Spatially-explicit modelling of ecological processes in complex agricultural landscapes: connecting 'artificial' landscapes with 'reality'

Florian Schwarzmueller, **Andrew Hulthen, Justine Murray & Hazel Parry**

CSIRO, GPO Box 2583 Brisbane QLD 4001
Email: florian.schwarzmueller@csiro.au

Abstract: Artificial landscapes are often used in landscape ecological models to investigate questions around movement of individuals, the spread of invasive plants and diseases or the stability of meta-populations. Using these artificial landscapes has some strong benefits such as simplification of landscape complexity, a means to replicate landscape scenarios or the potential to study systematic gradients. Communicating the results to non-theoreticians, however, is often difficult because there is often no quantitative comparison made to link them with reality. Other models that directly use digitized real landscapes as input do not have this barrier, but they are often lacking generalizability or the potential to forecast the effect of changes in the environment.

In this study, we propose a method to generate artificial landscapes with key parameters derived *a priori* from real landscapes through analyses based on Geographical Information Systems (GIS) data: surface cover, spatial aggregation and patch size. Existing methods of generating artificial landscape often do only *a posteriori* comparisons to prove that their landscapes are 'realistic'. We show that, by estimating the parameters beforehand, we can generate artificial landscapes incorporating multiple land use types that can be directly compared to existing landscapes in a quantitative manner. This allows for more targeted landscape generation and vastly reduces the parameter space that needs to be covered. At the same time, this method does not reduce the potential of the models in terms of being reproducible and transferable.

We show an application of this method with two contrasting, complex agricultural landscapes from eastern Sub-Saharan Africa and South-east Australia. We artificially generated landscapes using key parameters from an analysis of digitized real landscapes and found that the algorithm allows flexibility in single target parameters while retaining 'realism'. Realism is assessed in different way. For the Australian data, we compare ranges in land use cover and aggregation between real and generated landscapes. For the African data, we found a log-linear relationship between these two variables in the empirical data that we then used to generate the realistic artificial landscapes. By creating a measurable link between real and artificial landscapes this method will help reduce communication barriers between theoretical scientists and the general public, increasing the impact of our science.

Keywords: *Artificial landscapes, landscape generation algorithm, ecological modelling, Geographical Information System*
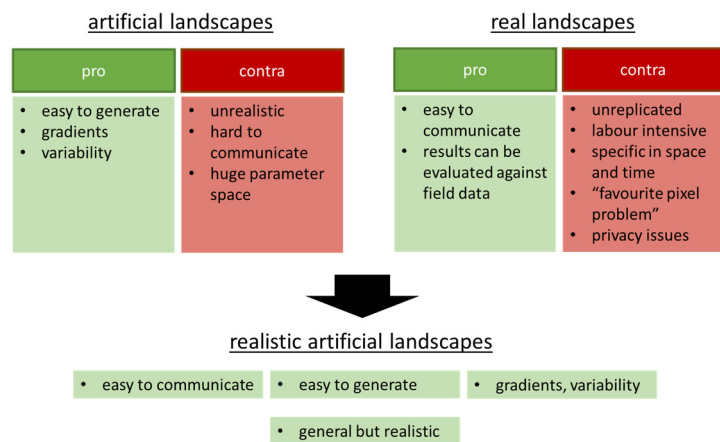
1. INTRODUCTION

Spatially-explicit modelling is used in landscape ecology to look at emerging regional patterns from locally determined rules. Its application covers a very broad range and it has, for example, been used to look at (1) behavioral factors that influence the spatial distribution of species (Parry et al., in press); (2) the success of weeds or diseases spreading in a landscape (Rees et al., 2009); and (3) the stability of whole meta-populations (Tromeur et al., 2016). Naturally, most of the effort in these studies is put on species specific parameters of the model often covering a huge parameter space (for example in Tromeur et al., 2016). However, the other half of the model, the underlying landscape, is often chosen less carefully. This makes it difficult to compare the results of two very specific models or between general models and empirical studies.

Spatially-explicit ecological models either work with artificial landscapes or digitized 'real' landscapes. The choice between the two types is generally influenced by the research question, but also by other factors such as data availability. There is no right or wrong choice, but each method has its pros and cons. In this paper, we present a method to artificially generate realistic landscapes; one that combines benefits from both landscape types.

The use of artificial landscapes is often of interest when investigating general or theoretical concepts (for example Parry et al., in press). They are easy to generate and replicable. By generating many landscapes, modelers can investigate systematic gradients in landscape attributes and thus increase the predictive power of their models. On the downside, results that are obtained from artificial landscapes are often hard to communicate because the landscapes are unrealistic. Even if they are 'realistic', their relatability to empirical landscapes is often not assessed (see examples below). From a modeller's perspective, dealing with purely artificial landscapes also creates a much bigger parameter space which includes landscapes that would never exist in reality. Although this might be interesting from a theoretical point of view, it might shift the focus away from potential real-world applications of the model.

The other possibility in landscape-ecological modelling is the use of digitized empirical landscapes. This is often done by using GIS layers that are then converted in some way into modelling input parameters (see for example Parry et al., in press and 2013). Having a 'real' landscape as input into a model makes it easy to communicate the obtained results. Generally, model findings can also easily be evaluated against field data from these landscapes (for example population dynamics Parry et al., in press). However, real landscapes are generally hard to obtain because they require intense on-the-ground work (although remote sensing can be used for some very general classifications, see Martínez and Mollicone, 2012). Consequently, the number of replicates is often limited and these often only represent snapshots in time (Parry et al., in press). Treating a 'real' landscape as absolute truth can therefore be misleading. To account for this and to assess model sensitivity towards this, researchers would need to introduce some sort of variability into the 'real' landscapes (Thierry et al., 2017). Depending on the research question, there might also be some privacy issues around the use of real data (especially when modelling pest dynamics in a landscape). Finally, models using 'real' landscapes might experience what is called the 'favorite pixel problem' when communicating results. This means that the model quality will be assessed by results for single pixels that the reader knows very well. Ultimately, this can undermine trust in the model if the predictions do not match with the experience of the local expert.

So, when it comes to applied landscape ecological modelling, neither of the two extremes might be a good choice. Rather, input landscapes should combine benefits of both, artificial and real landscapes (Fig. 1). There are some examples in the literature that claim to generate realistic landscapes but they either never assess realism (Hiebeler, 2007), or base it on the 'realistic look' of landscapes (Jackson and Fahrig, 2012). Here we provide a method that analyses existing landscape data beforehand and then uses values of land use cover and aggregation as input parameters for the landscape generation algorithm. By doing

Figure 1. The concept of realistic artificial landscapes

this, we can easily and systematically generate artificial landscapes without losing realism and relatability. In the end, this will reduce communication barriers and potentially increase the impact of the respective models.

2. SYSTEMS AND QUESTIONS

2.1 Queensland Fruit Fly (Qfly, *Bactrocera tryoni*) in Australia's Horticulture

Qfly is one of the major pests for Australia's horticulture sector and once established, the consequences can be dramatic. Adults lay their eggs in fruit across a very broad range of crops, causing decay and premature fruit drop, but most importantly the need for additional phytosanitary measures to prevent the closure of national and international markets (Clarke et al., 2011). Spatially-explicit ecological modelling is used to look at Qfly population dynamics in Australia's horticultural areas and assess the potential of different management options.

There are two main aspects in the biology of Qfly that require corresponding features in the model-landscapes: (1) Qfly is an extremely polyphagous insect. Its host range covers over 150 varieties of commercial fruits and vegetables plus a multitude of native fruit trees. Especially in horticultural areas, where highly suitable fruit is abundant, Qfly populations will utilize several of these hosts and build up populations over the summer. Population decrease, as a consequence of overwintering mortality, will largely be determined by the availability and quality of sheltering sites (better microclimate). In order to account for this, spatially explicit population models require input-landscapes that consist of a multitude of land use types, of which each can potentially have a different resource quality, seasonality and sheltering suitability.

(2) There is some evidence that, when local resource density is high, the movement of Qfly might be very local to almost non-existent (Balagawi et al., 2012). However, when adult females cannot find ripe fruit nearby, they will fly several hundred meters to another, more suitable patch. They might also forage for shelter habitats, adult food resources or overwintering sites (such as urban backyards or a dense forest). In order to investigate this potential sequential use of different resources, input-landscapes have to account for different levels of spatial aggregation.

Although we do have very good land use data, we refrained from using the real landscapes because of two main reasons: (1) identifying individual properties in the context of pest populations might violate individual privacy and (2) the aim of this project is to develop general guidelines on area-wide Qfly management. The use of very special cases might impose additional communication barriers.
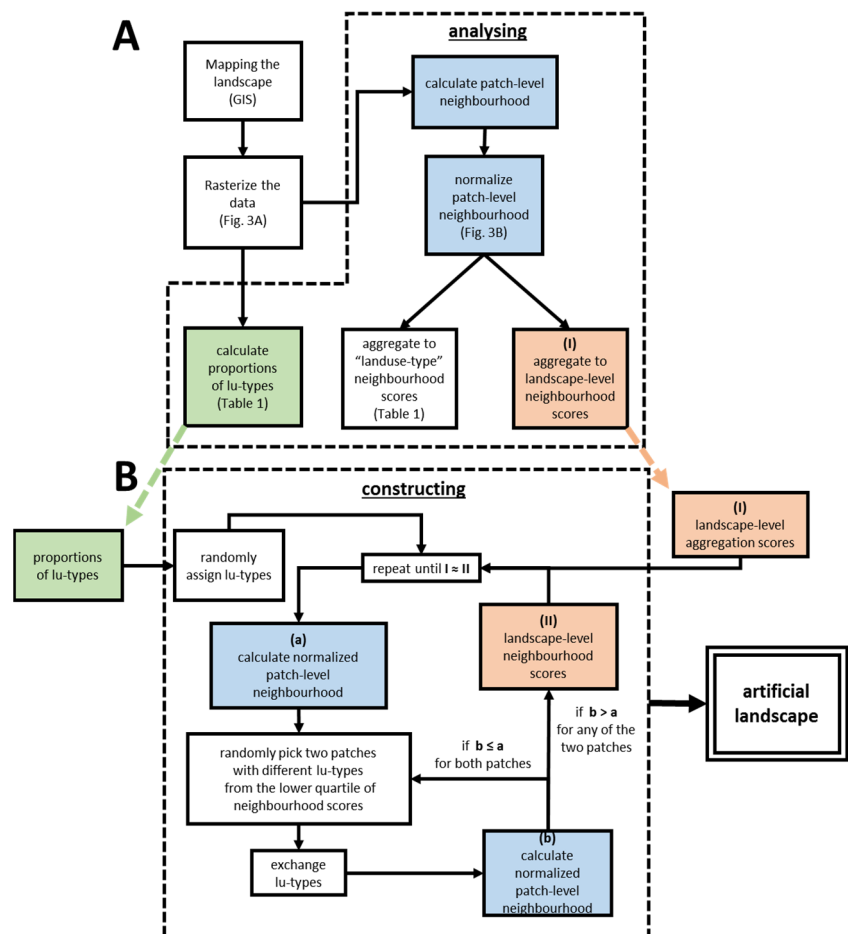

Figure 2. Flow charts that show the steps involved in **(A)** analyzing real landscapes and **(B)** constructing artificial landscapes

2.2 Cassava whitefly in East Sub-Saharan Africa

Cassava is a staple food for people across Sub-Saharan Africa. Over the last 20 years, there has been an increase in the frequency of outbreaks of 'Cassava whitefly' *Bemisia tabaci* species complex (Hemiptera: Aleyrodidae). The species in the Cassava whitefly complex can reduce yields through both direct feeding on and indirect damage to Cassava, in particular as a vector of viruses associated with two important plant diseases: Cassava Mosaic Disease and Cassava Brown-Streak Disease.

Schwarzmueller F. *et al.*, Spatially-explicit modelling of ecological processes in complex agricultural landscapes: connecting 'artificial' landscapes with 'reality'

The species of *Bemisia tabaci* found on Cassava in Africa are thought to be strongly associated with their Cassava host (e.g. Legg et al., 2014; Sseruwagi et al., 2006). Thus, the spatial extent and aggregation of Cassava, along with the management of the crop over time, can be considered to be a key driver of Cassava whitefly dynamics. A spatial simulation model has been constructed to explore how the spatial extent and configuration of Cassava influences this population dynamic (building on De Barro, 2012), as well as different temporal planting regimes (such as the cropping period and synchrony of cropping).

Unlike the case of Qfly in Australia, we have very limited landscape data (and few resources with which to obtain data) at the resolution required to answer these research questions across east Sub-Saharan Africa. A field team has spent two field seasons in Malawi, Tanzania and Uganda collecting data on the Cassava whitefly and mapping crops at a radius of ~100m around a focal Cassava field. This area is too small to meaningfully provide input data to simulate pest behavior across 'landscapes', however we can analyse the digitized data on land cover at this small scale and use it to generate artificial landscapes with our algorithm. This allows us to extrapolate the existing data to much larger areas, and explore key factors in the landscape configuration of Cassava, along with management scenarios, that may influence simulated population dynamics of Cassava whitefly.
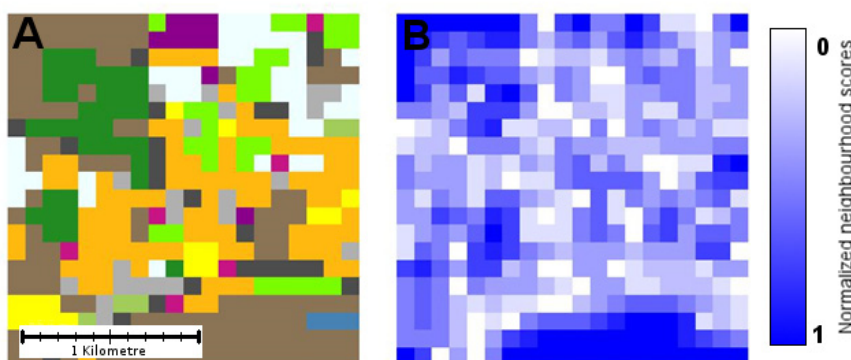
3. METHODS

The landscape generation algorithm needs two input parameters: the proportions of land use types in the landscape and the desired landscape-level aggregation score. These can either be derived from empirical landscapes (as in Fig. 2A and described in section 3.2) or can be chosen freely, for example when generating complexity gradients. In the following, we will briefly describe the type of input data needed by describing the examples used in this study. Then we will go on and describe the features of the proposed algorithm: the landscape analysis and generation.

3.1 Data sources and preparation

For the Australian data, we used vector layers of crop survey data from private organizations comprising data in fruit growing areas from three study regions. This was combined with vector data of the catchment scale state land use layers from the Australian Collaborative Land Use Management Program (ACLUMP, Australian Bureau of Agricultural and Resource Economics and Sciences) to also capture land cover in the surrounding areas. The combined layer was converted into a continuous grid surface with a 100m resolution.

In cassava growing areas in east Africa, land use was digitized in the field using ArcGIS Collector (ESRI 2015. ArcGIS Desktop: Release 10.4. Redlands, CA: Environmental Systems Research Institute) and assigned to one of 15 land use types at a radius of 100m from the edge of each focal cassava field. Vector data from each region was then converted to a grid using ArcGIS. Cell size of the grid can be chosen depending on the desired size and resolution. These are often dependent on the specific biological questions but a good proxy is to select a value in the lower quartile of patch sizes in the landscape measured using a GIS.

For our two examples, we picked a 100x100m resolution for the Australian data and a 10x10m resolution for the African data (fields generally are much smaller in Africa than Australia).

3.2 Analyzing

We used the rasterized data as input for the neighborhood analysis (Fig.2A). Neighborhood scores were calculated for each grid cell p as the

Figure 3. Example of a real Australian landscape (2x2 km) with a landscape-level aggregation score of 0.46. **(A)** shows the rasterized land use data and **(B)** the normalized neighborhood scores for each cell. The percentages of cover for each land use-type and the respective 'land use-type'-level aggregation scores are given in Table 1.

number of neighboring cells with the same land use type as p. These neighborhood scores are then normalized by the total number of neighbors. In our examples, we used a square grid with an eight-neighbor rule. This means, neighborhood scores were divided by eight for inner cells, five for edge cells, and three for the four corner-cells, respectively. The normalized neighborhood scores are then aggregated to 'land use-type'

neighborhood scores or a landscape-level aggregation score by calculating the respective means. Table 1 shows the means of aggregation scores of different land use types in the example landscape presented in Fig.3A.

Table 1. Cover and aggregation of the different land use types in the example landscape presented in Fig.3. Land use types include: Native vegetation, Pasture, Trees, Vegetables, Other fruit trees, Grapevines, N/A (empty), other hosts, Structures, Urban, Water, Citrus, Summer Citrus, Stonefruit, Pomefruit.

Land use type	Nat. veg.	Past.	Trees	Veg.	Oth. fruit	Grap.	N/A	oth. hosts	Struc.	Urb.	Wat.	Citr.	Sum. Citr.	Stone fruit	Pome fruit
color code															
% of cells	0	30.3	9.25	2.25	1.5	7.25	9.5	1.25	6.0	5.5	0.75	22.5	4.0	0	0
Aggregation	NA	0.65	0.53	0.46	0.00	0.33	0.41	0.12	0.08	0.15	0.19	0.48	0.36	NA	NA

3.3 Constructing artificial landscapes

In a first step, land use types are assigned randomly to each cell in a landscape. By doing this with a given probability (for example the proportions in a real landscape), generated landscapes already include some variability. The first landscape (generated at random) is then analyzed in terms of its overall aggregation. If the value is below the desired aggregation score (usually values of random landscapes are very low, see Fig. 5), rearrangement begins. Two cells with different land use types are selected at random from the lower quartile of cell-level neighborhood scores and their land use types are exchanged. In this way, we ensure that the overall proportions of these land use types do not change.
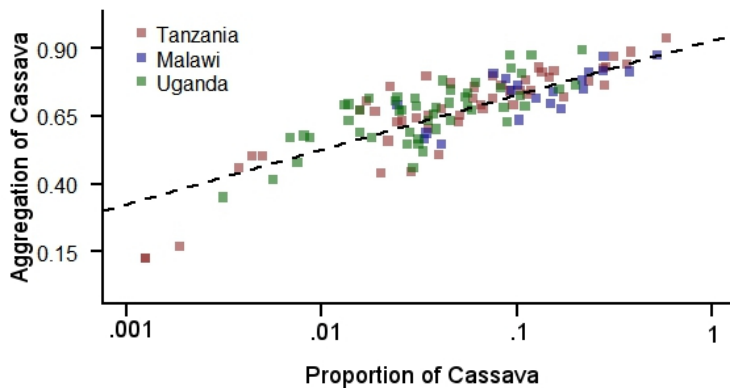

Figure 4. Relation between the proportion of Cassava and its aggregation in the real African landscapes. Different colors represent the different countries in which the landscapes were sampled. The dotted line is the result from a linear model with y= 0.2x + 0.93 (x= log10(proportion of Cassava)).

Then, the algorithm evaluates whether the exchange was a step towards a higher aggregation. To reduce computational effort and to reduce the possibility of a dead-end in the optimization, this evaluation is done on the scale of the two exchanged cells. If the local neighborhood score for any of the two cells was increased by the exchange, the new arrangement is accepted. If the new local neighborhood scores are the same or smaller for both cells, the change is rejected and two new cells are selected. After every tenth optimization step, the new landscape-level aggregation score is calculated and compared to the input value. This is repeated until the score of the new landscape is greater than or equal to the input value.

The output generated by the algorithm includes: a list of the land use type of each cell in the new landscape; a map of the new landscape and the neighborhood scores for each cell (Fig.6); the percent of cover for each land use type and the respective 'land use-type' aggregation scores (Table 2); and finally, the new landscape-level aggregation score.
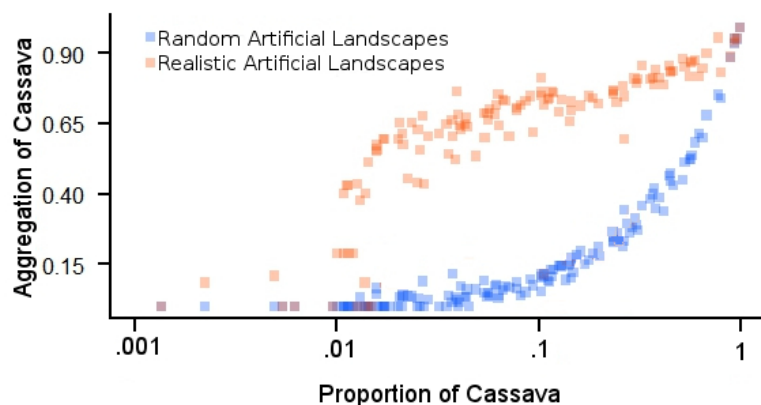
Figure 5. relationship between total proportion of a land use-type in a landscape and its aggregation score based on 150 random landscapes compared to their 'realistic' counterparts using the relationship in Fig.4

961

Schwarzmueller F. *et al.*, Spatially-explicit modelling of ecological processes in complex agricultural landscapes: connecting 'artificial' landscapes with 'reality'

4.   RESULTS

The analysis part of the tool (Fig. 2A) can be used to characterize landscapes in terms of their composition and aggregation. This can, for example, be used to interpolate from a limited sample-size. In the example of the African data, we used it to look at the relationship between the amount of Cassava in the landscape and its aggregation (Fig. 4). We found a log-linear relationship between them, within and across regions. We used this relationship, allowed 10% deviation from the exact value, and generated 150 'realistic' artificial landscapes

from this relationship. Those vary dramatically from purely random landscapes with the same proportion of Cassava as input variable (see Fig.5). By using the new algorithm, we can increase the number and size of landscapes for our simulations while maintaining the empirical relationship found in the real data.
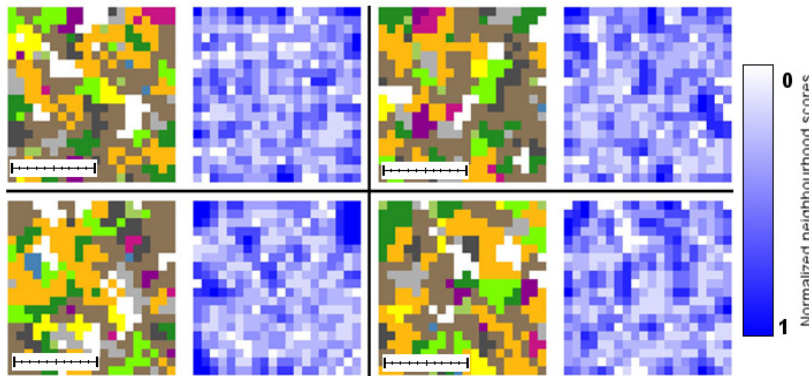
Figure 6. Four examples of artificial landscapes generated from the real landscape presented in Fig.3. All landscapes have an overall landscape-level aggregation score of around 0.46 (0.16-0.19 for random landscapes). The range of corresponding percentages of cover for each land use type and the respective "land use type" aggregation scores are given in Table 2. Scale indicates 1

For the Australian data, we generated ten realistic artificial landscapes for each real landscape. As an example, we present four artificial landscapes generated from the same landscape in Fig.3 (see Fig. 6). The range of

values for cover and aggregation of the ten landscapes are given in Table 2. By using these as input landscapes for our spatially explicit population models, we increase the generalizability of our modelling results. Increasing the variability allows for the testing of model sensitivity which also will help us to account for uncertainty in the data (such as errors or changes in the land use data). On top of this, we can ensure that we are not violating privacy by giving out sensitive information. However, we still keep communities and stakeholders engaged because they can relate to the landscapes and we directly asses in which aspects these represent reality.

Table 2. Ranges in land use type cover and aggregation for the ten artificial landscapes, generated from the real landscape presented in Fig.3. Rows three and four indicate the range of these values across the ten generated landscapes (top= minimum, bottom=maximum).

Landuse type	Nat. veg.	Past.	Trees	Veg.	Oth. fruit	Grap.	N/A	oth. hosts	Struc.	Urb.	Wat.	Citr.	Sum. Citr.	Stone fruit	Pome fruit
color code															
% of cells	0	25.8 – 36.5	7.5 – 10.8	0.75 – 3.0	0.75 – 3.0	5.75 – 9.75	6.7 – 10.8	1.0 – 2.25	4.75 – 8.75	4.25 – 7.25	0.25 – 1.50	18.5 – 27.8	2.0 – 6.25	0	0
Aggregation	NA	0.51 - 0.59	0.34 - 0.54	0 - 0.46	0 - 0.58	0.33 - 0.49	0.36 - 0.46	0 - 0.38	0.27 - 0.48	0.30 - 0.46	0 - 0.4	0.46 - 0.58	0.23 - 0.46	NA	NA

5.   DISCUSSION AND CONCLUSIONS

The use of artificial landscapes is extremely common in spatially-explicit ecosystem models. In order to communicate findings to non-theoreticians it can help to make these landscapes 'realistic'. Here, we present a method that estimates landscape metrics from empirical landscapes and uses them as an input into a landscape generation algorithm. By using the same technique to analyze both real and artificial landscapes, we can quantify the 'realism' of the newly generated landscapes in terms of landuse cover and aggregation.

This direct evaluation is what makes this method uniquely different from already existing landscape generation algorithms such as the one proposed in Hiebeler (2007). Hiebeler also generates landscapes at random, then selects two cells, exchanges their land use types and finally evaluates whether to keep or reject the new landscapes. The method in Hiebeler (2007), however, only considers two land use types - 'suitable' and 'unsuitable' – and upscaling it to multiple types is not easily achieved. When generating landscapes with a multitude of different land use types (such as the Australian data), the output landscapes show substantial

Schwarzmueller F. *et al.*, Spatially-explicit modelling of ecological processes in complex agricultural landscapes: connecting 'artificial' landscapes with 'reality'

variability in the aggregation score of different land use types. If the focus of the modelling is on the clusteredness of one particular land use type, changing the method accordingly is possible. This can, for example, be done by a hierarchical landscape generation where first only one land use type is assigned accordingly and when its aggregation score is reached, the rest of the landscape is assembled.

The landscape generation algorithm can potentially reach an 'optimization dead-end' when the further exchange of land use types cannot lead to a higher clustering score. This is especially true for very ordered landscapes. However, in our work, this only occurred at the high end of aggregation gradients and not when using empirical data as input. If this happens regularly, one could change the rule after which the exchanged cells are selected, or switch to other algorithms for landscape generation that can create more ordered structures (such as in Jackson and Fahrig (2012)).

Artificial landscapes are a useful tool in landscape ecological modelling. Generating them in a way that allows quantification of their realism, as proposed in this paper, will help reducing communication barriers and can potentially increase the impact of adaptive models.

ACKNOWLEDGEMENTS

REFERENCES

Balagawi, S., Jackson, K., Hamacek, E.L., Clarke, A.R. (2012). Spatial and temporal foraging patterns of Queensland fruit fly, *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae), for protein and implications for management: Qfly foraging for protein in the field. Aust. J. Entomol. 51, 279–288. doi:10.1111/j.1440-6055.2012.00863.x

Clarke, A.R., Powell, K.S., Weldon, C.W., Taylor, P.W. (2011). The ecology of Bactrocera tryoni (Diptera: Tephritidae): what do we know to assist pest management? Ann. Appl. Biol. 158, 26–54. doi:10.1111/j.1744-7348.2010.00448.x

De Barro, P.J. (2012). Getting the most out of Eretmocerus hayati, an effective natural enemy of silverleaf whitefly. (Final Report No. VG0B051). Horticulture Australia Ltd.

Hiebeler, D.E. (2007). Competing populations on fragmented landscapes with spatially structured heterogeneities: improved landscape generation and mixed dispersal strategies. J. Math. Biol. 54, 337–356. doi:10.1007/s00285-006-0054-6

Jackson, H.B., Fahrig, L. (2012). What size is a biologically relevant landscape? Landsc. Ecol. 27, 929–941. doi:10.1007/s10980-012-9757-9

Legg, J.P., Shirima, R., Tajebe, L.S., Guastella, D., Boniface, S., Jeremiah, S., Nsami, E., Chikoti, P., Rapisarda, C. (2014). Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. Pest Manag. Sci. 70, 1446–1453. doi:10.1002/ps.3793

Martínez, S., Mollicone, D., 2012. From land cover to land use: a methodology to assess land use from remote sensing data. Remote Sens. 4, 1024–1045. doi:10.3390/rs4041024

Parry, H.R., Bianchi, F., Schellhorn, N.A. (2013). The use of models to explore IPM strategies and design pest suppressive landscapes for sustainable agricultural practice. 20th Int. Congr. Model. Simul. Modsim2013 1847–1853.

Parry, H.R., Paull, C., Zalucki, M.P., Ives, A., Hulthen, A., Schellhorn, N.A., in press. Estimating the landscape distribution of eggs by Helicoverpa spp.,with implications for Bt resistance management. Ecol. Model.

Rees, E.E., Pond, B.A., Cullingham, C.I., Tinline, R.R., Ball, D., Kyle, C.J., White, B.N. (2009). Landscape modelling spatial bottlenecks: implications for raccoon rabies disease spread. Biol. Lett. 5, 387–390. doi:10.1098/rsbl.2009.0094

Sseruwagi, P., Maruthi, M. N., Colvin, J., Rey, M. E. C., Brown, J. K., Legg, J. P. (2006). Colonization of non-cassava plant species by cassava whiteflies (Bemisia tabaci) in Uganda. Entomol. Exp. Appl. 119, 145–153. doi:10.1111/j.1570-7458.2006.00402.x

Thierry, H., Vialatte, A., Choisis, J.-P., Gaudou, B., Parry, H., Monteil, C. (2017). Simulating spatially-explicit crop dynamics of agricultural landscapes: The ATLAS simulator. Ecol. Inform. 40, 62–80. doi:10.1016/j.ecoinf.2017.05.006

Tromeur, E., Rudolf, L., Gross, T. (2016). Impact of dispersal on the stability of metapopulations. J. Theor. Biol. 392, 1–11. doi:10.1016/j.jtbi.2015.11.029