UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

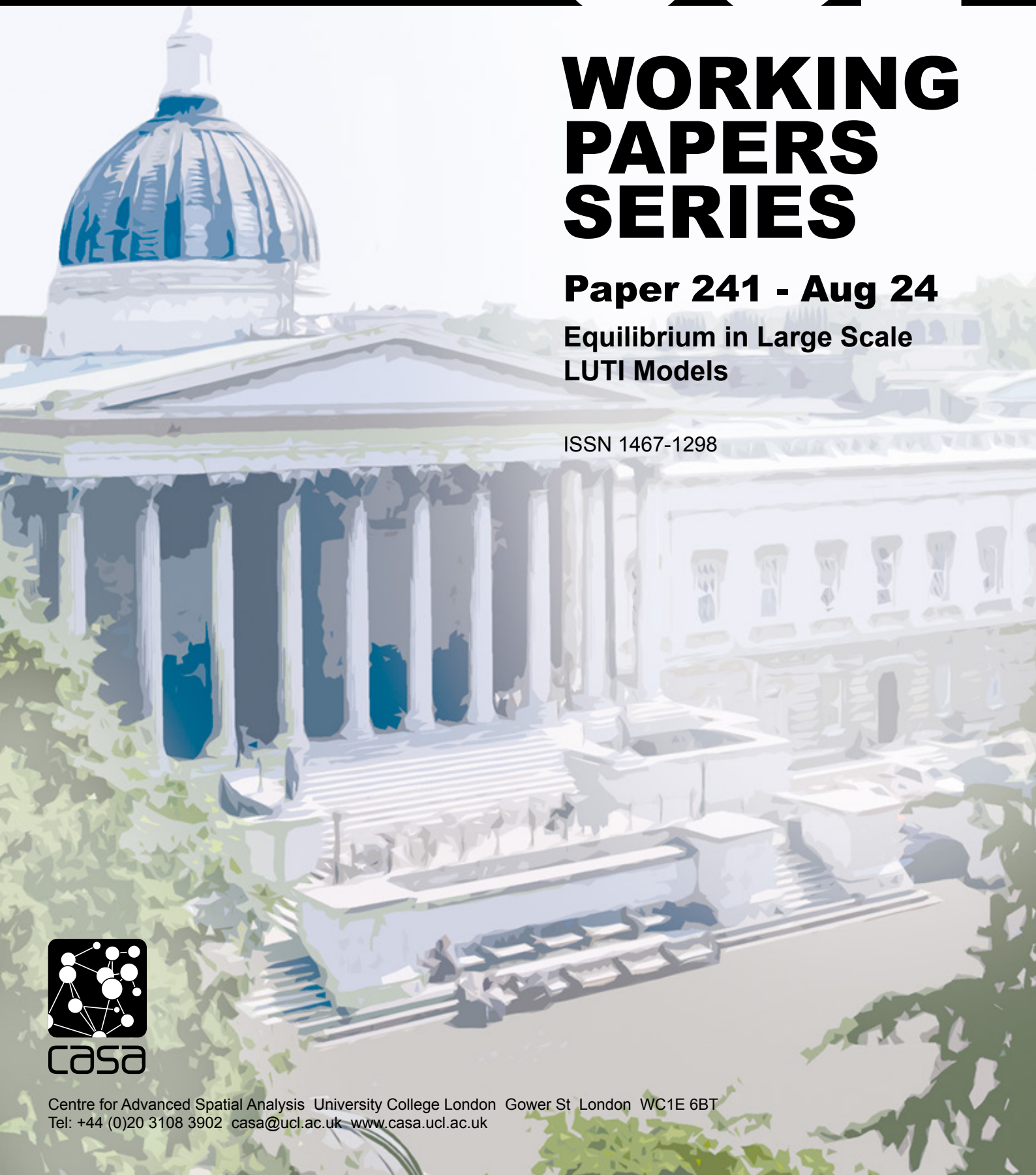
UCL

WORKING PAPERS SERIES

Paper 241 - Aug 24

Equilibrium in Large Scale LUTI Models

ISSN 1467-1298



casa

Centre for Advanced Spatial Analysis University College London Gower St London WC1E 6BT
Tel: +44 (0)20 3108 3902 casa@ucl.ac.uk www.casa.ucl.ac.uk

Equilibrium in Large Scale LUTI Models

Michael Batty, Willow Liu, and Jens Kandt

Centre for Advanced Spatial Analysis,
University College London, 90 Tottenham Court Road,
London W1T 4TJ, UK

m.batty@ucl.ac.uk, zhixuan.liu.21@ucl.ac.uk ; j.kandt@ucl.ac.uk
X@jmichaelbatty https://orcid.org/0000-0002-9931-1305

12 August 2024

Abstract

Most computable urban models, designed to replicate the locational distribution of socio-economic activities, are based on aggregate patterns of employment, population and spatial interaction. Such structures often called LUTI models can now be built very rapidly at scale so that they can be used to simulate the impact of large-scale economic and demographic change on the way a city or system of cities is able to embrace locational economies of agglomeration. We first explore the way such models can be articulated as simultaneous interactions of employment and population, and we demonstrate how they can be solved iteratively to mirror a system in equilibrium. We apply the model to the CAMKOX Corridor (the Cambridge-Milton-Keynes-Oxford Arc) and empirically investigate the model's properties, indicating how the model can be used to predict the agglomeration economies of changes in that region, and then illustrating how an array of possible solutions can be generated as different varieties of digital twin. As the aggregate LUTI models we develop can now be run hundreds, if not thousands, of times, we illustrate how sensitivity testing, scenario generation and changes in locational behaviour can now be tested routinely, thus developing simulations that bound define the wider solution space of different model types.

Key Words

Land Use Transport Interaction (LUTI) Models, Digital Twins, Convergence to Equilibrium, Iterative Solutions, the CAMKOX Corridor, Generating Scenarios,

Acknowledgements

Funding for this project has come from the ESRC's UCL,
Bloomsbury and East London Doctoral Training Partnership (UBEL DTP).

Equilibrium, Statics and Dynamics in Urban Models

If we examine the city in history, it is clear that its physical form is relatively stable from generation to generation (Mumford, 1963). Cities usually grow around some central location that represents a confluence of forces defining a marketplace or commercial cluster, now generally called a central business district which represents the main sources and sinks for the economic energy that drives urban growth. Cities usually reach out from their cores to encapsulate as much space as possible, but their activities are constrained by the abilities of their residents and workers to interact at a distance. The forces which push and pull activities in successive ways lead to compaction and urban sprawl on the one hand and patterns of radial growth on the other and together, these define the circular, tree-like morphology that we see in most city plans. The general assumption is that cities preserve this morphology and that new activities absorbed into their fabric as they grow and evolve, maintain an equilibrium in which activities such as jobs, housing, and transport, balance each other out.

There are many ways of representing cities and a simple but a convenient mode is to examine their patterns of employment at work and their associated population by residence. At any point in time, we assume these patterns are in equilibrium, if only because movement or interaction between their various sectors must be conserved. However, to explain the patterns of equilibrium that emerge, we invariably have to break into the various feedback loops that define how any set of activities depends on any other. This approach to defining a city tends to destroy its equilibrium, and many researchers arbitrarily adopt such a break, thus privileging one set of activities over another. The consequence of this is that the observed equilibrium is often ignored, and it is hard, if not impossible, to put the equilibrium back together again from the models that are produced.

The first attempts at building simulation models assumed that cities manifested such an equilibrium whilst exhibiting a dynamic that continually changes in time, thus representing an evolution of this equilibrium. Models that embraced both statics and dynamics have been rare because data pertaining to processes of urban change has been very difficult to collect, at least until the recent development of the online world and big data. Models that articulated the city system as a comparative static structure have thus become the norm. It is easy to define this bias as anti-dynamics, but this is only a superficial reaction to the lack of focus on time. So far there has been no synthesis of different approaches from comparative statics that dominated the first computer models of cities to later forms of evolutionary dynamics that introduced many new ideas about change in terms of temporal discontinuities associated with chaos and bifurcation. Moreover, the notion of urban dynamics as being intrinsic to complexity science now suggests that statics and dynamics are part of a much wider paradigm often referred to as involving systems that are ‘far-from-equilibrium’ (Batty, 2005).

The earliest urban models did in fact broach the notion of simulating a city in equilibrium. In the context of a model where employment E is dependent on population P and population P dependent on employment E, the generic model is to ensure that an integrated structure is built where $P = f(E)$ and $E = g(P)$ whose reduced form is $P = f\big(g(P)\big)$ and $E = g(f(E))$. For example, Lowry (1964) in his *Model of Metropolis* suggested that such an equilibrium structure could be solved iteratively. At the same time, a variety of econometric models, in particular the EMPIRIC model by Hill et al. (1965), were initially proposed and solved using linear simultaneous equations whose solution was exact and independent of its starting values. Many other partial solution procedures have been suggested since then, but a detailed inventory of techniques has not been developed and most of what have been explored are theoretical schemes, based on a linear chaining of models with no direct closure. Some attempts have been

made to map these iterative processes onto pseudo-dynamic forms, but these have not been applied empirically and remain idiosyncratic to the main field (Batty, 1984).

As soon as the first comparative static models were developed, there were attempts at making them dynamic and this was achieved by simply beginning at a cross-section in time, simulating the entire city at that point and then modelling subsequent increments or decrements of activities in small time internals usually from one to five or ten years. These models avoided problems of equilibrium by assuming that each time period was its own equilibrium. Since these first models, there has been little change in structure with respect to making these models fully dynamic; for example, the generic state-of-the-art LUTI (Land Use Transport Interaction) model currently developed in the UK for London (the so-called LonLUTI model), operates in this incremental fashion with only feed forward between the temporal intervals and no feedback to ensure a traditional equilibrium. In so far as these models embrace a wider dynamics, they have been linked to several other models that provide an integrated modelling environment, as in the suite called MOTION https://tfl.gov.uk/corporate/publications-and-reports/strategic-transport-and-land-use-models.

In fact, dynamic models have been linked to urban planning in a very different way, by developing different types of models which do not emphasise equilibrium at all. Other features that have dominated the development of computer models focus on changing the spatial as well as the temporal scale. Cellular automata (CA) models that encapsulate land development have emerged, and these contain a simple dynamics that relates to how the spatial units – the cells – change with respect to interactions with other cells. No equilibrium is assumed as growth (or decline) measured by the state of the cells – their attributes – can continue indefinitely. If we further disaggregate the cells to individual point locations and consider that each might contain an agent, then we can develop agent-based models (ABM) where agent or individual behaviours interact across space and time with respect to their spatial decision-making (Batty, 2005). In the case of land development, an agent might be a household which requires a housing structure, or a landlord, a developer, and so on. We will not discuss these further here but note that these provide key methods for building dynamic models where households, travellers, vehicles and other physical objects all act as agents in various kinds of microsimulation (MS). There are other computable structures that pertain to urban dynamics and these build on chaos and bifurcation theory, elements of which can be embedded into the different urban forecasting models we have just indicated. In fact, as the field of urban modelling has matured, more and more features from different modelling types – LUTI, CA, ABM, MS and so on – are being merged into a rich but rather chaotic landscape of ideas that urgently requires clarification.

Although we will be largely exploring the equilibrium properties of a generic type of LUTI model in this paper, we will encounter two other themes that are beginning to dominate models of cities. First there is the computational environment that is continuing to change very rapidly. It is now possible to build and continually explore quite large-scale urban models on the desktop, generating immediate outcomes from such simulations and thus providing an environment where the model builder is receiving instant feedback from his or her own application of the model. In this sense, large models that until quite recently took days to completely explore, can now be run instantly with thousands of runs being possible during the working day. In fact, large-scale models such as QUANT which represents the British urban system by over 8436 zones and over 71 million interactions can now be run in seconds using the latest computational architectures based on GPUs (Batty and Milton, 2023). The models we have developed here for the Oxford-Milton-Keynes-Cambridge Arc based on some 400 zones runs ‘instantly’ and this means we can generate many variants that hitherto were impossible. This changes the entire way we are able to build models, and it opens up a

perspective on how we develop both theory and computation in the context of the emerging science of cities.

The fact that we can now build more than one model is fast becoming the dominant style of application (Page, 2019). We can now elaborate and detail LUTI models in many different ways giving real focus to the idea that we are able to build 'digital twins', another theme that weaves its way through our exploration of models here. Digital twins are only possible if we are able to construct them in computational environments that allow us to explore and control many variants that enrich our understanding of the model and the real system, learning ever more about the model and its physical twin – the real system – to which it is being applied (Caldarelli et al., 2023). This also raises the question of not only constructing more than one model of the same system but also generating variants of the real system, reflecting the way different elements of a city, for example, can be fashioned in terms of different models. To date, the environment in which urban models continue to be developed has barely changed over the last half century apart from the fact that the models involved run faster but the organisational structures which are used to make them applicable in practice, is still based on the assumption that large models are difficult to build, take time to apply, require scientific skills that are in short supply, and have to take account of the inertia of current organisational and decision-making structures. Much of this is about to change.

In the next section of this paper, we will outline a generic land use transportation interaction (LUTI) model that we will elaborate with respect to its equilibrium properties in the third section. We will then introduce the application which here we base on data for the **CA**mbridge-**M**ilton-**K**eyes **OX**ford (CAMCOX) arc of urban development which is widely regarded as one of very few regions of the UK with the most potential for rapid economic growth. We then explore the fit of a pilot version of the LUTI model to this region and then illustrate how the full model can be solved in terms of its stable equilibrium. We develop different equilibria from the basic model and show that as we iterate the solution towards a long-term equilibrium, the performance of the model degrades substantially. This suggests that if we are to use such models with respect to their equilibrium solutions, we need to choose only modest iteration and thus accept solutions that are not stable, as indeed are most empirical applications of LUTI models to date which are not formulated in equilibrium terms.

A Generic Land Use Transportation Interaction (LUTI) Model

The generic LUTI model was first developed by linking together the location of activities or land uses, typically employment and population sectors but also extending to retail, leisure, educational and health facilities, using spatial interaction models that simulate the flows of people, materials, and income between these sectors. The simplest models generated population in residential locations from locations where that population worked using models relating the size or attraction of the relevant locations to the cost or travel time associated with the distance between those locations. The typical interaction model used was initially a gravitational model which was often elaborated into various utility-based discrete choice variants, sometimes subject to income and related economic constraints. Here we will first subdivide the spatial system into a set of zones $\Omega = i, j, k, \ldots$ noting that we begin with employment E_i from which we generate P_j. In turn, we then generate employment E_k which establishes the loop from which we continue to compute the equilibrium values of working population and employment. The essence of the model are the equilibrium conditions that we

noted in the first section above where we presented this in reduced form terms as $P_j = f\left(g\left(P_j\right)\right)$ and $E_k = g(f(E_k))$.

We will define two sectors – first employment E_i at workplaces i which generates population in residential locations P_j and second population at residential locations P_j which generates demand for employment at workplace locations E_k. In this variant of the model, we will disaggregate the employment into L different types which we define as E_i^ℓ, $\ell = 1, 2, \ldots, L$ but we do not disaggregate the population into the same or different types because in the model here, we do not have these same employment attributes for the population. The residential location model which simulates the flow of workers from workplaces to residences predicts T_{ij}^ℓ and can be stated as

$$T_{ij}^\ell = A_i E_i^\ell F_j \exp\left(-\beta^\ell c_{ij}\right) = E_i^\ell \frac{F_j \exp\left(-\beta^\ell c_{ij}\right)}{\sum_z F_j \exp\left(-\beta^\ell c_{iz}\right)} \quad , \tag{1}$$

where F_j is a measure of residential attraction for workers, c_{ij} is the travel cost, time or distance (here distance) between i and j, and β^ℓ is the parameter on distance which varies for each type of employment ℓ. In spatial interaction theory, this model is referred to as singly-constrained (Wilson, 1971) where the model always respects the origin constraint which we define as

$$E_i^\ell = \sum_j T_{ij}^\ell \quad , \tag{2}$$

while the model predicts the working population in residential zones as

$$P_j = \sum_{j\ell} T_{ij}^\ell \quad . \tag{3}$$

Note that this model structure in equations (1) to (3) can be replicated as a module for any number of sectors where it is possible to define interactions between two sets of locations, origins or destinations, sources or sinks although in practice, the number of sectors tends to reflect the number of distinguishable land uses or activities which in aggregate terms, is rarely more than 10, often not more than 5.

The reverse flow from residential locations of the population to employment at workplaces sets up a different pattern of demand E_{jk}^ℓ which is not symmetric with the flow pattern T_{ij}^ℓ. This is the pattern of demand for products workers make at their workplace which in turn is what the population consumes. The flow from population back to employment is thus from zones j to k and the equivalent model is

$$E_{jk}^\ell = B_j P_j F_k^\ell \exp\left(-\gamma^\ell c_{jk}\right) = P_j \frac{F_k^\ell \exp\left(-\gamma^\ell c_{jk}\right)}{\sum_z F_z^\ell \exp\left(-\gamma^\ell c_{zk}\right)} \quad , \tag{4}$$

where F_k^ℓ is a measure of attraction of the employment location for industry type ℓ, c_{jk} is the travel cost, time or distance, and γ is the parameter on distance which varies for each type of employment ℓ. The model is subject to a constraint on the new origins of population at j and it predicts employment demand at the new destinations k which are the previous origins i. These constraints are derived as

$$P_j = \sum_{k\ell} E_{jk}^\ell \quad , \quad \text{and} \tag{5}$$

$$E_k^\ell = \sum_j E_{jk}^\ell \quad . \tag{6}$$

Although these two models interface with one another through population and while also noting that they could be stated the other way around with the interface being though employment,

there is nothing in equations (1) to (6) to ensure that the employment input and then predicted from each sub-model are the same.

In short, for equilibrium, equation (2) must give the same answer as (6) and equation (3) the same as (5). In fact, in general they do not being reflected in the inequalities

$$\left.\begin{array}{l} E_i^\ell = \sum_j T_{ij}^\ell \neq E_k^\ell = \sum_j E_{jk}^\ell \\ P_j = \sum_{j\ell} T_{ij}^\ell \neq P_j = \sum_{k\ell} E_{jk}^\ell \end{array}\right\} \quad , \quad i = k \quad . \tag{7}$$

To achieve this balance, we need to couple the two models explicitly, and to do this, we define the relevant variables using an iteration index τ which can be interpreted as an index of computer time. Beginning with the residential location model in equation (1), we now write this as

$$T_{ij}^\ell(\tau) = A_i E_i^\ell(\tau)\, F_j \exp\left(-\beta^\ell c_{ij}\right) = E_i^\ell(\tau) \frac{F_j \exp\left(-\beta^\ell c_{ij}\right)}{\sum_z F_j \exp\left(-\beta^\ell c_{iz}\right)} \quad , \tag{8}$$

where we begin with $\tau = 1$. Then from the predicted working population $P_j(\tau) = \sum_{j\ell} T_{ij}^\ell(\tau)$, we compute the new demand for employment as

$$E_{jk}^\ell(\tau+1) = B_j^\ell P_j(\tau) F_k^\ell \exp\left(-\gamma^\ell c_{jk}\right) = P_j(\tau) \frac{F_k^\ell \exp\left(-\gamma^\ell c_{jk}\right)}{\sum_{\ell z} F_z^\ell \exp\left(-\gamma^\ell c_{jz}\right)} \quad , \tag{9}$$

while the new total employment is

$$E_k^\ell(\tau+1) = \sum_j E_{jk}^\ell\,(\tau+1) \quad . \tag{10}$$

We now substitute equation (10) into (8) and continue the iteration until a convergence threshold based on one of many measures we define below such as, say, a root mean squared error or difference

$$\theta(\tau) = \left[\sum_j \left(P_j(\tau+1) - P_j(\tau)\right)^2 / N\right]^{1/2} < \varepsilon \quad , \tag{11}$$

is reached.

We will not prove that this model converges to a stable solution in terms of its predicted variables, but it is intuitively obvious this must occur due to the manner in which the endogenous variables $\{T_{ij}^\ell(\tau)\}$ and $\{E_{jk}^\ell(\tau+1)\}$ are constructed. The key issue of course is how close the computed equilibrium is to what is actually observed. We can in fact write the complete model in its most compact form as a location rather than an interaction model and this serves to highlight the critical coupling and feedbacks involved. This is then

$$\left.\begin{array}{l} P_j(\tau) = \sum_{i\ell} A_i^\ell E_i^\ell(\tau) F_j \exp\left(-\beta^\ell c_{ij}\right) \\ E_k(\tau+1) = \sum_{j\ell} B_j P_j(\tau) F_k^\ell \exp\left(-\gamma^\ell c_{jk}\right) \end{array}\right\} \quad . \tag{12}$$

There are several additional features that can be added to these kinds of LUTI model. First the attractor variables $\{F_j\}$ and $\{F_k^\ell\}$ can reflect economies of scale by parameterising their form as (F_j^θ) and $(F_j^\ell)^\vartheta$ where if θ and ϑ are greater than 1, this incorporates agglomeration economies that mean that the larger the floorspace in the location in question, the more than proportionate is the attraction to locate there. In the same way, it can be argued that the negative exponential functions in equations (12) reflects diseconomies of scale. In fact, the structure of the coupled model is such that it is designed to incorporate both kinds of scaling effect, which are critical to how the model generates urban growth in employment and population.

There are several variants of the equilibrium LUTI model that we will define and test in later sections, once we have presented the empirical application. All depend on coupling the residential location model in equations (1) and (2) to the employment demand model in equations (3) and (4). The simplest structure is to run these two models separately, uncoupled, while the most basic coupling is to simply chain the two models in one step starting from equations (1) and (2) which drive the model in equations (3) and (4). No equilibrium is attempted with this coupling. We could also reverse this sequence, beginning with equations (3) and (4) and then moving to equations (1) and (2), and we could iterate these models starting in this order and continuing the sequence (3) and (4) to (1) and (2) and then back to (3) and (4) until full convergence is reached. We will explore all these variants in the sequel but at this point we need to introduce the empirical application as there are several features from the case study that exploit different elements within the model and these first need identifying.

An Initial Application: A Pilot Model for the CAMKOX Corridor

The region we use for testing our equilibrium pilot model stretches from Oxford, which is north-west of Greater London, then north-east to the new town of Milton Keynes, and continuing in this direction to Bedford and Cambridge. The region contains about 1.5 million employees (see https://nic.org.uk/app/uploads/Partnering-for-Prosperty.pdf) which scales to a population of about 3.3 million. In size, it is about 90 miles east-west and 50 miles north-south and it is widely considered to be one of the most prosperous in the UK, or at least one with the greatest potential for economic growth (ITRC, 2020; Lomax, Smith, Archer, Ford and Virgo, 2022). We have defined the basic spatial units of the model from the Population Census's standard geography called middle layer super-output areas (MSOAs) which gives 397 zones which we show in Figure 1(a). From the Office for National Statistics, we are able to define four different categories of employment and their associated floorspace which we show in the heatmap alongside the spatial subdivision of the region in Figure 1(b). The physical morphology of the region is essentially a landscape of relatively small towns but in the last 50 years, it has become highly urbanised, and the key settlements are beginning to merge into one another particularly in the central belt of the region whose orientation is north-south along the M1 and A1(M) motorways and the West Coast Mainline (rail). The east-west orientation of the region has only begun to gain significance in the last decade with the proposal that a new East-West Rail line be restored and rebuilt between Oxford and Cambridge.

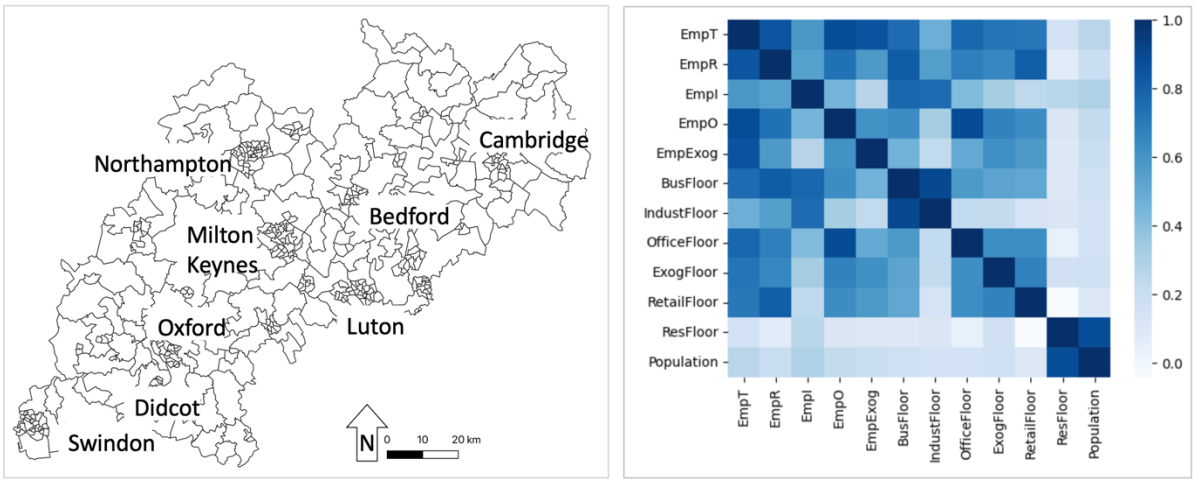


Figure 1: The Basic Data (a) The CAMKOX Corridor (b) Correlation of the Model Variables

The distribution of employment and population is shown in Figures 2(a) and (b) and it is quite clear that employment is much more concentrated and polarised than population. The average employment in each zone is similar to the population (3865 compared to 3793) but the maximum employment is 28928 compared to a maximum population of 8123. If we plot the rank size distributions of employment ($E_{r(i)}$ compared to its rank order r) and population ($P_{r(j)}$ compared to r), the differences are even more marked with employment being much closer to the classic power law compared to population which has a much flatter profile. These graphs are shown first in absolute terms in Figure 3(a) where the correlation between population and rank is -0.952 and employment and rank -0.749. When these are log-linearised, the correlations rise to -0.934 and -0.939 for population and employment and their logarithmic transforms in Figure 3(b). We also indicate the linearised rank size relations $\log P_{r(j)} = 11.411 - 0.686 \log r$ ($R^2 = 0.884$) and $\log E_{r(i)} = 9.349 - 0.228 \log r$ ($R^2 = 0.873$) in Figure 3(b) where it is clear that the region has a very well-defined polycentric, fractal-like structure that mirrors its evolution and provides a sense of how the region might continue its future growth (Batty, 2005).

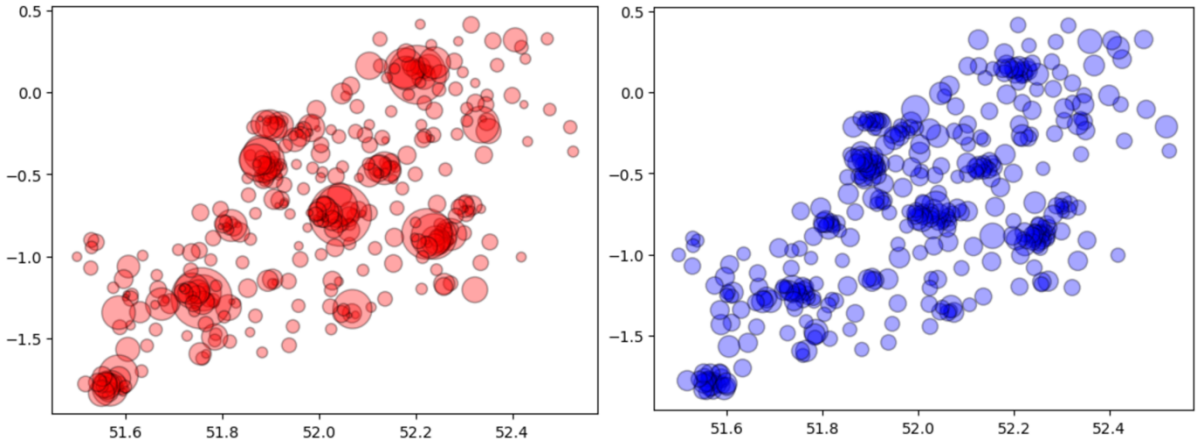


Figure 2: Employment and Population (a) Total Employment (b) Population Volumes as Overlapping Proportional Circles, Indicative of Intensity

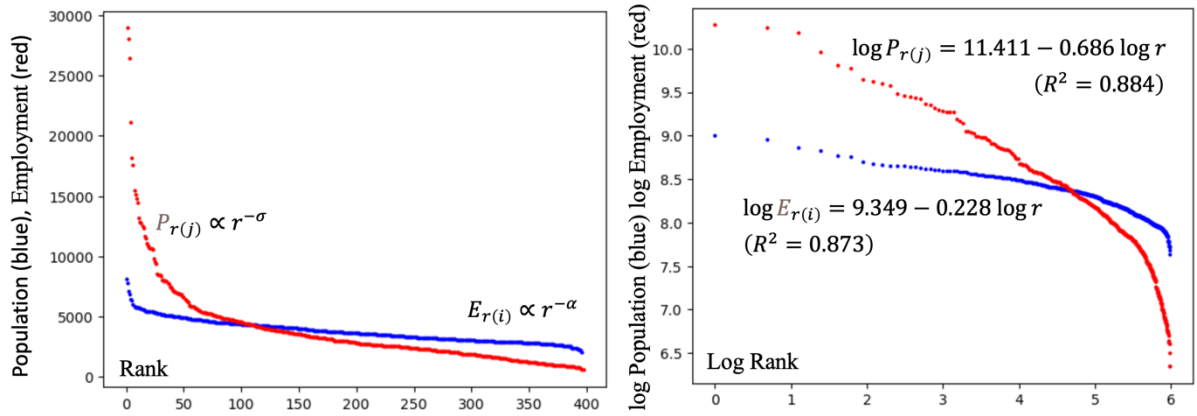


Figure 3: Employment and Population Rank Size Functions (a) As Power Laws (b) As Log Linear Functions

The employment distribution is composed of four different activity types as shown in the heat map in Figure 1(b). The four types are: Retail Employment (**EmpR**), Industrial Employment (**EmpI**), Office Employment (**EmpO**), and Exogenous Employment (**EmpX**) which are matched against the relevant floorspaces variables: Business (Retail) Floorspace (**BusF**), Industrial Floorspace (**IndF**), Office Floorspace (**OffF**), and Exogenous Floorspace (**EXF**). We show the four types in Figures 4(a) to 4(d) where their locational differences reflect different physical and spatial requirements. The correlations between these four variables and others that we include in the heat map in Figure 1(b) are not as high as we initially assumed. Population and its related residential floorspace are very highly correlated but their correlations with other variables (that in fact are mainly employment-related) are quite low. Again, this reinforces the fact that in this region, the correlations between employment and population are complex in terms of their size although the pattern of all these variables reflects the overall morphology of the region that we commented on above.

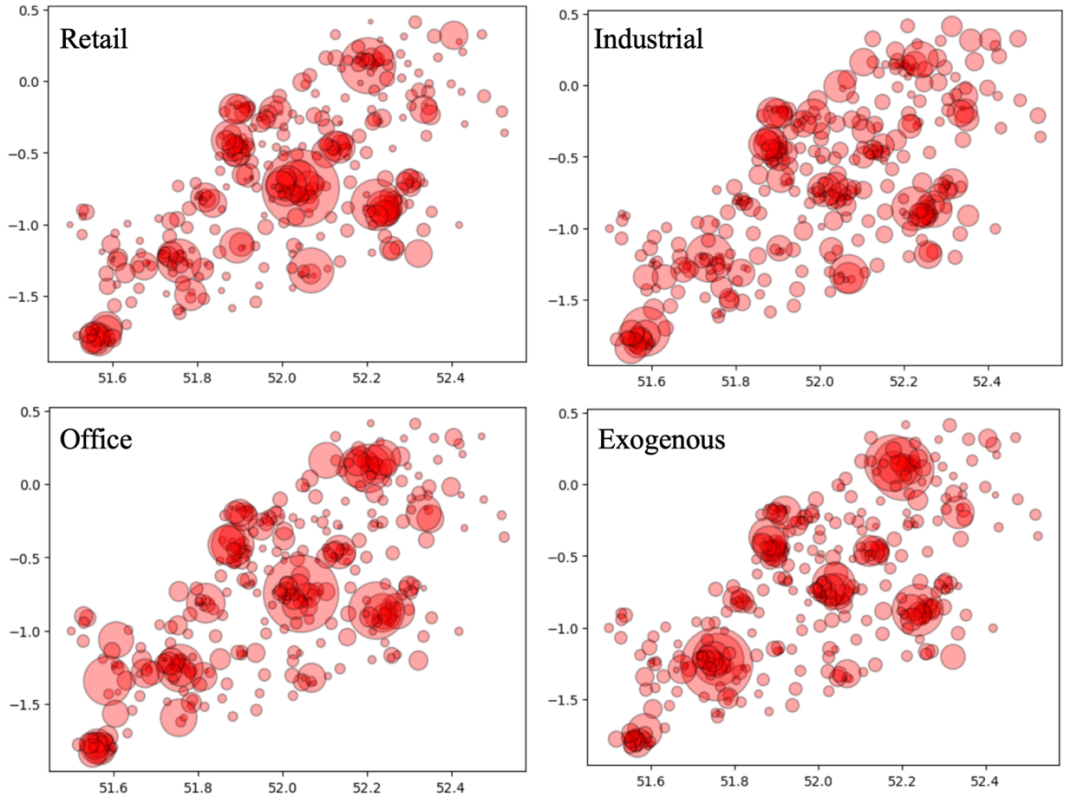


Figure 4: The Distribution of Employment Types

Our first test of the model is based on one iteration of the coupled residential with the employment location model as we will illustrate for the disaggregated version of the model in equations (1) to (4). We refer to this as the pilot model where essentially the population $P_j(1)$ is first generated from the employment $E_i(1) = \sum_\ell E_i^\ell\,(1)$ and then the next iteration of employment $E_k(2)$ is generated from $P_j(1)$. One cycle of the model is shown in Figure 5, and the predictions of total employment and population after one iteration are shown in Figures 6(a) and 6(b). These need to be compared to their observed equivalents in Figures 2(a) and 2(b) where it is immediately clear that the first cycle of employment $E_k(2)$ differs substantially from the observed employment $E_i(1)$ in comparison to the lesser differences between the predicted and observed populations $P_j(1)$ and $P_j(0)$.

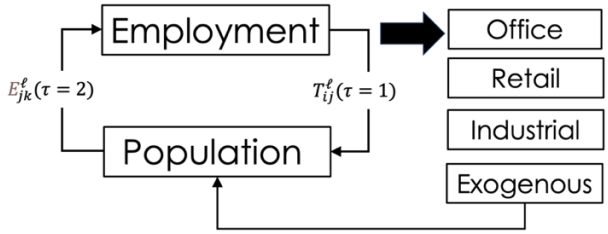


Figure 5: The First Cycle on the Path to Equilibrium

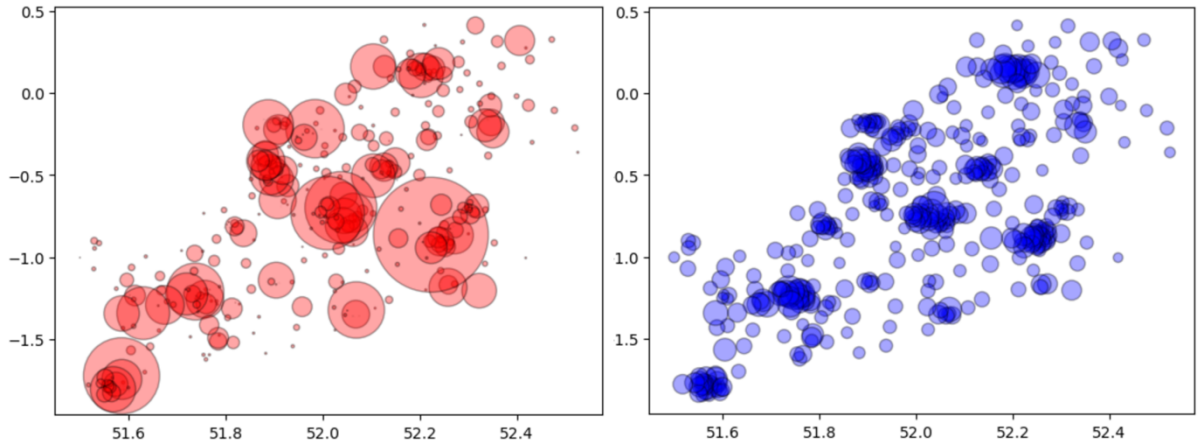


Figure 6: (a) Predicted Employment and (b) Predicted Population After One Cycle on the Path to Equilibrium $\tau = 1, 2$

We have not calibrated this version of the model but simply set the parameters β^{ℓ} and γ^{ℓ} equal to the inverse of the observed means $\beta^{\ell} = 1/\bar{T}^{\ell}$ and $\gamma^{\ell} = 1/\bar{E}^{\ell}$ which are usually good estimates of their orders of magnitude (Hyman, 1969). The correlations between observed and predicted total employment and population are 0.658 (R^2 =0.433) and 0.575 (R^2 =0.331) respectively where the employment fit is a little better the population. In fact, this is consistent with the fact that the population is much more spread out across locations whereas observed employment is more polarised. Our last analysis of this phenomenon from the pilot model relates to the fact that the generic model generates strong agglomeration economies as we will now demonstrate.

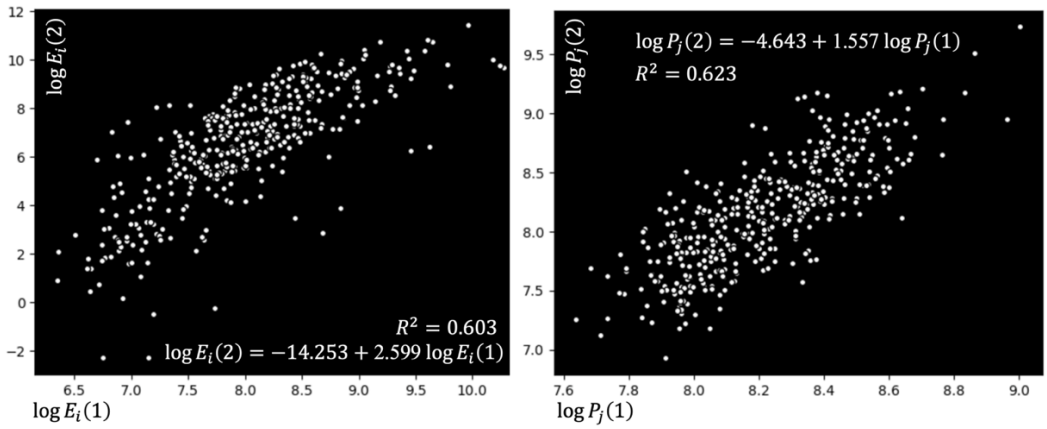


Figure 7: Agglomeration from Predicted Employment and Population at the First Iteration

If we examine the predicted employment $E_i(\tau+1)$ and population $P_j(\tau+1)$ with respect to their observed values, there is a clear relationship which implies that the model generates more than proportionate activity with respect to its inputs. This is strongest for the relationship between predicted employment $E_i(\tau+1)$ and the observed employment $E_i(\tau)$ where the relation follows the super-linear allometric form. This is associated with agglomeration economies and has the form $E_i(\tau+1) = \varphi E_i(\tau)^{\phi}$ where we assume that $\phi > 1$ if there are any agglomeration economies associated with the generation of activities by the model. The idea of agglomeration in this form dates back to Marshall (1890) but it has been resurrected and formally linked to allometry by Bettencourt et al. (2007) (see D'Acci, 2025) and it is intrinsic to the way LUTI models handle attraction and deterrence in terms of spatial interaction.

If we compare the data from the observed baseline $E_i(\tau=1)$ with that which is predicted after the first iteration $E_i(\tau+1)$, then the log linear relation $\log E_i(\tau+1) = \log\varphi + \phi \log E_i(\tau=1)$ can be estimated as $\log\varphi = -14.253$ and ϕ =2.599 with a goodness of fit $R^2 = 0.603$. We show the scatter in Figure 7(a) and this suggests that as employment increases, it generates much more than proportionate additional increases in the same variable, more than 2.5 times for one additional unit of baseline employment. If we examine the same relation for population as in Figure 7(b) where $\log P_j(\tau+1) = \log\zeta + \eta \log P_j(\tau)$, then the parameters are $\log\zeta = -4.643$ and η =1.557 with the goodness of fit $R^2 = 0.623$. To an extent, the fact that employment generates bigger 'more than proportionate' returns compared to population is consistent with the fact that employment is more polarised, and that increases in productivity and growth itself are more likely to be directly associated with economic than demographic variables. These effects appear to be quite stable when the pilot model is run through more and more iterations to equilibrium but further exploration of these coupling, feedback and confounding effects is required to get a clear understanding of how we can simulate the processes that lead to a stable equilibrium. We broach this in the next section.

Transitions to Equilibrium

Our main model involves iterating the two sub-models to a convergence which satisfies the equilibrium equations in (12). It is easy to see how the two processes associated with residential then employment location convolute so that every location and interaction links with every other, thus generating a strongly non-linear structure that emerges as the model converges to an equilibrium. In some respects, this process of continued iteration is a kind of machine learning in that it is reminiscent of continual feedback and feedforwards with the goal of ensuring that the model converges on a stable solution from continued reinforcement of the predicted locational structure so far. In some respects, the process is more complicated than iterative learning in that the focus on outputs is continually changing with new features being derived at each iteration. In fact, it might be possible to extract explicit weights with respect to how these iterations converge on stable solutions, but this is a direction which we have not explored as yet.

We will explain this process visually with respect to the spatial pattern of employment and population activities below but first it is worth noting the nature of this convergence. We have defined 12 different measures which all show convergence to stable values as the computational index τ increases and we can list the statistics associated with these measures as follows. The first set of 6 measures pertain to the sum and average of the differences between predictions on each iteration. We divide these into two sets of differences: first between the total population at iteration $\tau+1$ and the previous iteration τ which get progressively smaller

as the iterations continue, and second, differences between iteration $\tau + 1$ and the observed values at $\tau = 1$ which converge to stable values. All these statistics are applied to total population, total employment and the four employment types although we will only focus on population here with respect to their convergence due to our need to keep the argument in this paper crisp. The statistics which look at differences between the variables at iterations $\tau + 1$ and τ get smaller while the differences between the predictions at $\tau + 1$ and the observed measures at $\tau = 1$ converge to stable values. These measures and their trajectories are fairly similar for each of the activities – population, employment and their four types – so we will only show those for population.

The first statistic $\Gamma(1)$ is the sum of the absolute differences averaged over all zones with second statistic $\Gamma(2)$ being these same differences normalised as percentage values. These are

$$\Gamma(1) = \sum_j \left(\left| P_j(\tau + 1) - P_j(\tau) \right| / N \right) \quad , \tag{13}$$

$$\Gamma(2) = \sum_j \left(\frac{\left| P_j(\tau+1) - P_j(\tau) \right|}{P_j(\tau)} \right) / N \quad . \tag{14}$$

The third and fourth statistics have the same form as equations (13) and (14) but with the differences being computed from the baseline observed values at $\tau = 1$. These statistics converge on stable values as the differences ultimately home in on these equilibria

$$\Gamma(3) = \sum_j \left(\left| P_j(\tau + 1) - P_j(1) \right| / N \right) \quad , \quad \text{and} \tag{15}$$

$$\Gamma(4) = \sum_j \left(\frac{\left| P_j(\tau+1) - P_j(1) \right|}{P_j(1)} \right) / N \quad . \tag{16}$$

The last two statistics are root mean square errors which provide an average error for a typical zone. Again we define these as differences between the populations at $\tau + 1$ and τ and then the same statistic but with the difference expressed as a percentage not an absolute value. These statistics are

$$\Gamma(5) = \left[\sum_j \left(P_j(\tau + 1) - P_j(\tau) \right)^2 / N \right]^{1/2} \quad \text{and} \tag{17}$$

$$\Gamma(6) = \left[\sum_j \left(P_j(\tau + 1) - P_j(1) \right)^2 / N \right]^{1/2} \quad , \tag{18}$$

and we will demonstrate these in the convergence trajectories which we graph below.

There are three other statistics that we have computed and we also use to illustrate the convergence and performance of the model. We can correlate the predicted variable – population – with its previous value and with its observed value using the following equations

$$\Gamma(7) = \rho[P_j(\tau + 1), P_j(\tau)] \quad , \tag{19}$$

$$\Gamma(8) = \rho[P_j(\tau + 1), P_j(1)] \quad . \tag{20}$$

We can also compute the classic Kullback-Leibler (1951) information difference with respect to successive values of population $P_j(\tau + 1)$ normalised to probabilities $p_j(\tau + 1)$ and we can also do this with respect to the observed values $P_j(1)$. These equations are

$$\left. \begin{array}{l} \Gamma(9) = \sum_j p_j(\tau + 1) log \frac{p_j(\tau+1)}{p_j(\tau)} \\ \sum_j p_j(\tau + 1) = \sum_j p_j(\tau) = 1 \end{array} \right\} \quad , \quad \text{and} \tag{21}$$

$$\left.\begin{aligned} \Gamma(10) &= \sum_j p_j(\tau+1) log \frac{p_j(\tau+1)}{p_j(1)} \\ \sum_j p_j(\tau+1) &= \sum_j p_j(1) = 1 \end{aligned}\right\} \quad . \qquad (22)$$

The final set we have computed are derived from the Sorenson-Dice statistics which determine the counts of the differences between population $P_j(\tau+1)$ and $P_j(\tau)$ and $P_j(\tau+1)$ and $P_j(1)$ which differ from one another and are thus defined as

$$\Gamma(11) = \frac{2\sum_j \hat{P}_j(\tau+1) + \sum_j \hat{P}_j(\tau)}{\sum_j P_j(\tau+1) + \sum_j P_j(\tau)} \ , \ \hat{P}_j(\tau+1) > P_j(\tau) \ , \ \hat{P}_j(\tau 1) > P_j(\tau+1) \ , \quad (23)$$

$$\Gamma(12) = \frac{2\sum_j \hat{P}_j(\tau+1) + \sum_j \hat{P}_j(1)}{\sum_j P_j(\tau+1) + \sum_j P_j(1)} \ , \ \hat{P}_j(\tau+1) > P_j(1) \ , \ \hat{P}_j(1) > P_j(\tau+1) \ . \quad (24)$$

We define the variant of this statistic here in analogy to its use in evaluating the goodness of fit in gravitational models (see Masucci, Serras, Johansson, and Batty, 2013).

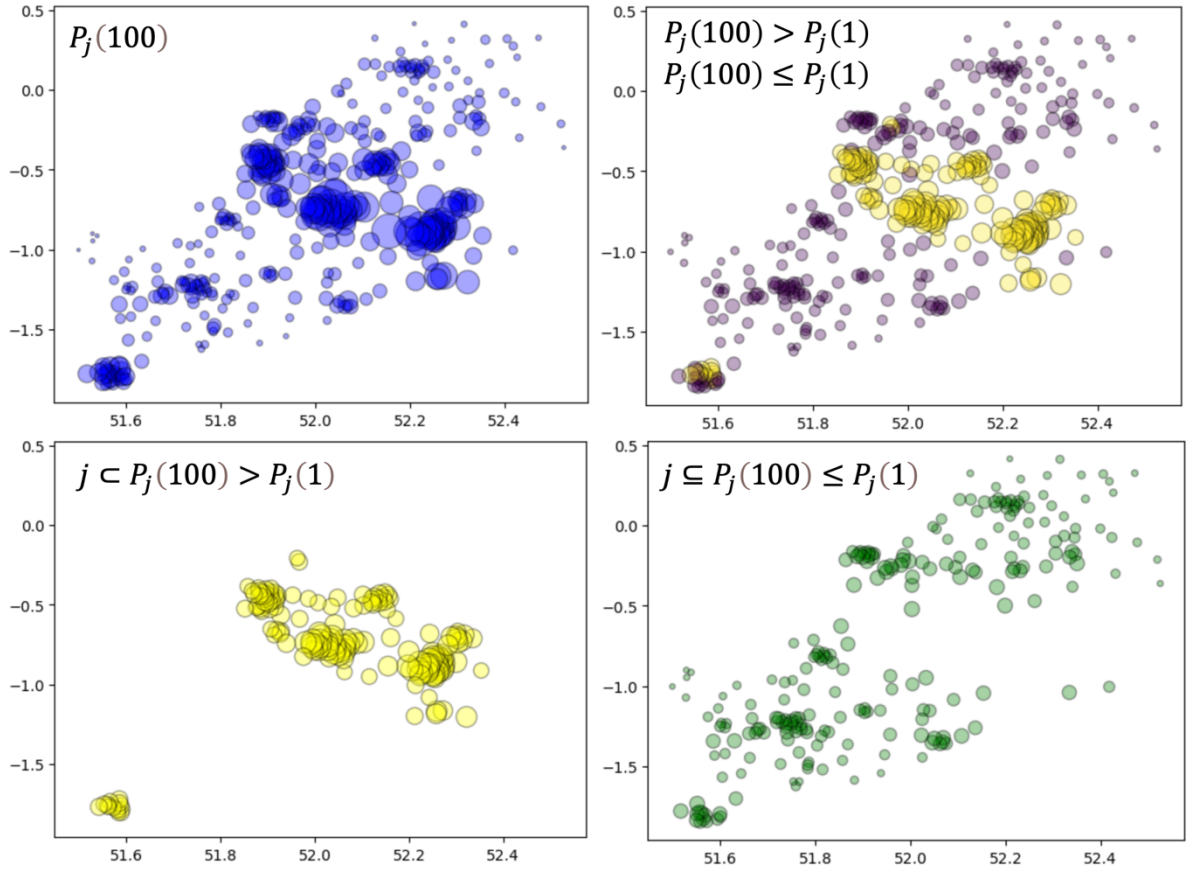


Figure 8: Long Term Equilibrium Spatial Distributions of Population

We have already illustrated the first cycle of iteration associated with the equilibrium conditions shown in equation (12) and we mapped the predicted and observed population and employment in Figures 2(a) and (b) and Figures 6(a) and (b) respectively. We have already noted the fact that the first cycle of the model generates more than proportionate employment relative to observed population, reflecting economies of scale which are a feature of the model. When we continue the iterations, these effects are reinforced and immediately it becomes clear that convergence of the activities on the final solution begins. In fact, we are able to end the convergence by defining thresholds on any of the 12 goodness of fit statistics but here we have

run the model up to $\tau = 100$ which to all intents and purposes generate stable values of $P_j(100), E_k(100)$ and the four employment types $E_i^{\ell}(100)$. We show a complete equilibrium solution in Figure 8 for employment and Figure 9 for population. These need to be compared with the first iteration of the model which generated Figures 6(a) and 6(b) as well as the observed values in Figures 2(a) and 2(b). The correlations between the employment and population in the long term equilibria ($\tau = 100$) compared to their observed values (at $\tau = 1$) fall to 0.519 and 0.265 respectively which implies much more spreading out but no radical change in their distribution over the region, despite quite large changes in the volume of activity predicted in different locations.

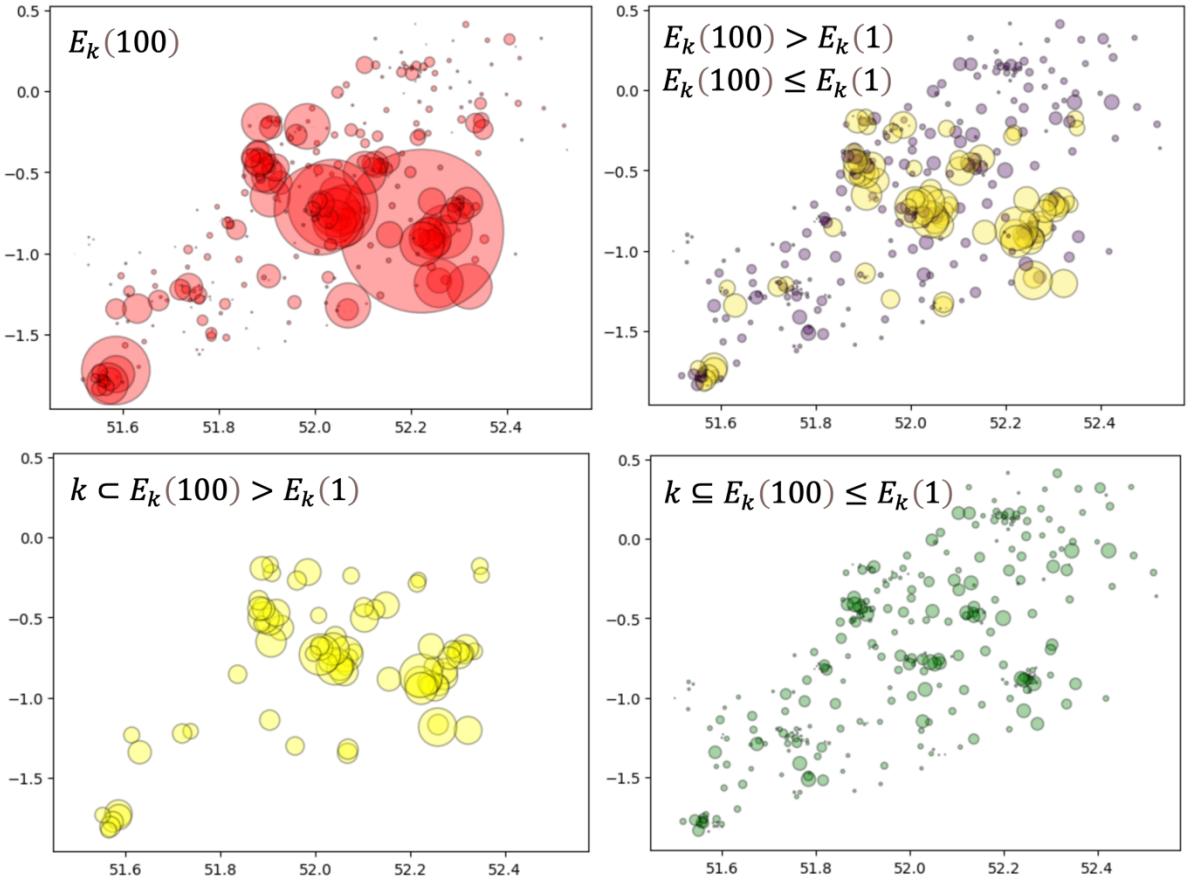


Figure 9: Long Term Equilibrium Spatial Distributions of Employment

In the equilibrium, we have separated out and then mapped the employment and population which are greater and less than the their observed values; that is, for employment, in Figure 8 we first map $E_k(100)$, the same variable with $E_k(100) > E_k(1)$ and $E_k(100) \leq E_k(1)$, and then these two sets of values on separate maps for $k \subset E_k(100) > E_k(1)$ and then $k \subseteq E_k(100) \leq E_k(1)$. In Figure 9, we plot the same types of maps for the relevant categories of population. Note that in all these maps we have used proportionate overlapping circles with a degree of transparency which give a much better sense of the fact that in this region, employments and populations are point rather than area locations and we consider these representations much more intuitively meaningful than their equivalent thematic maps.

We have plotted the 12 statistics for 100 iterations of the variable total population in Figure 10 and it is very clear that the convergence is smooth with no obvious discontinuities. For most measures, convergence is rapid at first and by iteration 30, most statistics are within 1% of their final values. The absolute and squared distance statistics $\Gamma(m), m = 1, ..., 6$ are easy to

interpret for these give the change in the average total count or percentage of population or employment per zone. The other six statistics – the correlations, the Kullback-Leibler information differences, and the Sorenson-Dice indices – all measure how close the solutions values are to their final solution as well as to their baselines – the observed values at $\tau = 1$. In the single cycle model, which is pictured previously in Figures 6(a) and 6(b), the model produced more than proportionate employment and population for the first prediction. When we move to the final iterative solution, Figures 8(a) and 9(a) show much the same degrees of agglomeration and the logarithmic scatter graphs based on comparisons of $E_k(100)$ against $E_k(1)$ and $P_j(100)$ against $P_j(1)$ are quite similar to the graphs in Figure 7(a) and (b). We will simply present the non-linear relationships, not the graphs as the fit; the estimated equations are fairly similar, with the equations estimated as $\log E_i(100) = -13.687 + 2.494 \log E_i(1)$ with a goodness of fit $R^2 = 0.548$ and $\log P_j(100) = -3.982 + 1.445 \log P_j(1)$ with the goodness of fit $R^2 = 0.623$. The implication of these results is that the economies of scale generated by the model do not vary much throughout the equilibrating process, but they remain highly significant.

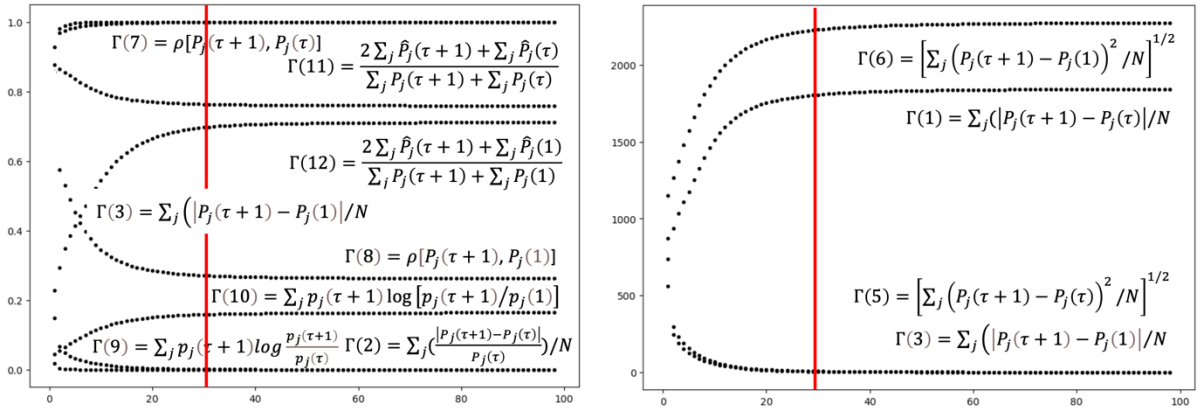


Figure 10: Convergence To Equilibrium: Population Trajectories

The last and perhaps most surprising feature of this process of reaching equilibrium is that continued iteration of the model equations leads to a succession of digital twins that might be considered different forecasting models in their own right. Moreover, continued iteration in the manner implied by equation (12) is like a sequence of forecasts which could be regarded as a system regenerating itself where the computable iterations τ are like time itself. In this sense, tracing the evolution of an equilibrium is like simulating the passage of time where the distribution of activities is continually changing. It is even possible to inject new exogenous variables into the mixture from external sources which always occur in any forecasting situation. To illustrate this, we simply modify the various inputs to the model by adding increments to any of the dependent variables but as the equilibrium model is composed entirely of dependent variables, the model becomes ever more complicated. To illustrate this, we take the first equation in (12) and modify this as $P_j(\tau) = \sum_{i\ell} A_i^\ell \{E_i^\ell(\tau) + \Delta E_i^\ell\} + F_j \exp(-\beta^\ell c_{ij})$ where we are adding a new component of employment ΔE_i^ℓ at each time period and for $E_k(\tau + 1) = \sum_{j\ell} B_j \{P_j(\tau) + \Delta P_j\} F_k^\ell \exp(-\gamma^\ell c_{jk})$ a new component of population ΔP_j. If these values were chosen as random variables, we could use this mechanism to simulate uncertainty in a probabilistic variant of the model, reflecting randomness in spatial behaviour. Such extensions, however, are currently beyond the scope of this paper but will be explored in later work

Developing Countless Digital Twins as Scenarios

To conclude our illustration of new simulation environments which we can use to explore many variants of our generic model, we will indicate how we can continually evolve new settlement patterns for the CAMKOX corridor by altering many features of the model itself and many of its inputs. These in turn can change the structure of the model and we already have some simple variants that we need to test to give some sense in which we can devise new structures that we can develop in this new environment of multiple simulations. In fact, we will not launch into a complete and comprehensive demonstration of how we might use these models continually on the desktop for we do not have the space here to present all this. But we can use this model to devise new plans for locating activities and constructing new transportation routes, thus generating one digital twin after another. To give some sense of this, there are three simple variants we have noted for which we can test different causal structures, first testing separate employment and residential models based on how population generates employment $E_k(\tau) = g[P_j(\tau)]$ and second how employment generates population $P_j(\tau) = f([E_k(\tau)]$. The first stage of the complete generic model generates population from employment, but we can also start this model structure by first generating employment from population, and as we have not yet demonstrated this, we will present this now.

We developed the full model in the previous section from $E_k(\tau + 1) = g\big[P_j(\tau)\big] = g[f[E_k(\tau)]]$ but we now also need to test the model starting from the population rather than the employment sub-model using $P_j(\tau + 1) = f[E_k(\tau + 1)] = f[g\big[P_j(\tau)\big]]$. We show the predictions of employment from the single residential location model $P_j(\tau) = f([E_k(\tau)]$ in Figure 11(a) and the correlation with observed employment is 0.586. The predictions from the full model starting from the same sub-model are shown in Figure 11(b) and the correlation for this prediction is much lower at 0.276. This prediction is not as radically different from the existing observed distribution of employment as in the generic model defined in equations (12) and we speculate that this is due to the fact that its population input data is much flatter in terms of its spatial clustering than employment.

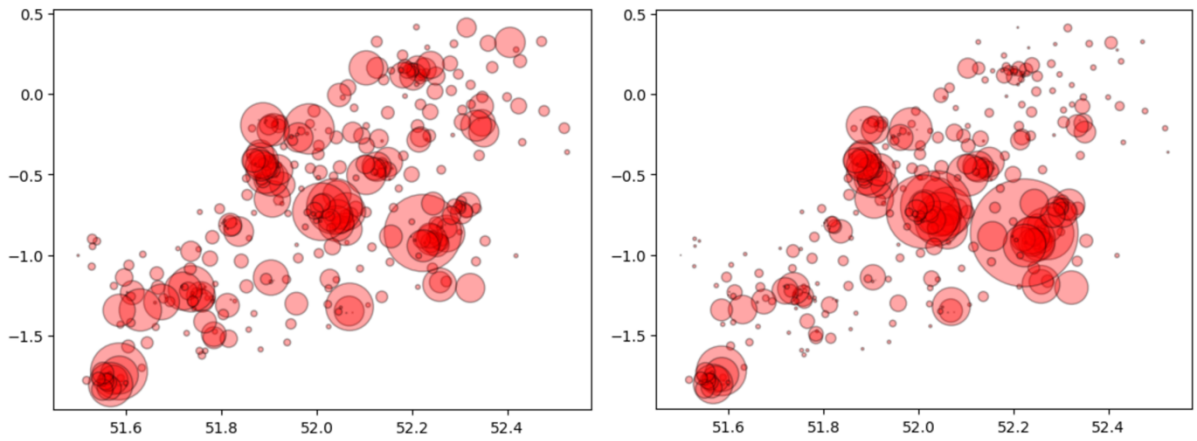


Figure 11: Variants of the Generic Model (a) Population Generating Employment in One Cycle (b) Population Generating Employment in 100 Time Cycles

We can also make many more changes to this model on-the-fly, and these give instant predictions but the advantage of the environment in which we can do these, is that we can input the data in single lines of code and generate immediate feedback that we can plot instantly using the simple nodal maps. One of the advantages in developing scenarios in this way is that it is very easy to

illustrate these predictions in the wider context of a dialogue between decision makers and planners. To impress these advantages, we will conclude with a simple example of looking at scenarios where single inputs of employment and population can be defined to test the impact of new growth poles or new towns in the wider region. To illustrate this, we note that in the CAMKOX corridor, there are two new towns, Stevenage started in 1946 (https://www.tcpa.org.uk/new-town/stevenage/), Milton Keynes (https://www.theplanformiltonkeynes.co.uk) established in 1969, a proposal for a new garden city in the Oxford area in 2014 (https://www.oxfordfutures.org.uk), and a plan for expanding Cambridge (https://www.cambridge-connect.uk/resources/cambridge-proposals/) first raised in 1996 We can simply reinforce these locations by adding 50,000 new jobs in the four locations and run the model which produces an immediate spread of these 200,000 jobs in the corridor. We anticipate that these jobs would reinforce the existing pattern of population because we have not changed the locational attraction factors or the transportation behaviours through the travel times between different places. If we simply add these jobs to these four locations, the model gives an immediate answer which we show in Figure 12(a) where the new jobs are located and in Figure 12(b) where the working population is generated. We first show the model predictions after one cycle of the model but then in the longer term equilibrium after 100 iterations, the patterns in 12(c) and 12(d) are quite different, reinforcing the central north-south corridor, as we also discovered in the various tests of the model in previous sections of the paper.

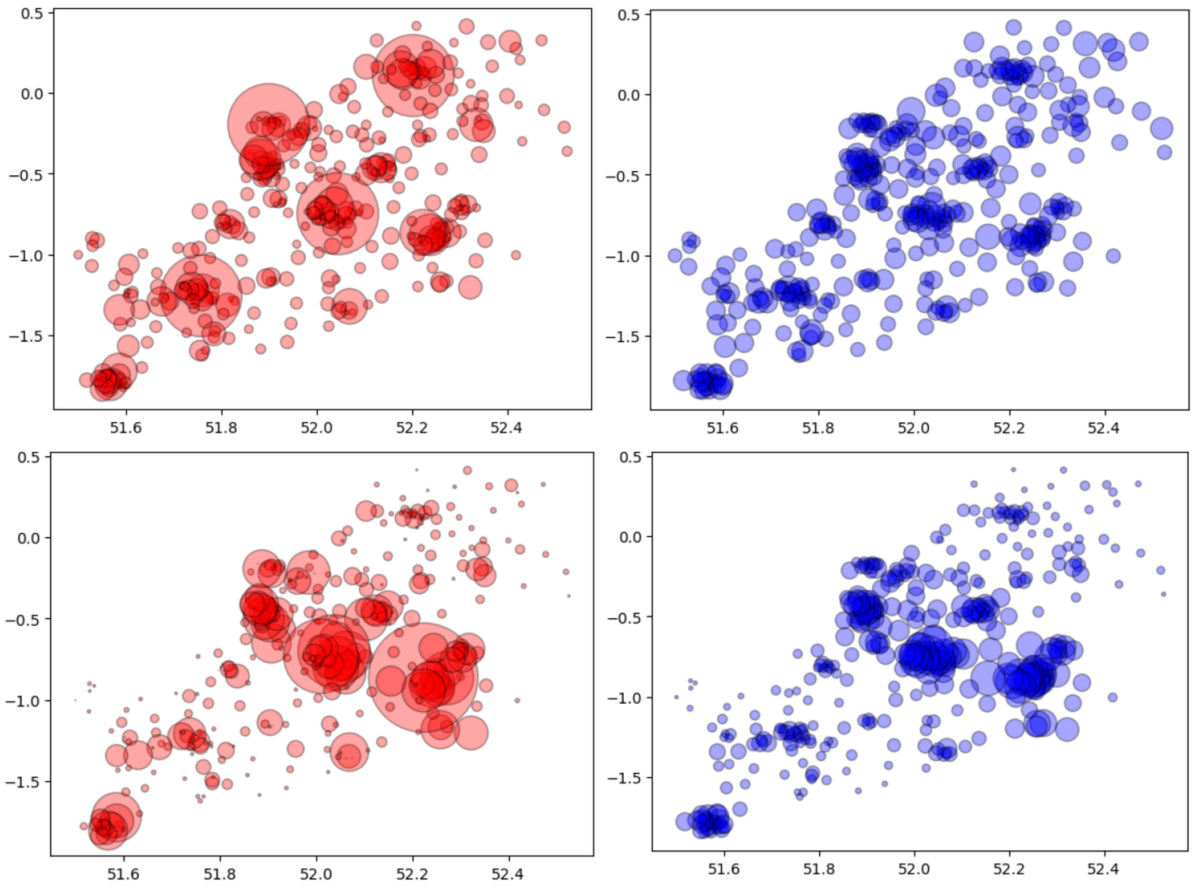


Figure 12: Intensifying the Four New Town Locations
(a) Employment $\tau = 1$ *(b) Population* $\tau = 1$ *(c) Employment* $\tau = 100$
(d) Population $\tau = 100$

In developing simulations in this way, we argue that we should begin to do this as much by trial and error guided by our own intuition as by any systematic variation of parameter values as we

have done in previous models (Batty et al., 2013). In short, the way we develop plans for cities before the digital age by manual sketch planning, can now be achieved on the desktop where the model builder can use his or her own imagination and intuition to explore multiple futures. Until we had environments in which to build relatively realistic models where we are able to simulate outcomes and scenarios almost instantly as we have done here using widely available open source software tools, it has not been possible to design and test anything but the most minimal of alternative futures using models of any kind. Until now we have not even catalogued the many ways, we are able to elaborate such futures as we have not had the ability to explore them other than in the most cursory way. As we have argued in this paper, this is now rapidly changing.

Every element of the model that we have constructed can be altered and the fact that we are dealing with an equilibrium model, means that all the model's variables are endogenous and can thus be manipulated; in short, this is because the inputs are also the outputs. There are some independent variables that we have assumed remain constant such as locational attractions and travel behaviours but even these can be varied to generate new scenarios. In terms of the equilibrium model which we defined in equations (12) as we are able to start it with assumptions about the location of employment and/or population, it is easy to begin to simulate alternative urban development patterns in a sequence which is guided by the outcomes that are simulated from changes in the model's inputs. We could continue our new town examples but there are many other patterns that we could test and in developing future variants of the model, this will be our goal.

References

d'Acci, L. S. (Editor) (2025) **Urban Scaling: Allometry in Urban Studies and Spatial Science**, Routledge, London.

Batty, M. (1984) **Pseudo-Dynamic Urban Models**, University of Wales Institute of Science and Technology, Cardiff, UK, http://www.casa.ucl.ac.uk/PhD.pdf

Batty, M. (2005) **Cities and Complexity: Understanding Cities with Cellular Automata**, Agent-Based Models, and Fractals, The MIT Press, Cambridge MA.

Batty, M., Vargas, C., Smith, D., Serras, J., Reades, J., and Johansson, A. (2013) SIMULACRA: Fast Land-Use–Transportation Models for the Rapid Assessment of Urban Futures, **Environment and Planning B**, **40**, 987 – 1002.

Batty, M., and Milton, R. (2023) Building a Digital Twin for British Cities, **Working Paper 236**, Centre for Advanced Spatial Analysis, University College, London, https://www.ucl.ac.uk/bartlett/casa/sites/bartlett_casa/files/casa_working_paper_236.pdf

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007) Growth, Innovation, Scaling, and the Pace of Life in Cities, **Proceedings of the National Academy of Sciences**, **104** (17) ,7301-7306, https://doi.org/10.1073/pnas.0610172104

Caldarelli, G., Arcaute, E., Barthelemy, M., Batty, M., Gershenson, C., Helbing. D., Mancuso, H., Moreno, Y., Ramasco, J., Rozenblat, C., Sánchez, A., & Fernández-Villacañas, J. L. (2023) Perspective: The Role of Complexity for Digital Twins of Cities, **Nature Computational Science**, https://doi.org/10.1038/s43588-023-00431-4

Hill, D. M. et al. (1965) A Growth Allocation Model for The Boston Region, **Journal of the American Institute of Planners**, **31**, 111–120

Hyman, G. M. (1969) The Calibration of Trip Distribution Models, **Environment and Planning**, 1, 105-114.

ITRC (2020). **A Sustainable Oxford-Cambridge Corridor? Spatial Analysis of Options and Futures for the Arc**, Infrastructure Transitions Research Consortium, Oxford, UK.

Kullback, S., and Leibler, R.A. (1951) On Information and Sufficiency, **Annals of Mathematical Statistics**, **22** (1), 79–86.

Lomax, N., Smith, A. P., Archer, L., Ford, A., and Virgo, J. (2022) An Open Source Model for Projecting Small Area Demographic and Land Use Change, **Geographical Analysis**, **54**, 599, 622.

Lowry, I. S. (1964) **Model of Metropolis**, RM-4035-RC, Rand Corporation, Santa Monica, CA.

Marshall, A. (1890, 2013) **Principles of Economics**, Palgrave Macmillan, London

Mumford, L. (1963) **The City in History: Its Origins, Its Transformations, and Its Prospects**, Secker and Warburg, London.

Page, S. E. (2019) **The Model Thinker**: **What You Need to Know to Make Data Work for You**, Basic Books, New York.

Wilson, A. G. (1971) A Family of Spatial Interaction Models, and Associated Developments, **Environment and Planning**, **3**(1), p1-32,