**Modelling spatial heterogeneity in land allocation: synthetic review of approaches and a new simulation technique using historical land use dynamics**

**1. Introduction**

Land-use change (LUC) is an issue of global concern and a major focus of sustainability research ([Aquilué et al., 2017](#_ENREF_2); [Gao and Bryan, 2017](#_ENREF_7); [Griggs et al., 2013](#_ENREF_8); [Veldkamp and Lambin, 2001](#_ENREF_30); [Yang et al., 2014](#_ENREF_35)). Predictive models which capture the complexities of land-use systems are important for analysing the drivers of LUC, and simulating future scenarios ([Cheng and Masser, 2004](#_ENREF_6); [Kamusoko and Gamba, 2015](#_ENREF_14)). Among these, so-called Cellular Automata models are perhaps the most commonly-used ([Batty and Xie, 1994](#_ENREF_5); [Ku, 2016](#_ENREF_16); [Santé et al., 2010](#_ENREF_25); [Tobler, 1979](#_ENREF_27); [Torrens, 2006](#_ENREF_28)). In these models, explanatory factors known to influence land use dynamics (model drivers) are used to compute a *transition potential* map, which determines the spatial allocation of land-use in each time step of the simulation ([Kolb et al., 2013](#_ENREF_15)). Model drivers typically include the proximity of land uses to each other (Neighbourhood ), the influence of infrastructure networks like roads, rivers or electricity transmission lines (Accessibility), and the underlying biophysical characteristics of the terrain (Suitability).

However, one fundamental problem with these kinds of models relates to the difficulty of simulating spatial variation in the pattern of land use between locations with apparently similar characteristics - known as *spatial heterogeneity*. When transition potential is calculated at each time step, there will typically be many highly suitable locations - a model will then allocate land to the highest potential cells until demand is exhausted. In this way, of these high potential locations, no one location will be favoured over any other, leading to an often unrealistically homogenous pattern of land use change. In reality, once high potential locations are identified (close to existing urban, accessible, flat land) development sites are chosen on the basis of socioeconomic criteria which are hard to capture in a spatial model – developers do not typically build five homes on ten equally suitable locations, they build fifty homes on one site, and then move on to others in subsequent years. A homogenous land allocation model is therefore quite unrealistic since it does not capture this dynamic. Since cellular automata models are iterative, and growth at one location in one time step then favours future growth at the same location, what may seem like a trivial problem can lead to major discrepancies between simulated allocation and real land use patterns. Thus, if spatial heterogeneity is not accounted for, a model will show many areas that grow slightly, rather than emergent new districts or expanding development corridors typically of the spatial dynamics of modern cities. In this way, a model which suggests that all ecosystems are slightly threatened is much less useful for managing environmental change than a model which identifies those one or two areas most susceptible to massive urban change and severe ecosystem degradation.

To resolve this problem, spatial constraints beyond the normal transition rules of neighbourhood, accessibility and suitability are typically applied. A wide range of approaches have been applied to this problem, each with their own individual shortcomings. One approach that has not so far been tackled, and which we present in this paper, is to use the endogenous pattern of growth embedded in the historic land use map time series to determine the most likely locations for future development, which we call *historical land use dynamics* (L).

To demonstrate our new approach, we first briefly review previous approaches to the problem of spatial heterogeneity and discuss their shortcomings (Section 2). Next, we describe the method we use to separate static from dynamic areas in a historical map time series and describe its integration with a cellular automata land use model in the R computing environment (Section 3). We demonstrate its effectiveness by applying the L factor to a *simplified model* configuration, in which the only other model drivers are Neighbourhood (N) and Stochasticity (α), and comparing the simulation results against both a *typical model* which includes Neighbourhood (N), Accessibility (A), Suitability (S) and Stochasticity (α), and a *combined model* comprising N, A, S, v, and L (Section 4). Finally, we discuss the results and their implications, as well as the limitations of our new approach (Section 5). Section 6 Concludes.

**2. Research Background**

Accounting for spatial heterogeneity in land use models is a longstanding problem for which various solutions have been proposed. We have identified six of the most common approaches, all of which address the problem to some extent, but have a number of shortcomings: 1) Zoning; 2) Stochasticity; 3) Neighbourhood weighting; 4) Geographically Weighted Regression (GWR); 5) Region overlay; 6) Actor-region overlay. These are discussed as follows:

2.1 Zoning

The conventional solution to this problem has been to include zoning information – detailed plans of controlled or restricted development prepared by the planning authority. Obviously, however, this approach can only be applied in cases where such information is actually available. Even if it is available, it is not always accurate, complete, or up-to-date. While it is probably true that most urban areas in most parts of the world are “zoned”, in the sense that city authorities have clearly defined rules about what kinds of development are permitted or restricted in particular locations, it does not follow that they will want (or be able) to share this information. This is especially likely to be the case in non-democratic countries where no legal obligation exists to put this information in the public domain, but it is also the case in some democratic countries too. Madrid, Spain, is one such prominent example where city authorities have repeatedly declined to make a zoning master plan publicly available (ref), perhaps because of politicians’ concerns that such a document might be used to curb this city’s enthusiasm for lightly controlled speculative development (ref). Finally, blanket exclusion or inclusion of areas on zoning plans can lead to an overfitted model which is too constrained to be useful for predicting future change.

2.2 Stochasticity

In some CA models (e.g. White and Engelen 1993, Hewitt et al 2013, Hewitt et al 2015), a stochastic term is also included to generate realistic variation between simulation runs. Beyond the intended effect of accounting for uncertainty in model drivers by increasing transition potential of certain locations at random, and thereby generating bifurcations (White et al 2014, p. X, Hewitt and Díaz Pacheco 2017) it also creates spatial heterogeneity, since clusters will form in some simulations and not in others. The disadvantage, however, is that the spatial heterogeneity this created is random, and does not reflect the real spatial variability of mapped land use changes.

2.3 Neighbourhood weighting

Feng and Tong (2019) approach the problem of spatial heterogeneity by using using hotspot analysis or gradients to weight the neighbourhood, usually the most important model driver, particularly in urban land change models (Shadman Roodposhti et al 2020). Useful though this approach clearly is, it is applicable to the cell neighbourhood only, so the inclusion of other model drivers, e.g. suitability or accessibility, risks diluting or negating the effect of the weighted neighbourhood maps.

2.4 Geographically weighted regression (GWR)

GWR also addresses this problem directly, and is a major improvement on conventional regression-based models of urban growth. However, this approach is reliant, like all land use modeling approaches, on correctly identifying the drivers of change. To be really effective, explanatory variables accounting for planners’ preferences (i.e. zoning) need to be identified and incorporated into the model, which returns us to the tricky problem of what to do if such information is not available. A further disadvantage is that GWR, like regression approaches in general, does not normally incorporate stochastic variation, and thus does not effectively address the complex characteristics of real cities, like bifurcation, emergence, or critical transitions (Hewitt and Díaz-Pacheco 2017).

2.5 Region overlay

This approach is found particularly in constrained cellular automata models like those of the popular Metronamica family (ref, ref). In these kinds of models, demand for urban land at the local level can be linked to population and economic variables at higher administrative levels, giving different land use demands for each region according to their demographic or economic characteristics. The region overlay approach has the great advantage of being able to replicate highly regionally dependent patterns of urban change, but this of course presupposes that good quality data on population and economy are available and that the links between these variables and economic growth are known. More problematically, this approach varies only the quantity of land allocated, but does not affect the transition potential within each region.

2.6 Actor-Region overlay

In this approach, used in the APoLUS model (Hewitt et al 2015, Hewitt 2018), overlay maps of administrative areas are used to weight the transition potential map according to the outcome of a simulated negotiation between different actors. For example, if actors are strongly in favour of development in a particular region, then this region will receive a higher transition potential than adjacent regions where actors preferences are not so strong. The actor-region overlay approach usually varies transition potential, rather than demand, within regions, such that regions where actors do not support a particular development type are less attractive overall for new development. Though this is realistic in theory, it relies on careful observation of actor behavior, which is difficult to do and requires extensive background research. This approach is more suited to a participatory exploration of future land use dynamics than for replicating observed patterns of change.

As we have seen, all of these approaches have advantages and disadvantages. All the methods except for the neighbourhood weighting approach are deductive, in that they introduce variables into the spatial model on the basis of some generalisable theory that is expected to improve model performance, or at least model theoretical coherence, in each specific case. The neighbourhood weighting approach is inductive, since it relies on the results of analysis of data in the case study area, and, on the basis of measurable improvements in simulation goodness-of.fit, claims to be useful in the general case (Feng and Tong 2019). This distinction is relevant to the strengths and weaknesses of the individual approaches – in general, the deductive methods seem intuitively correct, but may not necessarily improve goodness-of-fit in specific cases, while the neighbourhood weighting approach described by Feng and Tong (2019) would seem to improve goodness of fit in specific cases, but might not necessarily be generalisable beyond the study area or regional context in which it has been demonstrated.