EVALUATION AND CALIBRATION OF AN AGENT BASED LAND USE MODEL USING MACHINE LEARNING AND REMOTELY SENSED LAND COVER DATA

Bumsuk Seo, Calum Brown, and Mark Rounsevell

Land Use Change and Climate Research Group, Karlsruhe Institute of Technology,
Institute of Meteorology and Climate Research - Atmospheric Environmental Research

ABSTRACT

The evaluation of advanced land system models is critical for accurate assessments of global change scenarios. However, model evaluation is a complex task, especially in the case of agent-based and similar modelling approaches that provide flexible descriptions of land use decision-making. Machine learning has the potential to become a powerful technique for calibrating and evaluating these models, bringing large datasets to bear while exploring large areas of parameter space. We present an initial application of machine learning to land use model development, quantifying correspondence between simulated European land uses from a model (CRAFTY-EU) and remotely sensed MODIS land cover. We used grid searching and a more efficient searching algorithm to calibrate three behavioural parameters of the model on the basis of agreement in the spatial properties of land cover. The computational cost of the grid searching was high (n=486), while the searching algorithm yielded similar results with approximately 5% of the runs. We use our findings to propose a calibration and evaluation framework that holds promise for improving the accuracy, utility and transparency of land use decision models, and identify important areas for further work, including the design of more rigorous stopping rules and convergence statistics.

Index Terms— model evaluation, model calibration, machine learning, Model-Based Optimisation

1. INTRODUCTION

Understanding the likely and potential roles of land use in future global change is key to integrated assessments supporting international policy development [1, 2, 3, 4]. However, relatively few models are capable of simulating land use over the spatial and sectoral scales required, and these may include unrealistic representations of land system dynamics [5, 6, 7]. Furthermore, different models produce widely divergent projections, for reasons that are not entirely understood [8]. As a result, a broader range of rigorously-specified models is needed to deepen our understanding of potential land use change.

Behavioural or 'process-based' models such as agent-based models (ABMs) have received increasing attention because they allow the simulation of emergent social and economic conditions from underlying behavioural processes [9]. Unlike other models, ABMs do not attempt to achieve tractable aggregate model outcomes but represent land users as heterogeneous, boundedly-rational agents who act and interact according to characteristic protocols of communication and decision-making [10]. Thus, ABMs have a distinct and important role to play in complex land resource policy analysis. Nevertheless, the flexibility of these models means that they can be especially hard to calibrate and evaluate, because a great deal of data is required to constrain their parameter values without over-fitting to limited, unrepresentative or aggregate outcomes [11].

The development of satellite technology may provide a significant step towards satisfying these data requirements. In recent years, more sensors have become operational with higher spatial and spectral detail, and with shorter acquisition intervals. We propose to take advantage of these developments through a satellite-based evaluation framework for assessing land use models. In this study, we conduct an initial comparative analysis between simulated and remotely sensed land cover in order to explore the methodological and practical benefits of such a framework. Land use simulations were carried out using an ABM of European land use change (CRAFTY- EU) [12, 13]. For the reference data, we used a satellite-based land cover product. We then compared information contained in the simulated and remotely-sensed datasets concerning the spatial properties of different land covers. By using robust metrics, we try to overcome discrepancies in classification schemes and spatial resolutions between the two datasets. We then used grid searching and a Bayesian stepwise optimisation algorithm to identify the computationally feasible strategy of model calibration.

2. CALIBRATION EFFORTS IN LAND USE MODELLING

The calibration of land use models is challenging both conceptually and practically. As representations of complex human-environment systems, land use models invariably

have a relatively limited, artificially delineated area of focus, even within which they must make extensive simplifications. The great diversity of relevant theories in social, ecological and land use sciences allows the development of numerous alternative models of single aspects of study systems [11, 14], while the various contingencies and fundamentally behavioural nature of land use change makes the gathering and use of calibration data extremely difficult [15, 16, 11].

Many land use models, including many of the most established, overcome these problems to some extent by adopting tightly constrained approaches. Assumptions about the equilibrium of demand and supply or economic optimisation greatly increase model tractability and reduce the data and computational requirements of calibration. They also usually imply an interest in aggregate trends or patterns that represent clear 'targets' for calibration to reach. However, such approaches still face problems ranging from the identification of suitable observational data to the risk of over-fitting to historical trends that may not be relevant in the future [17].

The challenges of calibration are even greater for ABMs and similar models. These models include more detailed representations of underlying processes of land use change and therefore, in principle, can be more directly calibrated from real-world analogues [9, 17]. However, such models tend to have large numbers of parameters, not all (or any) of which necessarily have clear real-world counterparts, and can potentially be calibrated against numerous potentially unclear criteria. This makes spurious fits to observations easy to generate, and multiple observations and calibration methods necessary to avoid over-fitting and tease apart different processes [18, 17].

Currently, the dominant practice across land use modelling is to tune models to replicate specific observations, with limited exploration of alternative parameterisations or their implications for understanding of the modelled system [16, 19]. However, recognition of the need for greater behavioural realism along with advances in the computational efficiency of behavioural models such as ABMs has meant that new methods for robust model calibration are being sought. These methods must be capable of exploring large areas of parameter space efficiently, and can therefore benefit from statistical methods as well as those applied in other scientific fields. Of particular relevance may be techniques from economic modelling, in which ABMs and other models have been calibrated through various advanced computational techniques [20, 21, 22]. Perhaps most promising are calibration techniques based on machine learning. These techniques utilise recent developments in artificial intelligence to rigorously test alternative parameterisations, generating detailed understanding of model behaviour as well as allowing models to be tuned to replicate a range of different observations. Such techniques do not overcome the more conceptual questions of where 'sensible' or informative parameterisations are located, or how closely and reliably model outputs should approach particular obser-

vations [17, 23], but they arm the modeller with a great deal of information with which to tackle these questions.

Machine learning has been used to a limited extent in economic agent-based modelling [22]. It has also been used more extensively to derive land cover maps from satellite data and to establish simple land use models from these [24, 25, 26, 27, 28, 29]. However, the transfer of these techniques beyond relatively simple cellular automata models has been slow, with no previous examples of large-scale or agent-based land use models being calibrated in this way. We present an initial step towards the full utilisation of machine learning in land use modelling by performing a limited calibration of an ABM of European land use change.

3. MATERIALS AND METHODS

3.1. Land use simulation using CRAFTY-EU

The model CRAFTY-EU simulates the interplay among land use agents to represent land use decision-making processes. Land ownership is allocated via competition based on 'benefit' values that express the societal value of ecosystem service provision through agents' land uses. The model simulates EU-level land use represented by 17 agent functional types (AFTs) (Table 1). Initial AFT information was derived from the CORINE 2006 land cover data [30]. Full details of the operation of the CRAFTY modelling framework can be found in [12, 23, 31]

Table 1. Agent Functional Types (AFTs) in CRAFTY-EU[32]

Agent Functional Type	CODE	Description
Intensive arable farming	IA	Intensively farmed area
Intensive pastoral farming	IP	Intensively grass
Intensive agro-forestry mosaic	Int_AF	Intensively farmed, intensively grass, managed forest
Intensive farming	Int_Fa	Intensively farmed, intensively grass
Managed forestry	MF	Managed forest
Mixed farming	Mix_Fa	Intensively farmed, intensively grass, extensively grass
Mixed pastoral farming	Mix_P	Intensively grass, extensively grass, very extensively grass
Mixed forest	Mix_For	Managed forest, unmanaged forest
Extensive pastoral farming	EP	Extensively grass
Extensive agro-forestry mosaic	Ext_AF	Extensively grass, very extensively grass, managed forest
Very extensive pastoral farming	VEP	Very extensively grass
Multifunctional	Multifun	Four or more land uses in uncommon combination
Minimal management	Min_man	Very extensively grass, unmanaged forest, unmanaged land
Unmanaged land	UL	Unmanaged land
Unmanaged forest	UMF	Unmanaged forest
Peri-urban	P-Ur	Any combination with $> 40\%$ urban area
Urban	Ur	Urban

We focus on three major behavioural parameters that play important roles in the allocation process (Table 2). The 'Giving-in (GI)' threshold controls how easily an agent relinquishes land ownership to another agent in a competition. Second, the 'Giving-up (GU)' threshold represents how easily an agent abandons land ownership if its benefit value (akin to profit level but also incorporating non-monetary reward) becomes small. Thus these two parameters affect the speed, extent and nature of land use change. Third, we vary the 'Service Level Noise Maximum (SLN)' parameter, which controls the variation in levels of ecosystem service provision between individual agents. Each agent transforms their

land holding's '*capitals*' to '*services*' according to a defined production function, to which some noise is added [12]. The parameter 'Service Level Noise Maximum (*SLN*)' indicates the magnitude of the noise added in this step. The service level feeds back to the AFT allocation competition, thus indirectly affecting land ownership changes.

Table 2. Target parameters in the calibration. The threshold parameters directly control the land use allocation. The noise parameter indirectly affects the allocation by increasing uncertainty to the stochastic land use decision-making process[31]. Negative values of the thresholds indicate agent persistence even in the theoretical case of receiving disbenefit from land management.

Parameter	Description	Potential range	Target range
Giving-in threshold (*GI*)	Tendency of changing land use	$(-inf, inf)$	$[-2, 2]$
Giving-up threshold (*GU*)	Tendency of abandoning land	$(-inf, inf)$	$[-2, 2]$
Service Level Noise Maximum (*SLN*)	Noise in evaluating the value of land use	$[0, 1]$	$[0, 1]$

3.2. Model Evaluation using Mutual Information and Fractal Dimension

For EU-28 countries, we simulated AFT dynamics under static climatic and socio-economic conditions and compared resultant land use changes with MODIS Land Cover Type (2006–2013) (Fig. 1). We used the MODIS MCD12Q1 annual land cover product at 500 m resolution. Among the land cover type information enclosed in this product, we retrieved the IGBP 17-class land cover information. This land cover (LC) data was reprojected into a 1 km base grid using the nearest neighbor algorithm. The AFT data was projected into a 15-km base grid in ETRS89/ETRS-LAES space (EPSG:3035). The CRAFTY and MODIS datasets therefore differ in pixel size (15 km vs. 1 km) and land cover schemes (CRAFTY 17 AFTs vs. IGBP 17 land cover classes) (Fig. 1), which makes it difficult to use a traditional contingency table for comparison. Instead, we used Mutual Information (MI) [33], which is robust to such differences. MI is a general measure of dependency between random variables and used here to measure 'information' between the simulated and the observed land uses regardless of their unequal classification schemes [34]. Additionally, we calculated fractal dimension [35] of the land use datasets to measure the spatial complexity of it, and here we try to match these complexities between the two datasets.

3.3. Stepwise calibration of the land use model using Bayesian optimisation algorithm

Grid searching is generally infeasible for land use models due to its computational cost, especially at large scales. In this study, we also used an efficient parameter space searching algorithm to see whether it produced the same calibration result, and therefore whether we can safely replace a costly grid
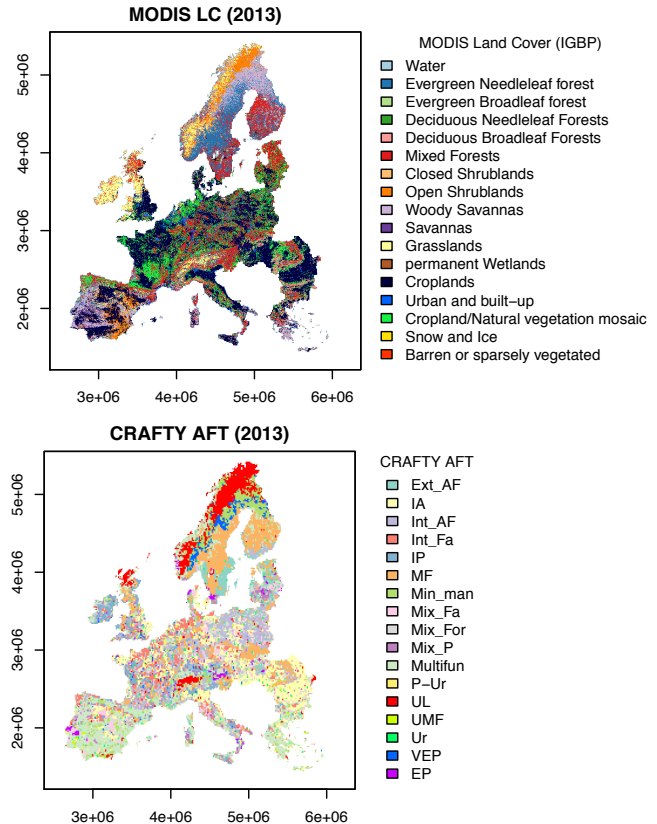

Fig. 1. MODIS MCD12Q1 Land Cover (upper) and simulated AFT (lower) for EU-28 countries

search with this algorithm. We used a Model-Based Optimisation (MBO) algorithm proposed in [36], designed to replace 'expensive' black-box functions. To minimize the cost of exploring a high-dimensional parameter space, the algorithm uses a specially designed Bayesian parameter proposal function, which identifies the parameter regions with higher probabilities of containing optimal parameter combinations. In its iterative process, the algorithm samples a new parameter proposal based on the '*infill criterion* (i.e., a sampling strategy designed to offer improvement)' [37] and the recent evaluation metrics. Then the proposed parameter set is evaluated and the result used to update the MBO model.

4. RESULTS

4.1. Parameter tuning using grid searching and stepwise calibration

We ran CRAFTY-EU for the eight-year period (2006–2013) and compared the simulated land uses under different parameterisations with the MODIS LC. For each year, we calculated MI between the two datasets at the MODIS 1 km grid and then averaged across cells. The total number of individual runs was 486 and it took 48 hours in a single node (Intel
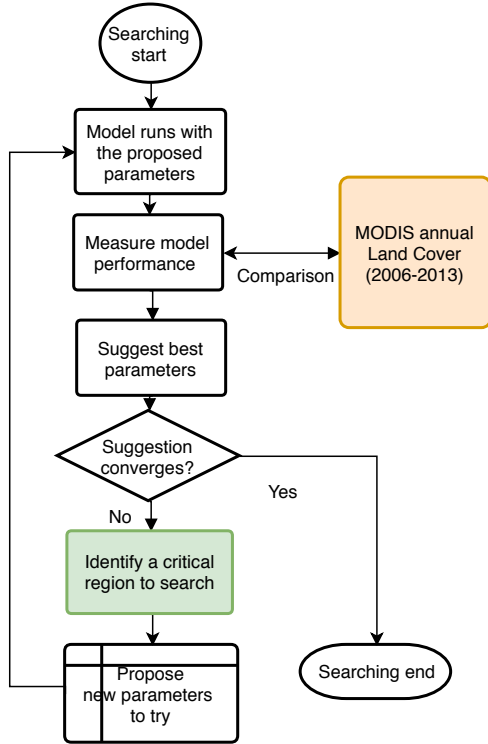
Fig. 2. Flow diagram of the stepwise calibration using a Bayesian Model-Based Optimisation (MBO) algorithm

Fig. 3. Average Mutual Information (MI) between MODIS Land Cover (MCD12Q1) and the simulated land cover for 2006–2013 and the 17 AFTs with respect to the three parameters: *GI*, *GU*, and *SLN*. The contours are estimated based on the runs from the grid searching (n=486).

Core i7 2.8Ghz). As shown in Fig. 3, the impacts of changing *GU* and *SLN* were pronounced in the Mutual Information, whereas the changes of *GI* did not seem to affect the Mutual Information systematically. This implies that land managers' individual production levels (*SLN*) and willingness to abandoning their land holdings (*GU*) are more important parameters than the tendency for changing land use (*GI*) in calibrating CRAFTY-EU. At the same time, a relatively wide range of parameter values had comparable levels of accuracy (Fig. 4). High MI was found around *GU*= 1, suggesting that moderate irrationality in GU is reasonable (i.e. agents abandon land even though they derive benefit from it). The multiple optima in the Figure indicate, though, that when *SLN* is high, perfect rationality (*GU*= 0) yielded good results.

To propose parameter values considering all three dimensions, we took the top-10 combinations with the highest Mutual Information and summarised them in Fig. 4. The parameter combination scoring the highest MI was *GI*=0.50, *GU*=0.50, and *SLN*=0.20. The mean parameter values of the top-10 samples were *GI*=1.15, *GU*=0.70, and *SLN*=0.48.

To minimize the cost of exploring a high-dimensional parameter space, we applied MBO algorithm to find the CRAFTY parameters maximising the Mutual Information between the simulated land use and the remotely sensed land cover (Fig. 2). At the beginning, 10 randomly sampled runs
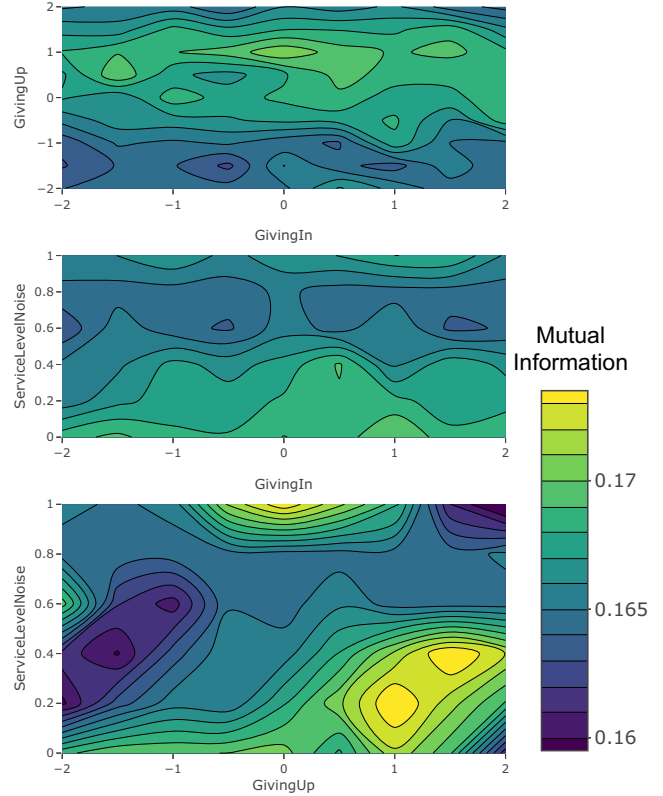
from the batch runs (n=486) were used to initialise the optimisation algorithm. The algorithm ran over 200 iterations and summarised the suggested parameter values based on the last 50 iterations. The proposals were evolved to converge to the those of the grid searching after approximately 30 iterations (Fig. 5). The proposed parameter values (*GI*=0.54, *GU*=0.75, and *SLN*=0.30) overlap well with the parameter values suggested by the grid searching (Fig. 4). The computing time was around $5(=30/486)\%$ of the time required for the grid searching.

4.2. Spatial complexity of the simulated land use

The fractal dimension of the simulated land use is calculated for the eight years of the comparison (2006–2013). Note that this complexity measure is calculated independently in the MODIS LC. The average fractal dimension of the 17 CRAFTY-EU AFTs for the eight year (Fig. 6) were generally lower than that of the MODIS data ($=1.015$). Although it is indecisive to say that the simulated land cover is spatially simpler than the MODIS land cover because the intrinsic spatial resolution were different (1 km MODIS vs. 15 km
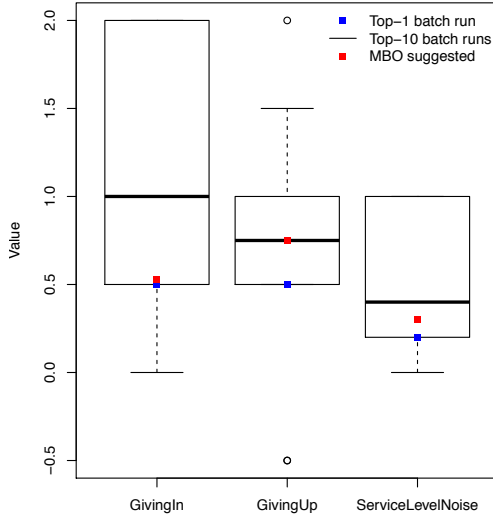
Fig. 4. Optimal parameter values suggested using the grid searching (black boxes) and the MBO algorithm (red squares). The top-10 runs with the highest Mutual Information were used to derive the optimal parameter candidates. The MBO-based values are the mean values from the last 50 iterations (Fig. 5).

CRAFTY-AFTs). The fractal dimension was affected by the parameters but with little variation, perhaps because the results account for the eight-year means. Once again, the *GI* parameter had especially small effects, and did not change fractal dimension systematically. Nevertheless, there are visible patterns: relatively high agreement was found around *GU¿* 1.5 and *SLN¡* 0.6. *SLN* made interesting patterns in fractal dimension as it significantly increased fractal dimension when moderate (0.2-0.6); i.e., when *SLN* was too high or too low, the landscape became simpler.

To examine the effect of parameter calibration on spatial complexity, the average fractal dimension of the model runs using calibrated parameters were compared (Fig. 7. The average fractal dimension of the Top-1 batch run was around the 75_{th} percentile of the fractal dimension of all batch runs (blue square). The fractal dimension of the MBO-suggested model run (red square) was above the median fractal dimension as well as the average fractal dimension of the Top-10 batch runs.

5. DISCUSSION

In this study, we proposed the use of new types of correspondence metrics for calibrating a complex land use model, and showed that these metrics can be used to identify best-performing values for behavioural parameters of the model. The machine learning approach adopted here holds considerable promise for the calibration and evaluation of land use
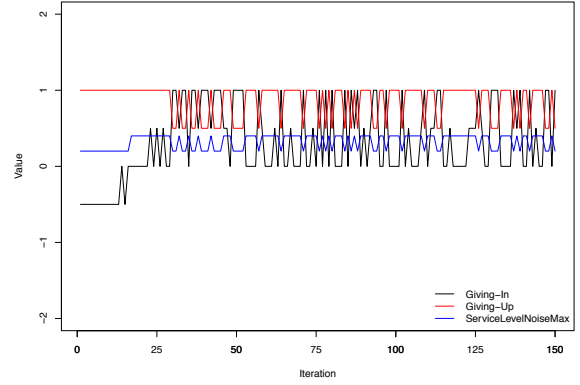

Fig. 5. Changes of the suggested parameter values using MBO algorithm. The parameter suggestion for '*GI*' threshold (black) converges to its stationary value (=0.5) after 30 iterations, approximately.
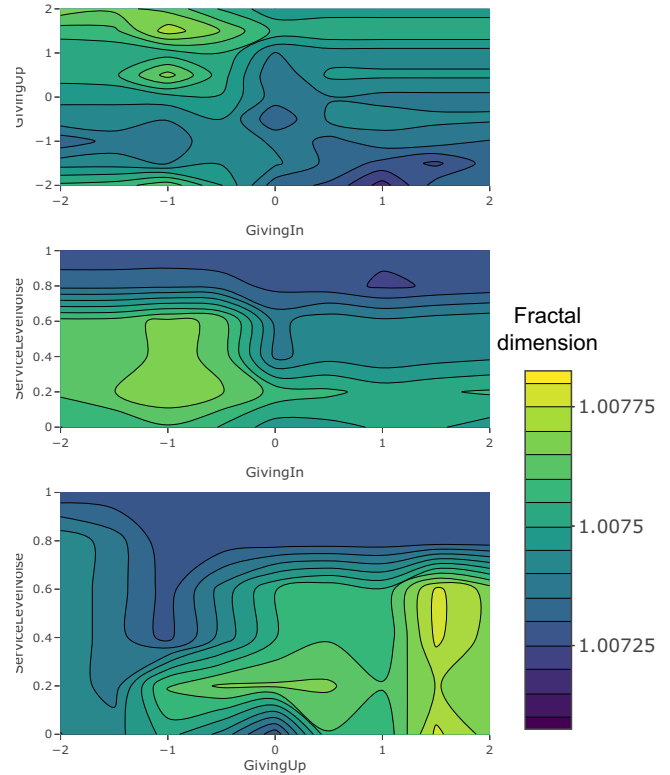

Fig. 6. Average fractal dimension of the simulated CRAFTY land cover for 2006–2013 and the 17 AFTs

models, allowing the application of large remotely-sensed datasets (or, in principle, other 'big' data concerning individual or social behaviour relevant to land system dynamics) to refine model parameterisations. Developing statistically sounds methods for utilising these data sources is a necessary step for rigorous understanding of land system dynamics, and
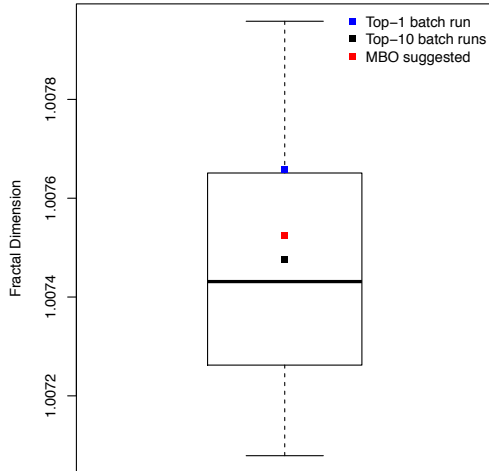
Fig. 7. Fractal dimension of the simulated land use with respect to the calibrated parameters. The box plot represents the disribution of the fractal dimension of the all individual runs used in the grid searching (n=486).

so for designing achievable policy objectives and interventions for the land system.

In the case of our example model, we made both methodological and practical findings. Methodologically, we investigated two alternative approaches to searching parameter space–grid searching and a more efficient searching algorithm. The MBO algorithm converged to the optimum parameter values derived by the expensive grid searching after a small number of iterations (approx. 30). The algorithm is designed to exhaustively search parameter space, thus represents a highly credible and efficient alternative to grid searching [36]. However, it still lacks a rigorous 'stopping rule', making our findings about convergence time only tentative. This issue is related to the important concept of 'budget' in data science [38]. Bayesian frameworks such as Approximate Bayesian Calculation [39] have been developed to use various convergence statistics on multiple Markov chains. It is important that such approaches are now investigated further to consolidate a formal calibration procedure with strong theoretical support.

In terms of model parameterisation, we found that considerable heterogeneity and/or 'irrationality' within our modelled agent population was necessary to improve agreement between simulated and remotely-sensed data. Where the ecosystem service production levels were homogeneous across agents, a high Giving-Up threshold that increased the rate of irrational land abandonment (abandonment even when non-negligible levels of benefit were being derived) gave the highest levels of agreement. Where production levels were heterogeneous, a low or even zero value for the Giving-Up threshold was superior. In both cases, a higher rate of turnover between agents would be expected than under more

'rational' settings, although the former settings would favour the most beneficial land uses in aggregate, while the latter settings would favour a more diverse mix of land uses (at least at the intermediate levels of heterogeneity that provided the best match to fractal dimension) [40]. These findings are intriguing, but are subject to confirmation in more detailed, longer-term comparisons that are clearly necessary to disentangle behavioural effects more fully.

Our analysis also raises more general issues. Clearly, the rigorous parameterisation of a single model, while an important objective, does not address more fundamental questions about land use modelling. For instance, model structure uncertainty–both in terms of the processes represented and the ways in which they are represented–is not considered here. Indeed, it remains very difficult to assess, not least because it is hard to constrain the range of alternative conceptualisations and representations of a given system. As a result, investigations of structural uncertainty are usually restricted to model comparisons, which are of limited robustness [8, 41]. However, more formal methods may be brought to bear, involving for example pedigree (independence) analysis, conceptual model analysis and sampled versus plausible models comparisons [42]. Most pertinent is the potential for structural uncertainty to be translated into alternative parameterisations, allowing it too to be investigated through machine learning [43].

Even in such a case, there remain clear dangers of overfitting models to available data; dangers that may be increased by easily applicable methods for maximising agreement between simulated and observed data. The balance between ensuring that models can replicate observations and ensuring that they can reveal unanticipated, novel dynamics (or between process and pattern accuracy) is one that is generally difficult to strike [11]. It is essential that technical advances in machine learning approaches do not overshadow this essential conceptual challenge, but contribute to appropriate handling of it where possible.

One way in which new methods may make this more general contribution is through the use of a wider variety of data sources. Large-scale statistics, grey literature, information on land management decision-making and social survey results should be used amongst others in improving, calibrating and evaluating land use models. Use of such supplementary data would allow modelling in finer resolution and give new insight into land system dynamics. More importantly, they will help to reveal a) how necessary behaviour (beyond economic optimisation) is to explain land use changes, and b) what forms and strengths behaviours take in different circumstances. Furthermore, the use of multiple goal functions across these data is clearly necessary, and can in principle help to disentangle the many processes at play in the land system. Given these further developments, the exploitation of novel data sources through machine learning could lead to a step-change in our capacity to model–and hence understand

and direct–the global consequences of land use change.

6. CONCLUSIONS

This study demonstrates the calibration of an agent-based land use model using machine learning and historical remote sensing data. The consistency of land use classes and spatial complexity between simulated and remotely sensed data were found to be informative about model performance. Using an efficient searching algorithm, it was possible to calibrate the complex land use model within a relatively short period of time. We therefore propose an operational framework for calibrating high-dimensional models of decision-making, capable of improving the quality of land use projections over large geographical extents. Further work is needed to develop rigorous stopping rules and convergence statistics as well as to provide broader theoretical support to the proposed framework. Nevertheless, machine-learning based approaches of this kind appear to hold great promise for supporting land system model development.

ACKNOWLEDGEMENT

7. REFERENCES

[1] Giacomo Grassi and Frank Dentener, "Quantifying the contribution of the Land Use sector to the Paris Climate Agreement," 2015.

[2] IPBES, "Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production," *The assessment report on Pollinators, pollination and food production*, pp. 91–105, 2016.

[3] United Nations, "United nations framework convention on climate change. The Paris Agreement.," *Paris, France*, 2016.

[4] United Nations, "United Nations Sustainable Development Goals," 2015.

[5] Timothy D Searchinger, Tim Beringer, and Asa Strong, "Does the world have low-carbon bioenergy potential from the dedicated use of land?," *Energy Policy*, vol. 110, pp. 434–446, 2017.

[6] P A Turner, C B Field, D B Lobell, D L Sanchez, and K J Mach, "Unprecedented rates of land-use transformation in modelled climate change mitigation pathways," *Nature Sustainability*, vol. 1, no. 5, pp. 240–245, 2018.

[7] Derek T Robinson, Alan Di Vittorio, Peter Alexander, Almut Arneth, C Michael Barton, Daniel G Brown, Albert Kettner, Carsten Lemmen, Brian C O'Neill, Marco Janssen, and Others, "Modelling feedbacks between human and natural processes in the land system," *Earth System Dynamics*, vol. 9, no. 2, pp. 895–914, 2018.

[8] Peter Alexander, Reinhard Prestele, Peter H. Verburg, Almut Arneth, Claudia Baranzelli, Filipe Batista e Silva, Calum Brown, Adam Butler, Katherine Calvin, Nicolas Dendoncker, Jonathan C. Doelman, Robert Dunford, Kerstin Engström, David Eitelberg, Shinichiro Fujimori, Paula A. Harrison, Tomoko Hasegawa, Petr Havlik, Sascha Holzhauer, Florian Humpenöder, Chris Jacobs-Crisioni, Atul K. Jain, Tamás Krisztin, Page Kyle, Carlo Lavalle, Tim Lenton, Jiayi Liu, Prasanth Meiyappan, Alexander Popp, Tom Powell, Ronald D. Sands, Rüdiger Schaldach, Elke Stehfest, Jevgenijs Steinbuks, Andrzej Tabeau, Hans van Meijl, Marshall A. Wise, and Mark D.A. Rounsevell, "Assessing uncertainties in land cover projections," *Global Change Biology*, vol. 23, no. 2, pp. 767–781, 2017.

[9] M. D. A. Rounsevell, A. Arneth, P. Alexander, D. G. Brown, N. de Noblet-Ducoudré, E. Ellis, J. Finnigan, K. Galvin, N. Grigg, I. Harman, J. Lennox, N. Magliocca, D. Parker, B. C. O'Neill, P. H. Verburg, and O. Young, "Towards decision-based global land use models for improved understanding of the Earth system," *Earth System Dynamics*, vol. 5, no. 1, pp. 117–137, 2014.

[10] John H Miller and Scott E Page, *Complex adaptive systems: An introduction to computational models of social life*, vol. 17, Princeton university press, 2009.

[11] Calum Brown, Ken Brown, and Mark Rounsevell, "A philosophical case for process-based modelling of land use change," *Modeling Earth Systems and Environment*, vol. 2, no. 2, pp. 50, 2016.

[12] D. Murray-Rust, C. Brown, J. van Vliet, S. J. Alam, D. T. Robinson, P. H. Verburg, and M. Rounsevell, "Combining agent functional types, capitals and services to model land use dynamics," *Environmental Modelling and Software*, vol. 59, pp. 187–201, 2014.

[13] Calum Brown, Dave Murray-Rust, Jasper Van Vliet, Shah Jamal Alam, Peter H. Verburg, and Mark D. Rounsevell, "Experiments in globalisation, food security and land use decision making," *PLoS ONE*, vol. 9, no. 12, pp. 1–24, 2014.

[14] P Meyfroidt, R Roy Chowdhury, A de Bremond, E C Ellis, K-H Erb, T Filatova, R D Garrett, J M Grove,

A Heinimann, T Kuemmerle, and Others, "Middle-range theories of land system change," *Global Environmental Change*, vol. 53, pp. 52–67, 2018.

[15] Rebecca A Kelly, Anthony J Jakeman, Olivier Barreteau, Mark E Borsuk, Sondoss ElSawah, Serena H Hamilton, Hans Jørgen Henriksen, Sakari Kuikka, Holger R Maier, Andrea Emilio Rizzoli, and Others, "Selecting among five common modelling approaches for integrated environmental assessment and management," *Environmental modelling & software*, vol. 47, pp. 159–181, 2013.

[16] Jasper van Vliet, Arnold K Bregt, Daniel G Brown, Hedwig van Delden, Scott Heckbert, and Peter H Verburg, "A review of current calibration and validation practices in land-change modeling," *Environmental Modelling & Software*, vol. 82, pp. 174–182, 2016.

[17] Claudius Graebner, "How to Relate Models to Reality? An Epistemological Framework for the Validation and Verification of Computational Models," *Journal of Artificial Societies and Social Simulation*, vol. 21, no. 3, pp. 8, 2018.

[18] Giorgio Fagiolo and Andrea Roventini, "Macroeconomic Policy in DSGE and Agent-Based Models Redux: New Developments and Challenges Ahead," LEM Papers Series 2016/17, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, 2016.

[19] Calum Brown, Peter Alexander, Sascha Holzhauer, and Mark D. A. Rounsevell, "Behavioral models of climate change adaptation and mitigation in land-based sectors," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 8, no. 2, pp. e448, 2017.

[20] Stanislao Gualdi, Marco Tarzia, Francesco Zamponi, and Jean-Philippe Bouchaud, "Tipping points in macroeconomic agent-based models," *Journal of Economic Dynamics and Control*, vol. 50, pp. 29–61, 2015.

[21] Leonardo Bargigli, Luca Riccetti, Alberto Russo, and Mauro Gallegati, "Network calibration and metamodeling of a financial accelerator agent based model," *Journal of Economic Interaction and Coordination*, pp. 1–28, 2018.

[22] Francesco Lamperti, Andrea Roventini, and Amir Sani, "Agent-based model calibration using machine learning surrogates," *Journal of Economic Dynamics and Control*, vol. 90, pp. 366–389, 2018.

[23] Calum Brown, Sascha Holzhauer, Marc J. Metzger, James S. Paterson, and Mark Rounsevell, "Land managers' behaviours modulate pathways to visions of future land systems," *Regional Environmental Change*, pp. 1–15, 2016.

[24] R S DeFries and Jonathan Cheung-Wai Chan, "Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data," *Remote Sensing of Environment*, vol. 74, no. 3, pp. 503–515, 2000.

[25] John Rogan, Janet Franklin, Doug Stow, Jennifer Miller, Curtis Woodcock, and Dar Roberts, "Mapping land-cover modifications over large areas: A comparison of machine learning algorithms," *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2272–2283, 2008.

[26] George P Petropoulos, Chariton Kalaitzidis, and Krishna Prasad Vadrevu, "Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery," *Computers & Geosciences*, vol. 41, pp. 99–107, 2012.

[27] Amin Tayyebi, Bryan C Pijanowski, Marc Linderman, and Claudio Gratton, "Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world," *Environmental Modelling & Software*, vol. 59, pp. 202–221, 2014.

[28] Hichem Omrani, Amin Tayyebi, and Bryan Pijanowski, "Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: an integrated ML-CA-LTM modeling framework," *GIScience & Remote Sensing*, vol. 54, no. 3, pp. 283–304, 2017.

[29] Hossein Shafizadeh-Moghadam, Ali Asghari, Amin Tayyebi, and Mohammad Taleai, "Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth," *Computers, Environment and Urban Systems*, vol. 64, pp. 297–308, 2017.

[30] European Environment Agency (EEA) and EEA, "Corine Land Cover (CLC) 2006, Version 18.5.1," 2016.

[31] Victor Blanco, Sascha Holzhauer, Calum Brown, Fredrik Lagergren, Gregor Vulturius, Mats Lindeskog, and Mark D.A. Rounsevell, "The effect of forest owner decision-making, climatic change and societal demands on land-use change and ecosystem service provision in Sweden," *Ecosystem Services*, vol. 23, no. December 2016, pp. 174–208, 2017.

[32] Ian Holman, Eric Audsley, Pam Berry, Calum Brown, Harald Bugmann, Liz Clarke, George Cojocaru, Rob Dunford, Stefan Fronzek, Paula A Harrison, Yasushi Honda, Victoria Janes, Sari Kovats, Valentine Lafond, Anastasia Lobanova, and Marianne Sloth, "Modelling

Climate Change Impacts , Adaptation and Vulnerability in Europe," pp. 1–100, 2017.

[33] A Kraskov, H Stögbauer, and P Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, pp. 066138, 2004.

[34] Christina Bogner, Bumsuk Seo, Dorian Rohner, and Björn Reineking, "Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea," *PLOS ONE*, vol. 13, no. 1, pp. 1–22, 2018.

[35] Kevin McGarigal, "Fragstats: Spatial pattern analysis program for categorical maps," Version 4, University of Massachusetts, 2015.

[36] Bernd Bischl, Jakob Richter, Jakob Bossek, Daniel Horn, Janek Thomas, and Michel Lang, *mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions*, 2017.

[37] Bernd Bischl, Simon Wessing, Nadja Bauer, Klaus Friedrichs, and Claus Weihs, "MOI-MBO: Multiobjective Infill for Parallel Model-Based Optimization," in *Learning and Intelligent Optimization*, Panos M Pardalos, Mauricio G C Resende, Chrysafis Vogiatzis, and Jose L Walteros, Eds., Cham, 2014, pp. 173–186, Springer International Publishing.

[38] Nikolaus Hansen, Anne Auger, Raymond Ros, Steffen Finck, and Petr Poš\'\ik, "Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009," in *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*. ACM, 2010, pp. 1689–1696.

[39] Elske van der Vaart, Mark A. Beaumont, Alice S A Johnston, and Richard M. Sibly, "Calibration and evaluation of individual-based models using Approximate Bayesian Computation," *Ecological Modelling*, vol. 312, pp. 182–190, 2015.

[40] Calum Brown, Sascha Holzhauer, Marc J. Metzger, James S. Paterson, and Mark Rounsevell, "Land managers' behaviours modulate pathways to visions of future land systems," *Regional Environmental Change*, pp. 1–15, 2016.

[41] Daniel Wallach, Sarath P Nissanka, Asha S Karunaratne, W M W Weerakoon, Peter J Thorburn, Kenneth J Boote, and James W Jones, "Accounting for both parameter and model structure uncertainty in crop model predictions of phenology: A case study on rice," *European Journal of Agronomy*, vol. 88, pp. 53–62, 2017.

[42] Jens Christian Refsgaard, Jeroen P van der Sluijs, James Brown, and Peter van der Keur, "A framework for dealing with uncertainty due to model structure error," *Advances in Water Resources*, vol. 29, no. 11, pp. 1586–1597, 2006.

[43] Laura Bojke, Karl Claxton, Mark Sculpher, and Stephen Palmer, "Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods," *Value in Health*, vol. 12, no. 5, pp. 739–749, 2006.