There’s no best model! Addressing limitations of land-use scenario modelling through multi-model ensembles

Richard J. Hewitt, Majid Shadman Roodposhti and Brett A. Bryan

ABSTRACT

Cellular automata models are popular tools for exploring future land change pathways. But simulation modelling approaches often focus too narrowly on calibration against historic reference maps, limiting the diversity of possible outcomes. We argue that, contrary to what is commonly believed, there is no 'best model', and that model specification and calibration accuracy depend on the objective of the research. We propose a multi-model ensemble approach, in which a wide range of models and calibration rules sets are systematically tested against multiple metrics. We apply our approach to a case study in Spain. No single model performed well for all statistics, illustrating the danger of cherry-picking statistics for best performance. In our case study, accounting for historic land changes in model design was useful for simulating compact urban development, but limited the variability of simulation outcomes. The accessibility model driver improved urban pattern replication, while suitability without accessibility was useful for simulating low-density development encroaching on natural areas. Rather than abandoning calibrations that show low agreement with reference maps based on a small number of metrics we should seek to understand what each metric is telling us and use this information to enrich the diversity of simulated outcomes.

1. Introduction

Spatially explicit land-use models are widely used for projecting and understanding future land-use change in the face of widespread global concern about the impact of human societies on the biosphere (Turner et al. 2007, Rounsevell et al. 2006, Gomes et al. 2020). Cellular Automata models are perhaps the most commonly used (Batty
and Xie 1994; Ku 2016; Santé et al. 2010; Tobler 1979; Torrens 2006). In these models, explanatory factors or variables 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). Model drivers typically include the proximity of land uses to each other – neighbourhood (N), the influence of infrastructure networks like roads, rivers, or electricity transmission lines – accessibility (A), and the underlying biophysical characteristics of the terrain – suitability (S). In such models, there are two main approaches to transition potential computation. In the first approach, the relationship between drivers is tested statistically, e.g. through regression-based approaches (Veldkamp and Fresco 1996, Schneider and Pontius 2001) or by applying a machine-learning algorithm like Artificial Neural Networks or Random Forest (Pijanowski et al. 2005, Kamusoku and Gamba 2015, Roodposhti et al. 2019). These approaches are widely used but have some important disadvantages, including, but not limited to, non-independence of drivers, or multicollinearity (Verburg et al. 1999, p. 124, Feng et al. 2019), vulnerability to choice of spatial scale or aerial units (Openshaw 1981, Overmars et al. 2003, Zhang and Kukadia 2005, Pan et al. 2010, Díaz-Pacheco et al. 2018), overfitting (especially machine-learning approaches, see, e.g. Pijanowski et al. 2005, p. 198, Jabbar and Khan 2015) and, perhaps most importantly, the fact that the transition rules thus derived are rarely understandable by humans – a model developed through these kinds of artificial intelligence techniques is effectively a black box able to discover and apply transition rules, but not explain them (Castelvecchi 2016). This is true even for approaches which apply sophisticated factor selection techniques to test the influence of different variables on transition potential (see, e.g. Rienow and Goetzke 2015, Liu et al. 2020, Ye et al. 2019). Though the most important variables can be identified, the reason why such variables are important must be inferred by the researcher.

In the second approach, the most important model drivers are identified, e.g. N, A and S, not by crunching the numbers but through prior knowledge. As drivers of urban development worldwide, an overwhelming body of literature confirms the importance of proximity to existing urban centres (Alonso 1960, Herbert and Stevens 1960, Mills 1967, Muth 1969, Brueckner and Fansler 1983, White and Engelen 1993), accessibility through the transport network (Hansen 1959, Gauthier 1966, Fairburn 1969, White et al. 1997), and the physical suitability of land for development, especially the slope of the terrain (Clarke et al. 1996, Wear and Bolstad 1998, Schneider and Pontius 2001, Lo and Yang 2002). The question, then, is not whether these factors are significant drivers, but to what degree they are influential in a particular geographical area. To answer this in general terms is a straightforward regression problem. But if we wish to discover the rules that may determine future urban land expansion in a given territory, we need to experiment more broadly. The resulting approach to determining transition potential is usually known as calibration, since the most appropriate settings for drivers, and the balance between them, are discovered by trying to simulate land-use change to a date for which a reference map is already available, and then comparing the simulation against the reference map using a range of statistical techniques (Roodposhti et al. 2020). Yet manual trial-and-error based calibration can be very tedious, since a large number of rules sets must be systematically
tested, and may sometimes be nearly impossible due to combinatorial problems, as in
the case with models with many land uses and multiple parameters. As a result of
this limitation, manual trial-and-error calibration approaches often obtain a degree of
fit against the reference map which is quite inferior to that obtainable by statistical or
machine-learning approaches.

The solution to this problem would seem to be an approach to transition potential
discovery that is at least partially automated, but at the same time, generates plausible
and satisfactory rules, not just perfect replication of a reference map. Such an
approach would take a much broader view of what constitutes ‘a good model’ by
properly evaluating and sensitivity-testing a wide range of alternatives in order to
understand the parameter space in which the model is operating. Such studies do
exist, but are certainly in the minority. One key example is that of Mustafa et al.
(2018), who employed a modified genetic algorithm to search for multiple near-
optimal calibration rules, rather than converging on a single optimum solution. These
authors show how different solutions are applicable for simulating different types of
urban development.

The goal of the present paper is to present and describe such an approach, which
we execute in R software using the freely available SIMLANDER modelling framework
(Hewitt et al. 2013). We build on an earlier proof-of-concept study (Roodposhti et al.
2020) which gets around the combinatorial problems and computational resource
issues presented by exhaustive approaches to autocalibration (Straatman et al. 2004,
Newland et al. 2020) through a random sampling procedure. We move beyond this
earlier study by testing not only the transition rules, but also a large range of model
drivers and driver combinations. The multi-model approach we propose is common in
machine-learning classification problems, where various algorithms, e.g. classification
and regression trees, random forest, or artificial neural networks, are applied succes-
sively and the best-fitting algorithm is chosen (Kamusoku and Gamba 2015,
Roodposhti et al. 2019, Liu et al. 2020). In our case however, the goal is more difficult
than simply producing the best replication obtainable, in that we seek also to discover
and understand the most appropriate rules for simulating future urban expansion in
our test case study, thus shedding light on the urban expansion process. In fact, differ-
ent combinations of drivers are effectively different models, see e.g. sensu Tobler
(1979); this is a usage we adopt in this paper. Though our approach does share some
points of similarity with earlier, partial approaches to calibration based on inclusion or
exclusion of key drivers (e.g. Barreira González et al. 2015) we believe our study is one
of the first to systematically apply a multi-model autocalibration approach to establish
the importance of key model drivers without relying on regression-based or machine-
learning methods.

We also apply a practice that has long been common in modelling within other dis-
ciplines, and which we argue is necessary in land-use modelling too, that of simulation
ensembles. Ensemble modelling is the practice of using multiple models to address
the same question from different angles or using different techniques (see e.g.
Semenov and Stratonovitch 2010). The main function of the multi-model ensemble
approach is to fully communicate the variability across different models, since out-
comes, and hence the degree of uncertainty in future projections, have been shown
to be highly dependent on modelling assumptions and algorithms (see e.g. Thuiller et al. 2019, Pontius et al. 2008, Warren and Seifert 2011, inter alia). By communicating the whole range of expected variability within the model ensemble, uncertainty in future outcomes can be reduced. Examining the variability of a given model’s outputs gives us a better idea of how it will perform for scenario simulation.

Recent work on autocalibration (Roodposhti et al. 2020) indicates that for every model, there are likely to be multiple transition rules that compare acceptably with reference datasets, and our autocalibration approach is designed to account for this, producing a ranked series of metrics for each model. Further, the inclusion of the stochastic factor (v) means that simulations carried out with the same rules set will be different to one another. The degree to which they are different is a reflection of the degree of bifurcation in the cellular system (Couclelis 1985, White et al. 2015). It has recently been argued that to simulate the truly complex spatial pattern of urban development, model constraints need to be relaxed sufficiently to allow each simulation to produce reasonably diverse outcomes (Hewitt and Díaz-Pacheco 2017). Thus, rather than searching for the ‘best’ calibration – we argue, in fact, that there is no such thing – we develop an ensemble of results from our multiple models, which we then use to simulate future urban areas. Further, each model simulation itself has multiple ‘replicates’, on account of the stochasticity (v) in all the models. The result is a model ensemble which reflects not our own ‘best guess’, but the true variation in the model space. This approach not only offers a way to manage uncertainty in simulation of scenario outcomes (Maier et al. 2016), but also comes closer to the random approximation approach envisaged by the pioneering theorists of cellular automata (Ulam 1952, Metropolis and Ulam 1949).

To this end, our study is orientated around two research questions. The first (RQ1) asks, in general terms, what multi-model auto-calibration approaches can contribute to the field of urban land-use simulation modelling. Does such an approach offer advantages over methods that simply seek to find a single well-fitting rules set for a predetermined set of parameters? Our second research question (RQ2), which emerges out of our experimental multi-model approach, is to answer how parsimonious a model of urban land change can be, and what are its key basic elements. Granted, N, A, and S are clearly relevant, as we noted earlier. But what about zoning – the local spatial planning rules or protected area restrictions? And what about historic land-use change? To take no account of this latter element means our model might miss areas that are evidently susceptible to redevelopment (due to having changed frequently in the past) but are not included in our calibration maps. Yet, on the other hand, to apply a model that too rigorously constrains new land use into areas of previous change might be excessively path dependent, and unable to simulate urban expansion in newly suitable areas. We discuss the implications of the choice of model component in the light of these, and other, relevant considerations, with a view to providing some useful guidelines for the land-use modelling community.

2. Methods

2.1. Study area

To answer our research questions, we built a series of LUC simulation models for a study area which has seen a high level of urban expansion in the recent past and
for which a long time sequence of high-resolution land-use maps is freely available. This was important for the purposes of our study, since large amounts of change allow for higher confidence in the tendencies of particular calibration rules (law of large numbers), and long time periods allow for more divergent outcomes, since there are sufficient time steps to allow bifurcations to become consolidated. Our chosen location is a 47162 km² area in the Spanish autonomous region of Andalusía, centred on the city of Seville (the regional capital) but taking in the city of Córdoba to the northeast, coastal areas and cities to the south like Marbella, Algeciras, Cádiz, and Huelva, as well as the historic towns of Jerez de la Frontera and Sanlúcar de Barrameda (Figure 1). Andalusía saw very strong urban expansion in the later 20th century, reaching a peak around 2010, before rapidly declining when the real estate bubble burst following the Global Financial Crisis of 2007-8 (OSE 2012, Romero et al. 2012). Urban and industrial land parcels have become less concentrated and dispersed across a wider area, particularly along transport routes and along the Mediterranean coast. This high rate of urban expansion and diversity of urban land patterns with older, high-density core city areas contrasting with newer, more dispersed development makes the Seville area and its hinterland especially challenging for land-use change modelling. Additionally, a large-scale, highly detailed and freely accessible cartographic database for multiple time periods is available for a period spanning c. 50 years (Moreira 2007). Protection of natural areas is a priority issue in this region, with well-known and long-standing disagreements over land use and natural protection, especially in the emblematic and internationally recognized Doñana protected area (see, e.g. Zorrilla-Miras et al. 2014).

2.2. LUC modelling drivers and data sources

Geographic data for urban land use and LUC drivers were obtained from three main sources (Table 1). Land-use maps were obtained from the Environmental Information Network for Andalusia (REDIAM – Moreira et al. 2009), a large-scale (1:25000) land-use and land-cover dataset obtained by visual interpretation from aerial photographs for the years 1956, 1977, 1999, and 2007. Land use was reclassified into two different land-use classes: urban and non-urban. Roads data (Accessibility) for the year 2017 were obtained from Google Earth and a digital elevation model (2018) for posterior calculation of slope (Suitability) was obtained from the Shuttle Radar Topography Mission (SRTM). ArcGIS software was used to generate Euclidean distance maps and slope maps from roads data and elevation data respectively. All data were resampled to a grid cell resolution of 90 metres in the ENVI environment, with the study area comprising 2468 × 2359 cells in total. Land-use simulation was carried out in the R environment using the dataset described above.

2.3. Model overview

To demonstrate our proposed multi-model ensemble approach, we used the autocalibration procedure described by Roodposhti et al. (2020) to build eight different CA models (Section 2.4) using the SIMLANDER modelling framework for the R
environment (Appendix 1). The LUC drivers used to build the different models (Table 1) can be categorised into six groups or blocks: neighbourhood dynamics (N), accessibility (A), suitability (S), randomness (v), historical land-use dynamics (L) and Inertia (I).

In SIMLANDER, following the approach of White and Engelen (1993), the evolution of the future cell state is determined by calculating the transition potential (TP) for

Figure 1. Location map and urban land use in Andalusia, Spain (source: own work, based on Andalusian Government datasets).
each cell to change. The formula for TP calculation normally includes N, A, S and \( v \), and sometimes Z, representing Zoning (see e.g. White et al. 2000), i.e.:

\[
P_{t+1}^{ij} = f\left(N_t^{ij}, v_t^{ij}, A_t^{ij}, S_t^{ij}, I_t^{ij}\right)
\]

where \( N_t^{ij}, v_t^{ij}, A_t^{ij}, S_t^{ij} \) and \( I_t^{ij} \) are the neighbourhood effect, randomness values, accessibility, suitability and inertia at time \( t \), and \( f \) is a transition function, typically, but not necessarily, either the sum (White and Engelen 1993, Roodposhti et al. 2020), or the product of all the terms (Hewitt et al. 2014; Hewitt and Díaz-Pacheco 2017). However, the TP computation shown in Equation 1 is simply a convention, and the application of this framework in the R environment means that any of the model drivers, e.g. N, A, S, can be omitted or substituted with others to create a new and different model. Unfortunately, this is not always well understood, and very few published modelling case studies (though see Barreira González et al. 2015) rigorously test whether adding or removing additional parameters actually improves the model performance. The basis of our multi-model approach rests on the recognition that a truly robust simulation approach needs to test a wide range of models, creating each new model by omitting particular parameters or adding others.

Four land-use maps for the years 1956, 1977, 1999, and 2007 were used in this study (Table 1, Figure 2). Historical land use dynamics (L), included in the transition potential computation for four of the models (Section 2.4), were obtained from land-use maps for 1956, 1977, and 1999. The differences between the eight models were evaluated by simulating land-use change for the period 1999–2007 for each model using the same 50 rule sets (400 simulation results), and comparing them against the reference map of land use in 2007. To ensure that differences between the models could not be due to stochasticity, the random parameter \( v \) was held constant using the \texttt{set.seed} function for all model runs. Since the aim of the study was to compare different models for the same amount of land change allocated, the amount of land-use change to be simulated, known as land-use demand (Appendix 1) was calculated following the historical linear growth tendency.

| Drivers            | Description                                                                 | Unit                | Data source |
|--------------------|-----------------------------------------------------------------------------|---------------------|-------------|
| Neighbourhood (N) | Number of neighbouring urban cells for each cell                            | –                   | REDIAM      |
| Randomness (v)     | Randomness effect by Weibull distribution                                     | –                   | REDIAM      |
| Suitability (S)    | Slope value                                                                  | Degrees             | SRTM DEM    |
| Accessibility (A)  | Distance to road networks                                                    | Metre               | Google Earth Engine |
| LUC dynamics (L)   | Three various cell state at times \( t-1 = 1956, t-2 = 1977 \) and \( t-3 = 1999 \) are applied to segregate dynamic and static land-use area | –                   | REDIAM      |
| Inertia (I)        | Inertia is the starting map at time \( t-3 = 1999 \) (for calibration) or \( t-4 = 2007 \) for future simulation. It is applied to all models to represent the strong persistence of the urban core. | –                   | REDIAM      |
| Land-use 1956, 1977, 1999, 2007 | Land-use state (urban/non-urban)                                             | –                   | REDIAM      |
Figure 2. Urban growth in major cities of the study area since 1956. Hatched borders represent restricted zones for urban growth. (source: own work, based on Andalusian Government datasets).
2.4. The eight models

2.4.1. The parsimonious model
Following the original design of White and Engelen (1993), we model land use using only cell neighbourhood rules, inertia, and a random parameter, such that:

\[ P_{ij}^{t+1} = \sum \left( I_{ij}^{t+1}, N_{ij}^{t}, V_{ij}^{t} \right) \] (2)

where \( I_{ij}^{t+1} \) is the inertia parameter, and \( N_{ij}^{t} \) and \( V_{ij}^{t} \) are the neighbourhood effect, and randomness values at time \( t \), respectively. This model does not account for the effect of accessibility (A), suitability (S) or historical land-use dynamics (L) on new land allocation.

2.4.2. The accessibility model
This model recognises the key role played by transport networks in determining the urban land pattern. This model introduces the accessibility parameter (A), such that:

\[ P_{ij}^{t+1} = \sum \left( I_{ij}^{t+1}, N_{ij}^{t}, A_{ij}^{t}, V_{ij}^{t} \right) \] (3)

where \( I_{ij}^{t+1} \) is the inertia parameter, and \( N_{ij}^{t}, A_{ij}^{t} \) and \( V_{ij}^{t} \) are the neighbourhood effect, accessibility and randomness values at time \( t \), respectively. This model does not account for the effect of suitability (S) or historical land-use dynamics (L) on new land allocation.

2.4.3. The suitability model
This model includes suitability (S), represented by the slope of the terrain, instead of accessibility, such that:

\[ P_{ij}^{t+1} = \sum \left( I_{ij}^{t+1}, N_{ij}^{t}, S_{ij}^{t}, V_{ij}^{t} \right) \] (4)

where \( I_{ij}^{t+1} \) is the inertia parameter, and \( N_{ij}^{t}, S_{ij}^{t} \) and \( V_{ij}^{t} \) are the neighbourhood effect, suitability and randomness values at time \( t \), respectively. This model does not account for the effect of accessibility (A) or historical land-use dynamics (L) on new land allocation.

2.4.4. The typical model
This model includes all of the previously mentioned parameters, such that:

\[ P_{ij}^{t+1} = \sum \left( I_{ij}^{t+1}, N_{ij}^{t}, A_{ij}^{t}, S_{ij}^{t}, V_{ij}^{t} \right) \] (5)

where \( I_{ij}^{t+1} \) is the inertia parameter, and \( N_{ij}^{t}, A_{ij}^{t}, S_{ij}^{t} \) and \( V_{ij}^{t} \) are the neighbourhood effect, accessibility, suitability and randomness values at time \( t \), respectively. This model, though extremely popular, may be over-specified (Bolstad 2011) in many cases, especially for urban land-use model applications.

2.4.5. The path-dependent model
This model includes the historical land change dynamics parameter, L, but eliminates the A and S parameters, such that:
\[ p_{ij}^{t+1} = \left( \sum l_{ij}^t, L_{ij}^t \right) \cdot \left( \sum N_{ij}^t, v_{ij}^t \right) \]  

where \( l_{ij}^t \) is the inertia parameter, \( L_{ij}^t \) is historical land change dynamics, and \( N_{ij}^t \) and \( v_{ij}^t \) are the neighbourhood effect, and randomness values at time \( t \), respectively. Note that \( L \) can be included as another element in a sum (see e.g. Equation 5), but this was found to lead to a rather weak effect. To strengthen the influence of \( L \), it was combined with the inertia parameter \( I \), and used as a multiplying constant.

\( L \) is a new parameter, which we introduce to account for the path-dependent nature of urban development (Brown et al. 2005). The existing \( N \) and \( I \) parameters make existing urban development and land nearby more attractive for future urban development, but do not account for past change. This aspect has been overlooked previously because of the difficulty of producing or sourcing long land-use time sequences. Nowadays, with multiple high-resolution land-use maps increasingly available for long time periods this aspect deserves more attention. In our own case, we choose to start our calibration efforts at 1999, but take into account the change areas in earlier land-use maps for the years 1956 and 1977.

A further justification for the introduction of historical land-use change is found in terms of the zoning parameter (Onsted and Chowdhury 2014). Zoning, usually understood as areas that are restricted or prioritised due to planning legislation, is an important criterion for understanding the possible location of future development. Since zoning maps are often unavailable, we wanted to test the possibility that the inclusion of historical land-use dynamics might effectively account for these missing areas, by strongly concentrating new urban land in areas of past change. Our hypothesis was that such an approach might be effective where very large areas of natural land, such as in the case of national parks or conservation areas, are off-limits to development, and we wished to test whether this would work in practice (RQ2).

To include historical land-use change, we used a map algebra routine to extract change areas for 1956-1999. The neighbourhood rules matrix was then applied separately to the change areas, denoted active land, to create the \( L \) parameter. The \( L \) parameter, added to the \( I \) parameter representing existing land use at the calibration start date (1999) is then used as a multiplier in the transition potential equation, which includes only \( N \) and \( v \) parameters. Thus:

\[ L_{tn} = \sum t_{hn} \]  

where \( t_n \) is the calibration starting date, and \( t_{hn} \) is all historic change preceding this date.

Clearly, the addition of historical land-use change to the inertia parameter and the existing land-use neighbourhood rules, means that the model including the \( L \) parameter is likely to limit considerably the possibility for land-use change beyond the existing and previously urbanised areas.

### 2.4.6. The dynamic path-dependent model

As with the path-dependent model, but with \( L \) being updated to include a feedback loop accounting for land change within the simulation runtime, such that \( L \) is continuously updated, as follows:
\[ L_{tn} = \sum t_{hn} + t_{n-1} \text{ where } t > 1 \]  

(8)

2.4.7. The combined model

In the combined model, the L parameter is added to the transition potential calculation sum, such that:

\[ P_{t+1}^{t+1} = \sum \left( L_{ij}^{t+1}, f_{ij}^{t+1}, N_{ij}^{t}, A_{ij}^{t}, S_{ij}^{t}, V_{ij}^{t} \right) \]  

(9)

2.4.8. The dynamic combined model

As with the combined model, but with L being updated to include a feedback loop accounting for land change within the simulation runtime (Equation 8).

2.5. Zoning

All of the calibration simulations for each of the 8 models described above were run without restricting urban growth inside the zoned areas (Figure 2). A raster algebra operation was used to extract the cells of urban land that coincided with the zoning map in each simulation for each model. The number of coincident cells for each simulation, known as the zoning intersection count (ZIC) was summed and exported as a table. The ZIC score allows the degree of coincidence of the simulation ensembles for each model to be evaluated, in order to discover which model allocated the most land into zoned areas. For the future scenarios (Section 3.7), the zoned areas were excluded from the transition potential map to ensure that land was not allocated in these locations.

2.6. Autocalibration and model testing procedure

Beyond the initial data preparation steps (Section 2.2), all operations in the workflow were carried out in the R environment, using an adapted version of the SIMLANDER cellular automata model script (Hewitt et al. 2013). All modelling work was carried out on a desktop computer with 24 GB of RAM and a 3000 Mhz Intel Xeon CPU (E5 1607-v2) processor, running Windows 10. To overcome memory limitations, test runs for each rule set were executed using a simple batch script, which ran Rscript, incrementing the iterator value after each run, ensuring that Rscript used the next rules matrix in the set. The Automatic Rule Detection (ARD) tool (Roodposhti et al. 2020) was used with the random option, but with set.seed parameter ensuring that each model was tested against the same set of matrices.

Our workflow, implemented entirely within two R scripts controlled by the batch script described above, was executed in the following way for each of the eight models described in Section 2.4.

Firstly, the ARD tool (Roodposhti et al. 2020) was used to generate 50 random neighbourhood rule matrices, all describing positively-skewed curves of varying steepness commensurate with known urban land attraction distance decay rules (Roodposhti et al. 2020) and varying in neighbourhood radius size from 1 (3×3 window, central cell and one
Moore neighbour) to 5 (11 x 11 window, central cell and five Moore neighbours). The choice of 50 rules sets therefore equates to 10 different curve shapes for each of the 5 neighbourhood sizes. This figure was arrived at due to the need to effectively test rules curves ranging from very steep to very shallow, which can be effectively achieved with 10 curve shapes (Roodposhti et al. 2020), and neighbourhood sizes from immediate proximity (1 cell, up to 90 m distance from the centre cell) to distances likely to be outside the range of a typical urban development (5 cells, up to 450 m away) (Díaz-Pacheco et al. 2018). Next, the land-use maps, model drivers and zoning areas were imported into R. Land-use maps were processed to extract urban land only. Accessibility and slope maps were computed using the settings described in the preceding section. Subsequently, each model was run 50 times (once for each matrix rule), and a variety of cell quantity agreement/disagreement and pattern statistics were calculated.

A separate file was generated in which autocalibration results for all models were assessed against 24 separate statistics ranging from cell-by-cell coincidence (e.g. hits, misses, hit rate) to spatial pattern quantification metrics (core area, perimeter), measures of aggregation (clumpiness) or shape complexity (fractal dimension) (see Supplementary Material; Data and Codes Section, after Conclusions). Some metrics that are habitually used in land-use modelling and remote sensing communities were found to be inappropriate to our study case. For example, since our models are allocation models (the specified number of cells must always be allocated), false alarms (FAs) do not provide any useful information on the performance of the model. Thus the FA score and any derived statistics that include it (i.e. precision, accuracy, AUC) are meaningless in our case. Further, our study area included a large amount of white space (non-urban land potentially available for allocation comprised >95% of the study area) and is thus highly class-imbalanced, so we could not informatively use any statistic that included true negatives (i.e. accuracy, AUC, kappa). On this basis, we chose the hit-rate, also known as recall (hits/hits + misses), to assess cell-by-cell coincidence. Similarly, many pattern metrics could be discarded as unhelpfully similar to each other. After careful consideration, we eventually settled on 12 separate metrics for systematic evaluation of model results (Figures 3 and 4). Mathematical specifications of the pattern metrics used are provided by McGarigal (2015); for descriptions and use cases of several key metrics (Shape Index, Fractal Dimension, Clumpiness and Edge Density) see also Roodposhti et al. 2020.

2.7. Future scenario development

On the basis of the results obtained from the multi-model autocalibration procedure, we discuss what our chosen metrics tell us about the characteristics of the simulation ensemble, choosing individual rule sets that represent the diversity of these characteristics to derive a series of future scenarios of urban land change. The chosen rules sets were then used to simulate 10 replicates in each case to the year 2050.

3. Results

In this section, we discuss the performance of each model against the benchmarks and in comparison with the other models with reference to Figures 3–6. The
3.1. General observations

Firstly, it can be seen that including path-dependency in any model (PDM, DPDM, CM and DCM) dramatically decreases its variability, as shown by the much narrower inter-quartile range for these models than for the others (Figures 3 and 4).

Secondly, the class level statistics (Figure 3 (B.1–B.5)) highlight some important differences between the simulations and the reference map, though no single model performs well across all statistics. For example, some of the models approach the

---

**Figure 3.** Calibration results for all models; cell-by-cell statistics (A) and class-level pattern metrics (B.1–B.5). For the class-level pattern metrics, the horizontal line indicates the score obtained for each statistic in the reference map (land-use in 2007).
reference dataset for edge density, clumpiness and total core area (especially the latter). For mass fractal dimension, and patch density, however, the match is much poorer for nearly all simulations. The patch level statistics (Figure 4) tell a similar story, with the simulation results for some of the models able to approximate the reference dataset for number of core cells, shape, fractal dimension and core area, but not for the perimeter and perimeter area ratio. This finding cautions of the dangers of cherry-picking statistics for the best performance.

3.2. Inter-model comparisons

There are important differences between the models. Firstly, the cell-by-cell agreement statistic, the hit rate (Figure 3A) shows that the highest hit rate was achieved...
by the path-dependent models (PDM, DPDM) without the inclusion of other parameters, and that when other parameters were included (CM, DCM), the hit rate was reduced. In this sense, the inclusion of historic land change areas in the model can be seen to be effective (RQ2), as long as the models are kept simple. The dynamic

![Figure 5. Scatter plot showing all calibration results for all models; left: A: Hit rate (x) vs clumpiness (y); right: B: Hit rate (x) vs fractal dimension (y). The horizontal line indicates the score obtained for the reference map for the measure given on the y axis. Numbers 1–3 indicate critical transitions for combined and dynamic combined models.](image1)

![Figure 6. Scatter plot showing all calibration results for all models; left: A: Zoning Intersection Count (x) vs clumpiness (y); right: B: Zoning Intersection Count (x) vs fractal dimension (y). The horizontal and vertical lines indicates the scores obtained for the reference map for the measures given on the x and y axes.](image2)
path-dependent model (DPDM) results show that including the feedback loop to account for land areas that change during the model runtime increases the hit rate further. Combining the path-dependent models with the other parameters (CM, DCM) brings the hit rate down, but in both cases most of the results show higher hit rates than for the typical (TM) and suitability model (SM) variants. The dynamic combined model (DCM) performs more poorly against the hit rate statistic than the other path-dependent models. The Parsimonious model (PM) and the Accessibility model (AM) did achieve comparable hit rates to the path-dependent models in a few cases (Figure 3A).

For pattern metrics, the simpler path-dependent models (PDM and DPDM), which produced the highest hit rates, show the poorest performance against both class level and patch level metrics, with the non-dynamic path-dependent model (PDM) performing only slightly better than the dynamic path-dependent model (DPDM). All the other models (TM, PM, AM, SM, CM and DCM) produced simulations which replicated patterns more effectively. The combined models (CM and DCM) performed better than the other models in general for clumpiness, edge density and patch density (class level) and shape (patch level), but less well for the core statistics (total core area at class level and number of core cells and core area at patch level) and for fractal dimension (patch level), with the TM, PM and AM performing best against this statistic.

Figure 5, which plots hit rates against a class level statistic – clumpiness (Figure 5A) and a patch level statistic – fractal dimension (Figure 5B), shows the relationship between hit rate and pattern metrics quite effectively. As patterns become less aggregated (lower clumpiness) and tend towards more complex geometric forms (higher fractal dimension), the model also scores fewer hits relative to misses (lower hit rate). This demonstrates a clear trade-off between pattern and predictive accuracy. The highest scoring rules sets in terms of predictive accuracy as measured by the hit rate are therefore the most aggregated (highest clumpiness) and have the least fractal complexity (lowest fractal dimension). For the combined models (CM, DCM), results cluster in three clearly-defined blocks (Figure 5, numbered 1–3), apparently indicating two critical transition thresholds or tipping points, behaviour associated with complexity (Hewitt and Díaz-Pacheco 2017). There is also some evidence for this behaviour in the TM and AM plots. Detailed exploration of this behaviour is a priority for future work.

Figure 5 confirms the impression that, after the path dependent models (PDM, DPDM), it is the parsimonious model (PM) that offers the overall best performance for the hit rate statistic. The parsimonious model, therefore, answers our second research question (how parsimonious a model of urban land change can be, and what are its key basic elements?), reminding us, that despite three decades of evolution, the simple model proposed by White and Engelen (1993), which our parsimonious model closely resembles, is an excellent and useful model, and should always be considered before adding additional elements, like accessibility, suitability or zoning. As would be expected, since the only driving factors are the cell neighbourhood (N), existing urban land-use (I) and the stochastic factor (v), this model is very sensitive to changes in the neighbourhood rules produced by the various calibration runs, as the high degree of variability in the calibration results shows. This variability is useful: the slope of both graphs is very
steep, indicating that very complex objects (fractal dimension index >1.5), with a high degree of cell scatter (clumpiness <0.6), reminiscent of a very dispersed, low-density land-use pattern can be obtained with the PM without substantial reductions in the hit rate. A small cluster of rules sets achieving a hit rate of around 73% match the aggregation pattern of the reference map very closely. The addition of Accessibility and Suitability parameters (AM and SM) reduces the predictive accuracy but gives a cluster of results that exactly replicate the fractal dimension of the reference map. The fact that this cluster does not perform as well in terms of clumpiness (PM provides closer matches for this metric) indicates that the clumpiness and fractal dimension metrics, though similar, are not exactly interchangeable. Fractal dimension, as an indicator of shape complexity is a more sensitive indicator of linearity at the patch level, and as such, is a better measure of the influence of the accessibility parameter than the clumpiness statistic.

3.3. Zoning

The results of the zoning test are shown in Figure 6. First, it should be observed that the zoning boundaries recorded by the zoning map do not indicate absolute exclusion of all urban development; 1105 cells of urban land were recorded within the zoned areas by the 2007 reference map, as indicated by the x-intercept in Figure 6. Second, we can see that the hypothesis that the introduction of path-dependency in the models (PDM, DPDM, CM and DCM) may help exclude zoned areas without the need for a zoning map is broadly correct. The more conservative and compact new land allocation that these models produce means that few cells encroach into areas beyond existing built-up land. For the other models (TM, PM, AM and SM), there is considerable variation, more so, in fact, than indicated in the previous plots. The models in which accessibility is included (TM, AM, CM, DCM) show a lower degree of overlap with the zoned areas in general than those that do not. The PM includes ~20 simulations with over 6000 cells of urban land falling into zoned areas, and in the case of the SM, most of the simulations have more than 6000 cells of urban land in zoned areas. The likely explanation is that the zoned areas are mostly distant from transport networks, so that simulations that include the accessibility parameter are more likely to avoid them than those that do not. These results highlight major differences between the SM and the other models which could not be appreciated in the previous figures (Figures 3, 4 and 5). It seems clear that physical suitability, represented in this model by the slope of the terrain, is a useful tool for simulating the effect of very dispersed development on natural protected areas, in a scenario where the future configuration of transport networks is unknown. The zoning comparison exercise tells us that while path-dependent models reliably avoid zoned areas, models that include the accessibility parameter (TCA, AM, CM, DCM) are also quite effective. Models that do not include either path-dependence or accessibility (PM, SM) are much less so.

3.4. Choosing the model ensemble

The systematic exploration of the multi-model calibration ensemble described above allows a series of land-use simulation scenarios to be developed that take full
advantage of the diverse sample population, rather than the usual approach of choosing a single calibration result from a single model, thus narrowly constraining future simulations. Table 2 shows a selection of different models and rules sets, obtained by ranking the calibration results against particular metrics selected by studying the distribution of results described in the previous sections (Figures 3, 4, 5 and 6). Where particular models performed well against more than one metric, a ranking procedure was used to select the rule set, using the geometric mean of the relevant metrics to ensure equal weighting of each metric. Where the aim was to include the tails of the ensemble distribution, the ranking procedure was not used, we simply selected the highest or lowest value from the ensemble. Figures 7–15 show the outcome of the scenarios in Table 2. Figure 16 provides a decision flowchart to assist readers with model selection based on the recommendations provided in this paper.

4. Discussion

The results of the application of the selected rules sets in Table 2 show not only the variability within a single model, but the variability across all models in the autocalibration ensemble. The results at the far ends of the distributions (see the scatter-plots; Figures 5 and 6), which are intended to allow the full variability of each model to be appreciated, often look unrealistic. Nevertheless, they provide important information about model behaviour. We argue that future scenario modelling outcomes should include these results as a way to communicate model uncertainty – i.e. to visualise the logical conclusion of the rules in each model at their most extreme. For example, Scenario 1a (Figure 7), generated with the DPDM, shows the logical conclusion of a model which prioritises historical path dependence over all other factors, and as a result, performed best on simple cell-by-cell accuracy (highest hit rate for

| Scenario no. | Scenario description | Model | Rule set |
|--------------|----------------------|-------|----------|
| 1a           | Compact, non-transport-led development, greater densification than reference map (RM) | DPDM  | Highest Hit Rate (HR) |
| 1 b_1        | Compact, transport-led development, greater densification than RM | AM     | Highest HR, highest Clumpiness (CL) |
| 1b_2         | Compact, transport-led development, greater densification than RM | AM     | Highest HR, lowest Fractal Dimension (FD) |
| 2a           | Scattered, non-transport-led development, much more scattered than RM | PM     | Lowest CL, Highest FD |
| 2b           | Scattered, transport-led development, much more scattered than RM | AM     | Lowest CL, Highest FD |
| 3a_1         | Replicate existing land pattern only | AM     | Total Core Area (Rank 1) |
| 3a_2         | Replicate existing land pattern only | CM     | CL (Rank 1) |
| 3a_3         | Replicate existing land pattern only | DCM    | Patch Density (Rank 1) |
| 3a_4         | Replicate existing land pattern only | TM     | Core Area Index (CAI), FD (Rank 1) |
| 3 b_1        | Replicate existing land pattern, trade off against accuracy | CM     | HR, CL, PD (Rank 1) |
| 3 b_2        | Replicate existing land pattern, trade off against accuracy | PM     | HR, CAI, FD (Rank 1) |
| 4            | Scattered development in protected areas | SM     | Lowest CL/Highest FD, Highest Zoning Intersection Count (ZIC) |
Scenario 1a: Compact, non-transport-led development, greater densification than reference map

Run 1 of 10, urban land in 2050

![Figure 7. Scenario 1a, DPDM, Highest HR.](image)

In this scenario, the DPDM has successfully optimised for hit rate, but offers a simulation of future urban land that seems too aggregated and has little variation between simulation runs (Figure 7). Scenario 1b tries to simulate the effect of transport networks by using the AM, while maintaining the highly clumped nature of the distribution. Here, the ranking procedure was used to select the best performing hit rates with the highest clumpiness (1b_1) or lowest fractal dimension (1b_2). The results (Figure 8), as expected, give a land pattern that is even more aggregated than Scenario 1a, but the effect of accessibility works to link the separate urban clusters. Though this seems visually unappealing, this information is useful as means of understanding the upper bounds of the distribution in terms of aggregation. Scenarios 2a and 2b, conversely, show the phenomenon of scattering, taken to its furthest extreme, as measured by rule sets giving the lowest scoring clumpiness and highest scoring fractal dimension, for a model in which transport networks are included (AM) and one in which they are not (PM). In Figure 9, we see that everything that was not already urban land in 2007 has been scattered, apparently at random, across the map, with each new simulation run choosing a slightly different pattern of dispersion. Though this seems implausible, it is the logical conclusion of a scenario in which all pretence at concentrating development in existing areas is abandoned. Scenario 2b (Figure 10) introduces transport networks, but maintains the excessive scattering, giving an odd ribbon-like pattern of development. Scenario 3a seeks to replicate the pattern of the reference map, useful for understanding how
Scenario 1b_1: Compact, transport–led development, greater densification than reference map
Run 1 of 10, urban land in 2050

Figure 8. Scenario 1b_1, AM, Highest HR, highest CL.

Scenario 2a: Scattered, transport–led development, much more scattered than reference map
Run 1 of 10, urban land in 2050

Figure 9. Scenario 2a, PM, lowest clumpiness, highest fractal dimension.
Scenario 2b: Scattered, transport-led development, much more scattered than reference map

Run 1 of 10, urban land in 2050

![Map of Seville and Costa del Sol](image)

Probability of urban land in 2050 (10 simulations), green: low; dark red: high

**Figure 10.** Scenario 2b, AM, lowest clumpiness, highest fractal dimension.

Far a future simulation result has departed from the starting map. The results, naturally enough, differ depending on the pattern metric (or combination of metrics) selected, and the model used. Here, however, we can see some convergence between the models (Figures 11 and 12) – indicating that careful choice of several closely-matching rules sets from several models is able to minimise the uncertainty. Trade-offs between different pattern metrics (Scenario 3a_4; Figure 13) and trade-offs between pattern similarity and cell-by-cell accuracy (Scenario 3b_1; Figure 14) produce similar results. Scenario 4 (Figure 15) which includes neighbourhood, randomness and suitability in the form of the terrain slope, is a very effective model of the sensitivity of protected areas to encroachment by low-density urban development, a phenomenon known as ‘naturbanization’, and well known in our study area (Prados Velasco 2012). It seems that in this case, by excluding the accessibility driver and choosing calibrations with very high indicators of dispersion (low CL, high FD), new low-density urban areas appear on any flat land, regardless of the low accessibility and distance from urban centres of such locations. These are exactly the characteristics sought by second home owners in natural beauty spots – they don’t live there all year round and are looking for the most beautiful location at the lowest cost. Though the model used for Scenario 4 could be further refined to make protected areas attractive to housing, we should remember that development in these areas is not actively encouraged. This model would seem to be a reasonable approximation of possible outcomes under a very dispersed and liberated planning system, with social dynamics driven by a desire to escape from urban centres,
Scenario 3a_1: Replicate existing land pattern only (Total Core Area)
Run 1 of 10, urban land in 2050

Figure 11. Scenario 3a_1 AM, best fit for total core area.

Scenario 3a_3: Replicate existing land pattern only (Patch Density)
Run 1 of 10, urban land in 2050

Figure 12. Scenario 3a_3, DCM, best fit for patch density.
Scenario 3a_4: Replicate existing land pattern only (Core Area Index and Fractal Dimension index)
Run 1 of 10, urban land in 2050

Figure 13. Scenario 3a_4, TM. Best fit for core area index and fractal dimension.

Scenario 3b_1: Replicate existing land pattern, trade off against accuracy (Hit Rate–Clumpiness–Patch Density)
Run 1 of 10, urban land in 2050

Figure 14. Scenario 3b_1, CM, Best for hit rate, clumpiness and patch density.
as has been recorded in some countries during the covid-19 pandemic (Gurrutxaga 2021).

**4.1. Can the approach we propose address some of the limitations of land use modelling methods?**

In the introduction, we cited a number of well-known disadvantages of existing land use modelling approaches. Statistical models, we noted, are vulnerable to issues like multicollinearity and modifiable reporting unit problems, machine learning tends to produce opaque transition rules (the black box problem), and non-automated trial and error approaches are tedious and risk seeming arbitrary. The approach we have described does avoid or solve many of these defects. Rigorous testing and comparison of various configurations of model drivers (the multi-model approach) addresses
problems of multicollinearity. Though this has been noted before (Feng et al. 2019), we would argue that our semi-automated trial-and-error approach is more transparent than statistical number crunching, and moreover, by studying the whole distribution of results against multiple metrics, we can avoid the vexed question of ‘best fit’ entirely. By testing the whole ensemble of model parameterisations, from the simplest to the most complicated, we can also avoid over-specification (Bolstad 2011). In support to the study by Feng et al. (2019), we show how very simple models can produce better outcomes by some metrics than more complex models, reinforcing the modelers’ mantra about parsimony, known as Ockham’s Razor. The modifiable reporting unit problem, or MAUP as it was originally known, is another aspect that our approach can potentially address, by systematic testing of different scales or resolutions. Our scripts (Supplementary Material) offer a means to automate the process originally described by Díaz-Pacheco et al. (2018) in which different sets of rules are tested at different resolutions and the results compared. Overfitting can be avoided in general by examining the whole distribution and picking different rules sets for future simulation of scenarios, e.g. either side of critical transition thresholds (see e.g. Figure 5, dynamic combined model results, indicated by numbers 1-3). The systematic testing

Figure 16. Decision flowchart to aid model selection from an ensemble.
of multiple transparently-described rules avoids the conceptual ‘black box’ problem presented by some machine-learning approaches (especially, for example, artificial neural networks). Finally, the problem of arbitrariness associated with trial-and-error approaches is addressed by automating the rules search by randomly sampling the parameter space (See e.g. Roodposhti et al. 2020) and presenting the whole distribution, e.g. as boxplots and scatterplots (e.g. Figures 3-6).

In recognition of the variability generated by the different parameter choices, and in view of the high level of uncertainty surrounding the way the choices are made, which may often depend on the technique adopted as much as the choice of variables, some recent authors have begun to adopt a more holistic approach to land use simulation, with multiple testing of different outcomes (see e.g. Feng et al. 2019), or approaches that seek several near-optimal calibration rules rather than converging on a single solution (Mustafa et al. 2018). Yet the academic rewards system does not facilitate synthesis, leading to a tendency to publish many highly similar papers addressing only small elements of wider problems. Our article has attempted to bring these disparate ideas together to bridge the gap between different spatial simulation modelling approaches and advance the discipline through a unifying methodological vision to guide future research: the multi-model ensemble approach.

5. Recommendations

Selecting a single model, i.e. from a favourite machine-learning technique or a tried-and-tested procedure is a good starting point for land-use simulation. But it is no more than a starting point. Testing approaches need to go beyond simply demonstrating that the model can simulate land-use change, and show that the model chosen is appropriate within a given range of options relevant to its intended use. This should not be understood as a recommendation to arbitrarily select several techniques and try to say which is ‘the best’, but to systematically test different models and rule sets based on what the model is intended to do. For example, we have argued that path dependency is an important factor and have thus tested the performance of a path-dependent model against other models which give lesser importance to this criterion. We also reflect on the importance of simplicity – i.e. Ockham’s Razor – and accordingly test a parsimonious model, in which other factors known to be important empirically, but not always justified in every case are deliberately omitted.

Further, when it comes to testing and comparing the performance of simulation results from different models, selecting the best scores based on a single metric, unfortunately still a very common practice in land-use simulation modelling, will lead us to reject many simulations which closely replicate land-use change reference maps. Optimising calibrations against a single statistic is not advisable, since, as our results show, this tends to lead us to the tail of a given distribution, making it likely that simulation results will be skewed to one end of a particular range. The results of optimisation always depend on what is being optimised, and no single statistic measures every property of a model. A useful analogy here might be a pliable material that can be stretched in various ways, e.g. chewing-gum, or bread dough. Stretching along a single axis – e.g. to optimise length –
will compromise the other properties of the material to an unacceptable extent, with the result that the whole may not be fit for purpose.

Unlike in other modelling fields, e.g. climate modelling, where it is well-known that different model specifications and measures of model performance offer different perspectives, land-use modelling is mostly still stuck in an earlier paradigm, where the goal is to find the ‘best model’. The problem is that there is really no such thing as a best model. Even where several map comparison statistics are used to assess model calibration performance (e.g. Roodposhti et al. 2020) most land-use modelling approaches still seek to narrow the options down to a single ‘best’ calibration, which is used to generate all future scenarios.

We recommend instead to use several different models, each calibrated using a wide range of rules sets, with particular models and calibration results being selected for scenario development based on the requirements of the modelling objective. For example, where the requirement is to simulate scattered urban development in a historically compact urban region, it makes little sense to select a model which matches historical data for pattern aggregation measures like clumpiness, edge density or fractal dimension. By contrast, to simulate increasing compactness, our study shows that including strong path dependence is an effective approach, but including accessibility drivers is less useful. In short, we should choose our model and calibration rules set on the basis of the land-use pattern we want to generate, not on the basis of agreement with arbitrarily chosen or irrelevant metrics.

Our results speak directly to the growing realisation among environmental scientists and policymakers that ‘best-guess’ approaches to highly uncertain futures in which social, environmental and technological factors interact with each other in unpredictable ways are no longer fit for purpose (Maier et al. 2016). As land-use modellers, we need to get better and smarter at maintaining coherence between change drivers and simulation results without basing all our assessments of future land use on a few model test runs measured against arbitrarily chosen indicators. Land-use modelling needs to employ a whole ensemble approach that accounts for uncertainty by transparently exploring the whole parameter space without exhaustive computation. Such approaches are indeed emerging, e.g. quantification of uncertainty through meta-modelling of sensitivity (Şalap-Ayça et al. 2018), or by finding lower and upper bounds for the parameter space under multiple objectives (Hildemann and Verstegen 2021). But it is also necessary to go beyond these emerging technical solutions, and enhance credibility and transparency of our modelling processes, given the need to communicate robust advice to policymakers in the face of multiple environmental crises. One way forward would be through more rigorous application of socio-participatory scenario approaches like the shared socio-economic pathways (e.g. Pedde et al. 2021), with different model rules sets corresponding to particular tendencies identified by stakeholders. The present paper has shown how such an approach might work in practice.

6. Conclusion

The land-use simulation modelling procedure we have presented in this paper is quite different to that typically employed where a statistical model (e.g. logistic regression)
or a learning algorithm (e.g. random forest) is used to fit or ‘train’ the model (analogous to the calibration step discussed here), with simulations run to a future date using the fitted or trained model. It is also different to most trial-and-error calibration approaches previously employed, in that we have not sought to identify an ideal or ‘best’ calibration for use in simulating future change outcomes. Following the findings of an earlier study (Roodposhti et al. 2020), which cautions against the idea of a single best model and rule set, instead, our process seeks to retain the broad characteristics of the whole multi-model rule-set ensemble by choosing from representative points within the entire population. In this way, the model addresses uncertainty concerns by bracketing possible outcomes within key maxima and minima. This provides some relief from a host of existential terrors that our model is over-specified, badly calibrated, overfitted etc, since we, and readers, reviewers and end-users of our model results, can just inspect the distribution, i.e. using boxplots (Figures 3 and 4) or scatter-plots (Figures 5 and 6), and select the model and rule sets which best correspond to the desired real-world dynamic.

Notes

1. For further discussion on the role of pattern metrics in land use simulation, please see, e.g. White (2006), Chen et al. (2014), Wang and Marceau (2013).

Acknowledgements

Richard J Hewitt gratefully acknowledges support provided by the European Union under Programme H2020-EU.1.3.2, MSCA-IF-2019 (INTRANCES Project, Ref 886050).

Author contributions

RH developed the model scripts, ran the models and wrote the manuscript. MS developed the model scripts, prepared figures and edited the manuscript, BB supported the model development process and edited the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data and codes availability statement

The data and codes that support the findings of this study are freely available at https://doi.org/10.6084/m9.figshare.17086643.v4

Notes on contributors

Richard J Hewitt (RH) is Ramón y Cajal Research fellow at Madrid Complutense University, Spain. His research focuses on integrated modelling for sustainable environmental policy and management, and in particular, cellular automata models of spatial change and diffusion.

Majid Shadman (MS) was a postdoctoral researcher at the Swinburne University of Technology, Melbourne, Australia, before recently moving to industry. Majid has been involved in natural
hazard modelling, urban growth simulation, spatial multi-criteria decision analysis and uncertainty assessment.

**Brett A. Bryan (BB)** is Alfred Deakin Professor of Global Change, Environment and Society at Deakin University, Melbourne, Australia. His work is focused on creating cost-effective policy and management solutions for the sustainability of coupled human and nature systems. He has expertise in the application and development of computational tools and analytical methods in integrated modelling and assessment of land use and ecosystem services under global change.

**ORCID**

Richard J. Hewitt [http://orcid.org/0000-0003-4169-8647](http://orcid.org/0000-0003-4169-8647)

Brett A. Bryan [http://orcid.org/0000-0003-4834-5641](http://orcid.org/0000-0003-4834-5641)

**References**

Alonso, W., 1960. *A model of the urban land market: Location and densities of dwellings and businesses*. Philadelphia, PA: University of Pennsylvania.

Barreira González, P., Aguilera-Benavente, F., and Gómez-Delgado, M., 2015. Partial validation of cellular automata based model simulations of urban growth: an approach to assessing factor influence using spatial methods. *Environmental Modelling & Software*, 69, 77–89.

Batty, M., and Xie, Y., 1994. From cells to cities. *Environment and Planning B: Planning and Design*, 21 (7), S31–S48.

Bolstad, P.V., 2011. The tyranny of complex spatial models. *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, (11), 1–4. [https://www.geofocus.org/index.php/geofocus/article/view/209](https://www.geofocus.org/index.php/geofocus/article/view/209)

Brown, D.G., et al., 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19 (2), 153–174.

Brueckner, J.K., and Fansler, D.A., 1983. The economics of urban sprawl: theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics*, 65 (3), 479–482.

Castelvecchi, D., 2016. Can we open the black box of AI? *Nature News*, 538 (7623), 20–23.

Chen, Y., et al., 2014. Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy. *International Journal of Geographical Information Science*, 28 (2), 234–255.

Clarke, K.C., Hoppen, S., and Gaydos, L., 1996, January. Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. In: Third International Conference/Workshop on Integrating GIS and Environmental, 21–25. [https://pubs.er.usgs.gov/pubs/18750](https://pubs.er.usgs.gov/pubs/18750)

Couclelis, H., 1985. Cellular worlds: a framework for modeling micro—macro dynamics. *Environment and Planning A: Economy and Space*, 17 (5), 585–596.

Díaz-Pacheco, J., Van Delden, H., and Hewitt, R., 2018. The importance of scale in land use models: experiments in data conversion, data resampling, resolution and neighborhood extent. In: M. T. Camacho Olmedo, M. Paegelow, J. F. Mas, and F. Escobar, eds. *Geomatic approaches for modeling land change scenarios*. Cham: Springer, 163–186.

Fairburn, D.T., 1969, January. Simulation of urban growth as a function of accessibility. In Proceedings of the 6th annual Design Automation Conference, 297–322. [https://pdf/10.1145/800260.809034](https://pdf/10.1145/800260.809034)

Feng, Y., et al., 2019. Urban expansion simulation and scenario prediction using cellular automata: comparison between individual and multiple influencing factors. *Environmental Monitoring and Assessment*, 191 (5), 291.

Gauthier, H.L., 1966. *Highway development and urban growth in Sao Paulo, Brazil: a network analysis*. PhD dissertation. Northwestern University. [https://www.proquest.com/openview/febe6ff215592659b1bdd551b6891834b/1?pq-origsite=gscholar&cbl=18750&diss=y](https://www.proquest.com/openview/febe6ff215592659b1bdd551b6891834b/1?pq-origsite=gscholar&cbl=18750&diss=y)
Gomes, L.C., et al., 2020. Land use and land cover scenarios: an interdisciplinary approach integrating local conditions and the global shared socioeconomic pathways. Land Use Policy, 97, 104723.

Gurrutxaga, M., 2021. Visualizing the rural population growth in Spain during 2020 triggered by the Covid-19 pandemic. Regional Studies, Regional Science, 8 (1), 305–307.

Hansen, W.G., 1959. Accessibility and residential growth. Doctoral dissertation. Massachusetts Institute of Technology.

Herbert, J.D., and Stevens, B.H., 1960. A model for the distribution of residential activity in urban areas. Journal of Regional Science, 2 (2), 21–36.

Hewitt, R., and Díaz-Pacheco, J., 2017. Stable models for metastable systems? Lessons from sensitivity analysis of a cellular automata urban land use model. Computers, Environment and Urban Systems, 62, 113–124.

Hewitt, R.J., et al., 2020. Impacts and trade-offs of future land use and land cover change in Scotland: spatial simulation modelling of shared socioeconomic pathways at regional scales (SSPs). SocArxiv.

Hewitt, R., Díaz Pacheco, J., and Moya Gómez, B., 2013. A cellular automata land use model for the R software environment. SocArxiv.

Hewitt, R., van Delden, H., and Escobar, F., 2014. Participatory land use modelling, pathways to an integrated approach. Environmental Modelling & Software, 52, 149–165.

Hildemann, M., and Verstegen, J. A., 2021. Quantifying uncertainty in Pareto fronts arising from spatial data. Environmental Modelling & Software, 141, 105069.

Jabbar, H., and Khan, R.Z., 2015. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Science, Communication and Instrumentation Devices, 70, 163–172.

Kamusoko, C., and Gamba, J., 2015. Simulating urban growth using a random forest-cellular automata (RF-CA) model. ISPRS International Journal of Geo-Information, 4 (2), 447–470.

Lo, C.P., and Yang, X., 2002. Drivers of land-use/land-cover changes and dynamic modeling for the Atlanta, Georgia metropolitan area. Photogrammetric Engineering and Remote Sensing, 68 (10), 1073–1082.

Ku, C.-A., 2016. Incorporating spatial regression model into cellular automata for simulating land use change. Applied Geography, 69, 1–9.

Liu, B., et al., 2020. Nonparametric machine learning for mapping forest cover and exploring influential factors. Landscape Ecology, 35 (7), 1683–1699.

McGarigal, K., 2015. FRAGSTATS help. Amherst, MA, USA University of Massachusetts, 182.

McGarigal, K., 2015. FRAGSTAS help. Amherst, MA, USA University of Massachusetts, 182.

Moreira, J.M., 2009. Sistema integrado de gestión y consulta del catálogo de información raster de la Red de Información Ambiental de Andalucía (REDIÁM). Seville, Spain: Junta de Andalucía. Available at: http://www.juntadeandalucia.es/ medioambiente/rediam

Mustafa, A., Ebaid, A., and Teller, J., 2018. Benefits of a multiple-solution approach in land change models. Transactions in GIS, 22 (6), 1484–1497.
Muth, R., 1969. *Cities and housing: the spatial patterns of urban residential land use*. Chicago: University of Chicago, 4, 114–123.

Newland, C.P., et al., 2020. A hybrid (semi) automatic calibration method for cellular automata land-use models: combining evolutionary algorithms with process understanding. *Environmental Modelling & Software*, 134, 104830.

Onsted, J.A., and Chowdhury, R.R., 2014. Does zoning matter? A comparative analysis of landscape change in Redland, Florida using cellular automata. *Landscape and Urban Planning*, 121, 1–18.

Openshaw, S., 1981. The modifiable areal unit problem. *Quantitative Geography: A British View*, 60–69. Andove, UK: Routledge and Kegan Paul.

OSE. 2012. *Informe de sostenibilidad en España 2012*. Available from: http://www.upv.es/contenidos/CAMUNISO/info/U0637061.pdf [Accessed 2016].

Overmars, K.D., De Koning, G.H.J., and Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164 (2–3), 257–270.

Pan, Y., et al., 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34 (5), 400–408.

Pedde, S., et al., 2021. Enriching the Shared Socioeconomic Pathways to co-create consistent multi-sector scenarios for the UK. *The Science of the Total Environment*, 756, 143172.

Pijanowski, B.C., et al., 2005. Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographical Information Science*, 19 (2), 197–215.

Pontius, R.G., et al., 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42 (1), 11–37.

Prados Velasco, M.J., 2012. Naturbanization and urban patterns in national parks in Andalusia. *Boletín De La Asociación De Geógrafos Españoles*, (60), 431–435.

Rienow, A., and Goetzke, R., 2015. Supporting SLEUTH—Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, 49, 66–81.

Romero, J., Jimenez, F., and Villoria, M., 2012. (Un) sustainable territories: causes of the speculative bubble in Spain (1996–2010) and its territorial, environmental, and sociopolitical consequences. *Environment and Planning C: Government and Policy*, 30 (3), 467–486.

Roodposhti, M.S., Aryal, J., and Bryan, B.A., 2019. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. *Environmental Modelling & Software*, 112, 70–81.

Roodposhti, M.S., Hewitt, R.J., and Bryan, B.A., 2020. Towards automatic calibration of neighbourhood influence in cellular automata land-use models. *Computers, Environment, and Urban Systems*, 79, 101416.

Rownsevell, M.D.A., et al., 2006. A coherent set of future land use change scenarios for. *Europe. Agriculture, Ecosystems & Environment*, 114 (1), 57–68.

Şalap-Aycça, S., et al., 2018. A meta-modeling approach for spatio-temporal uncertainty and sensitivity analysis: an application for a cellular automata-based urban growth and land-use change model. *International Journal of Geographical Information Science*, 32 (4), 637–662.

Santé, I., et al., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape and Urban Planning*, 96 (2), 108–122.

Schneider, L.C., and Pontius, R.G., Jr, 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, 85 (1-3), 83–94.

Semenov, M.A., and Stratonovitch, P., 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. *Climate Research*, 41 (1), 1–14.

Straatman, B., White, R., and Engelen, G., 2004. Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems*, 28 (1–2), 149–170.

Thuiller, W., et al., 2019. Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, 10 (1), 1–9.
Appendix 1. SIMLANDER detailed model description

In SIMLANDER, land use change is computed as a function of the key drivers: neighbourhood dynamics (N), accessibility (A), suitability (S), randomness (v). These are described as follows:

A1.1. Neighbourhood dynamics (N)

N are the core of any CA model and form the basis of the interactions between individual cells and their neighbours in a regular cell grid. In geospatial software, neighbourhood interactions are modelled by applying a moving window or kernel to urban land use at the simulation starting time $t_0$. 
The resulting map represents the susceptibility of urban land in $t_0$ to change to urban land in the next time step $t_1$, and is referred to as the neighbourhood transition potential. The neighbourhood transition potential map is then used (together with transition potential maps created from the interaction of other factors, e.g. A and S), to allocate new land use to the highest potential areas, creating a land-use map for time step $t_1$. The moving window operation is again applied to the new land allocation map produced in $t_1$, and so on, until the final time step of the simulation. $N$ is calculated after White et al. (2000) as:

$$N_{ij}^{t+1} = \sum \sum w_{kad}$$

Where $w_{kad}$ is the weighting parameter applied in land-use map $t$ to land-use $k$ at position $x$ in distance zone $d$ of the neighbourhood, and $N_{ij}^{t+1}$is the neighbourhood transition potential for the new urban land-use map for time $t + 1$.

A1.2. Randomness ($\nu$)

$\nu$ is included to ensure a level of variation between simulations, such as that found in the real world. The random factor ($\nu$) applied in this study is provided by a Weibull distribution function, as shown in the following equation (cf White and Engelen, 1993).

$$\nu = 1 + (-\ln(1 - \text{random})) \cdot \exp(\alpha)$$

where $\alpha$ is the scale factor, which controls the strength of the stochastic effect and random is a uniformly distributed pseudorandom number. The $\alpha$ factor can take any desired value. However, practical experience indicated that values in the range $-10 > \alpha < 1$ are usually appropriate, with $-10$ giving a very weak random effect and $1$ a very strong random effect. We used a value of 0.5 in all models, which gives a moderate random effect (see Roodposhti et al. 2020 for further discussion).

A1.3. Accessibility (A)

$A$ represents the increasing attractiveness of areas for new urban land on the basis of their proximity to roads, was applied using a monotonically decreasing sigmoid function representing distance decay (Roodposhti et al. 2020). Euclidean distances from major roads were calculated in GIS software and used as the input of the sigmoid function (Equation 3):

$$A_{ij}^t = \begin{cases} 
1, & \text{if } d_{ij} \leq a \\
0.5 \left(1 + \cos\left(\frac{\pi(d_{ij}-a)}{b-a}\right)\right), & \text{if } b > d_{ij} \leq a \\
0, & \text{if } d_{ij} \leq b
\end{cases}$$

where $d_{ij}$ is the distance value of the cell $a$ and $b$ are the nearest and furthest distances to the road network. Accessibility parameters were identical in all simulations in which they were used and did not change during model runtime.

A1.4. Suitability (S)

This factor is frequently included in land-use models to reflect the physical aptitude of different terrain types for urban development. In multiple land-use models suitability factors like soil type or lithology may be appropriate (e.g. for modelling crop allocation). In our case, the only suitability factor included is the slope angle of the terrain, a factor known to be an important driver
of urban development (see, e.g. Clarke et al. 1996; Schneider and Pontius 2001). The slope layer was first calculated in GIS software using a digital elevation model, and subsequently divided into six classes in the range 0–1 using a reclassify operation, with the flattest land (0° angle) taking value 1 (highly suitable), the steepest slopes (>20° up to the maximum value of 57°) taking value 0 (unsuitable), and intermediate slopes taking intermediate suitability values (0.7,0.5,0.3,0.1). Suitability parameters were identical in all simulations and did not change during model runtime.

A1.5. Inertia (I)

I is sometimes included in the transition potential calculation to ensure that urban land in t + 1 is always allocated first in areas which were already built up in time t. I was calculated simply by assigning value 1 to all urban land use in the simulation start year, and adding this map to the other parameters.

A1.6. Transition potential computation

Transition Potential for change in a given cell at a given date \( P_{ij}^{t+1} \) is calculated as follows:

\[
P_{ij}^{t+1} = f\left( N_{ij}^t, v_{ij}^t, A_{ij}^t, S_{ij}^t, I_{ij}^t \right)
\]

where \( N_{ij}^t, v_{ij}^t, A_{ij}^t, S_{ij}^t \) and \( I_{ij}^t \) are neighbourhood effect, randomness values, accessibility, suitability, and inertia at time \( t \), and \( f \) is a transition function, typically the product or the sum of all the terms (Hewitt et al. 2014; Hewitt and Díaz-Pacheco, 2017).

A1.7. Land use allocation

At each model time step, all cells in the model are allocated to the locations where the transition potential \( P_{ij}^{t+1} \) is highest. The number of cells which will be allocated at each timestep is known as the demand, and is allocated cumulatively. E.g. for the case of a final demand of 200 cells after 10 years of urban expansion, up from a starting total of 100 cells in time \( t = 1 \), 10 extra cells would be added at each timestep, i.e.

100 + 10 cells at \( t = 1 \), 100 + 20 cells at \( t = 2 \), 100 + 30 cells at \( t = 3 \), and so on until total demand is satisfied with 100 + 100 cells at time \( t = 10 \). In calibration, the demand is simply \( C_{\text{End}} - C_{\text{Start}} \) / \( n \); where \( C_{\text{End}} \) is the number of cells of urban land in the reference map which is to be simulated, \( C_{\text{Start}} \), is the number of cells of urban land in the initial map from which the simulation will begin, and \( n \) is the number of years between the two simulations. Naturally, this will produce incremental linear growth between the two time periods. Of course, if more information is available, demand could be allocated non-linearly instead by modifying this formula. In simulating future land-use change, future demand must be determined, and can be estimated in several ways, e.g. from the outputs of other models (e.g. Hewitt et al. 2020), or from population projections (Hewitt and Díaz-Pacheco 2017).

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE 2385