GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

EFForTS-LGraf: A Landscape Generator for Creating Smallholder-Driven Land-Use Mosaics

Jan Salecker, Claudia Dislich, Kerstin Wiegand, Katrin M. Meyer, Guy Pe'er

Bibliographische Information der Deutschen Nationalbibliothek

Abstract: Spatially-explicit simulation models are commonly used to study complex ecological and socio-economic research questions. Often these models depend on detailed input data, such as initial land-cover maps to set up model simulations. Here we present the landscape generator EFFortS-LGraf that provides artificially-generated land-use maps of agricultural landscapes shaped by small-scale farms. EFForTS-LGraf is a process-based landscape generator that explicitly incorporates the human dimension of land-use change. The model generates roads and villages that consist of smallholder farming households. These smallholders use different establishment strategies to create fields in their close vicinity. Crop types are distributed to these fields based on crop fractions and specialization levels. EFForTS-LGraf model parameters such as household area or field size frequency distributions can be derived from household surveys or geospatial data. This can be an advantage over the abstract parameters of neutral landscape generators. We tested the model using oil palm and rubber farming in Indonesia as a case study and validated the artificially-generated maps against classified satellite images. Our results show that EFForTS-LGraf is able to generate realistic land-cover maps with properties that lie within the boundaries of landscapes from classified satellite images. An applied simulation experiment on landscape-level effects of increasing household area and crop specialization revealed that larger households with higher specialization levels led to spatially more homogeneous and less scattered crop type distributions and reduced edge area proportion. Thus, EFForTS-LGraf can be applied both to generate maps as inputs for simulation modelling and as a stand-alone tool for specific landscape-scale analyses in the context of ecological-economic studies of smallholder farming systems.

Keywords: landscape generator; agent-based model; ABM; NetLogo; process-based; Indonesia

EFForTS-LGraf: A Landscape Generator for Creating Smallholder-Driven Land-Use Mosaics

Jan Salecker, Claudia Dislich, Kerstin Wiegand, Katrin M. Meyer, Guy Pe'er

EFForTS Discussion Paper Series No. 29

<u>May 2019</u>

EFForTS-LGraf: A Landscape Generator for Creating Smallholder-Driven Land-Use Mosaics

Jan Salecker[1*], Claudia Dislich[1], Kerstin Wiegand[1,2], Katrin M. Meyer[1], and Guy Pe'er[3,4,5]

[1]*Ecosystem Modelling, Faculty of Forest Sciences and Forest Ecology, University of Goettingen, Germany*
[2]*Centre of Biodiversity and Sustainable Land Use (CBL), University of Goettingen, Germany*
[3]*Synthesis Centre (sDiv) of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany*
[4]*UFZ - Helmholtz Centre for Environmental Research, Dept. Economics and Dept. Ecosystem Services, Leipzig, Germany*
[5]*University of Leipzig, Germany*

*Corresponding author: E-mail: jsaleck@gwdg.de

Abstract

Spatially-explicit simulation models are commonly used to study complex ecological and socio-economic research questions. Often these models depend on detailed input data, such as initial land-cover maps to set up model simulations. Here we present the landscape generator EFFortS-LGraf that provides artificially-generated land-use maps of agricultural landscapes shaped by small-scale farms. EFForTS-LGraf is a process-based landscape generator that explicitly incorporates the human dimension of land-use change. The model generates roads and villages that consist of smallholder farming households. These smallholders use different establishment strategies to create fields in their close vicinity. Crop types are distributed to these fields based on crop fractions and specialization levels. EFForTS-LGraf model parameters such as household area or field size frequency distributions can be derived from household surveys or geospatial data. This can be an advantage over the abstract parameters of neutral landscape generators. We tested the model using oil palm and rubber farming in Indonesia as a case study and validated the artificially-generated maps against classified satellite images. Our results show that EFForTS-LGraf is able to generate realistic land-cover maps with properties that lie within the boundaries of landscapes from classified satellite images. An applied simulation experiment on landscape-level effects of increasing household area and crop specialization revealed that larger households with higher specialization levels led to spatially more homogeneous and less scattered crop type distributions and reduced edge area proportion. Thus, EFForTS-LGraf can be applied both to generate maps as inputs for simulation modelling and as a stand-alone tool for specific landscape-scale analyses in the context of ecological-economic studies of smallholder farming systems.

Keywords: landscape generator, agent-based model, ABM, NetLogo, process-based, Indonesia

1 Introduction

Over the last decades simulation models have increased in complexity as a response to increasing demands on the robustness of model predictions (Grimm & Berger, 2016). Recent simulation models tend to increasingly consider more details of real world systems, such as environmental heterogeneity (Grimm & Berger, 2016). For example, recent spatially-explicit simulation models in land-use science typically incorporate environmental heterogeneity by modelling the spatio-temporal distribution and dynamics of land-cover types (e.g. Ayllón *et al.* (2016), Tietjen (2016)).

However, environmental data, such as land-cover maps, can often not be obtained at the necessary level of detail. This problem is especially prevalent in tropical regions, where constant cloud coverage limits remote sensing data usability. In such cases, as well as for systematic research analyses, it may be necessary to artificially generate land-cover maps that approximate reality. For this purpose, landscape generators (also called landscape simulators) have been developed (Langhammer *et al.*, 2019). A typical landscape generator creates landscapes consisting of a grid of cells, where each cell is assigned a given land-cover type. Landscape generators can also be used to systematically generate many similar land-cover maps and thereby allow for scenario-based control of landscape characteristics. Moreover, the option of producing a range of (also non-realistic) landscapes, e.g. by setting population density or field sizes to unrealistically high values, is an important tool for testing potential scenarios and improving the understanding of model processes.

Two approaches are known for landscape generators: pattern-based landscape generators and process-based landscape generators (Langhammer *et al.*, 2019). Pattern-based landscape generators allow for control over the amount and spatial distribution of cover types with relatively low model complexity and only a small number of parameters. Many pattern-based landscape generators are based on artificial (e.g. fractal) algorithms and the resulting landscapes are often known as neutral landscapes (Gardner *et al.*, 1987; With, 1997). However, real world land-cover patterns are often the result of historical land-use change by human interaction with the landscape and as such a result of the processes that shape these landscapes. Furthermore, the outcomes of these interactions often result in spatial patterns that differ substantial from those shaped by neutral processes. Thus, in contrast to pattern-based landscape generators, process-based generators try to simulate the processes that lead to the pattern (Schröder & Seppelt, 2006). The process-based approach can be more costly regarding the number of parameters and model complexity. However, parameters often have an empirical meaning and can be measured with surveys allowing for creation of artificial landscapes that mimic properties of real world landscapes of a certain study area. The resulting land-cover maps are often used to inform applied simulation modelling studies, oftentimes for model initialization.

Two key processes driving anthropogenic land-use changes are the rapid expansion of roads (Laurance *et al.*, 2009; Ibisch *et al.*, 2016) and the related expansion and intensification of agricultural fields (Laurance *et al.*, 2014). Models that incorporate these processes into models of landscape design, such as DYPAL (Gaucherel *et al.*, 2006) and G-RaFFe (Pe'er *et al.*, 2013), have been shown to successfully produce realistic ranges of landscape characteristics for anthropogenically-altered landscapes.

For some model applications, however a land-cover map alone is not sufficient. For example, EFForTS-ABM, an agent-based model studying land-use change in agricultural landscapes dominated by smallholders in Indonesia, requires additional information regarding fields (as agricultural units) and land ownership, because smallholder households are modelled as agents that own these fields (Dislich *et al.*, 2018). The determination of fields and allocation to specific households cannot be done using remote sensing data and is difficult to obtain using only

pattern-based landscape generators. In consequence, the human perspective is often lacking completely in landscapes used for ecological research (including modelling), and in fact, forming one of the key gaps between ecological and socioeconomic research. Thus, an extension of the process-based approach to landscape generation is needed that overcomes this mismatch and incorporates the human dimension of landscape ownership and management.

Here, we present a new model, EFForTS-LGraf, for creating land-cover maps that represent agricultural areas dominated by smallholders. We focus on smallholders because they comprise a large proportion of farmers in most parts of the world. The model allows for flexible parameterization of the main processes shaping these landscapes, i.e. road creation and field establishment by smallholder farming households. The model algorithms follow the assumption that the creation and expansion of agricultural land is connected to road establishment, as has been reported in several case studies (e.g., Gatto *et al.* (2015), Kirby *et al.* (2006), Soares-Filho *et al.* (2001)). We build upon G-Raffe, an existing landscape generator that simulates the process of field establishment along roads (Pe'er *et al.*, 2013) but does not explicitly incorporate the household dimension. We extended the algorithms of G-Raffe by using an agent-based modelling approach and introducing smallholder farming households that have a specific home-base and a given household size.

We present a detailed model description of EFForTS-LGraf, using the ODD (Overview, Design concepts and Details) protocol for describing agent-based models (Grimm *et al.*, 2006, 2010) and the ODD+D (ODD + Decision) extension of the protocol for describing agent-based models that involve human decisions (Müller *et al.*, 2013). We take smallholder farming in Jambi province, Indonesia, as an example to demonstrate model applications using three approaches (Details, see Section 3): (1) We use a sensitivity analysis to quantify effects of EFForTS-LGraf model parameters on properties of the landscapes generated. (2) We compare a classified land-use map from our study region to the landscapes generated with EFForTS-LGraf. (3) In an applied case study we investigate effects of household consolidation and specialization on landscape patterns. All approaches are based on quantification of landscape characteristics via landscape metrics.

2 Model description

2.1 Overview

2.1.1 Purpose

The general goal of the EFForTS-LGraf landscape generator is to create artificial maps of landscapes that are dominated, or strongly shaped, by agricultural activities. The grid-based maps include fields of various sizes and different crop types and other potential land-cover types as desired. These other potential land-cover types, such as forest, grassland, water bodies or degraded land are grouped into a single land-cover type (here, we used the general term 'others'). In addition to fields, the model considers land ownership by assigning each field to a farming household agent. Artificial land-cover maps produced by EFForTS-LGraf can be used as a template, or input, for other models which, for example, can simulate the effects of land-use types on ecological and/or economic functions. The resulting maps may also be used as a starting point to analyse how farmer decisions alter land-use and shape land-use changes.
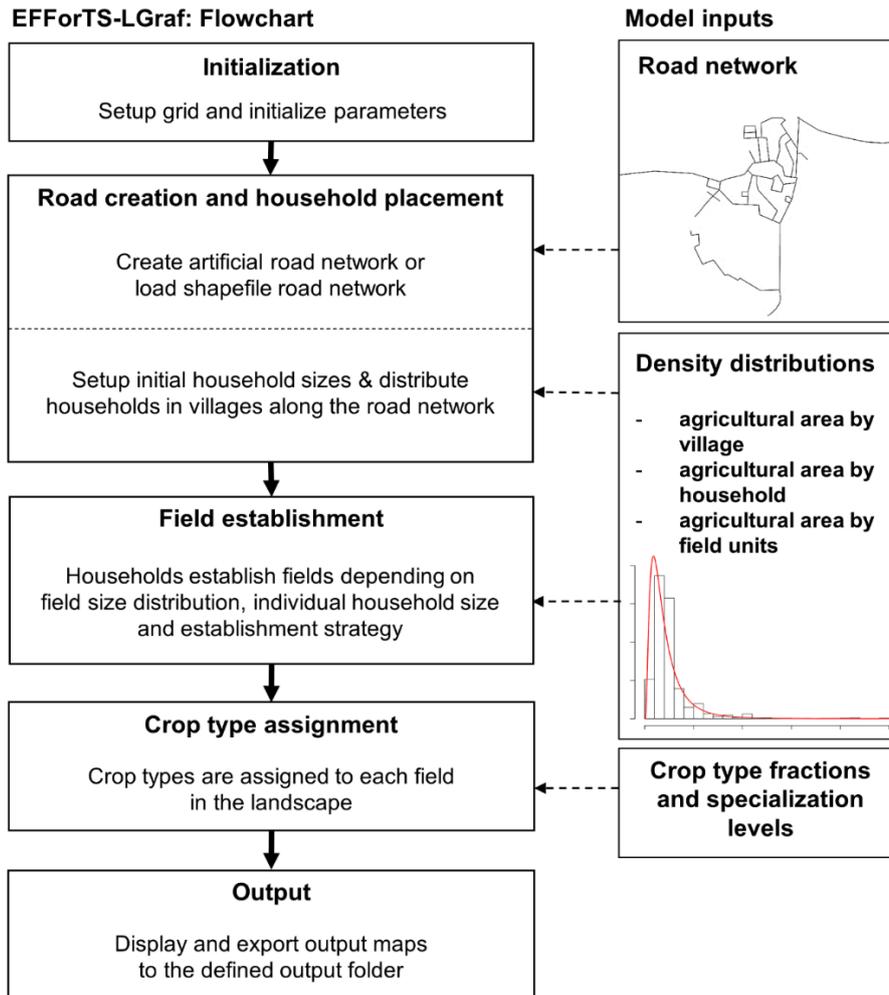
Figure 1: EFForTS-LGraf flowchart including process flow of main model processes and model inputs.

2.1.2 Entities, state variables and scales

The simulated landscape comprises several spatial units: cells, fields, households and the landscape. Roads are a non-spatial unit of EFForTS-LGraf as they are similar to polylines in a GIS context. Cells are the smallest spatial unit of EFForTS-LGraf. The landscape consists of a regular grid of these cells. The cell size can be set by the user and should approximately correspond to the smallest size of fields in the landscape. The extent of the landscape can also be set by the user. Cell attributes include its land-cover type, e.g. 'others' or 'field', and the household the cell belongs to (if any). Moreover, each cell has the attribute of whether it contains an intersecting road or not. The fields are composed of one or several contiguous cells that have all the same crop type and belong to (or are managed by) the same household. Each household has a home-base cell and owns one or several fields that do not need to occur next to each other. The household is represented by an agent that is establishing fields close to the household's home-base cell during the simulation process.

Figure 2: Output map examples of EFForTS-LGraf. All maps include patches of inaccessible area (brown color) and roads (black lines). In the field ownership map (upper left), hues indicate field owners. In the crop type map (upper right) colors indicate fields with different crops. The agriculture-non-agriculture map (lower left), is a binary map that differentiates agricultural cells (purple) from other cells (grey). The 'others' patches map (lower right) is similar to the agriculture-non-agriculture map but shows each separate patch of class 'other' in another color.
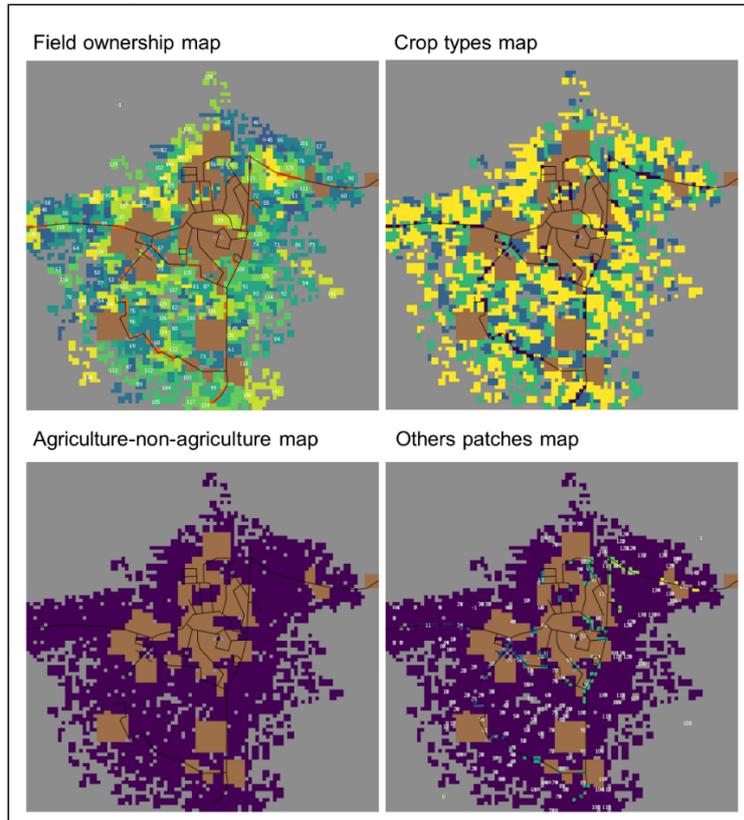
2.1.3 Process overview and scheduling

EFForTS-LGraf is initialized with a landscape completely covered by cells of the general class 'others' (Fig. 1). Within the procedure "Road creation and household placement" a network of roads is established either artificially or taken as input from a realistic road map (see Road creation and household placement in Section 2.3.2). The number of households in the landscape is determined and households are placed onto the map by assigning home-base cells. All home-base cells belong to the road network, i.e. home-base cells always have a road intersecting them. The spatial distribution of households may be completely random or aggregated in villages, depending on the village size distribution. In the "Field establishment" procedure, the households establish fields close to their home-base cell. This procedure is designed such that the resulting frequency distribution of field sizes as well as the resulting frequency distribution of household areas approximate the expected distributions set by the input parameters (see Field establishment in Section 2.3.2). Thereafter, the procedure "Crop type assignment" assigns crop types to the established fields based on input parameters such as fractions of crop types and specialization degree (see Crop type assignment in Section 2.3.2). Finally, different maps of the simulated

landscape are produced as model output, such as land-cover map and land-ownership map (see Output data in Section 2.3.2 and Fig. 2).

2.2 Design concepts

2.2.1 Theoretical and empirical background

The main processes in EFForTS-LGraf build upon the assumption that households settle along roads and establish fields in close vicinity to their home. Such patterns have been reported in several case studies (e.g., Gatto *et al.* (2015), Kirby *et al.* (2006), Soares-Filho *et al.* (2001)) and have successfully been implemented in other landscape generators (Pe'er *et al.*, 2013). EFForTS-LGraf does not incorporate any temporal dimension. Instead of simulating the process of land-use change over time explicitly, distributions of household and field sizes are used to create a landscape at one specific point in time. The model does not incorporate any environmental heterogeneity and all considered crop types are equally suitable throughout the landscape. However, the amount and distribution of crop types can be controlled by adjusting landscape proportions and household specialization levels for each crop type.

2.2.2 Individual decision-making

Households establish fields based on several decisions. They search for an unoccupied cell in the vicinity of their home-base and try to establish a field of a given size there. If this action is not successful, e.g. because the area between cells that are already occupied is too small, they continue to try in a different location. Finding an unoccupied cells depends on the current field establishment strategy (for Details see Field establishment in Section 2.3.2). If a household fails to establish a field for a given number of times under the current strategy, it switches the establishment strategy. Current establishment strategies include searching for unoccupied cells in vicinity of the home-base, in vicinity of already established fields, in neighboring cells of class 'others' with gradually increasing search radius, and searching for unoccupied cells that are surrounded only by unoccupied cells. If the attempt to establish a field of the given size fails for all potential strategies, a new field size is drawn from the field size distribution.

2.2.3 Individual sensing

As in reality, households are assumed to have full knowledge on land-use types and ownership of all cells in the landscape. A household cannot expand a new field into an already existing field, whether it is owned by a different household or by the same household.

2.2.4 Interaction

Households interact indirectly by land ownership, e.g. one household cannot expand a new field into the field of a different household.

2.2.5 Heterogeneity

Household agents are heterogeneous in their expected and realized household area. Households also differ in number, size and shape of established fields. Additionally, households may cultivate different proportions of crop types and some households may also specialize on one specific crop type. Households are aggregated in villages and have different initial home-base positions. The initial size of the village and home-base position may affect field establishment by the household, which may for instance result in higher distances between home-base and fields (in densely populated areas). These aspects of heterogeneity are mostly imposed by external inputs that determine the overall distribution of field sizes, household areas and village areas.

2.2.6 Stochasticity

The location of household home-base cells along the road network is random. However, aggregation of households at certain locations can be controlled by providing a village size distribution (see Road creation and household placement in Section 2.3.2). Field sizes and household areas are drawn from a given probability distribution and are therefore also stochastic. In addition, in the case that the user chooses to create the road network within EFForTS-LGraf, the algorithm for road generation randomly assigns the location, direction and length of each road segment (see Section 2.3.2).

2.2.7 Observation/Emergence

The spatial arrangement of fields is an emergent property of the model and will thus differ between simulations. The emerging patterns can be compared to real maps or used to generate a set of virtual land-use maps. Aggregated model outputs that can be generated from the maps include spatially implicit data such as maximum distance between roads and fields.

2.3 Details

EFForTS-LGraf was implemented in NetLogo 6.0.2.

2.3.1 Input data and Initialization

Table 1. EFForTS-LGraf model parameters

id	Name on GUI	Unit	Description
t_s	setup-type	[-]	*households*: number of villages and agricultural area are approximated by providing a fixed number of households; *villages*: number of households and agricultural area are approximated by providing a fixed number of villages; *area*: number of households and villages are approximated by providing the total agricultural area
$n_{s,h}$	number-of-farmers	[-]	number of farming households in landscape
$n_{s,v}$	number-of-villages	[-]	number of villages in landscape
$n_{s,a}$	prop-agricultural-area	[-]	proportion of agricultural area in landscape
$n_{s,c}$	households-per-cell	[-]	maximum number of household home-bases in one cell
$seed_s$	rnd-seed	[-]	random seed of the simulation, only used when rep_s is true
rep_s	reproducable?	[-]	if true, the user-set random seed $seed_s$ is used
w_s	width	cell	width of the landscape grid
h_s	height	cell	height of the landscape grid
c_s	cell-length-meter	cell	side length in meter of one cell of the landscape grid
t_r	road-type	[-]	type of road algorithm (*shapefile, artificial.perlin* or *artificial.graffe*)
$i_{r,shp}$	road-map-nr	[-]	number of road map file (only used when t_r is *shapefile*)
$n_{r,art}$	total-road-length	cell	total number of road cells in landscape (only used when t_r is *artificial.perlin* or *artificial.graffe*)

$m_{r,art}$	min-dist-roads	cell	minimum distance [cells] between two roads (only used when t_r is *artificial.perlin* or *artificial.graffe*)
$p1_{r,perl}$	perlin-octaves	cell	octaves parameter for the perlin algorithm (only used when t_r is *artificial.perlin*)
$p2_{r,perl}$	perlin-persistence	cell	persistence parameter for the perlin algorithm (only used when t_r is *artificial.perlin*)
$p3_{r,perl}$	cone-angle	cell	cone-angle parameter for the perlin algorithm (only used when t_r is *artificial.perlin*)
$p4_{r,perl}$	dist-weight	cell	distance versus elevation weighting for the perlin algorithm (only used when t_r is *artificial.perlin*)
d_v	vlg-size-distribution	[-]	type of distribution for village size (*constant, uniform, normal, log-normal*)
μ_v	vlg-size-mean_ha	hhs	mean of village size distribution
σ_v	vlg-size-sd_ha	hhs	standard deviation of village size distribution
d_v	vlg-min-distance	cell	minimum distance between villages
d_h	hh-area-distribution	[-]	type of distribution for household area (*normal, lognormal*)
μ_h	hh-area-mean_ha	ha	mean of household area distribution
σ_h	hh-area-sd_ha	ha	standard deviation of household area distribution
n_i	inaccessible-areafraction	[-]	fraction of landscape covered by inaccessible area (e.g. large-scale plantations, protected area)
l_i	inaccessible-arealocation	[-]	location of inaccessible areas (either *random* or *roadconnected*)
d_i	inaccessible-areadistribution	[-]	type of distribution for inaccessible area (*constant, uniform, normal*)
μ_i	inaccessible-area-mean	ha	mean of inaccessible area distribution
σ_i	inaccessible-area-sd	ha	standard deviation of inaccessible area distribution
t_f	field-type	[-]	*distribution*: field sizes are drawn from the provided distribution; *percentage*: mean of the field size distribution is adjusted to the percentage set by μ_h
d_f	field-size-distribution	[-]	type of distribution for field sizes (*constant, uniform, normal, log-normal*)
μ_f	field_size_mean_ha	ha	mean of field size distribution
σ_f	field_size_sd_ha	ha	standard deviation of field size distribution
p_f	field-size-percentage	[-]	sets percentage to adjust μ_f by multiplication with μ_h (only used when t_f is *percentage*)
s_f	field-shape-factor	[-]	controls if fields are mostly rectangular (value 1) or narrow (higher values)
t_{strat}	strategies-type	[-]	type of field strategies selection; *manual*: choose establishment strategies by manual selecting; *id*: use a predefined list of strategies
$s1_{strat}$	s1.homebase	[-]	field establishment strategy 1: establishment close to own home-base (true/false)

$s2_{strat}$	s2.fields	[-]	field establishment strategy 2: establishment close to own fields (true/false)
$s3_{strat}$	s3.nearby	[-]	field establishment strategy 3: establishment in nearby 'others' cell (true/false)
$s4_{strat}$	s4.avoid	[-]	field establishment strategy 4: establishment in nearby 'others' cell surrounded by 'others' cells (true/false)
n_{strat}	change-strategy	[-]	number of unsuccessful tries for field establishment after which search strategy is changed
I_{strat}	field-strategies-id	[-]	overwrites manual strategies selection and determines a pre-specified list of search strategies for field establishment (e.g. s1, s2, s4)
t_l	land-use-assignment	[-]	crop type assignment algorithm; *landscape-level-fraction*: approximate fractions of crop types; *landscape-levelspecialization*: approximate fractions of crop types and fractions of household specialization
id_l	LUT-l-name (l=1,2,3,4,5)	[-]	name of crop types l (l=1,2,3,4,5)
fr_l	LUT-l-fraction (l=1,2,3,4,5)	[-]	fraction of agricultural area under crop type l (l=1,2,3,4,5); fractions must sum up to 1
$fr_{l,spec}$	LUT-l-specialize (l=1,2,3,4,5)	[-]	minimum fraction of area under crop type l (l=1,2,3,4,5), which is farmed by specialist households
f_l	LUT-fill-up	[-]	crop type (ID) to fill up fractions if sum of fr_l is smaller than one

A potential external input for EFForTS-LGraf is a map of an existing road network (option real-road-map). Most of the model parameters (Table 1) can be estimated from empirical studies or remote sensing data. In addition to model parameters (numeric), switches (logical, categorical) allow turning on and off certain model processes (Table 1). Several distribution parameters affect model behaviour and outcomes: (1) A household size distribution, from which the total agricultural area of each household is sampled during model initialization. (2) A village size distribution, from which the village sizes (number of households per village) are sampled during model initialization. (3) A field size distribution, from which the areas of the single field units are sampled during field establishment. (4) Inaccessible area distribution, from which areas of each inaccessible area patch are sampled during model initialization. Different distribution types are possible and include constant, uniform, normal and log-normal shapes. Additionally, road parameters, crop type proportions, crop type specialization levels and field establishment strategies need to be set. Most of the model parameters can be estimated from empirical studies or remote sensing data.

The initialization procedure first updates world dimensions according to the parameters w_s and h_s. All cells in the landscape are initialized as 'others' cells, with no household or fields. At that stage, cells are not owned by anybody. Global variables are set according to the user inputs and the output plots of the model are refreshed. Next, a sample of the household size distribution is drawn to initialize household properties. While the creation and placement of household agents

is performed by the submodel "Road creation and Household placement" (see Section 2.3.2), the properties of the household agents are already determined during initialization. EFForTS-LGraf provides three options for household initialization via the parameter *setup-type* (t_s). Depending on the chosen *setup-type*, the user provides either (1) a fixed number of households $n_{s,h}$ (option: households). The number of villages and the proportion of agricultural area are then approximated by using the defined village size and household size distributions; (2) the number of villages $n_{s,v}$ (option: villages). The number of households and the proportion of agricultural area are then approximated by using the defined household size distribution; (3) the proportion of agricultural area $n_{s,a}$ (option: area). The number of households and the number of villages are then approximated by using the defined household size and village size distributions.

Each of the above options generates a preliminary list of households. Each household has three properties: household-ID, household size and village-ID. The household sizes approximate the defined household size distribution, whereas the village-IDs are assigned in such a way that resulting village sizes approximate the defined village size distribution.

2.3.2 Submodels

Road creation and household placement

After initialization, first all roads are created and then households are placed along the established road network. Roads are treated as landscape items without a dimension, i.e. like polylines in GIS. Just as any other cell, cells with an intersecting road have a land-use type. There are three options for road creation: (1) a road network is created based on an existing road map in an input file (option: real.shapefile); (2) a road network is artificially created based on a random elevation model (option: artificial.perlin); or (3) a road network is artificially created based on the straight road creation algorithm of the G-Raffe landscape generator (option: artificial.graffe). For details on road creation and household placement see section 1.1.1 in Appendix I.

Once the road network is established, households are created and placed along the road network. First, the algorithm determines the number of villages depending on the pre-generated list of households from the initialization procedure. Then, based on the number of villages and village-IDs, village centers are created on random road network cells complying with the minimum distance between village centers, d_v. Then the households are placed randomly on road cells around village centers matching the corresponding village-IDs, i.e. each household establishes a home-base cell at the assigned cell. There is a cap to the number of household home-bases on one cell $n_{s,c}$.

Inaccessible areas are an optional landscape feature of EFForTS-LGraf, defining patches of areas that are not available for use by smallholder agriculture. This option allows defining either areas belonging to large-scale company plantations or protected (forest) areas. Given the overall fraction of the landscape covered by inaccessible area, *frac$_p$* (Table 1) and the inaccessible area size distribution, patch sizes are drawn from the distribution until the total size of inaccessible area patches matches the defined landscape fraction. From this list, each patch is then created by first selecting a starting location, which can be either a random cell in the landscape (option *random* for l_i) or a random road cell (option *road-connected* for l_i). From this starting location, a square-shaped field of the given size is created, following the basic field establishment rules (details see section 1.1.2 in Appendix I).

Field establishment

The field establishment procedure determines the size and spatial location of fields, but does not yet determine the actual crop within fields. This is assigned in the next procedure (see "Land-use

assignment" below). Fields are established by household agents and an attempt to establish a field comprises three steps: (1) deciding on the field size, (2) moving to a potential location, and (3) making sure there is enough space to establish a field of the desired size in this location (for details see section 1.1.2 in Appendix I). In case of a successful attempt, the household gains ownership of these established field cells. By gaining ownership, the realized household area increases and field establishment continues until each household realizes its expected household size that was determined during initialization.

In other words, to realize the expected household size distribution, the procedure loops over all households that are marked as still growing, i.e. all households where the realized area of owned fields does not yet exceed the expected household size. At the first stage, these are all households because each household establishes at least one field. In every iteration, each relevant household draws a field size from the field-size distribution. If the projected household area, including the additional field, is below the expected household size, then the household attempts to establish a field of this size. If the projected area exceeds the expected household size, an attempt for establishment would take place only if the absolute difference between projected household area and expected household area is smaller after establishing the field. Otherwise, the household does not establish the field and becomes a non-growing household. After each establishment loop, households that have reached their final size become non-growing households, namely are excluded from the next iteration of field establishment. At the end of each loop, if the total realized agricultural area exceeds the expected cover of agricultural area, the field establishment procedure is halted.

Crop type assignment

After all fields are established, crop types are assigned to them. The current model version supports up to five different crop types and two alternative ways to distribute the crop types (parameter t_l). The first option (*landscape-level-fraction*) distributes crop types randomly among fields according to fractions of overall crop types. The second option (*household-level-specialization*) aims at additionally incorporating specialization for crop types at the household level. In this latter case, an additional input parameter is used for each crop type which describes the specialization by households for this particular crop type. The level of specialization is primarily a proportion ranging from 1 (all households that cultivate this crop type would cultivate this crop type exclusively) to zero (no preference for specialization in this crop type). For example, a specialization value of 0.7 for crop type 1 would mean that 70% of households that cultivate crop type 1 have only fields of crop type 1 and no field of any other crop type. The remaining 30% of households have fields of various crop types. Note that realized specialization levels can be higher than the input specialization levels, since the value determines a field-level outcome, and hence, all households with only one field are specialists by default.

Output data

The landscapes produced by EFForTS-LGraf contain information on various spatial scales (landscape level, household level, fields, cells) and can be visually inspected in different formats:
- Land-use maps: depict land uses in different colors (classes: agriculture, road, inaccessible, home-base, 'others')
- Crop type maps: similar to land-use map with additional classification of crop types
- Agriculture-non-agriculture map: depicts the distinction between agricultural cells and cells of class 'others'
- Field-patches map: depicts the different fields on the map in different colors
- Household-patches map: differentiates fields based on the different households they belong to

- Habitat-patches map: depicts clusters of cells belonging to the same patch of class 'others' in different colors. Roads, fields and inaccessible areas function as separators for patches.

Maps are produced at the resolution cell-length-meter c_s. Cell labels such as patch-IDs and owner-IDs, and spatial elements such as roads, home-bases and households can be selected to be drawn on top of these maps. All spatial outputs can also be stored as raster maps (ASCII) to allow using the generated landscapes for other model applications (Details on raster output see section 1.1.3 in Appendix I). An additional feature ("create-3D-map" function) uses the NetLogo 3D functionality to create a 3D rendered map that displays trees, crops and buildings using 3D-shapes in realistic densities (Fig. 4 in Appendix I).

3 Scenarios and Parameterization

Table 2. Landscape metrics description

landscape metric	short	range	description
landscape-shape-index	LSI	$LSI \geq 1, without\ limit$	measure of class aggregation or clumping
largest-patch-index	LPI	$0 < LPI \leq 100$	percentage of total landscape area comprised by the largest patch of a class
mean-patch-area	-	$\geq 0, without\ limit$	mean patch area of all patches of a class
n-patches	-	$\geq 0, without\ limit$	total number of patches of a class
patch-cohesion-index	PCI	$0 \leq PCI \leq 100$	physical connectedness of patches of a class

We demonstrate the capabilities and potential uses of EFForTS-LGraf based on the example of smallholder-dominated agricultural areas in Jambi province in Sumatra, Indonesia. During the last decades, this region has faced severe land-use changes, mainly deforestation and agricultural expansion (e.g., Koh (2011)) and loss of ecosystem-functioning of the transformed landscapes (e.g., Dislich *et al.* (2017)). In order to provide agricultural maps that incorporate smallholder households and field ownership, we parameterized EFForTS-LGraf using household data from a smallholder survey of relatively large size (701 farming households) that was performed in Jambi province (Fig. 3) (Euler *et al.*, 2017; Krishna *et al.*, 2017). Jambi province is characterized by small villages with farming households that are mostly of relatively small size (median survey data 3.5 ha) and within these villages, small-scale fields with mostly oil palm and rubber. By parameterizing EFForTS-LGraf for a specific study region, the generated maps can be used to inform scenario-based studies such as the application of the agent-based simulation model EFForTS-ABM (Dislich *et al.*, 2018). To show EFForTS-LGraf's model and output capabilities, we followed three approaches. In approach 1 (Sobol sensitivity analysis), we determined how variability in the landscapes generated by EFForTS-LGraf can be apportioned to the different model parameters. In approach 2 (Validation), we assessed the characteristics of typical landscapes in the Harapan region in Jambi province and applied an optimization algorithm in order to generate artificial landscapes with similar characteristics. In approach 3 (Applied case study), we present an applied simulation scenario that investigates effects of household consolidation and crop specialization on landscape characteristics. For all approaches, we quantified the landscape characteristics using five landscape metrics that are either class-based or aggregated on class-level: landscape-shape-index (LSI), largest-patch-index (LPI), mean-patch-

area, the number of patches (n-patches) and patch-cohesion-index (PCI) (Table 2, for a more detailed description see FRAGSTATS manual (McGarigal *et al.*, 2012)). We used the R-package *SDMTools v1.1* to calculate landscape metrics of all landscapes (VanDerWal *et al.*, 2014).
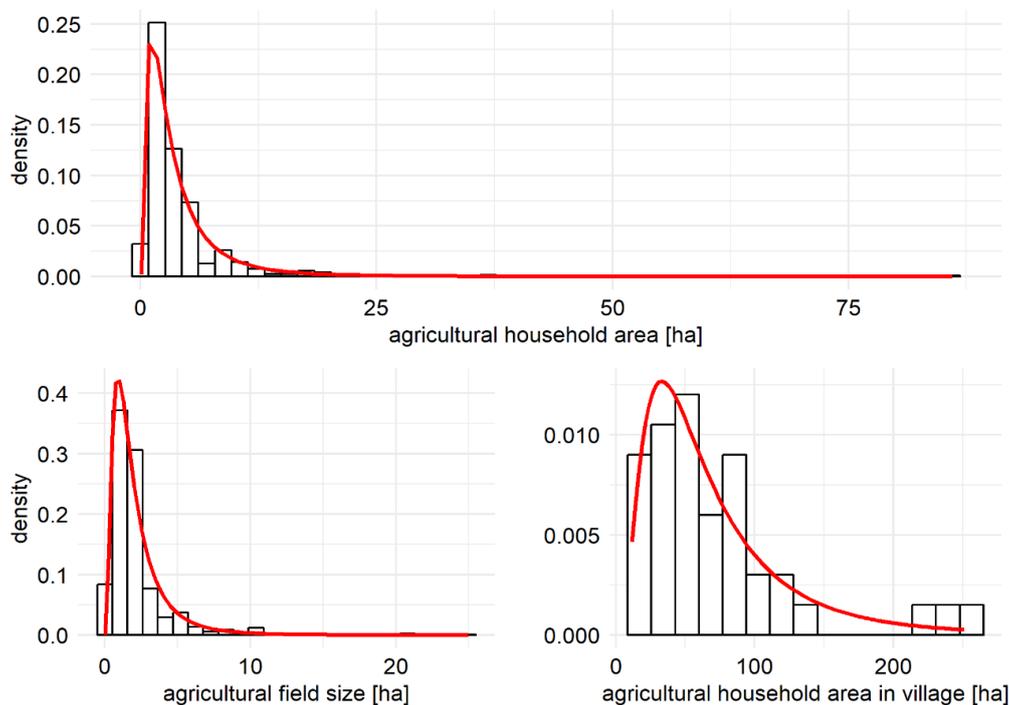

Figure 3: Distribution of household areas, village areas and field sizes, based on household surveys carried out in our study area in Jambi province, Sumatra, Indonesia.

For approach 1, we conducted a Sobol sensitivity analysis, i.e. a global variance-based sensitivity analysis for all model parameters including the parameters of the artificial road creation algorithm of EFForTS-LGraf (for parameterization details, see section 1.2.1 in Appendix I). The Sobol method measures direct effects and interaction effects of model parameters on model output (here: landscape metrics) (Sobol, 1990; Chan *et al.*, 2000; Saltelli *et al.*, 2010; Jansen, 1999). By applying a Sobol parameter sampling design, we generated 8000 different landscapes that cover a large parameter range of EFForTS-LGraf. Such analysis helps understand model processes and may be useful to reproduce certain landscape features. For instance, if one is interested in generating landscapes along a gradient of characteristics, the sensitivity analysis allows identifying which parameters can realize that gradient and how.

For approach 2 (Validation), we used a land-cover map (classified satellite image from 2013) from the Harapan region in Jambi province that features a large gradient of land-use intensities (Melati, 2017) (see Fig. 4). The original landcover map has a spatial resolution of 5 ×5 m and an overall classification accuracy of 68.4% (Melati, 2017). In order to allow for comparisons with our artificially-generated landscapes, we scaled the land-cover map to the same resolution as our generated EFForTS-LGraf landscapes, which is 50×50 m cells. The original classified land-cover map consisted of 9 classes which we reclassified into two final land-cover types: fields and 'others' (fields consist of original classes rubber and oil palm; 'others' consist of original classes secondary dryland forest, shrub, bare land, settlement, water body, cloud, shadow). While we did not explicitly assess the overall classification accuracy of the final reclassified and rescaled map,

overall classification accuracy is expected to improve trough reclassification into more general classes. From the reclassified land-cover map, we sampled 3 randomly placed landscapes, 100 × 100 cells in size (no overlaps) and calculated the five landscape metrics for each of these sampled landscapes and each land-cover type (fields and other). For each sample, we performed a genetic algorithm in order to recreate these samples with EFForTS-LGraf (Kumar *et al.*, 2010; Willighagen & Ballings, 2015). The algorithm uses the proportion of agricultural area from the samples, but varies all road, household, village and field establishment parameters. For each generated landscape, a fitness value is calculated by comparing the landscape metrics of the generated landscape to the landscape metrics of the current sample. The algorithm then tries to minimize the total deviance by repeated creation of landscapes with adjusted parameters. The genetic algorithm was set up with 50 different initial parameterizations per sample (population size) and 25 iterations. In approach 2 (validation), we also compared the ranges of landscape metrics of the 8000 generated landscape from the Sobol sensitivity analysis (approach 1) to landscape metrics of 100 landscapes. These landscapes were randomly sampled (allowing overlaps) from the reclassified land-use map (results, see section 1.2.2 in Appendix I).

EFForTS-LGraf not only allows us to generate land-cover maps, but also to conduct applied modelling studies (approach 3). Empirical data from Jambi, Indonesia indicates an increase over time in the area owned by households (Euler *et al.*, 2016). Smallholder households also tend to specialize on one specific crop type (Klasen *et al.*, 2016). Increasing household area and specialization on one crop may cause drastic changes in landscape composition and configuration. We analyzed whether EFForTS-LGraf can reproduce these changes by generating a set of landscapes with different levels of crop specialization and household area using a Latin hypercube sampling design with 500 samples (McKay *et al.*, 1979). We randomly selected one village from the household survey in Jambi and cropped a road polyline shapefile based on a spatial layer covering the road network of Jambi. We used the classified satellite image of the village to estimate the proportions of oil palm and rubber in the agricultural landscape (*oilpalm* = 0.5, *rubber* = 0.5).

To mimic the increase in household area, we set the proportion of agricultural area in the landscape to a fixed value of 50% and varied the mean of the log-normal household area distribution from 1 ha to 3 ha within the Latin hypercube sampling design. The total number of households in the landscape was then estimated based on the proportion of agricultural land. With increasing mean values of the household area distribution, this yielded fewer households but the same total agricultural area in the landscape. We used two crop types (*oil-palm* and *rubber*) and varied the specialization levels for *oil-palm* from 0 (specialist by chance) to 1 (always specialists) within the Latin hypercube sampling design. For crop type 2 *rubber* we set the specialization level to 0 (for parameterization details, see section 1.2.3 in Appendix I). We analyzed the resulting landscapes via the five selected landscape metrics from previous approaches (Table 2). We calculated linear regression models for each landscape metric and crop type combination and calculated standardized regression coefficients to estimate parameter and interaction effects on landscape metrics of the generated landscapes.

Execution of NetLogo simulations and output post-processing where performed with R and the R-package nlrx (R Core Team, 2018; Salecker & Sciaini, 2018).
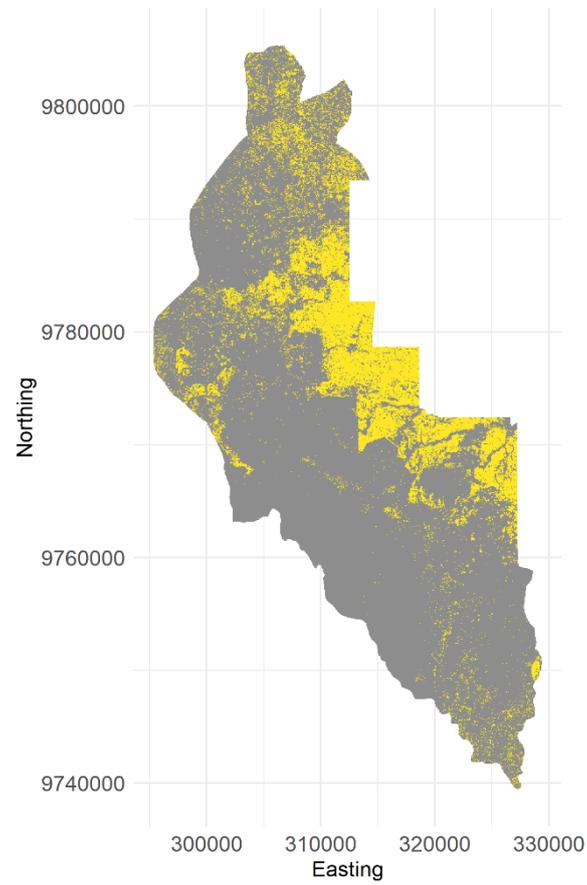
Figure 4: Snapshot of the reclassified satellite image of Harapan region in Jambi province. Grey cells indicate land-cover type 'others', which consist mostly of secondary forest but includes all other remaining nonagricultural land-cover classes, such as settlements and water bodies. Yellow cells indicate fields, which consist of oil palm and rubber plantations.

4 Results

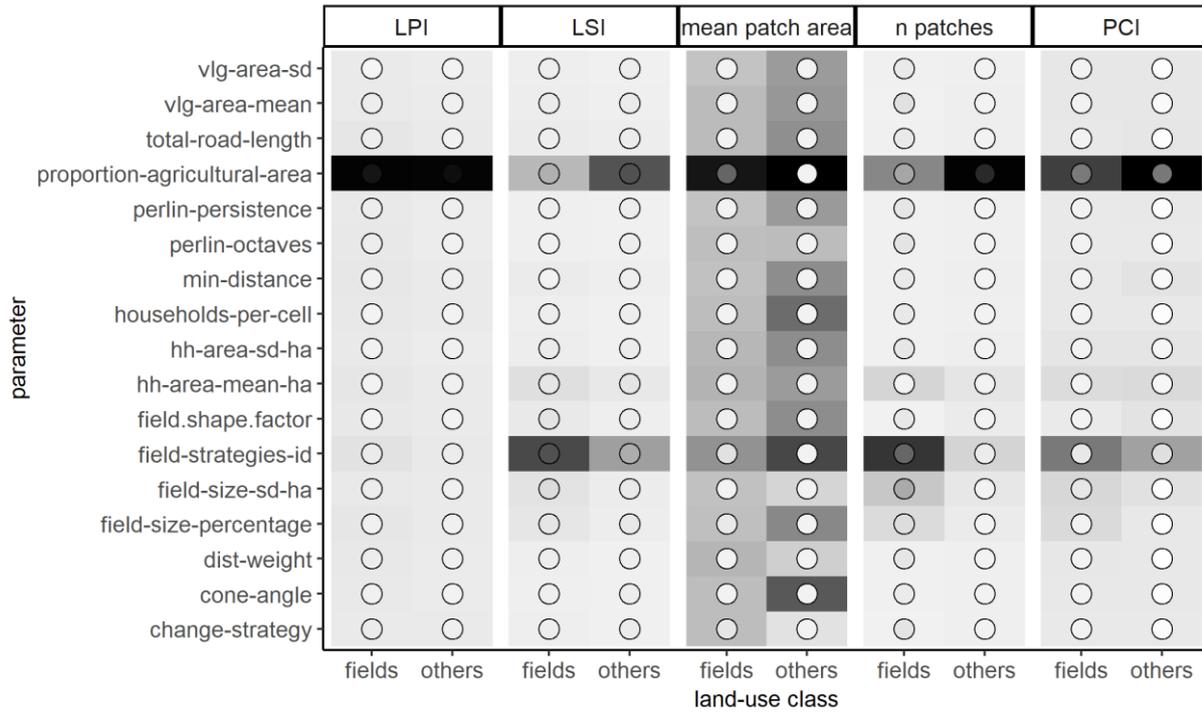## 4.1 Approach 1: Sobol Sensitivity Analysis

Figure 5: Approach 1, sensitivity analysis: Sobol total and main effects of EFForTS-LGraf model parameters on landscape metrics grouped by land-use classes fields and 'others'. Tile color of each parameter output combination indicates the total effect of parameter changes on the output metric. Colors of dots within each tile show the main effect of parameter changes on the output metric. Thus, tiles with dark color and a bright dot have a large total effect but a small main effect indicating strong interaction effects, whereas tiles with dark color and a dark dot indicate strong main effects. For abbreviations and model parameterizations, see section 1.2.1 in Appendix I.

Two parameters, *total-agricultural-area*, defining the resulting proportion of agricultural area in the generated landscapes and *field-strategies-id*, defining the set of field establishment strategies that is used by the households, had significant total (sum of direct and interaction effects) and main effects (direct effects without interaction) on a wide range of landscape metrics (Fig. 5). For some landscape metrics such as LPI or LSI these two parameters showed only main effects (indicated by dark tile and dot shading in Fig. 5). The landscape shape index (LSI) was only affected by the parameter *proportion-agricultural-area*. The mean patch area was the only output landscape metric that was significantly affected by all model parameters. For some output metrics (LSI, n patches, PCI), the *proportion-agricultural-area* had a larger effect on 'others' class indices, whereas the *field-strategies-id* had a larger effect on the fields indices. Interestingly this pattern was reversed for the output metric mean patch area, where the *proportion-agricultural-area* had a larger effect on the mean patch area of field patches and the *field-strategies-id* had a larger effect on the mean patch area of 'others' class patches.
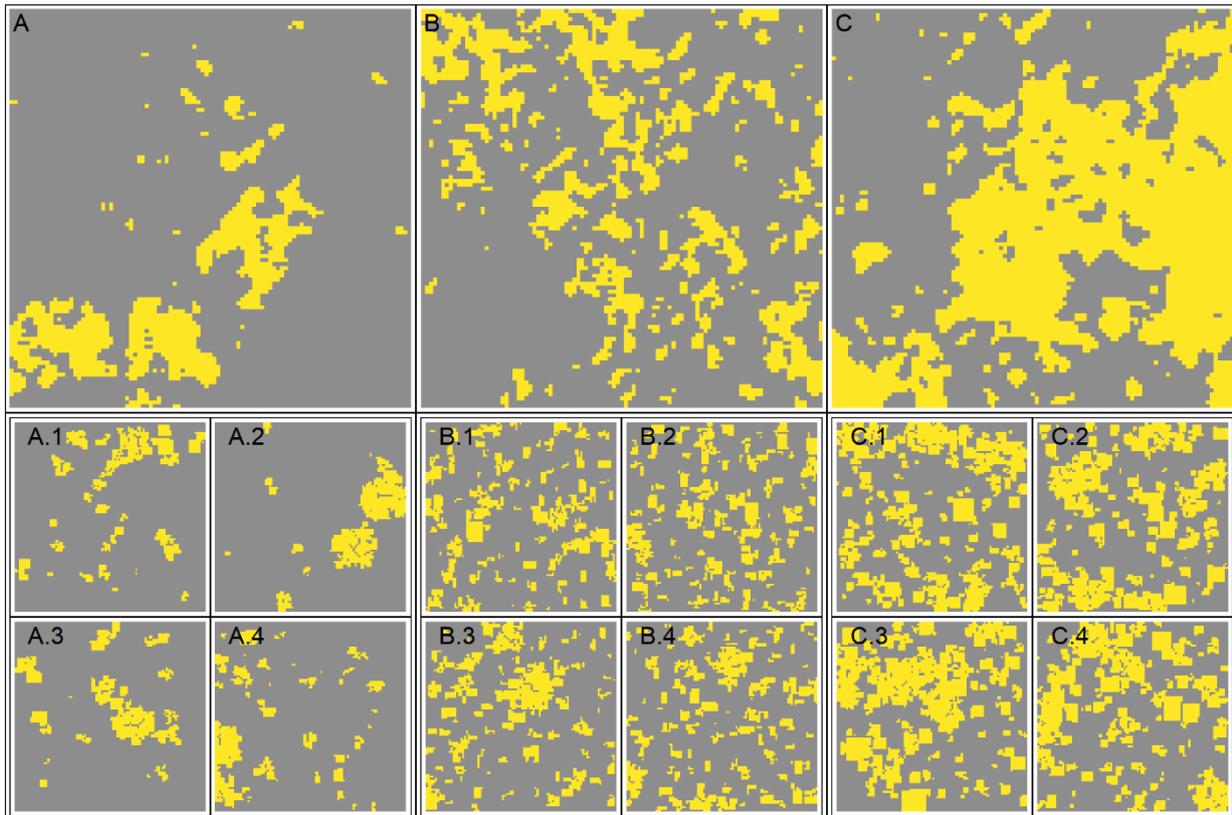
4.2 Approach 2: Validation

Figure 6: Approach 2, validation: A, B and C show sampled maps (100 × 100 cells, 50 m resolution) from the reclassified satellite image of the Harapan region, Jambi province, Indonesia. Yellow cells indicate agricultural area, grey cells indicate land-use class 'others'. We applied genetic algorithm optimization to tweak EFForTS-LGraf model parameters in order to recreate these map samples. This was done by calculating deviances in landscape metrics between the sampled map and the generated map and minimizing this deviance with each generation of the algorithm. We ran the algorithm for each map sample (A, B, C) individually and stored the final parameterization with the lowest deviation. Using these final parameterizations we generated 4 maps for each map sample to account for stochasticity during the map creation process (A.1-A.4, B.1-B.4, C.1-C.4). The generated maps have the same resolution as the map samples (100 × 100 cells, 50 m resolution) but are displayed at 1/4th size.

The re-classified satellite image of the Harapan region, Jambi, shows a large heterogeneity in the distribution of agricultural patches (Fig. 4). Large-scale clustered agricultural areas can be found in the North-East, whereas in the North, West and South, agricultural patches are more scattered. Harapan rainforest conservation area is located in the centre of the image and does not contain any agricultural fields at all.

The three landscapes that were sampled from this map reflect this gradient with agricultural proportions ranging from of 0.13 (sample 1, Fig. 6) over 0.23 (sample 2, Fig. 6) to 0.46 (sample 3, Fig. 6). All three samples also show a large heterogeneity in patch sizes, field sizes and distribution of fields (see Fig. 6). While the genetic algorithm was able to find parameterizations that recreate many properties of the sampled landscapes, not every detail could be matched (see Fig. 7). Especially large patches of agricultural area could not be recreated accurately (yellow ranges for mean.patch.area and n.patches, Fig. 7). However, even the highest deviation (mean.patch.area of agricultural patches) was still below 1.5%. The landscape shape index (LSI), patch cohesion index (PCI) and total area could be matched very accurately, both for agricultural patches and patches of class 'others'.
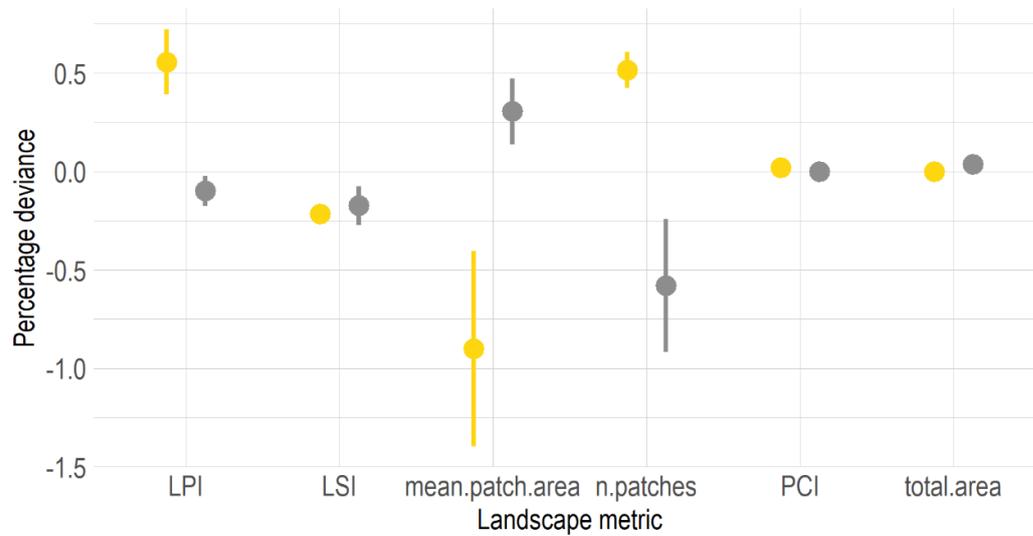
Figure 7: Approach 2, validation: Landscape metrics deviations of generated maps derived by application of a genetic algorithm (see maps A.1-A.4, B.1-B.4, C.1-C.4 in Fig. 6), to landscape metrics of original samples from the reclassified satellite image of the Harapan region, Jambi province (see maps A,B,C in Fig. 6). Yellow dots and line ranges represent landscape metrics of agricultural patches, grey dots and line ranges those of patches of class 'others'.

4.3 Approach 3: Applied case study

In all generated landscapes, fields were distributed mainly along the road network that was used to set up the model (Fig. 8). We observed inter-linking effects of household area and specialization on aggregation of crop types in the landscape by visual comparison of resulting land-use maps (Fig. 8). High specialization on oil palm led to much higher spatial aggregation of crop types when household area was larger, compared to smaller household area (Fig. 8).

Within the boundaries of our assumptions we did not find any significant effects of oil palm specialization (specialization, orange bars in Fig. 9) or interaction effects (size*specialization, blue bars in Fig. 9) on 'others' landscape metrics (see standardized regression coefficients for 'others' patches in Fig. 9). However, the landscape shape index (LSI), the mean patch area, the number of patches and the patch cohesion index (PCI) were significantly affected by the household size distributions (size, black bars in Fig. 9). For increasing household areas we found increased aggregation (negative landscape.shape.index coefficient), smaller patches (negative mean.patch.area coefficient), slightly higher total number of patches (positive n.patches coefficient) and fewer perimeter cells (positive patch.cohesion.index coefficient) of 'others' patches (Fig. 9).

The landscape metrics of agricultural patches of the two crop types where clearly affected by household area, specialization on oil palm and interaction effects of both parameters (see oil palm and rubber bars in Fig. 9). In a system with two mono-cultural crop types, high specialization on one crop type indirectly affects the spatial distribution of the other crop type. Patch aggregation was higher for both crop types with increasing specialization and household area (negative landscape.shape.index coefficients). Mean patch area was higher for high specialization and larger households (positive mean.patch.area coefficients). Accordingly, the number of rubber and oil palm patches strongly decreased with increasing household area and oil

palm specialization (negative n.patches coefficients). There were fewer perimeter cells for high specialization levels and larger household areas (positive patch.cohesion.index coefficients). Landscape shape index and mean patch area showed a considerable amount of interaction effects between both parameters suggesting non-linear relationships, which could also be observed from investigating raw data from the simulations (see Fig. 3 in section 1.2.3 in Appendix I).
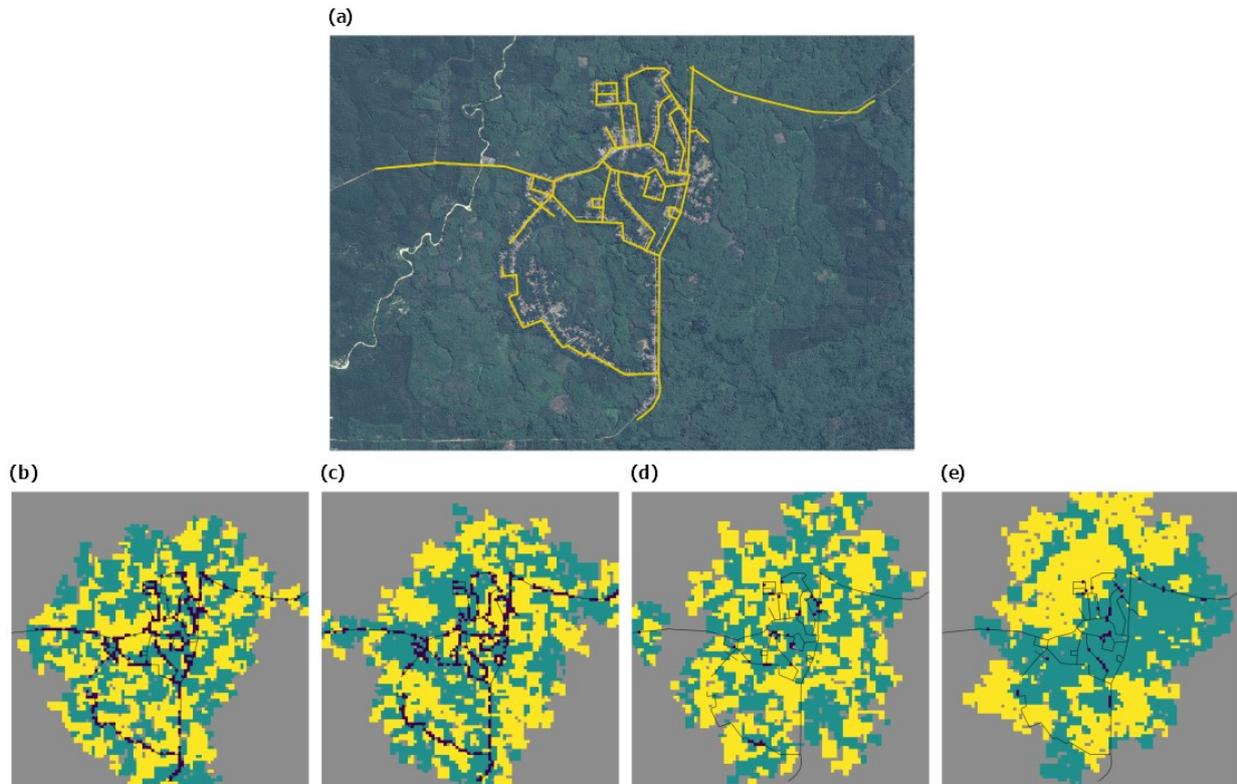
Figure 8: Approach 3, applied case study: (a) Google Satellite imagery, showing the village Lantak Seribu in Renah Pamenang District, Merangin Regency, Jambi. The road network of this village (yellow lines) was selected to generate examples of artificial agricultural smallholder landscape maps with EFForTS-LGraf for different household sizes and specialization levels. (b-e) Examples of artificial land-cover maps. Examples (b) and (c) consist of smaller households that own only some fields whereas households in (d) and (e) are larger and own more fields. In (b) and (d), land uses are distributed to fields completely at random, whereas in (c) and (e), households specialize completely on one land use. In (b - e), green cells indicate oil palm fields, yellow cells indicate rubber fields, grey cells indicate cells of class 'others', purple cells indicate household home-bases and black lines indicate roads.

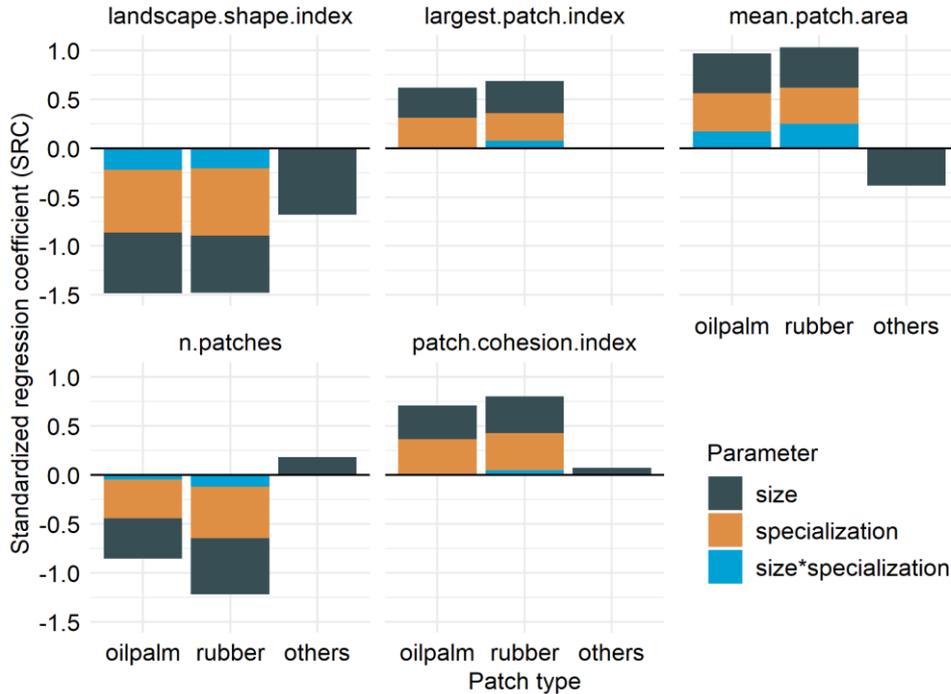Figure 9: Approach 3, applied case study: We generated land-cover maps with varying household area and specialization levels for oil palm and calculated five selected landscape metrics for the two crop types (oil palm and rubber) and patches of class 'others'. The colored bars illustrate the corresponding standardized regression coefficients (SRC) from linear model regressions. Bars display significant importances of household size (size), specialization level (specialization) and the importance of interactions between these two parameters on the selected landscape metrics. Parameter names and values are described in section 1.2.3 in Appendix I. Landscape metrics are described in Section 3, Table 2.

5 Discussion

Although various landscape generator approaches have been developed in the past, only few of them are process-based, have a distinct focus on agricultural land or incorporate any human dimension (Langhammer *et al.*, 2019). The main goal of our study was to fill this gap by developing the landscape generator EFForTS-LGraf. The process-based algorithms of EFForTS-LGraf explicitly reflect the linkage of agricultural expansion and deforestation to road and infrastructure development (Gatto *et al.*, 2015; Kirby *et al.*, 2006; Soares-Filho *et al.*, 2001). This allows not only to generate realistic-looking landscapes, but also links the model to one of the world's central sources of environmental pressures, i.e. road expansion (Ibisch *et al.*, 2016). Another key power of the model is explicit consideration of smallholder farming households and elements that characterize smallholder decisions and socio-economic patterns such as land ownership and farm economy – and how these shape landscapes. In this, the model helps overcoming an important barrier between ecological and socio-economic research, allowing one to explore the behavior of agricultural-natural frontiers as a socio-ecological system (Berkes *et al.*, 1998).

The process-based nature of the model algorithms and parameters allow to formulate and evaluate specific questions through adjustment of the model parameterization. Particularly the final analysis (approach 3) demonstrates how the model can be used to explore how factors that affect smallholder decisions and socio-economic processes shape landscape patterns and, thereby, biodiversity and ecosystem services. Notably, global processes of agricultural

intensification comprise two elements that are captured by the model, namely a trend of specialization toward monocultures and an increase in field sizes and area owned by fewer smallholders (Pe'er *et al.*, 2014), accompanied by an ongoing decline in rural employment (e.g. (Pe'er *et al.*, 2017) for the EU). Our simulation experiment revealed that household consolidation and crop specialization indeed had a large impact on various landscape characteristics. Although these effects were expected to some extent, our analysis showed that the inter-linkages between household-level processes and landscape characteristics were significantly affected by interaction effects of consolidation and specialization. Such identification of interactions and non-linear relationships may be important for a wide range of ecological studies.

We assessed the basic functionality, parameter sensitivity and validity of EFForTS-LGraf by performing a sensitivity analysis and a validation against real-world satellite imagery. The Sobol sensitivity analysis (approach 1) revealed that the proportion of agricultural area (*proportion-agricultural-area*) and the selection of establishment strategies (*field-strategies-id*) were the most important parameters across all landscape metrics. However, the effect of each parameter highly depended on the chosen landscape metric. For example, mean patch area was affected by nearly all model parameters, whereas the largest patch index was nearly exclusively affected by the proportion of agricultural area. We used the *setup-type 'area'* to initialize the households of each model simulation. Therefore, household area parameters by definition had little influence, since they were largely pre-set. This also means that we can expect larger household area parameter effects on landscape metrics when using the *setup-type 'household'*. The three different *setup-types* of EFForTS-LGraf allow for great flexibility in parameterizing the model. Depending on the application, it may be important to generate landscapes with the same proportion of agricultural area, but different household properties (approach 3). When the total number of households in the area is unknown but the typical village sizes and the number of villages in the area can be estimated, the *setup-type 'village'* can be used to approximate the number of households.

Our artificially-generated landscapes showed high capacity to match various landscape metrics when compared to classified satellite images and showed high flexibility to generate a broad range of maps along a gradient of spatial structures (see Fig. 4, 7 and section 1.2.2 in Appendix I). Both are important features of landscape generators. Depending on the approach, it may be important to recreate specific maps trough pattern-based optimization approaches or to generate many different maps along a gradient of specific landscape characteristics.

Besides these technical approaches, EFForTS-LGraf can be applied for a wide range of potential applications. First of all, the generated maps can be used to inform other modelling studies, as has been successfully done with the simulation model EFForTS-ABM (Dislich *et al.*, 2018). We are also planning to apply and validate EFForTS-LGraf to other agricultural regions where high quality remote sensing data are available, such as Central Europe. Additionally, EFForTS-LGraf can be used to perform policy-relevant applications, e.g., testing the future impacts of road expansion, especially in developing regions such as Indonesia and impacts of agricultural policies such as the CAP in the EU.

Currently, EFForTS-LGraf produces maps for one specific point in time. Considering the huge pressures of road expansion on natural habitats, particularly in developing countries including in Indonesia (Laurance *et al.*, 2014), incorporating a temporal component explicitly might be a useful extension to the model. Such an extension would allow to create time series of maps with roads and fields occurring gradually. The assumption that at the beginning of map generation, each cell is equally suitable for field establishment is another limitation of EFForTS-LGraf. A potential model extension could add heterogeneous land-use types to the initial state of the landscape (e.g. forest, grassland, peatland, instead of assigning type 'others' to all cells). By

consideration of differential pressure for land-use change depending on these land-use types, more complex landscape patterns could be created. We decided to develop EFForTS-LGraf with a clear focus on linking geospatial information, such as road polyline shapefiles and land-use fractions, with empirical data, such as household size and field size distributions.

When comparing model outcomes with reclassified land-cover maps from Harapan region (approach 2), we also revealed some systematic differences. Most prominently, the area of agricultural patches was underestimated in the generated landscapes, whereas the area of 'others' patches was overestimated. This mismatch may partly be explained by the homogeneous field establishment of model households. Although households are able to adjust their field establishment strategies if they are not successful, each household uses the same set of strategies and switches after the same number of unsuccessful tries. Adding household-level heterogeneity to field establishment strategies would allow for increased local field aggregation heterogeneity but would also add complexity to the model. In the current model version, crop types can be assigned to fields based on the user-set fractions and specialization levels. Future model extensions may also contain additional algorithms to control spatial clustering of certain crop types.

In conclusion, EFForTS-LGraf combines economic smallholder survey data and spatial information to generate landscapes featuring the characteristics of observed agricultural smallholder landscapes. EFForTS-LGraf is especially useful for applications where agricultural maps need to be provided in conjunction with corresponding economic household data which can not be obtained from remote sensing alone. Due to its flexibility, EFForTS-LGraf can be utilized for a wide range of applications, such as: (1) map generation by providing specific economic case study data, (2) application of pattern-matching approaches to match generated maps with realistic land-use maps, and (3) generating maps along gradients of properties on household or landscape-level. EFForTS-LGraf contributes to the set of already published landscape generators and fills an important gap through its application of process-based algorithms with a distinct focus on road expansion, agricultural land and explicit consideration of human dimensions of land-use change.

6 Acknowledgements

7 Funding

8 References

Ayllón, D., Railsback, S.F., Vincenzi, S., Groeneveld, J. & Almodóvar, A. (2016) InSTREAM-Gen: Modelling eco-evolutionary dynamics of trout populations under anthropogenic environmental change. *Ecological Modelling*, **326**, 36–53.

Berkes, F., Folke, C. & Colding, J. (1998) *Linking social and ecological systems : management practices and social mechanisms for building resilience*. Cambridge University Press.

Chan, K., Tarantola, S., Saltelli, A. & Sobol, I.M. (2000) Variance-Based Methods. A. Saltelli, K. Chan & E. Scott, eds., *Sensitivity Analysis*, pp. 167–197. Wiley, Chichester.

Dislich, C., Hettig, E., Salecker, J., Heinonen, J., Lay, J., Meyer, K.M., Wiegand, K. & Tarigan, S. (2018) Land-use change in oil palm dominated tropical landscapes—An agent-based model to explore ecological and socio-economic trade-offs. *PLOS ONE*, **13**, e0190506.

Dislich, C., Keyel, A.C., Salecker, J., Kisel, Y., Meyer, K.M., Auliya, M., Barnes, A.D., Corre, M.D., Darras, K., Faust, H., Hess, B., Klasen, S., Knohl, A., Kreft, H., Meijide, A., Nurdiansyah, F., Otten, F., Pe'er, G., Steinebach, S., Tarigan, S., Tölle, M.H., Tscharntke, T. & Wiegand, K. (2017) A review of the ecosystem functions in oil palm plantations, using forests as a reference system. *Biological Reviews*, **92**, 1539–1569.

Euler, M., Krishna, V., Schwarze, S., Siregar, H. & Qaim, M. (2017) Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia. *World Development*, **93**, 219–235.

Euler, M., Schwarze, S., Siregar, H. & Qaim, M. (2016) Oil Palm Expansion among Smallholder Farmers in Sumatra, Indonesia. *Journal of Agricultural Economics*, **67**, 658–676.

Gardner, R.H., Milne, B.T., Turnei, M.G. & O'Neill, R.V. (1987) Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology*, **1**, 19–28.

Gatto, M., Wollni, M. & Qaim, M. (2015) Oil palm boom and land-use dynamics in Indonesia: The role of policies and socioeconomic factors. *Land Use Policy*, **46**, 292–303.

Gaucherel, C., Giboire, N., Viaud, V., Houet, T., Baudry, J. & Burel, F. (2006) A domain-specific language for patchy landscape modelling: The Brittany agricultural mosaic as a case study. *Ecological Modelling*, **194**, 233–243.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., GossCustard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jorgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabo, R., Visser, U. & DeAngelis, D. (2006) A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, **198**, 115–126.

Grimm, V. & Berger, U. (2016) Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. *Ecological Modelling*, **326**, 177–187.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J. & Railsback, S.F. (2010) The ODD protocol: A review and first update. *Ecological Modelling*, **221**, 2760–2768.

Ibisch, P.L., Hoffmann, M.T., Kreft, S., Pe'er, G., Kati, V., Biber-Freudenberger, L., DellaSala, D.A., Vale, M.M., Hobson, P.R. & Selva, N. (2016) A global map of roadless areas and their conservation status. *Science (New York, NY)*, **354**, 1423–1427.

Jansen, M.J. (1999) Analysis of variance designs for model output. *Computer Physics Communications*, **117**, 35–43.

Kirby, K.R., Laurance, W.F., Albernaz, A.K., Schroth, G., Fearnside, P.M., Bergen, S., Venticinque, E.M. & da Costa, C. (2006) The future of deforestation in the Brazilian Amazon. *Futures*, **38**, 432–453.

Klasen, S., Meyer, K.M., Dislich, C., Euler, M., Faust, H., Gatto, M., Hettig, E., Melati, D.N., Jaya, I.N.S., Otten, F., Pérez-Cruzado, C., Steinebach, S., Tarigan, S. & Wiegand, K. (2016) Economic and ecological trade-offs of agricultural specialization at different spatial scales.

Koh, L.P. (2011) Balancing societies' priorities: An ecologist's perspective on sustainable development. *Basic and Applied Ecology*, **12**, 389–393.

Krishna, V., Euler, M., Siregar, H. & Qaim, M. (2017) Differential livelihood impacts of oil palm expansion in Indonesia. *Agricultural Economics*, **48**, 639– 653.

Kumar, M., Husian, M., Upreti, N. & Gupta, D. (2010) Genetic Algorithm: Review and Application. *International Journal of Information Technology and Knowledge Management*, **2**, 451–454.

Langhammer, M., Thober, J., Lange, M., Frank, K. & Grimm, V. (2019) Agricultural landscape generators for simulation models: A review of existing solutions and an outline of future directions. *Ecological Modelling*, **393**, 135–151.

Laurance, W.F., Goosem, M. & Laurance, S.G. (2009) Impacts of roads and linear clearings on tropical forests. *Trends in Ecology & Evolution*, **24**, 659–669.

Laurance, W.F., Sayer, J. & Cassman, K.G. (2014) Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, **29**, 107–116.

McGarigal, K., Cushman, S. & Ene, E. (2012) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps.

McKay, M.D., Beckman, R.J. & Conover, W.J. (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, **21**, 239.

Melati, D. (2017) *The use of remote sensing data to monitor land use systems and forest variables of the tropical rainforest landscape under transformation in Jambi Province, Sumatra, Indonesia*. Dissertation, Georg-August-Universität Göttingen.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H. & Schwarz, N. (2013) Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, **48**, 37–48.

Pe'er, G., Dicks, L.V., Visconti, P., Arlettaz, R., Baldi, A., Benton, T.G., Collins, S., Dieterich, M., Gregory, R.D., Hartig, F., Henle, K., Hobson, P.R., Kleijn, D., Neumann, R.K., Robijns, T., Schmidt, J., Shwartz, A., Sutherland, W.J., Turbe, A., Wulf, F. & Scott, A.V. (2014) EU agricultural reform fails on biodiversity. *Science*, **344**, 1090–1092.

Pe'er, G., Lakner, S., Passoni, G., Azam, C., Berger, J., Hartmann, L., Schueler, S., Müller, R., von Meyer-Höfer, M. & Zinngrebe, Y. (2017) Is the CAP Fit for purpose? An evidence-based, rapid fitness-check assessment. Technical report, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig.

Pe'er, G., Zurita, G.A., Schober, L., Bellocq, M.I., Strer, M., Müller, M. & Pütz, S. (2013) Simple Process-Based Simulators for Generating Spatial Patterns of Habitat Loss and Fragmentation: A Review and Introduction to the G-RaFFe Model. *PLoS ONE*, **8**, e64968.

R Core Team (2018) R: A Language and Environment for Statistical Computing.

Salecker, J. & Sciaini, M. (2018) nlrx: Setup, Run and Analyze NetLogo Model Simulations from R via XML.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. & Tarantola, S. (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, **181**, 259– 270.

Schröder, B. & Seppelt, R. (2006) Analysis of pattern–process interactions based on landscape models—Overview, general concepts, and methodological issues. *Ecological Modelling*, **199**, 505–516.

Soares-Filho, B.S., Assunção, R.M. & Pantuzzo, A.E. (2001) Modeling the Spatial Transition Probabilities of Landscape Dynamics in an Amazonian Colonization FrontierTransition probability maps indicate where changes may occur in the landscape, thus enabling better evaluation of the ecological consequences of lands. *BioScience*, **51**, 1059–1067.

Sobol, I.M. (1990) On sensitivity estimation for nonlinear mathematical models. *Matem Mod*, **2**, 112–118.

Tietjen, B. (2016) Same rainfall amount different vegetation—How environmental conditions and their interactions influence savanna dynamics. *Ecological Modelling*, **326**, 13–22.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L. & Storlie, C. (2014) SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises.

Willighagen, E. & Ballings, M. (2015) genalg: R based Genetic Algorithm.

With, K.A. (1997) The Application of Neutral Landscape Models in Conservation Biology. Aplicacion de Modelos de Paisaje Neutros en la Biologia de la Conservacion. *Conservation Biology*, **11**, 1069–1080.

9 Appendix

9.1 Model Details

9.1.1 Details on road creation and household placement

EFForTS-LGraf offers three options for road creation: (1) roads can be read in from an existing road map (option: real.shapefile); (2) a road network is artificially created based on a random elevation model (option: artificial.perlin); (3) a road network is created based on the straight road creation algorithm of the G-Raffe landscape generator (option: artificial.graffe).

To upload an existing **real road map**, one needs to store the respective shapefile of the roads, the projection file and a second shapefile that only includes the extent of the map. After loading the shapefile, the world is re-sized if necessary. All cells that have a road intersecting are identified as road cells.

The **Perlin** algorithm mimics situations in which roads are created depending on elevation gradients. First, a simple perlin noise elevation model is created by adding several random noise grids with decreasing weight (perlin parameters $p1_{r,perl}$, $p2_{r,perl}$, $p3_{r,perl}$) [1,2]. The first road is then created by connecting two random locations in the landscape. First, a road-building-agent is created on one of these two random locations. For each neighboring cell, the road-building-agent calculates the distance to the destination cell (second random location) and the elevation difference to the agents' current location. The total score of each cell is then determined by weighting the distance and elevation criterion depending on the continuous parameter $p4_{r,perl}$ ($\in[0,1]$). The road-building-agent then moves to the cell with the highest score, establishes a road-cell there and continues to move and create road-cells until it reaches the final destination. Thus, by adjusting the parameter $p4_{r,perl}$, the algorithm gives a higher weight to the distance (values near 1) of the connection allowing for straight roads or to the elevation gradient (values near 0) of the connection allowing for wiggling roads. If the $n_{r,art}$ is not reached yet, another road is created by selecting one random point on an already established road and another location in the landscape that is at least $m_{r,art}$ cells away from any other road.

Under the **G-RaFFe** algorithm all roads are straight and start along the left or lower edge of the simulation area and are directed either vertically, horizontally or diagonally within the landscape. Road length is drawn randomly within the interval $[1, \sqrt{w_s^2 + h_s^2}]$. The parameter $m_{r,art}$ determines the minimum parallel distance in cells between two roads. Roads are created until the number of road cells reaches $n_{r,art}$. This road creation option is in most parts adopted from the road creation algorithm of the G-RaFFe model [3].

9.1.2 Details on field establishment

During the EFForTS-LGraf field establishment procedure, households attempt to establish a field in three steps: first deciding on the field size, second moving to a potential location and third making sure there is enough space to establish a field of the desired size in this location.

Step one: The size f of the field is drawn randomly from the field size frequency distribution. Step two: Move to an 'others' start cell using one of four search strategies, which can change during the model run:

1. Random walk to an 'others' cell starting from the home base (*s1.homebase*)
2. Random walk to an 'others' cell starting from one of the fields of this household (*s2.field*)
3. Determine an 'others' cell within a defined radius from the home base (*s3.nearby*)
4. Determine the closest 'others' cell which is surrounded only by 'others' cells (*s4.avoid*)

The selection and order of these search strategies can be defined on the model interface. At the start cell, the household turns in a random compass direction (north, east, south or west) and starts with the field establishment procedure by trying to establish a first row of cells that will belong to the field. The household agent moves forward cell by cell until it either meets a cell which is already a field or a home base or an inaccessible area cell. To ensure reasonable field shapes, the minimum length l of this first row needs to be the side length of the smallest square that fits into the field size. The maximum length of this first row is then set to be the side length of the next larger square that fits into the field size. Because this algorithm restricts field sizes to rather quadratic shapes, the maximum length can be further modified by the parameter s_f which allows to proportionally increase the length of the first row. This results in narrower field shapes.

If the establishment of the first row is not successful, the other possible compass directions for the first row are tested. Once a first field row is established, the farmer tries to expand the field to the left or right of this row until the field reaches its predefined size f. If expansion to one side of the first row did not result in the designated field size, the field is extended also to the other side. If also this expansion does not yield the designated field size, all so far added field patches are removed, the household moves back to the start cell and starts over again with the first field row in a different direction. If all possible four compass directions do not lead to a successful field establishment, the farmer selects a new random start cell and tries again to establish a field of the designated field size f. The parameter n_{strat} defines in how many locations a household tries to establish fields before it switches the search strategy to the next strategy. If the household failed to establish fields for all selected strategies, the model will stop and report a warning message that field establishment was not successful. In such cases, the resulting landscape is incomplete because not all households reached their final size. This might especially happen under certain model parameterizations where households and field sizes tend to be very large without adjusting the total dimensions of the landscape.

9.1.3 Details on EFForTS-LGraf spatial output

The spatial information of generated landscapes can be exported as *.asc raster files in order to allow exchange to other model applications. EFForTSLGraf contains export functions for the following raster files:

- Road-raster: 0 for non-road cells, 1 for cells that have a road intersecting
- Homebase-raster: -1 for non-home base cells, household identity number (*p_homebase_id*) for home base cells. Multiple raster files are written if $n_{s,c} > 1$.
- Ownership-raster: -1 for cells that are not owned, household identity number (*p_owner*) for cells that are owned by a household. Note: inaccessible areas do not show up in this output.
- Land-use-type-raster: 0 for 'others' and road cells (i.e. also home bases), 1, 2, 3, 4, 5 for the different crop types, 1000 for inaccessible areas
- Patch-id-raster: -1 for cells that are not smallholder fields (also inaccessible areas), patch identity number (*p_id*) for smallholder fields
- Matrix-patch-raster: -1 for cells that have a road intersecting and that are not 'others' cells, 'others' patch identity number (*p_matrixcluster_id*) for 'others' cells.

Besides raster output, the write-road-shapefiles procedure creates a polyline shapefile from the current road network, which is useful if one of the artificial road creation algorithms has been used. Additionally, EFForTS-LGraf offers an output function that utilizes the 3D renderer of NetLogo, to create a three dimensional illustration of the generated landscape using trees, palms

and houses to visualize the different crop types (Fig. 1). With little code adjustments, this feature could in general be used to display other crop types as well.
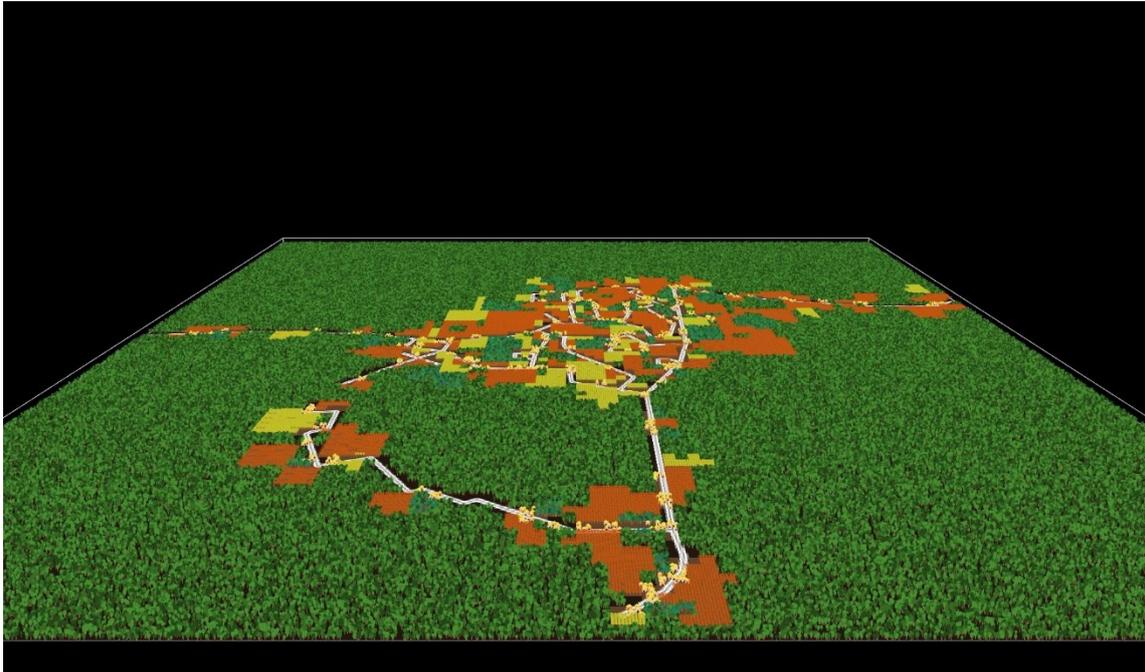

Figure 1: 3d rendered snapshot of EFForTS-LGraf. White lines represent the real road polylines derived from GIS data. 'Others' cells are illustrated as green shaded tree shapes. Oil palm plantations are illustrated as orange colored palm rows. Rubber plantations are illustrated as yellow colored tree rows. Agroforestry cells are illustrated as mixed rows and trees in blue and green color. Settlements are illustrated by small house symbols.

9.2 Analysis Details

9.2.1   Approach 1: Sobol sensitivity analysis

We created a parameter input matrix in order to calculate Sobol Sensitivity Indices for model parameters on different outputs (here: landscape metrics). The parameter matrix was created by variation of parameters in defined ranges (Table 1). The parameter matrix was created with 500 samples and 10 bootstrapping replicates which results in 9500 simulations. For calculating sensitivity indices we used the optimized Sobol estimator proposed by Jansen [4–7].

Table 1: Sobol sensitivity analysis parameter ranges

Id	Name on GUI	variation	value
t_s	setup-type	constant	"area"
$n_{s,h}$	number-of-farmers	not in use	-
$n_{s,v}$	number-of-villages	not in use	-
$n_{s,a}$	prop-agricultural-area	uniform	min=0.05, max=0.75
$n_{s,c}$	households-per-cell	uniform (integer)	min=3, max=6
$seed_s$	rnd-seed	set externally	-
rep_s	reproducable?	constant	true
w_s	width	constant	100
h_s	height	constant	100
c_s	cell-length-meter	constant	50
t_r	road-type	constant	"artificial.perlin"
$i_{r,shp}$	road-map-nr	not in use	-
$n_{r,art}$	total-road-length	uniform	min=500, max=1000
$m_{r,art}$	min-dist-roads	constant	5
$p1_{r,perl}$	perlin-octaves	uniform	min=2, max=12
$p2_{r,perl}$	perlin-persistence	uniform	min=0.1, max=0.9
$p3_{r,perl}$	cone-angle	uniform	min=90, max=180
$p4_{r,perl}$	dist-weight	uniform	min=0, max=1
d_v	vlg-size-distribution	constant	"uniform"
μ_v	vlg-size-mean_ha	uniform	min=10, max=20
σ_v	vlg-size-sd_ha	uniform	min=2, max=10
d_v	vlg-min-distance	uniform	min=1, max=10
d_h	hh-area-distribution	constant	"log-normal"
μ_h	hh-area-mean_ha	uniform	min=0.25, max=3
σ_h	hh-area-sd_ha	uniform	min=0.25, max=1
n_i	inaccessible-area-fraction	constant	0
l_i	inaccessible-area-location	not in use	-
d_i	inaccessible-area-distribution	not in use	-
μ_i	inaccessible-area-mean	not in use	-
σ_i	inaccessible-area-sd	not in use	-
t_f	field-type	constant	"percentage"
d_f	field-size-distribution	constant	"normal"
μ_f	field_size_mean_ha	not in use	derived as percentage from hh-area-mean_ha
σ_f	field_size_sd_ha	uniform	min=0.25, max=2
p_f	field-size-percentage	uniform	min=0.1, max=0.9
s_f	field-shape-factor	uniform	min=0.7, max=2
t_{strat}	strategies-type	constant	"id"
$s1_{strat}$	s1.homebase	not in use	derived from field-strategies-id

$s2_{strat}$	s2.fields	not in use	derived from field-strategies-id
$s3_{strat}$	s3.nearby	not in use	derived from field-strategies-id
$s4_{strat}$	s4.avoid	not in use	derived from field-strategies-id
n_{strat}	change-strategy	uniform	min=5, max=20
i_{strat}	field-strategies-id	uniform (integer)	min=1, max=9
t_l	land-use-assignment	constant	"landscape-level-fraction"
id_l	LUT-l-name (l=1,2,3,4,5)	constant	l=1: "fields"; l>1 not in use
fr_l	LUT-l-fraction (l=1,2,3,4,5)	constant	l=1: 1; l>1 not in use
$fr_{l,spec}$	LUT-l-specialize (l=1,2,3,4,5)	not in use	-
f_l	LUT-fill-up	constant	"LUT-1-fraction"

9.2.2 Approach 2: Validation

To study the range of landscape characteristics that EFForTS-LGraf can create, we performed a second validation approach in addition to the genetic algorithm validation approach (see Approach 2, main text). From the reclassified land-cover map, that was used for the genetic algorithm validation approach, we sampled 100 randomly placed landscapes, 100×100 cells in size (landscapes may overlap), and calculated the five landscape metrics for each of these sampled landscapes and each land-cover type (fields and other). We compared the landscape characteristics (landscape metrics) of these landscapes to the 9500 artificially generated landscapes from the Sobol sensitivity analysis (approach 1). To allow comparability, all agricultural crop types within the generated landscapes were aggregated to one class (*fields*) and the remaining area was classified as class 'others'.

Large heterogeneity was found in landscape metrics from 100 randomly placed sample landscapes (100 × 100 cells) within the classified land-cover map (point distributions in Fig. 2). Landscape metrics were calculated as a function of 'others' cells area, because most indices are very sensitive to class proportions. The landscape shape index (LSI) quantifies the spatial aggregation of patches for each land-cover type. Compared to our generated landscapes, the sampled landscapes covered a relatively small range of aggregation levels (Fig. 2). For landscapes with small 'others' area, our generated landscapes had higher LSI than the sampled landscapes. The largest patch index was almost linearly correlated with the 'others' area (negative for agricultural patches, positive for 'others' patches). With increasing 'others' area, the mean patch area was exponentially increasing for 'others' patches in both, the generated and the sampled landscapes. For fields, the mean patch area was almost constant at a very low level. Compared to our generated landscapes, the number of 'others' patches was relatively low, especially for landscapes with small 'others' area. The patch cohesion index (PCI) quantifies the shape complexity and perimeter density of patches and showed a decreasing trend for agricultural patches with increasing 'others' area. For 'others' patches, the patch cohesion index was almost constant across the 100 sampled landscapes.

Comparing the 9500 landscapes that were generated within the Sobol sensitivity analysis approach (approach 1) with the 100 realistic landscapes, many characteristics found in the sampled landscapes (point distributions in Fig. 2) were covered by our generated landscapes

(background shading in Fig. 2). The generated landscapes covered a wide range of landscape shape index (LSI) values. However, generated landscapes with small 'others' area showed systematically higher LSI values. The largest patch index (LPI) distribution was matched very accurately and all sampled landscapes lay within the boundaries of the 9500 generated landscapes. EFForTS-LGraf was also able to reproduce a wide range of mean patch area although the observed heterogeneity in the sampled landscapes was significantly lower. The variety of the total number of patches was reproduced for field patches, but did not fit well for 'others' patches, especially when 'others' area was small. For the patch cohesion index all sampled landscapes lay within the boundaries of our generated landscapes.
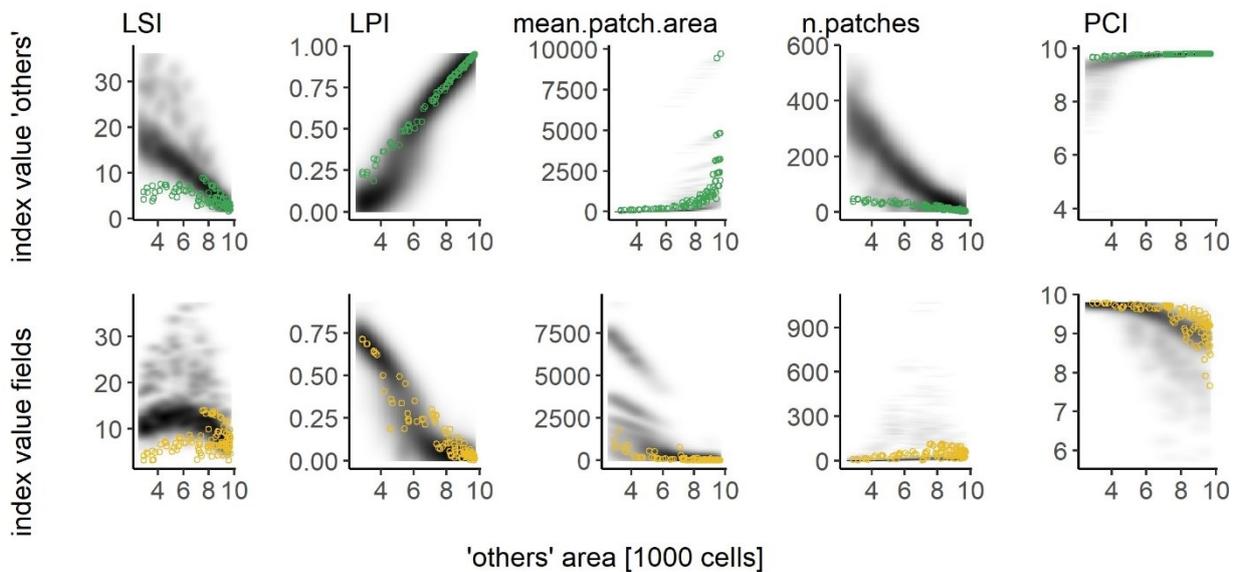

Figure 2: Points indicate distributions of landscape metrics from 100 landscape samples (100 × 100 *cells*) from the re-classified satellite image of the Harapan region. Background shading indicates distributions of landscape metrics from 9500 generated landscapes of the same size and resolution (sobol sensitivity analysis, approach 1). The darker the shading, the more often a landscape with the corresponding index value has been generated. Index distributions of 'others' cells and patches are presented in the upper row (green points, 'others'), whereas index distributions of field cells and patches are presented in the bottom row (yellow points, fields).

In order to visualize the similarity between sampled and generated landcover maps, we selected three different example landscapes with different 'others' cells area from the 100 samples of the Harapan land-cover map and three example landscapes from the 9500 generated artificial landscapes with matching 'others' area (Fig. 3). By visual comparison, the resulting landscapes from EFForTS-LGraf showed similar spatial clustering of patches and patch sizes although the landscapes from the Harapan land-cover map contained slightly more small field patches that were slightly more scattered across the 'others' area. This matched the observation from the landscape metrics comparison, where the sampled landscapes from the classified landuse map had a significantly greater number of field patches compared to our generated landscapes (Fig. 3, n patches).
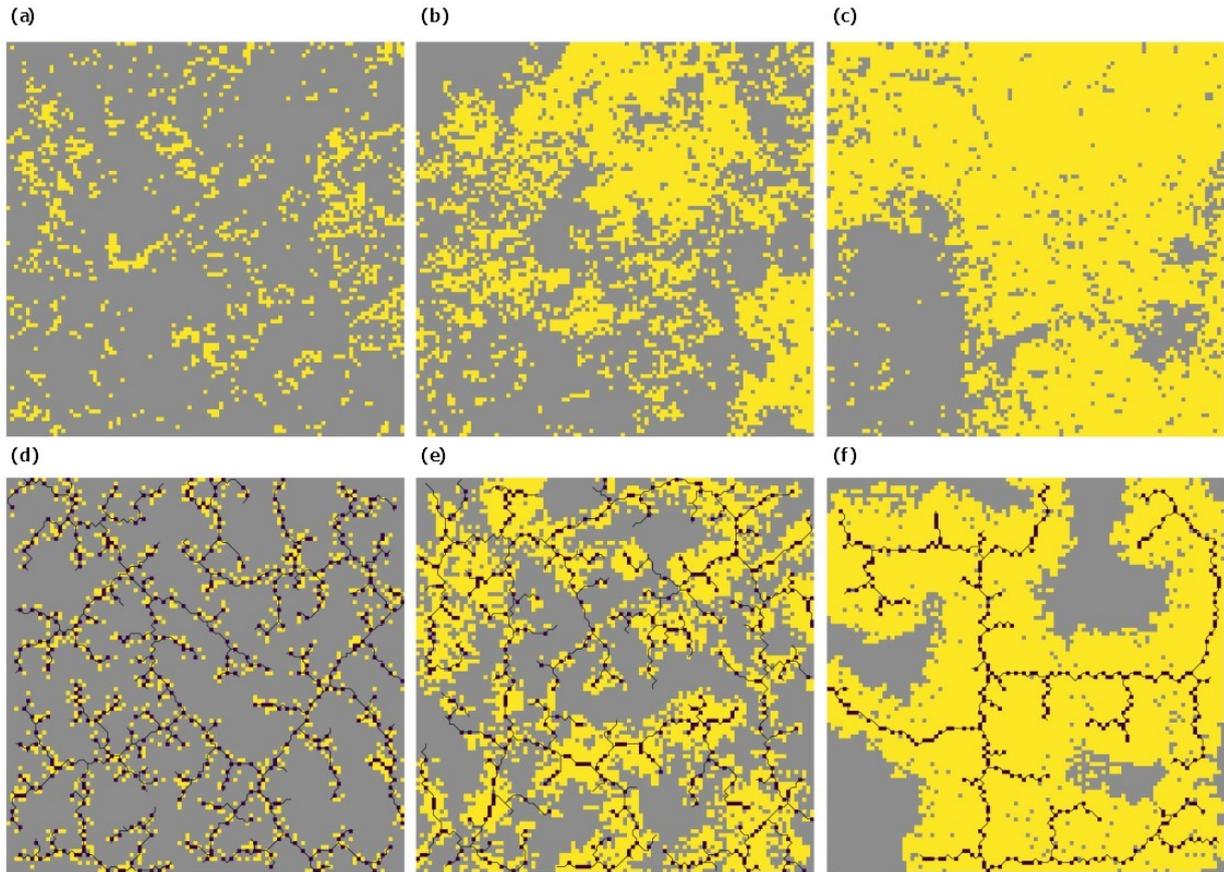
Figure 3: Three of the 100 sampled landscapes from the re-classified satellite image (a, b, c) and three generated EFForTS-LGraf landscapes from the Sobol sensitivity analysis (d, e, f). Grey cells indicate 'others' cells and yellow cells indicate fields. For the EFForTS-LGraf landscapes, black cells indicate household home bases and lines indicate artificially generated roads.

9.2.3 Approach 3: Applied case study

We created a parameter input matrix in order to calculate effects of household consolidation and household level specialization on different outputs (here: landscape metrics). The parameter matrix was created by variation of the two parameters *household-area-mean_ha* and *LUT-1-specialize* (Table 2). We generated the parameter matrix by using a Latin Hypercube sampling design with 500 samples [8]. We simulated three repetitions of the parameter matrix to account for stochastic effects. We analyzed the resulting landscapes via the five selected landscape metrics from previous approaches. We calculated linear regression models for each landscape metric and crop type combination and calculated standardized regression coefficients to estimate parameter and interaction effects on landscape metrics of the generated landscapes (coefficient results, see main text). Additionally, we investigated the raw data by plotting landscape metrics as a function of specialization level, grouped by household size (see Fig. 4).

Table 2: Applied case study parameter ranges

id	Name on GUI	variation	value
t_s	setup-type	constant	"area"
$n_{s,h}$	number-of-farmers	not in use	-
$n_{s,v}$	number-of-villages	not in use	-
$n_{s,a}$	prop-agricultural-area	constant	0.5
$n_{s,c}$	households-per-cell	constant	1
$seed_s$	rnd-seed	set externally	-
rep_s	reproducable?	constant	true
w_s	width	constant	100
h_s	height	constant	100
c_s	cell-length-meter	constant	50
t_r	road-type	constant	"real.shapefile"
$i_{r,shp}$	road-map-nr	constant	3
$n_{r,art}$	total-road-length	not in use	-
$m_{r,art}$	min-dist-roads	not in use	-
$p1_{r,perl}$	perlin-octaves	not in use	-
$p2_{r,perl}$	perlin-persistence	not in use	-
$p3_{r,perl}$	cone-angle	not in use	-
$p4_{r,perl}$	dist-weight	not in use	-
d_v	vlg-size-distribution	constant	"uniform"
μ_v	vlg-size-mean_ha	constant	15
σ_v	vlg-size-sd_ha	constant	6
d_v	vlg-min-distance	constant	10
d_h	hh-area-distribution	constant	"log-normal"
μ_h	hh-area-mean_ha	uniform	min=1, max=3
σ_h	hh-area-sd_ha	constant	0.91
n_i	inaccessible-area-fraction	constant	0
l_i	inaccessible-area-location	not in use	-
d_i	inaccessible-area-distribution	not in use	-
μ_i	inaccessible-area-mean	not in use	-
σ_i	inaccessible-area-sd	not in use	-
t_f	field-type	constant	"distribution"
d_f	field-size-distribution	constant	"log-normal"
μ_f	field_size_mean_ha	constant	0.49
σ_f	field_size_sd_ha	constant	0.77
p_f	field-size-percentage	not in use	-
s_f	field-shape-factor	constant	1
t_{strat}	strategies-type	constant	"id"
$s1_{strat}$	s1.homebase	not in use	derived from field-strategies-id
$s2_{strat}$	s2.fields	not in use	derived from field-strategies-id
$s3_{strat}$	s3.nearby	not in use	derived from field-strategies-id
$s4_{strat}$	s4.avoid	not in use	derived from field-strategies-id
n_{strat}	change-strategy	constant	10

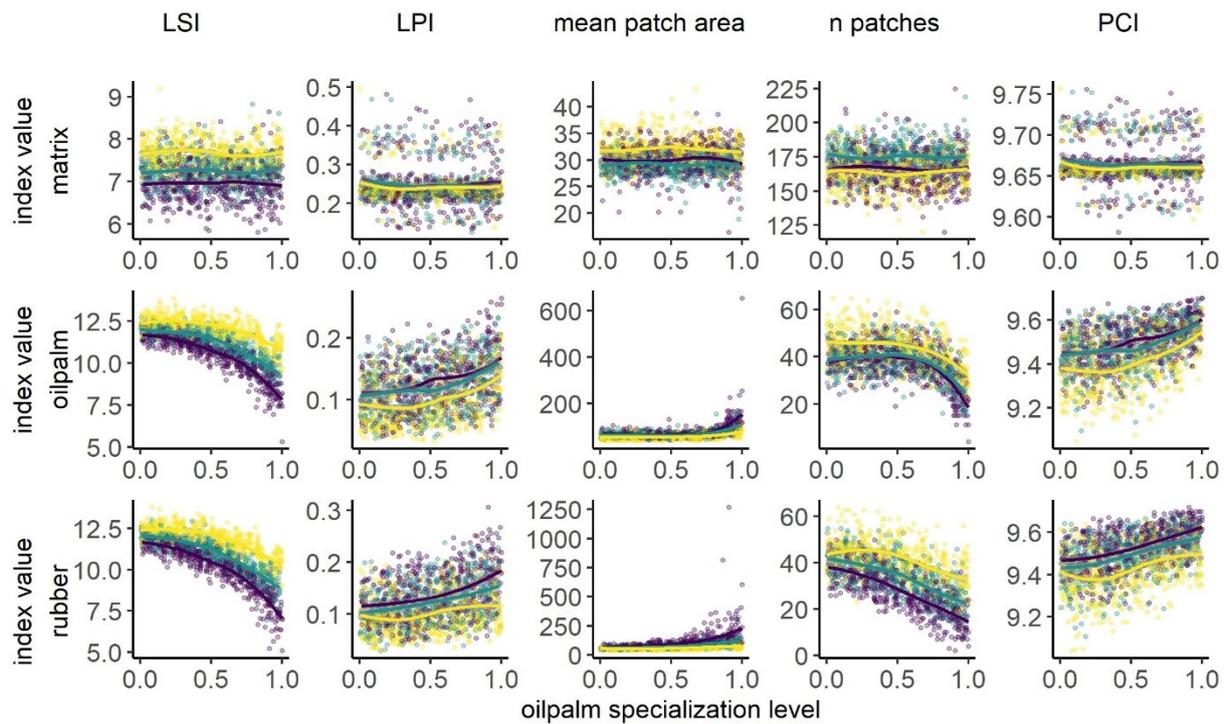| i_{strat}	field-strategies-id	constant	1
t_l	land-use-assignment	constant	"household-levelspecialization"
id_l	LUT-l-name (l=1,2,3,4,5)	constant	l=1: "oilpalm"; l=2: "rubber"; l>2 not in use
fr_l	LUT-l-fraction (l=1,2,3,4,5)	constant	l=1: 0.5; l=2: 0.5; l>2 not in use
$fr_{l,spec}$	LUT-l-specialize (l=1,2,3,4,5)	uniform/constant	l=1: min=0, max=1; l=2: 0, l>2 not in use
f_l	LUT-fill-up	constant	"LUT-2-fraction"

Figure 4: Five selected landscape metrics, for two crop types (oil palm and rubber) and 'others' patches (matrix) calculated for land-cover maps generated by the model using a latin hypercube sampling approach. Plots show the value of the landscape metrics versus household specialization level on oil palm. Differences in household size distributions are shown by colours with darker colors indicating larger household size. These are grouped into small (1 - 1.65 ha, yellow), medium (1.66 - 2.32 ha, green) and large (2.32 - 3 ha, purple). Lines show smoothed trends and standard error using the locally weighted scatterplot smoothing (LOESS) method.

10. References (Apendix)

1. Perlin K, Ken, Perlin, Ken. An image synthesizer. In: Proceedingsof the 12th annual conference on Computer graphics and interactive techniques - SIGGRAPH '85. vol. 19. New York, New York, USA: ACM Press; 1985. p. 287–296. Available from: http://portal.acm. org/citation.cfm?doid=325334.325247.

2. Ebert DS, Musgrave FK, Peachey D, Perlin K, Worley S, Mark WR, et al. Texturing & modeling : a procedural approach. Morgan Kaufmann; 2003. Available from: https://nyuscholars.nyu.edu/en/publications/ texturing-and-modeling-a-procedural-approach-third-edition.

3. Pe'er G, Zurita GA, Schober L, Bellocq MI, Strer M, Müller M,et al. Simple Process-Based Simulators for Generating Spatial Patterns of Habitat Loss and Fragmentation: A Review and Introduction to the G-RaFFe Model. PLoS ONE. 2013;8(5):e64968. doi:10.1371/journal.pone.0064968.

4. Sobol IM. On sensitivity estimation for nonlinear mathematical models. Matem Mod. 1990;2(1):112–118. doi:10.1029/94JC00847.

5. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S.Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications. 2010;181(2):259–270. doi:10.1016/J.CPC.2009.09.018.

6. Jansen MJW. Analysis of variance designs for model output. ComputerPhysics Communications. 1999;117(1-2):35–43. doi:10.1016/S00104655(98)00154-4.

7. Chan K, Tarantola S, Saltelli A, Sobol IM. Variance-Based Methods.In: Saltelli A, Chan K, Scott E, editors. Sensitivity Analysis. Chichester: Wiley; 2000. p. 167–197.

8. McKay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. 1979;21(2):239. doi:10.2307/1268522.