UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 158 - Aug 10

Exploring the possibility-cone of urban development

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Exploring the possibility-cone of urban development

Alan Wilson and Joel Dearden

Centre for Advanced Spatial Analysis, University College London

Abstract

It is well known that urban systems are complex and nonlinear and exhibit properties of multiple equilibria and path dependence. This makes model-based forecasting in a conventional sense impossible. In this paper we explore how to achieve a deeper understanding of path dependence by constructing a "possibility cone" - an envelope that contains feasible possible futures for a given urban system; and then exploring the consequences of this analysis for planning. Following Wilson (forthcoming) we characterise the factors that shape this cone as "urban DNA". In the planning context, we can then examine the idea of urban "genetic medicine" and ask the question: what interventions would be required to move an urban system into a new possibility cone? These ideas are explored using a stochastic version of the archetypal aggregate retail model using data for the South Yorkshire retail system. We produce a visual representation of the cone of development using a parallel coordinates plot to allow easier interpretation of the results.

Keywords: urban modelling complex systems path dependence visualisation

1 Forecasting the future of path dependent urban systems

It is well known that urban systems are complex and nonlinear and exhibit properties of multiple equilibria and path dependence. This makes model-based forecasting in a conventional sense impossible. Path dependence is a key concept - first explored by Arthur (1988) who showed that history can influence industrial location patterns if agglomeration economies affect that industry. Agglomeration, as a form of positive feedback, allows the possibility of multiple solutions to the industrial location problem. In a planning context, the traditional use of urban models has been for exploring "what if" questions, for example, for a given set of plans, what is the outcome? This paper shows how to interpret this goal given the consequences of nonlinearity through a focus on path dependence.

The value of each of the variables and parameters of an urban model at one time constitute a point in state space. Since, typically, many variables and parameters will be needed for an adequate description, this will be a very high dimensional space. We find it helpful to follow Wilson (forthcoming) and to think of the initial conditions - that are either exogenous in the model or part of the slow dynamics[1] - as the "DNA" of the system because these conditions largely determine the possible models of development. At each of a sequence of points in time, the state vector will constitute the initial conditions for the next step. There are then

[1] 'Fast dynamics' variables such as interaction arrays will be calculated within the model and should not be thought of as part of the DNA.

two ways in which the possibility-cone of development can be influenced and hence determined: (1) we can introduce stochastic variation into the dynamics of each model run from some time t – thus generating a set of varying outcomes at time t+1 and (2) we can vary the exogenous DNA in the initial conditions at time t to produce a range of outcomes at time t+1. The envelope of the initial conditions at time t and the outcomes at t+1 forms a section of the possibility-cone. Again, it should be emphasized that this is a 'cone' in a high dimensional space.

The initial conditions typically will not represent an equilibrium state of the system. However, at each point in time, there will be equilibrium states that will be influencing future development. That is, there will be basins of attraction for multiple solutions, particularly when positive feedback is present. There are then interesting questions to be explored. Which basins of attraction are within the possibility-cone and which not? In the case of 'not', and if such a state is a desirable one, is it possible to make an adjustment in the initial conditions that brings that attractor within the cone? That is potentially a perspective to be adopted by planners in a complex nonlinear world. This is in a way asking the question, "what can be achieved by planning?". It relates to the way in which urban systems can become "locked in" to a given development path by the initial conditions. In dynamical systems terms this represents being stuck near to an undesired basin of attraction. An example in the UK might be the seaside towns that are in decline and are experiencing negative "lock in" to a restricted set of possible futures. There is the potential here to explore what is necessary to break them out of a negative development path.

In section 2, we explore in more detail how to construct and represent a possibility-cone. We use an aggregate urban retail model to as a demonstrator – first in section 3 by introducing stochastic variation of the initial conditions and then in section 4 by tackling a hypothetical planning application.

2 Constructing the possibility-cone of an urban system

Meteorologists, being uncertain about the path that a hurricane will take across the surface of the earth, represent the diverging set of possible paths as cone (Figure 1) which illustrates the increasing uncertainty about the predictions. Similarly, it is useful to think of all the possible development paths of an urban system as a cone to highlight the potential for them to quickly diverge given a vast array of possible futures. We can use this idea to determine the likely envelope that contains the real future state of the system. As we have indicated, the "cone" exists in the very high dimensional state space of the system.

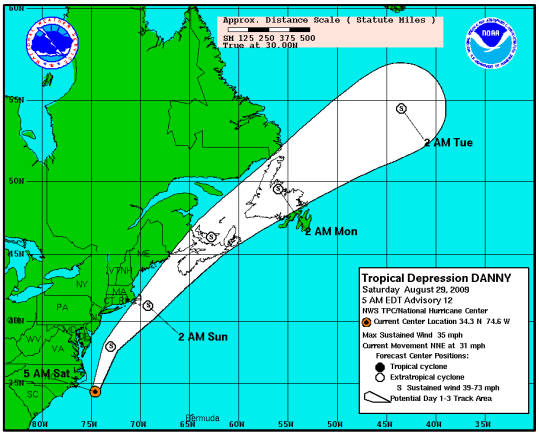

Figure 1. Forecast track for Hurricane Ernesto in 2009 (NOAA)

We have noted that if we take a given urban system at a particular moment in time then we can consider "the underlying structural variables to be the urban analogues of DNA" (Wilson forthcoming). The "physiology" of the city is its activity and development given the starting DNA – the fast dynamics predicted by the model. To define the cone in modelling terms we need to make a distinction between the exogenous DNA that our model does not directly adjust and endogenous DNA that our model adjusts and predicts. The cone of development we are envisioning here can be thought of as a rapidly diverging set of possibilities starting from a single set of endogenous DNA. We can begin by thinking of the many possibilities as a tree (Figure 2), each level of which represents one step through time. Each 'point' in this diagram is, of course, symbolic because it represents a high-dimension state vector.

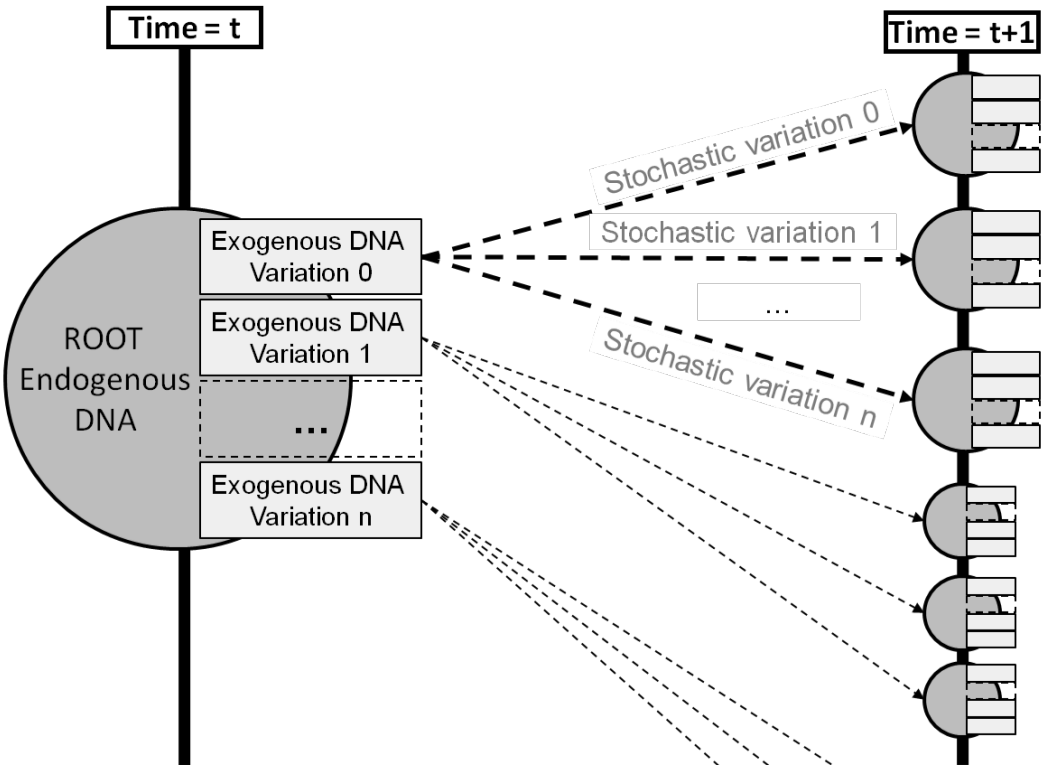

Figure 2. The tree structure inside the possibility-cone of development

At the root of the tree is the initial endogenous DNA representing the current state of the system of interest. A number of branches extend from the root node representing development paths of the system toward equilibrium. The branching, recall, is brought about in one (or both) of two ways: through noise, representing unpredictable events; or through adjustments to the exogenous DNA which represent either a planned intervention such as a new shopping centre or transport link; or an unplanned change such as an increase or decrease in petrol prices. Here we represent the planned interventions as a single binary choice at a specified time in the tree, e.g. we either build a new shopping centre or we don't. Unplanned changes are represented by two or more branches which are allowed to occur at every time step in the tree. One of the branches would be the "no effect" option while the others would represent each of the possible values that can have an effect on the system, e.g. fuel prices can increase, decrease or remain the same and can do so at any point in time.

Computer software is used to recursively construct the tree. Starting with the root node, and for each child node following it in the tree at time t, we generate a number of branches equal to (x_t * s) where: x_t is the number of different exogenous DNA sets available at time step t and s is the number of stochastically varying model runs we do with each exogenous DNA set. In order to represent a wide range of potential variations we would need to set s to a large value and so through time, the number of the branches in the tree will increase very rapidly. If m is the number of time steps in the tree then we can calculate the number of model runs required to construct the tree, N, as:

$$N = 1 + \sum_{t=0}^{m-1} \prod_{i=0}^{t} (x_i * s) \qquad (1)$$

Clearly a significant amount of computing power would be required to construct a deep cone for anything other than the simplest of models.

Each tree branch represents one model run from a complete DNA set at time t and this may or may not be iterated to equilibrium. The output from each model run represents a new and likely unique endogenous DNA set that forms a new node in the tree at the end of the branch. By choosing a path through the tree from the root to a leaf node we can see that the tree represents a wide range of possible development paths that cover a number of time steps. The envelope of this process is the cone of possible development.

Dealing with stochastic variation is relatively straightforward and we give an example in the next section. In order to represent a range of possible future events we can build a number of varying exogenous DNA sets for each time step in the tree. Each set of exogenous DNA will represent one permutation of the events that could possibly occur at the time step in question. An example set of future events that could occur at some time step t could be: fuel price fluctuations, planned development of a new shopping centre and construction of a new road. The set of exogenous DNA set we generate from this would then include every permutation of the three events (Table 1) and can be thought of as a set of possible future scenarios for the system at time t. Computer software can obviously be used to automatically generate these sets, given the set of possible events.

Exogenous DNA set index	Fuel prices…	Shopping Centre…	Road…
0	Increase	Not built	Not built
1	Increase	Not built	Built
2	Increase	Built	Built
3	No change	Not built	Not built
4	No change	Not built	Built
5	No change	Built	Built
6	Decrease	Not built	Not built
7	Decrease	Not built	Built
8	Decrease	Built	Built

Table 1. Example exogenous DNA variation set for one time step

It is easy to see that if we move from a small set of highly probable events to including increasingly unlikely events the possibility cone will grow and will be more likely to contain the real future state of the system. A balance needs to be struck however because there are obvious limits on the resources of time and computing power available to construct such a cone.

3 A stochastic version of the aggregate retail model

To construct a realistic possibility-cone of development for a particular town or city would likely require a comprehensive model that represented multiple sub-systems and the interdependencies that exist between them. For the purposes of demonstrating the techniques, we concentrate in this study on modelling a single sub-system - urban retail - however the approach could be applied to more comprehensive models and this is an intended area of future research.

In order to model a retail system we adapt the archetypal aggregate retail model developed by Harris and Wilson (1978). The model divides an urban system into retail zones j. We define W_j as the amount of retail floor space in each retail zone and this value also represents the attractiveness of that retail zone to consumers. We also define a number of residential zones i with population P_i and average spending power e_i. The combined spending power of all the consumers in residential zone i is e_iP_i. The cost of travelling from zone i to zone j is given by c_{ij}, for which we use Euclidian distance. We represent the varying public transport provision for each retail zone by adding an additional multiplier m_j which affects every travel cost c_{ij} into retail zone j. Here the parameter α represents the impact of retail zone size on consumer decisions about where to shop. We would expect to see path dependence wherever this parameter is greater than 1, representing increasing returns to scale and positive feedback, and giving rise to the possibility of multiple solutions. The β parameter represents the impact of travel cost on consumer shopping decisions. Figure 3 lists the parameters and variables used in this model, indicates how each evolves over time and identifies those that make up the system DNA.

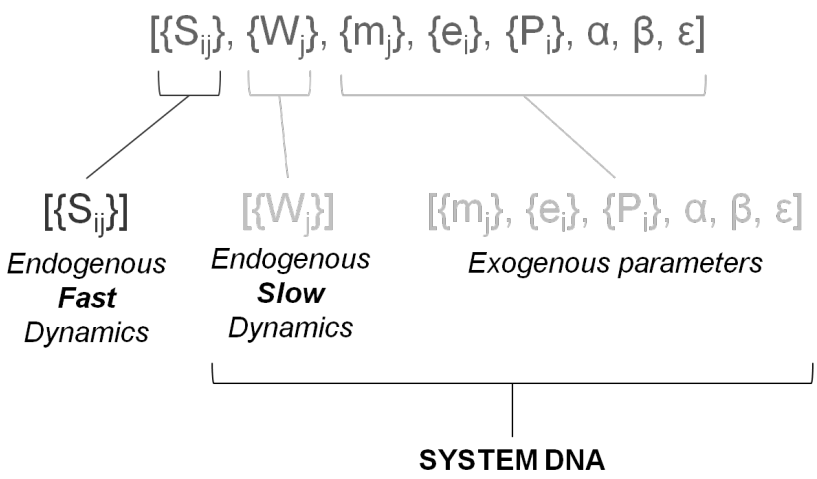

Figure 3. How the model parameters and variables

The flow of consumer spending from residential zone i to retail zone j is given by:

$$S_{ij} = A_i e_i P_i {W_j}^{\alpha} e^{(-\beta m_j c_{ij})} \tag{2}$$

where

$$A_i = \frac{1}{\sum_k {W_k}^{\alpha} e^{(-\beta m_k c_{ik})}} \tag{3}$$

to ensure that

$$\sum_j S_{ij} = e_i P_i \tag{4}$$

we can calculate the total flows into destinations as

$$D_j = \sum_i S_{ij} = \sum_i \left[\frac{e_i P_i {W_j}^{\alpha} e^{(-\beta m_j c_{ij})}}{\sum_k {W_k}^{\alpha} e^{(-\beta m_k c_{ik})}}\right] \tag{5}$$

A suitable hypothesis for representing the dynamics is (Harris and Wilson 1978):

$$\frac{dW_j}{dt} = \varepsilon\left(D_j - KW_j\right) \tag{6}$$

In order to represent the random events which can affect path dependence we add a stochastic term to the dynamics to give:

$$\Delta W_j = \varepsilon\left(D_j - KW_j\right) + W_j \varphi \tag{7}$$

Where φ is a stochastic[2] term drawn from a normal distribution with a mean (σ) of zero and a standard deviation of 0.02. This results in a low level of noise in the system that is proportional to the size of the retail zone. Here we assume that the greater the number of retail units in a zone the greater the potential for unpredictable change. This also keeps the total amount of stochastic variation in the system roughly constant (due to a constant total amount of floor space in the system). We draw the α and β parameters at each iteration from a normal distribution where our chosen α and β parameters are the mean and the standard deviation is 0.1. The stochastic variation in the β parameter represents minor fluctuations in travel cost and the fluctuations in α represent our uncertainty about the real level of returns to scale in the retail industry.

Each model run is iterated for a fixed number of iterations. This represents the fact that the system is being influenced by the solution(s) whose basin(s) it inside however it is unlikely to ever settle on one solution before conditions change. The iterative process represents the system converging towards an equilibrium state, however we choose here to explicitly represent the passage of time as a single step from a set of initial W_j values to the set of W_j values that exist at the end of a model run. These may or may not be equilibrium values depending on whether the model run converges.

In order to make the structure of the resulting possibility cone accessible for both analysis and communication purposes we visualise its structure. We are effectively plotting the tree from Figure 2 in the multidimensional state space of the system. Each node in the tree represents a state of the system and is represented by a $\{W_j\}$. This obviously presents some challenges because it is a high dimensional structure and contains a lot of information. Parallel coordinates (d'Ocagne 1885; Inselberg 1985) are one of the clearest and most intuitive of the multi-dimensional visualisation techniques available. This approach is able to clearly represent large numbers of model runs and also allow easy depiction of the range of values that appear. In our representation of the cone we add a third dimension to the parallel coordinates plot in order to represent time. We also plot splines rather than jagged lines to make it easier to distinguish each individual n-dimensional point plot (Moustafa and Wegman 2002).

We demonstrate the techniques developed so far by constructing and visualising a simple possibility cone for South Yorkshire, a metropolitan county in the UK containing one city and a number of large towns. We choose this region as our test case because it is small enough to be manageable while still containing some interesting and complex relationships between the competing retail zones. The data we use to initialise the model contains nineteen retail zones and ninety five residential zones. The retail zone position and floor space data is taken from the 2004 Town Centres Project (available online at www.planningstatistics.org.uk). The residential zone data is derived from the UK 2001 census – we use CAS Ward level population and boundary data together with CACI paycheck income data. A map showing the region is given in Figure 4 with the major retail

[2] Appendix 1 gives details of the random number generators used.

zones labelled. The heights of the bars in the figure illustrate the initial $\{W_j\}$ that represents the initial endogenous DNA for our example possibility cone.

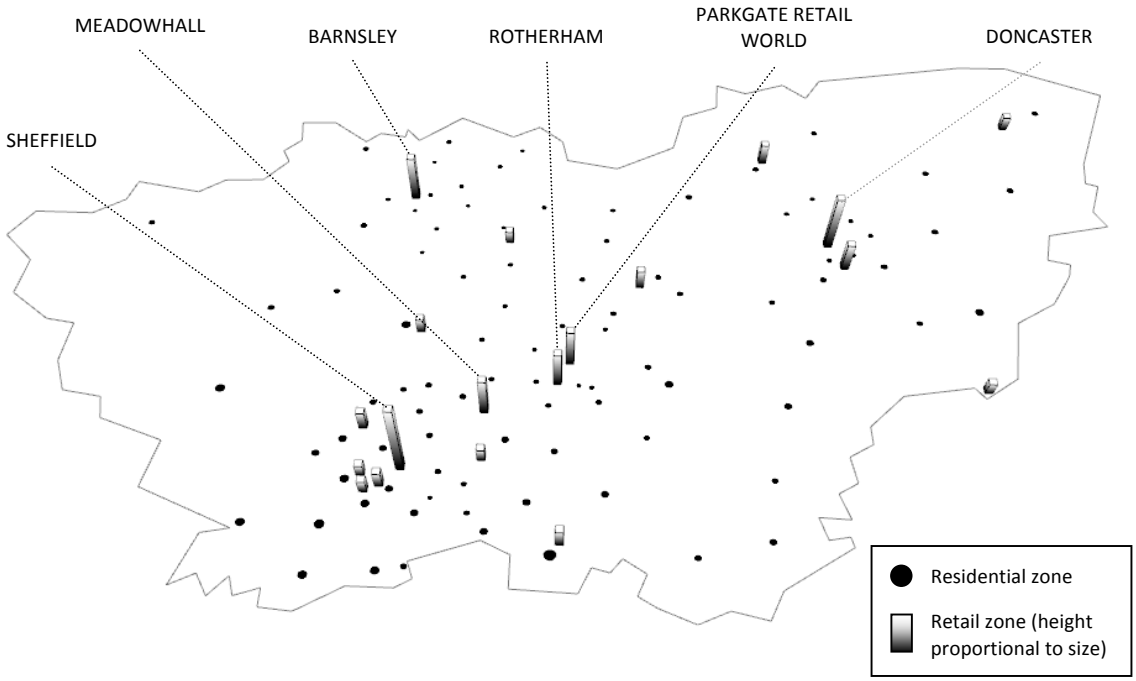

Figure 4. The South Yorkshire retail system

In order to keep the example as simple as possible we hold the exogenous DNA constant and rely on stochastic variation to differentiate each model run. We also model only a single time step into the future. The exogenous DNA was calibrated using a genetic algorithm: genes representing a subset of the exogenous model parameters (α, β, $\{m_j\}$) were evolved to maximise a fitness defined as the R-squared between the initial conditions and the equilibrium state of the system. A reasonable best fit was found with an R-squared of 0.89. Three hundred model runs were used to construct the cone and each branch evolves for 200 iterations towards equilibrium. The resulting cone of development can be seen in Figure 5.

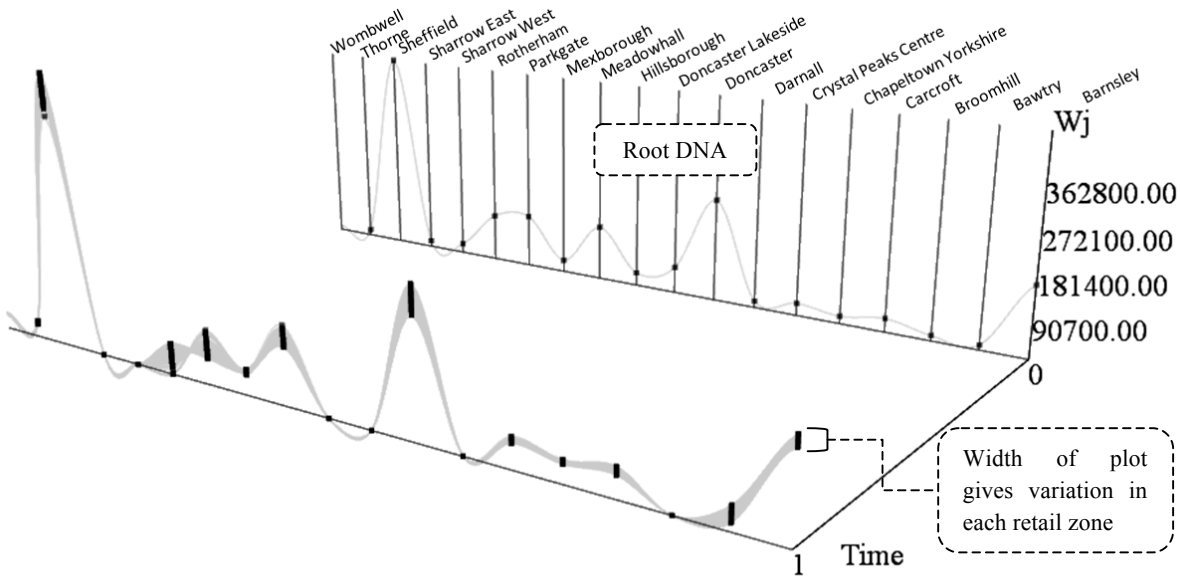

Figure 5. Parallel coordinates plot of a possibility cone for South Yorkshire

The figure shows that the output from all 300 model runs is fairly similar, however the range of values for each centre varies and does not appear directly correlated with size as one might expect given the set up of the model. The greatest variation appears at two retail zones: Rotherham town centre and an out of town shopping centre called Parkgate Retail World which are situated within about 3km of one another. In July 2009 Rotherham was in decline with a third of its shops closed (Addley 2009) due to recession and competition with out of town shopping centres like Parkgate Retail World. The plot here illustrates the interdependency between the size of the two centres, with Parkgate often swallowing up much of the retail business in Rotherham.

It is clear that the stochastic variation in the model is causing path dependence to have an impact on the system and suggests that the system is alternating between the basins of attraction of a number of different solution sets. The visualisation system is helpful for interpreting the output of the system and gives us a rough idea of which basins of attraction are within the possibility-cone. The figures shown here are screen grabs from a visual analytics system which allows the user to move through the results, zoom into areas of interest and rotate the view. This capability plays an important part in making the results accessible.

4 Tackling a hypothetical planning application

We now add a additional layer of complexity by adding variation to the exogenous DNA. We can use this to represent both deliberate planned changes and unplanned events which may or may not occur. Within the constraints of the retail model we are using we can represent: fuel price fluctuations (affecting the β parameter), migration (affecting P_i) and recession and expansion of the economy (potentially affecting both e_i and β). Planned events we can model include: rail and bus network construction and road construction (causing an adjustment in

particular c_{ij} and / or m_j), retail zone construction (increasing a particular W_j), housing construction (affecting P_i if we assume the houses are occupied) and changes in taxation (affecting e_i).

We use the South Yorkshire region as our test case again and for simplicity's sake consider one planned event and one unplanned event. We take the planned event as the construction of Sheffield's Sevenstone development - planned for partial completion in 2011 which will add 79,000 m^2 of retail to the city. The unplanned event will be petrol price fluctuations. Detail of these events is given in Table 2.

Event name	Occurence	Alternative parameter values
Sevenstone construction	Time step 1	$W_{Sheffield}$ unmodified
		$W_{Sheffield}$ + 79,000
Petrol price fluctuation	Every time step	β unmodified
		β + 0.2
		β – 0.2

Table 2. Details of possible future events represented in the possibility cone

The planned cone of development will contain three time steps with the construction of the Sevenstone development on the second time step. We calculate all permutations of these events to generate multiple exogenous DNA variations for each time step in the cone (see Figure 6).

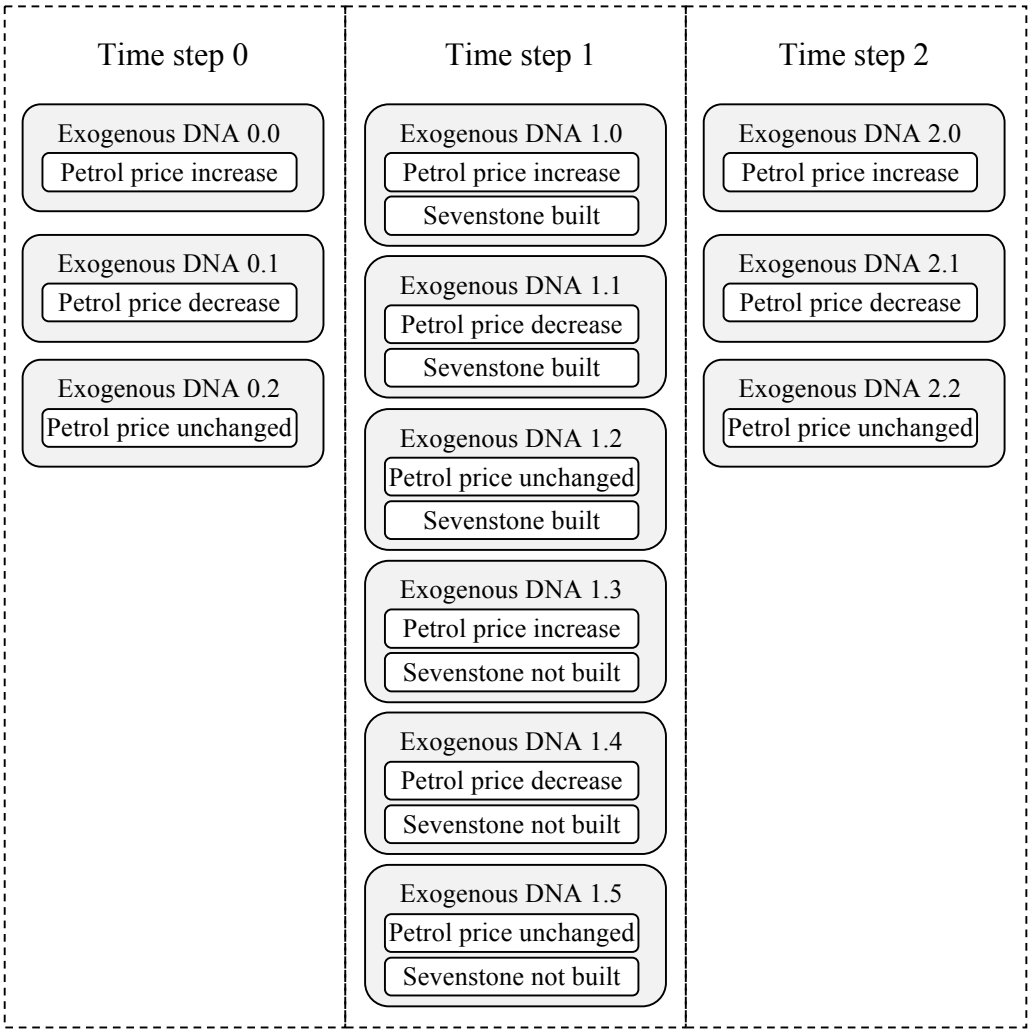

Figure 6. The exogenous DNA variation at each time step

Following the methodology given in Section 2 we set s equal to 3 giving a tree constructed from 1,630 model runs and representing 1,458 different development paths. Figure 7 shows the resulting cone of development. This cone of development shows greater variation across most centres and the possible development of Sheffield obviously contributes to this. The parallel coordinates plot helps to clearly differentiate discrete and continuous change: discrete change shows up a break in the plots (e.g. at Hillsborough) while continuous change shows up as a smoothly varying band. There is still a great deal of variation in both Rotherham and Parkgate and this appears to be in part because the resulting size of Rotherham is so uncertain. We can use a cone of development to explore the effect of modifying the system DNA - in effect urban "genetic medicine". If we want to make Rotherham more stable, one possible intervention might be to make it easier for consumers to reach the shops there by improving the public transport and offering free parking in the town centre. To represent this we reduce the m_j multiplier for Rotherham from 0.691 to 0.681 in all exogenous DNA sets. The resulting cone of development (Figure 8) actually shows a wider range of variation in Rotherham and Parkgate as well as some of the other nearby retail centres. Figure 9 shows the cone for Rotherham and Parkgate in detail before and after the intervention and clearly shows that the system is less stable after the intervention. The gaps in the plot indicate a

discontinuity which appears to represent a critical size for Rotherham below which it is not sustainable. This example illustrates the difficulty of intervening successfully in a nonlinear system such as this.

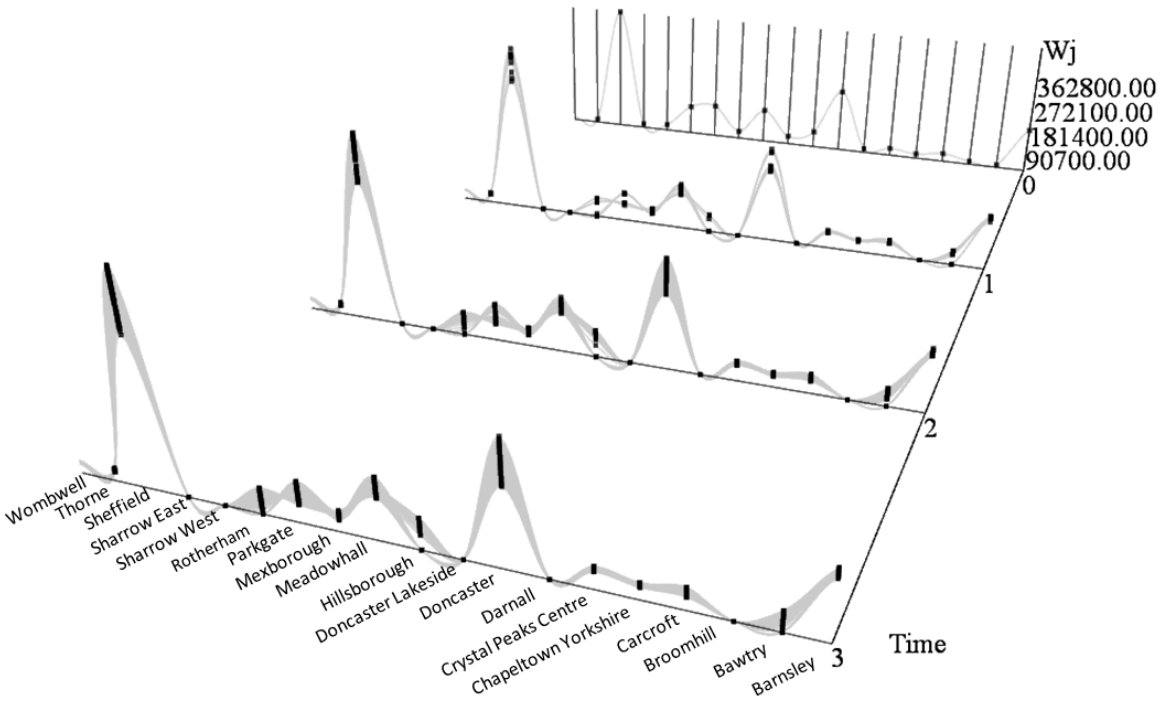

Figure 7. The cone of development for South Yorkshire taking into simple exogenous DNA variation

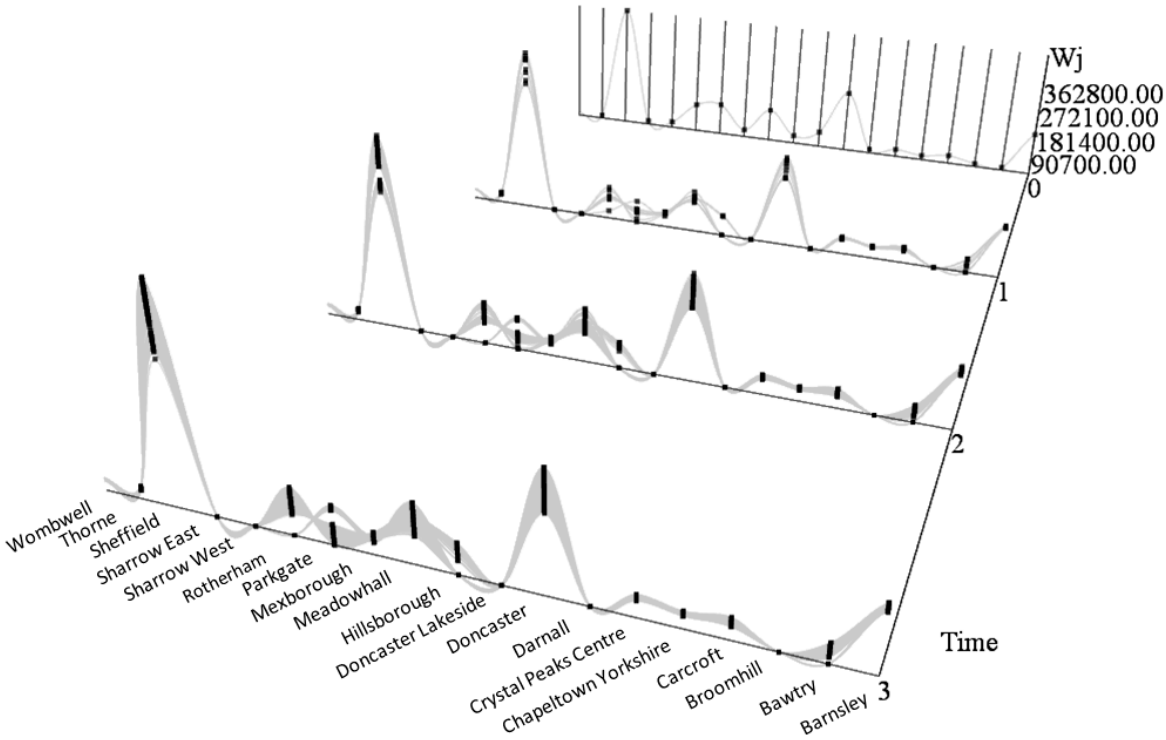

Figure 8. Possibility cone showing result of reduced travel cost into Rotherham

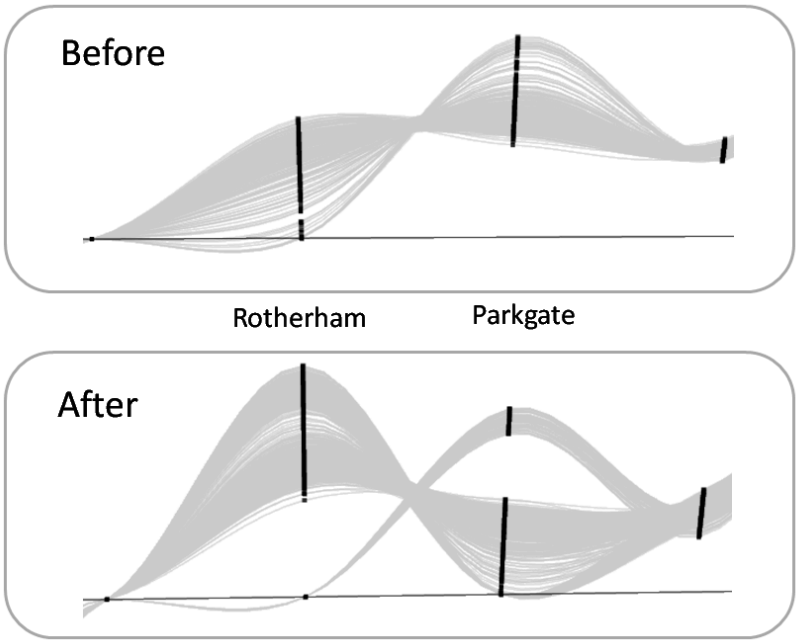

Figure 9. Detail of intervention to reduce travel cost into Rotherham

5 Conclusions

In this paper, we demonstrated a detailed methodology for constructing and visualising a multidimensional possibility-cone of development for an urban system. This is potentially a useful way of forecasting the future of nonlinear urban systems and of exploring stability and risk in planning. Here we have focused on retail in order to simplify our initial explorations; however to gain a full understanding of the possibility cone of urban development a more comprehensive model would be required. This presents a range of challenges due to the computing power required to build the cone. The output is also likely to be more complex than the system demonstrated here making visualisation more challenging. The idea that a system moves between the basin of attraction of a number of solutions is a useful way of thinking about path dependence and future research might look for ways to map out and visualise these multidimensional structures.

References

Addley, E. (2009), 'Empty, Unlet and Unloved: The New British High Street', *Guardian*, 25 July 2009, p. 17.

Arthur, W. B. (1988), 'Urban Systems and Historical Path Dependence', in *Cities and Their Vital Systems: Infrastructure, Past, Present and Future*, eds J. H. Ausubel & R. Herman, National Academy Press, Washington, D.C., pp. 85-97.

Box, G. E. P. and Muller, M. E. (1958), 'A Note on the Generation of Random Normal Deviates', *The Annals of Mathematical Statistics*, pp. 610-611.

d'Ocagne, M. (1885), *Coordonnées Parallèles Et Axiales-Méthode De Transformation Géométrique Et Procédé Nouveau De Calcul Graphique Déduit De La Considération Des Coordonnées Parallèles*, Gauthier-Villars, Paris.

Harris, B. and Wilson, A. G. (1978), 'Equilibrium Values and Dynamics of Attractiveness Terms in Production-Constrained Spatial-Interaction Models', *Environment and Planning A*, *vol. 10*, pp. 371-388.

Inselberg, A. (1985), 'The Plane with Parallel Coordinates', *The Visual Computer*, *vol. 1*, no. 2, pp. 69-91.

Matsumoto, M. and Nishimura, T. (1998), 'Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator', *ACM Transactions on Modeling and Computer Simulation*, *vol. 8*, no. 1, pp. 3-30.

Moustafa, R. E. and Wegman, E. J. (2002), *On Some Generalization of Parallel Coordinate Plots*, George Mason University, Fairfax, VA, USA.

Wilson, A. G. (forthcoming), 'Urban and Regional Dynamics from the Global to the Local: Hierarchies, 'DNA' and 'Genetic Planning'', *Environment and Planning, B*.

Appendix 1: details of the random number generator

We use the Mersenne Twister algorithm to generate uniformly distributed random numbers (Matsumoto and Nishimura 1998). These are then transformed into random numbers from a Gaussian distribution using the polar form of the Box-Mueller algorithm (Box and Muller 1958).