230414FF2_Final_reviewed.docx

Creating virtual forest around the globe: Forest Factory 2.0 and analysing the state space of forests

Authors: Hans Henniger*, Andreas Huth*, Karin Frank*, Friedrich Bohn*

* Helmholtz-centre of environmental research
Permoserstr. 15
04318 Leipzig
Germany
hans.henniger@ufz.de, andreas.huth@ufz.de, karin.frank@ufz.de, friedrich.bohn@ufz.de

Corresponding author: Hans Henniger

Highlights

- Recent model development, called "Forest Factory 2.0" which generates various virtual forest stands for different biomes on earth
- Data product (including 700,000 forest stands) to demonstrate the potential of this approach
- Gaining knowledge about forests by analyzing the state space of forests
- Systematic mechanistic analyses of structure-function relationships across biomes

Abstract

Forests, as one of the most important carbon sinks on earth, are more and more under stress by environmental changes. The dynamics of forests and consequently their functions in general, begin to change. We therefore present a recent model development, called "Forest Factory 2.0", which generates various virtual forest stands for different biomes on earth. This approach allows to generate forests using the architecture and processes of forest models (here we use individual based gap model FORMIND). Using Forest Factory 2.0, we generated 700,000 forest stands in seven different ecoregions. In contrast to the tradition of investigating the development of individual forest stands over time, we used the Forest Factory 2.0 as a tool to gain knowledge about forests by analyzing the state space of forests. We conducted a structural sensitivity analysis to compare the relationships between structural properties and Biomass, productivity, as well as Species evenness of forests. In this study we analyse the state space of forests in different biomes and demonstrate the potential of this approach for theoretical ecology.

Keywords

forest model, forest generator, ecosystem functions, productivity, forest biomass, forest factory

1. Introduction

Forests cover 25% of the global land surface (Gibson et al. 2011). They play a major role for the global carbon cycle because of their function as carbon storage and their contributions to global carbon fluxes (Grace et al. 2014, Bonan 2008). Forests are important for sustaining biodiversity and provide habitat for 70% of all animal species (Gibson et al. 2011, Myers et al. 2000, Pimm et al. 2014). Further, forests exhibit a diversity of spatial structure and change their structure due to natural succession, management or disturbances (Pan et al. 2013).

While forest ecosystems are exposed to environmental change, like all complex adaptive systems, they have a certain capacity to cope with it. However, if these change processes occur too frequently, on too large spatial scales, with too high intensity, the adaptive capacity of the forests may be exceeded. Global change processes such as climate change and related effects such as drought, heat waves, fire, storms or pest outbreaks (IPCC 2013), but also deforestation (IPCC 2013) and fragmentation (Taubert et al. 2018, Fischer 2021, FAO 2022) are accelerating and occur simultaneously. As a result, the dynamics of the forests would change as well as their tree species composition and structure. Therefore, forests appear to be under increasing pressure (McDowell et al. 2020), affecting forest biodiversity in general as well as the diversity of functions provided by forests. This shows the urgency of sustaining their functioning, understanding and enhancing their adaptive capacity, and appropriately adapting their management.

Prerequisite for addressing these challenges, however, is a sound understanding of structure-function relationships, esp. between the properties of forests (species-compositional and spatial-structural) and their functions (e.g. carbon flux and storage). To analyze forests ecosystems the perspectives of community ecology (organismal aspects, diversity of species and structure) and ecosystem ecology (matter and energy flux aspects, biogeochemical cycles) are not separable. A mechanistic understanding of the functioning of ecosystems can only be gained if these two perspectives are adequately linked to one another (Loreau 2010). However, though there are many empirical studies based on data on forest inventories, the number of available samples and plots or the lack of focus in monitoring make it difficult to create a sufficiently complete picture (Lindenmayer and Likens 2009, Lindenmayer et al. 2011). This challenge may be overcome by using remote sensing data, but relating and condensing this large scaled data to the local or individual scale remains a challenge (Ma et al. 2020). Another challenge is to capture the inherent spatial heterogeneity of environmental conditions and how they change in response to projected changing processes, especially among different biomes. Thus, there is a huge variety in the environmental factors which are supposed to influence forest properties (species composition and structure) and the shape of the structure-function relationship. The needed relevant variables are mostly not fully covered by the existing inventories and datasets.

Forest models can help to bridge the gap between multiscale field data and processes enabling a multivariate view of forests. Nevertheless, different types of models have different application fields. For example, global vegetation models have a focus on large spatial scales and time scales, whereas individual-based models focus on smaller scales, as they consider processes at tree level and can thus also analyze structural dynamics (Maréchaux et al. 2021). Thus, individual-based models are particularly suiTable for considering ecosystem dynamics as an emergent outcome from the interaction of processes at individual level. This allows the identification of structural properties and functional characteristics of forests at different spatial scales as they emerge from the assumed environmental conditions. This also opens up the opportunity for correlative analyses of the structure-function relationship (Roedig et al. 2018, Thurner et al. 2017). However, the causal relationships underlying them are not yet satisfyingly understood.

We introduce a new way of sensitivity analysis. The variation parameter values or the comparison of different scenarios is a prominent way of sensitivity analysis to gain causal understanding of relationships. We perform sensitivity analysis not by varying parameters, but by analyzing millions of initial states to gain understanding of the relationship between forest functions and forest structure for different biomes. Performing this way of sensitivity analysis is not common so far and methods for this are rare. Examples for such a powerful application are the use of landscape generators in the context of impact assessments of land use scenarios (Langhammer et al. 2019, Engel et al. 2012) and the use of weather generators in the frame of climate impact analyses (Friend 1998, Kumagai et al. 2004).

For forests, such a generator was developed by Bohn & Huth 2017, the so-called ‘forest factory approach’. In one of their studies they generated virtual forest stands that possibly could exist in Central Europe. This multivariate dataset enabled a multidimensional investigation of the relationships between structural properties, plant diversity and productivity (Bohn & Huth 2017, Bohn et al. 2018). This promising approach has shown on the basis of simple mechanisms that over a broad range of forest stands, several forest properties (biodiversity and structure) have to be considered to understand forest productivity. The forest factory approach establishes a new way to analyze forests which does not require simulating forests over long periods of time. Instead the focus of the analysis is on the state space of the forests (described by structure properties). Due to the regional limitations of the forest factory by Bohn and Huth (focus on European forests), it offers potential for further research. To realize the potential and to analyze a causal relationship between their structural -, diversity -, and productivity relationships for different biomes, a further development and extension of the forest factory approach is necessary.

In this study we present a novel software tool - the Forest Factory 2.0 - which creates millions of virtual forest stands, covering various species compositions and structural properties for different biomes.
Additionally, we provide a data product generated with the software tool to demonstrate the potential of this approach for systematic mechanistic analyses of structure-function relationships across biomes. The data product contains in total 700,000 forest stands including 12 forest properties. These forest stands consist of over 11 million individual trees with over 20 tree properties.

In this study, we show examples of ecological analysis based on the generated forests. First, we compare the state space (based on four structural properties) of forests between seven regions derived for different biomes. Second, we compare the relationships between the four structural properties of forests and (i) biomass (as a proxy for the carbon stock), (ii) aboveground wood production AWP (as a proxy for the carbon flux), and (iii) species evenness (as an example for a biodiversity index). With the analysis, we want to show the potential of the presented approach for a wide range of research questions.

2. Methodical concept

With the Forest Factory 2.0 we have developed a software tool that makes it possible to create virtual natural forests that could exist in nature. The Forest Factory 2.0 allows through its algorithm a fast and generic generation of forests. In contrast to forest simulations, the Forest Factory approach does not consider and simulate forests over a long period of time. It generates various forests describing different states of succession, as well as management and disturbed forest stands also for different species mixtures. The forests can be generated for different regions of the world. In this study we produced 700,000 forest stands in total for seven different ecoregions. The background knowledge, i.e. the information and processes for the generation of forests, is provided by forest inventories and studies which are represented in the parameterizations of forest models (here we use the forest model FORMIND). A large number of ecological properties can be calculated for each generated forest, which allows a detailed analysis of the relationships between forest properties. Comparison of forest stands for different ecoregions is made possible by using the same algorithm for each forest stand generated.

2.1 Forest Factory 2.0

For processes such as competition and productivity, the Forest Factory 2 uses the individual- and process-based forest model FORMIND. This forest model allows the simulation of species rich forests and also considers the complex age structure of their tree community. FORMIND has been extensively tested and applied to tropical forests (Köhler and Hutz 2004, Gutiérrez and Huth 2012, Huth and Ditzer 2001, Kammesheidt et al. 2001, Köhler et al. 2003, Köhler and Huth 2007, Rüger et al. 2008, Fischer et al. 2014, Rödig et al. 2019) and temperate forests (Bohn et al. 2014, Bruening et al. 2021, Rüger et al. 2007) and grasslands (Taubert et al. 2012). It is an individual-based model which means that the growth of every single tree is simulated. The model considers four main process groups: growth of single trees (increment of tree biomass, stem diameter and height), mortality, recruitment, and competition (e.g. for light and space). FORMIND is also used for large scale simulations (Paulick et al. 2017, Rödig et al. 2018) e.g. forest-wide carbon balances in the Amazon. The Forest Factory 2.0, is implemented as an independent module of FORMIND in C++ language and uses processes of the forest model FORMIND (like competition for light and allometries). The processes of the forest model can be modified independently of the Forest Factory 2.0. It is possible to combine the Forest Factory 2.0 with other forest models.

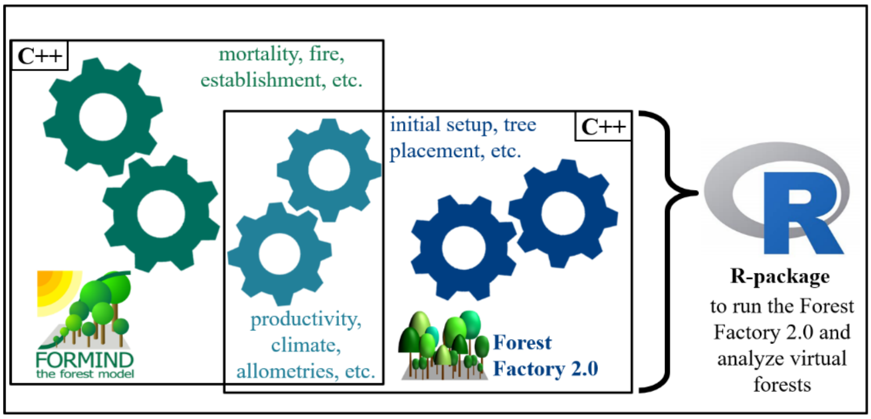

Figure 1: Interdependencies of FORMIND, Forest Factory 2.0 and the R package. The Forest Factory use processes of the forest model FORMIND. The R package (wrapper of C++ code) helps to run the Forest Factory 2.0 and process the generated forest stands to a data product. It also prevents some features for the analysis of the data product.

The methodology of the Forest Factory 2.0 follows the Forest Factory (Bohn 2017), that generated forest stands for the temperate zone and was implemented in the language R. In this paper, we introduce a new version of the Forest Factory that includes important new components and extensions that make it applicable on a global scale. We also provide an R package that facilitates the use of Forest Factory 2.0 (Figure 1). This package allows analysis of the data product that we publish or that users generate themselves.
The Forest Factory 2.0 can produce a large number of virtual forest stands (20m x 20m base area and funnel shape) for each available parameterization, which is representing an ecoregion. Every tree in the generated forest stand must have a positive productivity (gross primary production > respiration). In FORMIND a negative productivity causes the dying of trees. To calculate the productivity, we calculate the biomass increment of every placed tree over one year, which results from the different ecoregion-specific parameterizations (e.g. climate). To create forests for an ecoregion the Forest Factory needs information on climate conditions and a parameter set which consists of species-specific parameters e.g. concerning the tree geometry, productivity and species pool (see Section 2.3. for details) which are representative for an ecoregion.

As an initial information, which is valid for all generated forest stands, the Forest Factory 2.0 assumes a minimum and maximum height of the trees $H_{min}\ and\ H_{max}$, an overall maximum total crown volume ρ_{max} and an initial species pool. The overall maximum total crown volume ρ_{max} is the maximum sum of crown volume of all trees valid for every forest stand. The species pool is defined by the parameterization of each ecoregion and each species/plant functional type is representing a species or group of species with similar functional and morphological characteristics. This initial information is required to start the Forest Factory 2.0 (Figure 2).
Once started, the Forest Factory 2.0 pre-selects for each forest stand a minimum and maximum height of trees $h_{min}\ and\ h_{max}$ (from the initial $H_{min}\ and\ H_{max}$), a maximum total crown volume ρ and a group of plant functional types. The pre-selection for h_{min}, h_{max} and the maximum total crown volume ρ is done by random assuming uniform distributions (the

boundaries are $[H_{min}, H_{max}]$ and $[0, \rho_{max}]$). The pre- selection of h_{min}, h_{max} for every forest stands also allows the generation of even aged forests. The pre-selection of the species pool for each forest stand is done by random assuming a uniform distribution to select the number of species (more details in Appendix).

After the pre-selection for the forest stand is done, one tree after another is planted. The explicit position of a tree in the forest stand is not important due to the spatially implicit approach of forest gap models, where the position is randomly chosen at the end of the tree placement procedure. A tree height for the tree to be planted is selected from a predefined height distribution ($X \sim Exp(-0.05), h = X | X \in [h_{min}, h_{max}]$). The selected species pool is used to determine randomly the species type of a new tree (each species has an equal probability). For each tree these two attributes (height and species type) are selected and are used to calculate other attributes of the tree. Attributes are derived from processes and the parameter input of the used forest model (here we use FORMIND for different forest biomes). For the tree placement, it is checked if: a) each tree has a positive productivity, b) there is still space for the canopy of this tree (in each height layer, all tree crowns together must not exceed the boundaries of the forest stand) and c) the maximum total crown volume is not exceeded (we allow a certain maximum density in three-dimensional space: the maximum total crown volume). The selection rules for tree height and tree species is the same for each tree. If b) or c) is violated, the tree will not be considered, the tree placement for this forest is terminated and the created forest stand is saved in a database. If a) is violated and the calculated productivity over one year is negative, an attempt is made to replace the tree with a tree of a different species (with the same height and out of the selected species pool for this forest stand). If the tree has now a positive productivity, it is placed, if not, the tree placement for this forest is terminated (and the forest stand is saved). Every time a new tree is placed the annual productivity of all previous planted trees have to be recalculated (e.g. due to the change of light availability). If one or more trees have a negative productivity the algorithm try to replace them with tree(s) of another species and if this doesn't work the tree placement for the forest is terminated. The generation of a new forest stand starts.

In this way, the Forest Factory can be used to generate millions of forests for different ecoregions and climates (by considering input parameterizations). The forests describe different states of succession (e.g. by differentiate H_{min} and H_{max}), as well as managed (e.g. even aged forests by the selection of H_{min} and H_{max} values with a small difference) or disturbed forest stands (e.g. by selecting a low overall maximum total crown volume ρ_{max}) including different species mixtures. The goal is to generate as many potential forest states as possible. For specific analyses of e.g. even-aged forests or late-successional forests, the virtual forests must be filtered according to the desired attributes.

For the derived forest stands a large number of properties and characteristics can be calculated by using the methods of the forest model e.g. for leaf area, diameter increment, LAI per height layer, size distribution, biomass, maintenance respiration, gross primary production (GPP), net ecosystem carbon exchange (NEE). Since we simulate the productivity of each forest over only one year, we here do not focus on temporal evolution, but on states and benefit from the knowledge contained in widely applied and long-established forest models (here FORMIND).

The Forest Factory 2.0 enables the possibility for a coupling with other forest models. The coupling setup would run an iterative process. The Forest Factory provides tree and forest stand information to the corresponding forest model. The calculation of productivity and tree attributes (e.g. due to allometry) takes place in the forest model, and is reported back to the Forest Factory.

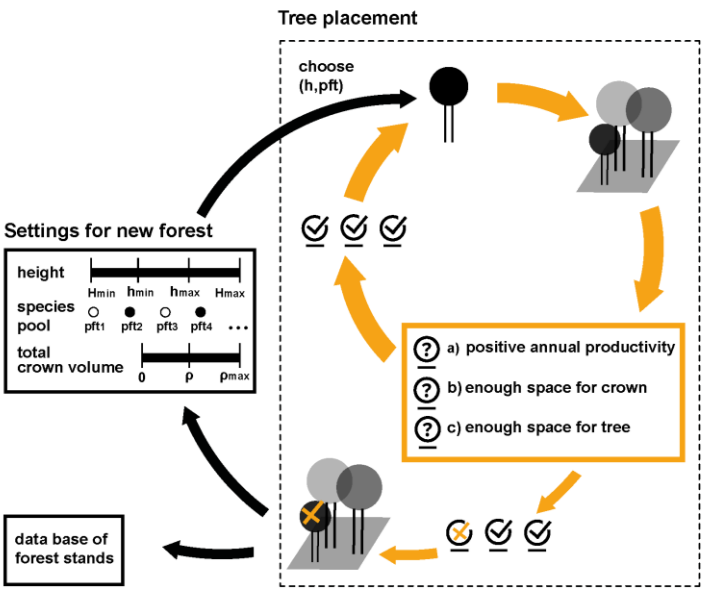

Figure 2: Concept of the Forest Factory 2.0. For each forest, Forest Factory 2.0 pre-selects a minimum and maximum height of trees, a group of species and a maximum total crown volume (sum of crown volume of all trees). Each tree is determined by a height (random from height distribution) and a species (random from the species pool). A new tree is added to the forest stand until the new tree has no positive productivity or space. Then the tree is deleted, the forest stand is saved in the data base and a new forest stand is generated.

2.2 Forest Factory (Bohn and Huth 2017) vs. Forest Factory 2.0

In this Section, we will explain the main differences between the Forest Factory by Bohn and Huth (2017) and the Forest Factory 2.0 and show how we have significantly extended the approach.

Forest Factory (Bohn, Huth 2017)	**Forest Factory 2.0**
10 000 forest stands per hour (standard notebook)	3 mio forest stands per hour (standard notebook)
programming language R	C++ and integration in the actual forest model (here FORMIND)
only temperate forests	forests in different biomes (ecoregions)
15 pre-defined stem diameter distributions	one continuous height distribution for tree placement
only stem diameters up to 0.5 m	no restrictions for stem diameter
planting trees until they are non-productive	replacing non-productive trees (by other species)
algorithm produces clustered sampling	more equally distributed sampling
	open source code and open data product

Table 1: Improvements of the Forest Factory 2.0 in comparison to the Forest Factory (Bohn and Huth 2017)

One important advantage of ForestFactory 2.0 compared to the Forest Factory by Bohn and Huth (Table 1) is a significant speed increase (3 million forest stands per hour, 30 times faster), which allows the creation of a huge number of forest stands. Further, it is now possible to create forests for all regions of the world for which parameter sets are available (here for the forest model FORMIND).

As the Forest Factory 2.0 is a part of the FORMIND model repository, functional model improvements are automatically available for the forest factory. This allows the Forest Factory 2.0 to use recently developed sub-modules of the forest model. For example, lidar waveforms or light reflectance spectra can be calculated for the generated forest stands. It is also possible to use the generated forests directly as input for simulations, to analyze future development of these forests.

Compared to the Forest Factory by Bohn and Huth, in the Forest Factory 2.0 the input parameters were reduced. There is only one function for tree height distribution to derive stem diameter as input (in the old version there were 15 fixed stem diameter distributions). This increases the flexibility and the possibilities for analysis. With the new Forest Factory 2.0, we can investigate the state space of the forests more evenly, i.e., different characteristics of the forest structure occur with sufficiently equal abundance (Appendix Figure A1).

We provide an open source R + Phyton package and a data product of forest stands to enable accessibility to a wide range of users. The R package (https://git.ufz.de/angermue/forestfactory) represents an interface which makes it possible to operate with the Forest Factory 2.0 from the R platform. An overview of the forest dataset is given in Appendix (Table A1).

2.3. Study Sites

Name in Paper	Biome short description	Number of pfts	Paper
Amazon	entire tropical forest in the Amazon using plant functional typed	3	Rödig et al. 2017
Panama	tropical lowland rainforest on Barro Colorado Island	4	Knapp et al. 2018
Germany	temperate forest in central Europe	8	Bohn, Huth 2014
US	temperate forest within the Northeast US	9	Bruening 2021
Ecuador	tropical evergreen montane rain forest in southern Ecuador	7	Dislich 2009
Malaysia	Southeastern Asian tropical rainforest (North Borneo, Malaysia)	4	update von Köhler Huth 2004
Tanzania	tropical submontane and lower montane rainforest at Mt. Kilimanjaro	6	Fischer et al. 2015

Table 2: Overview of the ecoregions, parameterizations and climate used for this study

The parameterizations (representing ecoregions in Table 2) represent the synthesis of information of many field measurements and inventories, not only concerning the species-specific allometric tree attributes but also concerning tree growth and productivity. Due to this we use for all ecoregions the same kind of information only with different values. The parameterizations can therefore be interpreted as a kind of recipe with always the same ingredients, in different quantities. The cooking process - the algorithm of the Forest Factory 2 - works for all parameterizations according to the same principle. The used parameterizations belong to different forest stands in different regions and we decided to use the names of the ecoregions in the paper to make clear where the investigated forests are located. For the generation of temperate forests in Germany we use a daily based climate data set of the Hainich National Park (Thuringia, Germany) for the year 2007. For the other regions we used reduced climate information which is described in the Appendix (Section 3).

2.4. Simulations & Analysis

In this study we present results for forest stands in seven different ecoregions (see Table 2). The ecoregions consist of two temperate regions (one in North America, one in Europe) and five tropical regions (two in South America, one in central America, one in Africa, one in Asia). For each region we generated 100,000 forest stands with the Forest Factory 2.0 (initial parameters: $H_{min} = 5\ m$, $H_{max} = 65\ m$ and the overall maximum total crown volume $\rho_{max} = 0.78$). Each region provides an initial species pool. We analyzed all forest stands for structural attributes (basal area, LAI, height heterogeneity, maximum height) and functional characteristics (above-ground wood productivity AWP, aboveground biomass and species evenness as an indicator for biodiversity. Species evenness is calculated by the Shannon Equitability Index (Heip 1974, Peet 1975). The Shannon Index (Shannon 1948) is normalized by the logarithm of the maximum number of species (we treat pfts as species here).

In a first step we explored under which structural conditions forests can exist in different ecoregions. For this, we use a state space approach. This space is determined here by four structural variables: maximum height, basal area, height heterogeneity, and LAI.
In Chapter 3.1 we investigated this state space of forests (mentioned above), by using diagrams (Figure 3) similar to the classical diagrams of Whittaker, in which he analyzed the relation between climate (average annual temperature and precipitation) and vegetation types (Whittaker 1970). Instead of climatic attributes we investigate here four structural properties (two in each Figure). We analyzed maximum height of trees (this corresponds to the forest height) and basal area which are typical properties to describe the structure of forests. Additionally, we investigated the role of tree height variability (here by using the standard deviation of the tree heights which we define as height heterogeneity) and leaf area index. Each generated forest stand can be represented as a point in the state space by a combination of these structural properties. We generated 700,000 forest stands, each representing a possible state, resulting in 700,000 points in the state space (100,000 for each ecoregion). To analyze the state space of the generated forest stands with positive productivity (Section 3.1.), we examined the

area which is covered by 100,000 forest stands of the same ecoregion by calculating the envelope around the points (each point represents one forest stand). These envelopes are calculated with the R package Concaveman (which uses convex hulls with concavity, for details please see Appendix Section 4).

To investigate the relationship between different forest properties and characteristics (Section 3.2.) we derived heatmaps (Figure 4,5) where the x and y axis describe structural properties and the color describe functional characteristics: biomass (carbon stock), AWP (carbon flow) and evenness (biodiversity). The maps are rastered so one cell contains information of several forest stands with the same structural properties. The shown value for a cell represents the mean value over these forest stands. We also derived maximum value and standard deviation for these analyses (shown in the Appendix).

To allow direct comparison of forests between the seven ecoregions, we examined forests by their functions (biomass, AWP, evenness) that are similar in all four structural properties (Figure 6). For these similar structured forests, we calculated the mean value of their functional characteristics and compared them in a 1:1 graph for three different regions. Additionally, the regression line and the adjusted R^2 were calculated.

For all analysis we considered only forest stands with a basal area under 100. In the Appendix the analysis of the maximum values and the standard deviation of the biomass, AWP and species evenness have been added (see Appendix Figures A7 – A11).

3. Results

3.1. Analysis of forest structure in different ecoregions

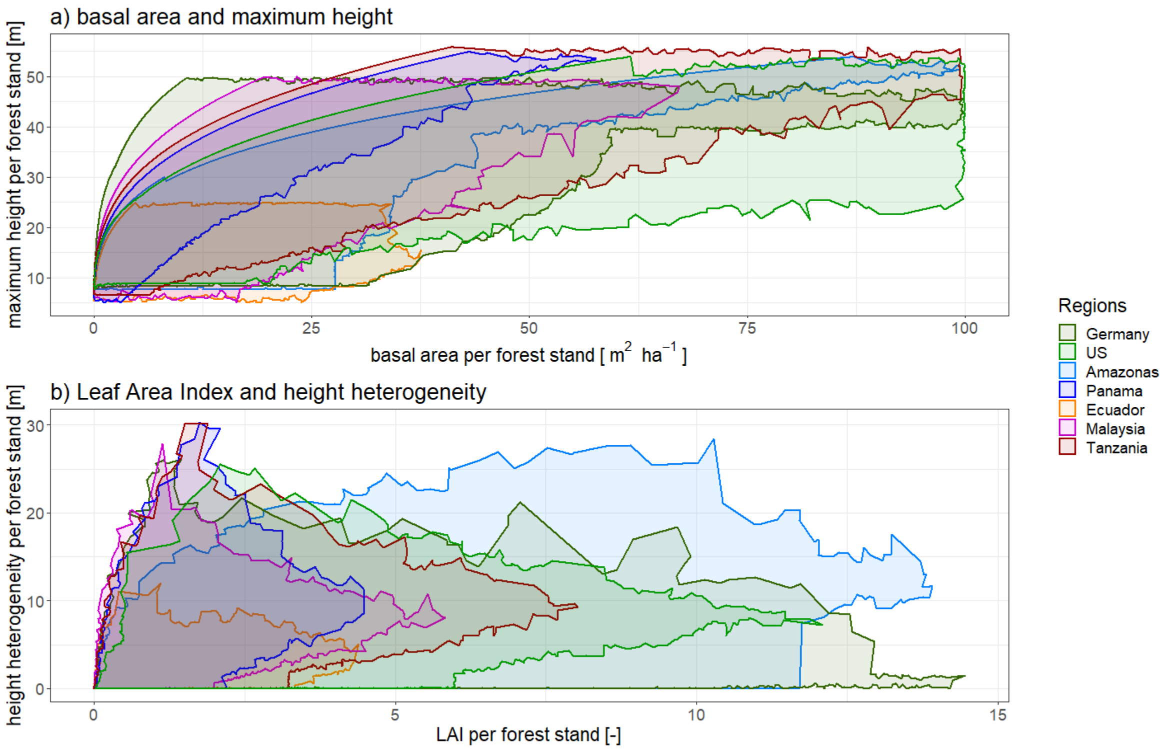

Figure 3: Analysis of the state space of generated forests. We examined the area which is covered by 100,000 forest stands of the same ecoregion by calculating the envelope around the points (for more details please see Appendix Section 4). Each generated forest stand can be represented as a point in the state space by a combination of the structural properties: a) maximum height and basal area; b) tree height heterogeneity and LAI. Different colors are indicating different ecoregions.

In the first step, we are looking at the structural characteristics of forest stands for the different ecoregions created by the Forest Factory 2.0 (Figure 3) by calculating the basal area, maximum height, height heterogeneity and LAI for each forest stand. The analysis in Figure 3 shows which combinations of maximum height/basal area and height heterogeneity/LAI lead to forests with positive productivity. Forests with properties outside the envelope line, don't have positive productivity. We observe mostly similarly-shaped envelopes with different sizes for the different ecoregions (represented by the different colors).

The analysis shows typical limitations of forest stands. Forest stands with a high basal area and low or moderate maximum height (empty area at the right bottom in Fig. 3a) do not occur. Large trees have large crowns. This tree allometries in combination with limited space restrict the abundance of these trees and also influences the resulting basal area (empty area at top left). The physiological and species-specific allometric interactions result in a typical shape in the state space that curves to the right.

Our forest stands can also be analyzed in a different state space, consisting of the LAI and the height heterogeneity. In most ecoregions, the largest values for tree height heterogeneity occur for forest stands with low LAI values, while the highest LAI values occur in forests with low to moderate tree height heterogeneity (Fig. 3b). As expected, the Brazilian Amazon has a large diversity of forest stands, and the shape of the envelope is quite different compared to other regions, e.g., without a peak at the top left (high height heterogeneity, low LAI).
The smallest area within the envelopes in both Figures (3a + 3b) is found for mountain forests of Ecuador (low maximum tree height).
We also investigated the frequency distributions of the forest properties of the forests within the illustrated areas (Appendix Figure A1).

3.2. Relationship between forest structure and ecosystem functions in different ecoregions

The Forest Factory 2.0 allows us also to analyze how structural properties (maximum height and basal area) affect functional characteristics (biomass, above-ground wood productivity (AWP), and evenness in the species composition (normalized Shannon Index as proxy for biodiversity). The biomass-related plots (Fig. 4: a, b, c) reveal a structure-function relationship that is quite similar for all investigated regions. Biomass is largely determined by the basal area and maximum height.

The analysis of the German forest stands (Fig. 4a) shows some interesting details for forests with high biomass and high basal area (top right area). Forests with a lower maximum height (40 m - 45 m) have on average a higher biomass than forests with a larger maximum height (> 45 m). With the Forest Factory, it is possible to analyze each individual tree of the corresponding forest stands. They all consist of trees of the species Picea abies. This is the tree species with the largest maximum height in the analysis for this region, but it has a low wood density, which leads to a lower forest biomass.

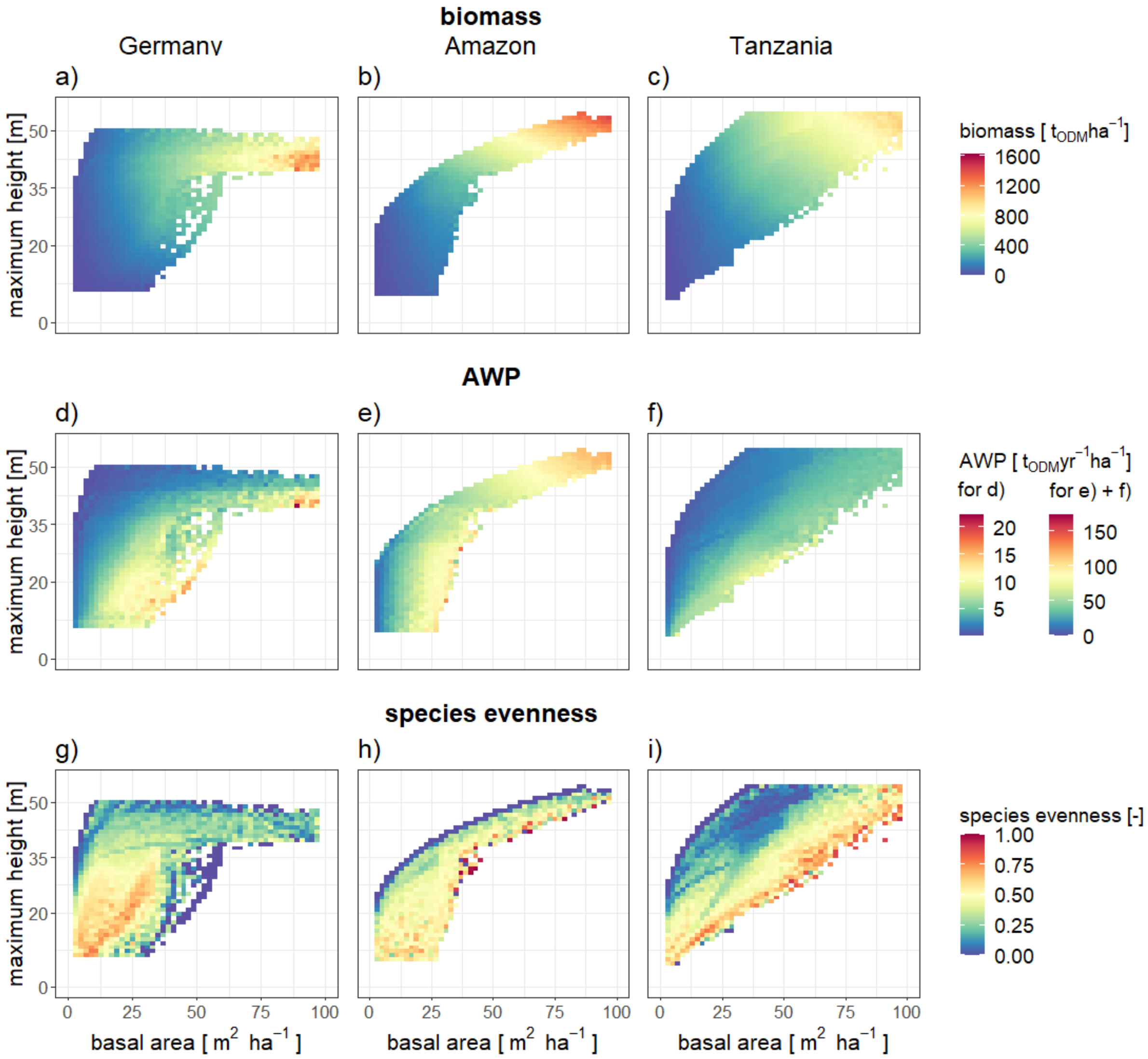

Figure 4: Relationship between structural properties (basal area and maximum height per forest stand) and biomass (a-c), above-ground productivity (d-f) and species evenness (g-i) for three selected ecoregions (Germany, Amazon, Tanzania). The color of each cell in the graph represents the mean value of the investigated property of all forest stands within one cell. Note that the AWP axes are scaled differently between the ecoregions (d-f). All other ecoregions, maximum value and standard deviation per property you can find in the Appendix (Figures A5 – A7).

In all three regions, we observe that AWP increases with the basal area and decreases with maximum height, while shape and strength of the combined effects are region-specific. Also, the range of AWP values differs due to climate variations between the temperate (Germany, Fig. 4d) and the tropic regions (Fig. 4: e, f), which leads to lower AWP values for the German forest stands. Nevertheless, we observe that for Germany (Fig. 4a) and the Amazon (Fig. 4b), forest stands with high AWP have a high maximum height and basal area. In all three ecoregions, there occur also forest stands with high AWP values that have only moderate basal area and height.

Concerning the evenness of species, forest stands (Fig. 4: g, h, i) show a similar structure-function relationship for the Amazon (Fig. 4h) and Tanzania (Fig. 4i). Species evenness is increasing with basal area but decreasing with maximum height. For the temperate forests in Germany (Fig. 4g), the situation is more intricate. Forest stands between low and medium maximum height (0 m - 35 m) and with medium basal area have high evenness values.

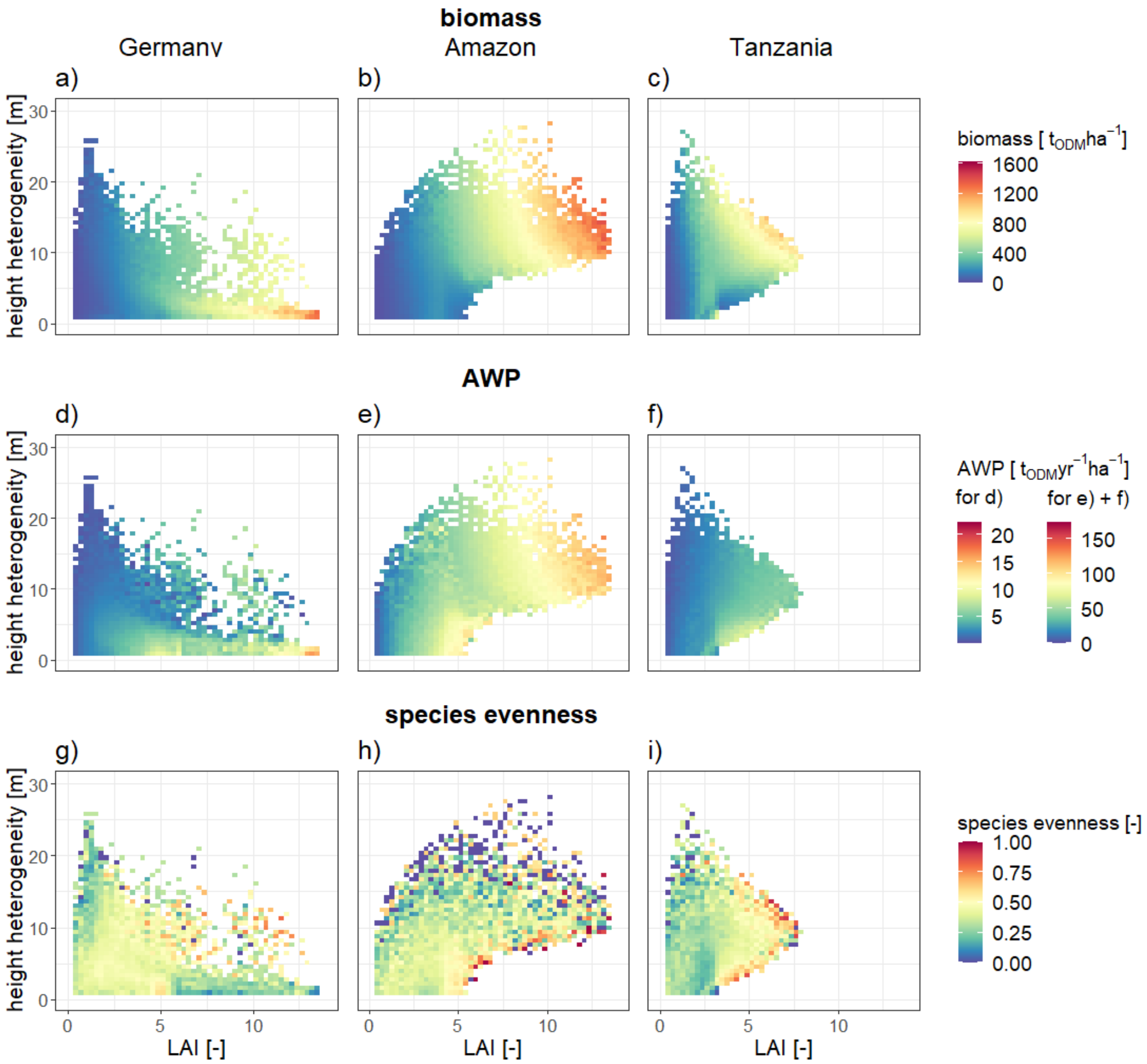

Figure 5: Relationship between structural properties (LAI and tree height heterogeneity) and biomass (a-c), above-ground productivity (d-f) and evenness (g-i) for the generated forests for three selected ecoregions (Germany, Amazon, Tanzania). The color of each cell in the graph represents the mean value of the investigated property of all forest stands within one cell. Note that the AWP axes are scaled differently between the ecoregions (d-f). All other ecoregions, maximum value and standard deviation per property you can find in the Appendix (Figures A8 – A10).

In a second step, we analyzed how two other structural properties (here: height heterogeneity and LAI) affect the functional characteristics of forest stands for the three investigated ecoregions (Fig. 5).

In all cases, forest stands with large biomass values (Fig. 5: a, b, c) can only be found if the LAI is high. Additionally, in the Amazon and Tanzania, these forest stands also need height heterogeneity values above 8 m.

We also analyzed forest productivity (Fig. 5: d, e, f). Forest stands with high AWP values have one pattern in common. High productivity goes along with low height heterogeneity and medium LAI in the Amazon and Tanzania (4-7 $t_{odm} yr^{-1} ha^{-1}$) and high LAI in Germany (5-13 $t_{odm}\ yr^{-1} ha^{-1}$). Lower AWP values of forest stands in Germany can be explained by the shorter vegetation period. In contrast to the other ecoregions for the forests in the Amazon, we also observe highly productive forests with large LAI (>10) and high height heterogeneity (>10 m), analogous to the biomass.

For Germany, it is remarkable that we observe only a few forest stands with medium height heterogeneity (6 m - 10 m) and medium LAI (2-10). Forest stands with these properties have low AWP values. These forests are next to an area in the state space without forest stands (white area), possibly due to the negative productivity of trees.

For species evenness (Fig. 5: g, h, i), we got no clear trends (like in the first and second row). Above a certain LAI (>3), we find Tanzanian forests with high species evenness (0.5 - 1). The highest evenness values (>0.7) can be found at the outer edge of the envelope.

The analysis of biomass, AWP and species evenness in German forests (Fig. 5: a, d, g) shows that forests with height heterogeneity smaller than 3 m and LAI larger than 6 have on average a lower evenness, besides all these forests have a high biomass and a high productivity. Results for the standard deviation and maximal biomass values, productivity and species evenness can be found in the Appendix (in Fig. 5, we analyzed mean values; for details, see methods and Appendix Figures A8 – A10).

3.3 Comparison of structure function relationships for different ecoregions

In the previous Sections, we examined structure-function relationships for different ecoregions. Here, we directly compare the structure-function relationships for three ecoregions (Amazonian, German and Tanzanian forests) to explore how generally the derived relationships apply (for comparisons for all ecoregions see Appendix Figures A11 - A13). Specifically, we compare mean biomasses (blue points in Fig. 6: a, b, c), mean AWPs (red points in Fig. 6: d, e, f) and mean species evennesses (green points in Fig. 6: g, h, i) of forest stands that have similar states (according to the four structural properties used in the Figures above) but are from different regions. We show them in 1:1 graphs. We consider forest states as similar if they have similar maximal height, height heterogeneity, LAI and basal area (details in Section 2.4.).

We observe a strong correlation for the biomass (high R^2 value). The biomass of forest stands with similar properties are not identical (not on the 1:1 line).

For the AWP (Figure 6: d, e, f), we see a good correlation. The four structural dimensions are sufficient to find relations between AWP for different regions but less effective than between the biomass.

We see no correlation in the evenness relationships for the different regions (Figure 6: g, h, i). That indicates that we may need more information in addition to structural properties to get a better correlation.

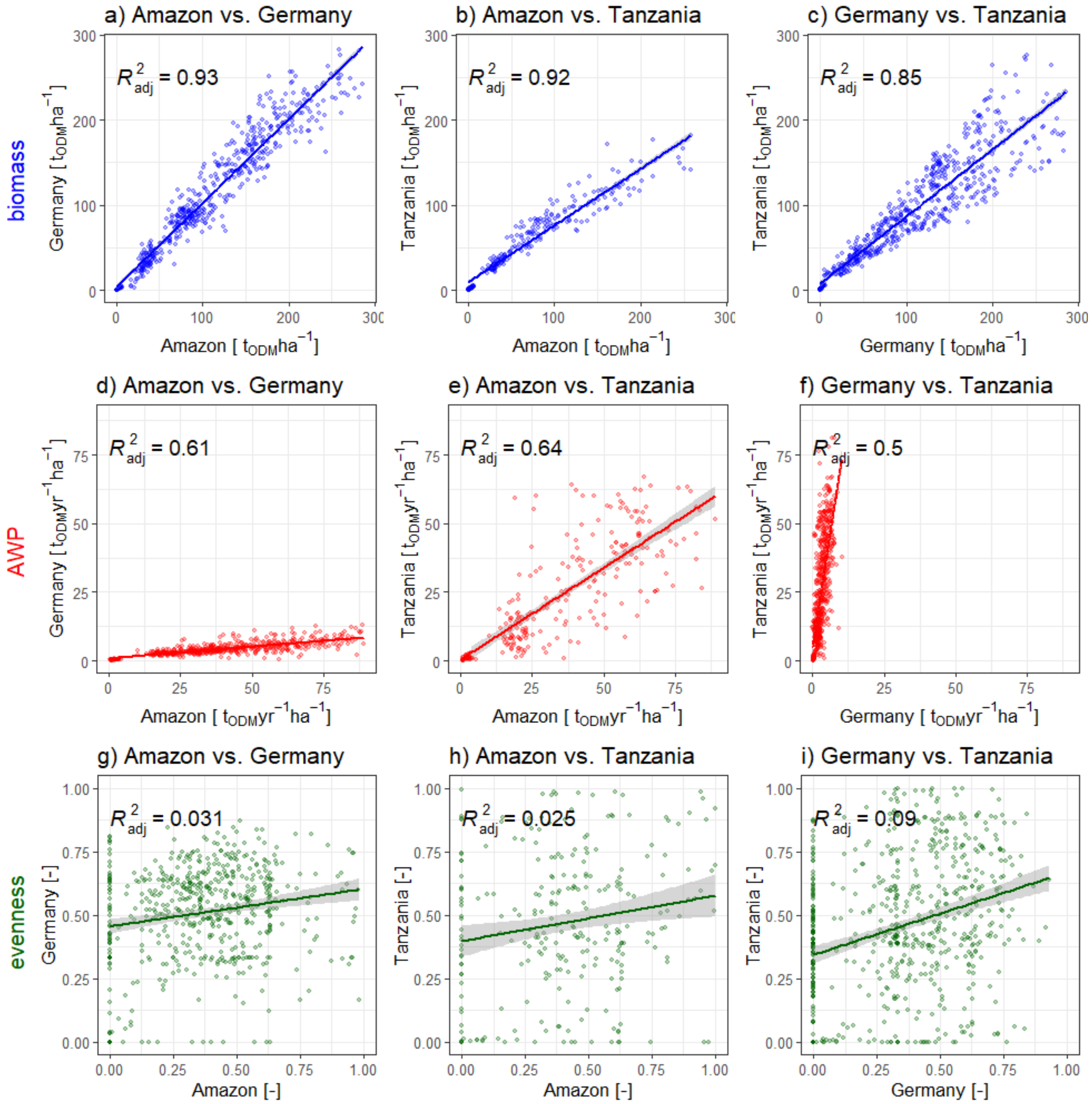

Figure 6: Comparison of biomass, aboveground productivity and species evenness derived from forests with a similar state space (by a 2% quantile of the four structural properties LAI, basal area, tree height heterogeneity and maximum tree height). Each graph compares forest stands out of two ecoregions. We show the pairwise comparisons for three illustrative ecoregions. Each point represents the mean values of the investigated functional characteristics.

4. Discussion

In this paper, we explored the Forest Factory 2.0 a new open source software tool to simulate and analyze forests from different biomes on earth. We demonstrated several benefits of the approach and provide insights into how this method can increase our knowledge on structure-function relationships of forests and overall forest functioning. Breaking with the tradition of investigating the development of individual forest stands over time, we used the Forest Factory 2.0 as a tool to gain knowledge about forests by analyzing the state space of forests, resulting from species pool and environmental factors.

The simple algorithm of the Forest Factory allows comparison of a large number of forest stands from different biomes (3 million forest stands per hour) generated with the same process-driven architecture. This also provides a causal understanding of forest structure-function relationships (as we showed in Section 3). In this manuscript, we present a method to investigate the relationship between structure (maximum height, basal area, LAI, height heterogeneity) and productivity (biomass and AWP) or biodiversity (species evenness) of forests. With the Forest Factory 2.0, it is also possible to analyze other forest properties, such as diameter increment or net ecosystem exchange. Additionally, it enables us to explore the role of other more complex structural characteristics like stem size distribution or height-layer-specific information. With this systematic approach we could investigate the causes of the differences and similarities of forest stands e.g. why forests with similar structure show different or similar biomass or productivity values. This could allow us to calculate transfer functions for structure-function relationships of forests from one ecoregion to another (outlined in Section 3.3). Here, we generated forest datasets for seven forest regions to illustrate the approach. It is also possible to use other parameterizations from other forest models to generate forests for additional regions.

With this approach it is not only possible to create forests that already exist but also could occur. Using the Forest Factory 2.0 to create forest states beyond the currently existing ones provides a fuller understanding of forests beyond the constraints of empirical data such as national forest inventories or remote sensing data. Some of these forest states may be due to current changes in disturbance regimes or management, and for some forest states it may not even be clear which successional or disturbance pathways will lead to them. In addition to the promising research area of realistic forest selection, it is also interesting to study forests that have almost no or even negative productivity. This analysis can be used to identify stressed forest stands (in forest inventories) or generally describe and understand the state space of stressed forests. This might help to detect potential regime shifts and to explore adaptive capacities of forests and forest ecosystems. As seen in Figure 3d, there are forests with medium height heterogeneity and LAI that have low productivity. These forests are next to an area in the state space without forest stands (white area). White areas may indicate that forest stands in this area of the state space have negative productivity. Such forests are not generated by the algorithm due to the productivity condition. Further analysis could reveal if this white area represents a transition from forests with low positive to forests with negative productivity. This area in the state space could give information on the limits of coping capacity of forest stands. If this were the case, these forests could change from being a carbon sink to being a carbon source.

In this study we presented how the Forest Factory 2.0 can be used to study region-specific patterns and the ecological mechanisms behind them. Every forest stand consists of many individual trees that are modeled by the selected forest model (here FORMIND). For each single tree, additional information is available (data product of the Forest Factory 2.0). This allows the analysis of specific forest attributes by analyzing the productivity or other properties of each tree in the forest, hence yielding a deeper understanding of forest dynamics. The Forest Factory 2.0 also offers the possibility for jointly addressing research questions from community ecology (organismal aspects, diversity of species and structure) to ecosystem ecology (matter and energy flux aspects, biogeochemical cycles) (Loreau 2010).

For making Forest Factory 2.0 easier to use for different user groups it might be useful to generate forests with only certain tree species for user groups that want to generate lidar data with Forest Factory 2.0. At the moment, forests with certain tree species can of course be sorted out of the data product or generated by changing the parameterization. Possible tree species should be selected during the initialization of Forest Factory 2.0. Perhaps users only want to study multilayer forests, so it would be interesting to allow other height distributions that make these forests more likely (e.g., bimodal height distributions), even if they already exist in the data product. Another direction would be to allow different spatial resolutions for Forest Factory 2.0 if users want to create larger contiguous forests without filtering and rearranging the ones already generated. An additional extension could be to allow different mechanisms for tree placement. It would be possible to remove trees in the virtual forest stands to mimic interventions. Also, we could implement mechanisms which guarantee a denser packing of forest stands and may widen the envelopes in Figure 3. Nevertheless, the presented envelopes show that we can already cover a broad range of different forest structures with the current approach.

Furthermore, the coupling of the Forest Factory with other modules of FORMIND allows us to explore additional properties and characteristics of the generated forest, for example to derive typical remote sensing data and indexes based on radiative transfer models. For instance, Bruenning et al. (2021) use the Forest Factory 2.0 to explore the relationship between lidar profiles and aboveground biomass. It is also possible to combine radiative transfer models with the Forest Factory 2.0 to generate reflection spectra for a huge number of forest stands. Virtual forests are also used in studies by the remote sensing community (Frazer et al. 2005, Frazer et al. 2011, Widlowski et al. 2015). In addition to the typical remote sensing forest variables (point clouds, lidar profiles), the generated forests allow the calculation of additional properties (basal area, LAI, AWP, net ecosystem exchange) also at the tree level. Thus, the presented approach can help to downscale the satellite-imagery-based data and to translate the remote sensing measurements available for large areas to the level of individual trees.

It is also possible to combine the Forest Factory 2.0 approach with other forest models. The new approach of looking at forests in terms of states rather than simulations over time, along with Forest Factory 2.0's free coupling possibility, offers a promising path to compare forest models and learn more about their capabilities and limitations. Specifically, it opens up the possibility of using different forest models to generate different databases of forest stands, as shown in this study with FORMIND, and then analyzing these comparatively using the methods presented. In addition to the possibility of combining the Forest Factory 2.0 with other forest models, the Forest Factory 2.0 is also an additional test for parameterizations. We can analyze forests that cannot be created by the forest succession for which the parameterization was made. These forests may be possible under different environmental conditions (like climate change) or due to disturbances (e.g. fallen trees).

Another possible application is the use of generated forest stands to initialize models simulating forest development (for different forest models) over a longer period of time. With this application it is possible e.g. to analyze the further behavior of these forests under climate change or management scenarios (natural extinction processes or implementation of new species). Again, the advantage is that we can simulate forests with states beyond those that currently exist and gain information that we cannot obtain from inventory or remote sensing observations. With forest models we can analyze the development of these forest stands which allows new ways of analysis. For example, we can explore forest states that are more resilient to climate change and should be pursued in forest management.

The presented way of analyzing forests in a digital universe of processes and mechanisms also offers new possibilities for data scientists. The freely available datasets of generated forest stands can be used to train artificial intelligence (AI) that estimates additional forest/tree attributes from just a few attributes of forest stands. The resulting relationships could be used to gain a deeper understanding at the level of individual trees from large-scale remote sensing observations. In addition, all relationships shown in the graphs and the data product could be condensed into equations by AIs.

With the Forest Factory 2.0, researchers can generate virtual forests for their needs or use the open-source forest data to analyze a digital forest universe of forest states.

Data accessibility. The method and dataset supporting this article has been uploaded to the R repository: https://git.ufz.de/angermue/forestfactory.

Authors' contributions. H.H., A.H., K.F. and F.J.B. conceived of the study. A.H. and H.H supervised the research. H.H. implemented and analysed the simulation model. F.J.B. provided guidance and technical support. All authors contributed to the interpretation of the results. H.H., A.H. and F.J.B developed the first draft of the manuscript. All Co-Authors reviewed and edited the manuscript. All authors gave final approval for publication.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements. This study was supported by the Collaborative Research Centre AquaDiva of the Friedrich Schiller University Jena, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1076 – "Project Number 218627073". We thank Samuel Fischer for providing many helpful suggestions and comments. We also thank the Department of Bioclimatology of the University Göttingen and the Max Planck Institute of Biogeochemistry for providing climate data and the administration of Hainich National Park for permission to conduct research there.

LITERATURE

Bruening, J. M., Fischer, R., Bohn, F. J., Armston, J., Armstrong, A. H., Knapp, N., ... & Dubayah, R., 2021. Challenges to aboveground biomass prediction from waveform lidar. *Environmental Research Letters*, *16*(12), 125013. https://doi.org/10.1088/1748-9326/ac3cec

Bohn, F. J., Frank, K., & Huth, A., 2014. Of climate and its resulting tree growth: Simulating the productivity of temperate forests. *Ecological Modelling*, *278*, 9-17. https://doi.org/10.1016/j.ecolmodel.2014.01.021

Bohn, F. J., & Huth, A., 2017. The importance of forest structure to biodiversity–productivity relationships. *Royal Society open science*, *4*(1), 160521. https://doi.org/10.1098/rsos.160521

Bohn, F. J., May, F., & Huth, A., 2018. Species composition and forest structure explain the temperature sensitivity patterns of productivity in temperate forests. Biogeosciences, 15(6), 1795-1813. https://doi.org/10.5194/bg-15-1795-2018

Bonan, G. B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *science*, *320*(5882), 1444-1449. https://doi.org/10.1126/science.1155121

Dislich, C., Günter, S., Homeier, J., Schröder, B., & Huth, A., 2009. Simulating forest dynamics of a tropical montane forest in South Ecuador. *Erdkunde*, 347-364.

Engel, J., Huth, A., & Frank, K., 2012. Bioenergy production and S kylark (A lauda arvensis) population abundance–a modelling approach for the analysis of land- use change impacts and conservation options. *Gcb Bioenergy*, *4*(6), 713-727. https://doi.org/10.1111/j.1757-1707.2012.01170.x

FAO, 2022. *The State of the World's Forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies*. Rome, FAO.

Frazer, G. W., Wulder, M. A., & Niemann, K. O., 2005. Simulation and quantification of the fine-scale spatial pattern and heterogeneity of forest canopy structure: A lacunarity-based method designed for analysis of continuous canopy heights. *Forest ecology and management*, *214*(1-3), 65-90. https://doi.org/10.1016/j.foreco.2005.03.056

Frazer, G. W., Magnussen, S., Wulder, M. A., & Niemann, K. O., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment*, *115*(2), 636-649. https://doi.org/10.1016/j.rse.2010.10.008

Friend, A. D., 1998. Parameterisation of a global daily weather generator for terrestrial ecosystem modelling. *Ecological Modelling*, *109*(2), 121-140. https://doi.org/10.1016/S0304-3800(98)00036-2

Fischer, R., Armstrong, A., Shugart, H. H., & Huth, A., 2014. Simulating the impacts of reduced rainfall on carbon stocks and net ecosystem exchange in a tropical forest. *Environmental modelling & software*, *52*, 200-206. https://doi.org/10.1016/j.envsoft.2013.10.026

Fischer, R., Ensslin, A., Rutten, G., Fischer, M., Schellenberger Costa, D., Kleyer, M., ... & Huth, A., 2015. Simulating carbon stocks and fluxes of an African tropical montane forest with an individual-based forest model. *PLoS One*, *10*(4), e0123300. https://doi.org/10.1371/journal.pone.0123300

Fischer, R., 2021. The long-term consequences of forest fires on the carbon fluxes of a tropical forest in Africa. *Applied Sciences*, *11*(10), 4696. https://doi.org/10.3390/app11104696

Gibson, L., Lee, T. M., Koh, L. P., Brook, B. W., Gardner, T. A., Barlow, J., ... & Sodhi, N. S., 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, *478*(7369), 378-381. https://doi.org/10.1038/nature10425

Grace, J., Mitchard, E., & Gloor, E., 2014. Perturbations in the carbon budget of the tropics. *Global Change Biology*, *20*(10), 3238-3255. https://doi.org/10.1111/gcb.12600

Gutiérrez, A. G., & Huth, A., 2012. Successional stages of primary temperate rainforests of Chiloé Island, Chile. *Perspectives in Plant Ecology, Evolution and Systematics*, *14*(4), 243-256. https://doi.org/10.1016/j.ppees.2012.01.004

Heip, C., 1974. A new index measuring evenness. *Journal of the Marine Biological Association of the United Kingdom*, *54*(3), 555-557. https://doi.org/10.1017/S0025315400022736

Huth, A., & Ditzer, T., 2000. Simulation of the growth of a lowland Dipterocarp rain forest with FORMIX3. *Ecological Modelling*, *134*(1), 1-25. https://doi.org/10.1016/S0304-3800(00)00328-8

Huth, A., & Ditzer, T., 2001. Long-term impacts of logging in a tropical rain forest—a simulation study. *Forest Ecology and management*, *142*(1-3), 33-51. https://doi.org/10.1016/S0378-1127(00)00338-8

IPCC, 2013. Climate change 2013: the physical science basis. In: Stocker ,T.F. ,et al. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p.1535.

Kammesheidt, L., Köhler, P., & Huth, A., 2001. Sustainable timber harvesting in Venezuela: a modelling approach. *Journal of applied ecology*, *38*(4), 756-770. https://doi.org/10.1046/j.1365-2664.2001.00629.x

Knapp, N., Huth, A., Kugler, F., Papathanassiou, K., Condit, R., Hubbell, S. P., & Fischer, R., 2018. Model-assisted estimation of tropical forest biomass change: A comparison of approaches. *Remote Sensing*, *10*(5), 731. https://doi.org/10.3390/rs10050731

Köhler, P., Chave, J., Riéra, B., & Huth, A., 2003. Simulating the long-term response of tropical wet forests to fragmentation. *Ecosystems*, *6*(2), 0114-0128. https://doi.org/10.1007/s10021-002-0121-9

Köhler, P., & Huth, A., 2004. Simulating growth dynamics in a South-East Asian rainforest threatened by recruitment shortage and tree harvesting. *Climatic Change*, *67*(1), 95-117. https://doi.org/10.1007/s10584-004-0713-9

Köhler, P., & Huth, A., 2007. Impacts of recruitment limitation and canopy disturbance on tropical tree species richness. *Ecological Modelling*, *203*(3-4), 511-517. https://doi.org/10.1016/j.ecolmodel.2006.11.023

Kumagai, T. O., Katul, G. G., Porporato, A., Saitoh, T. M., Ohashi, M., Ichie, T., & Suzuki, M., 2004. Carbon and water cycling in a Bornean tropical rainforest under current and future climate scenarios. Advances in Water Resources, 27(12), 1135-1150. https://doi.org/10.1016/j.advwatres.2004.10.002

Langhammer, M., Thober, J., Lange, M., Frank, K., & Grimm, V., 2019. Agricultural landscape generators for simulation models: A review of existing solutions and an outline of future directions. *Ecological Modelling*, *393*, 135-151. https://doi.org/10.1016/j.ecolmodel.2018.12.010

Lindenmayer, D. B., & Likens, G. E., 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in ecology & evolution*, *24*(9), 482-486. https://doi.org/10.1016/j.tree.2009.03.005

Lindenmayer, D. B., Likens, G. E., Haywood, A., & Miezis, L., 2011. Adaptive monitoring in the real world: proof of concept. *Trends in ecology & evolution*, *26*(12), 641-646. https://doi.org/10.1016/j.tree.2011.08.002

Loreau, M., 2010. Linking biodiversity and ecosystems: towards a unifying ecological theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 49-60. https://doi.org/10.1098/rstb.2009.0155

Ma, X., Migliavacca, M., Wirth, C., Bohn, F. J., Huth, A., Richter, R., & Mahecha, M. D., 2020. Monitoring plant functional diversity using the reflectance and echo from space. Remote Sensing, 12(8), 1248. https://doi.org/10.3390/rs12081248

Maréchaux, I., Langerwisch, F., Huth, A., Bugmann, H., Morin, X., Reyer, C. P., ... & Bohn, F. J., 2021. Tackling unresolved questions in forest ecology: The past and future role of simulation models. *Ecology and evolution*, *11*(9), 3746-3770. https://doi.org/10.1002/ece3.7391

McDowell, N. G., Allen, C. D., Anderson-Teixeira, K., Aukema, B. H., Bond-Lamberty, B., Chini, L., ... & Xu, C., 2020. Pervasive shifts in forest dynamics in a changing world. *Science*, *368*(6494), eaaz9463. https://doi.org/10.1126/science.aaz9463

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853-858. https://doi.org/10.1038/35002501

Pan, Y., Birdsey, R. A., Phillips, O. L., & Jackson, R. B., 2013. The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 593-622. https://doi.org/10.1146/annurev-ecolsys-110512-135914

Paulick, S., Dislich, C., Homeier, J., Fischer, R., & Huth, A., 2017. The carbon fluxes in different successional stages: modelling the dynamics of tropical montane forests in South Ecuador. *Forest Ecosystems*, *4*(1), 1-11. https://doi.org/10.1186/s40663-017-0092-0

Peet, R. K., 1975. Relative diversity indices. *Ecology, 56(2)*, 496-498. https://www.jstor.org/sTable/1934984

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *science*, *344*(6187), 1246752. https://doi.org/10.1126/science.1246752

Rödig, E., Cuntz, M., Heinke, J., Rammig, A., & Huth, A., 2017. Spatial heterogeneity of biomass and forest structure of the Amazon rain forest: Linking remote sensing, forest modelling and field inventory. *Global ecology and biogeography*, *26*(11), 1292-1302. https://doi.org/10.1111/geb.12639

Rödig, E., Cuntz, M., Rammig, A., Fischer, R., Taubert, F., & Huth, A., 2018. The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environmental Research Letters*, *13*(5), 054013. https://doi.org/10.1088/1748-9326/aabc61

Rödig, E., Knapp, N., Fischer, R., Bohn, F. J., Dubayah, R., Tang, H., & Huth, A., 2019. From small-scale forest structure to Amazon-wide carbon estimates. *Nature communications*, *10*(1), 1-7. https://doi.org/10.1038/s41467-019-13063-y

Rüger, N., Gutiérrez, A. G., Kissling, W. D., Armesto, J. J., & Huth, A., 2007. Ecological impacts of different harvesting scenarios for temperate evergreen rain forest in southern Chile—a simulation experiment. *Forest Ecology and Management*, *252*(1-3), 52-66. https://doi.org/10.1016/j.foreco.2007.06.020

Rüger, N., Williams-Linera, G., Kissling, W. D., & Huth, A., 2008. Long-term impacts of fuelwood extraction on a tropical montane cloud forest. *Ecosystems*, *11*(6), 868-881. https://doi.org/10.1007/s10021-008-9166-8

Shannon, C. E. (1948). A mathematical theory of communication, Bell Systems Tech. *J, 27,* 379-423.

Taubert, F., Frank, K., & Huth, A., 2012. A review of grassland models in the biofuel context. *Ecological Modelling*, *245*, 84-93. https://doi.org/10.1016/j.ecolmodel.2012.04.007

Taubert, F., Fischer, R., Groeneveld, J., Lehmann, S., Müller, M. S., Rödig, E., ... & Huth, A., 2018. Global patterns of tropical forest fragmentation. *Nature*, *554*(7693), 519-522. https://doi.org/10.1038/nature25508

Thurner, M., Beer, C., Ciais, P., Friend, A. D., Ito, A., Kleidon, A., ... & Carvalhais, N., 2017. Evaluation of climate‐ related carbon turnover processes in global vegetation models for boreal and temperate forests. *Global Change Biology*, *23*(8), 3076-3091. https://doi.org/10.1111/gcb.13660

Whittaker, R. H., 1970. Communities and ecosystems. *Communities and ecosystems.*

Widlowski, J. L., Mio, C., Disney, M., Adams, J., Andredakis, I., Atzberger, C., ... & Zenone, T., 2015. The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing. *Remote Sensing of Environment*, *169*, 418-437 https://doi.org/10.1016/j.rse.2015.08.016

APPENDIX

I. Additional information for method Section

1. Pre-defined forest stand attributes

For each forest stand we select a minimum and maximum tree height h_{min} and h_{max} (a), a species pool (b) and a total crown volume of trees (c).
The Forest Factory 2.0 receives as an initial input a minimum and maximum height of the trees H_{min} and H_{max} (this applies to all trees in all forest stands). From the range of H_{min} and H_{max} a forest stand specific h_{min} and h_{max} (a) is chosen randomly for each forest stand. We assume an equally distributed probability distribution in the mentioned range.
The parameterization of each ecoregion defines the total species pool for Forest Factory 2.0. For each forest stand, a forest stand-specific species pool (b) is chosen. For this purpose, a number between 1 and the number of species in the total species pool is chosen uniformly distributed. It determines how many species the forest stand-specific species pool should contain. Each species has an equal probability of being included in the forest stand-specific species pool until the next to last species (selected number of species -1) is selected. To ensure that it is possible to plant trees within the selected height range between h_{min} and h_{max} (which is different for each forest stand) we check whether at least one of the species selected so far has a maximum attainable height greater than or equal to h_{max}. If this is not the case, the last species is selected so that a tree with h_{max} could be placed. Forest Factory 2.0 receives as input also a maximum total crown volume. This total crown volume can be seen as a kind of crown density (proportion of crown volume to forest stand volume). None of the generated forest stands have a total crown volume above this input value. For each forest stand, between 0 and the maximum total crown volume, a forest stand specific maximum total crown volume (c) is randomly chosen, assuming an equal probability distribution.

2. Normalized Shannon Index

The Shannon Index H (Shannon 1948) and the species evenness E_H (Shannon equitability Index by Heip 1974, Peet 1975) is calculated by:

$$\mathrm{H} = \sum_{i=1}^{S} p_i \cdot ln p_i$$

$$E_H = \frac{H}{lnS}$$

with p_i *…Proportion of trees of species i in the total number of trees*
S*…Set of all species in the initial species pool derived by the parameterization of the ecoregion (we treat pfts as species here)*

3. Information about data product

Region	Number Forest Stands	Total number of Trees	# of pfts	mean basal area $[m^2 ha^{-1}]$	max basal area $[m^2 ha^{-1}]$	mean biomass $[t_{odm} ha^{-1}]$	max biomass $[t_{odm} ha^{-1}]$
Amazon	100 000	1,379,989	3	26.5	99.3	272	1430
Panama	100 000	611,013	4	15.6	57.6	166.3	752

Germany	100 000	1,258,893	8	26.5	99.8	163.7	1462
US	100 000	816,839	9	41.4	99.9	383.7	1623
Ecuador	100 000	3,851,850	7	20.7	37.5	116.9	228
Malaysia	100 000	1,623,988	4	19.9	67.2	269.9	1067
Tanzania	100 000	1,222,930	6	32.8	99.7	300.6	1091

Table A1: Description of the Data product of 700 000 forests stands from 7 different ecoregions. Each forest patch has an area of 20m x 20m. The total number of trees is accumulated over all forest stands from one ecoregion. The species mix is described through the number of pft's. In addition, the mean and maximum values of basel Area and biomass are shown.

4. Reduced climate information

The climate information for Hainich climate (Germany) includes daily temperature values, radiation values and precipitation values. In the case of reduced climate information, other values are used for productivity calculation: mean yearly light intensity above canopy during day-length, length of daily photosynthetic active period, i.e. day-length, relative length of wet and dry season. More information about these variables can be found in respective studies and parameterizations (see main text Table 2).

5. Realization of envelopes in R

For the analysis of the state space of forests with the help of envelopes we use the function geom_mark_hull of the R package R/mark_hull.R. It uses the package concaveman (https://github.com/mapbox/concaveman) which allows to adjust concavity of the resulting hull. We choose the following parameters: con. Cap = 0 and concavity = 2.

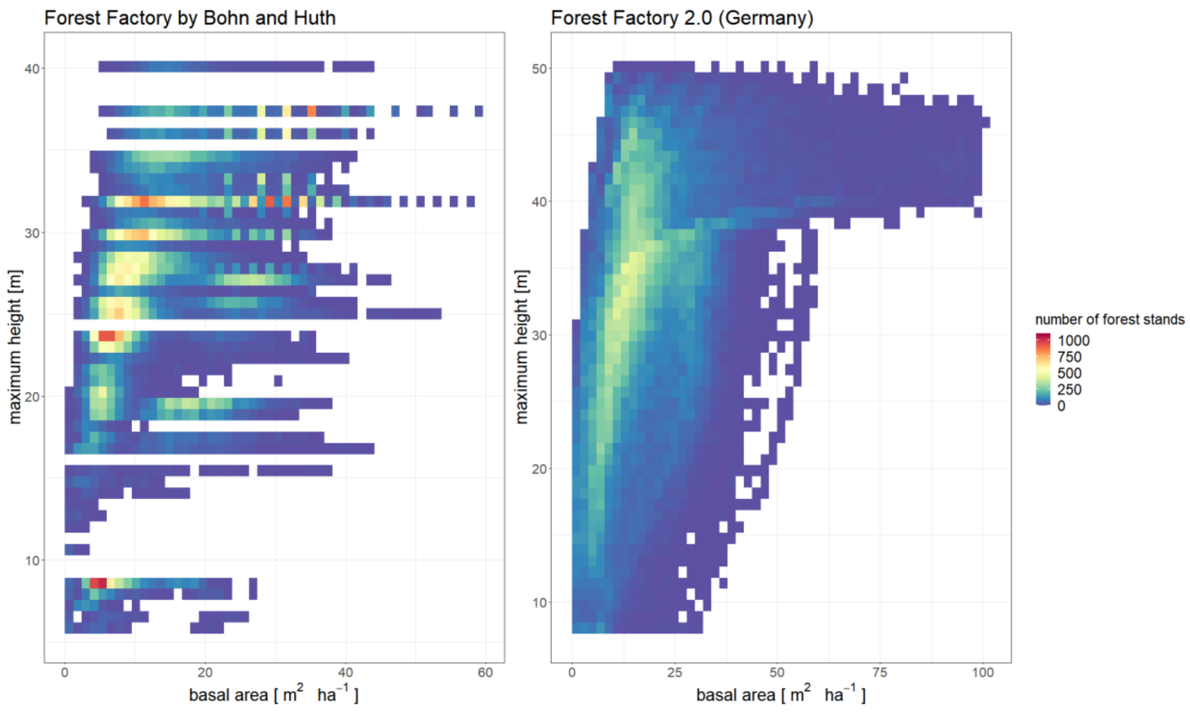

Figure A1: Comparison of Forest Factory by Bohn and Huth and Forest Factory 2.0. Distribution of forest properties for the generated forests of the Forest Factory by Bohn and Huth (2017) (left) and Forest Factory 2.0 (right). We used here as structural properties basal area and maximum height per forest stand. The color of each cell in the graph represents the number of forest stands with the respective value of the properties.

II. Additional results for biomass, productivity and species evenness of forest stands

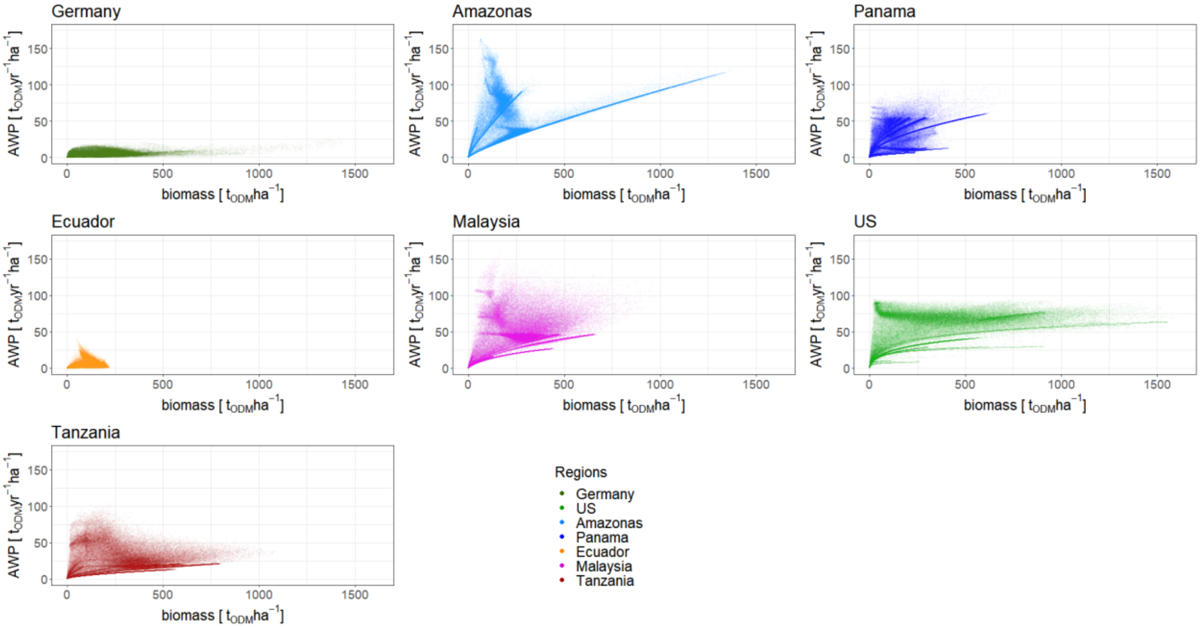

Figure A2: Relationship between biomass and above-ground productivity for analyzed ecoregions. Each point represents a forest stand in the respective ecoregion (100,000 per ecoregion).

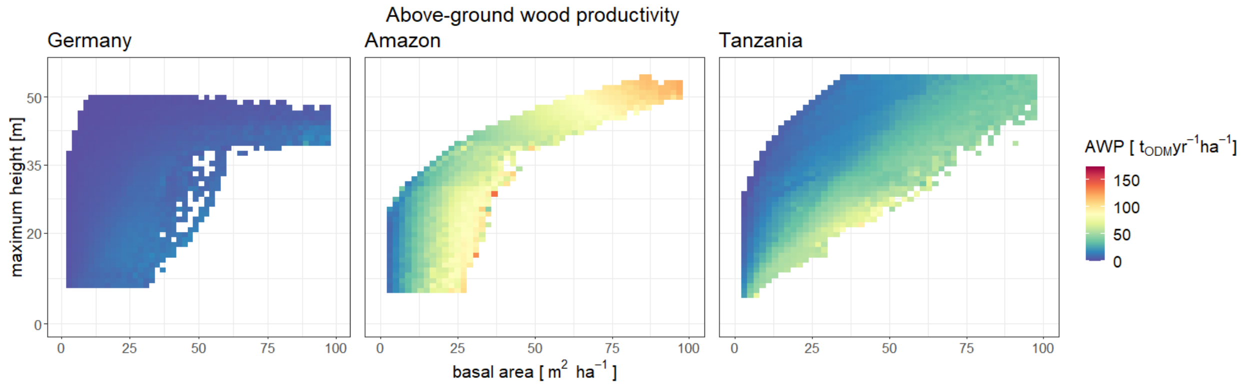

Figure A3: Relationship between structural properties and above-ground wood productivity. We analyzed here structural properties (basal area and maximum height per forest stand) and above-ground productivity for three selected ecoregions (Germany, Amazon, Tanzania). The color of each cell in the graph represents the mean value of the AWP of all forest stands within one cell. In difference to Figure 2 (main text), we use for AWP always the same color legend (for all three ecoregions).

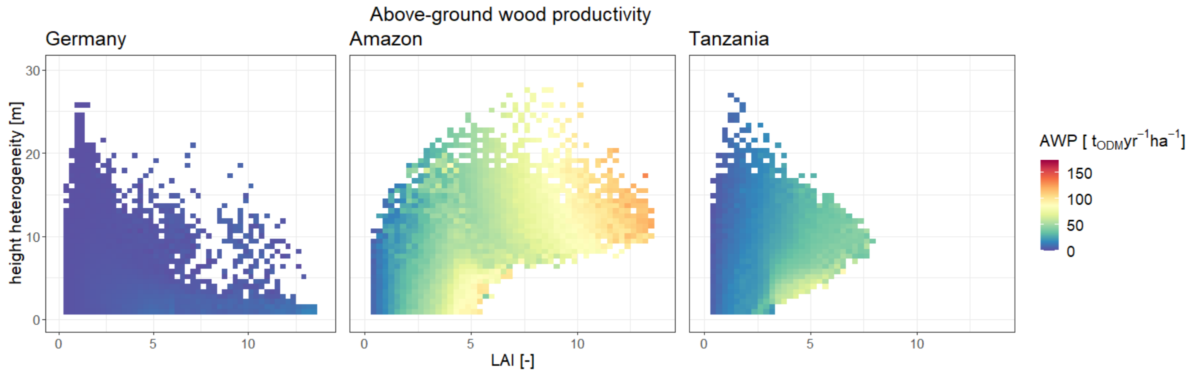

Figure A4: Relationship between structural properties and above-ground wood productivity. We analyzed here structural properties (height heterogeneity and LAI per forest stand) and above-ground productivity for three selected ecoregions (Germany, Amazon, Tanzania). The color of each cell in the graph represents the mean value of the AWP of all forest stands within one cell. In difference to Figure 2 (main text), we use for AWP always the same color legend (for all three ecoregions).

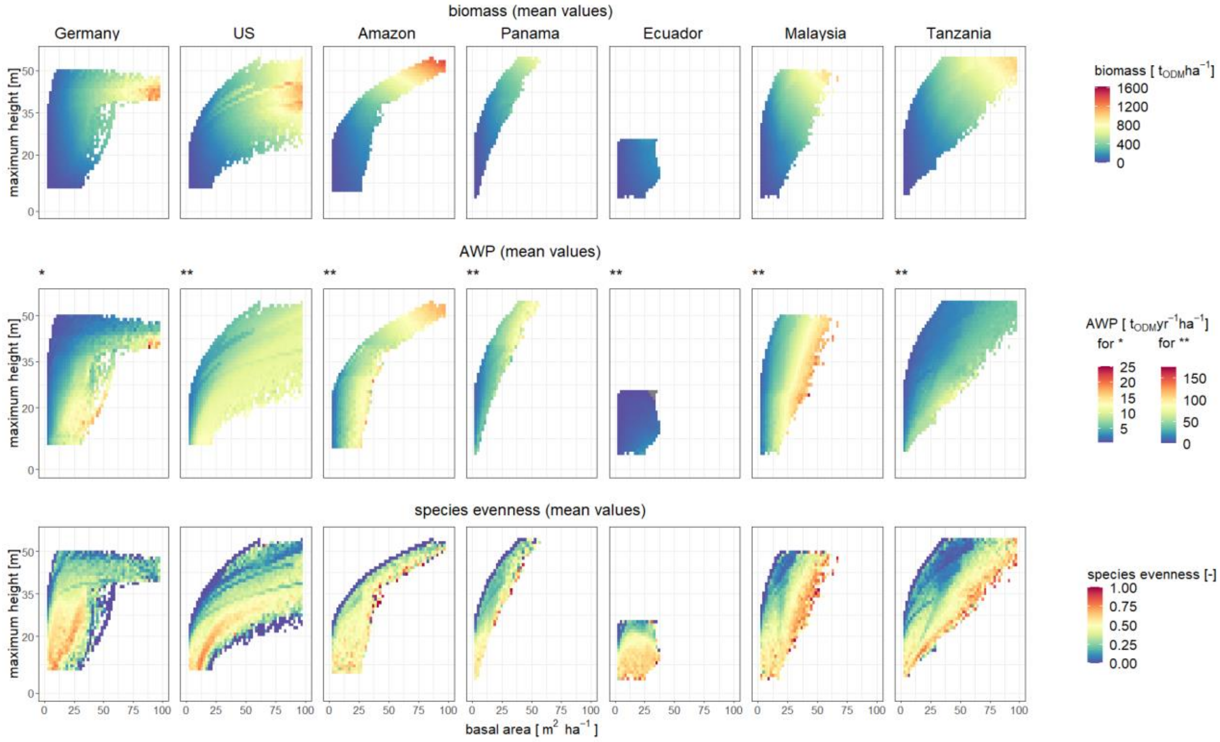

Figure A5: Relationship between structural properties and functional characteristics of forest stands. Relationship between structural properties (basal area and maximum height per forest stand) and biomass (first row), above-ground productivity (second row) and species evenness (third row) for all analyzed ecoregions. The color of each cell in the graph represents the mean value of the investigated property of all forest stands within one cell.

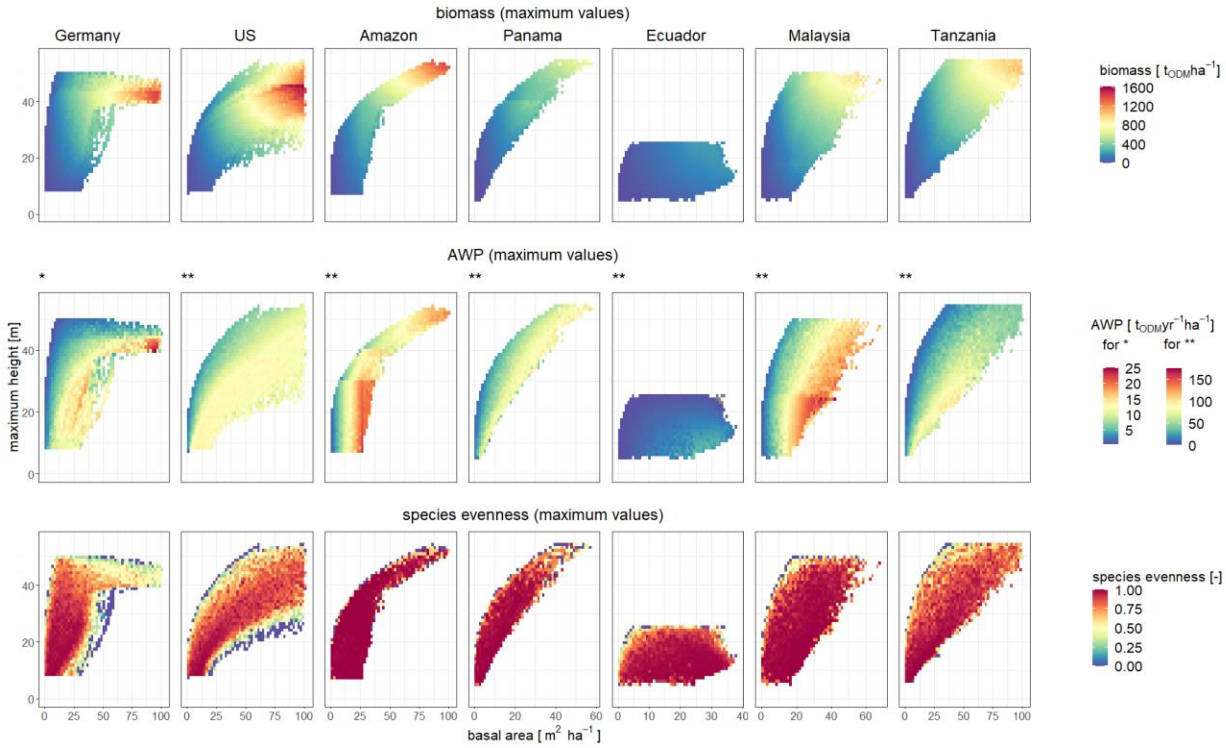

Figure A6: Maximum values of functional characteristics of forest stands. Relationship between structural properties (basal area and maximum height per forest stand) and biomass (a-g), above-ground productivity (h-n) and species evenness (o-u) for all analyzed ecoregions. The color of each cell in the graph represents the maximum value of the investigated property of all forest stands within one cell.

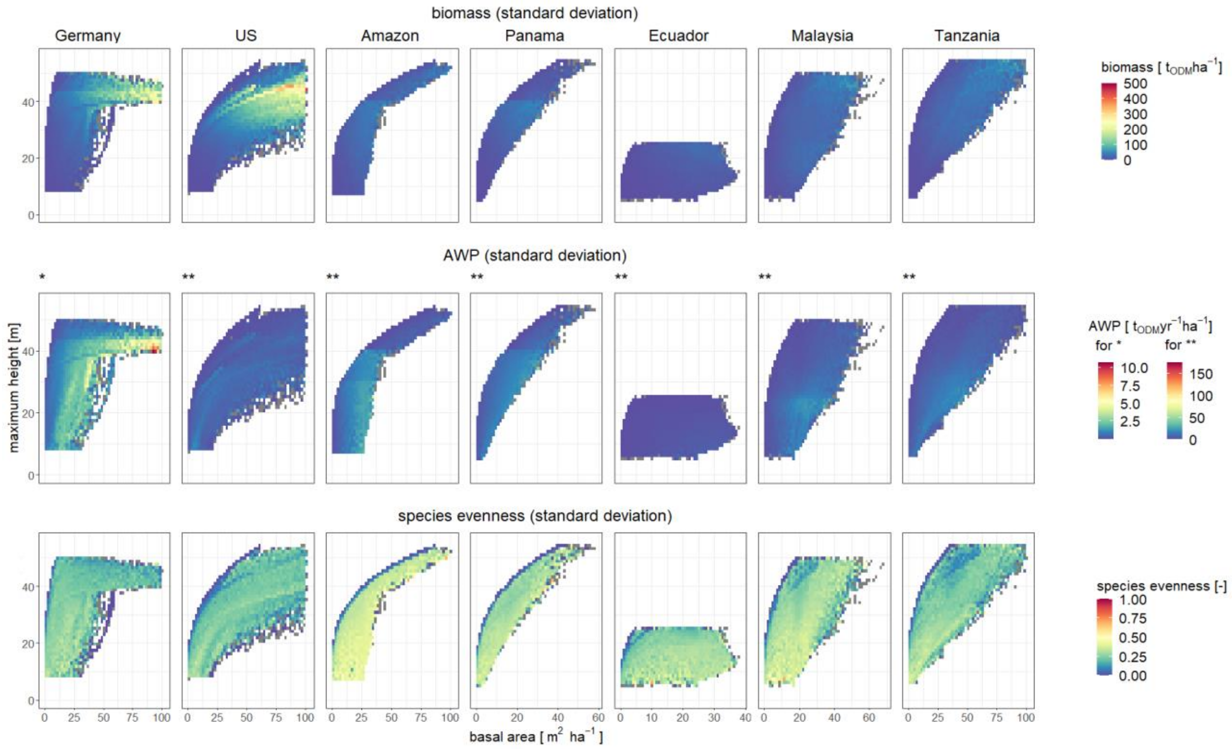

Figure A7: Standard deviation of functional characteristics of forest stands. Relationship between structural properties (basal area and maximum height per forest stand) and biomass (a-g), above-ground productivity (h-n) and species evenness (o-u) for all analyzed ecoregions. The color of each cell in the graph represents the standard deviation of the investigated property of all forest stands within one cell. Cells consisting of only one forest stand has no standard deviation (grey color).

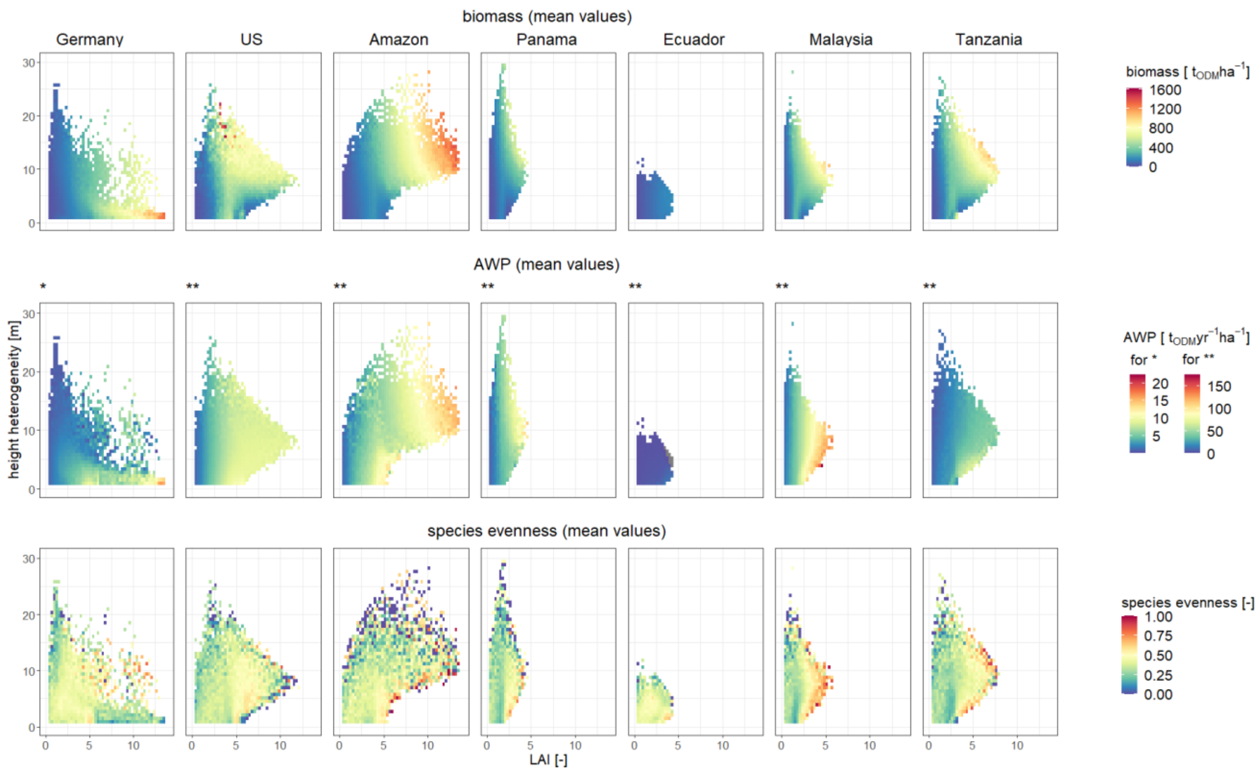

Figure A8: Relationship between structural properties and functional characteristics of forest stands. Relationship between structural properties (LAI and height heterogeneity per forest stand) and biomass (a-g), above-ground productivity (h-n) and species evenness (o-u) for all analyzed ecoregions. The color of each cell in the graph represents the mean value of the investigated property of all forest stands within one cell.

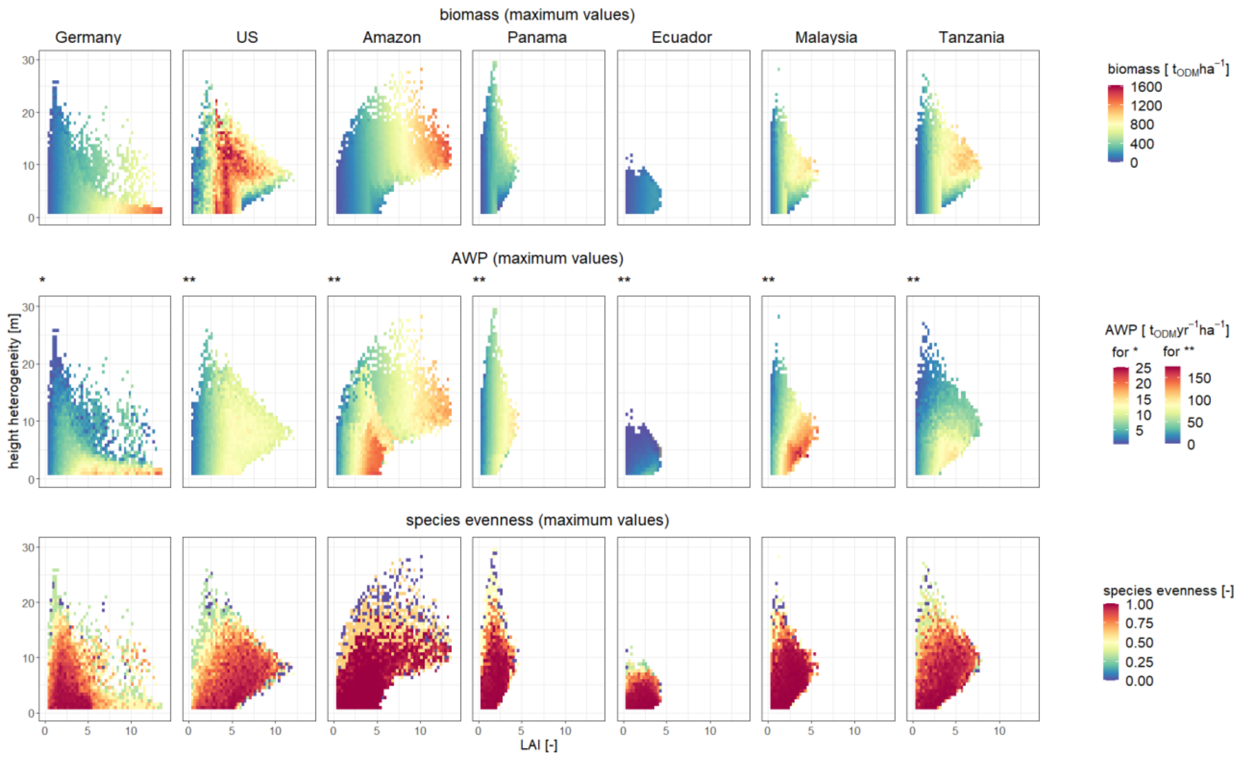

Figure A9: Maximum values of functional characteristics of forest stands. Relationship between structural properties (LAI and height heterogeneity per forest stand) and biomass (a-g), above-ground productivity (h-n) and species evenness (o-u) for all analyzed ecoregions. The color of each cell in the graph represents the maximum value of the investigated property of all forest stands within one cell.

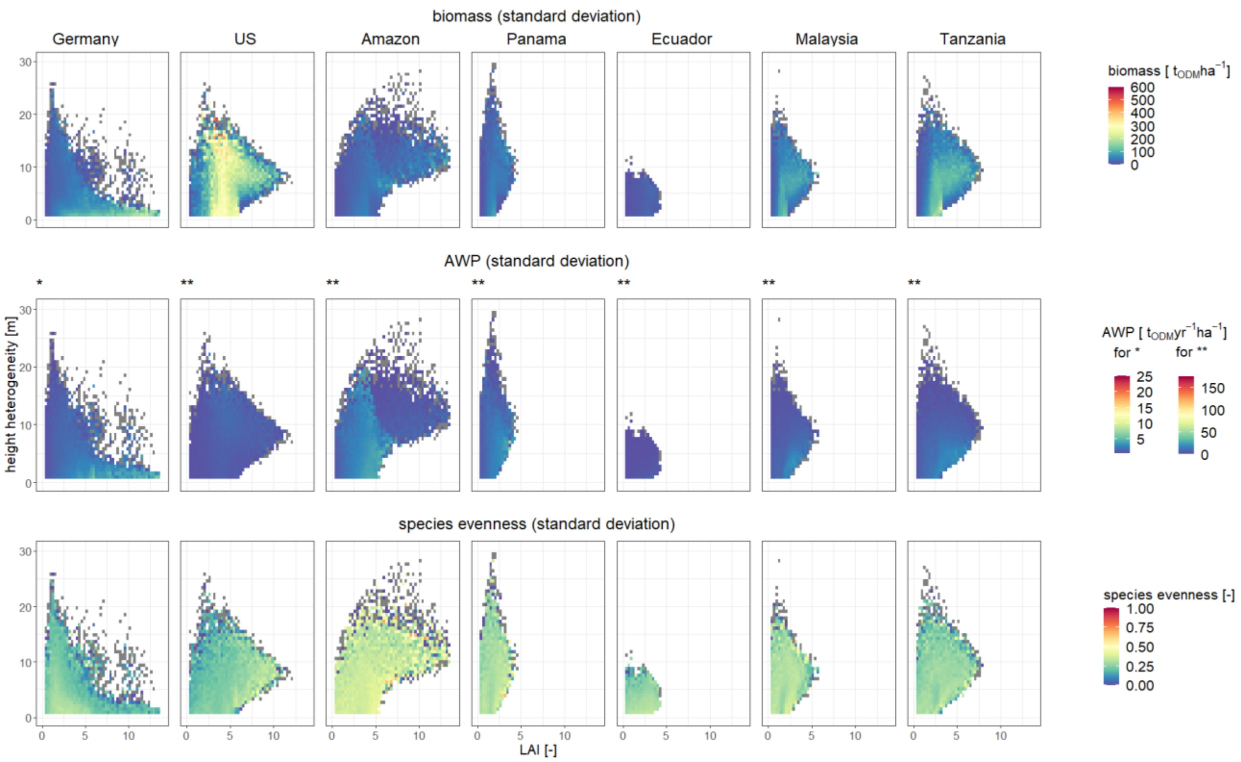

Figure A10: Standard deviation of functional characteristics of forest stands. Relationship between structural properties (LAI and height heterogeneity per forest stand) and biomass (a-g), above-ground productivity (h-n) and species evenness (o-u) for all analyzed ecoregions. The color of each cell in the graph represents the standard deviation of the investigated property of all forest stands within one cell. Cells consisting of only one forest stand has no standard deviation (grey color).

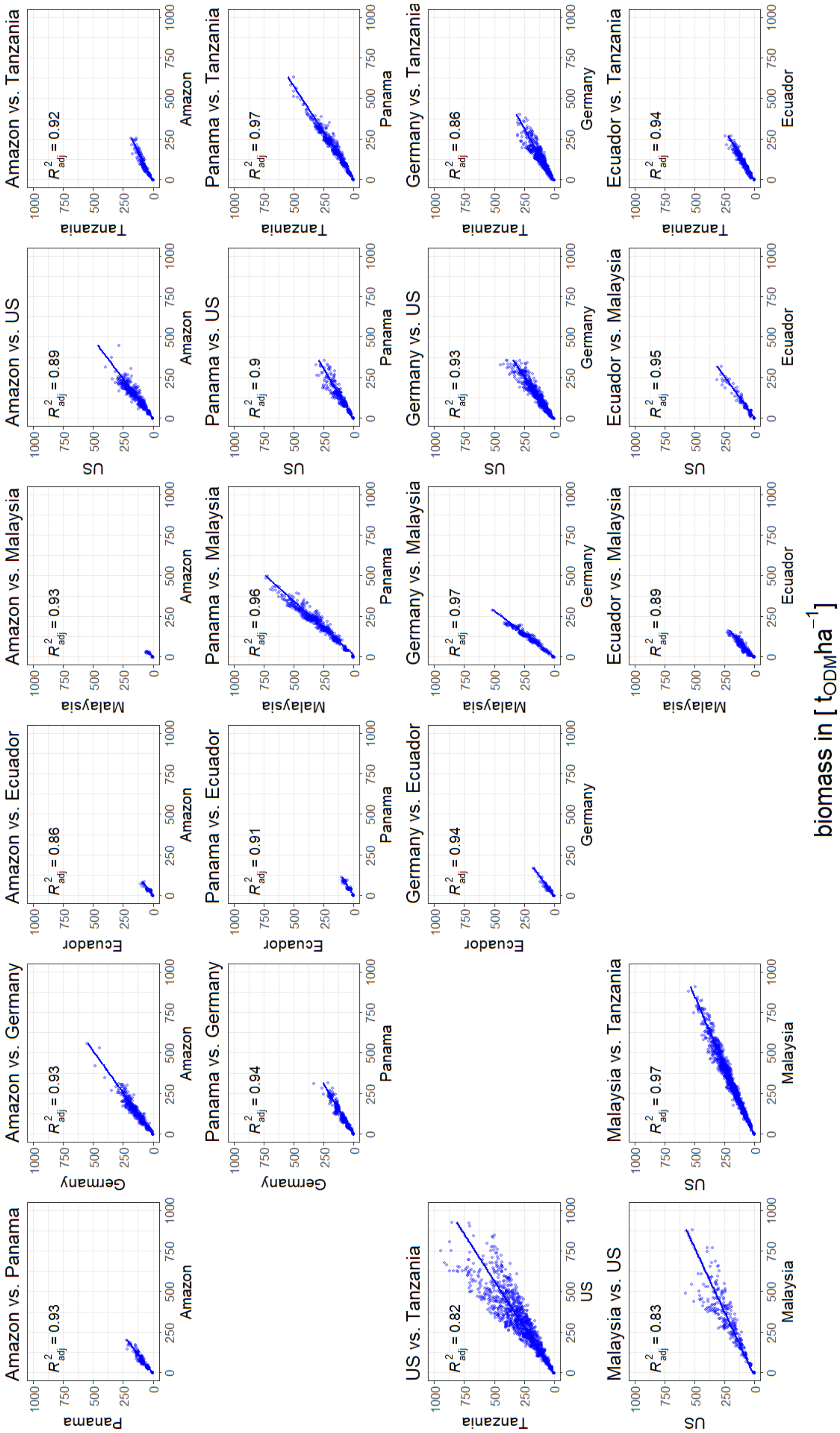

Figure A11: Comparison of biomass from different regions derived from forests with a similar state space (by a 2% quantile of the four structural properties LAI, basal area, tree height heterogeneity and maximum tree height). Each graph compares forest stands out of two ecoregions. We show the pairwise comparisons for all ecoregions. Each point represents the mean values of biomass $[t_{odm}ha^{-1}]$ for both regions.

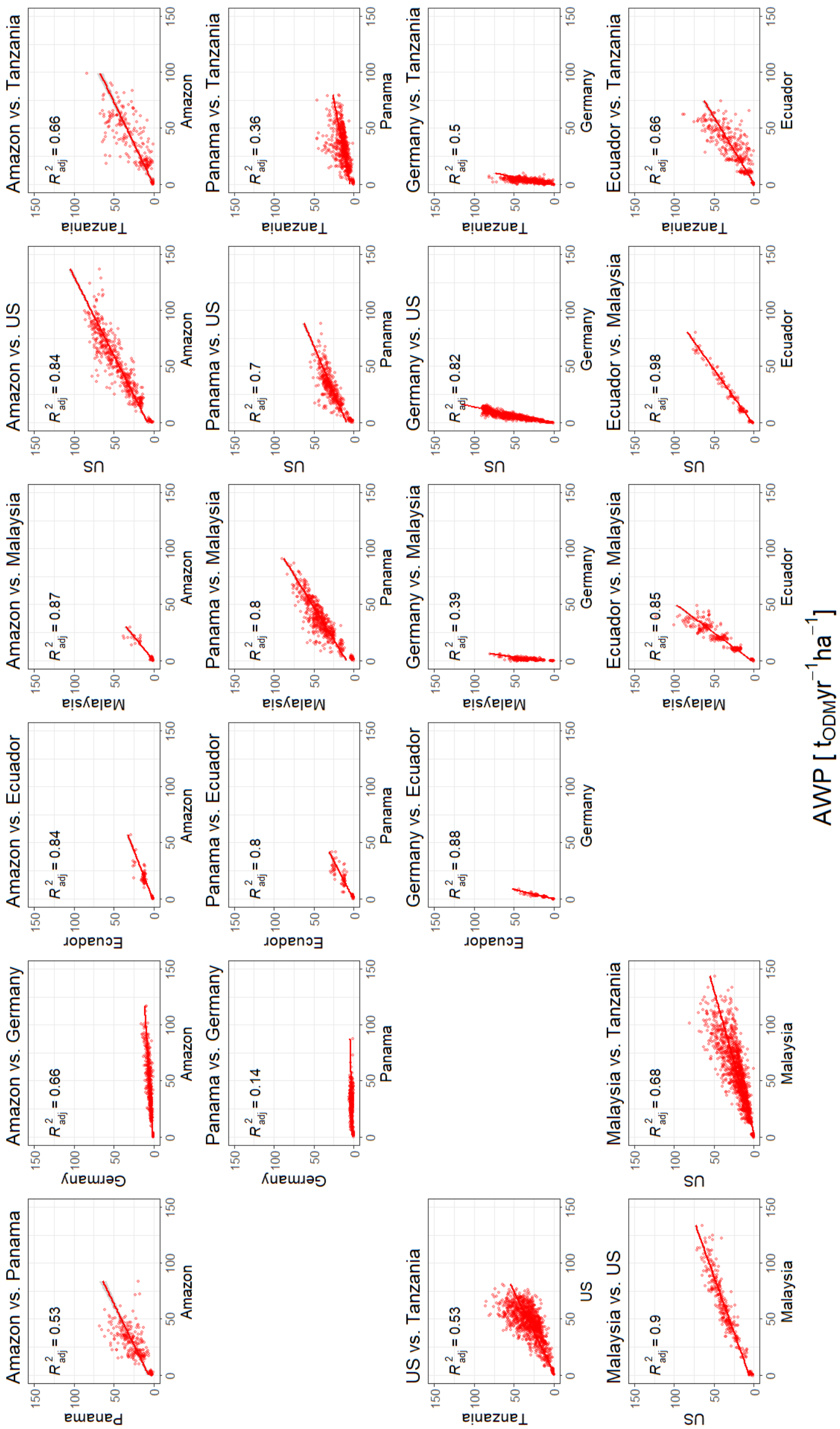

Figure A12: Comparison of aboveground productivity derived from forests with a similar state space (by a 2% quantile of the four structural properties LAI, basal area, tree height heterogeneity and maximum tree height). Each graph compares forest stands out of two ecoregions. We show the pairwise comparisons for three illustrative ecoregions. Each point represents the mean values of AWP $[t_{odm}yr^{-1}ha^{-1}]$ for both regions.

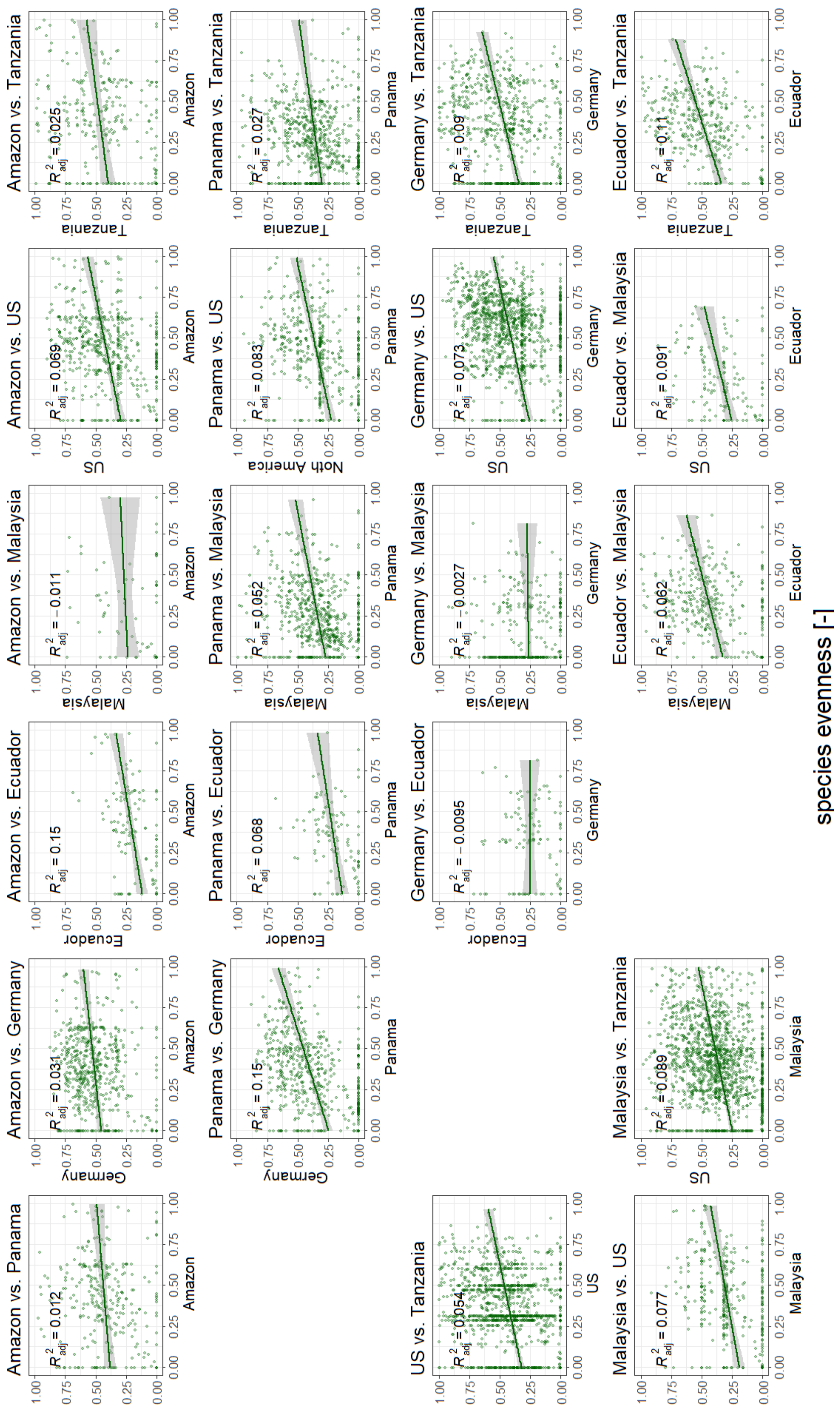

Figure A13: Comparison of species evenness derived from forests with a similar state space (by a 2% quantile of the four structural properties LAI, basal area, tree height heterogeneity and maximum tree height). Each graph compares forest stands out of two ecoregions. We show the pairwise comparisons for three illustrative ecoregions. Each point represents the mean values of species evenness $[-]$ for both regions.