Dear Editors,

Please find below a point-by-point response to all referee comments specifying all changes in the revised manuscript. The revised manuscript including changes marked in red can be found at the end of this document.

On behalf of the co-authors,

Anja Rammig

Response to Executive Editor Comment:

In my role as Executive editor of GMD, I would like remind you of our Editorial, which states that "All papers must include a section, at the end of the paper, entitled 'Code availability'. Here, either instructions for obtaining the code, or the reasons why the code is not available should be clearly stated. It is preferred for the code to be uploaded as a supplement or to be made available at a data repository with an associated DOI (digital object identifier) for the exact model version described in the paper. Alternatively, for established models, there may be an existing means of accessing the code through a particular system. In this case, there must exist a means of permanently accessing the precise model version described in the paper. In some cases, authors may prefer to put models on their own website, or to act as a point of contact for obtaining the code. Given the impermanence of websites and email addresses, this is not encouraged, and authors should consider improving the availability with a more permanent arrangement. After the paper is accepted the model archive should be updated to include a link to the GMD paper." Thus please add for all used models explicitly how to access the code. Additionally, the reference of Brienen et al. 2014 is unfortunately not part of the reference list.

Response: Thank you for your comment. We have included in section 6: "All models are described in more detail in the supplementary material. The model code for LPJmL is available through PIK's gitlab server at https://gitlab.pik-potsdam.de/lpjml/LPJmL and archived under https://doi.org/10.5880/pik.2018.002. The model code for ORCHIDEE is available under http://dx.doi.org/10.14768/06337394-73A9-407C-9997-0E380DAC5597. The model code for JULES is

1

available from the JULES FCM repository: https://code.metoffice.gov.uk/trac/jules (registration required). The model code for INLAND is available from http://www.ccst.inpe.br/wp-content/uploads/inland/inland2.0.tar.gz. The permanent archive of the observational data from Mitchard et al. (2014) can be accessed at http://dx.doi.org/10.5521/FORESTPLOTS.NET/2014_1, see also Lopez-Gonzales et al. (2014). The inventory data from Brienen et al. (2015) are available at http://dx.doi.org/10.5521/ForestPlots.net/2014_4, see also Brienen et al. (2014)."

We have included the reference to Brienen et al. 2014 in the reference list.

Response to Referee Comment #1:

I have found this paper quite interesting as how to approach scale mismatch. The mathematical/statistical procedures are rigorous, and appropriate to derive indexes.

Response: Thank you for the very positive evaluation of our manuscript. Below we provide a point-to-point reply to your comments.

I only suggest including brief comments or modifying the following:

Do you think that DGVM models might increase resolution by using remote sensing data in order to improve spatial variability? Such as MODIS GPP/NPP at 1km.

Response: The resolution of DGVMs is always defined by the resolution of the climate input data. Spatial variability in DGVMS could hence only be increased by using input climate data at higher resolution (e.g., from regional climate models). We have explained this in more detail on p.6, l.5. As simulated GPP and NPP are a model output, MODIS GPP and NPP could only be used to evaluate model output. However, note that MODIS is derived from NDVI which tends to saturate in dense forests such as the Amazon and, thus, has limited capability to identify spatial variations (Hall et al 2011). As pointed out in the conclusion (page 9, line 30) the integration of forest structural information derived from upcoming Lidar remote sensing missions (such as GEDI or BIOMASS) may reveal more information on spatial variability than previous optical remote sensing (NDVI).

I see more emphasis in spatial variability than temporal variability. How to assess the variability of the observed/estimated forest properties (as biomass) due to gap dynamics? Is this more important than edaphic properties or Site Index?

Response: Thank you for pointing this out. Gap dynamics clearly have an effect on the spatial variability of biomass (e.g. see Rödig et al. 2017, 2018) and this is discussed in our manuscript on page 9, l.15-18. However, we emphasise that most gaps in tropical forests are much smaller than the one hectare scale of the plot data used for model validation (e.g. 99% of gaps on BCI are <0.04 ha; Hubbell et al. 1999), and so the plots are generally representative of the disturbance regime in each landscape. Additionally, gap dynamics and site properties cannot be separated easily as site characteristics strongly influence forest dynamics (e.g. Quesada et al. 2011). We have made this more clear in the introduction by rephrasing p.2, l.18-22. We agree that it would be interesting to assess the temporal variability and the respective model behaviour, but this is out of the scope of this study.

Do you consider that some of the forest plots you used were subject to selective logging or natural gaps? and for this reason a cause of mismatch.

Response: Thank you for pointing this out. The plot data are taken from Brienen et al. (2014, 2015) and Mitchard et al. (2014) who excluded all plots that were subject to anthropogenic disturbances, including selective logging, so this is not a cause of the mismatch. As above, the spatial scale of the plot data means that they incorporate most size classes of natural gaps, particularly as the plot data were averaged across sites occurring within the same pixel. Across the plot network, the biases introduced into estimate of carbon balance by one hectare plots not sampling the very largest and rarest natural gaps are in fact very small (Espirito-Santo et al. 2014). We have revised on p.5 the section "2.2.1 Observed data at point scale: Description of site-level data".

Can you provide a reference of studies calculating a "pixel-wise within-pixel variability"? See line 1 page 10.

Response: To our knowledge, there is no previous study that calculates pixel-wise within-pixel variability. This is the reason why we propose this in our manuscript. We have revised the text in l. 1, p.10 to "… which could be used to calculate a pixel-wise within-pixel variability based on our approach." to make this more clear.

Page 2, Line 27, do you mean "uneven spatial distribution and -temporal- variability due to natural gap dynamics"

Response: We mean "uneven spatial distribution and spatial variability due to natural gap dynamics" and have changed it accordingly in the manuscript.

Page 3, line 10, there is a typo error.

Response: Thank you, is corrected.

Page 3 line 24, what do you mean by "resolution of the study area"

Response: Thank you, this is indeed not fully clear in our text. We have revised the sentence to "Landscape variability depends on the extent and heterogeneity of the study area (Turner et al., 2001). Point measurements within a pixel of larger spatial scale, for example, may reveal small-scale spatial variabilities within the pixel."

I found figure 1 misleading. At some point, the distance between plots (in b) should depict what "corresponds to the size of the grid cell" (in c). Or maybe explaining this better in the caption.

Response: Yes, we agree that Figure 1b is misleading. We have revised Figure 1 so that the red arrows indicate better that they correspond with the size of the grid cell and we have expanded the caption of Figure 1.

Page 5, line 30. I assume is 2.1.3 instead of 1.3

Response: Yes, agreed, we have corrected that.

Response to Referee Comment #2:

The present paper addresses an important problem in model data comparisons, namely how to compare measurements at the plot scale with pixel level model predictions, given the well-known spatial heterogeneity in the measured variables. The authors propose a straightforward statistical framework that can take into account within pixel variations. The authors exemplify the use of this model using aboveground biomass (AGB) measurements and model predictions in the Amazon region. They show that by using the new metric for data variation it is easier to show whether the model predictions match the observed variability.

This core part of the paper is very clear and, in my opinion, very valuable for future modelling studies. The paper in general is very well written. There are however some parts of the paper that are not so well developed, in particular the model comparison with the other datasets. In my opinion, the authors

can go one of two ways: either cut down to the core of their method and the AGB data, or extend the less well developed and explained parts of their paper.

Response: Thank you very much for the positive feedback on our manuscript. We appreciate the suggestion for improving the part of the manuscript about the model comparison with other datasets and have expanded it. Please see below our response to your detailed comments.

Detailed comments

Description of observed data (section 2.2.1) I find the data description in the main text of the paper extremely short. While the concept of AGB might be widely used and easily understandable, woody productivity and woody loss are not and a brief definition and description of how these were measured/calculated would greatly help the reader understand the subsequent analyses.

Response: Yes, we agree and we thank the reviewer for pointing this out. We have described the observational data in more detail in section 2.2.1, in particular how woody productivity and woody loss were measured and calculated. Brienen et al. (2015) derived "…forest woody productivity… from the sum of biomass growth of surviving trees and trees that recruited (that is, reached a diameter ≥ 100 mm),and mortality [=woody loss] from the biomass of trees that died between censuses. We have added this accordingly in our methods section on p. 5, sect. 2.2.1

Analysis of the woody productivity and loss. The AGB data and mode predictions are analysed in detail and presented in four different figures, while the other two datasets have one figure each and one joint table. I would find it interesting to see a bit more detail about these observations too, especially since one of the strong discussion points (section 4.3) revolves around the model's inability to predict productivity and loss.

Response: Thank you for your comment. We agree that woody productivity and loss are also interesting. However, we decided to focus on mainly AGB in the main manuscript as an example to present our approach. We prefer to keep the detailed figures in the supplementary material.

Different allometric models. While I fundamentally understand why the choice of allometric model is important for estimates of AGB, it does not feel like this additional dimension adds to the central message of the study. Most of the detail for the allometric models is buried in the supplementary

material and, as far as I understand, the majority of the analysis has been done with only one of the allometric equations.

Response: Thank you for pointing this out. However, we feel that the uncertainty introduced by allometric relations for the comparison of field observations and DGVM simulations is an aspect that needs to be part of the discussion of our paper. We have now added more detail on the allometric equations in the methods section (p. 5, sect. 2.2.1) and we have improved the discussion about the relevance of the allometric equations on p. 8 and 9.

A generic pixel-to-point comparison for simulated large-scale ecosystem properties and ground-based observations: an example from the Amazon region

Anja Rammig[1,*], Jens Heinke[2,*], Florian Hofhansl[3], Hans Verbeeck[4], Timothy R. Baker[5], Bradley Christoffersen[6], Philippe Ciais[7], Hannes De Deurwaerder[4], Katrin Fleischer[1], David Galbraith[5], Matthieu Guimberteau[7], Andreas Huth[8], Michelle Johnson[5], Bart Krujit[9], Fanny Langerwisch[2], Patrick Meir[10,11], Phillip Papastefanou[1], Gilvan Sampaio[12], Kirsten Thonicke[2], Celso von Randow[12], Christian Zang[1], Edna Rödig[8]

[1]Technical University of Munich, TUM School of Life Sciences Weihenstephan, Hans-Carl-von-Carlowitz-Platz 2, 85356 Freising, Germany, Anja.Rammig@tum.de
[2]Potsdam Institute for Climate Impact Research, Potsdam, Germany
[3]IIASA International Institute for Applied Systems Analysis; Schlossplatz 1, A-2361 Laxenburg, Austria
[4]CAVElab Computational & Applied Vegetation Ecology, Department of Applied Ecology and Environmental Biology, Faculty of Bioscience Engineering, Gent, Belgium
[5]School of Geography, University of Leeds, Leeds UK
[6]Department of Biology, The University of Texas Rio Grande Valley, Edinburg, USA
[7]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[8]Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
[9]ALTERRA, Wageningen-UR, Wageningen, The Netherlands
[10]School of Geosciences, University of Edinburgh, Edinburgh, UK
[11]Research School of Biology, Australian National University, Canberra, Australia
[12]INPE, Sao Jose dos Campos, SP, Brazil
* both authors contributed equally to the manuscript.

Correspondence to: Anja Rammig (Anja.Rammig@tum.de)

Abstract. Comparing model output and observed data is an important step for assessing model performance and quality of simulation results. However, such comparisons are often hampered by differences in spatial scales between local point observations and large-scale simulations of grid-cells or pixels. In this study, we propose a generic approach for a pixel-to-point comparison and provide statistical measures accounting for the uncertainty resulting from landscape variability and measurement errors in ecosystem variables. The basic concept of our approach is to determine the statistical properties of small-scale (within-pixel) variability and observational errors, and to use this information to correct for their effect when large-scale area averages (pixel) are compared to small-scale point estimates. We demonstrate our approach by comparing simulated values of aboveground biomass, woody productivity (woody net primary productivity, NPP) and residence time of woody biomass from four dynamic global vegetation models (DGVMs) with measured inventory data from permanent plots in the Amazon rainforest, a region with the typical problem of low data availability, potential scale mismatch and thus, high model uncertainty. We find that the DGVMs under- and overestimate aboveground biomass by 25% and up to 60%, respectively. Our comparison metrics provide a quantitative measure for model-data agreement and show moderate to good agreement with

the region-wide spatial biomass pattern detected by plot observations. However, all four DGVMs overestimate woody productivity and underestimate residence time of woody biomass even when accounting for the large uncertainty range of the observational data. This is because DGVMs do not represent the relation between productivity and residence time of woody biomass correctly. Thus, the DGVMs may simulate the correct large-scale patterns of biomass but for the wrong reasons. We conclude that more information about the underlying processes driving biomass distribution are necessary to improve DGVMs. Our approach provides robust statistical measures for any pixel-to-point comparison, which is applicable for evaluation of models and remote sensing products.

1 Introduction

The rate of environmental change in tropical South America and in particular in the Amazon region has been unprecedented in the last decades (e.g. Lewis *et al.* 2011; Davidson *et al.* 2012). Estimates of the amount of carbon stored in tropical rainforest biomass differ strongly (Avitabile et al., 2016;Baccini et al., 2012;Saatchi et al., 2011;Mitchard et al., 2014). In addition, estimated carbon release to the atmosphere from land-use change is uncertain (e.g. Houghton et al., 2012;Baccini et al., 2017;Song et al., 2015;Harris et al., 2012). Nonetheless, a successful implementation of protection incentives, e.g. for reducing emissions from deforestation and degradation (REDD+), requires both accurate estimates of existing regional carbon stocks as well as improved projections of future scenarios (e.g. Langner et al., 2014).

Dynamic global vegetation models (DGVMs) are important tools to estimate impacts of climate and land-use change on the carbon cycle (e.g. Cramer *et al.* 2004; Sitch *et al.* 2008). To correctly capture carbon dynamics in tropical forests, DGVMs need to improve the simulation approach of drought-related mortality and other types of tree mortality that control stand density (Pillet *et al.* 2018), they need to include more detailed gap dynamics influencing stand dynamics (Espírito-Santo et al., 2014;Rödig et al., 2017), and they need to incorporate how nutrient availability limits woody productivity (Quesada *et al.* 2012; Johnson *et al.* 2016). At the same time, model evaluation based on available data is necessary for which the primary source are ground-based observations of AGB obtained from forest census data (Lopez-Gonzalez et al. 2011; Brienen et al. 2014; Lopez-Gonzalez et al. 2014; Mitchard et al. 2014; Brienen et al. 2015; Johnson et al. 2016). When conducting data-model comparisons at the plot scale, the spatial resolution of both get into focus. The size of forest plots is typically in the order of 1 ha or less, whereas average DGVM grid-cell resolution is determined by the available gridded climate data set, which is usually about several thousand square kilometers (>100,000 ha). Plot observations are affected by observational errors, uneven spatial distribution (Saatchi et al. 2015) and spatial variability due to natural gap dynamics (Chambers *et al.* 2013), and thus are likely to exhibit substantial deviation from average large-scale properties. The problem of comparing point data with model results obtained at grid-cell (pixel) size occurs in many applications of remote sensing and ecological modelling.

So far, we lack a reliable and objective method to compare simulation results from DGVMs at grid-cell scale (pixel) and plot (point) observations. Several studies, that evaluated patterns of interpolated maps from plot data and model simulations,

concluded that the observed and simulated spatial patterns do not match (e.g. Johnson *et al.* 2016). Here, we complement these findings by providing quantitative statistical measures for such comparisons and present an approach for performing pixel-to-point comparisons while accounting for different statistical properties of point and pixel values and its uncertainties. The basic concept of the approach is to determine statistical properties from small-scale variability and observational errors in ecosystem

5   variables, in order to account for these effects when comparing large-scale area averages (pixels) and small-scale plot estimates (points).

We apply our approach by comparing point estimates of ecosystem properties obtained from forest inventories (Mitchard *et al.* 2014; Brienen *et al.* 2015) to corresponding simulated pixel values from four state-of-the-art DGVMs. Similar to Johnson et al. (2016), we focus on three ecosystem properties that are well defined and represented in both, DGVMs and forest

10   inventories: (i) aboveground biomass (AGB, in Mg C ha^{-1}); (ii) aboveground woody productivity (WP, in Mg C ha^{-1} yr^{-1}); and (iii) residence time of woody biomass (τ in years).

We evaluate the accuracy of the spatial pattern of these three ecosystem properties and provide statistical measures for the quality of model simulations in comparison to observations, thereby accounting for small-scale landscape variability and associated measurement errors. We demonstrate the strength of our approach by highlighting its applicability for model

15   evaluation and model benchmarking.

In particular, we address three research questions:

1)   How well do models represent variations in aboveground biomass across the Amazon region? We expand the pure (visual) qualitative comparison by deriving three statistical metrics for a quantitative comparison.

2)   How can we evaluate differences between observed and simulated spatial biomass patterns (based on the presented

20   metrics), in particular when considering different allometric equations? We discuss the effects of inadequate data and associated model uncertainty.

3)   What can we learn from the spatial heterogeneity and underlying drivers of spatial biomass patterns? We analyze simulated and observed patterns of WP and τ.

2 Methods

25   Landscape variability depends on the extent and heterogeneity of the study area (Turner et al., 2001). Point measurements within a pixel of larger spatial scale, for example, may reveal small-scale spatial variabilities within the pixel. We derive a "within-pixel variability" that so far has not been accounted for in earlier approaches. We present three steps to calculate three metrics that provide a measure on the best achievable correlation between point and pixel values (see Fig. 1).

3

2.1 A generic method for point-to-pixel comparisons

2.1.1 Calculate the "global variability" across the region of interest

Assume, we have a dataset X with a number N of point observations x_i at location i (e.g. plot observations from inventory data). In the first step, we calculate the mean $\bar{x}$ and variance σ_x^2 across all plots in a region (e.g. across the Amazon region).

5   The variance σ_x^2 denotes the global variability (i.e. the variability across the whole Amazon region) at point scale (Fig. 1a).

2.1.2 Calculate within-pixel variability

In the second step, we identify within-pixel variability from point measurements. With coarser pixel resolution, the spatial variability (here: global variability) is reduced. In order to compare pixel values against point values, global variability at point scale needs to be reduced by the within-pixel variability (variability component ε_i; Fig. 1b).

10   The variability component ε_i is assumed to be normally distributed with zero mean and variance σ_ε^2:

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2) \tag{1}$$

Based on the variability component ε_i, we estimate the within-pixel variance σ_ε^2 from the point observations by analysing their covariance, which is equivalent to the nugget effect (i.e. the sum of variance caused by small-scale variability and observation error) in a semivariogram (see SI Methods). Due to the limited amount of inventory data, we assume here, that σ_ε^2

15   is stationary across the region of interest (for details on that assumption see discussion and SI Results).

The global variance at point scale σ_x^2 now differs from the corrected global variance at pixel scale, $\sigma_{x,corr}^2$, as variances add quadratically, assuming that the (small scale) variability component ε_i has errors uncorrelated to the global distribution of x :

$$\sigma_{x,corr}^2 = \sigma_x^2 - \sigma_\varepsilon^2 \tag{2}$$

2.1.3 Metrics for the comparison of two datasets with different spatial resolutions

20   In a third step, we compare the point data x_i with simulated data y_i at pixel scale. Similar to the above procedure, we calculate the mean $\bar{y}$ and variance σ_y^2 for the simulated pixels that contain point observations (hereby we assign each point observation the pixel value in which the point is located). We then compare the simulation results by applying three metrics:

1) Mean bias (MB): the ratio of means $\bar{y}/\bar{x}$ across the whole region as a measure of the mean bias in the patterns which is not affected by small scale variability;

25   2) Pattern amplitude (PA): the ratio of standard deviations $\sigma_y/\sigma_{x,corr}$ using the corrected global variability (i.e. removed within-pixel variability) and serves as a measure of differences in pattern amplitude or in the variability of the simulated and observed data;

3) Similarity of pattern (SP): We use r_{corr} as a measure of the similarity of the 'shape' of spatial patterns, i.e. the spatial correlation of simulated and observed data (see SI). Accordingly, we can calculate the maximum achievable

correlation coefficient r_{max}, which is derived from correlating the observational data set at point scale to the same observational data set at pixel scale (see Fig. 1a, b and SI).

The limited number of point observations and their non-random spatial distribution in the Amazon region affects the accuracy of the comparison. We therefore estimate confidence intervals for the comparison metrics MB, PA and SP, respectively, by applying a bootstrapping technique (10,000 repetitions). Because the estimation of σ_ε^2 is based on the analysis of the spatial correlation structure of the data, a block-bootstrapping is performed (Politis and Romano, 1994). For each permutation, the domain of observations is randomly divided into 100 tiles (random orientation and offset, ca. pixel size) from which a random recombination is drawn with replacement. This technique assures that the spatial correlation structure of the data remains intact.

2.2 Application of the pixel-to-point comparison to simulated and observed data from the Amazon region

2.2.1 Observed data at point scale: Description of site-level data

The observed data at point scale are forest census-based plot measurements across the Amazon region, in which all plots that were subject to anthropogenic disturbances, including selective logging were excluded (Brienen et al. 2015). The average plot size is ~1.2 ha (Brienen et al. 2015) so that the plots incorporate most size classes of natural gaps, particularly as the plot data were averaged across sites occurring within the same pixel. Across the plot network, the biases introduced into estimate of carbon balance by one hectare plots not sampling the very largest and rarest natural gaps are in fact very small (Espirito-Santo et al. 2014). We use datasets of AGB (Lopez-Gonzalez et al., 2011;Lopez-Gonzalez et al., 2014;Mitchard et al., 2014), WP and woody loss (WL; Brienen et al., 2014). WP and WL are derived "[…] from the sum of biomass growth of surviving trees and trees that recruited (that is, reached a diameter $\geq$ 100 mm), and mortality [=woody loss] from the biomass of trees that died between censuses" (Brienen et al. 2015). We convert AGB, WP and WL from dry biomass to carbon mass (see SI Methods). For the calculation of AGB, we use different allometric equations that account e.g. for regional differences in wood density or tree height (Table S1). We exemplify our comparison metrics based on AGB calculated from the three parameter moist tropical forest allometry from Chave et al. (2005), where tree height is estimated from DBH individually for each stem based on the region-specific Weibull models from Feldpausch et al. (2012). Wood density is estimated for each stem using the mean value for the species in the Global Wood Density Database (Chave et al., 2009;Zanne et al., 2009), or the mean for the genus using congeneric taxa from Mexico, Central America and tropical South America if no data were available for that species (Mitchard et al. 2014). We here evaluate the principal AGB dataset ($K_{DH\rho}$) from Mitchard et al. (2014) in more detail (the other allometric equations are presented in the supplement).

2.2.2 Simulated data at pixel scale: Description of DGVM simulations

We use outputs from four state-of-the-art DGVMs, namely the Lund-Potsdam-Jena DGVM for managed Land (LPJmL, Bondeau et al., 2007;Gerten et al., 2004;Sitch et al., 2003), the Joint U.K. Land Environment Simulator (JULES), v. 2.1. (Best

et al., 2011;Clark et al., 2011), the INtegrated model of LAND surface processes (INLAND) model (a development of the IBIS model, Kucharik et al., 2000) and the Organising Carbon and Hydrology In Dynamic EcosystEms (ORCHIDEE) model (Krinner et al., 2005). A short description of each of the applied models is provided in the Supplementary Information (SI Methods). The models were applied to the Amazon region covering the area of 88°W to 34°W and 13°N to 25°S at a spatial resolution of 1°x 1° lat/lon. The resolution of the DGVMs is defined by the resolution of the climate input data for which we here used bias-corrected NCEP meteorological data (Sheffield et al., 2006). Model runs were performed based on the standardized Moore Foundation Andes-Amazon Initiative (AAI) modelling protocol (Zhang et al., 2015). The same set of models and output variables was analysed in Johnson et al. (2016).

2.2.3 Comparing inventory and simulation results

In our application, data set *X* corresponds to the inventory measurements at point scale (Fig. 1a). For this dataset, we have to derive the within-pixel variability (Fig. 1b). Data set *Y* corresponds to the simulated pixel values (Fig. 1c). Hence, the pixel scale is defined by the resolution of the model simulation (1°x1°, approximately 12,200 km², Fig. 1c). We calculate the three metrics (sect. 2.1.3) from the observed and simulated ecosystem variables AGB, WP and τ.

3 Results

3.1 Comparison of aboveground biomass (AGB)

The visual comparison indicates that the spatial pattern of AGB from the plots (Fig. 2a and Fig. S1) differs from the spatial patterns of AGB simulated by either DGVM (Fig. 2c-f). In addition, the DGVM patterns are vastly different among each other. Mean $\bar{x}$ and global variability σ_x of AGB for all plot observations across the Amazon region (Fig. 3a) range from 134-153 and 36-50 MgC ha^{-1}, respectively (Tab. S2), depending on the allometric equation applied. Within-pixel variability σ_ε, as calculated from Eq. S1, ranges between 28 and 36 MgC ha^{-1}. The corrected variability of observed AGB at pixel-scale ($\sigma_{x,corr}$) is thus substantially lower than the global variability and ranges between 22-39 MgC ha^{-1} (Fig. 4a, Tab. S2). Based on these estimates we calculate the maximum achievable coefficients r_{max} for a comparison between pixel averages and point estimates of 0.61 to 0.78 for different allometric equations (Fig. 5).

The models simulate a continuous cover of biomass across the Amazon region at a spatial resolution of 1°x1° pixel size. For our comparison, we only use the simulated pixel values of AGB at each plot location. Thus, the estimated statistical properties are not representative for the entire Amazon region but only for a relatively small subset of pixels (i.e. 98 pixels as in Fig. 2a/b). For simulated AGB from the four DGVMs, we estimate a mean $\bar{y}$ of 114 MgC ha^{-1} for INLAND, 151 MgC ha^{-1} for JULES, 217 MgC ha^{-1} for ORCHIDEE and 170 MgC ha^{-1} for LPJmL (Fig. 3a, Tab. S3). Depending on the allometric equation applied to calculate observed biomass (Tab. S1), INLAND underestimates mean AGB by 15-25%. LPJmL and ORCHIDEE overestimate AGB by 11-26 and 42-62%, respectively. JULES deviates only by 1% from AGB derived from the 2-parameter

allometric equations (kdr2p, kd2p), but overestimates AGB derived from all 3-parameter allometric equations by 12% (Tab. S3).

Mean global variability of simulated AGB, σ_y, ranges between 13 MgC ha^{-1} for JULES and 62 MgC ha^{-1} for ORCHIDEE (Fig. 4a and Tab. S3). Without correcting for small-scale variability σ_ε in the point-to-pixel comparison, we would conclude that

5   the pattern amplitude simulated by ORCHIDEE and LPJmL agree quite well with observed patterns (Fig. 4b). However, when accounting for the lower corrected variability ($\sigma_{x,corr}$), because the error of observation-based estimates at pixel-level is smaller, it becomes easier to falsify models with uncertain data. We find that LPJmL and ORCHIDEE both overestimate the observed spatial amplitude by 43% and 62%, respectively (Fig. 4c and see Tab. S3 for other allometric equations). For INLAND and JULES, on the other hand, we find a corresponding underestimation of pattern amplitude by 14 and 65%,

10   respectively. We also note that confidence intervals for $\sigma_y/\sigma_{x,corr}$ are large in particular for ORCHIDEE and LPJmL (Fig. 4c).

Correlation coefficients indicating the similarity of simulated and observed patterns of AGB range from 0.25 - 0.53 (corrected) across all models (Tab. S3). The highest similarity of pattern (i.e. best correlation values r_{corr}) is found for ORCHIDEE, lowest similarity of pattern for LPJmL. Across the three models INLAND, JULES and LPJmL, generally, higher similarity of pattern

15   is found for the allometric equations that include regional height models and mean or species specific wood density (K$_{DH\rho}$, K$_{DH}$; Fig. 5).

3.2 Comparison of woody productivity (WP)

Mean $\bar{x}$ and variability at pixel-scale $\sigma_{x,corr}$ of observed WP are 2.57 and 0.38 Mg C ha^{-1} yr^{-1}, respectively. There seems to be a weak spatial pattern in the plot estimates at pixel level (Fig. 6a), which is not reflected by the models (Fig. 6c-f). The DGVMs

20   display a distinct pattern of WP across the region that strongly differs among the four models (Fig. 6c-f).

Mean WP simulated by the DGVMs ($\bar{y}$) is between 4-5 Mg C ha^{-1} yr^{-1} for LPJmL and JULES, and 8-9 Mg C ha^{-1} yr^{-1} for INLAND and ORCHIDEE, respectively (Tab. S4). All DGVMs strongly overestimate mean WP (Tab. 1A). In addition, most models overestimate the pattern amplitude, and the simulated variability ranges between 0.72 and 1.6 Mg C ha^{-1} yr^{-1} (Tab. S4). Pattern similarity of observed and simulated data is low ranging from 0.03 to 0.50 (Tab.1A), even with a relatively low

25   maximum achievable correlation of 0.65 (Tab. 1A).

3.3 Comparison of residence time of woody biomass (τ)

Mean $\bar{x}$ and variability at pixel-scale $\sigma_{x,corr}$ of observed τ are 74 and 28 years, respectively. Again the visual comparison shows that the simulations do not match the observations (Fig. 7a vs. Fig. 7c-f). The simulated mean $\bar{y}$ of τ ranges between 15 (INLAND) to 35 (LPJmL) years with a variability of 3 (INLAND) to 8 (LPJmL) years. This is displayed in our comparison

30   metrics: Mean bias results in very low values (i.e. strong underestimation of 53 to 80%; Tab. 1B) and pattern amplitude is strongly underestimated by 65 to 87% (Tab. 1B). The similarity of pattern is very low for all models (Tab. 1B).

4 Discussion

We here present a novel approach for a pixel-to-point comparison. We account for the reduced observed variability when going from point to pixel scale by evaluating three indicators, i.e. the mean bias, the pattern amplitude and the similarity of spatial pattern (section 1.3). We use an example from the Amazon region by comparing model output from four DGVMs and
5   forest inventory data. In the following, we discuss our findings of substantial discrepancies between simulated and observed patterns of AGB, WP and τ across the Amazon region.

4.1 How well do model simulations represent observed biomass patterns across the Amazon?

Interpolated biomass maps from plot observations (e.g. Johnson et al., 2016;Malhi et al., 2006) should be treated with caution since plot observations may not be representative at the landscape-scale (Chave et al., 2004). As a result, a direct and
10   meaningful comparison of observed and simulated maps is currently not feasible but reliable biomass estimates are necessary for implementation of protection incentives and future projections of vegetation biomass. Our results demonstrate that most models are in good agreement and deviate from mean observational biomass by less than 20% (i.e. low mean bias c.f. Fig. 3) and their variability at landscape-scale deviates by about 40% (i.e. pattern amplitude, c.f. Fig. 4). Such relatively good agreement was also found in simulation runs from Delbart et al. (2010) and Johnson et al. (2016). Our results even yield
15   relatively high similarity in observed and simulated spatial patterns of AGB at pixel scale (except LPJmL; c.f. Fig. 5), given the fact that the maximum achievable correlation in the data itself is only 0.6-0.8 (Fig. 5). This indicates that there is considerable uncertainty in the data, which needs to be considered in point-to-pixel comparisons and which we elaborate in the following paragraph.

4.2 How to evaluate differences between observed and simulated patterns of biomass (based on the presented metrics),
20   in particular when considering different allometric equations?

As discussed by several authors (e.g. Baker et al., 2004;Chave et al., 2006;Chave et al., 2014;Réjou-Méchain et al., 2017), the methodology used to convert plot measurements to actual biomass may lead to differential biomass estimates depending on the respective assumptions of the allometric equations employed (i.e. using species-level or community mean wood density, and region-specific or basin-wide height models, see SI Tab. S1). As a result, we here find a more or less pronounced pattern
25   of biomass variability across the Amazon region based on respective assumption used (c.f. Fig. 2, Fig. S1). While mean global variability of biomass is highest for the allometric equations including species-level wood density (c.f. Figure S1, Table S2), highest within-pixel variability is found for biomass values estimated from two-parameter allometric equations (Tab. S2) excluding tree height (c.f. Table S1). This result is also reflected by the lower maximum achievable correlation coefficient (r_{max}), describing how observational data at point scale correlates with observational data at pixel scale, which is particularly
30   low for the two-parameter allometric equations (Fig. 5). Albeit the fact that three out of four DGVMs achieve a relatively good agreement between simulated and observed patterns at the pixel scale we find substantial uncertainty in the observational data due to spatial heterogeneity of local vegetation characteristics such as the structural and functional tree species composition

affecting biomass estimates across the Amazon (see also Rödig et al. 2017). The uncertainty resulting from conversion of raw inventory measurements into biomass from different allometric equations is generally neglected in model-data comparisons. However, it strongly affects our pixel-to-point comparison metrics, thereby remaining an important bottleneck for good model-data biomass comparisons. We suggest to include AGB estimates with associated uncertainties, e.g. using Bayesian inference procedures (see Réjou-Méchain et al., 2017) or to directly compare modelled allometries and related parameters in DGVMs with observational data.

4.3 What can we learn from including spatial heterogeneity and underlying drivers of biomass?

While our approach shows that some models could provide robust estimates for standing biomass stocks across the Amazon region (c.f. Fig. 2) it highlights that currently DGVMs do not represent productivity and related turnover correctly (i.e. the relation between productivity and residence time of woody biomass). As a result, the models might simulate the correct patterns for the wrong reasons as far as it can be derived from observational data. The four DGVMs applied in this study generally capture the observed pattern of AGB but strongly overestimate observed WP and underestimate τ, and, from a pixel perspective, do not show strong variability across the Amazon region, thereby not capturing observed gradients (c.f. Fig. 6 and Tab. 1). WP and τ are driving AGB and are calculated by different schemes in the four DGVMs, e.g., regarding carbon allocation and drivers of mortality. Ground observations suggest that forest structure, forest dynamics and species composition vary across the Amazon region, such that variations in geology and soil fertility/mechanical properties coincide with region-wide variations in aboveground biomass, growth and stem mortality rates (Johnson et al., 2016;Quesada et al., 2012). In accordance, recent studies highlight that variation in stem mortality rates determines spatial variation in AGB and conclude that mortality should be modeled on the basis of individual stems, since stem-size distributions and stand density are important for predicting variation in aboveground biomass (Johnson et al., 2016;Rödig et al., 2017;Pillet et al., 2018). Projected increasing disturbances with different sizes and frequency may be an important additional driver for further variations under future scenarios (Espírito-Santo et al., 2014;Rödig et al., 2017). Nonetheless, the mechanisms leading to stem mortality need to be implemented in models based on experimental data that are only recently becoming available (Meir et al., 2015;Rowland et al., 2015). Overall, the DGVMs are able to reproduce the observed spatial pattern of AGB across the Amazon region, whereas for WP the model performance is less good and reproduction of the spatial pattern in mortality is generally very poor (Fig. 2, 6, 7). This suggests that models need to account for processes such as WP and mortality more mechanistically by including factors associated with resource limitation and disturbances regimes (see also Johnson et al. 2016). Recent efforts aiming at improving simulated Amazon forest biomass and productivity by including spatial variation in biophysical parameters (such as τ and Vc_{max}) have found that using single values for key parameters limits simulation accuracy (Castanho et al., 2013). Thus, we conclude that a more mechanistic representation of the processes driving the spatial variability of carbon stocks and fluxes, forest structure and tree demographic dynamics is necessary to improve simulation accuracy (Rödig et al., 2018).

5 Future applications of the methodological approach and outlook

In general, we assume that the basic concept of our method is applicable to any comparison between two datasets that are characterized by differences in spatial scale. If the process that causes small-scale variability can be approximated as white noise, corrected statistics can be computed. Notwithstanding future developments of next generation DGVMs, the most relevant step of the presented approach is to account for the within-pixel variability σ_ε from the point data to allow for a comparison of observational and simulated data. Due to relatively sparse plot data availability, we assume here that σ_ε is stationary across the Amazon region. To evaluate this assumption further, we have calculated a regional within-pixel variability σ_ε (Fig. S2) and find that it is in the range of the Amazon-wide σ_ε of 28 to 36 Mg C ha^{-1} (depending on the allometric equation used, see Tab. S2). Field studies show that forest dynamics vary locally, mostly due to variations in natural disturbance regimes, mortality and edaphic properties (e.g. Baker et al., 2004;Chambers et al., 2013;Chave et al., 2006;Malhi et al., 2006;John et al., 2007), which in turn strongly influences our calculated within-pixel variability and thus, the metrics of the pixel-to-point comparison. Recent regional studies, that combine observational plot data and remote-sensing products from applications such as LIDAR (regions of Peru: Marvin et al. (2014), French Guyana: Fayad et al. (2016), Congo: Xu et al. (2017)), have already proven to detect spatial variability at high spatial resolution, which could be used to calculate a pixel-wise within-pixel variability based on our approach. Upcoming remote sensing missions as the Global Ecosystem Dynamics Investigation Lidar (GEDI), the ESA BIOMASS mission, the NASA-ISRO Synthetic Aperture Radar (NISAR) mission, or the proposed Tandem-L mission (Moreira et al., 2015) will have the potential to provide non-stationary values of within-pixel variability for all regions of the Amazon. Thus, it is desirable to include regionally or locally specific estimates of σ_ε^2 in our analyses, which could be derived e.g. from above mentioned remote sensing data or from individual tree-based high-resolution simulations (e.g. Rödig et al., 2017). In any case, we conclude that upcoming model-data comparison studies should at least account for stationary within-pixel variability when comparing simulated spatial data to data from discrete observational networks.

6 Code and data availability

All models are described in more detail in the supplementary material. The model code for LPJmL is available through PIK's gitlab server at https://gitlab.pik-potsdam.de/lpjml/LPJmL and archived under https://doi.org/10.5880/pik.2018.002. The model code for ORCHIDEE is available under http://dx.doi.org/10.14768/06337394-73A9-407C-9997-0E380DAC5597. The model code for JULES is available from the JULES FCM repository: https://code.metoffice.gov.uk/trac/jules (registration required). The model code for INLAND is available from http://www.ccst.inpe.br/wp-content/uploads/inland/inland2.0.tar.gz. The permanent archive of the observational data from Mitchard et al. (2014) can be accessed at http://dx.doi.org/10.5521/FORESTPLOTS.NET/2014_1, see also Lopez-Gonzales et al. (2014). The inventory data from Brienen et al. (2015) are available at http://dx.doi.org/10.5521/ForestPlots.net/2014_4, see also Brienen et al. (2014).

7 Author contribution

AR and JH conceived the ideas and designed methodology, and contributed equally to the paper; AR, JH, ER, PP and CZ analysed the data; FL, KT, MG, CR, BC, GS performed simulation runs; all authors contributed to writing the paper.

8 Acknowledgements

9 References

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N.,
10 Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., and Willcock, S.: An integrated pan-tropical biomass map using multiple reference datasets, Global Change Biology, 22, 1406-1420, doi:10.1111/gcb.13139, 2016.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A.,
15 Samanta, S., and Houghton, R. A.: Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps, Nature Climate Change, 2, 182-185, 10.1038/nclimate1354, 2012.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., and Houghton, R. A.: Tropical forests are a net carbon source based on aboveground measurements of gain and loss, Science, 358, 230-234, 10.1126/science.aam5962, 2017.

Baker, T. R., Phillips, O. L., Malhi, Y., Almeida, S., Arroyo, L., Di Fiore, A., Erwin, T., Higuchi, N., Killeen, T. J., Laurance, S. G., Laurance,
20 W. F., Lewis, S. L., Monteagudo, A., Neill, D. A., Vargas, P. N., Pitman, N. C. A., Silva, J. N. M., and Martinez, R. V.: Increasing biomass in Amazonian forest plots, Phil. Trans. R. Soc. B, 359, 353-365, 2004.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677-699, 10.5194/gmd-4-677-2011, 2011.

25 Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Global Change Biology, 13, 679-706, 2007.

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S. L., Vasquez Martinez, R., Alexiades, M., Alvarez Davila, E., Alvarez-Loayza, P., Andrade, A., Aragao, L. E. O. C.,
30 Araujo-Murakami, A., Arets, E. J. M. M., Arroyo, L., Aymard C, G. A., Banki, O. S., Baraloto, C., Barroso, J., Bonal, D., Boot, R. G. A., Camargo, J. L. C., Castilho, C. V., Chama, V., Chao, K. J., Chave, J., Comiskey, J. A., Cornejo Valverde, F., da Costa, L., de Oliveira, E. A., Di Fiore, A., Erwin, T. L., Fauset, S., Forsthofer, M., Galbraith, D. R., Grahame, E. S., Groot, N., Herault, B., Higuchi, N., Honorio Coronado, E. N., Keeling, H., Killeen, T. J., Laurance, W. F., Laurance, S., Licona, J., Magnussen, W. E., Marimon, B. S., Marimon-Junior, B. H., Mendoza, C., Neill, D. A., Nogueira, E. M., Nunez, P., Pallqui Camacho, N. C., Parada, A., Pardo-Molina, G., Peacock, J., Pena-
35 Claros, M., Pickavance, G. C., Pitman, N. C. A., Poorter, L., Prieto, A., Quesada, C. A., Ramirez, F., Ramirez-Angulo, H., Restrepo, Z., Roopsind, A., Rudas, A., Salomao, R. P., Schwarz, M., Silva, N., Silva-Espejo, J. E., Silveira, M., Stropp, J., Talbot, J., ter Steege, H., Teran-Aguilar, J., Terborgh, J., Thomas-Caesar, R., Toledo, M., Torello-Raventos, M., Umetsu, R. K., van der Heijden, G. M. F., van der Hout, P., Guimaraes Vieira, I. C., Vieira, S. A., Vilanova, E., Vos, V. A., and Zagt, R. J.: Plot Data from: "Long-term decline of the Amazon carbon sink", ForestPlots.NET, DOI: 10.5521/ForestPlots.net/2014_4, 2014.

40 Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S. L., Vasquez Martinez, R., Alexiades, M., Alvarez Davila, E., Alvarez-Loayza, P., Andrade, A., Aragao, L. E. O. C., Araujo-Murakami, A., Arets, E. J. M. M., Arroyo, L., Aymard C, G. A., Banki, O. S., Baraloto, C., Barroso, J., Bonal, D., Boot, R. G. A., Camargo, J. L. C., Castilho, C. V., Chama, V., Chao, K. J., Chave, J., Comiskey, J. A., Cornejo Valverde, F., da Costa, L., de Oliveira, E. A., Di Fiore, A., Erwin, T. L., Fauset, S., Forsthofer, M., Galbraith, D. R., Grahame, E. S., Groot, N., Herault, B., Higuchi, N., Honorio

Coronado, E. N., Keeling, H., Killeen, T. J., Laurance, W. F., Laurance, S., Licona, J., Magnussen, W. E., Marimon, B. S., Marimon-Junior, B. H., Mendoza, C., Neill, D. A., Nogueira, E. M., Nunez, P., Pallqui Camacho, N. C., Parada, A., Pardo-Molina, G., Peacock, J., Pena-Claros, M., Pickavance, G. C., Pitman, N. C. A., Poorter, L., Prieto, A., Quesada, C. A., Ramirez, F., Ramirez-Angulo, H., Restrepo, Z., Roopsind, A., Rudas, A., Salomao, R. P., Schwarz, M., Silva, N., Silva-Espejo, J. E., Silveira, M., Stropp, J., Talbot, J., ter Steege, H., Teran-

5   Aguilar, J., Terborgh, J., Thomas-Caesar, R., Toledo, M., Torello-Raventos, M., Umetsu, R. K., van der Heijden, G. M. F., van der Hout, P., Guimaraes Vieira, I. C., Vieira, S. A., Vilanova, E., Vos, V. A., and Zagt, R. J.: Long-term decline of the Amazon carbon sink, Nature, 519, 344-348, 10.1038/nature14283
http://www.nature.com/nature/journal/v519/n7543/abs/nature14283.html#supplementary-information, 2015.
Castanho, A. D. A., Coe, M. T., Costa, M. H., Malhi, Y., Galbraith, D., and Quesada, C. A.: Improving simulated Amazon forest biomass

10 and productivity by including spatial variation in biophysical parameters, Biogeosciences, 10, 2255-2272, 10.5194/bg-10-2255-2013, 2013.
Chambers, J. Q., Negron-Juarez, R. I., Magnabosco Marrac, D. M., Di Vittorioa, A., Tewse, J., Roberts, D., Ribeiro, G. H. P. M., Trumbore, S. E., and Higuchi, N.: The steady-state mosaic of disturbance and succession across an old-growth Central Amazon forest landscape, Proceedings of the National Academy of Sciences, 110, 3949–3954, 2013.
Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., and Perez, R.: Error propagation and scaling for tropical forest biomass estimates,

15 Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359, 409-420, 10.1098/rstb.2003.1425, 2004.
Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J. P., Nelson, B. W., Ogawa, H., Puig, H., Riéra, B., and Yamakura, T.: Tree allometry and improved estimation of carbon stocks and balance in tropical forests, Oecologia, 145, 87–99, 2005.
Chave, J., Muller-Landau, H. C., Baker, T. R., Easdale, T. A., Steege, H. t., and Webb, C. O.: REGIONAL AND PHYLOGENETIC

20 VARIATION OF WOOD DENSITY ACROSS 2456 NEOTROPICAL TREE SPECIES, Ecological Applications, 16, 2356-2367, 10.1890/1051-0761(2006)016[2356:RAPVOW]2.0.CO;2, 2006.
Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., and Zanne, A. E.: Towards a worldwide wood economics spectrum, Ecology Letters, 12, 351-366, 2009.
Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman,

25 R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., Saldarriaga, J. G., and Vieilledent, G.: Improved allometric models to estimate the aboveground biomass of tropical trees, Global Change Biology, 20, 3177-3190, 10.1111/gcb.12629, 2014.
Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon

30 fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701-722, 10.5194/gmd-4-701-2011, 2011.
Delbart, N., Ciais, P., Chave, J., Viovy, N., Malhi, Y., and Le Toan, T.: Mortality as a key driver of the spatial distribution of aboveground biomass in Amazonian forest: results from a dynamics vegetation model. , Biogeosciences, 7, 3017-3039, 2010.
Espírito-Santo, F. D. B., Gloor, M., Keller, M., Malhi, Y., Saatchi, S., Nelson, B., Junior, R. C. O., Pereira, C., Lloyd, J., Frolking, S., Palace, M., Shimabukuro, Y. E., Duarte, V., Mendoza, A. M., López-González, G., Baker, T. R., Feldpausch, T. R., Brienen, R. J. W., Asner, G. P.,

35 Boyd, D. S., and Phillips, O. L.: Size and frequency of natural forest disturbances and the Amazon forest carbon balance, Nature Communications, 5, 3434, 10.1038/ncomms4434
https://www.nature.com/articles/ncomms4434#supplementary-information, 2014.
Fayad, I., Baghdadi, N., Guitet, S., Bailly, J.-S., Hérault, B., Gond, V., El Hajj, M., and Tong Minh, D. H.: Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data, International Journal of Applied Earth

40 Observation and Geoinformation, 52, 502-514, https://doi.org/10.1016/j.jag.2016.07.015, 2016.
Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., Monteagudo Mendoza, A., Lopez-Gonzalez, G., Banin, L., Abu Salim, K., Affum-Baffoe, K., Alexiades, M., Almeida, S., Amaral, I., Andrade, A., Aragão, L. E. O. C., Araujo Murakami, A., Arets, E. J. M. M., Arroyo, L., Aymard C, G. A., Baker, T. R., Bánki, O. S., Berry, N. J., Cardozo, N., Chave, J., Comiskey, J. A., Alvarez, E., de Oliveira, A., Di Fiore, A., Djagbletey, G., Domingues, T. F., Erwin, T. L., Fearnside, P. M., França, M. B., Freitas, M. A., Higuchi, N., C,

45 E. H., Iida, Y., Jiménez, E., Kassim, A. R., Killeen, T. J., Laurance, W. F., Lovett, J. C., Malhi, Y., Marimon, B. S., Marimon-Junior, B. H., Lenza, E., Marshall, A. R., Mendoza, C., Metcalfe, D. J., Mitchard, E. T. A., Neill, D. A., Nelson, B. W., Nilus, R., Nogueira, E. M., Parada, A., Peh, K. S. H., Pena Cruz, A., Peñuela, M. C., Pitman, N. C. A., Prieto, A., Quesada, C. A., Ramírez, F., Ramírez-Angulo, H., Reitsma, J. M., Rudas, A., Saiz, G., Salomão, R. P., Schwarz, M., Silva, N., Silva-Espejo, J. E., Silveira, M., Sonké, B., Stropp, J., Taedoumg, H. E., Tan, S., ter Steege, H., Terborgh, J., Torello-Raventos, M., van der Heijden, G. M. F., Vásquez, R., Vilanova, E., Vos, V. A., White, L.,

50 Willcock, S., Woell, H., and Phillips, O. L.: Tree height integrated into pantropical forest biomass estimates, Biogeosciences, 9, 3381-3403, 10.5194/bg-9-3381-2012, 2012.
Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., and Sitch, S.: Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model, Journal of Hydrology, 286, 249-270, 2004.
Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., Hansen, M. C., Potapov, P. V., and Lotsch, A.: Baseline Map

55 of Carbon Emissions from Deforestation in Tropical Regions, Science, 336, 1573-1576, 10.1126/science.1217962, 2012.

Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., De-Fries, R. S., Hansen, M. C., Le Quéré, C., and Ramankutty, N.: Carbon emissions from land use and land-cover change, Biogeosciences, 9, 5125–5142, 2012.

John, R., Dalling, J. W., Harms, K. E., Yavitt, J. B., Stallard, R. F., Mirabello, M., Hubbell, S. P., Valencia, R., Navarrete, H., Vallejo, M., and Foster, R. B.: Soil nutrients influence spatial distributions of tropical tree species., Proceedings of the National Academy of Sciences, 104, 864–869, 2007.

Johnson, M. O., Galbraith, D., Gloor, M., De Deurwaerder, H., Guimberteau, M., Rammig, A., Thonicke, K., Verbeeck, H., Von Randow, C., and al., e.: Variation in stem mortality rates determines patterns of above-ground biomass in Amazonian forests: implications for dynamic global vegetation models, Global Change Biology, 22, 3996-4013, doi: 10.1111/gcb.13315, 2016.

Krinner, G., Viovy, N., de Noblet-Ducoudre, N., Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere biosphere system, Global Biogeochemical Cycles, 19, GB1015, 2005.

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., and Ramankutty, N.: Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure, Global Biogeochemical Cycles, 14, 795-825, 2000.

Langner, A., Achard, F., and Grassi, G.: Can recent pan-tropical biomass maps be used to derive alternative Tier 1 values for reporting REDD+ activities under UNFCCC?, Environmental Research Letters, 9, 124008, 2014.

Lopez-Gonzalez, G., Lewis, S., Burkitt, M., and Phillips, O.: ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data, Journal of Vegetation Science, 22, 610-613, 2011.

Lopez-Gonzalez, G., Mitchard, E., Feldpausch, T., Brienen, R., Monteagudo, A., Baker, T., Lewis, S., Lloyd, J., Quesada, C., Gloor, E., ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragao, L., Arroyo, L., Aymard, G., Banki, O., Bonal, D., Brown, S., Brown, F., Ceron, C., Chama Moscoso, V., Chave, J., Comiskey, J., Cornejo, F., Corrales Medina, M., Da Costa, L., Costa, F., Di Fiore, A., Domingues, T., Erwin, T., Fredericksen, T., Higuchi, N., Honorio Coronado, E., Killeen, T., Laurance, W., Levis, C., Magnusson, W., Marimon, B., Marimon-Junior, B., Mendoza Polo, I., Mishra, P., Nascimento, M., Neill, D., Nunez Vargas, M., Palacios, W., Parada-Gutierrez, A., Pardo Molina, G., Pena-Claros, M., Pitman, N., Peres, C., Poorter, L., Prieto, A., Ramirez-Angulo, H., Restrepo Correa, Z., Roopsind, A., Roucoux, K., Rudas, A., Salomao, R., Schietti, J., Silveira, M., De Souza, P., Steiniger, M., Stropp, J., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., Van Andel, T., van der Heijden, G., Vieira, I., Vieira, S., Vilanova-Torre, E., Vos, V., Wang, O., Zartman, C., de Oliveira, E., Morandi, P., Malhi, Y., and Phillips, O.: Amazon forest biomass measured in inventory plots. Plot Data from "Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites." ForestPlots.net. , doi: 10.5521/FORESTPLOTS.NET/2014_1, 2014.

Malhi, Y., Wood, D., Baker, T. R., Wright, J., Phillips, O. L., Cochrane, T., Meir, P., Chave, J., Almeida, S., Arroyo, L., Higuchi, N., Killeen, T. J., Laurance, S. G., Laurance, W. F., Lewis, S. L., Monteagudo, A., Neill, D. A., Nunez Vargas, P., Pitman, N. C. A., Quesada, C. A., Salamao, R., Silva, J. N. M., Torres-Lezama, A., Terborgh, J., Vasquez Martinez, R., and Vinceti, B.: The regional variation of aboveground live biomass in old-growth Amazonian forests, Global Change Biology, 12, 1107-1138, 2006.

Marvin, D. C., Asner, G. P., Knapp, D. E., Anderson, C. B., Martin, R. E., Sinca, F., and Tupayachi, R.: Amazonian landscapes and the bias in field studies of forest structure and biomass, Proceedings of the National Academy of Sciences, 111, E5224-E5232, 10.1073/pnas.1412999111, 2014.

Meir, P., Mencuccini, M., and Dewar, R. C.: Drought-related tree mortality: addressing the gaps in understanding and prediction, New Phytologist, 207, 28-33, doi:10.1111/nph.13382, 2015.

Mitchard, E. T. A., Feldpausch, T. R., Brienen, R. J. W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., Lewis, S. L., Lloyd, J., Quesada, C. A., Gloor, M., ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragao, L. E. O. C., Arroyo, L., Aymard, G., Banki, O., Bonal, D., Brown, S., Brown, F. I., Ceron, C. E., Chama Moscoso, V., Chave, J., Comiskey, J. A., Cornejo, F., Corrales Medina, M., Da Costa, L., Costa, F. R. C., Di Fiore, A., Domingues, T. F., Erwin, T. L., Frederickson, T., Higuchi, N., Honorio Coronado, E. N., Killeen, T. J., Laurance, W. F., Levis, C., Magnusson, W. E., Marimon, B. S., Marimon Junior, B. H., Mendoza Polo, I., Mishra, P., Nascimento, M. T., Neill, D., Nunez Vargas, M. P., Palacios, W. A., Parada, A., Pardo Molina, G., Pena-Claros, M., Pitman, N., Peres, C. A., Poorter, L., Prieto, A., Ramirez-Angulo, H., Restrepo Correa, Z., Roopsind, A., Roucoux, K. H., Rudas, A., Salomao, R. P., Schietti, J., Silveira, M., de Souza, P. F., Steiniger, M. K., Stropp, J., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., van Andel, T. R., van der Heijden, G. M. F., Vieira, I. C. G., Vieira, S., Vilanova-Torre, E., Vos, V. A., Wang, O., Zartman, C. E., Malhi, Y., and Phillips, O. L.: Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites, Glob. Ecol. Biogeogr., 23, 935-946, 10.1111/geb.12168, 2014.

Moreira, A., Krieger, G., Hajnsek, I., Papathanassiou, K., Younis, M., Lopez-Dekker, P., Huber, S., Villano, M., Pardini, M., Eineder, M., De Zan, F., and Parizzi, A.: Tandem-L: A Highly Innovative Bistatic SAR Mission for Global Observation of Dynamic Processes on the Earth's Surface, IEEE Geoscience and Remote Sensing Magazine, 3, 8-23, 10.1109/MGRS.2015.2437353, 2015.

Pillet, M., Joetzjer, E., Belmin, C., Chave, J., Ciais, P., Dourdain, A., Evans, M., Hérault, B., Luyssaert, S., Poulter, B., and Zhou, S.: Disentangling competitive vs. climatic drivers of tropical forest mortality, Journal of Ecology, 106, 1165-1179, doi:10.1111/1365-2745.12876, 2018.

Politis, D. N., and Romano, J. P.: The stationary bootstrap, J. Am. Stat. Assoc., 89, 1303–1313, 1994.

Quesada, C. A., Phillips, O. L., Schwarz, M., Czimczik, C. I., Baker, T. R., Patino, S., Fyllas, N. M., Hodnett, M. G., Herrera, R., Almeida, S., Alvarez Davila, E., Arneth, A., Arroyo, L., Chao, K. J., Dezzeo, N., Erwin, T., di Fiore, A., Higuchi, N., Honorio Coronado, E., Jimenez, E. M., Killeen, T., Lezama, A. T., Lloyd, G., Lopez-Gonzalez, G., Luizao, F. J., Malhi, Y., Monteagudo, A., Neill, D. A., Nunez Vargas, P.,

Paiva, R., Peacock, J., Penuela, M. C., Pena Cruz, A., Pitman, N., Priante Filho, N., Prieto, A., Ramirez, H., Rudas, A., Salomao, R., Santos, A. J. B., Schmerler, J., Silva, N., Silveira, M., Vasquez, R., Vieira, I., Terborgh, J., and Lloyd, J.: Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate, Biogeosciences, 9, 2203-2246, 10.5194/bg-9-2203-2012, 2012.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., and Hérault, B.: biomass: an r package for estimating above-ground biomass and
5   its uncertainty in tropical forests, Methods in Ecology and Evolution, 8, 1163-1167, 10.1111/2041-210X.12753, 2017.

Rödig, E., Cuntz, M., Heinke, J., Rammig, A., and Huth, A.: Spatial heterogeneity of biomass and forest structure of the Amazon rain forest: Linking remote sensing, forest modelling and field inventory,, Glob. Ecol. Biogeogr., 26, 1292–1302, 2017.

Rödig, E., Cuntz, M., Rammig, A., Fischer, R., Taubert, F., and Huth, A.: The importance of forest structure for carbon fluxes of the Amazon rainforest, Environmental Research Letters, 13, 054013, 2018.

10 Rowland, L., da Costa, A. C. L., Galbraith, D. R., Oliveira, R. S., Binks, O. J., Oliveira, A. A. R., Pullen, A. M., Doughty, C. E., Metcalfe, D. B., Vasconcelos, S. S., Ferreira, L. V., Malhi, Y., Grace, J., Mencuccini, M., and Meir, P.: Death from drought in tropical forests is triggered by hydraulics not carbon starvation, Nature, 528, 119-+, 10.1038/nature15539, 2015.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A.: Benchmark map of forest carbon stocks in tropical regions across three continents,
15 Proceedings of the National Academy of Sciences of the United States of America, 108, 9899-9904, 10.1073/pnas.1019576108, 2011.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088-3111, 10.1175/jcli3790.1, 2006.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplans, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation
20 model, Global Change Biology, 9, 161-185, 2003.

Song, X.-P., Huang, C., Saatchi, S., Hansen, M., and Townshend, J.: Annual Carbon Emissions from Deforestation in the Amazon Basin between 2000 and 2010, PLoS ONE, 10, e0126754, 2015.

Turner, M. G., Gardner, R. H., and O'Neill, R. V.: Landscape Ecology in theory and praxis, edited by: Turner, M. G., Gardner, R. H., and O'Neill, R. V., Springer-Verlag, New York, 401 pp., 2001.

25 Xu, L., Saatchi, S. S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., Bastin, J.-F., Banks, N., Boeckx, P., Verbeeck, H., Lewis, S. L., Muanza, E. T., Bongwele, E., Kayembe, F., Mbenza, D., Kalau, L., Mukendi, F., Ilunga, F., and Ebuta, D.: Spatial Distribution of Carbon Stored in Forests of the Democratic Republic of Congo, Scientific Reports, 7, 15030, 10.1038/s41598-017-15050-z, 2017.

Zanne, A. E., Zanne, A. E., Lopez-Gonzalez, G., Coomes, D. A., Ilic, J., Jansen, S., Lewis, S. L., Miller, R. B., Swenson, N. G., Wiemann, M. C., and Chave, J.: Global wood density database, http://dx.doi.org/10.5061/dryad.234, 2009.

30 Zhang, K., de Almeida Castanho, A. D., Galbraith, D. R., Moghim, S., Levine, N. M., Bras, R. L., Coe, M. T., Costa, M. H., Malhi, Y., Longo, M., Knox, R. G., McKnight, S., Wang, J., and Moorcroft, P. R.: The fate of Amazonian ecosystems over the coming century arising from changes in climate, atmospheric CO2, and land use, Global Change Biology, 21, 2569-2587, 10.1111/gcb.12903, 2015.
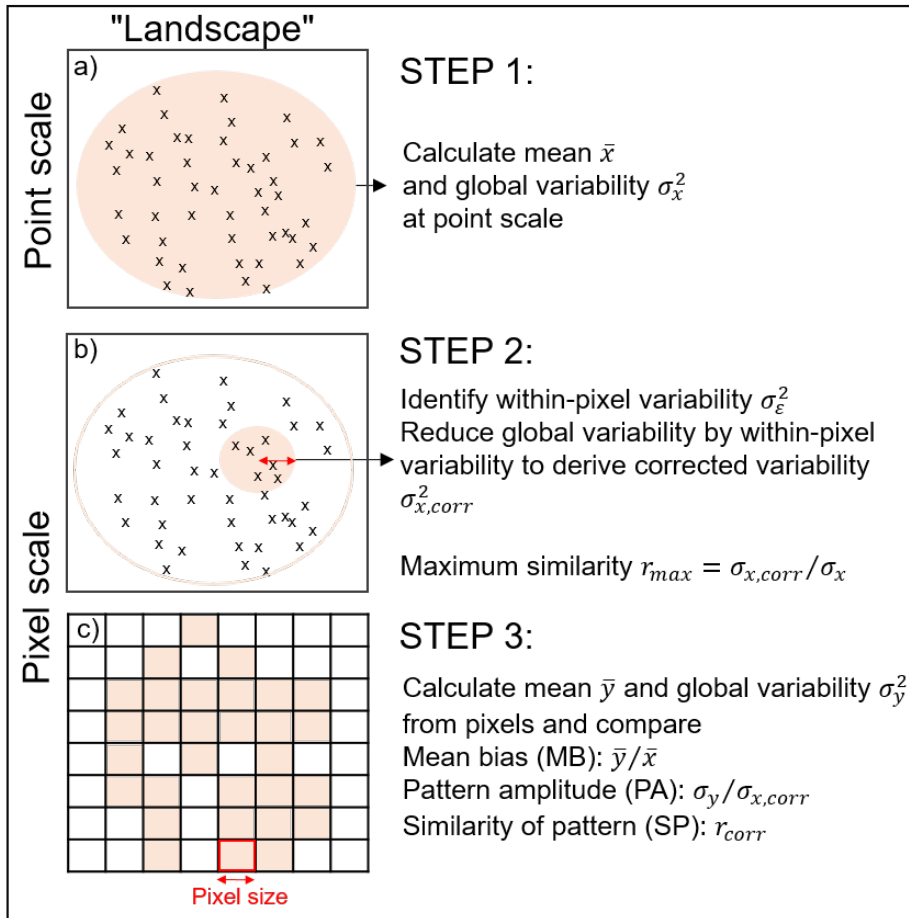
14

Figures

Figure 1: Schematic overview of the three steps in the pixel-to-point comparison. Note that we refer to "landscape" as the region of interest for which point and pixel data are available. (a) the mean and global variability at point scale are calculated from all plots across the landscape; (b) the within-pixel variability is calculated from all plots within the distance of the pixel-size (i.e. the red arrow corresponds with pixel size); (c) the mean and global variability is calculated from the pixel values and the three comparison metrics are derived. Note: A block bootstrapping with 10000 repetitions is performed to derive confidence intervals of the comparison metrics. The detailed set of equations to calculate maximum similarity, PA and SP can be found in the SI.

15
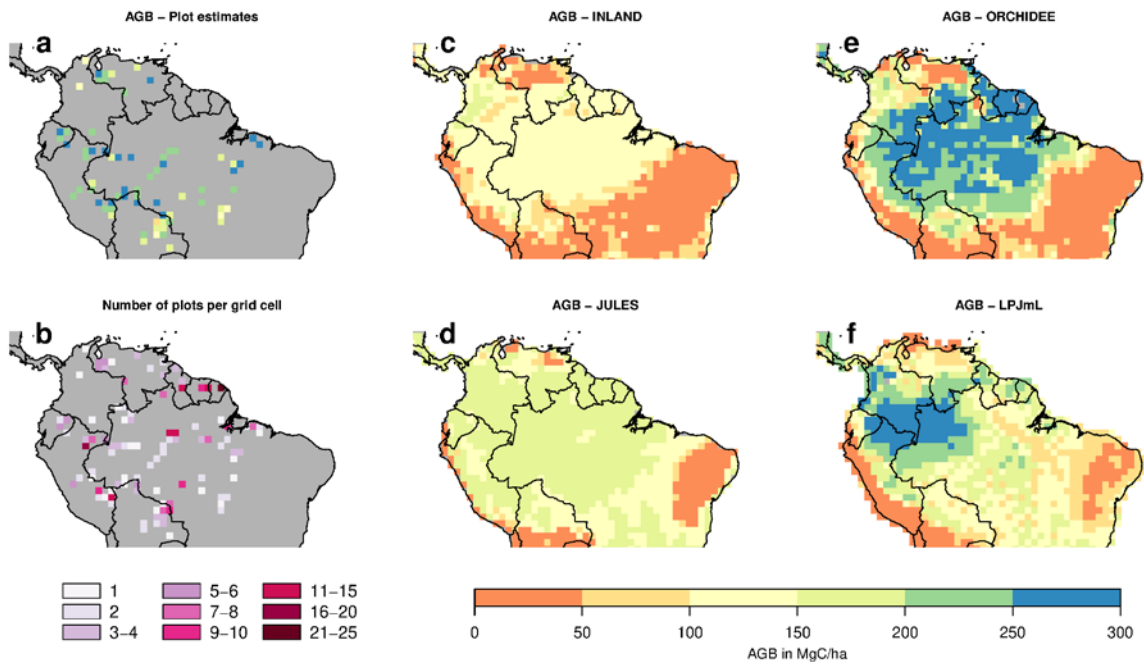
Figure 2. Estimates of aboveground biomass (AGB) from forest plots in 1° x 1° pixels. (a) Mean AGB per pixel derived from inventory data based on one allometric equation ($K_{DH\rho}$, see SI for explanation and other allometric equations). (b) Number of plots per pixel and (c-f) simulated AGB from four DGVMs.
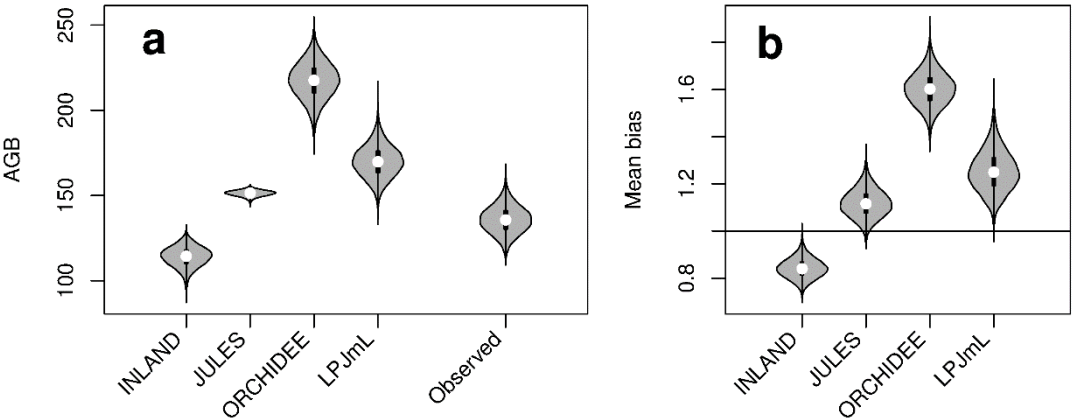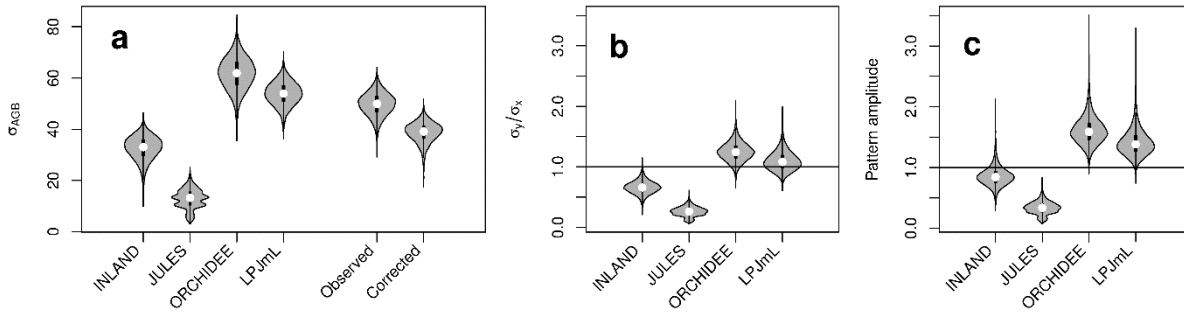
5

Figure 3. Distribution of aboveground biomass (AGB in MgC/ha) from the four DGVMs and from the observational plots (see also Table S2 and S3). The figure shows (a) the mean value (white dot) and distribution from bootstrapping of absolute AGB values from the four simulations and observed (grey violins). b) Mean bias as the ratio of mean simulated and mean observed AGB ($\bar{y}/\bar{x}$).

16

Figure 4: (a) Standard deviations of AGB (in MgC/ha) for the four models and observational data (for the other allometric equations see also Table S2, S3). For the observational data, the global variability at point scale ("observed") and the corrected variability at pixel scale ("corrected") is given; **(b)** ratio of standard deviations without correcting for within-pixel variability, **(c)** corrected metrics of pattern amplitude ($\sigma_y/\sigma_{x,corr}$).
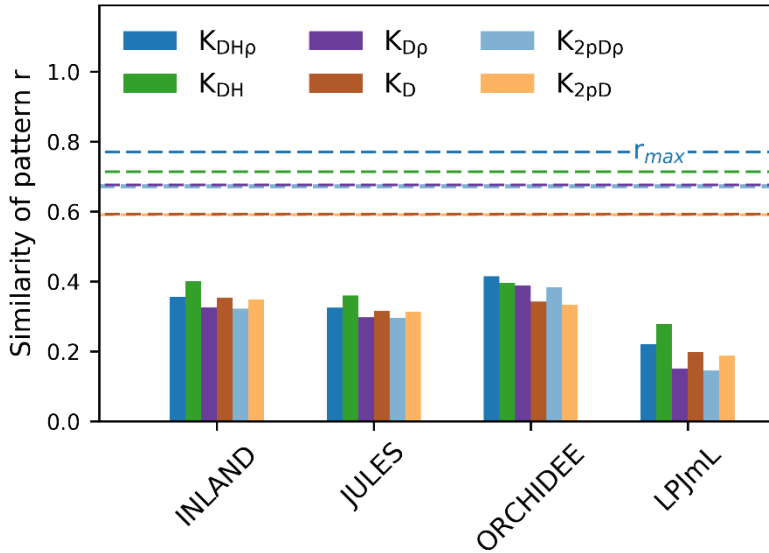
Figure 5: The similarity of the observed vs. simulated spatial pattern of AGB at pixel scale (as indicated by *r* given in bars). The similarity is calculated for different versions of observed AGB derived from six allometric equations (indicated by the colors, see Tab. S1). The dashed line shows the maximum achievable correlation coefficients r_{max} from the observational data and for the different allometric equations.
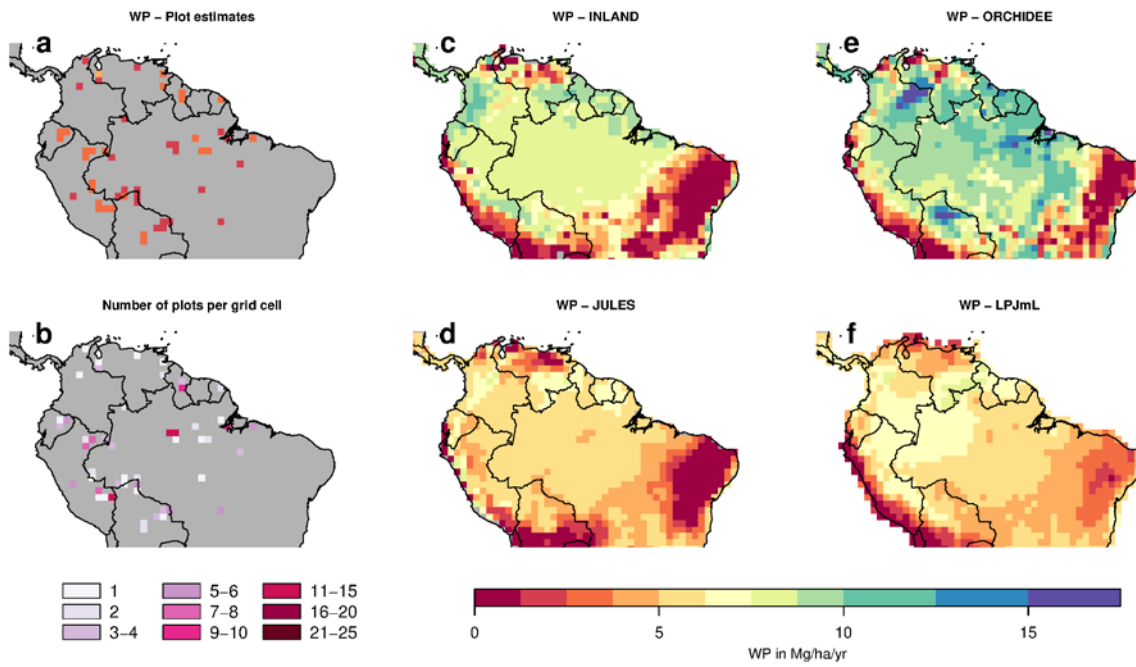
Figure 6. Estimates of aboveground woody productivity (WP) from forest plots in 1° x 1° pixels. (a) Mean WP from inventory plots. (b) Number of plots per pixel and (c-f) simulated WP from four DGVMs.
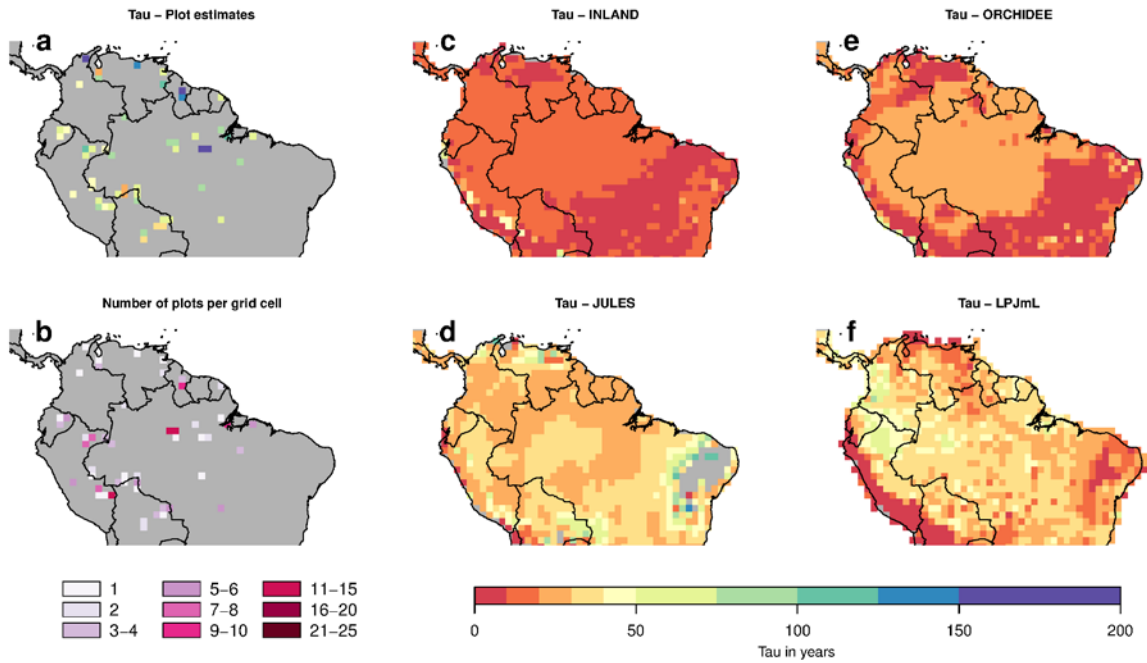
Figure 7. Estimates of woody biomass residence time (τ) from forest plots in 1° x 1° pixels. (b) Mean residence time from inventory plots. (b) Number of plots per pixel and (c-f) simulated τ from four DGVMs.

Tables

Table 1. Results of the point-to-pixel comparison for (A) woody productivity (WP) and (B) residence time of woody biomass (τ). In brackets, the 5% and 95% confidence intervals are given. Blue boxes indicate when models overestimate observed values, red boxes indicate underestimation.

A) Woody productivity (WP)	Mean $\bar{x}$ (Mg/ha/yr)	Corrected global variability $\sigma_{x,corr}$ (Mg/ha/yr)	Max. achievable correlation r_{max}
Observed	2.57	0.38	0.67
	Mean bias ($\bar{y}/\bar{x}$)	Pattern amplitude ($\sigma_y/\sigma_{x,corr}$)	Similarity of pattern (r_{corr})
INLAND	3.11 (2.91 – 3.31)	2.91 (1.75 – 4.83)	0.36 (0.11 – 0.35)
JULES	2.01 (1.88 – 2.14)	1.91 (1.08 – 3.25)	0.38 (0.07 – 0.37)
ORCHIDEE	3.55 (3.21 – 3.96)	4.26 (2.64 – 6.96)	0.03 (-0.25 – 0.01)
LPJmL	1.74 (1.63 – 1.83)	1.99 (1.36 – 3.16)	0.50 (0.27 – 0.50)
B) Residence time (τ)	$\bar{x}$ (years)	$\sigma_{x,corr}$ (years)	r_{max}
Observed	73.84	28.04	0.64
	Mean bias ($\bar{y}/\bar{x}$)	Pattern amplitude ($\sigma_y/\sigma_{x,corr}$)	Similarity of pattern (r_{corr})
INLAND	0.20 (0.17 – 0.24)	0.13 (0.06 – 0.25)	0.01 (-0.30 – 0.02)
JULES	0.42 (0.35 – 0.51)	0.24 (0.07 – 0.46)	-0.23 (-0.68 – -0.22)
ORCHIDEE	0.35 (0.29 – 0.42)	0.24 (0.12 – 0.45)	0.08 (-0.22 – 0.08)
LPJmL	0.47 (0.38 – 0.59)	0.35 (0.19 – 0.61)	-0.18 (-0.46 – -0.18)