High-resolution simulations of population-density change with an activity-based cellular automata land-use model

T. Crols*[1,2,3], R. White[4], I. Uljee[3], L. Poelmans[3], G. Engelen[3] and F. Canters[2]

[1]School of Geography, University of Leeds, Leeds, LS2 9JT, UK
[2]Cartography and GIS Research Group, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[3]Environmental Modelling Unit, VITO – Flemish Institute for Technological Research, Boeretang 200, 2400 Mol, Belgium
[4]Department of Geography, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada
*Email: t.w.crols@leeds.ac.uk

Abstract

The MOLAND model is a cellular automata (CA) land-use change model that has often been applied to simulate urban growth. A more recent alternative model makes the simulations more multifunctional by also computing different activities (population and employment) for every cell. However, the equation to update population density in time in this activity-based CA model could not deal with high population growth rates in some existing urban centres. Therefore, we experimented with two alternative equations. A semi-automated calibration routine was used to compare errors of the different model versions at a continuous range of resolutions in two study areas: the Greater Dublin Region, Ireland, and Flanders and Brussels, Belgium. The two new population density equations turn out to solve the particular problem of fast changes in high-density neighbourhoods and generally improve regional errors in the Belgian application, but can unfortunately introduce larger errors in low-density areas or in the land-use simulations.

Keywords: Population density, Cellular automata, Densification, Land-use change, Semi-automated calibration.

1. Introduction

Cities and their suburbs grow fast in many regions around the world. This urban growth can be studied with a whole range of land-use change models (Koomen and Stillwell, 2007). Cellular automata (CA) models are dynamic and spatially explicit (Poelmans and Van Rompaey, 2010; Santé et al., 2010). They simulate the influence that different land uses have on each other in their neighbourhood. The MOLAND model couples a constrained CA model to a gravity-based model of regional demographics and economics and has many applications worldwide (White et al., 1997, 2015; Engelen et al., 2007). More recently, an alternative to MOLAND was developed in which the population density and employment in different economic sectors are simulated as activities in the CA model in order to omit the coupling with the regional model and to enable multifunctional urban growth simulations at the level of individual cells (White et al., 2012). This paper presents an update of how this activity-based CA model (ACA model) can deal with changing population densities in different urban environments. We applied the model both to the Greater Dublin Region for the period 1990-2000, when this region experienced a fast monocentric urban growth (Williams and Shiels, 2002), and to the northern regions of Belgium (Flanders and the Brussels Capital Region) for the period 2001-2013, when these regions

experienced scattered suburban growth together with a large population increase in existing urban cores (Crols et al., 2017).

2. The activity-based cellular automata model

In the ACA model, activities are associated with active land uses but can also be present in any other land use. Influences of activities on each other are taken into account at all distances within the study area. To make this computationally feasible, cell values are grouped into super-cell values in a variable grid structure: the further away from the focus cell for which the neighbourhood effect is being computed, the larger the super-cell (Figure 1) (White, 2006). Variable grid levels are numbered upwards from 0 (unit cell resolution).

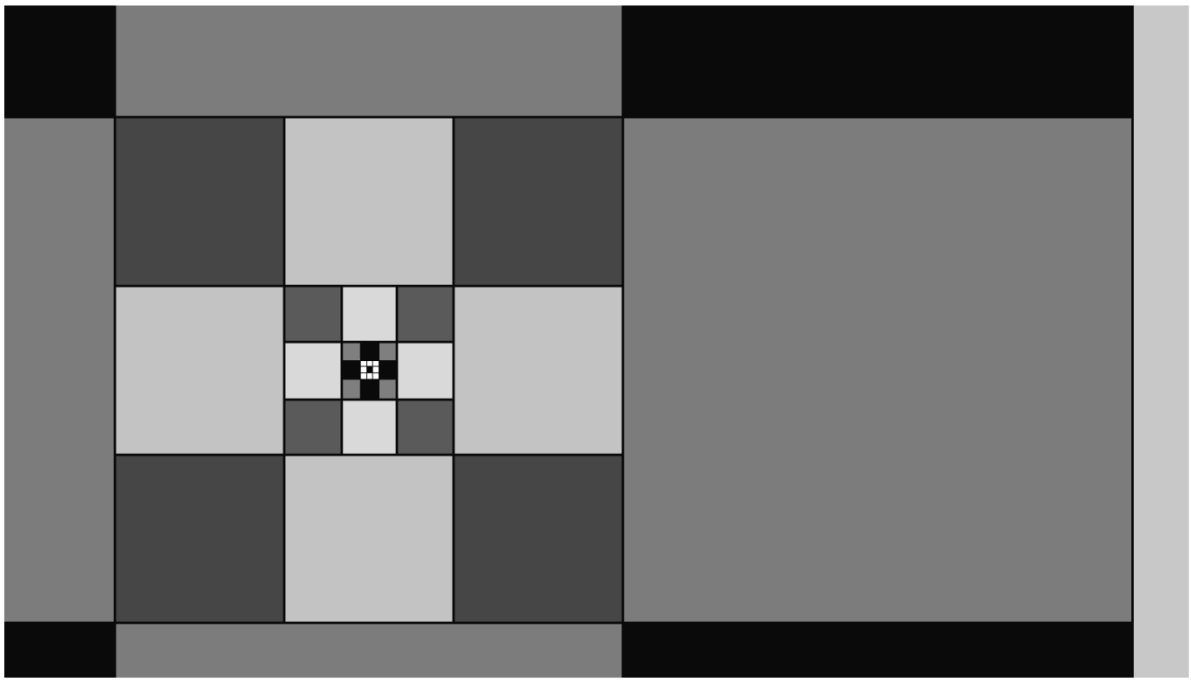

Figure 1: Structure of the variable grid.

Crols et al. (2015) updated the variable grid approach and have set up influence rules depending on travel time for all distances beyond the local environment (typically from +/- 1 km).

Transition potentials V_{Ki} on a cell *i* are computed for each activity *K* for each annual time step:

$$V_{Ki} = r\, Z_{Ki}\, S_{Ki}\, X_{Ki}\, N_{Ki} \qquad \text{Equation 1}$$

with *r* a random perturbation term, Z_{Ki} the zoning status, X_{Ki} the accessibility to the transport network, S_{Ki} the physical suitability, and N_{Ki} the neighbourhood effect. The neighbourhood rules are a function of the Euclidean distance – or travel time for long distances – between the focal cell and a variable grid cell that influences the focal cell:

$$N_{Ki} = \Sigma_J \Sigma_j\, W_{JK,dij}\, (T_{Jj} / T_J) \qquad \text{Equation 2}$$

with $W_{JK,dij}$ the weight given by the influence function f_{JK} for the influence of activity *J* on activity *K* at (time) distance d_{ij}, T_{Jj} the total activity *J* on variable grid cell *j*, and T_J the total activity *J* in the study area. Travel times are computed between the centres of gravity of population of the variable grid cells.

The land-use transition potential VT_{Ki} for the associated active land use U_K on cell i is calculated as:

$$VT_{Ki} = D_{Ki}\,(V_{Ki})^{mK} + (I_K)^{\rho} \qquad \text{Equation 3}$$

with D_{Ki} the diseconomies of agglomeration, representing the effect of high land prices and congestion on locational decisions, and m_K parameters to be calibrated. Next, I_K is the inertia value, which deals with the tendency of land uses to remain fixed at a location, with ρ a parameter to decrease inertia outside the associated land use of activities (inside, ρ = 1 by definition). All cells are ranked by their highest potential for a land use. The transition rule gives each cell the land use for which it has the highest potential until there is no more demand for that land use.

Activity values T_{Ki} are updated to T'_{Ki} in two steps, followed by a rescaling operation to ensure that total demand values are respected. Firstly, the allocation of activity in cells with a changed land-use state is in proportion to the relative value of the activity potential V_{Ki} within all cells of this land-use state. Next, activities should be updated in all cells. The original equation of White et al. (2012) uses relative changes in activity potential in comparison with the previous time step:

$$T'_{Ki} = \left[\frac{(V_{Ki})^{m_K} / \sum_i ((V_{Ki})^{m_K})}{(V_{(t-1)Ki})^{m_K} / \sum_i \left((V_{(t-1)Ki})^{m_K}\right)}\right]^{\tau_K} T_{Ki} \qquad \text{Equation 4}$$

with $V_{(t-1)Ki}$ the transition potential for activity K on cell i in the previous time step, m_K as in equation 3, and τ_K parameters to be calibrated.

In this study, we experimented with two alternative equations to replace equation 4, which turned out to have too weak an effect in fast-growing urban environments. Both alternatives increase the direct impact of the relative size of the neighbourhood effect on population, the most important factor in cities to determine the location of more rapidly densifying areas. The first alternative makes use of a densification exponent χ_K to directly modify the activity values of cells with $N_{pop,i} > \zeta_K \overline{N_{pop}}$. Then:

$$T'_{Ki} = (T_{Ki})^{\chi K} \qquad \text{Equation 5}$$

with $\overline{N_{pop}}$ the average neighbourhood effect on population in the whole study area, and ζ_K and χ_K parameters to be calibrated. The second alternative makes activity growth directly dependent on the relative neighbourhood effect on population:

$$T'_{Ki} = [\,(N_{pop,i} / \overline{N_{pop}})^{\tau K}]\; T_{Ki} \qquad \text{Equation 6}$$

with τ_K parameters to be calibrated.

A detailed model description can be found in White et al. (2012, 2015) and Crols et al. (2015).

3. Semi-automated calibration and multiple-resolution errors

The different equations were evaluated with a semi-automated calibration framework under development (Crols et al., 2016). Expert-based knowledge has been found important to obtain reliable calibrations of the model to avoid non-sensical, over-calibrated combinations of transition rules (Engelen and White, 2007). Therefore, a calibration procedure was developed in which the modeller

can iteratively define the limits and a best guess of all influence weights and parameters considered. The optimisation is carried out with a genetic algorithm (GA), based on NSGA-II in GAUL (Deb et al., 2002; Adcock, 2005).

The GA fitness function is the inverse of a weighted sum of error functions. We used multiple-resolution error functions, based on Costanza (1989) and Pontius et al. (2004, 2008). Specifically, the RMSE of all activities and of the number of land-use cells of all active categories, is computed at all resolutions of the variable grid, and possibly for administrative regions too, which we did for the Dublin application.

4. Results and discussion

Using density equation 4, the calibration framework succeeded in lowering the regional population error in the Dublin application by 29% compared to the best result of White et al. (2012). The northern and central neighbourhoods of Dublin largely had a declining population in the 1990s while population grew in the southern neighbourhoods. All simulations have difficulties to generate this spatial pattern (Figure 2). Equations 5 and 6 do not cause big changes in Dublin in comparison with equation 4. At the local scale, equation 6 generates a slightly higher overestimation of the population in central Dublin, but a slightly lower residential land-use error. At the regional scale, scattered residential growth in rural areas is overestimated, but the error of population is slightly lower since changes inside smaller remote towns are better predicted. Generally, equation 4 still provides the lowest errors in population at the regional level and at most resolutions (Table 1).

An initial manual calibration of the Belgian application indicated that the population growth was systematically underestimated in the largest cities (around 30% in Brussels) when equation 4 was used. The semi-automated approach was then applied to equations 5 and 6 (Figure 3). Simulations with both equations resulted in a lower and spatially balanced regional error pattern, while the same local error patterns appeared. Within cities, population growth is somewhat underestimated in high-density neighbourhoods and overestimated in low-density neighbourhoods, just as in Dublin. Residential land-use growth in rural areas is more adjacent to existing built-up cells than in reality. Both equations have important disadvantages too. Equation 5 only allocates extra activity to cells that exceed the threshold, which only happens in the cores of the largest cities. Hence, too many other town cores have an underestimated population because of a strong rescaling effect to get the correct population total. Nevertheless, the errors in population at the larger scales are lower (Table 2). Alternatively, equation 6 generates even better calibration results for the population in urban cores (including smaller towns), which decreases the population errors at local resolutions. Unfortunately, the power function is difficult to calibrate and easily leads to unstable behaviour and exaggerated residential growth in some parts of the study area.

To conclude, the best population density equation seems to depend on the study area. The new alternatives solve the specific issue with high growth rates in urban cores but introduce underestimations elsewhere or worse suburban land-use growth predictions. More research is still needed to update the model equations.

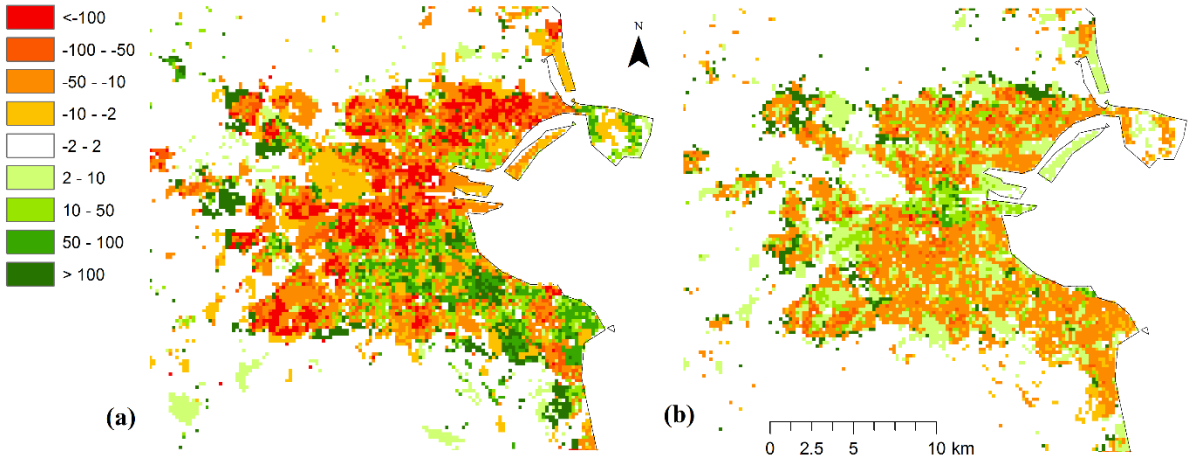

Figure 2: Evolution of population density (per 200 m cell) in the city of Dublin between 1990 and 2000 in (a) reality, and (b) a simulation with the activity-based CA model using equation 4.

Density equation	R	E_1	E_2	E_3	E_4	E_5	E_6
4	0.0077	1.9124	1.3033	0.7707	0.3621	0.0870	0.0104
5	0.0103	1.9688	1.2701	0.7638	0.3736	0.0862	0.0100
6	0.0144	1.9169	1.2857	0.7987	0.3769	0.0838	0.0096

Table 1: Coefficient of variation of the regional RMSE of the different counties of the Greater Dublin Region (R) and errors E_L at different resolutions of the variable grid, given by their level number L (going upwards from the local scale (600 m) with a factor of 3 per level) for simulations with the different population density equations.

Density equation	R	E_1	E_2	E_3	E_4	E_5	E_6
5	0.0141	0.4208	0.2329	0.1172	0.0405	0.0107	0.0026
6	0.0278	0.3665	0.2117	0.1132	0.0518	0.0187	0.0056

Table 2: Coefficient of variation of the regional RMSE of the different counties of Flanders and the Brussels Capital Region (R) and errors E_L at different resolutions of the variable grid, given by their level number L (going upwards from the local scale (300 m) with a factor of 3 per level) for simulations with the different population density equations.

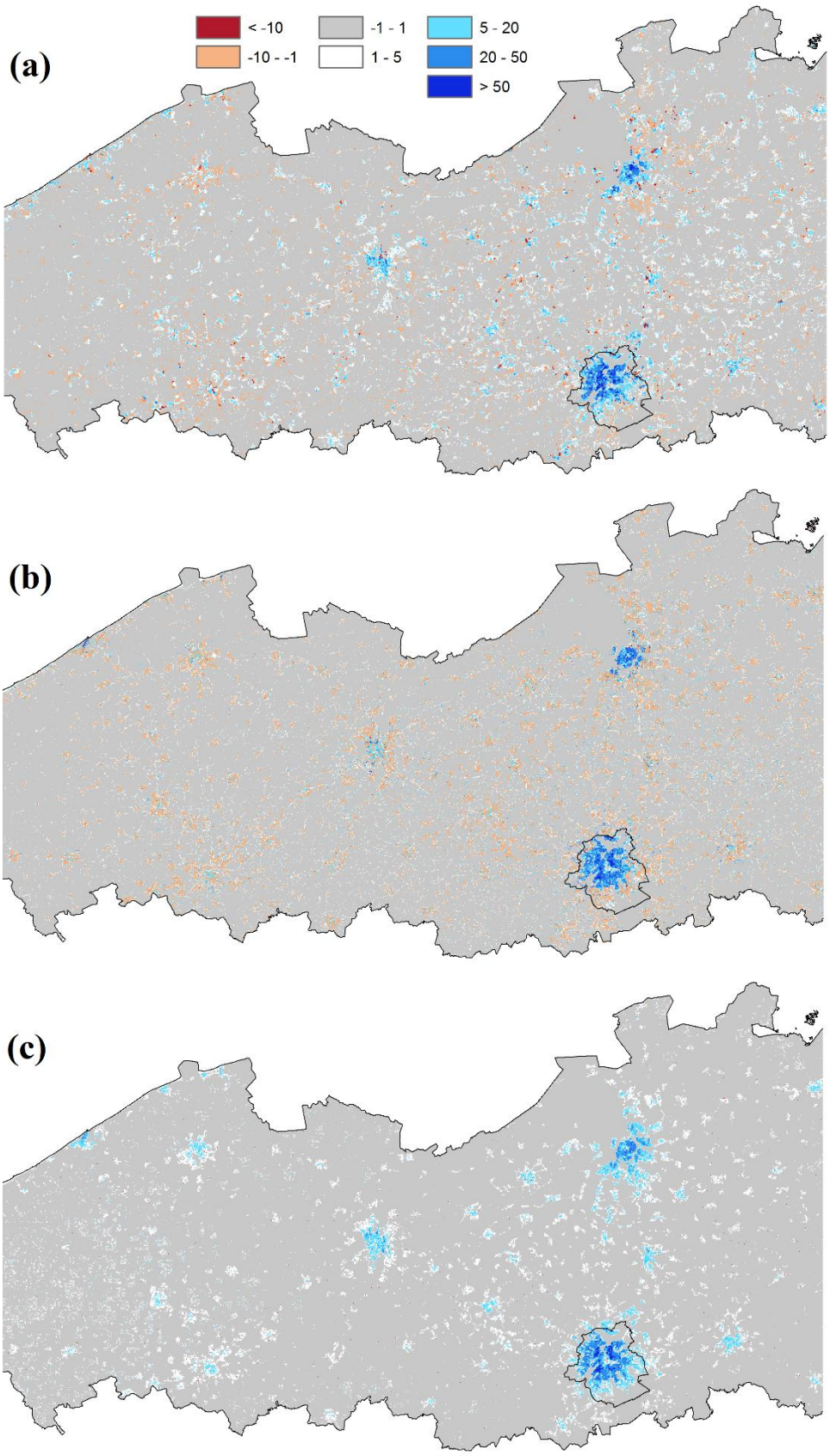

Figure 3: Evolution of population density (per 100 m cell) in Flanders and Brussels between 2001 and 2013 in (a) reality, and simulations with the activity-based CA model with equation (b) 5 or (c) 6.

5. Acknowledgements

This work was supported by a PhD-scholarship financed by the Flemish Institute for Technological Research (VITO), Environmental Modelling Unit, Mol, Belgium.

6. References

Adcock, S. 2005. *The Genetic Algorithm Utility Library (GAUL)*. [Online]. [Accessed 11 May 2016]. Available from: http://gaul.sourceforge.net/index.php

Costanza, R. 1989. Model goodness of fit: a multiple resolution procedure. *Ecological Modelling*. **47**(3), pp.199–215.

Crols, T., Vanderhaegen, S., Canters, F., Engelen, G., Poelmans, L., Uljee, I. and White, R. 2017. Downdating high-resolution population density maps using sealed surface cover time series. *Landscape and Urban Planning*. **160**, pp.96–106.

Crols, T., van der Meulen, M., White, R., Uljee, I., Engelen, G., Poelmans, L. and Canters, F. 2016. Semi-automated calibration of an activity-based CA model. In: Mostafavi, M.A. and Pinto, N.N. eds. *Proceedings of the 2nd International Symposium on Cellular Automata Modeling for Urban and Spatial Systems (CAMUSS 2016), 21-23 September 2016, Québec, Canada*. [Online]. pp. 12–17. Available from: http://www.camuss2016.ulaval.ca/

Crols, T., White, R., Uljee, I., Engelen, G., Poelmans, L. and Canters, F. 2015. A travel time-based variable grid approach for an activity-based cellular automata model. *International Journal of Geographical Information Science*. **29**(10), pp.1757–1781.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*. **6**(2), pp.182–197.

Engelen, G., Lavalle, C., Barredo, J.I., van der Meulen, M. and White, R. 2007. The MOLAND modelling framework for urban and regional land use dynamics. In: Koomen, E., Stillwell, J., Bakema, A. and Scholten, H.J. eds. *Modelling Land-Use Change: Progress and Applications* Dordrecht, Netherlands: Springer, pp. 297–320.

Engelen, G. and White, R. 2007. Validating and calibrating integrated cellular automata based models of land use change. In: Albeverio, S., Andrey, D., Giordano, P. and Vancheri, A. eds. *The Dynamics of Complex Urban Systems*. Heidelberg: Physica-Verlag, pp. 185–211.

Koomen, E. and Stillwell, J. 2007. Modelling land-use change: Theories and methods. In: Koomen, E., Stillwell, J., Bakema, A. and Scholten, H.J. eds. *Modelling Land-Use Change: Progress and Applications* Dordrecht, Netherlands: Springer, pp. 1–21.

Poelmans, L. and Van Rompaey, A. 2010. Complexity and performance of urban expansion models. *Computers, Environment and Urban Systems*. **34**(1), pp.17–27.

Pontius, R.G. Jr., Huffaker, D. and Denman, K. 2004. Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*. **179**(4), pp.445–461.

Pontius, R.G. Jr., Thontteh, O. and Chen, H. 2008. Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics*, **15**(2), pp.111–142.

Santé, I., García, A.M., Miranda, D. and Crecente, R. 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*. **96**(2), pp.108–122.

White, R. 2006. Modelling multi-scale processes in a cellular automata framework. In: Portugali, J. ed. *Complex artificial environments*. Berlin: Springer-Verlag, pp. 165–178.

White, R., Engelen, G. and Uljee, I. 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*. **24**(3), pp.323–343.

White, R., Engelen, G. and Uljee, I. 2015. *Modeling cities and regions as complex systems. From theory to planning applications*. Cambridge, MA, USA and London, UK: The MIT Press.

White, R., Uljee, I. and Engelen, G. 2012. Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *International Journal of Geographical Information Science*. **26**(7), pp.1251–1280.

Williams, B. and Shiels, P. 2002. The expansion of Dublin and the policy implications of dispersal. *Journal of Irish Urban Studies*. **1**(1), pp.1–20.