Environmental Modelling and Software 159 (2023) 105574

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

# Re-considering the status quo: Improving calibration of land use change models through validation of transition potential predictions

Benjamin Black [a,*], Maarten J. van Strien [a], Antoine Adde [b], Adrienne Grêt-Regamey [a]

[a] Planning of Landscape and Urban Systems, Swiss Federal Institute of Technology (ETH), Stefano-Franscini-Platz 5, Zurich, 8093, Switzerland
[b] Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, Lausanne, 1015, Switzerland

ARTICLE INFO

*Keywords:*
Land use change modelling
Cellular automata
Random forests
Land transition potential
Predictor variable selection
Land use change model calibration

ABSTRACT

The increasing complexity of the dynamics captured in Land Use and Land Cover (LULC) change modelling has made model behaviour less transparent and calibration more extensive. For cellular automata models in particular, this is compounded by the fact that validation is typically performed indirectly, using final simulated change maps; rather than directly considering the probabilistic predictions of transition potential. This study demonstrates that evaluating transition potential predictions provides detail into model behaviour and performance that cannot be obtained from simulated map comparison alone. This is illustrated by modelling LULC transitions in Switzerland using both Logistic Regression and Random Forests. The results emphasize the need for LULC modellers to explicitly consider the performance of individual transition models independently to ensure robust predictions. Additionally, this study highlights the potential for predictor variable selection as a means to improve transition model generalizability and parsimony, which is beneficial for simulating future LULC change.

## Software and data availability

All data preparation, modelling and analysis was conducted in R 4.0.5 (R core team, 2021) and the processed data and scripts to replicate the results of this research have been made available alongside this publication (https://doi.org/10.5281/zenodo.6912914).

## 1. Introduction

Over the last three decades, a wide variety of modelling approaches have been developed to simulate Land Use and Land Cover Change (LULCC), such as agent-based models, econometric models and cellular automata (Lambin 1997; Schaldach and Priess 2008; van Schrojenstein Lantman et al., 2011; Ren et al., 2019). What all of these models have in common is that they all attempt to capture the dynamics by which land changes from one use, or state, to another. In many of these LULCC models, the methods by which this is achieved have become increasingly complex over time (Brown et al., 2013), expanding their capacity to represent non-linear and non-stationary aspects of the LULCC system (Santé et al., 2010; Verstegen et al., 2016). However, this complexity can come at a cost, as it can deter, or even hinder, users from understanding the nature of the relationships being modelled (Sohl and Claggett, 2013), which in turn can make the process of model calibration, to produce accurate results, inefficient and in-transparent (van Vliet et al., 2016). At the same time, numerous approaches by which these complex models can be better explored have been expounded; although their adoption is still limited (Tong and Feng 2020).

The development of cellular automata models of LULCC (LULCC-CAs) are no exception to the trend of increasing model complexity leading to in-transparency and inefficiency in calibration (Mas et al., 2018). The basic premise of LULCC-CAs is that the study area is abstracted to a finite grid of cells of different LULC states. The likelihood of cells to change state is calculated (often as a probability) based on (1) their previous state, (2) the influence of surrounding cells' states (neighbourhood effect) and (3) transition rules or transition potential (TP) models that encode the relationship between state transitions and external driving variables (Tobler 1979; White and Engelen, 1997; White et al., 2012). LULCC is then simulated over discrete time steps with cellular transitions typically allocated on the basis of rates of change derived through Markov chain analysis (Mas et al., 2014).

Since early examples of LULCC-CAs for the simulation of urban development in the 1990s (Batty and Xie, 1994; White and Engelen,

* Corresponding author. Planning of Landscape and Urban Systems (PLUS) Institut für Raum- und Landschaftsentwicklung (IRL), HIL H 52.1, Stefano-Franscini-Platz 5, Zurich, 8093, Switzerland.
*E-mail addresses:* bblack@ethz.ch (B. Black), vanstrien@ethz.ch (M.J. van Strien), antoine.adde@unil.ch (A. Adde), gret@ethz.ch (A. Grêt-Regamey).

https://doi.org/10.1016/j.envsoft.2022.105574
Received 12 August 2022; Received in revised form 21 October 2022; Accepted 2 November 2022
Available online 9 November 2022

1993), many aspects of the approach have been expanded. Notably, the statistical modelling techniques used for TP models have become substantially more complex over time, moving from logistic regression (Kolb et al., 2013) towards approaches such as neural networks (Li and Yeh 2002), support vector machines (Yang et al., 2008), bayesian weights of evidence (Rodrigues and Soares-Filho, 2018) and random forests (RF; Kamusoko and Gamba, 2015; Du et al., 2018). One thing that many of these approaches have in common is that they are supervised learning techniques (i.e., models trained directly on observational data), which means that it is possible for their performance to be evaluated directly after model fitting. In the context of LULCC-CA modelling, such evaluation is referred to by different terms, with Paegelow et al. (2018) dubbing it "soft prediction validation" and Tong and Feng (2020) grouping it under the term "procedure assessment". Neither of these terms, represent a succinct description and as such we will henceforth refer to the direct validation of supervised TP models simply as 'stage 1 validation'. Although stage 1 validation can thus be considered the most direct way of validating TP models, the majority of LULCC-CA studies do not include stage 1 validation as part of model calibration (van Vliet et al., 2016; Tong and Feng 2020).

Instead of performing a stage 1 validation, many studies typically focus on the evaluation of TP models' outputs only after they have been used to produce simulated LULC maps at the end of the CA process, which are validated against observed historical maps. This procedure is referred to as "hard" prediction validation (Paegelow et al. 2014, 2018) or "result assessment" (Tong and Feng 2020) although we will refer to it as 'stage 2 validation'. The need for such stage 2 validation is unquestionable as it represents confirmation of the final outputs of the model. However, relying only on this validation approach is sub-optimal in two respects. Firstly, it makes for an inefficient calibration process (Mas et al., 2018), because it requires simultaneously considering all CA parameters. The number of these parameters vary with the particular model, but examples include factors related to the perturbation of transition probabilities to incorporate uncertainty or policy regulations affecting land use as well as parameters pertaining to the allocation process used to assign transitions (Mas et al., 2014). When these parameters are considered in conjunction with the aspects of the TP models that must be calibrated, such as the choice of statistical model (e.g. logistic regression vs. random forests) and attendant hyperparameters (e.g. number of layers in neural networks or trees in random forests); model scale; and the optimization of the selection of predictor variables (Mas et al., 2018), then the number of possible parameter combinations and model specifications to test becomes substantial (Newland et al., 2018a). This is typically dealt with using a brute force calibration approach where the model is systematically re-run by altering one parameter value whilst the others are held constant (Torrens, 2011), or by only addressing some of the aspects related to the TP models and ignoring others. The second reason why only employing stage 2 validation for calibration is sub-optimal is because validation is being performed on discrete classifications (i.e. allocated as either a transition or non-transition) of probabilistic predictions that have been binarized and as such this does not provide information as to the certainty of these decisions. For example, it allows no insight into the distribution of transition probability values amongst the cells that were assigned to transition at a given time point. Having access to this distribution may show that in order to meet LULCC demand some transitioning cells in fact exhibit a low modelled transition probability, this knowledge could prompt further decisions such as re-assessing the demand component of the model. Given that stage 2 validation cannot provide such insights this is a clear example of in-transparency in the modelling process.

Both of these limitations of stage 2 validation can be improved by better utilising the opportunity for stage 1 validation. Firstly, using stage 1 validation to test a greater range of specifications for the TP models is more efficient than doing so with only a stage 2 validation, because the stage 1 calibration is occurring in isolation from the other LULCC-CA parameters and the transition allocation procedure. Secondly, including stage 1 validation allows for the performance of each TP model to be immediately assessed separately. In contrast, as the predictions of the TP models have already been combined into a single output map for stage 2 validation, the performance of individual TP models is difficult to assess. Assessing TP model performance separately is useful because it allows for exploration of the range in performance across different transitions and more easily identify causes of poor performance, such as low sample sizes or high imbalance in the numbers of instances of transitions vs. persistence in the data.

In addition to this, stage 1 validation creates the opportunity to utilise a wider range of model performance metrics to provide insights into TP model behaviour than stage 2 validation (Paegelow et al., 2014). This is due to the fact that it is possible to validate the cellular transition values as either probabilities or as discretized, binary values (by applying a classification threshold). When validation of the transition probabilities is performed, it is typically with 'non-threshold' dependent measures such as the Receiver Operating Characteristic (ROC) approach (Paegelow et al., 2018), although there is potential to utilise other complementary metrics such as the Boyce Index (Boyce et al., 2002; Hirzel et al., 2006), which as a 'presence only' measure, focuses only on the instances of observed LULC transitions and not persistence. This index is particularly applicable because LULCC datasets are typically strongly skewed towards the latter (given that the majority of the landscape does not change).

A further benefit of stage 1 validation is the fact that it allows for independent validation, i.e. for model accuracy to be assessed using instances (data points) that have not been used to train the models. This is often not possible for stage 2 validation because the CA allocation process of binarizing probabilistic predictions is not robust to being performed on a subset of the instances (For example, allocation algorithms in Dinamica EGO utilise the spatial aggregation of instances: (Rodrigues and Soares-Filho, 2018). Independent validation is useful as it can provide insights as to the generalizability of models (Bishop, 2006). In the fields of machine learning and data mining, generalizable models are those that exhibit less overfitting on training data and as such perform better in the prediction of new, unseen, data (Abu-Mostafa et al., 2012). Thus, generalizability is desirable trait for TP models given that they are supplied with temporally dynamic data to simulate future LULC (Mas 2004; Soares-Filho et al., 2013). One means of improving model generalizability that can be leveraged through stage 1 validation is by incorporating processes of predictor variable selection to remove redundant variables that constitute noise in the data and can lead to models becoming overfit, i.e. non-generalizable (Guyon et al., 2006). Furthermore, an additional incentive to use predictor variable selection is to produce parsimonious TP models, i.e. models optimized to have compact and non-redundant predictor sets whilst still having an acceptable level of accuracy. Parsimonious TP models, with less predictors, minimize the need to acquire or extrapolate temporally dynamic predictor values or adopt stationarity assumptions when performing future LULCC simulations.

Despite these potential benefits, the incorporation of stage 1 validation as part of the calibration of LULCC-CAs still remains underutilized (Tong and Feng 2020). As such, the aim of this study is to highlight the utility of stage 1 validation as a means of providing insights into TP model behaviour and performance that can be used to improve the efficiency of the calibration of LULCC-CAs. To this end, we present an applied example of the stage 1 validation of TP models for LULCC in Switzerland between 2009 and 2018. We use multiple model validation metrics to highlight differences in the performance of TP models under different specifications and apply a two-step approach to predictor variable selection and how this can be used to improve TP model generalizability and parsimony.

## 2. Methods

The methodological process of this study is presented in Fig. 1. First,

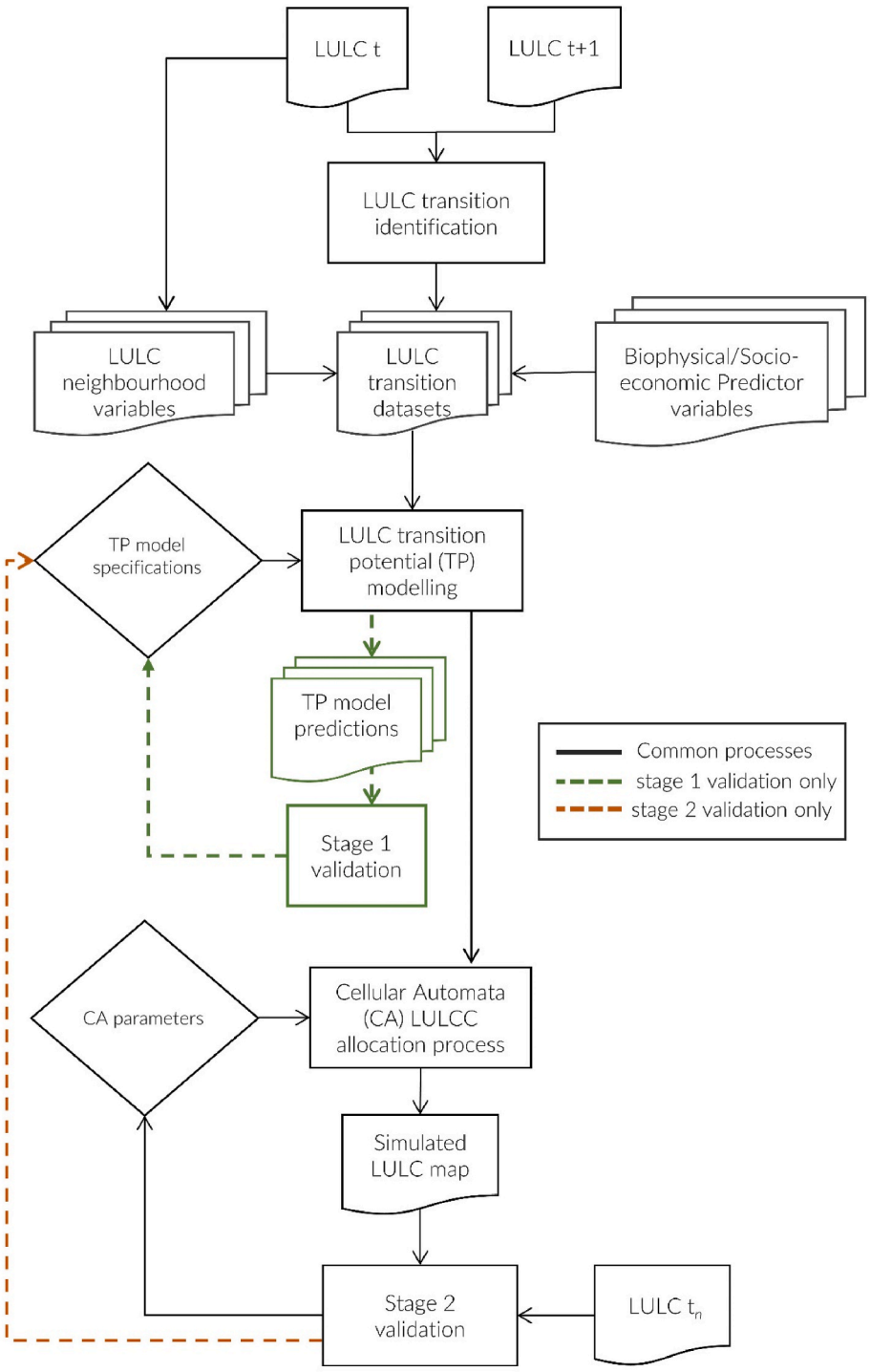


**Fig. 1.** Generalized schematic of the calibration of transition potential models within land use land cover (LULC) change cellular automata highlighting the differences between stage 1 and 2 validation.

LULC transitions in the form of changes from a specific 'initial' LULC class to another specific 'final' class (i.e. LULC class x to LULC class y) were identified and combined with predictors to form transition datasets (see sections 2.1-2.2). These datasets where then used to prepare models of transition potential under different specifications (see section 2.3) which were subjected to stage 1 validation (see section 2.4). It is important to note that this study does not address any of the subsequent steps below the stage 1 validation presented in Fig. 1. However, for the purpose of discussing the efficiency benefits of incorporating stage 1 validation, these latter steps have been deliberately included in the figure.

### 2.1. LULC transition dataset creation

#### 2.1.1. LULC data preparation

Analysing LULCC requires two historical LULC maps to identify change between. For this, we utilized the Swiss area statistics (Swiss Federal Office for Statistics and Geoinformation, 2021) for the periods of 2004–2009 and 2013–2018. These are classified LULC datasets derived from aerial photography using a 100 m point grid covering the entirety

of Switzerland (41,285 km$^2$). Whilst a date range is given for each period this is the extent of the data-collection flying period, and in fact each product represents a single LULC map. As such, for simplicity, the datasets will henceforth be referred to as 2009 and 2018 respectively, meaning that the time period for analysis is between these two dates. The original datasets include 72 LULC classes, but, following the example of previous LULCC studies in Switzerland (Price et al., 2015; Gago-Silva et al., 2017; Gerecke et al., 2019), these were aggregated to 10 classes as presented in Table 1. Finally, in order to be intersected with the predictor data, the LULC transition point datasets were converted to rasters with a resolution of 100 m.

#### 2.1.2. Transition dataset identification

To identify the specific LULC transitions to be modelled, we created a cross-tabulation matrix of the areal changes in the 10 aggregated LULC classes that occurred between the 2009 and 2018 datasets. This matrix was first filtered to remove illogical transitions, for example sealed surfaces such as motorways are unlikely to be converted to semi-natural surfaces. As highly imbalanced data (i.e. having many more transitions than non-transitions or vice-versa) can decrease model robustness, further filtering was applied to exclude any LULC transitions that resulted in an areal change of less than 0.5% of the total area for the initial class. This left a final list of 30 viable LULC transitions, which was used to identify transition datasets by evaluating all pixels from the two LULC data layers on the basis of two criteria for each LULC transition:

i. Pixels that displayed the initial LULC class x and final LULC class y corresponding to the LULC transition were assigned a value of 1 to represent 'change' pixels.
ii. All other pixels displaying initial LULC class x to any other final class than y were assigned a value of 0 to represent 'non-change' pixels.

**Table 1**
Aggregation of the Swiss area statistics land use land cover (LULC) classes.

| Aggregated LULC class | Area statistics LULC classes |
|---|---|
| Alpine pastures | Favourable alpine pastures; Brush alpine pastures; Rocky alpine pastures; Sheep pastures |
| Closed forest | Normal dense forest; Forest strips; Afforestations; Felling areas; Brush forest |
| Glacier | Glaciers, perpetual snow |
| Grassland or meadows | Meadows; Farm pastures; Brush meadows and farm pastures; Alpine meadows |
| Intensive agriculture | Arable land |
| Open Forest | Damaged forest areas; Open forest (on agricultural areas); Open forest (on unproductive areas); Groves, hedges; Clusters of trees (on agricultural areas); Clusters of trees (on unproductive areas) |
| Shrubland | Scrub vegetation; Unproductive grass and shrubs |
| Permanent crops | Intensive orchards; Field fruit trees; Vineyards; Horticulture |
| Settlement/urban/amenities | Industrial and commercial buildings; Surroundings of industrial and commercial buildings; One and two-family houses; Surroundings of one and two-family houses; Terraced houses; Surroundings of terraced houses; Blocks of flats; Surroundings of blocks of flats; Public buildings; Surroundings of public buildings; Agricultural buildings; Surroundings of agricultural buildings; Unspecified buildings; Surroundings of unspecified buildings; Parking areas; Construction sites; Unexploited urban areas; Public parks; Sports facilities; Golf courses; Camping areas; Garden allotments; Cemeteries |
| Static | Motorways; Green motorway environs; Roads and paths; Green road environs; Sealed railway areas; Green railway environs; Airports; Airfields, green airport environs; Energy supply plants; Waste water treatment plants; Other supply or waste treatment plants; Dumps; Quarries, mines; Lakes; Rivers; Flood protection structures; Avalanche and rockfall barriers; Wetlands; Alpine sports facilities; Rocks; Screes, sand; Landscape interventions |

Pixels identified by these criteria formed the units of analysis or instances for a given LULC transition dataset. This resulted in 30 Switzerland-wide transition datasets. As transition probability can be strongly dependent on region-specific conditions, these transition datasets were sub-divided by the extent of the six biogeographical regions of Switzerland: Jura, Plateau, Northern Prealps, Southern Prealps, Western Central Alps and Eastern Central Alps (Gonseth et al., 2001). This left a total of 174 regionalized transition datasets (some transitions involving glaciers were not present in particular regions).

### 2.2. Predictor selection

#### 2.2.1. Conceptual prediction model

Predictor variables within the TP models of LULCC-CAs are commonly grouped into three categories: suitability, accessibility and neighbourhood variables (Escobar, 2018). Suitability predictors are typically biophysical and socio-economic predictors that are perceived to be related to the suitability for a given land use or land use change e.g. elevation, precipitation, human population density etc. Accessibility predictors, as the name suggests, capture the spatial proximity of individual instances to infrastructure or landscape features e.g. distance to roads or urban centres. Neighbourhood predictors represent the influence of surrounding LULC on the likelihood of a given instance (cell) to undergo a LULC transition which can be quantified with a range of approaches (Verburg et al., 2004; Santé et al., 2010; Roodposhti et al., 2020). Given this diversity of approaches, neighbourhood influence is typically one of the most extensively calibrated aspects of LULCC-CAs (van Vliet et al., 2013; Newland et al., 2018b).

Many LULC-CAs utilise some combination of the predictor categories described above but may differ in terms of the weighting ascribed to each based on perceived importance. Given that this study focuses on statistical models of TP, we operate on a simple conceptual model of LULCC that considers all categories of predictors equally, illustrated mathematically as follows:

$$ {}^{t}P_{j,i} = f\left({}^{t}S_{j,i}, {}^{t}A_{j,i}, {}^{t}N_{j,i}\right) \quad \text{Eqn. 1} $$

where $P$ is the probability for LULC transition $j$ to occur at the location of instance $i$ at time $t$, given the values of suitability (S), accessibility (A) and neighbourhood (N) predictor variables.

#### 2.2.2. Suitability and accessibility predictors

Suitability and accessibility predictors for this study were chosen based on those employed by previous LULCC studies in Switzerland (Price et al., 2015; Gago-Silva et al., 2017; Gerecke et al., 2019). Table 2 below details the names of the predictors utilized and their data sources. All predictor data was resampled to rasters with 100 × 100 m cell size and then combined with the LULC transition datasets.

#### 2.2.3. Neighbourhood predictors

To incorporate the effect of surrounding LULC (neighbourhood influence) on LULC transitions, first 5 'active' LULC classes were identified based on their perceived influence on transitions, these were: Settlement/urban/amenities, Intensive agriculture, Alpine pastures, Grassland/meadows and Permanent crops. Following this, we adopted the approach of Roodposhti et al. (2020) by creating a set of 25 Pythagorean matrices of varying size (9–121 cells) with randomized central values and decay rates. These matrices were then applied as moving focal windows across the rasters of active LULC classes, whereby the values in the matrix accumulate according to the locations of the active LULC class pixels. Further details of this process, including exemplar matrices, have been included in Appendix A. This resulted in 125 neighbourhood predictor layers (25 per active LULC class) that capture different realizations of the influence of active LULC classes on their surroundings. These were natively produced as 100 m rasters and were also combined with the suitability and accessibility predictors.

**Table 2**
Suitability and accessibility predictor variables employed.

| Variable name | Data source |
|---|---|
| Distance to roads | Swiss Federal Office of Topography, 2011 |
| Continentality | Descombes et al. (2020) |
| Light | Descombes et al. (2020) |
| Soil pH | Descombes et al. (2020) |
| Soil nutrients | Descombes et al. (2020) |
| Soil moisture | Descombes et al. (2020) |
| Soil moisture variability | Descombes et al. (2020) |
| Soil aeration | Descombes et al. (2020) |
| Soil humus | Descombes et al. (2020) |
| Mean elevation | Wiederkehr and Möri (2013) |
| Aspect | Wiederkehr and Möri (2013) |
| Slope | Wiederkehr and Möri (2013) |
| Hillshade | Wiederkehr and Möri (2013) |
| Noise pollution index | Swiss Federal Office of the Environment (2009) |
| Distance to lakes (mean minimum dist to all lakes of different categories) | Swiss Federal Office of Topography, 2022 |
| Distance to river (mean minimum distance (agg. From 25 m data to rivers of all Strahler classes) | Swiss Federal Office of Topography, 2007 |
| Average annual mean air temperature between 2004 and 2009 | Broennimann (2018) |
| Average annual precipitation between 2004 and 2009 | Broennimann (2018) |
| Average growing degree days heat sum above 0 °C between 2004 and 2009 | Broennimann (2018) |
| Average growing degree days heat sum above 3 °C between 2004 and 2009 | Broennimann (2018) |
| Average growing degree days heat sum above 5 °C between 2004 and 2009 | Broennimann (2018) |

### 2.3. Transition potential modelling

#### 2.3.1. Predictor variable selection

Predictor variable selection techniques can be categorized as filter, wrapper and embedded approaches (Stańczyk, 2015). In this study, we utilized a two-step approach to predictor variable selection that combined filter and embedded approaches (A. Adde, University of Lausanne, 2022; personal communication). The first filter-based step involved using univariate regression models to rank all predictors and then use pairwise Pearson's correlation testing to iteratively remove predictors whose correlation exceeded a threshold of 0.7 (Dormann et al., 2013). In the second step, the filtered sets of predictors were then subject to further variable selection using a model embedded approach in the form of the Guided Regularized Random Forests (GRRF) algorithm. GRRF is an adaptation to the Random Forests algorithm developed for the purpose of selecting "compact" (i.e. non-redundant) subsets of predictors (Deng and Runger 2013), thereby giving rise to parsimonious models. For brevity, the full details of both predictor variable selection steps are presented in Appendix B.

#### 2.3.2. Random forests

Random Forests (RF) is an ensemble decision tree algorithm capable of being utilized for binary or multi-class classification or regression problems in either a supervised or unsupervised context. Given its widespread usage, we will not detail the RF algorithm here, instead readers should refer to the seminal work of Breiman (2001).

For this study, we created RF models for two variations of each regionalized LULC transition dataset (see 2.1.2). One model for each dataset without predictor variable selection being applied (RF_full) and one model with predictor variable selection applied (RF_reduced). RF classification models were fitted using the 'randomForest' R package (Cutler et al., 2022). The specifications for the models under both conditions were largely the same: the minimum size of terminal nodes ('nodesize') and number of variables randomly sampled as candidates at each split ('mtry') were both set to the default values for classification, whereas the optimum number of trees was determined through testing (Appendix C) to be 500 trees. A systematic down-sampling approach was used to address class imbalance utilising the option for tree-level proportional sampling through the 'sampsize' argument based on the degree of imbalance in the dataset (Appendix C). Overall, this resulted in the creation of 174 RF models under each condition (RF_reduced vs. RF_full) that were used in model comparison.

#### 2.3.2. Logistic regression

Logistic Regression (LR), despite acknowledged limitations (Mustafa et al., 2018), is still widely considered to be the most popular and benchmark technique for TP modelling in LULCC-CAs (Feng et al., 2018). We created two variations of LR models for comparison to the RF models: LR models for each regionalized LULC transition dataset subject to predictor variable selection (LR_reduced) and without predictor variable selection (LR_full). All LR models were fitted using the base R "glm" function with the family argument set as "binomial" (R core team, 2021).

### 2.4. Model validation

For all RF and LR models, fitting was performed for five replicates using a split-sample approach (70:30 training and test set splits using proportional random sampling without replacement) to allow for hold-out validation (i.e. model performance is validated using the test set).

The Relative Operating Characteristic (ROC) method is a commonly applied technique for the validation of probabilistic TP model outputs (Paegelow et al., 2018). The ROC approach involves plotting a curve of the rate of true positives (i.e. correct predictions of change) versus the rate of false positives from the comparison of observed LULC transitions with predicted Boolean transition values generated by applying multiple classification thresholds to the probabilistic predictions from the TP model (Pontius and Parmentier 2014). The benefit of the ROC approach is that the area under the curve (AUC) can be calculated as a single-valued metric representing the degree of association between high predicted probabilities of LULC transition and actual observed transitions. A complementary metric to the AUC ROC that is not commonly applied, but is useful for TP model validation, is the Boyce Index (Boyce et al., 2002; Hirzel et al., 2006). Calculating the Boyce Index involves separating the instances of observed LULC transitions into classes according to the probabilistic prediction values assigned to them, then calculating class-wise ratios of the frequency of instances predicted to fall into the class vs. the expected frequency of instances in the class under a random distribution (predicted-to-expected (P/E) ratio: Hirzel et al., 2006). If models are well-fitted, then the P/E ratio values of the classes should exhibit a monotonically increasing curve as the value of prediction probability increases. As such, the value of the Boyce Index is the Spearman rank correlation coefficient between the P/E ratios and the probability classes.

We calculated AUC ROC and Boyce Index values (using the "ROCR" and "ecospat" packages respectively with the moving window approach for the 'ecospat.boyce' function using 1000 windows: Sing et al. (2020); Broennimann et al. (2022)) for all variations of the LR and RF TP models. Values for each metric were calculated individually for each of the test and training sets under the five replicates before being aggregated into average values across the replicates, for each model. Finally, in order to present a single validation metric, we re-scaled the AUC value to the same range as the Boyce Index (−1 to 1) and took an average of the two, which we will refer to as the 'model score'.

In evaluating models using the ROC approach an AUC value of 0.7 is considered to be a general threshold value for acceptable performance (Hosmer and Lemeshow, 2000) whereas no such value has been proposed for the Boyce Index. As such, to evaluate models of the different model specifications (RF_reduced, RF_full, LR_reduced and LR_full) we re-scaled this AUC threshold value according to the range of the Boyce index and model score to give a threshold value of 0.4 for these metrics. In terms of statistical comparisons between model specifications, this

represented an un-replicated complete block design. Given that the AUC, Boyce and model score results for the models all violated the normality assumption for parametric repeated measures ANOVA (Shapiro-Wilks test of between group residuals), we utilized the non-parametric Friedman test (Pereira et al., 2015) to test for significant differences in performance between models both with and without the removal of outliers. Following the Friedman test, post-hoc testing was completed using pairwise Conover's all-pairs comparisons test (with Bonferroni correction of the p-values: Conover 1999).

We utilized the results of this analysis to select TP models that exhibited poor performance to demonstrate how the Boyce Index and ROC-AUC can be used to provide more detailed information on model behaviour. In this regard we produced Boyce index curves of the ratio of the predicted vs. expected frequency of instances across prediction probability classes as well ROC curves.

Finally, to demonstrate the benefits of predictor variable selection in terms of model generalizability, we calculated the differences in average model score between the test and training datasets for the LR and RF models under both the reduced and full specifications. The magnitude of the difference in performance between the test and training datasets represents an approximation of the generalization error (Roelofs, 2019) and as such is a simplistic gauge of generalizability, whereby the smaller the difference in performance the more generalizable (less likely to be overfit) the model is. Whereas to highlight benefits in terms of model parsimony we calculated the differences in average model score between TP models under the RF_reduced vs. RF_full and LR_reduced vs. LR_full specifications.

## 3. Results

### 3.1. Transition datasets

A table detailing the LULC classes and the number of instances in each of the Switzerland-wide and bioregional LULCC transition datasets is presented in Appendix D.

### 3.2. Model performance

Fig. 2 shows that the performance of the TP models in terms of the average model score, AUC and Boyce index values, varied under the different specifications (LR_full, LR_reduced and RF_full, RF_reduced). Simple visual comparison between the distributions of values suggest that both of the RF specifications outperformed the LR specifications and indeed this was confirmed with statistical testing (Appendix E).

More importantly, Fig. 2 shows that under all specifications there were a number of models that exhibited values of the performance metrics below the respective threshold values indicating poor performance. Specifically, for the AUC metric, for which the threshold value of 0.7 is well recognised, Fig. 2 shows that even under the best performing RF model specifications there were 36 transition models below the threshold. However, the benefit of utilising non-threshold metrics such as the Boyce Index and the ROC approach is that the behaviour of these models can be explored in greater detail through the graphical representation of model performance across the prediction probability gradient. In this regard, Fig. 3 presents Boyce and ROC curves along with the corresponding values of the metrics for a single replicate (randomly selected) for each of the RF and LR reduced models, for the LULC transition of Intensive Agriculture to Grassland in the Southern Pre-Alps region, which was one of the transitions that performed below the thresholds (Fig. 2).

The Boyce Index values in Fig. 3A suggest that both the LR and RF models for this transition exhibited relatively similar performance ($\rho$ = 0.28 and 0.47 respectively) however the curves highlight some discrepancies between them. Whereas well fitted models should exhibit a monotonically increasing curve of P/E ratio to predicted probability, the curve for the LR model shows a plateau of low P/E ratio at prediction

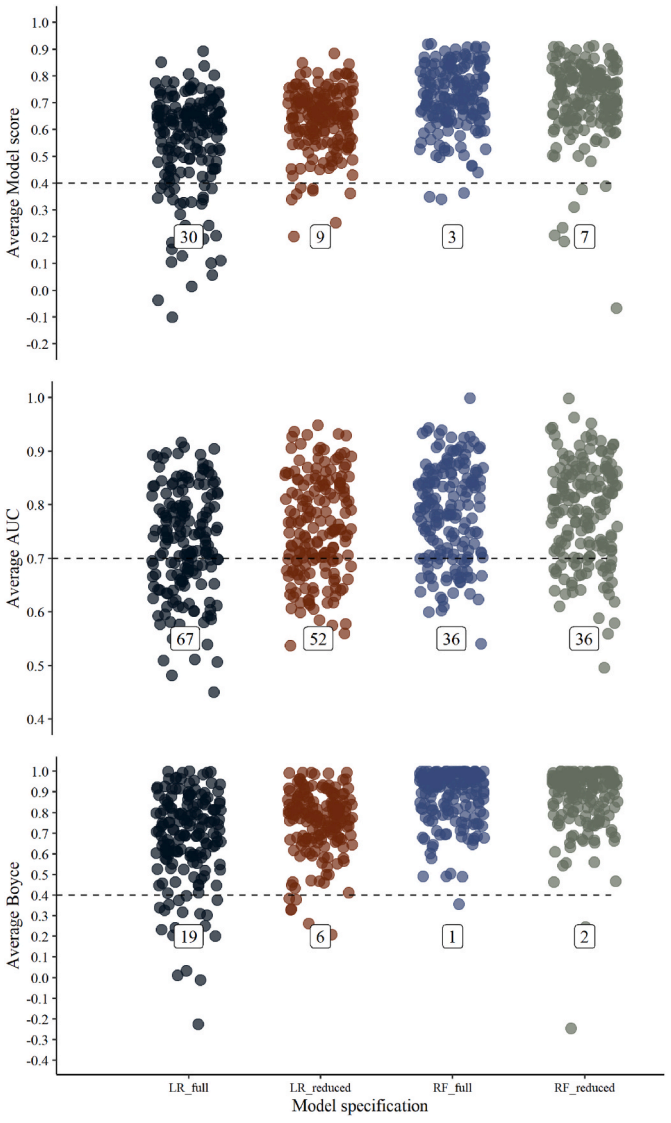


**Fig. 2.** Comparison of average values of model score, Area Under Curve (AUC) and Boyce Index across the replicates for transition models under different specifications with labels indicating the number of models below respective performance thresholds (Dashed lines: model score and Boyce Index = 0.4, AUC = 0.7). Note: horizontal perturbation applied to avoid overplotting.

probability of 0.22–0.24 and a subsequent drop in P/E ratio at probabilities >0.3. A similar pattern is exhibited by the curve for the RF model albeit at comparatively greater values of predicted probability. In both cases, the large drop in P/E ratio at the upper bounds of the predicted probability values is notable because this means that at these high probabilities, where we should expect the greatest frequencies of transitions to be predicted the model is in fact predicting fewer transitions than should be expected under a random distribution (P/E ratio values < 1). In addition to this, the fact that the highest predicted probability values from either model in Fig. 3A do not exceed 0.5 would mean that if a threshold of 50% predicted probability was applied to select cells to transition in the CA then none of the instances of transitions in this dataset would be selected, which is of course erroneous. This indicates that these TP models are not capable of strongly discriminating between instances of transitions and persistence.

This is further supported by the ROC curves in Fig. 3B, for which the curve of the LR model shows that there is range of prediction probability thresholds for which the true positive rate is approximately equal to the false positive rate (i.e. where the curve crosses the dashed no

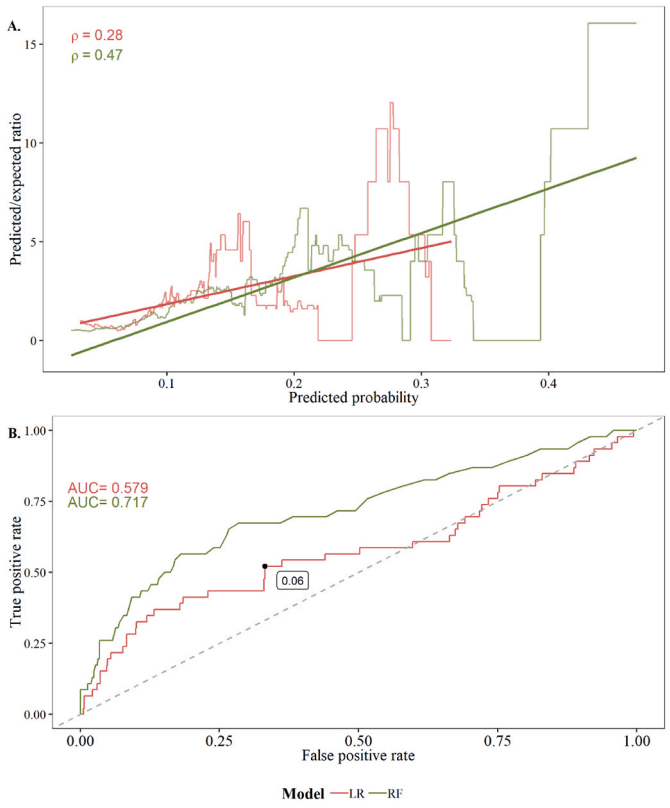


**Fig. 3.** A. Boyce curves with corresponding linear regression lines and Boyce index values (Spearman's rank correlation ($\rho$)). B. Receiver-Operating Characteristic (ROC) curves and Area Under Curve (AUC) values for the Random Forest (RF) and Logistic Regression (LR) reduced models for the transition of Intensive agriculture to Grassland in the Southern Pre-Alps region (Note the dashed line in B is the theoretical 'no discrimination' line representing performance of a random model and the labelled value is the prediction probability threshold at which there is the greatest difference between the True positive and False positive rate for the LR model).

discrimination line in the upper right quadrant of the plot). This indicates that at these values of prediction probability the model is no better at discriminating between instances of transitions and persistence than a random model. By comparison the prediction probability at which the model demonstrates the best discrimination (i.e. greatest difference between true and false positive rates) is a value of 0.06 or a 6% likelihood of transition (labelled). This is problematic as ideally the model should exhibit better discrimination at high probability values as cells with these values are more likely to be selected to transition in the CA allocation process.

### 3.3. Impacts of predictor variable selection

#### 3.3.1. Model generalizability

Fig. 4 shows a comparison of the difference in average model score between the training and test datasets of the models without vs with predictor variable selection (LR_full vs. LR_reduced, RF_full vs. RF_reduced). Fig. 4 shows that under LR the reduced models displayed significantly lower differences in average model score as compared to the full models (median: −0.02 vs. −0.15 respectively). This indicates that when using LR the use of predictor selection resulted in more generalizable models. The same general trend is true for the RF models with the reduced models exhibiting a median difference in average model score value of −9.25e-03 as compared to a median value of −0.04 for the full models. However, for the RF models, the difference was non-significant under the Wilcoxon signed rank test.

#### 3.3.2. Model parsimony

Fig. 5 shows the differences in average model score between the models with and without predictor variable selection (i.e. reduced model – full model) for both the LR and RF models. From Fig. 5, it is clear that predictor variable selection had different effects on performance depending on the type of model. For LR there was a substantially greater number of models that showed positive values of differences in average model score as compared to negative values (124:50 respectively) indicating that predictor variable selection tended to improve performance. Furthermore, the mean value of the difference in average model score was also comparatively greater for models exhibiting positive values (0.129) than negative values (−0.067), which highlights that even when predictor variable selection reduced model performance the effect was not as pronounced as the positive effect. As for RF the number

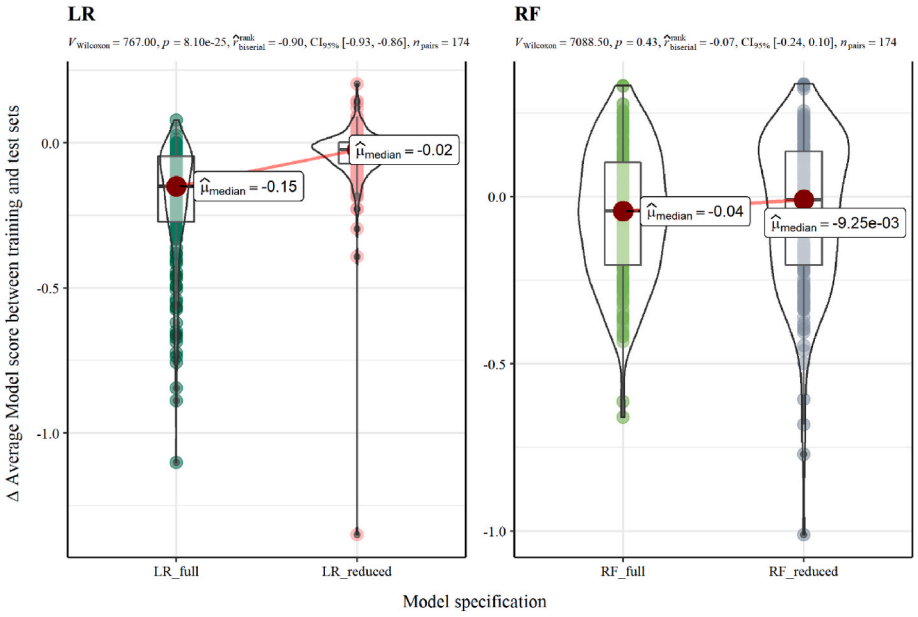


**Fig. 4.** Violin plots of the differences (Δ) in the average model score between the training and test sets for reduced and full models under both Logistic Regression (LR) and Random Forests (RF) including statistical comparisons (Wilcoxon signed-rank test).

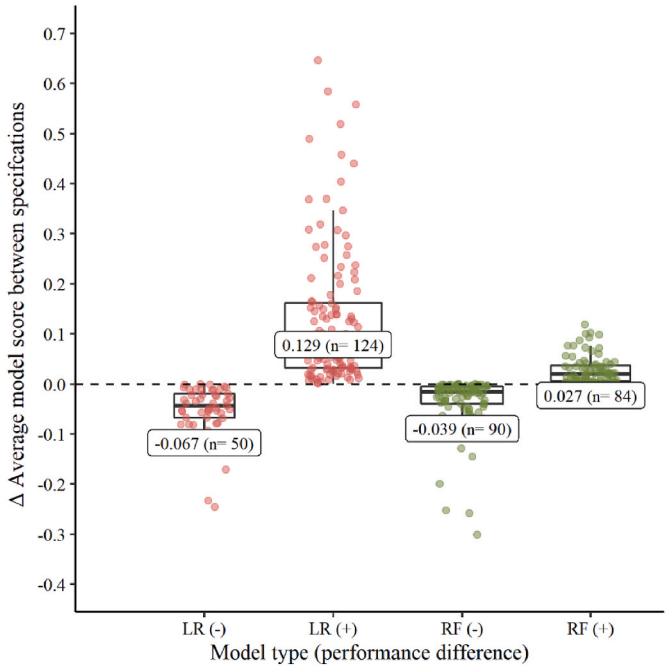


**Fig. 5.** Differences (Δ) in the average model score metric between the reduced and full models under both Logistic Regression (LR) and Random Forests (RF) with separate box plots and labels (mean Δ average model score (freq)) for models showing decreased (<0 = '-') vs. increased performance (>0 = '+').

of models exhibiting positive vs. negative values of difference in average model score were fairly even (84:90) with the means of each group also being of a similar magnitude. This indicates that predictor variable selection had less of impact on the performance of RF models, which is supported by the statistical pairwise comparisons presented in Appendix E.

## 4. Discussion

In this study, we have demonstrated how stage 1 validation allows for more comprehensive insights into TP model performance and behaviour. By comparing individual TP models' performance under multiple metrics, we highlighted that many TP models were not of a 'threshold' quality sufficient for accurate prediction of data from the time period they were fitted upon. This is especially problematic considering that these models are required to be used in future simulations on unseen data. Regardless, the capability to provide this insight is a clear advantage of stage 1 validation, whereas with stage 2 validation it would have been necessary to perform disaggregation of the validation results to achieve this level of detail.

Following this, we showed how stage 1 validation can be used to understand how TP models may be performing poorly through inspection of the Boyce and ROC curves which is not possible with stage 2 validation given that the model predictions have already been discretized into binary values. These curves offer insights into performance across the prediction probability gradient and whilst the ROC approach is frequently used in LULCC-CA studies that do incorporate stage 1 validation our application of the Boyce curve to this domain is novel. We demonstrated how the Boyce curve can be used to identify prediction probabilities at which TP models are predicting less transitions than should be expected to occur by chance. Whilst this insight does not directly indicate how TP models can be improved, it is useful in informing further investigation such as mapping the instances at prediction probabilities with low P/E ratio values to identify possible causative factors.

Ultimately, poor performance of TP models must be addressed if they are to be utilized within a LULCC-CA. In this regard, there are two general options: to try and improve the models by keeping the algorithm the same, but altering parameters (Du et al., 2018), or to use a different algorithmic approach or a combination of multiple methods (e.g. ensemble modelling: Shafizadeh-Moghadam 2019). We have demonstrated elements of both solutions, through hyper-parameter tuning and attempting to address class imbalance (Appendix C), but primarily by exploring the benefits of predictor variable selection which has not been covered extensively within previous LULCC-CA studies using RF (Kamusoko and Gamba 2015; Du et al., 2018; Gounaridis et al., 2019; Roodposhti et al., 2019; Zhang et al., 2019; Chen et al., 2019; Li and Chen 2020; Rienow et al., 2021).

We demonstrated that predictor variable selection not only improved TP model generalizability but also improved performance in a substantial proportion of cases, whilst sometimes removing over 50% of predictors (see Figure B1). These findings have important consequences for TP models being used within LULCC-CAs to make predictions of future LULC, because the models remain stationary but are fitted with new data and hence less generalizable models will produce less accurate simulations (Soares-Filho et al., 2013). However, it is important to highlight that these benefits of predictor variable selection differed between the model types utilized, with the process having less impact on RF than LR. This is likely because the regular RF algorithm is inherently robust to redundant variables to an extent, due to the fact that variables are chosen during node splitting based on a measure of importance (Kubus, 2018). Also, the fact that the majority of predictors that were removed were neighbourhood predictors which, given that they represented different realizations of the same phenomena (Appendix A), could have been disproportionately contributing to the overfitting of the models.

At the same time, it is important to note that the conception of generalizability, and the means of quantifying it, employed in this research are not definitive. Indeed, an alternative approach to quantifying generalizability would be to compare the performance of models trained on data from a given time period when used to predict transition potential for data from a subsequent period, sometimes referred to as external validation (Ho et al., 2020). The decision to instead utilise independent validation to quantify generalizability in this research was intended to demonstrate that this is possible with stage 1 validation but not stage 2 validation. This is a useful capability of stage 1 validation given that future LULCC simulation modelling typically only utilizes TP models for the most recent time period available and hence we have an interest in quantifying whether these specific models are generalizable and of course this cannot be done using external validation because subsequent time period data does not exist. However, further research should be performed to compare the estimates of generalizability produced by independent versus external validation approaches.

Regardless of how the improvement of TP model performance is pursued a final benefit of stage 1 validation is that it makes the process more efficient. This is because it allows for the comparison of different specifications without the need to instantiate other parameters in the CA and run the allocation process to produce simulated LULC maps (Fig. 1). In the context of this study, only using stage 2 validation for hyper-parameter testing (four different RF ensemble sizes for both reduced and full datasets: Appendix C) and the two additional LR specifications of TP models would have required the allocation process to be run a minimum of 10 times. The time required for allocation varies dependent on the specific LULCC-CA being utilized but generally it can be expected to scale with the size of the study area and the number of LULC transitions being modelled. Given that we are modelling LULCC at the scale of the whole of Switzerland with 100 m resolution and 174 LULC transitions, the fact that we did not have to prepare the CA model or run the allocation process to calibrate the TP models represents a substantial improvement in efficiency.

Despite the benefits of utilising stage 1 validation as shown by this study and acknowledged by other authors (Kolb et al., 2013), it still

remains under-utilized in LULCC-CA studies (Tong and Feng 2020). The reason for this is difficult to attribute, but one suggestion is because extensive validation of TP models is often not central to the aims of many LULCC-CA studies (*ibid.*). For example, where the goal of research is to devise and simulate a range of future LULCC scenarios it is understandable that model calibration is given limited attention. Another possible explanation could stem from the fact that many popular LULCC-CAs are proprietary software (e.g. Dinamica EGO, Land Change Modeller) and conduct the fitting of the TP models internally, thereby removing some of the onus on the user to specify and interpret the models. Furthermore, whilst many proprietary LULCC-CAs do include the option to calculate performance metrics for stage 1 validation these are limited in comparison to the means for stage 2 validation (Paegelow et al., 2018) and by no means do they force users to explicitly consider the performance of the TP models in isolation. On the other hand, the fact that LULCC-CAs, such as Dinamica EGO, offer a flexible, modular and graphical modelling environment makes them accessible to a greater range of users. In this regard, it is clear that promoting increased scrutiny of TP models through stage 1 validation should not come at the expense of the usability of LULCC-CA software.

A possible solution to this, in order to increase the adoption of stage 1 validation within LULCC-CAs studies could be the establishment of a universal protocol for the calibration of TP models that is generalizable across the popular LULCC-CA models. In this regard, the framework proposed by Moulds et al. (2015) represents some progress; however, it is largely a software focused approach. Instead, such a protocol should specifically elucidate the steps involved in preparing and evaluating TP models and the relevant aspects that should be considered, such as the use of different model scales, algorithmic techniques, predictor variable selection, addressing imbalanced datasets, model uncertainty and the merits of different performance metrics. Whilst we do not present such a protocol, we hope that by sharing the data and scripts that allow our research to be replicated we are contributing in a small way towards its development.

## 5. Conclusion

Our research has shown that directly evaluating probabilistic predictions of LULC transition models, which we dub stage 1 validation, has the potential to improve the efficiency and transparency of the calibration process for LULCC-CAs. As highlighted, the potential to utilise stage 1 validation is not novel, rather it has been overlooked as part of the status quo approach for LULCC-CA calibration. We hope that this introspective approach to highlighting existing opportunities to improve practice could serve to stimulate similar efforts other fields of land use change modelling.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We gratefully acknowledge financial support from the Swiss Federal Office for the Environment (FOEN) under the project "ValPar.CH: Values of the ecological infrastructure in Swiss parks" of the Action Plan of the Swiss Biodiversity Strategy. We would also like to thank Philipp Brun (Swiss Federal Institute for Forest, Snow and Landscape Research) for providing R scripts and functions for species distribution modelling that were adapted for this research.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105574.

## References

Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H.-T., 2012. Learning from Data. AMLBook.
Batty, M., Xie, Y., 1994. From cells to cities. Environ. Plann. Plann. Des. 21, S31–S48. https://doi.org/10.1068/b21S031.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning, Information Science and Statistics. Springer, New York.
Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. Ecol. Model. 157, 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Broennimann, O., 2018. CHclim25: A High Spatial and Temporal Resolution Climate Dataset for Switzerland. Ecospat laboratory, University of Lausanne, Switzerland.
Broennimann, O., Cola, V.D., Petitpierre, B., Breiner, F., Scherrer, D., D'Amen, M., Randin, C., Engler, R., Hordijk, W., Mod, H., Pottier, J., Febbraro, M.D., Pellissier, L., Pio, D., Mateo, R.G., Dubuis, A., Maiorano, L., Psomas, A., Ndiribe, C., Salamin, N., Zimmermann, N., Collart, F., Guisan, A., 2022. Ecospat: Spatial Ecology Miscellaneous Methods.
Brown, D., Band, L.E., Green, K.O., Irwin, E.G., Jain, A., Lambin, E.F., Pontius, R.G., Seto, K.C., Turner II, B.L., Verburg, P.H., 2013. Advancing Land Change Modeling: Opportunities and Research Requirements. The National Research Council Press, Washington.
Chen, Y., Li, X., Liu, X., Zhang, Y., Huang, M., 2019. Tele-connecting China's future urban growth to impacts on ecosystem services under the shared socioeconomic pathways. Sci. Total Environ. 652, 765–779. https://doi.org/10.1016/j.scitotenv.2018.10.283.
Conover, W.J., 1999. Practical nonparametric statistics. In: Wiley Series in Probability and Statistics. Applied Probability and Statistics Section, third ed. Wiley, New York.
Cutler, F. original by L.B. and A. Wiener, R. port by A.L. and M., 2022. randomForest: Breiman and Cutler's Random Forests for Classification and Regression.
Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. Pattern Recogn. 46, 3483–3489. https://doi.org/10.1016/j.patcog.2013.05.018.
Descombes, P., Walthert, L., Baltensweiler, A., Meuli, R.G., Karger, D.N., Ginzler, C., Zurell, D., Zimmermann, N.E., 2020. Spatial modelling of ecological indicator values improves predictions of plant distributions in complex landscapes. Ecography 43, 1448–1463. https://doi.org/10.1111/ecog.05117.
Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.
Du, G., Shin, K.J., Yuan, L., Managi, S., 2018. A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of Greater Tokyo Area. Int. J. Geogr. Inf. Sci. 32, 757–782. https://doi.org/10.1080/13658816.2017.1410550.
Escobar, F., 2018. The NASZ model. In: Camacho Olmedo, M.T., Paegelow, M., Mas, J.-F., Escobar, Francisco (Eds.), Geomatic Approaches for Modeling Land Change Scenarios, Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham, pp. 461–464. https://doi.org/10.1007/978-3-319-60801-3_29.
Feng, Y., Liu, Y., Tong, X., 2018. Comparison of metaheuristic cellular automata models: a case study of dynamic land use simulation in the Yangtze River Delta. Comput. Environ. Urban Syst. 70, 138–150. https://doi.org/10.1016/j.compenvurbsys.2018.03.003.
Gago-Silva, A., Ray, N., Lehmann, A., 2017. Spatial dynamic modelling of future scenarios of land use change in Vaud and Valais, Western Switzerland. ISPRS Int. J. Geo-Inf. 6, 115. https://doi.org/10.3390/ijgi6040115.
Gerecke, M., Hagen, O., Bolliger, J., Hersperger, A.M., Kienast, F., Price, B., Pellissier, L., 2019. Assessing potential landscape service trade-offs driven by urbanization in Switzerland. Palgrave Commun 5, 109. https://doi.org/10.1057/s41599-019-0316-8.
Gonseth, Y., Wohlgemuth, T., Sansonnens, B., Buttler, A., 2001. Die biogeographischen Regionen der Schweiz.
Gounaridis, D., Chorianopoulos, I., Symeonakis, E., Koukoulas, S., 2019. A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. Sci. Total Environ. 646, 320–335. https://doi.org/10.1016/j.scitotenv.2018.07.302.
Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (Eds.), 2006. Feature Extraction: Foundations and Applications, Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-35488-8.
Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Model. 199, 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017.
Ho, S.Y., Phua, K., Wong, L., Bin Goh, W.W., 2020. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns 1, 100129. https://doi.org/10.1016/j.patter.2020.100129.
Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, second ed. John Wiley & Sons, Inc., Hoboken, NJ, USA https://doi.org/10.1002/0471722146.
Kamusoko, C., Gamba, J., 2015. Simulating urban growth using a random forest-cellular automata (RF-CA) model. ISPRS Int. J. Geo-Inf. 4, 447–470. https://doi.org/10.3390/ijgi4020447.

Kolb, M., Mas, J.-F., Galicia, L., 2013. Evaluating drivers of land-use change and transition potential models in a complex landscape in Southern Mexico. Int. J. Geogr. Inf. Sci. 27, 1804–1827. https://doi.org/10.1080/13658816.2013.770517.
Kubus, M., 2018. The problem of redundant variables in random forests. Acta Univ. Lodz. Folia Oeconomica 6, 7–16. https://doi.org/10.18778/0208-6018.339.01.
Lambin, E.F., 1997. Modelling and monitoring land-cover change processes in tropical regions. Prog. Phys. Geogr. Earth Environ. 21, 375–393. https://doi.org/10.1177/030913339702100303.
Li, X., Chen, Y., 2020. Projecting the future impacts of China's cropland balance policy on ecosystem services under the shared socioeconomic pathways. J. Clean. Prod. 250, 119489 https://doi.org/10.1016/j.jclepro.2019.119489.
Li, X., Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int. J. Geogr. Inf. Sci. 16, 323–343. https://doi.org/10.1080/13658810210137004.
Mas, J., 2004. Modelling deforestation using GIS and artificial neural networks. Environ. Model. Software 19, 461–471. https://doi.org/10.1016/S1364-8152(03)00161-0.
Mas, J.-F., Kolb, M., Paegelow, M., Camacho Olmedo, M.T., Houet, T., 2014. Inductive pattern-based land use/cover change models: a comparison of four software packages. Environ. Model. Software 51, 94–111. https://doi.org/10.1016/j.envsoft.2013.09.010.
Mas, J.F., Paegelow, M., Camacho Olmedo, M.T., 2018. LUCC modeling approaches to calibration. In: Camacho Olmedo, Teresa, María, Paegelow, Martin, Mas, J.-F., Escobar, F. (Eds.), Geomatic Approaches for Modeling Land Change Scenarios, Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham, pp. 11–25. https://doi.org/10.1007/978-3-319-60801-3_2.
Moulds, S., Buytaert, W., Mijic, A., 2015. An open and extensible framework for spatially explicit land use change modelling: the lulcc R package. Geosci. Model Dev. (GMD) 8, 3215–3229. https://doi.org/10.5194/gmd-8-3215-2015.
Mustafa, A., Rienow, A., Saadi, I., Cools, M., Teller, J., 2018. Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. Eur. J. Remote Sens. 51, 391–401. https://doi.org/10.1080/22797254.2018.1442179.
Newland, C.P., Maier, H.R., Zecchin, A.C., Newman, J.P., van Delden, H., 2018a. Multi-objective optimisation framework for calibration of Cellular Automata land-use models. Environ. Model. Software 100, 175–200. https://doi.org/10.1016/j.envsoft.2017.11.012.
Newland, C.P., Zecchin, A.C., Maier, H.R., Newman, J.P., van Delden, H., 2018b. Empirically derived method and software for semi-automatic calibration of Cellular Automata land-use models. Environ. Model. Software 108, 208–239. https://doi.org/10.1016/j.envsoft.2018.07.013.
Paegelow, M., Camacho Olmedo, M.T., Mas, J.F., 2018. Techniques for the validation of LUCC modeling outputs. In: Camacho Olmedo, Teresa, María, Paegelow, Martin, Mas, J.-F., Escobar, F. (Eds.), Geomatic Approaches for Modeling Land Change Scenarios, Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham, pp. 53–80. https://doi.org/10.1007/978-3-319-60801-3_4.
Paegelow, M., Camacho Olmedo, M.T., Mas, J.-F., Houet, T., 2014. Benchmarking of LUCC modelling tools by various validation techniques and error analysis. Cybergeo Rev. Eur. Géographie Eur. J. Geogr.
Pereira, D.G., Afonso, A., Medeiros, F.M., 2015. Overview of Friedman's test and post-hoc analysis. Commun. Stat. Simulat. Comput. 44, 2636–2653. https://doi.org/10.1080/03610918.2014.931971.
Pontius, R.G.J., Parmentier, B., 2014. Recommendations for using the relative operating characteristic (ROC). Landsc. Ecol.
Price, B., Kienast, F., Seidl, I., Ginzler, C., Verburg, P.H., Bolliger, J., 2015. Future landscapes of Switzerland: risk areas for urbanisation and land abandonment. Appl. Geogr. 57, 32–41. https://doi.org/10.1016/j.apgeog.2014.12.009.
R core team, 2021. R: A Language and Environment for Statistical Computing.
Ren, Y., Lü, Y., Comber, A., Fu, B., Harris, P., Wu, L., 2019. Spatially explicit simulation of land use/land cover changes: current coverage and future prospects. Earth Sci. Rev. 190, 398–415. https://doi.org/10.1016/j.earscirev.2019.01.001.
Rienow, A., Mustafa, A., Krelaus, L., Lindner, C., 2021. Modeling urban regions: comparing random forest and support vector machines for cellular automata. Trans. GIS 25, 1625–1645. https://doi.org/10.1111/tgis.12756.
Rodrigues, H., Soares-Filho, B., 2018. A short presentation of Dinamica EGO. In: Camacho Olmedo, M.T., Paegelow, M., Mas, J.-F., Escobar, F. (Eds.), Geomatic Approaches for Modeling Land Change Scenarios. Springer International Publishing, Cham, pp. 493–498. https://doi.org/10.1007/978-3-319-60801-3_35.
Roelofs, R., 2019. Measuring Generalization and Overfitting in Machine Learning. UC Berkeley.
Roodposhti, M.S., Aryal, J., Bryan, B.A., 2019. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environ. Model. Software 112, 70–81. https://doi.org/10.1016/j.envsoft.2018.10.006.
Roodposhti, M.S., Hewitt, R.J., Bryan, B.A., 2020. Towards automatic calibration of neighbourhood influence in cellular automata land-use models. Comput. Environ. Urban Syst. 79, 101416 https://doi.org/10.1016/j.compenvurbsys.2019.101416.
Santé, I., García, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. Landsc. Urban Plann. 96, 108–122. https://doi.org/10.1016/j.landurbplan.2010.03.001.
Schaldach, R., Priess, J.A., 2008. Integrated models of the land system: a review of modelling approaches on the regional to global scale. Living Rev. Landsc. Res. 2 https://doi.org/10.12942/lrlr-2008-1.
Shafizadeh-Moghadam, H., 2019. Improving spatial accuracy of urban growth simulation models using ensemble forecasting approaches. Comput. Environ. Urban Syst. 76, 91–100. https://doi.org/10.1016/j.compenvurbsys.2019.04.005.
Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., Unterthiner, T., Ernst, F.G.M., 2020. ROCR: Visualizing the Performance of Scoring Classifiers.
Soares-Filho, B., Rodrigues, H., Follador, M., 2013. A hybrid analytical-heuristic method for calibrating land-use change models. Environ. Model. Software 43, 80–87. https://doi.org/10.1016/j.envsoft.2013.01.010.
Sohl, T.L., Claggett, P.R., 2013. Clarity versus complexity: land-use modeling as a practical tool for decision-makers. J. Environ. Manag. 129, 235–243. https://doi.org/10.1016/j.jenvman.2013.07.027.
Stańczyk, U., 2015. Feature evaluation by filter, wrapper, and embedded approaches. In: Stańczyk, U., Jain, L.C. (Eds.), Feature Selection for Data and Pattern Recognition, Studies in Computational Intelligence. Springer, Berlin, Heidelberg, pp. 29–44. https://doi.org/10.1007/978-3-662-45620-0_3.
Swiss Federal Office for Statistics (SFSO), Section Geoinformation, 2021. Areal Statistics According to Nomenclature 2004, Surveys 1979-1985, 1992-1997, 2004-2009, 2013-2018.
Swiss Federal Office of the Environment (FOEN), 2009. SonBase - the GIS Noise Database of Switzerland. FOEN, Bern.
Swiss Federal Office of Topography (Swisstopo), 2022. Swiss Map Vector 25. Swisstopo, Wabern.
Swiss Federal Office of Topography (Swisstopo), 2011. swissTLM3D. Swisstopo, Wabern.
Swiss Federal Office of Topography (Swisstopo), 2007. VECTOR25 Hydrographic Network GWN07. Swisstopo, Wabern.
Tobler, W.R., 1979. Cellular geography. In: Gale, S., Olsson, G. (Eds.), Philosophy in Geography. Springer Netherlands, Dordrecht, pp. 379–386. https://doi.org/10.1007/978-94-009-9394-5_18.
Tong, X., Feng, Y., 2020. A review of assessment methods for cellular automata models of land-use change and urban growth. Int. J. Geogr. Inf. Sci. 34, 866–898. https://doi.org/10.1080/13658816.2019.1684499.
Torrens, P.M., 2011. Calibrating and validating cellular automata models of urbanization. In: Yang, X. (Ed.), Urban Remote Sensing. John Wiley & Sons, Ltd, Chichester, UK, pp. 335–345. https://doi.org/10.1002/9780470979563.ch23.
van Schrojenstein Lantman, J., Verburg, P.H., Bregt, A., Geertman, S., 2011. Core principles and concepts in land-use modelling: a literature review. In: Koomen, E., Borsboom-van Beurden, J. (Eds.), Land-Use Modelling in Planning Practice. GeoJournal Library. Springer Netherlands, Dordrecht, pp. 35–57. https://doi.org/10.1007/978-94-007-1822-7_3.
van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S., Verburg, P.H., 2016. A review of current calibration and validation practices in land-change modeling. Environ. Model. Software 82, 174–182. https://doi.org/10.1016/j.envsoft.2016.04.017.
van Vliet, J., Naus, N., van Lammeren, R.J.A., Bregt, A.K., Hurkens, J., van Delden, H., 2013. Measuring the neighbourhood effect to calibrate land use models. Comput. Environ. Urban Syst. 41, 55–64. https://doi.org/10.1016/j.compenvurbsys.2013.03.006.
Verburg, P.H., de Nijs, T.C.M., Ritsema van Eck, J., Visser, H., de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. Comput. Environ. Urban Syst. 28, 667–690. https://doi.org/10.1016/j.compenvurbsys.2003.07.001.
Verstegen, J.A., Karssenberg, D., van der Hilst, F., Faaij, A.P.C., 2016. Detecting systemic change in a land use system by Bayesian data assimilation. Environ. Model. Software 75, 424–438. https://doi.org/10.1016/j.envsoft.2015.02.013.
White, R., Engelen, G., 1993. Cellular automata and Fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environ. Plan. Econ. Space 25, 1175–1199. https://doi.org/10.1068/a251175.
White, R., Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environ. Plann. B Plann. Des. 24, 235–246. https://doi.org/10.1068/b240235.
White, R., Uljee, I., Engelen, G., 2012. Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. Int. J. Geogr. Inf. Sci. 26, 1251–1280. https://doi.org/10.1080/13658816.2011.635146.
Wiederkehr, M., Möri, A., 2013. Swissalti3D : a New Tool for Geological Mapping. https://doi.org/10.5169/SEALS-391140.
Yang, Q., Li, X., Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. Comput. Geosci. 34, 592–602. https://doi.org/10.1016/j.cageo.2007.08.003.
Zhang, D., Liu, X., Wu, Xiaoyu, Yao, Y., Wu, Xinxin, Chen, Y., 2019. Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng, China. GIScience Remote Sens. 56, 282–308. https://doi.org/10.1080/15481603.2018.1507074.