Foundations of Urban Science

Recitation 4 – November 22nd
Topics: Spatial Modeling and Systems Interactions
Assignment: Assignment #4

---

## The Assignment: Exploring Spatial Interaction Models

What I want you to do in this assignment is write a simple program for computing a spatial interaction-gravity model in a spreadsheet (or in any programming language you care to use) and then compare the results with some real data that consists of the worktrips between employment origins and residential destinations in Greater London. This consists of 33 boroughs, each somewhat smaller than the 5 boroughs that make up NYC. I provide all the data in the attached spreadsheet that you can download from the blog.

The model I am proposing you fit can be written as

$$T'_{ij} \propto O_i D_j d_{ij}^{-\alpha} = K O_i D_j d_{ij}^{-\alpha}$$

where $T'_{ij}$ are the 'predicted' work trips between zone $i$ and zone $j$, $O_i$ is the origin activity, in this case employment in the workplace in origin zone $i$ and $D_j$ is the destination activity, in this case employment in the residential place in zone $j$. $d_{ij}$ is the distance from zone $i$ to zone $j$, $\alpha$ is a parameter sometimes called the friction of distance to be chosen to fit the model best, and $K$ is a normalising constant which can be scaled to ensure the total trips generated are equal to those observed.

Now from the data matrix of trips $T_{ij}$, we note that the sum over origins and the sum over destinations gives us a measure of the mass at each which we can assume is data that we use in the model; that is

$$O_i = \sum_j T_{ij} \quad \text{and} \quad D_j = \sum_i T_{ij}$$

What you are required to do is compute the predicted trips $T'_{ij}$ by choosing a value of the parameter $\alpha$ and then compute a goodness of fit statistic between the observed and predicted data. There are many such statistics such as the correlation but one which is easy to understand is the sum of the squared differences which we can write for the observed and predicted trips as

$$\Delta = \sum_{ij} (T_{ij} - T_{ij})^2$$

We seek to choose the parameter $\alpha$ to minimise this.

***The Data:*** In the spreadsheet I have given you the following

1. a list of the 33 boroughs by name
2. the $X_i, Y_i$ coordinates of each borough (these are latitude and longitude – I do not know what projection these are in but this is of no consequence at this scale)
3. The origin $O_i$ and the destination $D_j$ activity for each zone/borough but note that these are actually calculated from the row and column sums of the trip matrix respectively
4. The work trip matrix whose rows are origins – the employment workplaces, and whose columns are destinations – employment at the residential living place

What you have to do is as follows:

1. compute the raw distances between all origins and destinations from

   $$d_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}$$
2. Note that $d_{ii} = 0$. We need to ensure that this is positive and thus if you examine the distance matrix and work out what the mean distance is from any zone to its immediate neighbours and then take one third of this distance that would be a first approximation to this intrazonal distance.
3. You now have all the data and you can compute the predicted trips from the formula above but you need to choose a value for $\alpha$
4. You can then work out the squared differences and sum these
5. Now what you might like to do it rerun the model with a few different parameter values say $\alpha = 0.5, 1, 2, \ldots$ and see how the sum of squared differences changes. This essentially is the process of calibrating the model
6. Once you have got this far you can sum the rows and columns of the predicted matrix and see how good these are relative to the observed rows and columns
7. For those of you who are really adventurous you could visualise the observations and predictions in some other software or even in the spreadsheet itself.

All the data and instructions are on the attached spreadsheet. You could do the entire exercise in the spreadsheet and if you do, then you will need to produce new tables for distances and then for predicted trips and then take the difference of observed and predicted and square these for another table, and then add over all the values to get the goodness of fit.

*Mike Batty (m.batty@ucl.ac.uk)*
*Tuesday, 19 November 2013*