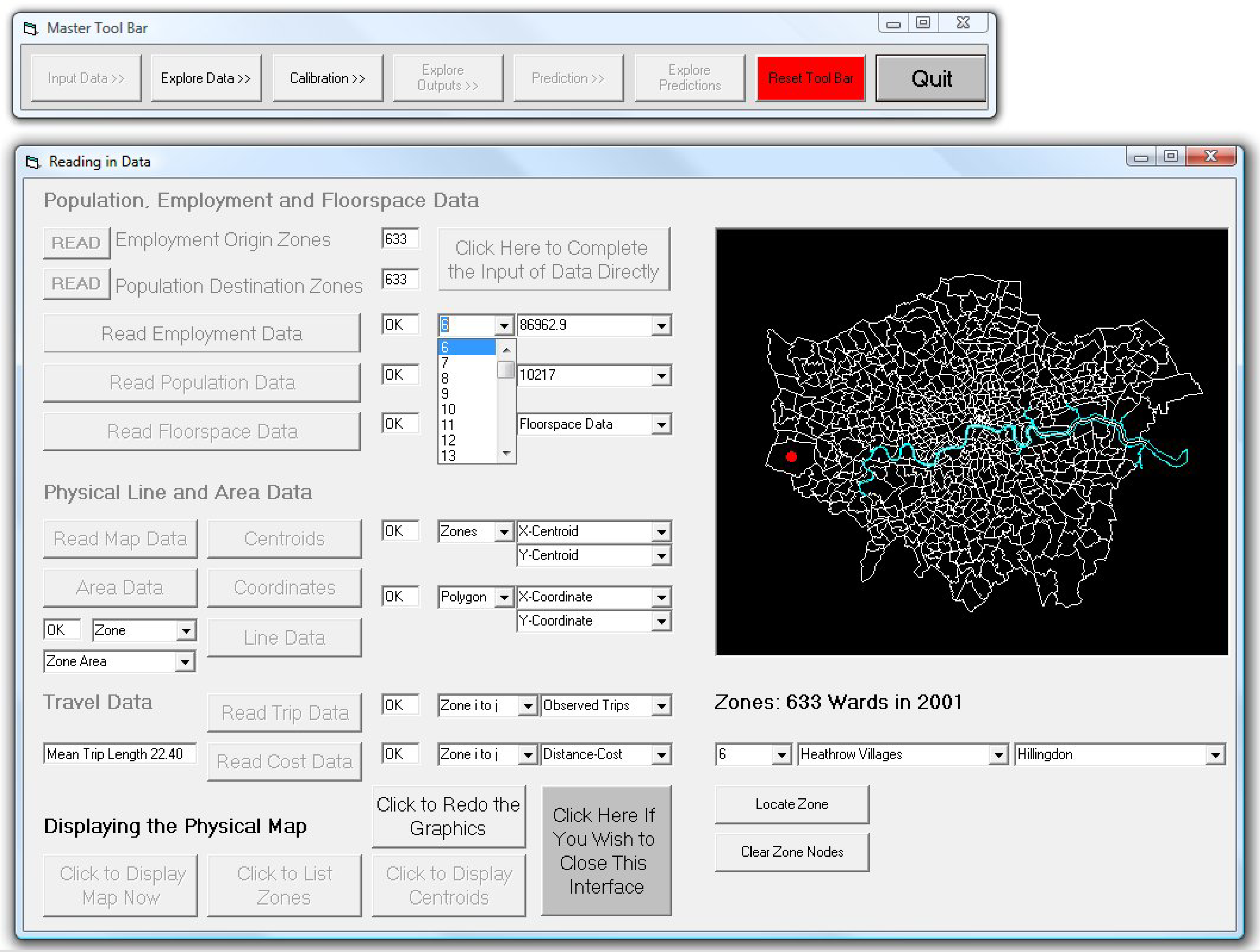
The programme enables the user to read in the data and explore it spatially, to calibrate the parameters of the model and explore its outputs spatially and to engage in various predictions ranging from the typical' business as usual scenarios' to much more radical changes posed limits on spatial behaviour which either result from climate change and, or mandated by government. The predictions and scenarios are intended to go out to 2100 and thus the model is largely designed as a sketch planning tool.

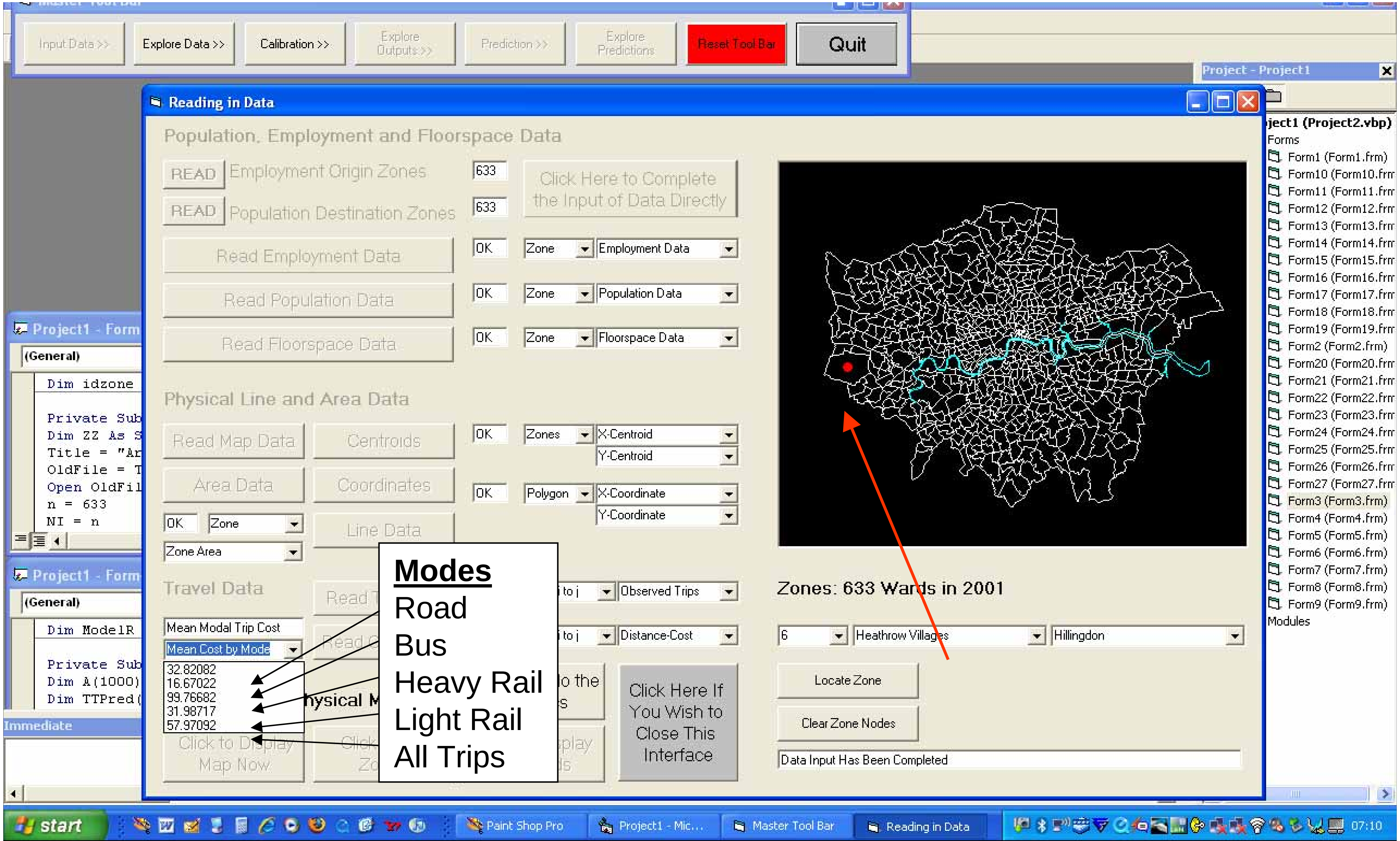
These various stages of the model contained in a master tool bar which is activated when the GO! button is pressed on this screen. The master tool bar enables the users to proceed through the various stages indicated and to display outputs in map and statistical form at any stage.

with GLA ECONOMICS LONDON

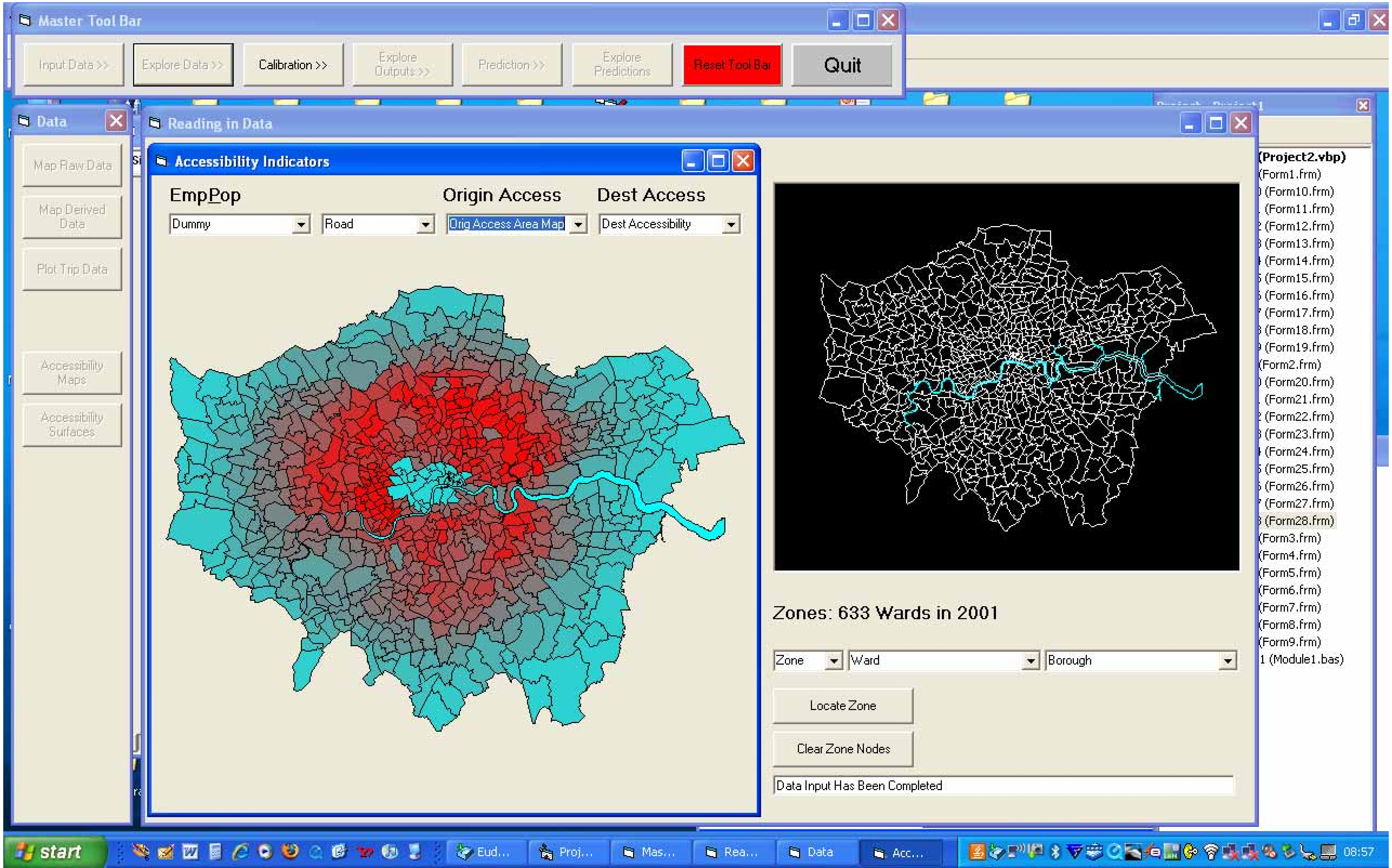
GO!

Program Manual

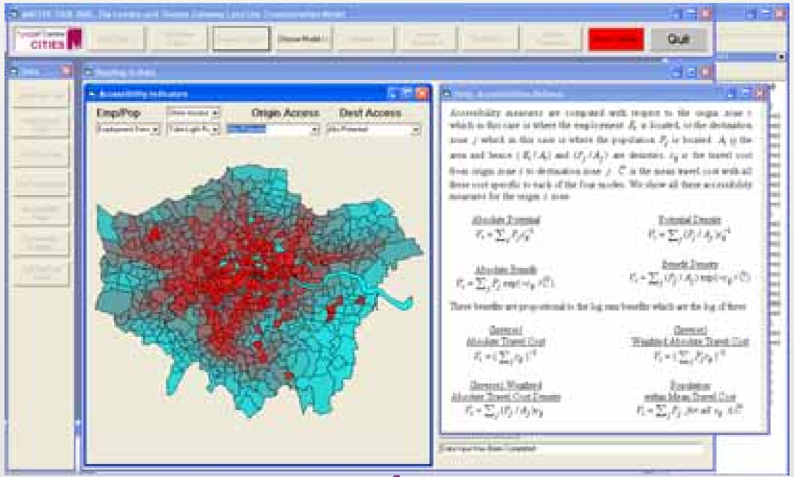

Master Tool Bar
Input Data >>
Explore Data >>
Calibration >>
Explore Outputs >>
Prediction >>
Explore Predictions
Reset Tool Bar
Quit
Reading in Data
Population, Employment and Floorspace Data
READ
Employment Origin Zones
633
READ
Population Destination Zones
633
Click Here to Complete the Input of Data Directly
Read Employment Data
OK
86962.9
Read Population Data
OK
10217
Read Floorspace Data
OK
Floorspace Data
Physical Line and Area Data
Read Map Data
Centroids
OK
Zones
X-Centroid
Y-Centroid
Area Data
Coordinates
OK
Polygon
X-Coordinate
Y-Coordinate
OK
Zone
Zone Area
Line Data
Travel Data
Read Trip Data
OK
Zone i to j
Observed Trips
Mean Trip Length 22.40
Read Cost Data
OK
Zone i to j
Distance-Cost
Displaying the Physical Map
Click to Redo the Graphics
Click Here If You Wish to Close This Interface
Click to Display Map Now
Click to List Zones
Click to Display Centroids
Zones: 633 Wards in 2001
6
Heathrow Villages
Hillingdon
Locate Zone
Clear Zone Nodes



Road: 38%; Bus: 12%: Heavy Rail: 12%: Light Rail 19%; Other (Walk, Bike, Fly): 19%


Master Tool Bar
Input Data >>
Explore Data >>
Calibration >>
Explore Outputs >>
Prediction >>
Explore Predictions
Reset Tool Bar
Quit
Data
Map Raw Data
Map Derived Data
Plot Trip Data
Accessibility Maps
Accessibility Surfaces
Reading in Data
Accessibility Indicators
EmpPop
Origin Access
Dest Access
Dummy
Road
Orig Access Area Map
Dest Accessibility
Zones: 633 Wards in 2001
Zone
Ward
Borough
Locate Zone
Clear Zone Nodes
Data Input Has Been Completed
(Project2.vbp)
(Form1.frm)
(Form10.frm)
(Form11.frm)
(Form12.frm)
(Form13.frm)
(Form14.frm)
(Form15.frm)
(Form16.frm)
(Form17.frm)
(Form18.frm)
(Form19.frm)
(Form2.frm)
(Form20.frm)
(Form21.frm)
(Form22.frm)
(Form23.frm)
(Form24.frm)
(Form25.frm)
(Form26.frm)
(Form27.frm)
(Form28.frm)
(Form3.frm)
(Form4.frm)
(Form5.frm)
(Form6.frm)
(Form7.frm)
(Form8.frm)
(Form9.frm)
1 (Module1.bas)
start
Eud...
Proj...
Mas...
Rea...
Data
Acc...
08:57

Accessibility from the LUTM model Many different accessibility measures, 8 in all



Help: Accessibilities Defined

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

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

Long Term Scenarios Based on the Impact of Changes in Employment, Residential Floorspace, and Transport Costs

Predictions with the model involve forecasting the location of small area populations and the trip patterns associated with the four modes used to distribute employment as population to these small (residential) areas. This involves changing the input variables - employment and residential floorspace by small area, and the travel costs associated with each mode of transport, which in turn imply changes to the transport infrastructure. The user also has control over the parameter values on the friction of Travel Cost or travel cost associated with each mode. This can be changed in value to reflect changes in the average Travel Cost or cost travelled on each mode.

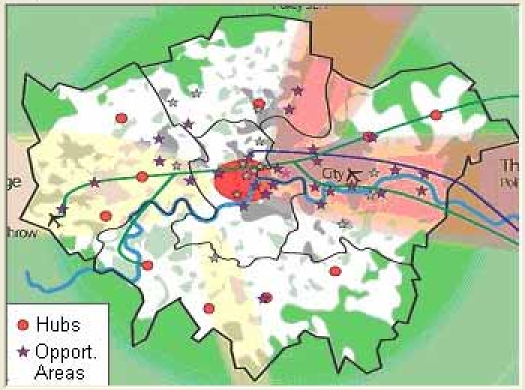
Users have a choice of inputting a preset scenario in which all these variables are changed exogenously or a process of changing these variables interactively, on screen. The interactive process can involve many thousands of changes and is probably best used to input data which reflects 'what-if' scenarios which require a small number of rather simple changes in the inputs reflecting substantial or radical change.

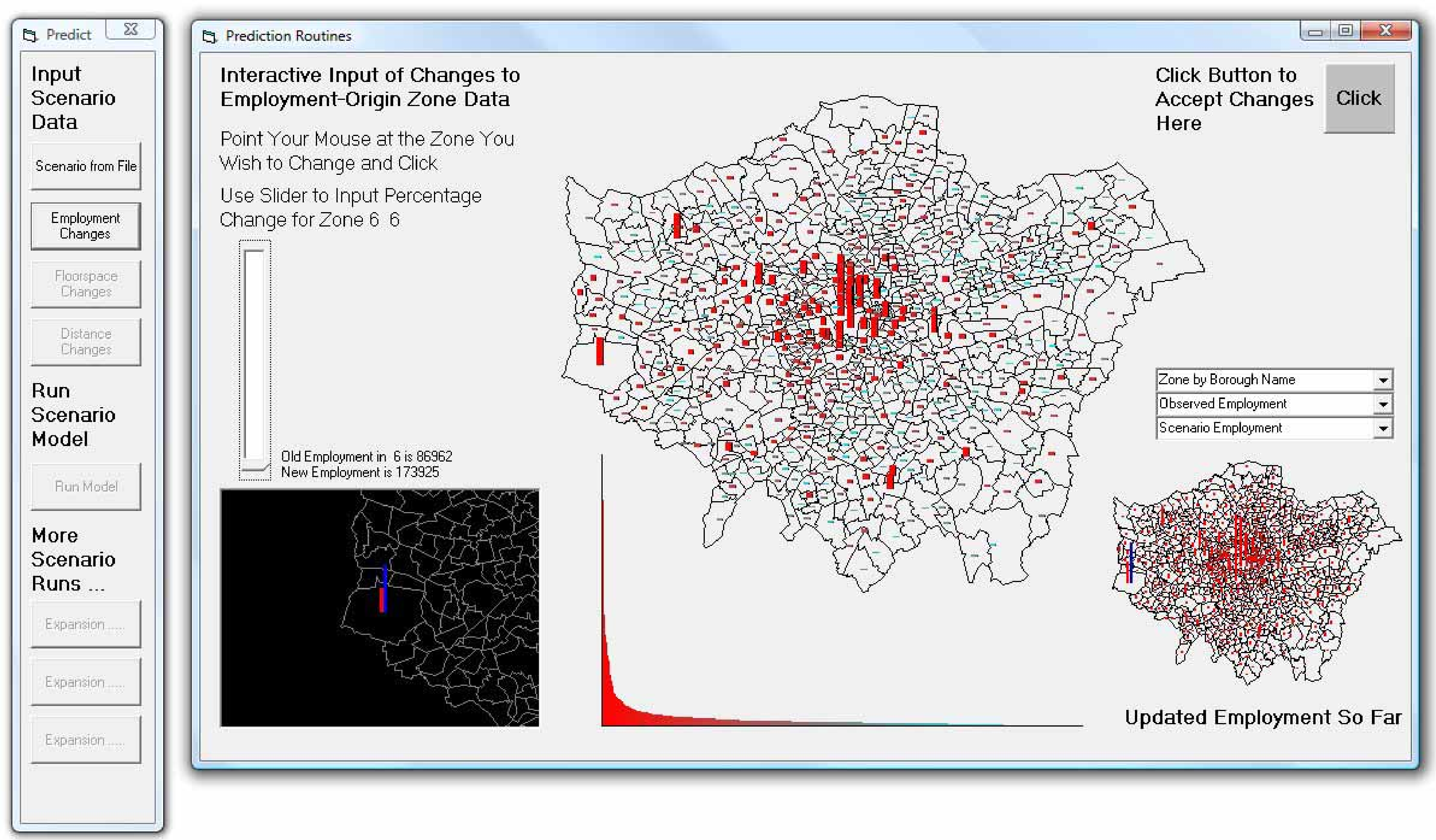
By clicking the 'Scenario from File' button in the toolbar to the left, a preset scenario is loaded and the user is then taken to the point where the model must be run. Alternatively if the user clicks the Employment Changes button, the user activates a screen where each employment zone can be identified by pointing the mouse at it and clicking. Then the user can use a slider bar to increase the value of employment in that zone by up to 100 percent or decrease it by up to 100 percent. As many zones as required can be changed using this method. When the user is satisfied with the employment scenario which has been developed, a button accepting these changes can be clicked. The same can then be done for floorspace activated by clicking the relevant button from the toolbar to the left.

Finally the travel cost on any link by any mode from one zone to another can be changed using the same method. An origin and then a destination zone need to be clicked and then reduced or increased travel cost (by up to 100 percent) made using the slider bar. The user must choose the mode each time and the program then recomputes all the shortest routes implied by these changes once the changes are accepted.

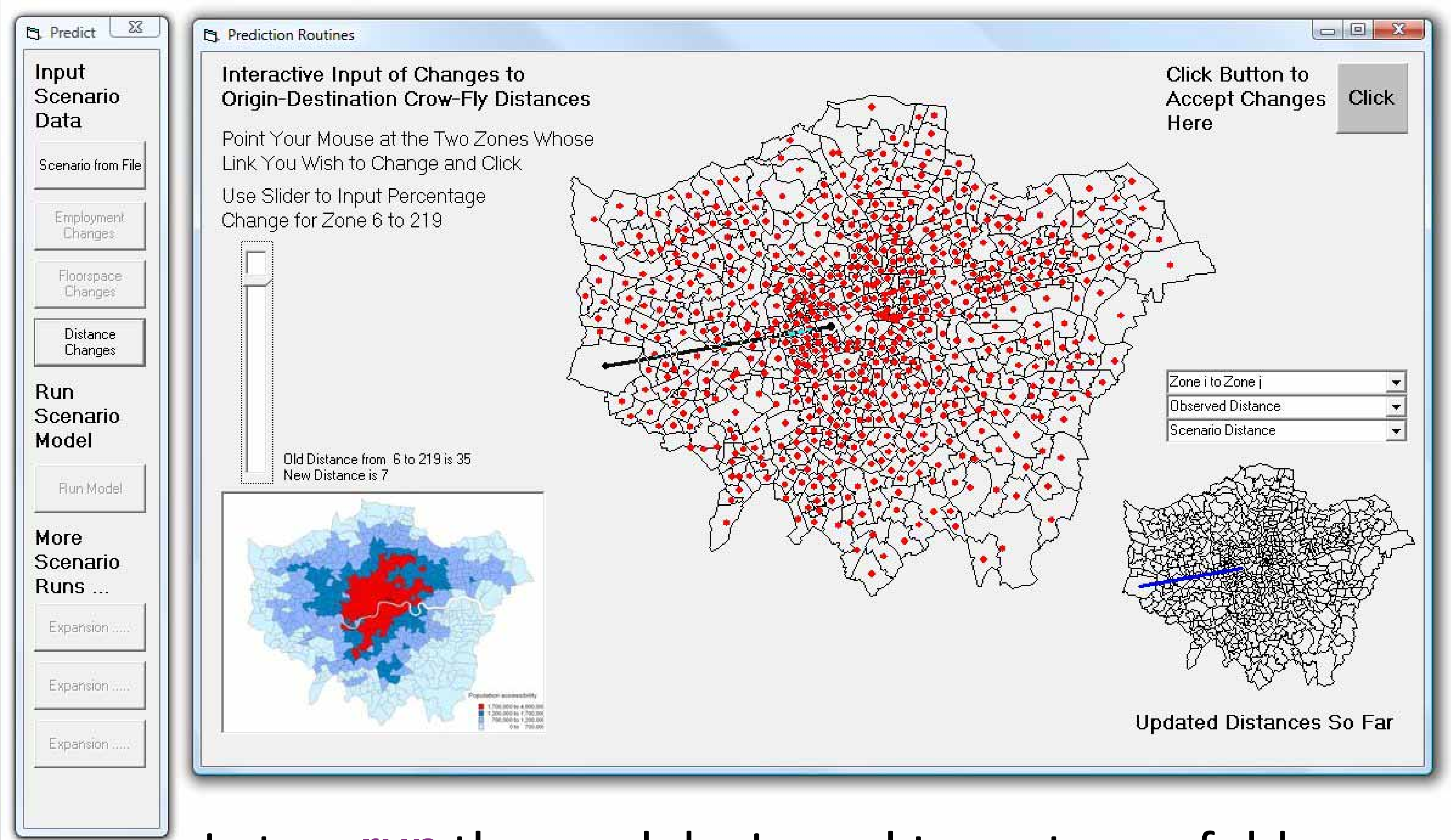
The user then proceeds to run the model as for the 'Scenario from File' option and once this is done, the outputs can be visualised using the same system for exploring the data and calibration results.

Key Elements of the London Plan to 2025 Shown Below.





Predict
Input Scenario Data
Scenario from File
Employment Changes
Floorspace Changes
Distance Changes
Run Scenario Model
Run Model
More Scenario Runs ...
Expansion .....
Expansion .....
Expansion .....
Prediction Routines
Interactive Input of Changes to Employment-Origin Zone Data
Point Your Mouse at the Zone You Wish to Change and Click
Use Slider to Input Percentage Change for Zone 6 6
Old Employment in 6 is 86962
New Employment is 173925
Click Button to Accept Changes Here
Click
Zone by Borough Name
Observed Employment
Scenario Employment
Updated Employment So Far



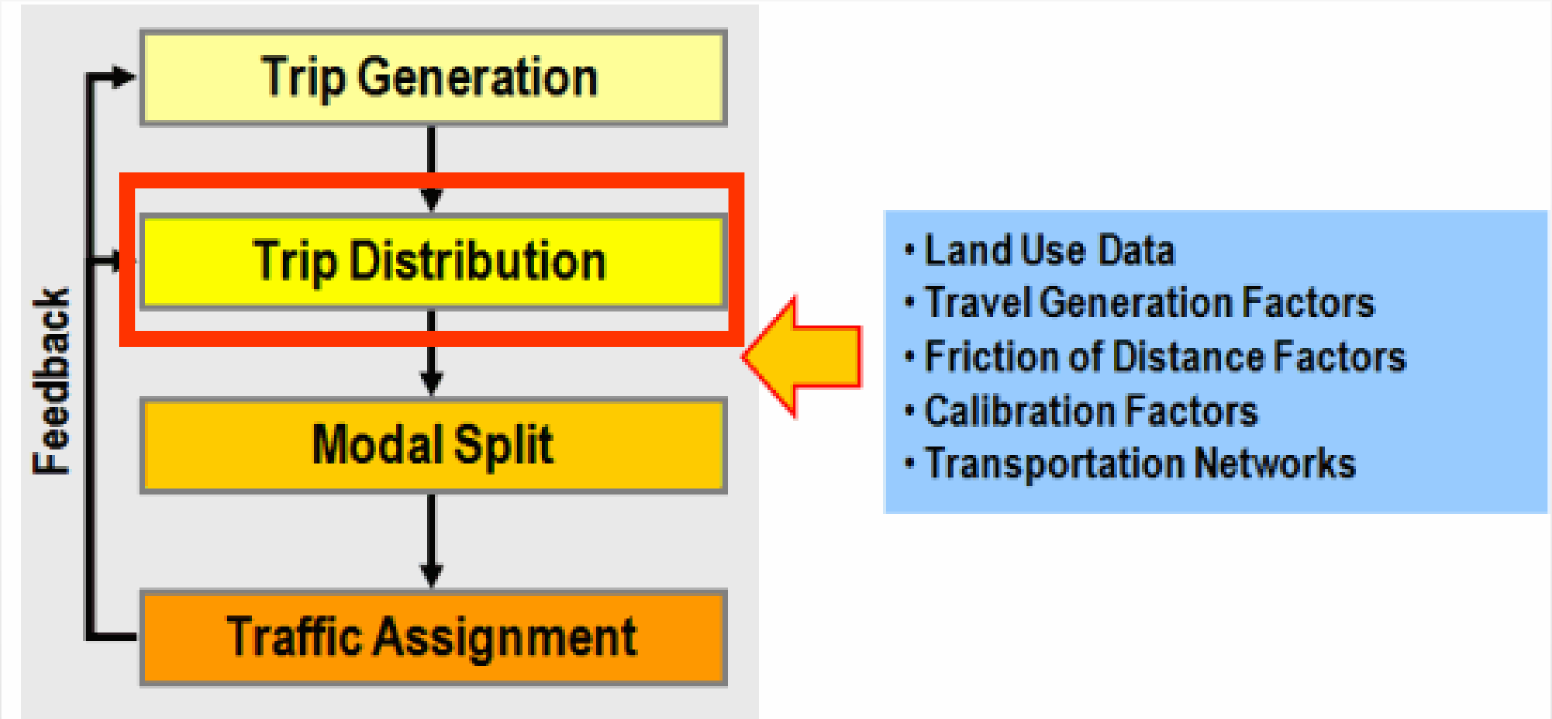
Let us *run* the model… I need to go to my folder…>>

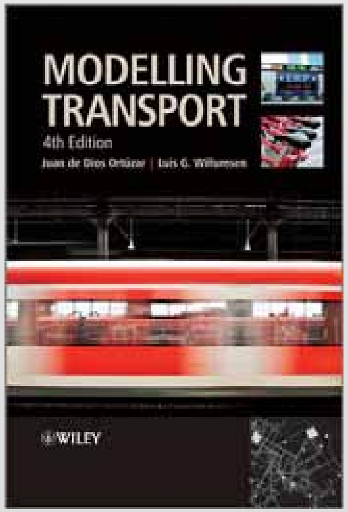
Run

Transportation Modelling: The Four Stage Process

I should make a brief point about transport modelling – we have included transport and location together here but traditionally the transport model is based on a four stage process that involves generation, distribution, modal split and assignment

The other issue is that in the standard transport modelling process, once trips are assigned to the network, then one can assess whether the network can take the load – this is matching travel demand against supply and if not then the model is iterated to match demand to supply. This is another generic issue in urban modelling – demand and supply and the way the market resolves this.





Juan de Dios Ortúzar, Luis G. Willumsen 2011 ***Modelling Transport***, 4th Edition, Wiley, New York

Next Time: Modular Modelling: Coupled Spatial Interaction

Now we have a module for one kind of interaction – consider stringing these together as more than one kind of spatial interaction

Classically we might model flows from home to work and home to shop but there are many more and in this sense, we can use these as building blocks for wider models. This is for next time too

What we will do is illustrate how we might build such a structure taking a journey to work model from Employment to Population and then to Shopping

Lectures on Urban Modelling January 2017

Thanks – More Next Thursday, same time

Michael Batty

m.batty@ucl.ac.uk

@jmichaelbatty

13 January, 2017

http://www.spatialcomplexity.info/