

**Georg-August-Universität Göttingen**

Faculty of Forest Sciences and Forest Ecology

**Modelling avian  $\alpha$ -diversity in the land use transformation  
systems of Indonesia**

Modellierung der  $\alpha$ -Diversität von Vögeln in den Landnutzungs-  
Transformationssystemen Indonesiens

Mats Mahnken

Master Thesis

Göttingen, August 2018

**Georg-August-Universität Göttingen**

**Faculty of Forest Sciences and Forest Ecology**

**Büsgen-Institute, Department of Ecosystem Modelling**

Modelling avian  $\alpha$ -diversity in the land use transformation systems of Indonesia

Modellierung der  $\alpha$ -Diversität von Vögeln in den Landnutzungs-Transformationssystemen Indonesiens

**Master Thesis**

submitted to

**Prof. Dr. Kerstin Wiegand**

Department of Ecosystem Modelling

Büsgen-Institute

Faculty of Forest Sciences and Forest Ecology

University of Göttingen

second examiner:

**Prof. Dr. Holger Kreft**

Biodiversity, Macroecology & Biogeography

Faculty of Forest Sciences and Forest Ecology

University of Göttingen

submitted by

**B.Sc. Mats Mahnken**

21166058

Göttingen, August 2018

**Table of contents**

Abstract.....	i
Kurzfassung .....	ii
List of figures.....	iv
List of tables.....	iv
List of abbreviations .....	v
1 Introduction .....	1
1.1 Background and motivation .....	1
1.2 Research objectives.....	2
2 Theoretical background .....	4
2.1 Drivers and effects of land use transformation .....	4
2.2 Biological diversity .....	5
2.2.1 Measurement of biological diversity .....	5
2.2.2 Avian diversity.....	6
2.3 Regression modelling in ecology.....	7
3 Methodology.....	9
3.1 Data collection and preparation .....	9
3.1.1 Data collection .....	9
3.1.2 Extraction of predictors from remote sensing data.....	10
3.1.3 Data preparation and bird count data analysis .....	12
3.2 Model construction.....	12
3.2.1 Construction of regression models for avian species richness .....	12
3.2.2 Selection of appropriate statistical models .....	16

3.3	Model implementation and validation .....	17
3.3.1	Integration of predictors into land use change model (EFForTS-ABM)..	17
3.3.2	Model validation .....	18
4	Results .....	20
4.1	Analysis of remotely sensed predictors.....	20
4.2	Descriptive data analysis and correlation analysis.....	22
4.3	Model construction and selection.....	25
4.4	Model implementation .....	28
4.5	Model validation .....	30
5	Discussion.....	33
5.1	Results against the background of the research objective.....	33
5.2	Results against the background of other relevant studies .....	36
5.3	Methodological limitations .....	37
5.4	Model and data limitations.....	38
5.5	Conclusions and outlook .....	39
6	Literature .....	41
7	Appendix .....	47
7.1	Running the model implementation in NetLogo.....	47
	Acknowledgements.....	51
	Statutory declaration .....	52

## Abstract

The global challenge of ecosystem degradation and biodiversity loss affected by agricultural crop expansion poses major problems for many countries in tropical Asia. Indonesia faces the issues linked to large scale lowland rainforest conversion into cash crop systems of oil palm (*Elaeis guineensis*) and rubber (*Hevea brasiliensis*). Identification of synergy and trade-off effects for ecological and socioeconomic ecosystem functions is crucial to assure adequate biodiversity and ecosystem function conservation. Modelling avian  $\alpha$ -diversity as one facet of biodiversity is at the focus of the present study. Since birds provide a variety of ecosystem functions (e.g. pest control, seed dispersal, pollination), they are important entities to be considered in forests and agricultural systems.

Generalized linear regression models are used to derive statistical relationships between bird species richness and habitat as well as species richness of other taxa. A dataset collected in Jambi, Indonesia, of more than twenty potential predictor variables consisting of environmental, habitat and biodiversity variables is investigated in a regression analysis by a stepwise variable selection procedure. The best fit model describing total bird species richness has a deviance explained of 57.2 % with a relative root-mean-square-error of 26.08 %. The constructed model predicts the bird species richness from the species richness of beetles, ants, the canopy openness and the surrounding land cover diversity.

After the regression model construction, the model is implemented in an already existing land use change model (EForTS-ABM) in NetLogo. For this, predictor variables are simulated over the model landscape and the regression model is applied in each cell. The simulation is done by randomly assigning values to cells from smoothed distributions of the predictors from field data grouped by land use system. Additionally, a species spillover effect at land use borders is implemented for the predictor variables ant and beetle species richness. The model output is validated by comparison to an independent validation dataset. The validation yields no significant differences between model output means and validation data means except for jungle rubber agroforestry systems.

The present study shows that avian  $\alpha$ -diversity can be modelled by habitat and biodiversity of other taxa with relatively high goodness-of-fit, compared to other similar studies. The model constructed may be used to investigate specific land use scenarios in order to explore their effect on bird diversity.

## Kurzfassung

Die weltweite Ökosystemzerstörung und der Biodiversitätsverlust durch die Expansion landwirtschaftlicher Produktionssysteme stellt ein Problem für viele Länder im tropischen Asien dar. Indonesien sieht sich mit den Folgen groß angelegter Landnutzungs-Transformation von Tiefland-Regenwald zu Ölpalm- (*Elaeis guineensis*) und Kautschuk-Plantagen (*Hevea brasiliensis*) konfrontiert. Die Identifizierung von Synergie- und Trade-Off-Effekten der ökologischen und sozioökonomischen Ökosystemfunktionen ist wesentlich für die Sicherstellung eines angemessenen Biodiversitätsschutzes. Die Modellierung der Vogel- $\alpha$ -Diversität, als ein Aspekt der Biodiversität, steht im Fokus der vorliegenden Arbeit. Da Vögel eine Vielzahl an Ökosystemdienstleistungen (z.B. Schädlingskontrolle, Samenverbreitung, Bestäubung) bereitstellen, stellen sie wichtige Einheiten im Wald und in landwirtschaftlichen Ökosystemen dar.

Generalisierte lineare Modelle werden angewendet, um statistische Beziehungen zwischen Vogel-Artenvielfalt und Habitat, sowie Artenvielfalt anderer Taxa abzuleiten. Ein Datensatz aus Jambi, Indonesien, mit mehr als zwanzig potenziellen Prädiktoren aus Umwelt-, Habitat- und Biodiversitäts-Variablen wird mithilfe der Regressionsanalyse und einer schrittweisen Variablenselektion untersucht. Das bestangepasste Modell, das die Gesamt-Vogel-Artenvielfalt beschreibt, hat eine erklärte Devianz von 57,2 % und einen relativen *root-mean-square-error* von 26,08 %. Das konstruierte Modell schätzt die Anzahl der Vogelarten in Abhängigkeit von der Artenvielfalt von Käfern, Ameisen, der Kronenoffenheit und Landnutzungs-Diversität der Umgebung.

Nach der Regressionsanalyse wird das selektierte Modell in ein bereits existierendes Landnutzungsmodell (EFForTS-ABM) in NetLogo implementiert. Dazu werden die unabhängigen Variablen über die Modelllandschaft simuliert und danach in jeder Zelle das Regressionsmodell angewandt. Die Prädiktoren werden durch geglättete Verteilungen der unabhängigen Variablen aus den Felddaten, gruppiert nach Landnutzungssystem, simuliert. Außerdem wird ein *Spillover-Effekt* für die Prädiktoren Käfer-Artenvielfalt und Ameisen-Artenvielfalt an Landnutzungsgrenzen implementiert. Die Validierung des Modell-Outputs wird durch den Vergleich mit einem unabhängigen Validierungsdatensatz durchgeführt. Es können bis auf Werte aus Kautschuk-Wald-Mischsystemen keine statistisch signifikanten Unterschiede zwischen Modell-Output-Mittelwerten und Validierungsdatensatz-Mittelwerten festgestellt werden.

Die vorliegende Arbeit zeigt, dass die Vogel- $\alpha$ -Diversität durch Habitat und Artenvielfalt anderer Taxa mit relativ hoher Anpassungsgüte, verglichen zu anderen ähnlichen Studien, modelliert werden kann. Das erstellte Modell kann dazu genutzt werden Landnutzungsszenarien zu untersuchen, um deren Effekt auf die Vogel-Diversität festzustellen.

## List of figures

Fig. 1: Map of the research area depicting the sample sites ( <i>core plots</i> ).....	10
Fig. 2: Correlation of plot bird species richness to NDVI metrics collected in increasingly large circular buffers around sample plots.....	20
Fig. 3: Correlation of plot bird species richness to land use system proportions and land use diversity collected in increasingly large circular buffers around sample plots. ....	21
Fig. 4: Boxplots of bird species richness grouped by land use system.....	22
Fig. 5: Correlation matrix indicating <i>Pearson's correlation coefficient</i> between all variables in the dataset. ....	23
Fig. 6: Scatterplot for bird species richness and beetle species richness.....	24
Fig. 7: Scatterplot for bird species richness and land use diversity.....	25
Fig. 8: Model diagnostic for model 1 with field-observed bird species richness (y) plotted against modelled bird species richness (x). ....	28
Fig. 9: NetLogo landscape model output with land use (A) and corresponding cell bird species richness (B) for two different land use patterns (1 and 2).....	29

## List of tables

Tab. 1: Overview of potential predictor variables. ....	15
Tab. 2: Model diagnostics overview.....	25
Tab. 3: Model coefficient estimates with significance codes indicating thresholds for <i>p</i> -values. ....	27
Tab. 4: Comparison of descriptive statistics for model validation grouped by land use type. ....	31



## List of abbreviations

ABM	agent-based model
AIC	Akaike information criterion
BD	Bukit Duabelas (research region)
CRC 990	Collaborative Research Center 990
EFForTS	Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems
GAM	generalized additive model
GLM	generalized linear model
GLMM	generalized linear mixed model
HR	Harapan (research region)
NDVI	normalized difference vegetation index
RMSE	root-mean-square-error
VIF	variance inflation factor

# 1 Introduction

## 1.1 Background and motivation

Expansion of agricultural crop systems as a result of increasing human population paired with a growing per person consumption has led to a worldwide decline in biodiversity and a degradation of ecosystems (DIRZO AND RAVEN 2003; GIBBS *et al.* 2010). In tropical Asia, a major hotspot of biodiversity and endemism (MYERS *et al.* 2000) coincides with a rapidly growing population and associated land use conversion of large areas (SODHI *et al.* 2010; JONES 2013). The process of land conversion from rainforest into agricultural cash crops, like oil palm (*Elaeis guineensis*) and rubber (*Hevea brasiliensis*) has accelerated in the past decades in Indonesia and poses a substantial challenge for biodiversity conservation (MARGONO *et al.* 2014).

Biodiversity and linked ecosystem functions are at risk due to the progressing land use transformation (SALA *et al.* 2000). Knowledge about effects of land use conversion and impacts of potential future scenarios is crucial to mitigate negative consequences for biodiversity, ecosystem functions and human livelihoods. Therefore, different aspects of biodiversity need to be evaluated under ongoing land use change. This study aims at modelling avian species diversity, namely the bird species richness, as one facet of biological diversity.

Models, representing simplified abstractions of reality, provide valuable tools to investigate biological systems and understand their underlying processes or simulate specific scenarios. By application of regression analysis bird species richness is modelled by independent predicting environmental, habitat and biodiversity variables. After implementation in NetLogo (WILENSKY 1999) the constructed regression model is used to estimate bird species richness under simulated land use conversion in an artificial landscape. Critical evaluation of the constructed regression model and validation of the outcome of the prediction are crucial to gain inference on the quality of the model.

Birds provide a variety of ecosystem functions including pest control, seed dispersal and pollination, that entail great benefits for humans (KARP *et al.* 2013; MAAS *et al.* 2016). Some of these ecosystem functions have direct effects on yields in agricultural crop systems, such as the control of arthropod abundance (MAAS *et al.* 2013). Apart from the direct benefits of ecosystem functions that can be accessed by humans, it is important to investigate the diversity of species itself in order to safeguard that species are not lost

during the habitat degradation described above. Although the diversity of species is not a good that is directly marketable, societies still recognize the importance of preserving biodiversity (e.g. Convention on Biological Diversity, (UNITED NATIONS 1992)) attributing value to biological diversity itself.

A regression model of bird species richness enables the spatially explicit estimation of bird species richness in artificial landscapes as well as real landscapes. The outcome of the model indicates hot spots of bird  $\alpha$ -diversity as well as areas of low bird species diversity. The model also allows for the investigation of effects of different land use composition and configuration to identify scenarios that benefit bird species diversity.

The present study is part of the *Collaborative Research Center 990 (CRC 990): Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)* (EFForTS), which is an extensive international and interdisciplinary research project investigating implications of large scale land use transformation in the province of Jambi in Indonesia and whether ecological and socioeconomic synergy effects may be accomplished.

## 1.2 Research objectives

The overarching goal of this thesis is to establish a statistical regression model that predicts avian species richness in the study area based on a set of predictor variables. Additionally, the constructed regression model is implemented, and model quality as well as model behavior are assessed.

Regarding the presented thematic background and motivation, research questions which shall be answered in the course of this study may be derived:

- i. *Which environmental, habitat and biodiversity variables determine bird species richness in the land use transformation systems of Indonesia?*

Discovering the underlying ecological principles regulating the number of species at a given spatial point is not the aim of this investigation but identifying relationships between species richness and independent variables. The detection and description of these relationships is guided by theoretical considerations and empirical findings from the literature and base upon the statistical analysis of the data.

- ii. *How can environmental, habitat and biodiversity predictor variables be combined into a statistical regression model to predict bird species richness?*

Identification of a suitable model type and methodology for the prediction of bird species richness as well as the adequate selection of predictor variables and their combination in the model are substantial parts of the analysis.

- iii. *Can bird species richness be predicted by environmental, habitat and biodiversity variables in a sufficiently precise manner?*

Determining the quality of a model and assessing the power of the model to predict the target variable is important to provide information for potential users. The level of precision that is sufficiently precise may be derived from other studies investigating bird species richness by similar methods.

- iv. *What is the link between land use and bird species richness?*

The investigation of a direct connection between avian species richness and land use system (e.g. secondary forest, oil palm plantation, etc.) is of special interest since the rapid transformation of land use is the main research subject of the EF-ForTS project and an important driver of habitat alteration in the study area. If possible and reasonable the land use is implemented as a predictor variable in the constructed regression model.

The search for answers to these questions guides the analysis in the subsequent chapters. Encountering responses to the stated research questions is the ultimate objective of this study.

## 2 Theoretical background

### 2.1 Drivers and effects of land use transformation

The conversion of forest into agricultural crop systems which has substantial ecological and socioeconomic consequences is a phenomenon that can be found in large areas of tropical Asia (ZHAO *et al.* 2006). Due to the land use transformation anthropogenic CO<sup>2</sup> emissions have increased (HOUGHTON 2003), air and water quality have deteriorated (AKIMOTO 2003), the regional climate is altered (PIELKE 2005) and biodiversity loss has increased (PIMM *et al.* 1995). Indonesia faces the same effects of land use conversion as many other tropical Asian countries since large areas of tropical forest have been transformed into agricultural cropping systems (MARGONO *et al.* 2014).

Drivers for the conversion of tropical lowland rainforest to agricultural cropping systems in Indonesia include policy changes and socioeconomic factors (GATTO *et al.* 2015). Policy as implemented by the Indonesian government aiming at the transmigration from highly populated islands to more sparsely populated areas of the country with a coupled incentive for oil palm cultivation and other economic activities has led to an increase of cash crop production area since the 1980s (LARSON 1996).

The primary land use systems that may be found in the research region Jambi are tropical lowland rainforest (forest), jungle rubber agroforestry systems (jungle rubber), rubber monoculture plantations (rubber) and small as well as large scale oil palm monoculture plantations (oil palm) (DRESCHER *et al.* 2016). These land use systems present different habitat structures for the species occupying them.

Effects of land use conversion from forest to agricultural plantations, like oil palm, include severe ecosystem function alterations because of a decrease in multitrophic biodiversity that often leads to a reduction in functional diversity and a decline of the resilience of communities (BARNES *et al.* 2014; DRESCHER *et al.* 2016; DISLICH *et al.* 2017). However, the cultivation of cash crops increases the amount of produced sellable goods per area (DISLICH *et al.* 2017), which in turn secures livelihoods of small scale farmers.

## 2.2 Biological diversity

### 2.2.1 Measurement of biological diversity

Biological diversity or biodiversity as the variation of life forms ranges from genetic diversity to the diversity of ecosystems (MAGURRAN 2011). In the present study, species in an assemblage of a given taxon, birds, are at the focus. Species diversity comprises multiple facets that may be investigated. Since a species consists of individuals, the variation of the number of individuals between species can be considered as well, which reveals assemblage structures of evenly distributed or dominating species. The evenness or dominance of species in an assemblage may be expressed as biodiversity indices, like the *Shannon-Weaver* index (SHANNON AND WEAVER 1949) or the *Simpson* index (SIMPSON 1949). The number of species in an assemblage may be considered as the simplest measure of species diversity, which does not account for evenly distributed or dominating species and can be derived from occurrence and abundance data (MAGURRAN 2011).

Sampling of individuals for estimates of species diversity is a demanding task since low evenness, which is frequently found in assemblages, means that some species are present with very few individuals that are difficult to sample, while other species with high abundance are sampled often. This results in a condition where more time and effort put into sampling increases the number of species found. Hence, the species richness estimate is highly dependent on sampling intensity. In addition, the detectability of individuals varies between species considering different diurnal and seasonal activity rhythms, representing another challenge for sampling (MAGURRAN 2011).

The absolute number of species in an assemblage may be estimated from the sample by different non-parametric techniques like the *Chao*-estimator (CHAO 1984), the *Jackknife* statistics (BURNHAM AND OVERTON 1978) or the bootstrap method (SMITH AND VAN BELLE 1984). Furthermore, the number of species may be estimated by extrapolating species accumulation curves that consist of the accumulated number of species which show a constantly increasing trend with growing sample size (MAGURRAN 2011). These species richness estimation techniques can only be applied if multiple independent observations from a community were gathered.

Accordingly, the measurement or estimation of species diversity at a given location from one observation at a sample plot is a challenging task. If the objective is not estimating the total number of species at a location but comparing communities between locations,

a constant sampling intensity over sampling sites or rarefaction techniques facilitate the comparability and a fully inventoried community is not required (MAGURRAN 2011).

### **2.2.2 Avian diversity**

Most bird's primary mode of locomotion is flight. Therefore, this taxon of vertebrates is highly mobile. A quick way of travelling and the affiliated great mobility of individuals pose problems for the detection and sampling of species for a given spatial location at a small scale.

Following the example of DARRAS (2016), sampling techniques for bird individuals consist of visual and aural point count surveys. A given site is sampled for a predetermined time (e.g. 20 minutes) and each bird individual detected by the person surveying is recorded. The sampling may be repeated (e.g. four times) at a location to enhance the completeness of the sample. Audio recordings can be analyzed for complementation of the visually detected individuals. Daytime of sampling has to be considered because of the diurnal activity differences bird species exhibit.

Avian community diversity is determined by climate and by habitat, while climate plays a larger role on macrogeographical scales and habitat structure is more important on a local scale (TELLERÍA *et al.* 1992). Climatic parameters may control the "theoretical pool of species" (IBID.) but habitat structure determines which of the species can actually be found in a given area (IBID.). Indicators for habitat diversity, especially linked to diversity of vegetation seem to play a vital role in the capacity of a habitat to host a larger number of bird species (MACARTHUR AND MACARTHUR 1961; TELLERÍA *et al.* 1992). A positive correlation between habitat diversity and species diversity can be explained by a larger number of ecological niches in a diverse habitat that can be occupied by more species (ROSENZWEIG 2002).

The modelling of bird species richness, as an indicator for avian species diversity and associated functional diversity, demands for a distinct definition of what species richness at a given location means. If a given species is present in an area, whether it can be detected or not, it may be included as a species in the biodiversity pool of that location since it is contributing to ecological functions in that area as a part of the biocenosis. This means that irrelevant of the fact if the species is residential and breeding in that area or passing by in a migration event, it is included in the community. Hence, the avian species pool is different from what can be found in the field because of detectability and implications of sampling stated above. Therefore, the species richness assumed in the present study has

to be considered a proxy for the actual species richness which in turn is a measure incompletely describing avian biodiversity.

### 2.3 Regression modelling in ecology

Statistical regression modelling in ecology is used to determine and mathematically describe interactions between ecological entities. Regression models in ecology are either explanatory, predictive or a combination of both. Explanatory models investigate the underlying ecological processes that produce the observed patterns, whereas predictive models seek to provide a statistical relationship between dependent and independent variables in order to estimate the dependent variable from observations of the independent variable (GUISAN *et al.* 2002).

The modelling of bird species diversity can be achieved by different techniques. Total species richness can be estimated by summing several species richness predictions from species distribution models at a given geographical area (LEHMANN, LEATHWICK *et al.* 2002; RION 2010). Alternatively, species richness can also be directly estimated with regression models including total species richness as dependent variable (RION 2010). The present study focuses on the approach of directly modelling species richness as a function of correlated environmental, habitat and biodiversity predictors.

In modern species distribution modelling general linear models (GLMs) and general additive models (GAMs) are widely applied because they do not assume distributions that the data cannot meet and they are able to fit to the data more flexibly than other regression techniques (GUISAN *et al.* 2002). Due to their inherent empirical nature, regression models incorporate information about realized niche in contrast to fundamental niche, including biotic interactions and stochastic effects. This hampers the comparability between locations, even for identical species, since the biotic interactions and stochastic effects are unequal between sites which is ignored in the regression model (GUISAN AND ZIMMERMANN 2000).

Several studies exist that use regression analysis techniques for bird species diversity modelling. KOH (2008) models bird species richness of oil palm plantations in Malaysia with generalized linear mixed effects models (GLMMs), which can be seen as extensions of GLMs, and includes predictor variables describing habitat and sampling effort. LUOTO *et al.* (2004) model bird species richness with GLMs as a function of remotely sensed predictor variables describing the vegetation. Although GAMs are able to describe non-



linear relationships between target and predictor variables, the interpretation of the results is more difficult than for linear models, which is important in cases where the model is used as an explanatory tool rather than used for predictive estimations.

Statistical regression models pose some inherent problems that hamper model quality in terms of how well the model depicts real interactions or even effects of independent variables on dependent variables. Omitted-variable bias is one of the issues that occurs when one or more relevant predictor variables are left out of the regression model, resulting in a bias of the coefficient estimates of the predictors included in the regression model (FAHRMEIR *et al.* 2013). Connected to the omitted-variable bias is the problem of confounding factors in regression analysis that occurs if a variable exists that influences both predictor and target variable, hence it is called a confounding variable because the effect may be attributed to that variable and not the included predictor leading to incorrect inference (VANDERWEELE AND SHPITSER 2013).

### 3 Methodology

#### 3.1 Data collection and preparation

##### 3.1.1 Data collection

The research area of the EFForTS-project extends over the province of Jambi on Sumatra, Indonesia. Jambi's landscape stretches from a mountainous area to wide lowlands. The local climate in the lowlands is tropical humid, consisting of two rainy seasons in March and December with a dry season from July to August (DRESCHER *et al.* 2016).

The area has an extended history of rainforest usage by traditional agroforestry practices and logging as well as usage of non-timber forest products (ANDAYA 1993; KATHIRITHAMBY-WELLS 1993). Since commercial logging concessions were first issued in the 1970s (SUYANTO *et al.* 2000) governmental policies with a linked population migration aim at promoting economic growth in the area (ELMHIRST 2011; GATTO *et al.* 2015) which in turn increases pressure on forest resources. Intensively managed agricultural plantations of cash crop trees such as rubber and oil palm have replaced former rainforest areas (DRESCHER *et al.* 2016). Expanding agricultural activity has led to a land use change from rainforest to agricultural land that resulted in a rainforest cover of 30 %, an agricultural land cover of 55 % and a cover of 10 % degraded lands mostly pending for conversion into agricultural land in 2013 (IBID.)

The data analyzed in the present thesis was collected on sample plots, named *core plots* (see Fig. 1) that were established in the region to facilitate the purpose of the project of investigating effects of land use change in an interdisciplinary way. The research area consists of two spatially separated subareas, Harapan (HR) and Bukit Duabelas (BD). Observations made by different subparts of the project were carried out on these *core plots* and thus are linked to each other via the location. A total of 32 sample plots were established with eight plots located in each of the land use systems forest, jungle rubber, rubber and oil palm. The sample sites were chosen following an experimental scheme as replications within each land use type.

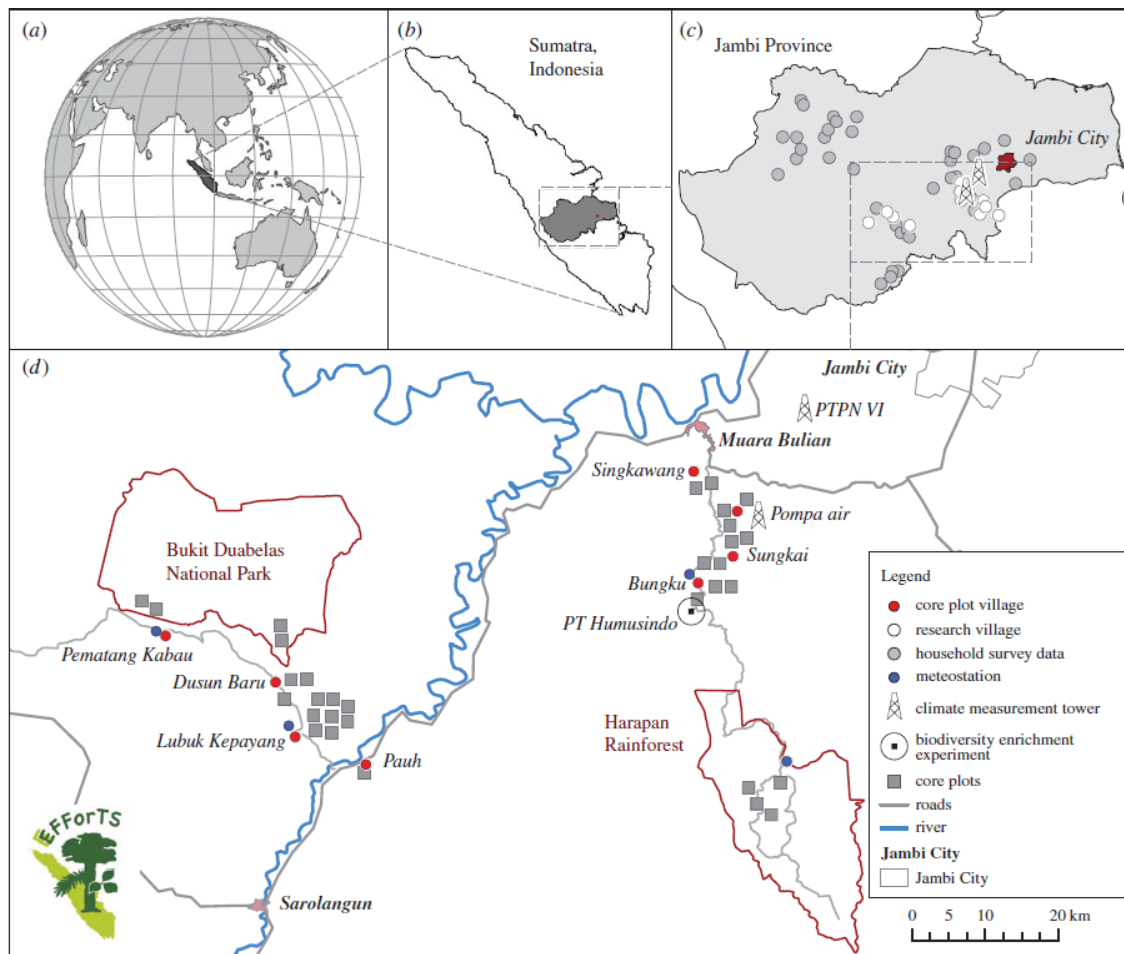


Fig. 1: Map of the research area depicting the sample sites (*core plots*). (illustration from DRESCHER *et al.* 2016)

Field data from *core plots* surveyed in 2013 is gathered in the present study, which include environmental variables, variables that characterize the habitat and species diversity (richness) of different taxa. In addition, remote sensing products were analyzed to gather more potentially predicting variables describing the habitat.

### 3.1.2 Extraction of predictors from remote sensing data

Remote sensing variables, like the normalized difference vegetation index (NDVI), have been shown to correlate with avian species richness and abundance (LUOTO *et al.* 2004; NIETO *et al.* 2015). Therefore, the NDVI and other remote sensing statistics, namely land use information from satellite data classification, is included into the regression analysis. Because of the high mobility of birds, the surrounding of the sample plots is considered to have an effect on the number of bird species that can be found on the sample plot. Consequently, the remote sensing variables are not only investigated on the sample plot itself, but also in the surrounding on increasingly large concentric areas.

A land use map from 2013 (MELATI 2017) is analyzed to gather information on the surrounding of the *core plots* that may influence the species richness found on the corresponding plot. The *Shannon-Weaver index* (SHANNON AND WEAVER 1949) (see below) for land use is calculated for multiple increasingly large concentric circles around the plot location to investigate land use diversity. Additionally, land use system proportions around the plot center are derived from the land use map.

$$\text{Shannon-Weaver index: } H' = - \sum_{i=1}^C p_i * \ln(p_i) \quad , \text{ with } p_i = \frac{n_i}{N}$$

and

$C$  = number of land use classes,

$p_i$  = relative proportion of land use class  $i$ ,

$n_i$  = number of pixels/cells in land use class  $i$ ,

$N$  = total number of pixels/cells.

*Pearson's correlation coefficient* (see below) is then calculated for avian species richness to inspect at which distance the relation is strongest between surrounding land use and the target variable to support a decision which buffer distance is used to derive potential predictors for the regression analysis. The correlation coefficient is used to determine relationships between random variables and describes the linear correlation ranging from - 1 (strong negative correlation) over 0 (no correlation) to + 1 (strong positive correlation) (CHATTERJEE AND HADI 2006):

$$\text{Pearson's correlation coefficient: } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

with

$n$  = sample size,

$x_i, y_i$  = individual observations of variable  $x$  and  $y$  indexed by  $i$ ,

$n_i$  = number of pixels/cells in land use class  $i$ ,

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  = sample mean of variable  $x$ ,

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  = sample mean of variable  $y$ .

Satellite data (satellite constellation from RapidEye/Blackbridge) from 2013 with a spatial resolution of 5 m is used to investigate the vegetation in the surrounding of the sample plots. The remotely sensed imagery is used to calculate the NDVI per pixel in the image, which is then used to check for relationships between surrounding NDVI and bird species

richness on the sample plot. The variability of NDVI values (pixel standard deviation) and mean pixel NDVI are investigated in increasingly large concentric circles around the plot with the same technique as for land use (see above). Then the correlation to avian species richness is checked for the identification of correlation peaks.

### 3.1.3 Data preparation and bird count data analysis

The analyzed dataset consists of a variety of preprocessed variables that were collected by different working groups using a wide range of observation techniques and processed to be available for the analysis in other research studies within the CRC 990. Initially, variables are removed, that are not expected to affect bird species richness according to the literature. Accordingly, soil characteristics, like nutrient availability for individual elements, or biotic soil activities, like microbial or fungal activity data are removed from the dataset. Then the variables are converted into a unified dataset with *core plot* as a unique identifier to merge observations by location. For this purpose, biodiversity data available as individual species abundance data are converted into species presence-absence data and finally translated into species richness per plot, which is performed for multiple taxa.

Descriptive statistics and figures are produced to investigate the data at hand and explore potential relationships in the data. Scatterplots help to detect potentially useful relations between target and predicting variables. For information about multicollinearity in the predictors a correlation table graph is used. The correlation is measured by *Pearson's r*. A *Kruskal-Wallis-test* for comparison of independent samples of non-normally distributed random variables (KRUSKAL AND WALLIS 1952) is applied to check for statistically significant differences between bird species richness in different land use systems and in the two different research landscapes (HR and BD). If significant differences in the set of groups are found by the *Kruskal-Wallis-test*, the determination of which groups show statistically significant differences between each other is done by a post-hoc-test (*Nemenyi-test*) (NEMENYI 1963).

## 3.2 Model construction

### 3.2.1 Construction of regression models for avian species richness

Statistical models in the context of ecology are used to gain inference about underlying processes and ecological principles as well as to predict dependent variables from

observations of correlated independent variables. They are used as a “basis for interpretation” to explore “fit” as well as to investigate the “strength” between the variables, that is statistical significance of the relationship (GUISAN *et al.* 2002). In addition to the general subject of the investigation of underlying ecological causalities, mathematical (statistical) models are used to simulate and predict respective response variables where they are not or cannot be measured (IBID.).

The principle of parsimony guides the search for a model since the simplest explanation for a problem is usually the best (*Occam’s Razor*). In the case of statistical models, the simplest model means the simplest model type and fitting methodology with the least amount of predictor variables (e.g. a simple linear model fitted with the ordinary least squares method), since these models are easier to construct, analyze, validate and convey to other people. In fact, this principle keeps the models as interpretable as possible, since excessively complex models are harder and sometimes impossible to interpret. Nevertheless, some ecological questions cannot be investigated with simple linear regression because the data does not fit assumptions needed for the respective models. Then more complex modelling techniques are applied which may compromise the concept of easily interpretable models but ensures that the data can be analyzed in the best way possible under the given circumstances (GUISAN *et al.* 2002).

In the course of the present study the constructed statistical model is used as a prediction tool and means to investigate underlying patterns. The relationship between predictor variables and species diversity of birds as depicted in the data is used to predict species diversity using the environmental, habitat or biodiversity variables as explaining variables for modelling bird species diversity in areas where the bird species richness was not measured.

GUISAN *et al.* (2002) summarize the use of statistical regression models in ecology and species distribution modelling and indicate the benefits and drawbacks of specific modelling techniques. Linear regression models of the form

$$Y = \alpha + X^T \beta + \varepsilon_i$$

with

$Y$  = response variable,

$\alpha$  = intercept,

$X$  = vector of predictor variables,

$\beta$  = vector of regression coefficients,

$\varepsilon$  = errors/residuals,

are limited through their assumptions that must be met to properly apply the methodology in an appropriate context. These assumptions are:

1. identically and independently distributed errors  $\varepsilon_i$  over all observations as well as a homoscedastic response variable  $Y$ ,
2. normal distribution in the errors  $\varepsilon_i$ ,
3. and a linear relationship between predictor and response variable.

If these assumptions are not met, the data is usually transformed via various techniques that either alter values of the response variable or the predictor variable. To overcome these limiting assumptions other modelling techniques can be applied that eliminate the need for the data to fit into these premises (GUISAN *et al.* 2002).

Regression methods used for modelling species richness as a function of environmental variables include GLMs (GUISAN *et al.* 2002; LEHMANN, OVERTON *et al.* 2002; HILDEBRANDT 2015). GLMs provide a possibility to overcome the restrictive assumption of normally distributed target variables of linear regression models. Since species richness is a discrete variable that is in the case of sampling an outcome of counting individuals and species, a Gaussian normal distribution cannot be assumed. A Poisson distribution must be assumed and incorporated in the regression methodology. Hence, GLMs account for the deviance from normal distribution (FAHRMEIR *et al.* 2013).

Considering a log-link function in the GLM because of the Poisson distributed target variable, the regression model formula becomes:

$$Y = e^{\alpha + X^T \beta + \varepsilon_i}$$

with

$Y$  = response variable,

$\alpha$  = intercept,

$X$  = vector of predictor variables,

$\beta$  = vector of regression coefficients,

$\varepsilon$  = errors/residuals.

The regression analysis is performed for different sets of response variables. These include total bird species richness, bird species richness separated for research area (HR and BD) and bird species richness of different diet groups of birds. Potential predictor variables that are included in the regression analysis are shown in Tab. 1.

Tab. 1: Overview of potential predictor variables. (own illustration)

predictor variable	short	variable type	description
air temperature	<i>temp</i>	continuous	mean below-canopy air temperature (°C)
mean humidity	<i>hum</i>	continuous	mean below-canopy air humidity (%)
elevation	<i>elev</i>	continuous	elevation above sea level (m)
canopy openness	<i>canop</i>	continuous	percentage of ground area not covered by tree crowns (%)
litterfall	<i>lit</i>	continuous	leaf litterfall (Mg/ha*yr)
biomass carbon	<i>biom</i>	continuous	total tree biomass carbon (Mg/ha)
plant species richness	<i>plant</i>	discrete	number of understorey plant species observed
ant species richness	<i>ant</i>	discrete	number of ant species observed
oribatida species richness	<i>orib</i>	discrete	number of oribatida species observed
protist species richness	<i>prot</i>	discrete	number of protist species observed
invertebrate species richness	<i>inv</i>	discrete	number of invertebrate species observed in leaf and litter
spider species richness	<i>spid</i>	discrete	number of spider species observed in leaf and litter
chilopoda species richness	<i>chil</i>	discrete	number of chilopoda species observed
beetle species richness	<i>beetle</i>	discrete	number of beetle species observed in leaf and litter
butterfly species richness	<i>butter</i>	discrete	number of butterfly species observed
land use diversity	<i>ludiv</i>	continuous	land use diversity in plot surrounding of a radius of 2300 m calculated with <i>Shannon-Weaverr index</i> (see 3.1.2)
NDVI	<i>ndvi</i>	continuous	mean normalized difference vegetation index (NDVI) in plot surrounding of a radius of 2300 m (see 3.1.2)
NDVI variability	<i>ndvivar</i>	continuous	standard deviation of the NDVI in plot surrounding of a radius of 2300 m (see 3.1.2)
land use intensity	<i>lui</i>	continuous	intensity of management practices (SAHNER 2016)
stand structural complexity	<i>ssc</i>	continuous	measure for overall structural vegetation complexity (EHBRECHT <i>et al.</i> 2017)
effective number of layers	<i>enl</i>	continuous	measure of vertical structure (EHBRECHT <i>et al.</i> 2016)



Following this methodology, multiple models are constructed in the statistical computing software R (R DEVELOPMENT CORE TEAM 2018) that are then subject to selection based on model diagnostic criteria. Initially, a full model is constructed including all available data from the two landscapes HR and BD with the target variable total bird species richness. Additionally, two separate models for HR and BD are constructed with  $n=15$  and  $n=12$  respectively to look at differences between the landscapes. Adjustment of the target variable allows for further integration of additional dietary information (adopted from DARRAS (2016)), that may improve the model. Therefore, a diet group is assigned to most of the bird species that can be used to model the bird species richness on the level of a diet group. The four separate diet groups are insectivorous birds, omnivorous birds and alternative-diet birds of either frugivorous, granivorous, nectarivorous or carnivorous diet as well as one group of bird species with unknown diet type. The diet groups are selected in the presented way to assure equally sized groups of species and no diet groups of very few species. The diet-separated bird species models are aggregated so that a model for the total species richness is constructed. Diet-separated bird species richness models are constructed for the whole research area with 27 observations.

### 3.2.2 Selection of appropriate statistical models

Model fit needs to be assessed to select appropriate regression models and discard models that are not well suited to predict bird species richness. Therefore, the goodness-of-fit of the model is checked. One measure for assessing the goodness-of-fit of the model is the *Akaike information criterion* (AIC) (AKAIKE 1974) that is used to rank models according to their fit assuming a lower AIC value indicates a better model fit.

A stepwise variable selection procedure (RION 2010) is applied to pick predictor variables that are best suited for the prediction of the target variable. AIC is used as a measure to compare between models constructed with the same dataset, but different predictor variables, favoring the model with the lower AIC. Accordingly, the first model investigated is the full model with all potential predictor variables included. In cases where the number of observations is lower than the number of potential predictors, the stepwise procedure is repeated with multiple starting models consisting of different sets of predictor variables. The stepwise variable selection then starts to eliminate predictors until the lowest possible AIC is reached. Re-inclusion of a predictor after exclusion is possible if the produced model would have a lower AIC (FAHRMEIR *et al.* 2013).

The variance inflation factor (VIF) (BRAUNER AND SHACHAM 1998) is calculated for models with multiple independent variables to account for potential multicollinearity in the predictors. This approach is applied for the elimination of collinearity in predictors that can compromise model quality. If the VIF exceeds the threshold value of 10 (FAHRMEIR *et al.* 2013) the predictor variable that shows the highest VIF is discarded from the set of predictors.

For further selection of a model for the implementation from the presented set of models for total bird species richness, research site separated bird species richness and diet type separated bird species richness the adjusted  $R^2$ , deviance explained, root-mean-square-error (RMSE) between the observed and the modelled bird species richness as well as the number of predictors are considered.

### **3.3 Model implementation and validation**

#### **3.3.1 Integration of predictors into land use change model (EForTS-ABM)**

The EForTS-ABM (DISLICH *et al.* 2018) is an agent-based model that describes the effects of tropical land use change on ecological and socio-economic functions. The model is parameterized for the study region in Jambi and includes ecological and economic sub-models implemented in an artificially generated mosaic landscape of land use systems that is controlled by households. The model landscape consists of 100 x 100 quadratic cells (also called patches) representing each 50 x 50 m. The predictor variables integrated in the selected regression model for bird species richness estimation are implemented in the EForTS-ABM since they are prerequisite to apply the model in order to predict bird species richness.

Independent variables are simulated by random sampling from a predefined probability density distribution computed from the field observations for each of the land use systems. For the predictor probability density distributions, the observed sample histogram is smoothed with kernels to generate an averaged distribution of the predictor. Kernel density estimation is performed in order to estimate the probability density distribution of all predictor variables by application of a Gaussian kernel with band width estimated after SILVERMAN (1986). The probability density distribution estimated with the kernels is then discretized to 512 values from which the weighted random sampling is done. For each cell in the model a weighted random selection from the predictor probability density distribution then defines the specific value.

Landscape metrics of land use composition in the surrounding of a cell are computed in the same way as done for the extraction of the predictor. Therefore, each cell is considered as the center of a 2300 m circular neighborhood that is used to compute the land use metrics (see 3.1.2).

Furthermore, spillover effects at the borders of land use systems may be included, depending on the predictors implemented. The spillover may happen for species richness of other taxa, assuming that more species can be found at the edges to land use systems that host a larger number of species than the local system. This may be backed up by findings from NURDIANSYAH (2016) in oil palm plantations.

After predictor simulation, the constructed regression model is applied to each of the cells in the model landscape to estimate the cell bird species richness from the set of predictor values. Mean overall cell bird species richness and land use grouped mean cell bird species richness are computed. Additionally, a graph displaying the distributions of cell bird species richness grouped for land use system is generated for the inspection of variability in cell bird species richness.

### 3.3.2 Model validation

Model output validation is done to gain inference on how well the implemented model represents the observations in the field (that were also used to construct the model) and to compare the output to an independently collected dataset of bird counts, collected in 2015 in the research area, but on other sample plots. For this purpose, a *virtual ecologist approach* (ZURELL *et al.* 2010) is applied that simulates sampling from the landscape-output of the model. The sampling that is identical to field sampling in the independent dataset is repeated 100 times and averaged statistics are compared to the validation dataset and model input dataset.

For output comparison to the model validation dataset the sampling approach of 28 sample plots per land use type is implemented by a repeated random selection of 28 cells of the model landscape per land use type and computation of the averaged means with associated confidence interval of certainty level  $\alpha \approx 0.95$  as well as averaged standard deviations to be compared to the field sample. The confidence interval is constructed using twice the standard error to either side of the averaged mean that can be produced by calculating the standard deviation of multiple sample means from the same population.

Multiple land use scenarios with different land use configurations are simulated and bird species richness is estimated over the artificial landscape to gain inference on model behavior in the ABM under different scenarios.

## 4 Results

### 4.1 Analysis of remotely sensed predictors

Potential remotely sensed predictor variables computed by analysis of satellite imagery and a classified land use map are investigated for their correlation to the observed bird species richness on the plot. Because of restrictions of the map area, the largest buffer size for the surrounding land use analysis is 3900 m and because of restrictions of the satellite data, the largest buffer size for the vegetation analysis by NDVI is 3500 m.

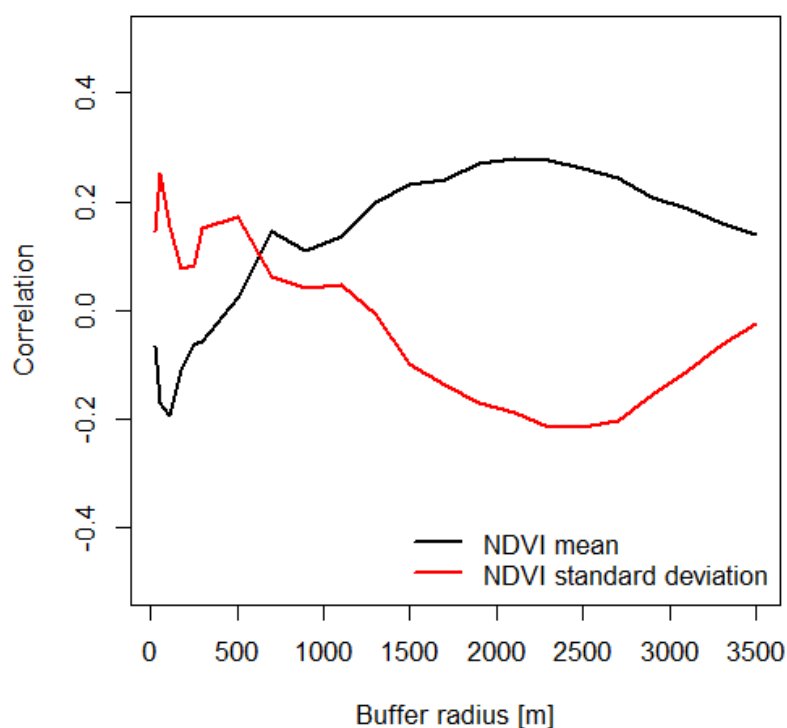


Fig. 2: Correlation of plot bird species richness to NDVI metrics collected in increasingly large circular buffers around sample plots. (own illustration)

The correlation of mean NDVI and NDVI diversity (cell standard deviation) to bird species richness exhibit opposing trends. A positive correlation of bird species richness with NDVI diversity corresponds to a negative correlation with mean NDVI at small buffer radii. Between 500 m and 1000 m buffer radii, the correlation of NDVI diversity and mean NDVI to bird species richness change sign and indicate an inverse trend for larger buffer sizes.

The increasingly large circular buffers around the sample plots show peaks of correlation for mean NDVI and NDVI diversity at a small distance to the plot center and at around

2300 m radius to the plot (see Fig. 2). Because of the peaks at 2300 m the computed values for mean cell NDVI and NDVI cell standard diversity are included into the potential set of predictor variables that is later used in the regression analysis.

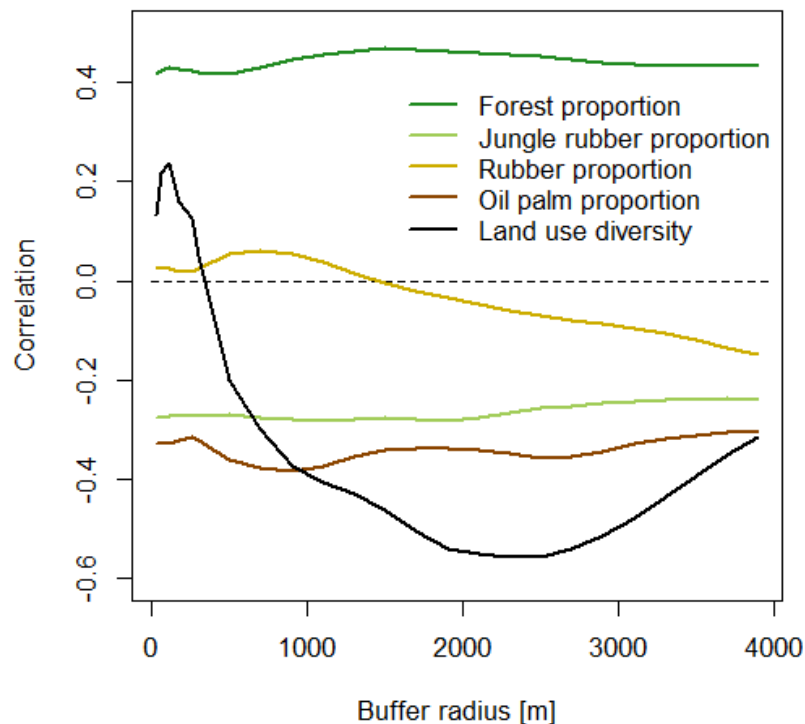


Fig. 3: Correlation of plot bird species richness to land use system proportions and land use diversity collected in increasingly large circular buffers around sample plots. (own illustration)

Land use proportions exhibit varying degrees and tendencies of correlation to plot bird species richness. Independent of the buffer radius, forest, jungle rubber and oil palm show constant correlations to bird species richness, where forest correlates strongly positive (+ 0.4) and jungle rubber and oil palm correlate moderately negative (- 0.25). The low correlation of rubber varies over increasing buffer size to a small degree, where first a positive correlation may be seen that becomes negative at a buffer size of around 1500 m. The correlation of land use diversity to plot bird species richness varies over increasing buffer size. In small surroundings to the sample plot a low positive correlation may be found that decreases until a buffer radius of 2300 m, where a strong negative correlation (- 0.55) may be found. After that, the correlation slowly increases with growing buffer radius.

The strongest absolute correlation may be found for plot bird species richness and land use diversity when considering a circular area of radius 2300 m around the plot (see Fig. 3). Therefore, the derived land use diversity in a buffer of 2300 m around the sample plot is included in the set of potential predictors for the regression analysis.

## 4.2 Descriptive data analysis and correlation analysis

Bird species richness and variability between observations vary between land use systems (see Fig. 4). The highest observed median bird species richness can be found in forest with 18 (mean: 18.86) species per plot. Jungle rubber and rubber host a median bird species richness per plot of 15 (mean: 13.16) and 11.5 (mean: 13.88) respectively, while oil palm plots host the lowest median number of bird species per plot of 9 (mean: 9.14). Variability between observations is highest in rubber (standard deviation: 5.74) and forest (standard deviation: 5.14), intermediate for jungle rubber (standard deviation: 4.02) and lowest for oil palm (standard deviation: 1.77). The *Kruskal-Wallis-test* indicates statistical differences between land use types that are uncovered via the post-hoc-test indicating significant differences ( $p = 0.002$ ) between forest and oil palm land use systems in terms of bird species richness.

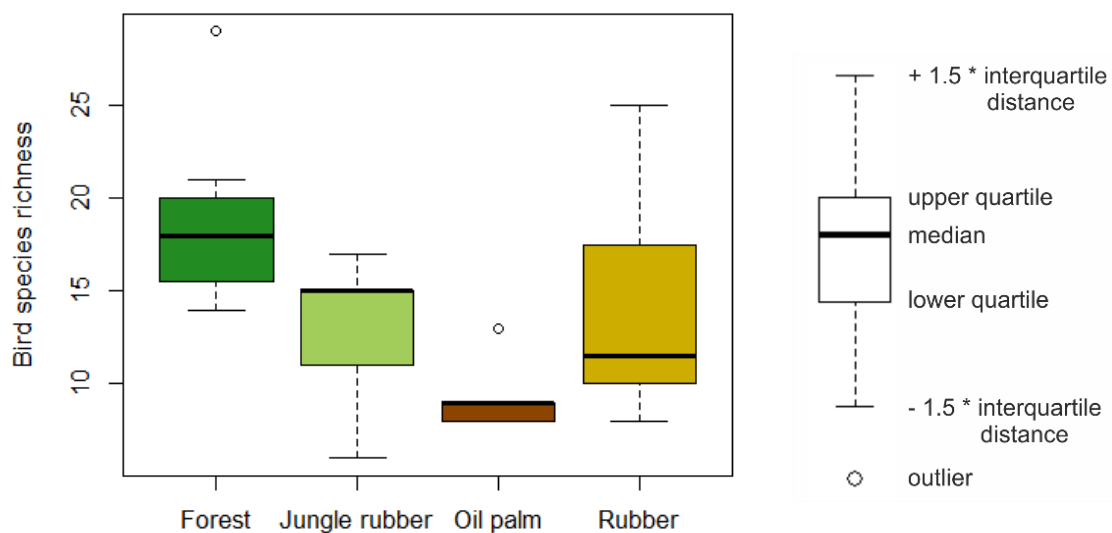


Fig. 4: Boxplots of bird species richness grouped by land use system. (own illustration)

The observations of overall bird species richness per plot and variability of observations between the two landscapes BD and HR differs to a small extent. The research landscape BD presents a higher average bird species richness per plot of median 15 (mean: 14.85), while the median bird species richness in HR is 11 (mean: 12.87). The variability of observations is similar between BD and HR with standard deviations of 5.41 and 5.63 respectively. The *Kruskal-Wallis-test* indicates no significant difference ( $p = 0.21$ ) between the two landscapes in terms of overall bird species richness.

*Pearson's correlation coefficient* reveals the strength of the linear relationship between the target variable total bird species richness and the potential predictor variables as well as the correlation between the predictor variables (see Fig. 5).

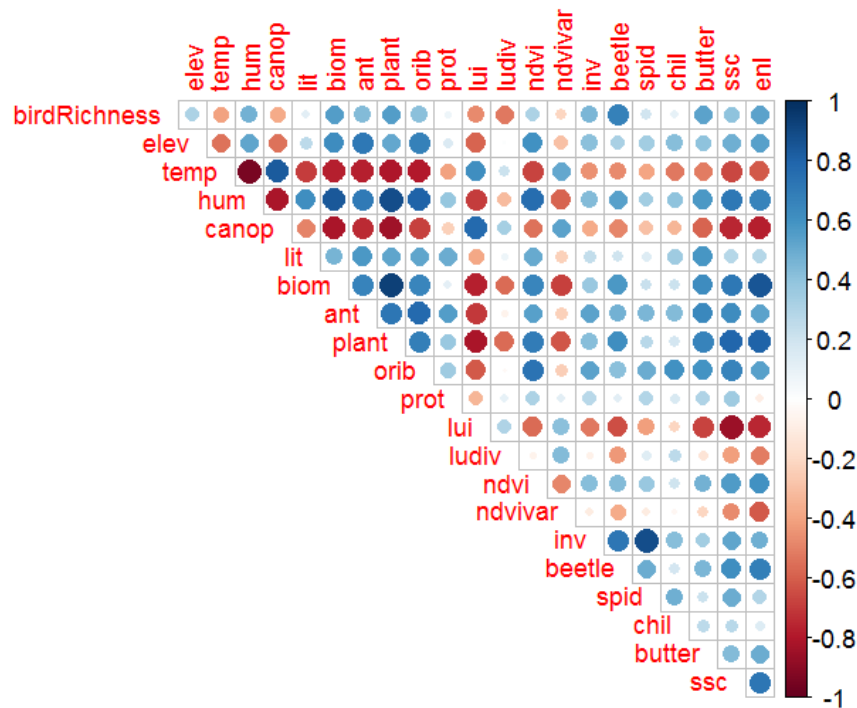


Fig. 5: Correlation matrix indicating *Pearson's correlation coefficient* between all variables in the dataset. (own illustration)

Bird species richness shows strong positive correlation with beetle species richness, biomass carbon and plant species richness while strong negative correlation can be found for land use diversity and land use intensity. Less strong positive correlations may be found for elevation, humidity, ant species richness, oribatida species richness, invertebrate species richness, butterfly species richness, stand structural complexity and effective number of layers. Weak negative correlations are found between bird species richness and temperature as well as canopy openness. No correlation may be found between bird species richness and litterfall, protist species richness, spider species richness, chilopoda species richness and mean NDVI as well as NDVI diversity. A large amount of multicollinearity can be found between some of the predictors, for example temperature and effective number of layers exhibit strong linear correlations with many of the potential predictor variables.



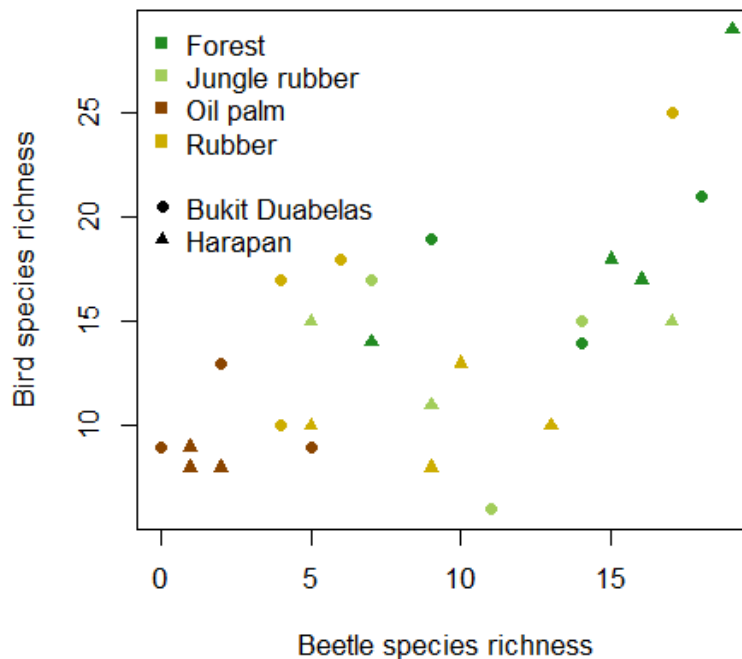


Fig. 6: Scatterplot for bird species richness and beetle species richness. (own illustration)

Bird species richness and beetle species richness show a strong positive correlation (*Pearson's*  $r = 0.65$ ) indicating an increasing number of bird species with increasing number of beetle species on a plot (see Fig. 6). Forest sample plots show both a high richness of bird species and high richness of beetle species, while oil palm plots show lower numbers of bird and beetle species. Although the overall trend is positive, looking at the observations in land use types independently a larger beetle richness correlates negatively with bird species richness in rubber plantations if the largest observation is considered as an outlier.

Fig. 7 shows the relationship between bird species richness and land use diversity in a surrounding of 2300 m. Bird species richness negatively correlates with land use diversity (*Pearson's*  $r = -0.53$ ) indicating that a low land use diversity in the surrounding relates to a high bird species richness. All sample plots in the HR research area located in forest show very low surrounding land use diversity and medium to high bird species richness, while forest plots in BD show medium bird species richness, but intermediate to high surrounding land use diversity. Oil palm sample sites of both research areas show low bird species richness, but high land use diversity in their surroundings. Jungle rubber and rubber sample plots scatter between forest and oil palm with an intermediate to low number of bird species with intermediate to high land use diversity in their surroundings.

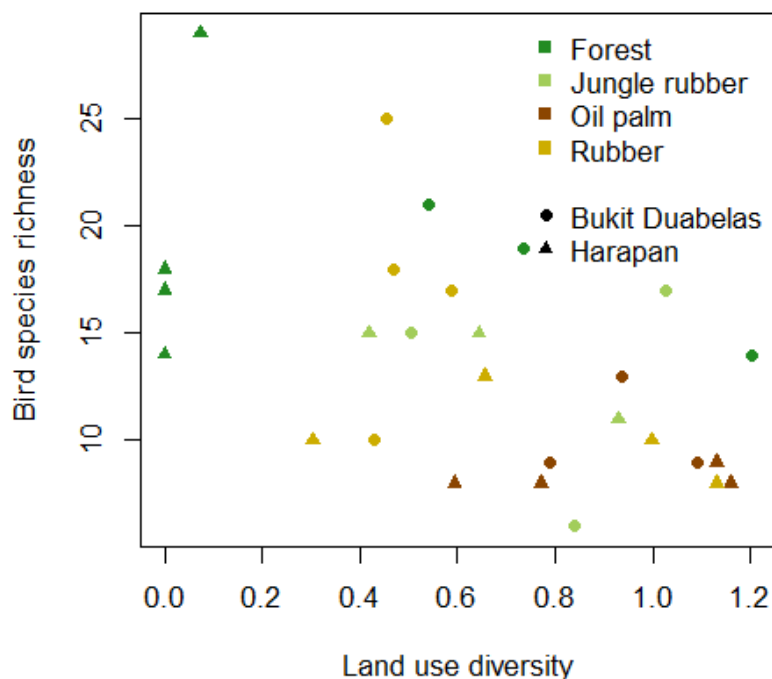


Fig. 7: Scatterplot for bird species richness and land use diversity. (own illustration)

### 4.3 Model construction and selection

After excluding missing values, a dataset with 21 potential predictors each consisting of 27 observations is produced that is then analyzed for regression relationships to the target variable avian species richness.

Tab. 2: Model diagnostics overview, with insect = insectivores; omni = omnivores; alt = alternative -diet species; unkn = unknown diet species. (own illustration)

	model ID							
	1	2	3	4a	4b	4c	4d	4
RMSE	3.5410	2.0253	4.1093	1.9839	1.2150	0.7307	1.6953	3.4634
RMSE %	26.08	15.74	28.25	39.68	30.09	65.51	49.53	25.51
adj. R <sup>2</sup>	0.513	0.838	0.319	0.217	0.307	0.434	0.550	-
deviance explained (%)	57.2	86.8	39.8	29.6	40.2	36.8	54.2	-
no. of predictors	4	2	2	2	3	2	2	7
response data	total all	total HR	total BD	insect all	omni all	alt all	unkn all	total all
sample size (n)	27	15	11	27	27	27	27	27

The outcome of the described methodology are 8 models in total of which 3 models have the target variable total bird species richness (see Tab. 2). The adjusted  $R^2$  of the models range from 0.217 to 0.838 and the relative RMSE ranges from 15.74 % to 65.51 % with a number of predictors included in the models from 2 to 7.

The best fit model, following adjusted  $R^2$ , deviance explained and RMSE is regression model 2 that describes the relationship between total bird species richness and ant species richness as well as effective number of layers for the research landscape HR. In contrast to the model describing bird species richness in BD which has a far worse fit.

The diet separated bird species richness models have variable fits for different diet groups that range from an adjusted  $R^2$  of 0.217 for insectivores to 0.55 for unknown diet bird species. The relative RMSE indicates that the model estimating the number of species of omnivores shows the lowest error between the field observations and the modelled species richness, while bird species richness of alternative diets is modelled with the highest relative average difference between field observation and model output.

Model 1 and 4 both model the total bird species richness including all data, not separated for research landscapes. While model 1 estimates the total bird species richness directly by the predictor variables, model 4 consists of the sum of all diet-separated models. Both models perform similar in terms of fit, while model 4 that incorporates information on the diet of a species, that is not considered in model 1, shows a slightly better RMSE. However, this may be attributed to the larger number of predictor variables included in the regression model of 7, versus 4 in model 1.

The best fit model for total bird species richness with all data available not separated for diet groups (model 1), has an adj.  $R^2$  of 0.513 and a relative RMSE of 26.08 % with 4 predictors.

The most important predictor variables are, in descending order of inclusions in models, land use diversity, beetle species richness, ant species richness and stand structural complexity (see Tab. 3). Canopy openness, effective number of layers, invertebrate species richness, elevation, plant species richness and chilopoda species richness were selected in one model as predictor. The rest of the potential predictors (Tab. 1) is not included in any of the constructed regression models.

Some coefficient estimates for predictors included in the constructed regression models do not exhibit statistically significant differences from 0, hence their  $p$ -value is higher than a given threshold. This is true for example in the case of canopy openness in model

1. Nevertheless, the non-significant predictors are kept in the model because of the potential omitted-variable bias, that could occur if the predictor would be excluded.

Tab. 3: Model coefficient estimates with significance codes indicating thresholds for  $p$ -values: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '.' 1. (own illustration)

variable	model ID						
	1	2	3	4a	4b	4c	4d
(intercept)	2.078***	1.252***	2.679***	1.280***	1.206*	0.919	0.697*
<i>ludiv</i>	- 0.476**	-	- 0.388	- 0.373	- 0.599*	-	-
<i>beetle</i>	0.0271*	-	0.0331*	-	-	-	0.093***
<i>ant</i>	0.010 .	0.018*	-	-	-	-	-
<i>ssc</i>	-	-	-	-	- 0.109	- 0.429*	-
<i>canop</i>	2.201	-	-	-	-	-	-
<i>enl</i>	-	0.052*	-	-	-	-	-
<i>inv</i>	-	-	-	0.008*	-	-	-
<i>elev</i>	-	-	-	-	0.018*	-	-
<i>plant</i>	-	-	-	-	-	0.010**	-
<i>chil</i>	-	-	-	-	-	-	- 0.162

The model selected for further implementation from the set of models is model 1 due to the model fit statistics, a low number of included predictor variables and the target variable of overall bird species richness (see Tab. 2):

$$birdRichness = e^{2.781 + 0.0271*beetle + 0.0104*ant + 2.2019*canop - 0.4762*ludiv}$$

with

*birdRichness* = total number of bird species,

*beetle* = beetle species richness,

*ant* = ant species richness,

*ludiv* = neighborhood landcover diversity (2300 m around plot),

*canop* = canopy openness (%).

A land use-grouped scatterplot visualizing the relationship between observed bird species richness and modelled bird species richness according to the regression model selected is suitable to gain inference about the model performance in different land use systems (see Fig. 8). The optimal model that has the best fit possible would produce observation points on the dashed 45° line, points above the dashed line indicate model underestimation, points below indicate model overestimation.

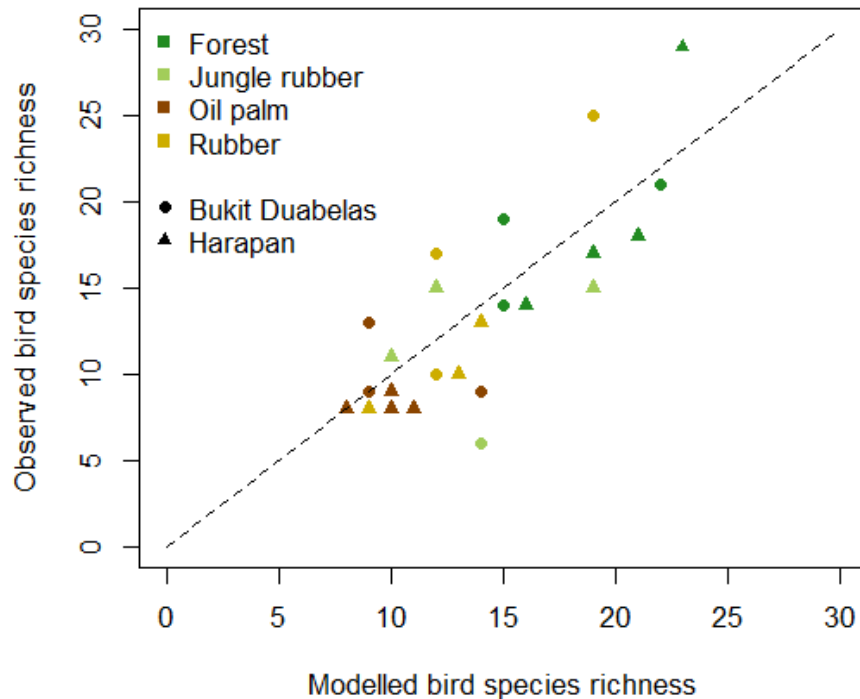


Fig. 8: Model diagnostic for model 1 with field-observed bird species richness (y) plotted against modelled bird species richness (x). (own illustration)

The average error of the model (RMSE) to the according field observation is 3.541 which represents a relative error of  $\pm 26.08\%$  normalized to the overall mean bird species richness. The model overestimates bird species richness in the majority of instances. Larger errors occur for all land use systems and the errors are distributed equally over the whole range of bird species richness modelled.

#### 4.4 Model implementation

The constructed statistical regression model is incorporated in the already existing EF-ForTS-ABM in NetLogo. NetLogo is chosen as a programming platform since it facilitates the implementation of spatially explicit agent-based systems. The inclusion of the bird richness model widens the scope of the ABM, including then another ecological facet, besides a carbon model. The predictor variables beetle species richness, ant species richness, canopy openness and land use diversity are simulated over the model landscape to generate observations of the independent variables in each cell that feed into the bird richness model in a subsequent step.

Beetle and ant species richness as well as canopy openness are simulated in each cell by random weighted sampling from a probability distribution constructed from the smoothed field data on the *core plots*.

Spillover effect of ants was found in the area, indicating that at interfaces between land use systems an increased amount of species may be found (NURDIANSYAH 2016). That is why a spillover effect for beetle and ant species is added to the model that increases the number of species found close to forest patches. All land use cells that are not forest, but border forest on at least one side are affected by the spillover effect. The strength of the spillover effect can be adjusted from 0 to + 50 % of the ant species richness and 0 to + 10 % of the beetle species richness. The spillover effect can be switched off entirely.

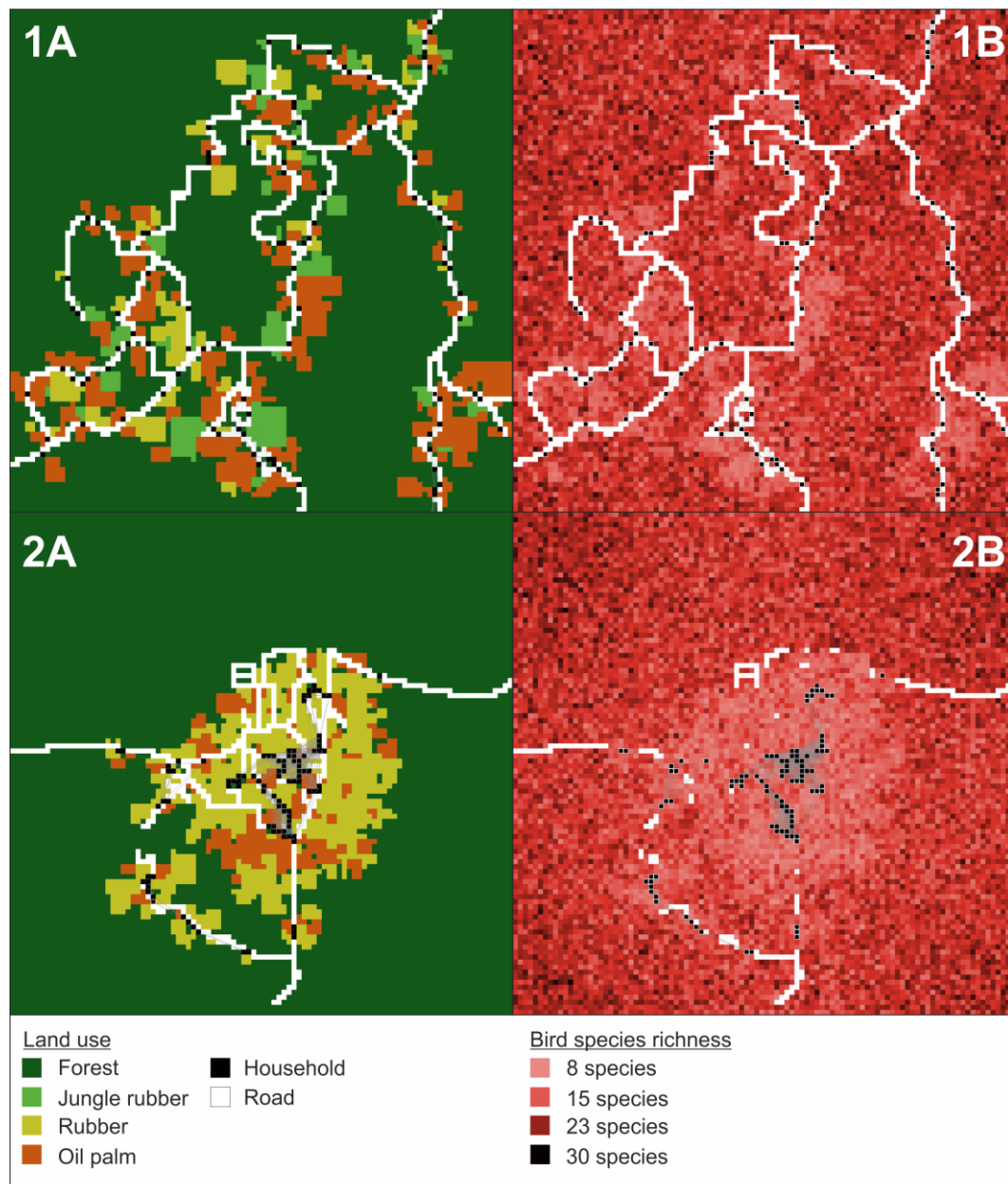


Fig. 9: NetLogo landscape model output with land use (A) and corresponding cell bird species richness (B) for two different land use patterns (1 and 2). (own illustration)

The overall mean cell bird species richness for model output 1 is 13.81 (model output 2: 15.67). Forest shows the highest mean cell bird species richness with 15.66 (model output

2: 16.99), jungle rubber shows the second highest mean cell bird species richness with 14.43. Rubber land use cells host an average of 12.82 (model output 2: 12.69), bird species and oil palm cells have the lowest mean cell bird species richness of 11.34 (model output 2: 11.27). Forest and jungle rubber show equally high variability (standard deviation) of cell bird species richness followed by rubber with intermediate variability and oil palm with the lowest variability of cell bird species richness.

Bird species richness differences in the modelled landscape output may be seen in the different land use systems, especially between forest and oil palm. Lower average bird species richness can be found in oil palm and rubber cells, while forest hosts a larger average number of bird species per cell. Additionally, higher cell bird species richness may be found in remote forest areas (see corners of 2B of Fig. 9) that are affected by the low land use diversity in the surrounding of the areas. Stochastic noise can be seen in the model output indicating the variability inside the land use systems (e.g. differences in bird species richness between forest cells).

#### **4.5 Model validation**

The comparison of the sampled model output to a validation dataset reveals the model performance. The comparison of sampled model output to model input indicates how well the model fits the data used for construction, which gives inference about how well the model implementation methodology is suited. Land use configuration in the artificial landscape influences the bird species richness estimation due to the fact that land use diversity in a certain cell neighborhood affects the cell bird species richness. Therefore, two different land use configuration scenarios are investigated and their output is compared to the validation dataset and to the model input.

Initially, the independent validation dataset needs to be compared to the model input dataset to gain information about differences and similarities to be able to correctly interpret why there are differences between the model output and the independent validation dataset. The comparison between independent dataset and model input dataset indicates statistically significant differences between the mean number of bird species only for oil palm, because the constructed confidence intervals do not overlap. The differences in mean bird species richness in all other land use systems can be explained by a low sampling effort or no real difference in the communities represented by the samples. The variability between the observations is higher in the validation dataset for forest and oil palm and lower in the validation dataset for rubber and jungle rubber.

Tab. 4: Comparison of descriptive statistics for model validation grouped by land use type (\* = shrub-bushland in validation data). (own illustration)

		model output 1 (n=28)	model output 2 (n=28)	validation data (n=28)	model input (n=8)
Forest	mean	15.66	16.99	17.39	18.86
	confidence interval	14.54 – 16.78	15.65 – 18.33	15.31 – 19.47	14.98 – 22.74
	standard deviation	3.05	3.44	5.50	5.15
Oil palm	mean	11.34	11.27	12.18	9.14
	confidence interval	10.72 – 11.96	10.71 – 11.83	10.92 – 13.44	7.80 – 10.48
	standard deviation	1.55	1.57	3.33	1.77
Rubber	mean	12.82	12.69	12.79	13.29
	confidence interval	11.86 – 13.78	11.75 – 13.63	11.15 – 14.61	8.81 – 17.77
	standard deviation	2.34	2.23	4.80	5.90
Jungle rub- ber*	mean	14.43	-	17.25	12.80
	confidence interval	13.15 – 15.71	-	15.75 – 18.75	8.88 – 16.72
	standard deviation	3.39	-	3.98	4.38

The mean of the sampled model output 1 (see Tab. 4) corresponding to landscape 1 of Fig. 9 compared to the validation dataset indicates small differences for oil palm and rubber land use systems. Mean cell bird species richness for forest and jungle rubber are underestimated in the model compared to the validation dataset by 1.73 and 2.82 species. The variability (standard deviation) of bird species richness is underestimated by the model in all land use systems presenting the highest difference of 2.46 in rubber and the lowest difference in jungle rubber of 0.59. Considering the standard errors of the estimates in the model validation dataset, construction of confidence intervals of certainty level of  $\alpha \approx 0.95$  around model input mean and model validation mean indicate no significant differences between model input and model validation data for forest, oil palm and rubber. Only the confidence intervals of jungle rubber do not overlap indicating significant differences that could also be accounted to the fact that the validation dataset for jungle rubber was sampled in a land use type called “shrub-bushland” that may differ to the jungle rubber observations collected and used as model input.

Comparison of the sampled model output 1 to the model input dataset indicates an average bird species richness overestimation of the model for oil palm and jungle rubber and an underestimation for forest and rubber. The variability of bird species richness (standard deviation) is underestimated in all cases, with a larger underestimation in rubber and forest, a smaller underestimation in jungle rubber and close to zero difference in oil palm.



Considering the confidence intervals of certainty level of  $\alpha \simeq 0.95$  around model input mean and model output mean indicate no significant differences between model input and model output for forest, rubber and jungle rubber. Only the confidence intervals of oil palm do not overlap indicating significant differences.

The sampled model output 2 (see Tab. 4) corresponding to landscape 2 of Fig. 9 shows similar relationships to the validation dataset and to the model input dataset as model output 1. Differences are that no jungle rubber land use cells exist and that the average forest cell bird species richness is increased but still not statistically significantly different to the validation or model input dataset.

## 5 Discussion

### 5.1 Results against the background of the research objective

Statistical modelling of bird species diversity in the highly complex ecosystems of the tropics under continuing land use and linked habitat change by regression analysis as performed in the present study and integrated in the EFForTS-ABM supports the search for synergy and trade-off effects between ecological and socioeconomic functions in the land use transformation systems of Indonesia.

The results show that regression modelling of bird species richness by habitat and biodiversity of other taxa is possible even though sample size is low. Additionally, the model output shows no significant differences to an independent validation dataset for land use systems forest, rubber and oil palm. The research objective of the present investigation is to construct a statistical regression model, implement the model in NetLogo, validate the model output and assess model behavior. For the evaluation of this objective, four research questions were formulated that shall be answered:

- i. *Which environmental, habitat and biodiversity variables determine bird species richness in the land use transformation systems of Indonesia?*

Beetle species richness, plant species richness, biomass carbon, land use intensity and land use diversity in the surrounding of 2300 m show the strongest correlations with bird species richness and can be considered to have a relationship to bird species richness. However, a causal effect cannot undoubtedly be associated to these relations. The habitat describing proxies plant species richness, biomass carbon, land use diversity and land use intensity may affect which niches are available at a given site, while beetle species richness may be an indicator for diversity of forage for birds. Environmental variables, like air humidity or temperature only show weak correlation to bird species richness, indicating that variables describing the small scale structure of the habitat may be more important factors determining bird species  $\alpha$ -diversity than environmental variables describing the climate, which is supported by the statements of TELLERÍA *et al.* (1992).

Considering the regression analysis, most important predictor variables for bird species richness remain habitat and biodiversity variables like land use diversity, beetle species richness, ant species richness and stand structural diversity. With respect to different sub areas of the research region and different target variables

- diet groups of bird species - the importance of habitat structure describing variables and forage diversity proxies persists.

- ii. *How can environmental, habitat and biodiversity predictor variables be combined into a statistical regression model to predict bird species richness?*

For the prediction of bird species richness a GLM of Poisson distribution with a log-link may be used following partly the methodologies of KOH (2008) and LUOTO *et al.* (2004). The best fit model that is most suitable to predict bird species richness in the considered environment uses habitat describing variables (canopy openness and surrounding land use diversity) as well as biodiversity measures of other taxa (beetle and ant species richness) as independent variables. These predictor variables are combined linearly without interacting effects.

Including additional information into the regression analysis, like species diet information, for the separation of total bird species richness into multiple individually modelled subparts of the overall species richness improves the model fit only marginally while increasing the number of predictors. Hence, the direct modelling of total bird species richness remains the best option.

Modelling of bird species richness by GAMs may improve fit and predictive power of the model because the data can be fit more flexibly. Since the main goal of the constructed model is to predict bird species richness, rather than explaining the underlying effects of the distribution of bird species richness, GAMs as “data defined [...] models” (GUISAN *et al.* 2002) may also be a viable option.

- iii. *Can bird species richness be predicted by environmental, habitat and biodiversity variables in a sufficiently precise manner?*

Bird species richness can be predicted by habitat and biodiversity variables with a deviance explained of 57.2 % in the research area. Comparison of the model output to an independent validation dataset yielded only small, mostly statistically insignificant differences.

KOH (2008) established bird species richness models for oil palm plantations in Malaysia that explain 5.7 % and 14.4 % of the variation in bird species richness at a local level and landscape level respectively. LUOTO *et al.* (2004) constructed a bird species richness model using remote sensing predictors that accounts for 60.8 % of the variation in bird species richness in southwestern Finland. Accordingly, the model constructed in the present study exhibits a relatively high

goodness-of-fit compared to other studies modelling bird species richness by regression analysis.

*iv. What is the link between land use and bird species richness?*

The link between land use and bird species richness remains uncertain. Statistically significant differences between the land use types regarding bird species richness were only found between forest and oil palm, while rubber and jungle rubber did not show significant differences between each other and to forest and oil palm. Hence, the sample only indicates that bird species richness is diminished in oil palm plantations compared to forest.

PRABOWO *et al.* (2016) analyzed the same dataset of 32 observations of bird species richness and found additional statistically significant differences for forest and rubber plantation in HR. Considering this finding, evidence exists that total bird species richness is reduced in cash crop plantations of rubber and oil palm compared to tropical lowland rainforest.

Furthermore, the remote sensing analysis and subsequent correlation and regression analysis show that the land use diversity in the surrounding is linked to bird species richness. The strong negative correlation of land use diversity in a surrounding of 2300 m with bird species richness suggests a negative effect. This strong correlation however may be masked by the isolated forest plots in the sample that show high bird species richness and low land use diversity in their surrounding because the sample locations are placed in remote national parks consisting of forest only. A higher land use diversity in the surrounding would implicitly represent an increased proportion of monoculture or agroforestry land, which is assumed to negatively affect the bird species richness. This can be backed up when looking at the correlation between different land use types and bird species richness on the plot (see Fig. 3) indicating a positive correlation for forest and negative correlation for oil palm and jungle rubber. Although the negative correlation of land use diversity in the surrounding is difficult to explain, the inclusion of the variable as an independent variable is improving the goodness-of-fit of the model drastically.

The implemented regression model produces expected outputs regarding the field data, indicating highest per cell bird species richness in forest and lowest bird species richness in oil palm. These findings are also backed up by the validation dataset. Model behavior

in different landscapes, representing varying land use composition and configuration also produce expected model outputs.

A strong link between land use and bird species richness exists for the land use diversity considering a circular surrounding of radius 2300 m. As mentioned before, this relationship may be an artefact of the isolated forest sampling plots, how the plots are located in the area and the uncertainties of the analyzed land use map. Why a surrounding of 2300 m has the largest correlation to bird species richness is difficult to explain. This specific distance could be linked to the travel distance of the birds. The distance could also be an artefact of the land use map and sample site selection.

## 5.2 Results against the background of other relevant studies

The outcome of the regression model can be assessed against the background of other recent studies investigating species diversity by regression analysis to derive differences and similarities.

KOH (2008) found that in Malaysia's oil palm plantations "the most important predictors of species richness [...] were [...] percentage cover of young secondary forests surrounding an estate for birds". This evidence can also be found in Jambi since on a small scale as well as on a large scale forest proportion has a strong positive linear correlation with bird species richness on the plot.

The same observations were made for rubber plantations. ZHANG *et al.* (2017) show in a study conducted in China that the extent of natural forest patches at landscape level is linked to bird species richness in rubber plantations. This finding corresponds to the high correlation of bird species richness in the analyzed data of the present study with forest proportion in the larger scale surrounding of the plot, indicating that the correlation between bird species richness and land use diversity is indeed caused by high forest proportions associated with high bird species richness and low surrounding land use diversity.

Regarding the goodness-of-fit, the constructed model can also be compared to other studies of different taxa. OVERTON *et al.* (2009) modelled snail species richness in New Zealand, based on 2330 community surveys and a range of environmental and vegetation variables. The species richness model was developed using a GAM with 14 predictor variables, explaining 27 % of variation in snail richness. Even though the sample size is larger, the degree of variation explained indicates that the goodness-of-fit of the regression model constructed in the present study is high compared to other taxa, like snails.

In a study on the species richness of forest passerines on the Iberian Peninsula by TELLERÍA *et al.* (1992), only tree density between several climatic and habitat variables investigated, was correlated with avian species richness. A comparable result may be found in the present study, since habitat describing variables, like plant diversity, canopy openness or tree biomass were also correlated to bird species richness.

In the present study, NDVI statistics have no strong correlation to bird species richness and were not included in the bird species richness model. MCFARLAND *et al.* (2012) conducted a study on the regression model based prediction of bird species richness by remotely sensed NDVI statistics in Arizona, USA, and come to the conclusion that the variability explained even by the best constructed regression model is very low. They conclude that NDVI statistics are not able to predict bird species richness in certain environments, which can be backed up by the findings in this study.

### 5.3 Methodological limitations

Considering the applied methodology, some limitations exist need to be kept in mind. For instance, stepwise variable selection procedures are considered to be high variance operations because small perturbations of the response data can sometimes lead to vastly different subsets of the variables (GUISAN AND ZIMMERMANN 2000). Therefore, the presented approach is only viable if the goal is to construct predicting regression models rather than explaining models, because the inference can be vastly different although the analyzed data differs only marginally.

*Pearson's correlation coefficient* and GLMs do not account for non-linear correlation between dependent and independent variables. Potential strong non-linear correlations that may improve goodness-of-fit of the regression model if implemented in a GAM are missed in the presented methodology. Consequently, GAMs, which provide the possibility to automatically estimate non-linear relationships between random variables and fit a generalized regression model may improve model fit and the power to estimate bird species richness.

The implemented parametrization of the spillover effect, i.e. the extent of spillover, into other land use systems cannot be backed up by specific empirical findings, only the presence of a spillover effect is based on empirical observations after NURDIANSYAH (2016). Consequently, the spillover effect may be parametrized incorrectly and can be turned off or adjusted in its extent.

Furthermore, it is important to develop a theoretical or conceptual model of the studied system that can explain the results seen in the statistical model, or better, to have a conceptual model before the statistical modelling starts that can be evaluated with the statistical model (CALE *et al.* 1983). This approach is most important if the model is used to explain relationships and effects rather than predicting the dependent variable from an independent variable. The “data-driven” approach that is applied in the present study is not able to reliably identify effects of independent variables on bird species richness and should therefore not be used for that task.

#### 5.4 Model and data limitations

Regarding the data and the constructed model, limitations exist that need to be kept in mind when using the model and inferring from the model and model output.

The model at hand is not suited to model the overall bird species richness that is present at a given location because of the mentioned restrictions of sampling and the correlation of sampling effort with bird individuals and species observed. Consequently, the model should only be used for the comparison between sites and considered to be constructed for areas of 2500 m<sup>2</sup> only with a given sampling effort.

Another constraint of the model is that it deals with anonymous species. The species are not modelled individually and therefore it is not included in the model which species is present at a given location, only how many different species. This approach has the downside that neither  $\beta$ - nor  $\gamma$ -diversity can be assessed. Also, functional diversity associated to specific bird species cannot be assessed due to the anonymous species modelling technique.

The experimental sampling design with sample sites intended to be replications in the different land use systems is not optimal for the use of the gathered data in regression analysis. An experimental design is used if the effect of a categorized variable, or at least discrete levels of a treatment, on a target variable is assessed. In the present case, the effect of land use system on a variety of target variables is investigated, which requires multiple replications in the different land use systems. Therefore, *core plots* are selected so that the conditions, i.e. the land use system, varies very little between the replications. For a regression analysis, as done in the present study, capturing as much variability of the target as well as the predictor variable is crucial to be able to construct valid regression

relationships. Consequently, the sampling design would have been different if the data was intended to be used in regression analysis only.

The analyzed dataset may be one restriction that needs to be kept in mind when using the model and interpreting results. This applies to all input data which include field observed data on the 32 *core plots* as well as remotely sensed data from the land use map. Since the data was collected by different people with a variety of methods, the assessment of data quality is also difficult. Additional data restrictions are also present for the validation dataset since the detection range, the distance at which an individual that is in the dataset was detected, could not always be recorded. Therefore, the size of the sample plot on which birds were observed in the sample dataset is not precisely defined and this may have effects on the comparability between model output and validation data. Nevertheless, validating the model output at all enables a rough estimation of the model performance, even though the validation data does not perfectly fit model input data.

Additionally, the analyzed land use map produced by MELATI (2017) poses another potential limitation to model quality. Although the overall accuracy is 81.8 %, indicating that the presented share of pixels in the map are classified correctly, potential misclassifications around the plots may lead to biased correlation estimates between surrounding land use and bird species richness.

## 5.5 Conclusions and outlook

The raw number of bird species found at a location does not provide accurate and comprehensive information about the diversity of the given taxon. Individual abundance, evenness or variability of species composition between sites ( $\beta$ -diversity) need to be considered also to gain full information about the present biodiversity.

When investigating ecosystem functions, functional diversity as an indicator can be considered since two different species may perform the same ecosystem function, while another species is the only species responsible for a variety of ecosystem functions (BARNES *et al.* 2014). Therefore, decreased bird species richness must not always result in diminished ecosystem functioning since there may be multiple species occupying the same functional role in a habitat. Nevertheless, BARNES *et al.* (2014) discovered that reductions in ecosystem functions exceed the loss of species diversity, which may be explained stating that a single species provides multiple ecosystem function.



The model constructed in the present study may be improved by some additions to the implementation in NetLogo. This includes the data-based parametrization of spillover effects and the elaboration of improved techniques of predictor simulation over the landscape. It may be assumed that the predictor variables show a trend over the lifecycle of a plantation that differs in oil palm, rubber and jungle rubber. This relation may be used to construct deterministic relationships in form of regression models between beetle and ant species richness, as well as canopy cover with plantation age.

A literature review shows that some predictors correlate with plantation age. Canopy cover positively correlates with oil palm plantation age (WANG AND FOSTER 2016). This is also backed up by data collected in Malaysia (KOH 2008) that may be used to derive a simple exponential regression model to determine the relationship between oil palm plantation age and canopy cover. Using this deterministic regression model, the oil palm canopy cover can be estimated using plantation age as a predictor. In addition, WANG AND FOSTER (2016) observe that in contrast to compositional diversity, the species richness of ground-foraging ants does not vary with increasing oil palm plantation age. This relationship can also be backed up by the field data used in the present study indicating no potential use of a deterministic model to estimate ant species richness as a function of plantation age in the simulation of that predictor.

The model constructed in the present study may be used to identify synergy and trade-off effects for avian  $\alpha$ -diversity and other ecological and socioeconomic ecosystem functions to secure human livelihoods as well as biodiversity. Keeping the stated restrictions and limitations in mind, the model implementation may be used to pinpoint areas of high bird species richness indicating areas of conservation priority and investigating different future land use scenarios. Ultimately, the presented findings contribute to the discussion on avian biodiversity in a changing world.

## 6 Literature

- Akaike, H. (1974). New look at statistical model identification. *IEEE Transactions on Automatic Control*, *AC19*(6), 716–723.
- Akimoto, H. (2003). Global air quality and pollution. *Science*, *302*(5651), 1716–1719.
- Andaya, B. W. (1993). To live as brothers: Southeast Sumatra in the seventeenth and eighteenth centuries. Honolulu, Hawaii: University of Hawaii Press.
- Barnes, A. D.; Jochum, M.; Mumme, S.; Haneda, N. F.; Farajallah, A.; Widarto, T. H. and Brose, U. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nature communications*, *5*, 5351.
- Brauner, N. and Shacham, M. (1998). Role of range and precision of the independent variable in regression of data. *AIChE Journal*, *44*(3), 603–611.
- Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, *65*(3), 625–633.
- Cale, W. G.; Oneill, R. V. and Shugart, H. H. (1983). Developments and applications of desirable ecological models. *Ecological Modelling*, *18*(3-4), 171–186.
- Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, *11*(4), 265–270.
- Chatterjee, S. and Hadi, A. S. (2006). Regression analysis by example. Hoboken, NJ, Wiley-Interscience.
- Darras, K. (2016). Bird Diversity, Functions and Services across Indonesian Land-use Systems. *Dissertation*, Georg-August Universität, Göttingen
- Dirzo, R. and Raven, P. H. (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, *28*(1), 137–167.
- Dislich, C.; Hettig, E.; Salecker, J.; Heinonen, J.; Lay, J.; Meyer, K. M.; Wiegand, K. and 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*(1).
- 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.; Mejjide, A.; Nurdiansyah, F.; Otten, F.; Pe'er, G.; Steinebach, S.; Tarigan, S.; Tölle, M. H.; Tschardtke, T. and Wiegand, K. (2017). A review of the ecosystem functions

in oil palm plantations, using forests as a reference system. *Biological reviews of the Cambridge Philosophical Society*, 92(3), 1539–1569.

- Drescher, J.; Rembold, K.; Allen, K.; Beckschäfer, P.; Buchori, D.; Clough, Y.; Faust, H.; Fauzi, A. M.; Gunawan, D.; Hertel, D.; Irawan, B.; Jaya, I. N. S.; Klarner, B.; Kleinn, C.; Knohl, A.; Kotowska, M. M.; Krashevska, V.; Krishna, V.; Leuschner, C.; Lorenz, W.; Mejjide, A.; Melati, D.; Nomura, M.; Pérez-Cruzado, C.; Qaim, M.; Siregar, I. Z.; Steinebach, S.; Tjoa, A.; Tschardtke, T.; Wick, B.; Wiegand, K.; Kreft, H. and Scheu, S. (2016). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1694).
- Ehbrecht, M.; Schall, P.; Ammer, C. and Seidel, D. (2017). Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate. *Agricultural and Forest Meteorology*, 242, 1–9.
- Ehbrecht, M.; Schall, P.; Juchheim, J.; Ammer, C. and Seidel, D. (2016). Effective number of layers: A new measure for quantifying three-dimensional stand structure based on sampling with terrestrial LiDAR. *Forest Ecology and Management*, 380, 212–223.
- Elmhirst, R. (2011). Migrant pathways to resource access in Lampung's political forest: Gender, citizenship and creative conjugality. *Geoforum*, 42(2), 173–183.
- Fahrmeir, L.; Kneib, T.; Lang, S. and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer.
- Gatto, M.; Wollni, M. and 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.
- Gibbs, H. K.; Ruesch, A. S.; Achard, F.; Clayton, M. K.; Holmgren, P.; Ramankutty, N. and Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16732–16737.
- Guisan, A.; Edwards, T. C. and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2-3), 89-100.

- Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), 147–186.
- Hildebrandt, A. (2015). Diet and functional diversity of birds in different rainforest transformation systems in Sumatra. *Master thesis*, Georg-August Universität, Göttingen.
- Houghton, R. A. (2003). Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. *Tellus B: Chemical and Physical Meteorology*, 55(2), 378–390.
- Jones, G. W. (2013). *The Population of Southeast Asia*. National University of Singapore, Singapore.
- Karp, D. S.; Mendenhall, C. D.; Sandí, R. F.; Chaumont, N.; Ehrlich, P. R.; Hadly, E. A. and Daily, G. C. (2013). Forest bolsters bird abundance, pest control and coffee yield. *Ecology Letters*, 16(11), 1339–1347.
- Kathirithamby-Wells, J. (1993). Hulu-hilir Unity and Conflict: Malay Statecraft in East Sumatra before the Mid-Nineteenth Century. *Archipel*, 45(1), 77–96.
- Koh, L. P. (2008). Can oil palm plantations be made more hospitable for forest butterflies and birds? *Journal of Applied Ecology*, 45(4), 1002–1009.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260), 583–621.
- Larson, D. F. (1996). Indonesia's palm oil subsector. The World Bank No. WPS1654.
- Lehmann, A.; Leathwick, J. R. and Overton, J. M. (2002). Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, 11(12), 2217–2238.
- Lehmann, A.; Overton, J. M. and Leathwick, J. R. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 157(2-3), 189–207.
- Luoto, M.; Virkkala, R.; Heikkinen Risto K. and Rainio, K. (2004). Predicting Bird Species Richness Using Remote Sensing in Boreal Agricultural-Forest Mosaics. *Ecological Applications*, 14(6).
- Maas, B.; Clough, Y. and Tschardtke, T. (2013). Bats and birds increase crop yield in tropical agroforestry landscapes. *Ecology Letters*, 16(12), 1480–1487.
- Maas, B.; Karp, D. S.; Bumrungsri, S.; Darras, K.; Gonthier, D.; Huang, J. C.-C.; Lindell, C. A.; Maine, J. J.; Mestre, L.; Michel, N. L.; Morrison, E. B.; Perfecto, I.; Philpott, S. M.; Şekercioğlu, Ç. H.; Silva, R. M.; Taylor, P. J.; Tschardtke, T.; van Bael,

- S. A.; Whelan, C. J. and Williams-Guillén, K. (2016). Bird and bat predation services in tropical forests and agroforestry landscapes. *Biological reviews of the Cambridge Philosophical Society*, 91(4), 1081–1101.
- MacArthur, R. H. and MacArthur, J. W. (1961). On Bird Species Diversity. *Ecology*, 42(3), 594–598.
- Magurran, A. E. (2011). Measuring biological diversity. Malden, Mass., Blackwell.
- Margono, B. A.; Potapov, P. V.; Turubanova, S.; Stolle, F. and Hansen, M. C. (2014). Primary forest cover loss in Indonesia over 2000–2012. *Nature Climate Change*, 4(8), 730.
- Mcfarland, T. M.; van Riper, C. and Johnson, G. E. (2012). Evaluation of NDVI to assess avian abundance and richness along the upper San Pedro River. *Journal of Arid Environments*, 77, 45–53.
- 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.
- Myers, N.; Mittermeier, R. A.; Mittermeier, C. G.; da Fonseca, G. A. and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858.
- Nemenyi, P. B. (1963). Distribution-free multiple comparisons. *Dissertation*, Princeton University, New Jersey.
- Nieto, S.; Flombaum, P. and Garbulsky, M. F. (2015). Can temporal and spatial NDVI predict regional bird-species richness? *Global Ecology and Conservation*, 3, 729–735.
- Nurdiansyah, F. (2016). Local and Landscape Management of Biological Pest Control in Oil Palm Plantations. *Dissertation*, Georg-August Universität, Göttingen.
- Overton, J. M.; Barker, G. M. and Price, R. (2009). Estimating and conserving patterns of invertebrate diversity: a test case of New Zealand land snails. *Diversity and Distributions*, 15(5), 731–741.
- Pielke, R. A. (2005). Atmospheric science. Land use and climate change. *Science* 310(5754), 1625–1626.
- Pimm, S. L.; Russell, G. J.; Gittleman, J. L. and Brooks, T. M. (1995). The future of biodiversity. *Science (New York, N.Y.)*, 269(5222), 347–350.

- Prabowo, W. E.; Darras, K.; Clough, Y.; Toledo-Hernandez, M.; Arlettaz, R.; Mulyani, Y. A. and Tschardtke, T. (2016). Bird Responses to Lowland Rainforest Conversion in Sumatran Smallholder Landscapes, Indonesia. *PLOS ONE*, 11(5), e0154876.
- R Development Core Team. (2018). R: A language and environment for statistical computing. Vienna.
- Rion, V. (2010). Modelling the plant species richness: a comparison between two approaches. *Master thesis*, Université de Lausanne, Lausanne
- Rosenzweig, M. L. (2002). Species diversity in space and time. Cambridge Univ. Press.
- Sahner, J. (2016). Impact of Rain Forest Transformation on Roots and Functional Diversity of Root-Associated Fungal Communities. *Dissertation*, Georg-August Universität, Göttingen.
- Sala, O. E.; Chapin, F. S. 3.; Armesto, J. J.; Berlow, E.; Bloomfield, J.; Dirzo, R.; Huber-Sanwald, E.; Huenneke, L. F.; Jackson, R. B.; Kinzig, A.; Leemans, R.; Lodge, D. M.; Mooney, H. A.; Oesterheld, M.; Poff, N. L.; Sykes, M. T.; Walker, B. H.; Walker, M. and Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459), 1770–1774.
- Shannon, C. E. and Weaver, W. (1949). The mathematical theory of communication. University of Illinois Press.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Boston, MA, Springer US.
- Simpson, E. H. (1949). Measurement of Diversity. *Nature*, 163, 688.
- Smith, E. P. and van Belle, G. (1984). Nonparametric Estimation of Species Richness. *Biometrics*, 40(1), 119–129.
- Sodhi, N. S.; Posa, M. R. C.; Lee, T. M.; Bickford, D.; Koh, L. P. and Brook, B. W. (2010). The state and conservation of Southeast Asian biodiversity. *Biodiversity and Conservation*, 19(2), 317–328.
- Suyanto, S.; Ruchiat, Y.; Stolle, F. and Applegate, G. (2000). The underlying causes and impacts of fires in South-east Asia.: Site3.Tanah Tumbuh, Jambi Province, Indonesia. Jakarta, Indonesia.
- Tellería, J. L.; Santos, T.; Sánchez, A. and Galarza, A. (1992). Habitat structure predicts bird diversity distribution in Iberian forests better than climate. *Bird Study*, 39(1), 63–68.

- United Nations. (1992). Convention on Biological Diversity.
- VanderWeele, T. J. and Shpitser, I. (2013). On the definition of a confounder. *The Annals of Statistics*, 41(1), 196–220.
- Wang, W. Y. and Foster, W. A. (2016). Ground-foraging ant communities vary with oil palm age. *Basic and Applied Ecology*, 17(1), 21–32.
- Wilensky, U. (1999). NetLogo. Evanston, IL.
- Zhang, M.; Chang, C. and Quan, R. (2017). Natural forest at landscape scale is most important for bird conservation in rubber plantation. *Biological Conservation*, 210, 243–252.
- Zhao, S.; Peng, C.; Jiang, H.; Tian, D.; Lei, X. and Zhou, X. (2006). Land use change in Asia and the ecological consequences. *Ecological Research*, 21(6), 890–896.
- Zurell, D.; Berger, U.; Cabral, J. S.; Jeltsch, F.; Meynard, C. N.; Münkemüller, T.; Nehrbaas, N.; Pagel, J.; Reineking, B.; Schröder, B. and Grimm, V. (2010). The virtual ecologist approach: simulating data and observers. *Oikos*, 119(4), 622–635.

## 7 Appendix

### 7.1 Running the model implementation in NetLogo

The bird species richness model implementation is realized in NetLogo 5.3.1. and based upon the EFForTS-ABM that is available as supplementary material from DISLICH *et al.* (2018). Additional information on the EFForTS-ABM and how the economic and ecological sub-models function can also be read in DISLICH *et al.* (2018).

The EFForTS-ABM must be extended for running the bird species richness model. Therefore, the model code nls-file must be included, code must be added to the EFForTS-ABM main code, predictor distribution csv-files must be added and the user interface has to be expanded. The following instructions guide through the process of implementing the model based on the readily available EFForTS-ABM in NetLogo 5.3.1 and indicate how the model can be used.

#### Inclusion of bird species richness submodel

The model code, available as *birdRichness\_submodel.nls*, is copied from the supplementary material of the present study to the directory */scr* of the EFForTS-ABM directory. `__includes ["scr/birdRichness_submodel.nls"]` has to be included in the main code to import the model code.

#### Inclusion of predictor distributions

The smoothed empirical predictor distributions from the field data are available as *values\_predictor\_landuse.csv* and *weights\_predictor\_landuse.csv* where *predictor* is one of *antRichness*, *beetleRichness* or *canopyOpenness* and *landuse* is one of *forest*, *oilpalm*, *junglerubber* or *rubber*. These 24 files are copied from the supplementary material of the present study to the directory */input/bird/predictor\_distributions* of the EFForTS-ABM directory. The inclusion into the model is done in the main code (see below).

#### New code to be added in the EFForTS-ABM main code

An additional extension needs to be included by `extensions [csv]`.

Global variables are added that store mean cell bird species richness and empirical smoothed distributions of predictor variables from the data:

```
globals [
  forest_bird_richness
  oilpalm_bird_richness
  rubber_bird_richness
  junglerubber_bird_richness
```



```

overall_bird_richness

values_anRichness_forest
weights_anRichness_forest
values_beetleRichness_forest
weights_beetleRichness_forest
values_canopyOpenness_forest
weights_canopyOpenness_forest

values_anRichness_oilpalm
weights_anRichness_oilpalm
values_beetleRichness_oilpalm
weights_beetleRichness_oilpalm
values_canopyOpenness_oilpalm
weights_canopyOpenness_oilpalm

values_anRichness_junglerubber
weights_anRichness_junglerubber
values_beetleRichness_junglerubber
weights_beetleRichness_junglerubber
values_canopyOpenness_junglerubber
weights_canopyOpenness_junglerubber

values_anRichness_rubber
weights_anRichness_rubber
values_beetleRichness_rubber
weights_beetleRichness_rubber
values_canopyOpenness_rubber
weights_canopyOpenness_rubber

]

```

Additionally, a patch variable is added, that stores the bird species richness in a given cell:

```

patches-own [
  p_bird_richness
]

```

Furthermore, the modelling procedure starter and inclusion of the predictor distributions is added by:

```

to model-biodiversity

set values_anRichness_forest csv:from-file "/input/bird/predictor_distributions/values_anRichness_forest.csv"
set values_anRichness_forest item 0 values_anRichness_forest
set weights_anRichness_forest csv:from-file "/input/bird/predictor_distributions/weights_anRichness_forest.csv"
set weights_anRichness_forest item 0 weights_anRichness_forest
set values_beetleRichness_forest csv:from-file "/input/bird/predictor_distributions/values_beetleRichness_forest.csv"
set values_beetleRichness_forest item 0 values_beetleRichness_forest
set weights_beetleRichness_forest csv:from-file "/input/bird/predictor_distributions/weights_beetleRichness_forest.csv"
set weights_beetleRichness_forest item 0 weights_beetleRichness_forest
set values_canopyOpenness_forest csv:from-file "/input/bird/predictor_distributions/values_canopyOpenness_forest.csv"
set values_canopyOpenness_forest item 0 values_canopyOpenness_forest
set weights_canopyOpenness_forest csv:from-file "/input/bird/predictor_distributions/weights_canopyOpenness_forest.csv"
set weights_canopyOpenness_forest item 0 weights_canopyOpenness_forest

set values_anRichness_oilpalm csv:from-file "/input/bird/predictor_distributions/values_anRichness_oilpalm.csv"
set values_anRichness_oilpalm item 0 values_anRichness_oilpalm
set weights_anRichness_oilpalm csv:from-file "/input/bird/predictor_distributions/weights_anRichness_oilpalm.csv"
set weights_anRichness_oilpalm item 0 weights_anRichness_oilpalm
set values_beetleRichness_oilpalm csv:from-file "/input/bird/predictor_distributions/values_beetleRichness_oilpalm.csv"
set values_beetleRichness_oilpalm item 0 values_beetleRichness_oilpalm
set weights_beetleRichness_oilpalm csv:from-file "/input/bird/predictor_distributions/weights_beetleRichness_oilpalm.csv"

```

```

set weights_beetleRichness_oilpalm item 0 weights_beetleRichness_oilpalm
set values_canopyOpenness_oilpalm csv:from-file "/input/bird/predictor_distributions/values_canopyOpenness_oilpalm.csv"
set values_canopyOpenness_oilpalm item 0 values_canopyOpenness_oilpalm
set weights_canopyOpenness_oilpalm csv:from-file "/input/bird/predictor_distributions/weights_canopyOpenness_oilpalm.csv"
set weights_canopyOpenness_oilpalm item 0 weights_canopyOpenness_oilpalm

set values_anRichness_junglerubber csv:from-file "/input/bird/predictor_distributions/values_anRichness_junglerubber.csv"
set values_anRichness_junglerubber item 0 values_anRichness_junglerubber
set weights_anRichness_junglerubber csv:from-file "/input/bird/predictor_distributions/weights_anRichness_junglerubber.csv"
set weights_anRichness_junglerubber item 0 weights_anRichness_junglerubber
set values_beetleRichness_junglerubber csv:from-file "/input/bird/predictor_distributions/values_beetleRichness_junglerubber.csv"
set values_beetleRichness_junglerubber item 0 values_beetleRichness_junglerubber
set weights_beetleRichness_junglerubber csv:from-file "/input/bird/predictor_distributions/weights_beetleRichness_junglerubber.csv"
set weights_beetleRichness_junglerubber item 0 weights_beetleRichness_junglerubber
set values_canopyOpenness_junglerubber csv:from-file "/input/bird/predictor_distributions/values_canopyOpenness_junglerubber.csv"
set values_canopyOpenness_junglerubber item 0 values_canopyOpenness_junglerubber
set weights_canopyOpenness_junglerubber csv:from-file "/input/bird/predictor_distributions/weights_canopyOpenness_junglerubber.csv"
set weights_canopyOpenness_junglerubber item 0 weights_canopyOpenness_junglerubber

set values_anRichness_rubber csv:from-file "/input/bird/predictor_distributions/values_anRichness_rubber.csv"
set values_anRichness_rubber item 0 values_anRichness_rubber
set weights_anRichness_rubber csv:from-file "/input/bird/predictor_distributions/weights_anRichness_rubber.csv"
set weights_anRichness_rubber item 0 weights_anRichness_rubber
set values_beetleRichness_rubber csv:from-file "/input/bird/predictor_distributions/values_beetleRichness_rubber.csv"
set values_beetleRichness_rubber item 0 values_beetleRichness_rubber
set weights_beetleRichness_rubber csv:from-file "/input/bird/predictor_distributions/weights_beetleRichness_rubber.csv"
set weights_beetleRichness_rubber item 0 weights_beetleRichness_rubber
set values_canopyOpenness_rubber csv:from-file "/input/bird/predictor_distributions/values_canopyOpenness_rubber.csv"
set values_canopyOpenness_rubber item 0 values_canopyOpenness_rubber
set weights_canopyOpenness_rubber csv:from-file "/input/bird/predictor_distributions/weights_canopyOpenness_rubber.csv"
set weights_canopyOpenness_rubber item 0 weights_canopyOpenness_rubber

calculate_patch_bird_richness
show_overall_patch_bird_richness_statistics

end

```

### Expansion of the user interface

For starting the model, a button has to be added with the command `model-biodiversity`.

For visualizing the cell bird species richness, a button has to be added with the command `visualize-bird-richness`.

For validation of the model, a button has to be added with the command `validate-model-output`.

For the spillover effect, a slider must be added for the global variable `spillover_effect`, that has a minimum of 0 and a maximum of 1.

For model output display, monitors displaying the mean cell bird species richness must be added for forest, oil palm, jungle rubber, rubber and overall bird species richness with the according reporters `forest_bird_richness`, `oilpalm_bird_richness`, `junglerubber_bird_richness`, `rubber_bird_richness` and `overall_bird_richness`.

Additionally, a plot needs to be added for cell bird species richness distribution display. The name may be set to *bird species richness per patch*, x axis label may be set to *patch species richness* with minimum of 0 and maximum of 35, the y axis label may be set to *no. of patches*. *Auto scale?* is to be checked. Four pen update commands need to be included for forest, jungle rubber, rubber and oil palm respectively: `do-histogramForest`, `do-histogramJunglerubber`, `do-histogramRubber` and `do-histogramOilpalm`.

### Using the model

Before using the bird species richness model, the EFForTS-ABM needs to be setup. Therefore, a variety of settings can be adjusted in the *EFForTS-ABM Parameters* tab. Afterwards the EFForTS-ABM can be initialized with the button *setup* in the model control section. Then the model can be started with *Go - loop* or *Go - once* in the model control section. The bird species richness model can then be applied at any given point in time and after any amount of ticks of the model since the bird species richness model is a static model. For application of the bird species richness model the button *biodiversity-model* is used after the strength of the *spillover\_effect* is defined by the slider, ranging from 0 (no spillover effect) to 1 (highest implemented spillover effect). The model is applied to each of the cells of the landscape. Afterwards, the distribution of bird species richness can be visualized in different shades of red by *visualize-bird-richness* and statistics of the model output can be generated and displayed in the command center to be compared to validation data by *validate-model-output*. Mean cell bird species richness is displayed by the monitors separated for land use types. The distributions of cell bird species richness separated for land use types are displayed in the plot.

The model can be used to investigate land use change under different scenarios by changing the EFForTS-ABM Parameters in the respective tab. Scenarios to investigate could include for example different land use compositions and configurations customizable via *which-map*.

## Acknowledgements

I would like to thank my thesis advisor Jan Salecker for his patience and the great counseling that has enabled me to write this final assignment of my studies. Without his help, compiling the data from the different parts of the EFForTS-project would have been far more difficult and I would not have learned nearly as much as I have during this thesis. Also, I would like to thank my supervisors Prof. Dr. Kerstin Wiegand and Prof. Dr. Holger Kreft, who gave me the opportunity to work on my thesis in their working groups. In addition, I would like to thank all people from the group of ecosystem modelling for giving me the opportunity to present and discuss my thesis in the department's seminar. Furthermore, I would like to thank all researchers from the EFForTS-project who put the effort into investigating the effects of land use transformation in Indonesia and collecting the information and data that the present thesis makes use of. Special thanks go to Dr. Kevin Darras and Prof. Dr. Teja Tscharntke for providing the bird count data used in the model construction of the present study, Dr. Kevin Darras for the bird count data used for model validation, Prof. Dr. Teja Tscharntke for the data on ant species richness, Dr. Paul Magdon and Prof. Dr. Christoph Kleinn for the elevation data and the NDVI data, Dr. Dian Melati and Prof. Dr. Christoph Kleinn for the land use map used in the present study, Prof. Dr. Alexander Knohl and Dr. Christian Stiegler for the data on air humidity and air temperature, Dr. Katja Rembold and Prof. Dr. Holger Kreft for the data on plant species richness, canopy openness, stand structural complexity as well as effective number of layers, Dorothee Sandmann and Prof. Dr. Stefan Scheu for the data on oribatida species richness, Prof. Dr. Stefan Scheu for the data on protist and chilopoda species richness, Dr. Jochen Drescher and Prof. Dr. Stefan Scheu for the data on butterfly species richness, Dr. Josephine Sahner and Prof. Dr. Andrea Polle for the data on land use intensity, Prof. Dr. Ulrich Brose, Dr. Andrew Barnes and Dr. Malte Jochum for the data on beetle and spider species richness, Prof. Dr. Ulrich Brose for the data on invertebrate species richness as well as to all the people involved in data collection for the study of DRESCHER *et al.* (2016) for the information on litterfall and biomass.

Moreover, I want to thank all the fellow students, lecturers and friends, that I had contact to during my master studies, who made the experience of studying so enjoyable, interesting and mind broadening. Finally, Wiebke Nieberg has my full gratitude not only for the great work reviewing the manuscript of the present study for formal mistakes.

## Statutory declaration

I declare that I have developed and written the enclosed master thesis completely by myself and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The master thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Göttingen, 31.07.2018

---

Location, date



---

Mats Mahnken