[Article title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

[Authors]

Siyuan Wang[1, 2], Hui Yang[1], Sujan Koirala[1], Matthias Forkel[2], Markus Reichstein[1,3], Nuno Carvalhais[1,3,4]

(swang@bgc-jena.mpg.de, huiyang@bgc-jena.mpg.de, skoirala@bgc-jena.mpg.de, matthias.forkel@tu-dresden.de, mreichstein@bgc-jena.mpg.de, ncarvalhais@bgc-jena.mpg.de)

[1] Max-Planck Institute for Biogeochemistry, Jena, Germany

[2] Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing, Dresden, Germany

[3] ELLIS Unit Jena at Michael-Stifel-Center Jena for Data-driven and Simulation Science, Jena, Germany

[4] Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

[Correspondence]

Nuno Carvalhais, Department Biogeochemical Integration, Max Planck Institute for Biogeochemistry; Hans-Knöll-Straße 10, 07745, Jena, Germany.

Email: nuno.carvalhais@bgc-jena.mpg.de

Telephone: +49 3641 57-6225

[Open Research Statement]

Data and code are archived in Zenodo: https://zenodo.org/record/8121119

[Keywords]

Biomass; Carbon Cycle Dynamics; Disturbance Regimes; Forest Mortality; Gap Dynamics.

Abstract

Natural and anthropogenic disturbances are important drivers of tree mortality, shaping the structure, composition, and biomass distribution of forest ecosystems. Differences in disturbance regimes, characterized by the frequency, extent, and intensity of disturbance events, result in structurally different landscapes. Characterizing different disturbance regimes through landscape-scale forest structure provides a unique perspective for diagnosing the impacts and potential carbon-climate feedbacks from terrestrial ecosystems. In this study, we design a model-based experiment to investigate the links between disturbance regimes and spatial biomass patterns. First, the effects of disturbance events on biomass patterns are simulated using a simple dynamic carbon cycle model based on different frequency, extent, and intensity of forest disturbance. We characterize the disturbance regimes via three parameters: μ, α, and β; that represent the probability scale, clustering degree and intensity of different disturbance events, respectively. We generate over 850 thousand biomass patterns, from 2,142 combinations of μ, α, and β under different primary productivity and background mortality scenarios. We characterize the emergent biomass patterns via synthesis statistics, including central tendency statistics; different moments of the distribution; information-based and texture features. We further follow a multi-output regression approach that takes the biomass synthesis statistics and gross primary production (GPP) as independent variables to retrieve the three disturbance regimes parameters. Results show confident inversion of all three “true” disturbance parameters, with Nash-Sutcliffe efficiency of 94.8% for μ, 94.9% for α, and 97.1% for β. And biomass distribution statistics primarily dominate the prediction of μ and β, while texture features have a stronger influence on α. Overall, these results demonstrate the association between biomass patterns and disturbance statistics that emerge from different underlying disturbance regimes. By doing so, it overcomes

the known issue of equifinality between mortality rates and total biomass. Given the increasing availability of Earth observation of biomass, our findings open a new avenue to better understand and parameterize disturbance regimes and their links with vegetation dynamics under climate change. Ultimately, at a large scale, this approach would improve our current understanding of controls and feedback at the biosphere-atmosphere interface in the current Earth system models.

1. Introduction

Mortality is one of the key processes of vegetation dynamics (Franklin et al. 1987; Runkle 2000) that dominates aboveground carbon turnover (Carvalhais et al. 2014; Thurner et al. 2016) and contributes to model uncertainties in the carbon cycle (Friend et al. 2014). The diverse range of natural (e.g. fires, droughts, wind-throws, pathogens and insects outbreaks) and anthropogenic disturbances (e.g. agricultural expansion, urbanization, and clearcutting) act as strong drivers of vegetation mortality, leading to the total or partial loss of biomass (McDowell et al. 2022; Hammond et al. 2022; Grime 1977). A better understanding of mortality and disturbance and their impacts on carbon dynamics is thus crucial for constraining future carbon cycling prognostics (Friend et al. 2014).

The mortality caused by disturbances, as well as primary productivity and allocation, play an important role in controlling the local distribution and the large scale spatial gradients of above-ground biomass (AGB, Delbart et al. 2010; Johnson et al. 2016). But diagnosing disturbances from primary productivity or biomass patterns it is still poorly understood due to equifinality issues (Ryan et al. 2011; Williams et al. 2013) and its highly stochastic nature (Chambers et al. 2004; Allen et al. 2010; Hammond et al. 2022). Characteristics of disturbances at the landscape level, i.e., disturbance regimes, are commonly inferred using metrics like size, frequency, intensity, and aggregation (Turner 2010), which describe the cumulative effects of all disturbance events in a given area and time period (Senf and Seidl 2021b). The disturbance regime ultimately leads to a shifting steady-state mosaic, represented by patches of distinct successional stages or carbon stocks over long time periods (Brokaw and Scheiner 1989).

Most research on quantifying disturbance regime parameters has been carried out either through observation-driven methods or by using model-data-integration. For example, remote sensing based on spectral bands, indices and segment outlines have been used to identify changes in vegetation that are associated with disturbance magnitude, duration, and rate of change (Chambers et al. 2013; Senf et al. 2021b). Two common disturbance parameters, determining the average probability and intensity of biomass loss, were retrieved using satellite biomass observation, though with a high level of equifinality (Williams et al. 2013). Alternatively, using successive biomass maps, some studies have been able to detect differences in patterns of intensity, ranging from deforestation to widespread low-intensity disturbance (Hill et al. 2015). Moreover, the clustering pattern of disturbance events has been recognized as a distinguishing attribute of different disturbance regimes and failure to resolve these patterns can lead to misestimation of average mortality and growth patterns (Fisher et al. 2008). These studies have been motivating the exploration, and highlighting the significant potential, of using biomass observations to understand and quantify disturbance regimes. In parallel, the emergence of up-to-date biomass observations (Saatchi et al. 2011; Santoro et al. 2021) opens novel pathways for comprehending and investigating the clustering patterns of disturbances, alongside other facets of the disturbance regimes, at a more intricate scale. Ultimately, disturbance parameters based on observations could be incorporated in process-based models (e.g., Friend et al., 2013) or individual-based models (Bugmann 2001; Bossel et al. 1994; Yan et al. 2005; Köhler et al. 1998) as stochastic model components that allow the quantification of the impacts of disturbances on the forest carbon cycle from local and global scales.

The goal of our study is to comprehensively characterize disturbance regimes and investigate their connection with resulting biomass patterns through a model-based experiment. Specifically, we focus on the methodology used to simulate the impact of three distinct disturbance regime attributes - extent, frequency, and intensity of disturbance events - on biomass dynamics. The simulations were conducted under varying scenarios of photosynthetic and background mortality rates. The emerging patterns in biomass are then summarized into distribution statistics, information theory and textural features across the simulations in order to retrieve the prescribed disturbance regimes via a multi-output regression approach.

Below, in section 2, we introduce in detail the carbon model, the parameters that control the simulated disturbance regimes, describe the approach to generate disturbance cubes, and the methodology for retrieving disturbance parameters. In Section 3, we present the results of these experiments, namely on the varying patterns of biomass emerging from different disturbance regimes, and the performance of the regression approach to invert the underlying disturbance parameters. This is followed by a discussion on the robustness of the framework and an outlook in Section 4. We present the conclusions of our research in Section 5.

2. Methods

2.1 Dynamic carbon cycle model

We simulate the carbon cycle at the patch level. Each patch is specified as a homogeneous forest stand, representing the smallest computing unit during the simulation (Fisher et al. 2008). The changes in aboveground vegetation carbon (C, in $kgC \cdot m^{-2} \cdot yr^{-1}$) are determined by the difference between aboveground carbon gains (via photosynthesis, NPP_{AGB}) and losses (L, Eq. 1) at the annual scale as,

$$\frac{dAGB}{dt} = NPP_{AGB} - L \qquad Eq.1$$

The aboveground carbon gain, NPP_{AGB}, is calculated from gross primary production (GPP), the losses of C to growth respiration ($1 - Y_G$, Amthor 2000), and the fraction of biomass that is aboveground (f_{AGB}) as,

$$NPP_{AGB} = GPP \times (1 - Y_G) \times f_{AGB} \quad Eq.2$$

To simplify the experiments, the transfer ratio from GPP to NPP_{AGB} ($(1 - Y_G) \times f_{AGB}$) is fixed to a value of 0.5, representing 2/3 of C allocation to AGB and a growth respiration ratio of 0.25 (Amthor 2000).

The GPP dynamics are represented as a simple saturating exponential function of biomass. We acknowledge that the variations in the relationship between GPP and AGB for which we introduce a varying parameter G. Changes in G lead to different recovery trajectories and to a variability in the maximum photosynthetic capacity, representing the impact of species and local edaphoclimatic conditions in primary productivity dynamics between different landscapes.

$$GPP = \frac{100}{G + e^{-\frac{AGB}{1200}}} \quad Eq.3$$

The total carbon loss includes the carbon losses during disturbance events (L_d) and by the background mortality (L_b, Eq. 4) as,

$$L = L_b + L_d \quad Eq.4$$

The L_b is assumed to be a constant proportion of AGB, implicitly including the average effects of litterfall, root exudates, and herbivory (Thurner et al. 2016) as,

$$L_b = AGB \times K_b \quad Eq.5$$

$$K_b = \frac{1}{\tau} \quad Eq.6$$

where K_b refers to the background mortality rate, a reciprocal of turnover time (τ). To account for the fact that K_b is also spatially variable (e.g. Thurner et al., 2016), we define a range of K_b between 0.025 and 0.2, with an interval of 0.025, representing background turnover times from 5 to 40 years (Table 1). The other part of the carbon loss is caused by disturbance, L_d, determined by the intensity of disturbance event covering the patch as,

$$L_d = AGB \times f_L \qquad Eq.7$$

where, parameter f_L represents fraction of carbon loss during the disturbance, and it depends on the size of the event and the intensity slope (β, see Section 2.2).

2.2 Modeling different disturbance regimes

We applied three parameters to describe different disturbance regimes: the probability of disturbances, μ; the clustering patterns of events, α; and the intensity slope, β. These parameters represent the fraction of the domain affected by disturbances, the number and size of disturbed clusters of patches, and the fraction of carbon loss during each event, respectively. For the purpose of distributing a sufficiently large and spatially random number of disturbance events, our experiment is set on a domain of 1,000 by 1,000 pixels to simulate the corresponding landscape-scale domains. As in Fisher et al. (2008), we assume that each pixel (patch) represents one single canopy tree square with a 10m-by-10m size, and the total domain size is of 100 km^2.

2.2.1 Parameterization of disturbance regimes

Parameter μ refers to the total disturbed fraction of the domain, where D refers to the total domain size and D_a is the area of the domain affected by disturbances as,

$$D_a = D \times \mu \qquad Eq.8$$

For parameter α, we followed Fisher's method by applying the scaling exponent α to determine the clustering degree of events (Fisher et al. 2008) as,

$$n_{zi} = A z_i^{-\alpha} \qquad Eq.9$$

where α is the scaling exponent for the disturbance event clustering degree with a dimensionless unit, and z_i is specific event size. A is the proportionality factor, manipulated by the size of the total disturbed area and the setting of events size series as,

$$A = \frac{D_a}{\sum(z_i \cdot z_i^{-\alpha})} \qquad \textit{Eq.10}$$

where z_i is the corresponding disturbance event size. We applied an event size ranging between a patch to half the size of the domain during the simulations (see Appendix S1).

For parameter β, we assumed that the intensity of disturbance (f_L, fraction of carbon loss per event) is proportional to the logarithm of event size ($log_{10}(Z_i)$, Chambers et al. 2013). This relationship between intensity and event size is controlled by parameter β (intensity slope, Figure 3), and b is a constant value:

$$f_L = \beta \cdot log_{10}(z_i) + b \qquad \textit{Eq.11}$$

More details on the parameterization setup related to α and β (Eq.9 - 11) can be found in Appendix S1.

2.2.2 Disturbance Parameter Ranges

The inclusion of the parameters μ, α, and β allows a flexible description of disturbance regimes in different landscapes, containing a wide range of all possible disturbance regimes without specifying disturbance type. For providing realistic simulation scenarios we consulted existing literature to set different ranges and intervals for the primary productivity parameter G, background mortality rate K_b, and the three disturbance parameters independently (Table 1).

In particular, the parameter G is specified as a range from 0.03 to 0.1, imposing a range in GPP of 1 and 4 $kgC \cdot m^{-2} \cdot yr^{-1}$ and an average steady state biomass from 10 to 40 $kgC \cdot m^{-2}$, without disturbances. The parameter μ is set in a range of 0.01 to 0.05 with an interval of 0.005; this range substantially spans the average value of 0.02 documented in forests by Moorcroft et al. (2001) and Malhi et al. (2004). The settings for the clustering parameter α comes from observed gap-size distributions from previous studies in the tropical and sub-humid forest ecosystem, which indicated a range between 1.1-1.6 (Araujo et al. 2021; Lawton et al. 1988; Jans et al. 1993; Nelson et al. 1994; Yavitt et al. 1995; Fisher et al. 2008). In the experiment, we have increased the documented range slightly from 1.0 to 1.8, with an interval of 0.05. For parameter *β,* we have considered the findings from Chambers et al.'s (2013) to establish a range between 0.03 and 0.5. The intervals within that range differ, with a value of 0.01 for the range [0.03, 0.09], 0.05 for [0.1, 0.25], and 0.1 for [0.3, 0.5] (see Figure 3).

2.3 Generation of Disturbance Events

The prescription of the different disturbance regime parameters required the design of a stochastic disturbance event generator that distributes all of the events within the domain. We followed Fisher et al. (2008)'s strategy that distribute all of the different size disturbance events in rectangular shapes and descending order. The disturbance event generator assigns center coordinates randomly for each event in each class size; it checks for any overlaps with surrounding placed events; if an overlap is detected it reassigns the event's location. The location of an event is also reassigned when an event's location is partly placed outside of the domain. In this way the algorithm ensures an accurate prescription of each disturbance regime.

A full factorial combination of the different parameter's ranges and intervals sums up to a total of 2,142 combinations of disturbance regimes. For every disturbance regime, we generate a reference disturbance dataset for a 1000 × 1000 domain, consisting of 200 annual time steps, in the form of a three-dimensional array (a data cube). As such, every snapshot of the reference cube is a 1000 × 1000 stochastic disturbance reference map, providing the spatial reference for simulating the effect of disturbances in carbon cycling dynamics. To further impose variability in the resulting biomass distributions, for the same disturbance regime, the disturbance dataset was shuffled ten times in the time domain, generating a total of more than 2×10^4 simulation scenarios. The emerging AGB distributions from the different scenarios result from the variations in disturbance regimes, in productivity (G), and in background mortality (K_b) levels (Appendix S2), upon which three types of statistic metrics were used to characterize the biomass patterns at steady state on the domain scale.

2.4 Characterizing the Biomass Patterns

The first type of statistics utilized in our study is based on the histogram distribution, including mean, median, variance, skewness, kurtosis, percentiles, as well as standard deviation and coefficient of variation. Previous literature suggests that some of these metrics, such as skewness and mean, are associated with the probability and intensity of disturbances (Williams et al. 2013). We have additionally introduced information theory based metrics: the Shannon entropy index (Shannon 1948), also called Shannon-Wiener index, which is a widely used indicator for describing the diversity level in ecosystems (Spellerberg et al. 2003). Furthermore, we included statistical properties based on the texture of biomass distribution. Texture provides information about the spatial arrangement of intensities in an image (e.g. continuity, contrast, smoothness), in our case, the emergent biomass map. We utilized Gray-Level Co-Occurrence Matrices (GLCMs), one of the most common texture feature extraction methods based on image statistics, to study the spatial correlation properties by using gray-scaled images (Haralick et al. 1973). To characterize texture on the biomass maps we applied four statistics from the GLCMs method, namely: contrast; correlation; energy; and homogeneity (Table 2). For doing so, all biomass maps were re-scaled to a range of 0 to 255, and AGB outliers were substituted with the nearest neighboring pixel prior to rescaling to avoid contamination in texture features. These outliers are individual isolated pixels with substantially higher AGB values resulting from incidentally undisturbed locations during the 200 years of simulations.

The simulation of biomass dynamics using the carbon cycling model involved a comprehensive examination of 85,680 parameter combinations. This comprised of 2,142 disturbance regimes in combination with 5 primary productivity scenarios and 8 background mortality rate scenarios. Each parameter combination was executed ten times, with the reference disturbance cube being shuffled in a different sequence for each run. Ultimately, 17 statistics were extracted for each run, including mean GPP at the end of simulation, as well as 16 other statistics related to the steady-state biomass distribution.

2.5 Prediction and validation

We used a multi-output random forest regression method in Scikit-learn (Pedregosa et al. 2011) to investigate the relationship between the simulated biomass statistics and the prescribed disturbance regimes. To avoid overfitting, we implemented three cross-validation strategies (Appendix S3): Completely Random 10-fold method (CR), the Leave-One-Sequence-Out method (LOSO), and the Leave-One-Parameter-Out method (LOPO).

We use the Nash-Sutcliffe efficiency (NSE) to evaluate the performance of our prediction model (Nash et al. 1970). The NSE measures the accuracy of the predicted disturbance regime parameters compared to the prescribed values in this study. A higher NSE indicates better accuracy. The formula for calculating NSE (D. N. Moriasi et al. 2007) is shown in Eq.12,

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pre}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - \overline{Y^{obs}}\right)^2} \qquad Eq.12$$

where the Y_i^{obs} is the i^{th} prescribed disturbance regime parameter, Y_i^{pre} is the i^{th} predicted value for the corresponding parameter, $\overline{Y^{obs}}$ is the mean of the prescribed parameter, and n is the total number of observations.

3. Results

3.1 Parameterization and dynamic carbon model

3.1.1 Parameterization of disturbance regimes

The clustering parameter α defines the relationship between the size and frequency of disturbance events across the domain establishing that the number of events decreases exponentially with the event size (Appendix S4: Figure S1). This relationship is diagnosed from the simulated cubes. The average size of events in the domain exponentially decays as the parameter α increases, conversely, the number of events tends to a logistic increase, which confirms that larger α corresponds to more and smaller discrete events rather than few and larger ones. With discrete but progressive α values, the disturbance data cubes provide a gradient in the relationship between the amounts and sizes of events across the domain (Figure 2).

We have introduced a range in the event intensity parameter, β, that results in an intensity gradient for events of the same size between the disturbance data cubes. The relationship between β and event size, as defined in Eq.9, results in a logarithmic increase in disturbance intensity as the event size increases. Figure 3 shows this relationship for all the assigned β values in our experiment. The results demonstrate the distinct gradient across all curves, each translating a steeper increase between intensity and event sizes and β increases. When β was greater than 0.2, the disturbance intensity was 100% for events larger than ~2×10^3. Notably, β values of 0.5 impose a full intensity disturbance (full loss) for events of any size (Figure 3).

3.1.2 Temporal carbon dynamics

By employing the parsimonious carbon model, we can analyze the dynamics of AGB given trends in GPP alongside the carbon losses due to the background mortality and to the disturbance regime (Figure 4). The range of values for parameter G, designed to represent variations in the relationship between photosynthesis and biomass, produce a gradient in maximum GPP at steady states, with and without taking disturbances into account. Such gradient directly compares to the gradients in maximum AGB, both across the GPP gradients driven by G, with noteworthy variations in average biomass across domains (from ~10 $kgC \cdot m^{-2}$ to 40 $kgC \cdot m^{-2}$) under a no disturbance scenario. When considering a disturbance regime with parameters μ=0.03, α=1.0, and β=0.2, the steady-state average levels of GPP and AGB clearly exhibit a decrease. Yet, and particularly for this disturbance regime, the reduction in AGB (~58%) was more than twice the reductions in GPP (~23%). Additionally, with the introduction of disturbance events, the original smooth growth curves are replaced by higher frequency fluctuations that become increasingly pronounced with higher values of G. The comparison of GPP and AGB evolution over time in different disturbance regimes is provided in Appendix S5: Figure S1.

3.1.3 Steady-state biomass distribution

Under the same photosynthetic capacity (fixed G) and background mortality rate (fixed K_b), the steady-state biomass (year 200) exhibits diverse spatial patterns under different disturbance regimes, as shown in Figure 5. Increasing parameter μ while keeping the other two disturbance regime parameters fixed (first row in Figure 5) imposes an increasing fraction of the domain affected by disturbance that leads to more areas with lower biomass. As the clustering parameter α increases (second row in Figure 5), the same fraction of disturbance (same μ) is imposed through a larger number of small events. As such, the spatial distribution of the biomass becomes more homogeneous as α increases, shifting from distinguishable disturbance effects to a uniformly distributed noisy mosaic (from left to right, second row in Figure 5). The impact of higher β values is also evident, with more intense events resulting in greater biomass removal and lower plant regrowth levels, resulting in more contrasting disturbance footprints. The combination of the three disturbance regime parameters resulted in a large diversity of spatial patterns and biomass distributions, ideal to explore the retrieval of disturbance regimes from potential AGB observations via dynamic vegetation modeling.

3.2 Retrieving Disturbance Regimes from Biomass Patterns

3.2.1 Cross-Validation

Figure 6 shows that all the strategies of cross-validation have a good performance for retrieving the three disturbance regime parameters under various primary productivity and background mortality conditions. Relying on the 16 statistics (Table 2) calculated from the biomass and the average GPP of the last year of the simulation, all the NSEs of μ, α, and β exceeded 0.94 in the CR and LOSO validation. In addition, the LOPO validation exhibits a high degree of precision for estimating μ and β, with an NSE approximately 0.82 and 0.85, and a moderately accurate prediction for α with NSE of 0.69. This suggests that the model has the ability in predicting target disturbance regime parameters when they are not present in the training set, although the level of precision may vary.

The scatter plots of different CV strategies confirmed the high accuracy and apparent gradient for the predictions across the cross-validation strategies (from left to right, Figure 6). The results from the random cross-validation (CR validation, Figure 6a) as well as for the LOSO cross-validation approach (Figure 6b) show robust and high NSE values, and the regression lines close to the 1:1 line, indicating the great accuracy and high correlation with the prescribed values. The LOSO validation (Figure 6b) maintains a similar prediction accuracy with CR validation, but the LOPO results show lower NSE values, larger scatters, and a regression line up to 20% away from the 1:1 line. The performance reduction is especially for parameter α. One possible reason is that, when training, if certain boundary values of the parameters are absent, the results show an evident bias. To investigate this further, we conduct a test by replacing the LOPO CV predictions by the LOSO CV predictions specially for the μ, α, and β that fall at the boundaries of the prescribed intervals. As a result, we found a significant increase in precision, with the NSE for μ and β exceeding 0.91 and the NSE for α increasing to 0.76 (Figure 6d). These results confirm the extrapolation challenge faced by our multi-output random forest model when predicting parameter combinations outside the prescribed bounds.

3.2.2 Feature Importance

According to the feature importance analysis in the multi-output random forest method, it is observed all types of statistics played significant roles in predicting the disturbance regimes (Figure 7). Additionally, information on mean GPP contributed around 13% for the predictions, ranking it in the third position among the statistical metrics on AGB (Figure 7a). GPP, along with the texture feature correlation, and the coefficient of variation, are the main contributors to the predictions of disturbance regimes (close to 60% of total feature importance). In fact, these three features alone can explain more than 80% of the variation in the different disturbance parameters (Figure 7b).

To elucidate the association between the individual disturbance regime parameters and statistics, we retrained a random forest model specifically for predicting each single parameter. Upon considering features whose contribution exceeds 5%, we see that μ is primarily governed by features characterizing the statistical distribution of AGB (feature importance ~55% of the total contribution, Figure 8a). However, texture feature, correlation, is the second most important, accounting for 23% of the contribution, while GPP contributes only 6% for predicting μ. The importance of GPP is more apparent in predicting α (14%) and β (13%). Texture features have a higher significance in predicting parameter α than μ and β, contributing approximately

37%. The feature that made the highest contribution was correlation, which accounted for 30% of the prediction (Figure 8b). For parameter β, the coefficient of variation has a dominant contribution for prediction (~60%, Figure 8c). The analysis shows that for achieving and accuracy over 75% for predicting each of the parameters μ, α, and β individually, would require the top 4, 3, and 2 features, respectively (Figure 8).

4. Discussion

Disturbance regimes are usually defined by their frequency, severity, and spatial coverage (Liu et al. 2011), and can vary significantly across the landscape (Nelson et al. 1994). Previous approaches have explored the ability to infer different properties of disturbance regimes in several ways using modelling and observations, for example: Fisher et al. (2008) have focused on the number, size, and distribution of disturbance events; Williams et al. (2013) derived disturbance probability and intensity; while Chambers et al. (2013), following Fisher et al. (2008) focused on return frequency, area and tree mortality intensity. A common challenge across the different approaches is the equifinality, the inability to distinguish different disturbance regimes using AGB integrals or only distribution statistics across the landscape or simple regression approaches. Our results show that this challenge can be overcome by exploring higher complexity metrics, using primary productivity for constraining the problem and exploring machine learning approaches for multi-output regression problems, while some aspects require further discussion.

4.1 Experimental factorial design

Building on antecedent research, here we synthesize disturbance regimes in three overarching parameters μ, α, and β, which control the average area affected by disturbances, their event size-frequency relationship, and the event intensity, respectively. The emergent biomass distribution is further determined by primary productivity, controlled by G, and by the background mortality, K_b. We implemented a full factorial design in the variation of these five parameters, imposing a relatively wide range to simulate a sufficiently large sample of disturbance regime scenarios following Fisher et al. (2008) for μ and α, Chambers et al. (2013), for β, Thurner et al. (2016) for K_b, and allowing G to generate a range in GPP according to the current ranges emerging from the FLUXNET eddy covariance network (Pastorello et al. 2020). Although we ensured that the experimental setup spans the parametric ranges found in literature, further findings expanding these intervals should necessarily involve expansion of experimental design. Such aspect is especially relevant given the limitations in predicting boundary values (as shown in Figure 6c). Furthermore, although the event clustering and intensity parameters (α and β) span widely, here we follow a prescribed relationship between even size, frequency, and intensity (as in Fisher et al., 2008 and in Chambers et al., 2013). Alternative formulations on the links between size, frequency and intensity may necessarily lead to different statistical model structures.

Temporal independence in disturbance events

Different types of events may have different size-frequency-intensity relationships, or be temporally correlated (e.g. drought and insect outbreaks, Anderegg et al. 2015). In this regard, the current prescription of disturbance events is stochastic and, given the lack of quantitative information, temporally independent. The experimental setup is also insensitive to changes in local edaphoclimatic conditions after disturbances. Disturbance events, such as fire, can modify the physical and chemical properties in soils or local microclimatic environments, creating ecological legacies that have cascading implications on carbon dynamics (Liu et al. 2003). The ability to quantify how different disturbances change the posteriori growth conditions and vary the probability of subsequent disturbance events may be very local. However, in a context of climate change and for prognostic purposes, it will support constraining the temporal dynamics of disturbance regimes and potentially provide more realistic projections of carbon cycle responses to climate.

Shapes of disturbance events

In our experiments, following also previous studies, for simplicity the disturbance events are prescribed as rectangular shapes (Fisher et al. 2008; Williams et al. 2013). Shapes of disturbance events are usually more complex as demonstrated by high-resolution remote sensing observations (Chambers et al. 2013; Senf and Seidl 2021a). For most statistical properties, such as distribution or information theory metrics, this aspect is irrelevant. However, the importance of texture features for the prediction of μ and α, especially correlation, may reflect limitations in the generalization of the approach. We additionally conducted a simple exercise including disturbances with irregular shapes to confirm the performance of the approach and that the variable importance is kept (See Appendix S6). These results suggest that the landscape texture patterns are mainly controlled by the frequency-size relationship rather than by event shape, and hence type, of disturbances.

Local biomass outliers

Occasionally, local and sporadic very high AGB grid cells emerge from the simulations, even after the 200-year simulation period. The outlier, defined as AGB value greater than three standard deviations from the mean value of the column in which the AGB pixel is located (Appendix S7: Table S1), was filled with the nearest nonoutlier value. Although increasing the simulation years may theoretically mitigate this phenomenon we found it computationally very inefficient and difficult to control. The presence of such outliers changes some of the statistical features, especially distribution statistics, which leads to a reduction in the ability to predict disturbance regimes. This translates mostly that these local outliers can significantly impact AGB extremes metrics and by that breaking down the relationship between AGB distributions and the underlying disturbance regimes.

Large scale controls on primary productivity

Another assumption is the dominant role of climate in controlling landscape scale GPP (Wang et al. 2014). In our experimental design is the relationship between GPP and biomass, determined by parameter *G*. We prescribed a gradient in *G* that corresponds to a gradient in maximum GPP at landscape level between 1000 and 4000 $gC \cdot m^{-2} \cdot yr^{-1}$. We acknowledge the heterogeneity of photosynthetic capacity within the domain by randomly perturbing ¼ G for each patch. These perturbations represent a variation of 37% on maximum GPP. The approach falls short not consider the high frequency environmental controls on photosynthesis, such as solar radiation, temperature, vapor pressure deficit or soil water availability. As such, the approach stands on the assumption that landscape scale GPP is mostly controlled by climate rather than by meteorology (Wang et al. 2014; Pastorello et al. 2020).

4.2 Performance of the multi-output regression approach

Overall, the results from the different cross-validation exercises demonstrate the robustness of the approach (Figure 6). The challenges in estimating the disturbance regimes underlying different biomass distributions has been previously highlighted by modelling exercises due to the high spatiotemporal stochasticity that characterizes disturbance events and the equifinality found between disturbance regimes and total biomass distributions (Fisher et al. 2008; Ryan et al. 2012; Williams et al. 2013). We argue that the equifinality issue can be addressed by (1) expanding the feature space to include texture, diversity, and more comprehensive distribution metrics on the emerging biomass patterns; and (2) including additional information about primary productivity.

We find that deriving parameters outside the training bounds is a limitation (Figure 6c) but extending the parameter intervals can reduce this problem.

4.3 Landscape properties emerging from disturbances

The multi-output regression shows that the most important variables to predict μ, α, and β relate to the spatial distribution of biomass, rather than the mean or any higher quantiles of the biomass distribution (Figure 7). Yet, the feature importance rankings change across the three disturbance metrics. The average domain fraction affected by disturbances, μ, is strongly linked to biomass distribution statistics (AGBcv, AGBskew, AGBvar, AGBstd) and also to the correlation (Figure 8a). Interestingly biomass itself does not emerge as highly important, although it is implicitly present in AGBcv and AGBstd. The event size-frequency clustering parameter, α, displays a notable link to text features, such as correlation, homogeneity, and contrast, which account for ~37% of the overall contributions of all features (Figure 8b). The parameter controlling the relationship between disturbance intensity and event size, β, is mainly dominated by AGBcv (Figure 8c), contributing ~60% to β's prediction, followed by GPP (~13%). This connection is the most obvious among the three parameters, indicating that the intensity would directly affect the biomass distribution and indirectly the GPP.

4.4 Modeling forest carbon dynamics

In this study we used a simple carbon dynamics model to simulate primary productivity and growth resulting from carbon gains and losses at patch level. The realistic annual trajectories in primary productivity and AGB dynamics during recovery (Figure 4), along with model tractability, translate a clear benefit in model simplicity. However, this supports the analysis of the direct disturbances' impact without considering other detailed physiological processes and allocation mechanisms. Realistic biomass dynamics may describe different biomass compartments such as leaves, branches, stems and roots (e.g. CASA, Potter et al. 1993; JSBACH Reick et al. 2021; DALEC, Williams et al. 2005); differentiate tree density from carbon stocks and include individual and community processes to describe forest dynamics (Bugmann 2022). Often these models include the effects of climate and other factors on plant growth recovery, such as the impacts of changes in atmospheric CO_2 and in rising temperature (Norby et al. 2001; Pan et al. 2010). The present assumption here is that under given factors promoting primary productivity, disturbances exert the dominant controls on the AGB patterns. Yet we acknowledge the importance of bringing forward more comprehensive mechanistic models with more detailed carbon compartments and plant physiological processes, along with multiple observed constraints for further testing the robustness of these results and for differentiating regional biomass dynamics and corresponding disturbance regimes.

Through the use of remote sensing data and ground-based networks, significant advances have been achieved in understanding, representing, scaling, and characterizing disturbances, ultimately leading to the development of the hypotheses in the process-based models, which can generally be categorized into compartment models and demography models (Liu et al. 2011). The compartment models, including the biogeochemical and ecophysiological ones (Parton et al. 1987; Running et al. 1991; Raich et al. 1991; McGuire et al. 1992; Chen et al. 2000; Liu et al. 2003; Bond-Lamberty et al. 2005), can integrate general stand information and meteorological data to simulate carbon cycling, and applied to simulate the biogeochemical processes of forests associated with disturbance (Brugnach 2005; Tatarinov and Cienciala 2006; Wang et al. 2009). And the demography models, also referred to as gap models, are more focused on the simulation of the impacts of disturbance on the forest composition, structure, and biomass in a relatively long term (Shugart et al. 1992; Hurtt et al. 1998; Bugmann 2001; Norby et al. 2001). To better capture the fine-tuning functionality of plant sub-compartments and the effects of disturbance on demographic dynamics, the complexity of this type of gap models is increasing (Needham et al. 2022).

4.5 Perspectives on Earth observation

Overall, most of these processed-based models at the moment rely on field or satellite observations to quantify and evaluate the impacts of disturbance on modelled carbon stocks or fluxes. But model-data comparisons are far from trivial and often the model-observations mismatch is due to missing information, such as the extent, type, and timing of disturbance events. However, the prescription of individual disturbances based on disturbance regimes metrics derived from observations would minimize this problem and support the analysis of forest dynamics at larger scales. This level of detail in describing disturbances can be also transferred to dynamic global vegetation models (DGVMs), such as LPJ (Haxeltine et al. 1996), HYBRID (Friend et al. 1997), IBIS (Foley et al. 1996), VECODE (Brovkin et al. 1997), and LM3V (Shevliakova et al. 2009). Such could be done, e.g., by prescribing only μ and intensity losses of carbon at landscape level, or even via lumped parameters (e.g. whole landscape turnover rates), to describe the large scale impacts of disturbance on carbon dynamics. The proliferation of high-resolution biomass from Earth observation, as those by Saatchi et al. (2011) and Santoro et al. (2021), offers a valuable prospect for distinguishing different disturbance regimes.

We would further argue that metrics on disturbance regimes hold information about the different natural and anthropogenic drivers. For example, clearcutting and forest thinning may result in similar spatial patterns of biomass, but with different biomass loss fractions (Appendix S8: Figure S1 a-b). This could be reflected in the similar μ and α, but different β, which could be the same case for wildfire and drought (Appendix S8: Figure S1 c-d). However, tornado and insect outbreak (Appendix S8: Figure S1 e-f) differ in the shape of the affected area, resulting in different biomass clustering patterns, which can be reflected by different α. Therefore, these natural or anthropogenic factors may cause differential biomass patterns, which are reflected in distinct disturbance parameters.

Furthermore, by analyzing the spatial patterns of biomass stocks over time we can also detect shifts in disturbance regimes and subsequent successional changes. For example, extended dryness periods could induce extensive drought mortality in some regions, leading to higher values of μ and β which would signal transitions in disturbance regimes. Similarly, more frequent wind-throws would result in changes in event clustering patterns over a larger area, presumably indicated by a higher value of μ but smaller values of α. Another example is the intensification in management activities, such as clearcutting, which would cause significant changes in the disturbance regimes, resulting in larger values of μ, α and β. Information on the transition in disturbance regimes is key for improving our capacity to diagnose relevant changes in forest dynamics and for understanding the relationship between vegetation mortality, the carbon cycle dynamics and climate variability. It can assist us in identifying susceptible regions that are particularly vulnerable to disturbances, thereby facilitating the implementation of effective climate adaptation and mitigation strategies, but also management strategies and conservation efforts to mitigate the impacts of disturbance on both ecosystem functioning and biodiversity.

5. Conclusion

This study presents a framework for simulating disturbance events using a wide range of disturbance regime attributes. Together with a simple carbon dynamics model, we simulated the effects of disturbance events, from combinations of three landscape-level disturbance regime parameters: μ (probability scale), α (clustering degree), and β (intensity slope) on biomass patterns. We observe how changes in extent, frequency, and intensity characterizing the different disturbance regimes indeed lead to significantly different biomass patterns even for analogous primary productivity and background mortality inputs.

The emerging spatial biomass patterns are summarized in a set of metrics containing the central tendencies, moments, as well as information-based and texture information. Based on a conceptual model-based experiment and machine learning regression we demonstrate that with this set of summary metrics of biomass, along with primary productivity constraints, the set of disturbance regime parameters can be reliably retrieved. The average fraction of the domain affected by disturbances, the event size clustering exponent, and the perturbation intensity could be retrieved with an accuracy of 94.8%, 94.9%, and 97.1%, respectively.

As Earth observation efforts evolve to retrieve space-borne estimates of vegetation structure and patterns such as GEDI (Stavros et al. 2017), NISAR (Rosen et al. 2015), and BIOMASS (Le Toan et al. 2011), and as photosynthesis patterns are being estimated at high resolution (Cogliati et al. 2015; Jung et al. 2020), the methods presented in this study will open up avenues to provide observation based long-term diagnostics on the terrestrial carbon cycle and background disturbance patterns, which could be used to constrain Earth system models.

Acknowledgements

SW acknowledges support from the International Max Planck Research School for Biogeochemical Cycles (IMPRS-gBGC). SW and HY acknowledge EEBIOMASS, Project Office BIOMASS, support from the German Federal Ministry of Economics and Technology (50EE1904).

Author Contributions

NC and SW designed the experiments. In close collaboration with HY, SK, and NC, SW conducted the analysis and prepared the first draft of the manuscript. All authors contributed to the research discussions and contributed to the manuscript.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Allen, Craig, Alison Macalady, Dominique Bachelet, Nate McDowell, Michel Vennetier, Thomas Kitzberger, Andreas Rigling, et al. 2010. “A Global Overview of Drought and Heat-Induced Tree Mortality Reveals Emerging Climate Change Risks for Forests.” *Forest Ecology and Management* 259 (January): 660–84. https://doi.org/10.1016/j.foreco.2009.09.001.

Araujo, Raquel Fernandes, Samuel Grubinger, Carlos Henrique Souza Celes, Robinson I. Negrón-Juárez, Milton Garcia, Jonathan P. Dandois, and Helene C. Muller-Landau. 2021. “Strong Temporal Variation in Treefall and Branchfall Rates in a Tropical Forest Is Related to Extreme Rainfall: Results from 5 Years of Monthly Drone Data for a 50 Ha Plot.” *Biogeosciences* 18 (24): 6517–31. https://doi.org/10.5194/bg-18-6517-2021.

Amthor, Jeffrey S. 2000. “The McCree–de Wit–Penning de Vries–Thornley Respiration Paradigms: 30 Years Later.” *Annals of Botany* 86 (1): 1–20. https://doi.org/10.1006/anbo.2000.1175.

Anderegg, William R. L., Jeffrey A. Hicke, Rosie A. Fisher, Craig D. Allen, Juliann Aukema, Barbara Bentz, Sharon Hood, et al. 2015. “Tree Mortality from Drought, Insects, and

Their Interactions in a Changing Climate.” *New Phytologist* 208 (3): 674–83. https://doi.org/10.1111/nph.13477.

Bao, Shanning, Thomas Wutzler, Sujan Koirala, Matthias Cuntz, Andreas Ibrom, Simon Besnard, Sophia Walther, et al. 2022. “Environment-Sensitivity Functions for Gross Primary Productivity in Light Use Efficiency Models.” *Agricultural and Forest Meteorology* 312 (January): 108708. https://doi.org/10.1016/j.agrformet.2021.108708.

Bond-Lamberty, Ben, Stith T. Gower, Douglas E. Ahl, and Peter E. Thornton. 2005. “Reimplementation of the Biome-BGC Model to Simulate Successional Change.” *Tree Physiology* 25 (4): 413–24. https://doi.org/10.1093/treephys/25.4.413.

Bossel, Hartmut, and Holger Krieger. 1994. “Simulation of Multi-Species Tropical Forest Dynamics Using a Vertically and Horizontally Structured Model.” *Forest Ecology and Management*, Contrasts between biologically-based process models and management-oriented growth and yield models, 69 (1): 123–44. https://doi.org/10.1016/0378-1127(94)90224-0.

Brokaw, Nicholas V. L., and Samuel M. Scheiner. 1989. “Species Composition in Gaps and Structure of a Tropical Forest.” *Ecology* 70 (3): 538–41. https://doi.org/10.2307/1940196.

Brovkin, Victor, Andrei Ganopolski, and Yuri Svirezhev. 1997. “A Continuous Climate-Vegetation Classification for Use in Climate-Biosphere Studies.” *Ecological Modelling* 101 (2): 251–61. https://doi.org/10.1016/S0304-3800(97)00049-5.

Brugnach, Marcela. 2005. “Process Level Sensitivity Analysis for Complex Ecological Models.” *Ecological Modelling* 187 (2): 99–120. https://doi.org/10.1016/j.ecolmodel.2005.01.044.

Bugmann, Harald. 2001. “A Review of Forest Gap Models.” *Climatic Change* 51 (3): 259–305. https://doi.org/10.1023/A:1012525626267.

Bugmann, Harald, and Rupert Seidl. 2022. “The Evolution, Complexity and Diversity of Models of Long-Term Forest Dynamics.” *Journal of Ecology* 110 (10): 2288–2307. https://doi.org/10.1111/1365-2745.13989.

Carvalhais, Nuno, Matthias Forkel, Myroslava Khomik, Jessica Bellarby, Martin Jung, Mirco Migliavacca, Mingquan Mu, et al. 2014. “Global Covariation of Carbon Turnover Times with Climate in Terrestrial Ecosystems.” *Nature* 514 (7521): 213–17. https://doi.org/10.1038/nature13731.

Chambers, Jeffrey Q., Niro Higuchi, Liliane M. Teixeira, Joaquim dos Santos, Susan G. Laurance, and Susan E. Trumbore. 2004. “Response of Tree Biomass and Wood Litter to Disturbance in a Central Amazon Forest.” *Oecologia* 141 (4): 596–611. https://doi.org/10.1007/s00442-004-1676-2.

Chambers, Jeffrey Q., Robinson I. Negron-Juarez, Daniel Magnabosco Marra, Alan Di Vittorio, Joerg Tews, Dar Roberts, Gabriel H. P. M. Ribeiro, Susan E. Trumbore, and Niro Higuchi. 2013. “The Steady-State Mosaic of Disturbance and Succession across an Old-Growth Central Amazon Forest Landscape.” *Proceedings of the National Academy of Sciences* 110 (10): 3949–54. https://doi.org/10.1073/pnas.1202894110.

Chen, Wenjun, Jing Chen, and Josef Cihlar. 2000. “An Integrated Terrestrial Ecosystem Carbon-Budget Model Based on Changes in Disturbance, Climate, and Atmospheric Chemistry.” *Ecological Modelling* 135 (1): 55–79. https://doi.org/10.1016/S0304-3800(00)00371-9.

Cogliati, S., W. Verhoef, S. Kraft, N. Sabater, L. Alonso, J. Vicent, J. Moreno, M. Drusch, and R. Colombo. 2015. “Retrieval of Sun-Induced Fluorescence Using Advanced Spectral Fitting Methods.” *Remote Sensing of Environment* 169 (November): 344–57. https://doi.org/10.1016/j.rse.2015.08.022.

D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. “Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations.” *Transactions of the ASABE* 50 (3): 885–900. https://doi.org/10.13031/2013.23153.

Delbart, N., P. Ciais, J. Chave, N. Viovy, Y. Malhi, and T. Le Toan. 2010. “Mortality as a Key Driver of the Spatial Distribution of Aboveground Biomass in Amazonian Forest: Results from a Dynamic Vegetation Model.” *Biogeosciences* 7 (10): 3027–39. https://doi.org/10.5194/bg-7-3027-2010.

Fisher, Jeremy I., George C. Hurtt, R. Quinn Thomas, and Jeffrey Q. Chambers. 2008. “Clustered Disturbances Lead to Bias in Large-Scale Estimates Based on Forest Sample Plots.” *Ecology Letters* 11 (6): 554–63. https://doi.org/10.1111/j.1461-0248.2008.01169.x.

Foley, Jonathan A., I. Colin Prentice, Navin Ramankutty, Samuel Levis, David Pollard, Steven Sitch, and Alex Haxeltine. 1996. “An Integrated Biosphere Model of Land Surface Processes, Terrestrial Carbon Balance, and Vegetation Dynamics.” *Global Biogeochemical Cycles* 10 (4): 603–28. https://doi.org/10.1029/96GB02692.

Franklin, Jerry F., H. H. Shugart, and Mark E. Harmon. 1987. “Tree Death as an Ecological Process.” *BioScience* 37 (8): 550–56. https://doi.org/10.2307/1310665.

Friend, A. D., A. K. Stevens, R. G. Knox, and M. G. R. Cannell. 1997. “A Process-Based, Terrestrial Biosphere Model of Ecosystem Dynamics (Hybrid v3.0).” *Ecological Modelling* 95 (2): 249–87. https://doi.org/10.1016/S0304-3800(96)00034-8.

Friend, Andrew D., Wolfgang Lucht, Tim T. Rademacher, Rozenn Keribin, Richard Betts, Patricia Cadule, Philippe Ciais, et al. 2014. “Carbon Residence Time Dominates

Uncertainty in Terrestrial Vegetation Responses to Future Climate and Atmospheric CO_2.” *Proceedings of the National Academy of Sciences* 111 (9): 3280–85. https://doi.org/10.1073/pnas.1222477110.

Grime, J. P. 1977. “Evidence for the Existence of Three Primary Strategies in Plants and Its Relevance to Ecological and Evolutionary Theory.” *The American Naturalist* 111 (982): 1169–94.

Hammond, William M., A. Park Williams, John T. Abatzoglou, Henry D. Adams, Tamir Klein, Rosana López, Cuauhtémoc Sáenz-Romero, Henrik Hartmann, David D. Breshears, and Craig D. Allen. 2022. “Global Field Observations of Tree Die-off Reveal Hotter-Drought Fingerprint for Earth’s Forests.” *Nature Communications* 13 (1): 1761. https://doi.org/10.1038/s41467-022-29289-2.

Haralick, Robert M., K. Shanmugam, and Its’Hak Dinstein. 1973. “Textural Features for Image Classification.” *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–21. https://doi.org/10.1109/TSMC.1973.4309314.

Haxeltine, Alex, and I. Colin Prentice. 1996. “BIOME3: An Equilibrium Terrestrial Biosphere Model Based on Ecophysiological Constraints, Resource Availability, and Competition among Plant Functional Types.” *Global Biogeochemical Cycles* 10 (4): 693–709. https://doi.org/10.1029/96GB02344.

Hill, T., C. Ryan, and M. Williams. 2015. “A Framework for Estimating Forest Disturbance Intensity from Successive Remotely Sensed Biomass Maps: Moving beyond Average Biomass Loss Estimates.” *Carbon Balance and Management*. https://doi.org/10.1186/s13021-015-0039-0.

Hurtt, George C., PauL. R. Moorcroft, Stephen W. Pacala And, and Simon A. Levin. 1998.

"Terrestrial Models and Global Change: Challenges for the Future." *Global Change Biology* 4 (5): 581–90. https://doi.org/10.1046/j.1365-2486.1998.t01-1-00203.x.

Jans, Luc, Lourens Poorter, Renaat S. A. R. van Rompaey, and Frans Bongers. 1993. "Gaps and Forest Zones in Tropical Moist Forest in Ivory Coast." *Biotropica* 25 (3): 258–69. https://doi.org/10.2307/2388784.

Johnson, Michelle O., David Galbraith, Manuel Gloor, Hannes De Deurwaerder, Matthieu Guimberteau, Anja Rammig, Kirsten Thonicke, et al. 2016. "Variation in Stem Mortality Rates Determines Patterns of Above-Ground Biomass in Amazonian Forests: Implications for Dynamic Global Vegetation Models." *Global Change Biology* 22 (12): 3996–4013. https://doi.org/10.1111/gcb.13315.

Jung, Martin, Christopher Schwalm, Mirco Migliavacca, Sophia Walther, Gustau Camps-Valls, Sujan Koirala, Peter Anthoni, et al. 2020. "Scaling Carbon Fluxes from Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach." *Biogeosciences* 17 (5): 1343–65. https://doi.org/10.5194/bg-17-1343-2020.

Köhler, Peter, and Andreas Huth. 1998. "The Effects of Tree Species Grouping in Tropical Rainforest Modelling: Simulations with the Individual-Based Model Formind." *Ecological Modelling* 109 (3): 301–21. https://doi.org/10.1016/S0304-3800(98)00066-0.

Lawton, Robert O., and Francis E. Putz. 1988. "Natural Disturbance and Gap-Phase Regeneration in a Wind-Exposed Tropical Cloud Forest." *Ecology* 69 (3): 764–77. https://doi.org/10.2307/1941025.

Le Toan, T., S. Quegan, M. W. J. Davidson, H. Balzter, P. Paillou, K. Papathanassiou, S. Plummer, et al. 2011. "The BIOMASS Mission: Mapping Global Forest Biomass to Better Understand the Terrestrial Carbon Cycle." *Remote Sensing of Environment*, DESDynI VEG-3D Special Issue, 115 (11): 2850–60. https://doi.org/10.1016/j.rse.2011.03.020.

Liu, Shuguang, Norman Bliss, Eric Sundquist, and Thomas G. Huntington. 2003. "Modeling Carbon Dynamics in Vegetation and Soil under the Impact of Soil Erosion and Deposition." *Global Biogeochemical Cycles* 17 (2). https://doi.org/10.1029/2002GB002010.

Liu, Shuguang, Ben Bond-Lamberty, Jeffrey A. Hicke, Rodrigo Vargas, Shuqing Zhao, Jing Chen, Steven L. Edburg, et al. 2011. "Simulating the Impacts of Disturbances on Forest Carbon Cycling in North America: Processes, Data, Models, and Challenges." *Journal of Geophysical Research: Biogeosciences* 116 (G4). https://doi.org/10.1029/2010JG001585.

Malhi, Y., O. L. Phillips, O. L. Phillips, T. R. Baker, L. Arroyo, N. Higuchi, T. J. Killeen, et al. 2004. "Pattern and Process in Amazon Tree Turnover, 1976–2001." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359 (1443): 381–407. https://doi.org/10.1098/rstb.2003.1438.

McDowell, Nate G., Gerard Sapes, Alexandria Pivovaroff, Henry D. Adams, Craig D. Allen, William R. L. Anderegg, Matthias Arend, et al. 2022. "Mechanisms of Woody-Plant Mortality under Rising Drought, CO2 and Vapour Pressure Deficit." *Nature Reviews Earth & Environment*, March, 1–15. https://doi.org/10.1038/s43017-022-00272-1.

McGuire, A. D., J. M. Melillo, L. A. Joyce, D. W. Kicklighter, A. L. Grace, B. Moore III, and C. J. Vorosmarty. 1992. "Interactions between Carbon and Nitrogen Dynamics in Estimating Net Primary Productivity for Potential Vegetation in North America." *Global Biogeochemical Cycles* 6 (2): 101–24. https://doi.org/10.1029/92GB00219.

Moorcroft, P. R., G. C. Hurtt, and S. W. Pacala. 2001. "A Method for Scaling Vegetation

Dynamics: The Ecosystem Demography Model (Ed).” *Ecological Monographs* 71 (4): 557–86. https://doi.org/10.1890/0012-9615(2001)071[0557:AMFSVD]2.0.CO;2.

Nash, J. E., and J. V. Sutcliffe. 1970. “River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles.” *Journal of Hydrology* 10 (3): 282–90. https://doi.org/10.1016/0022-1694(70)90255-6.

Needham, Jessica F., Gabriel Arellano, Stuart J. Davies, Rosie A. Fisher, Valerie Hammer, Ryan G. Knox, David Mitre, Helene C. Muller-Landau, Daniel Zuleta, and Charlie D. Koven. 2022. “Tree Crown Damage and Its Effects on Forest Carbon Cycling in a Tropical Forest.” *Global Change Biology* 28 (18): 5560–74. https://doi.org/10.1111/gcb.16318.

Nelson, Bruce W., Valerie Kapos, John B. Adams, Wilson J. Oliveira, and Oscar P. G. Braun. 1994. “Forest Disturbance by Large Blowdowns in the Brazilian Amazon.” *Ecology* 75 (3): 853–58. https://doi.org/10.2307/1941742.

Norby, Richard J., Kiona Ogle, Peter S. Curtis, Franz-W. Badeck, Andreas Huth, George C. Hurtt, Takashi Kohyama, and Josep Peñuelas. 2001. “Aboveground Growth and Competition in Forest Gap Models: An Analysis for Studies of Climatic Change.” *Climatic Change* 51 (3): 415–47. https://doi.org/10.1023/A:1012510619424.

Pan, Y., J. M. Chen, R. Birdsey, K. McCullough, L. He, and F. Deng. 2010. “Age Structure and Disturbance Legacy of North American Forests.” *Biogeoscience Discussions. 7: 979-1020.* 7: 979–1020. https://doi.org/10.5194/bgd-7-979-2010.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. “The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data.” *Scientific Data* 7 (1): 225. https://doi.org/10.1038/s41597-020-0534-3.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima. 1987. "Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands." *Soil Science Society of America Journal* 51 (5): 1173–79. https://doi.org/10.2136/sssaj1987.03615995005100050015x.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (85): 2825–30.

Potter, Christopher S., James T. Randerson, Christopher B. Field, Pamela A. Matson, Peter M. Vitousek, Harold A. Mooney, and Steven A. Klooster. 1993. "Terrestrial Ecosystem Production: A Process Model Based on Global Satellite and Surface Data." *Global Biogeochemical Cycles* 7 (4): 811–41. https://doi.org/10.1029/93GB02725.

Raich, J. W., E. B. Rastetter, J. M. Melillo, D. W. Kicklighter, P. A. Steudler, B. J. Peterson, A. L. Grace, B. Moore III, and C. J. Vorosmarty. 1991. "Potential Net Primary Productivity in South America: Application of a Global Model." *Ecological Applications* 1 (4): 399–429. https://doi.org/10.2307/1941899.

Reick, Christian H., Veronika Gayler, Daniel Goll, Stefan Hagemann, Marvin Heidkamp, Julia E. M. S. Nabel, Thomas Raddatz, Erich Roeckner, Reiner Schnur, and Stiig Wilkenskjeld. 2021. "JSBACH 3 - The Land Component of the MPI Earth System Model: Documentation of Version 3.2," February. https://doi.org/10.17617/2.3279802.

Rosen, Paul A., Scott Hensley, Scott Shaffer, Louise Veilleux, Manab Chakraborty, Tapan Misra, Rakesh Bhan, V. Raju Sagi, and R. Satish. 2015. "The NASA-ISRO SAR Mission - An International Space Partnership for Science and Societal Benefit." In *2015 IEEE Radar Conference (RadarCon)*, 1610–13.

https://doi.org/10.1109/RADAR.2015.7131255.

Runkle, James R. 2000. “Canopy Tree Turnover in Old-Growth Mesic Forests of Eastern North America.” *Ecology* 81 (2): 554–67. https://doi.org/10.2307/177448.

Running, Steven W., and Stith T. Gower. 1991. “FOREST-BGC, A General Model of Forest Ecosystem Processes for Regional Applications. II. Dynamic Carbon Allocation and Nitrogen Budgets1.” *Tree Physiology* 9 (1–2): 147–60. https://doi.org/10.1093/treephys/9.1-2.147.

Ryan, Casey M., Timothy Hill, Emily Woollen, Claire Ghee, Edward Mitchard, Gemma Cassells, John Grace, Iain H. Woodhouse, and Mathew Williams. 2012. “Quantifying Small-Scale Deforestation and Forest Degradation in African Woodlands Using Radar Imagery.” *Global Change Biology* 18 (1): 243–57. https://doi.org/10.1111/j.1365-2486.2011.02551.x.

Ryan, Casey M., and Mathew Williams. 2011. “How Does Fire Intensity and Frequency Affect Miombo Woodland Tree Populations and Biomass?” Ecological Applications 21 (1): 48–60. https://doi.org/10.1890/09-1489.1.

Saatchi, Sassan S., Nancy L. Harris, Sandra Brown, Michael Lefsky, Edward T. A. Mitchard, William Salas, Brian R. Zutta, et al. 2011. “Benchmark Map of Forest Carbon Stocks in Tropical Regions across Three Continents.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (24): 9899–9904. https://doi.org/10.1073/pnas.1019576108.

Santoro, Maurizio, Oliver Cartus, Nuno Carvalhais, Danaë M. A. Rozendaal, Valerio Avitabile, Arnan Araza, Sytze de Bruin, et al. 2021. “The Global Forest Above-Ground Biomass Pool for 2010 Estimated from High-Resolution Satellite Observations.” *Earth System*

Science Data 13 (8): 3927–50. https://doi.org/10.5194/essd-13-3927-2021.

Senf, Cornelius, and Rupert Seidl. 2021a. “Storm and Fire Disturbances in Europe: Distribution and Trends.” *Global Change Biology* 27 (15): 3605–19. https://doi.org/10.1111/gcb.15679.

———. 2021b. “Mapping the Forest Disturbance Regimes of Europe.” *Nature Sustainability* 4 (1): 63–70. https://doi.org/10.1038/s41893-020-00609-y.

Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shevliakova, Elena, Stephen W. Pacala, Sergey Malyshev, George C. Hurtt, P. C. D. Milly, John P. Caspersen, Lori T. Sentman, Justin P. Fisk, Christian Wirth, and Cyril Crevoisier. 2009. “Carbon Cycling under 300 Years of Land Use Change: Importance of the Secondary Vegetation Sink.” *Global Biogeochemical Cycles* 23 (2). https://doi.org/10.1029/2007GB003176.

Shugart, Herman H., Rik Leemans, and Gordon B. Bonan, eds. 1992. *A Systems Analysis of the Global Boreal Forest*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511565489.

Spellerberg, Ian F., and Peter J. Fedor. 2003. “A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the ‘Shannon–Wiener’ Index.” *Global Ecology and Biogeography* 12 (3): 177–79. https://doi.org/10.1046/j.1466-822X.2003.00015.x.

Stavros, E. Natasha, David Schimel, Ryan Pavlick, Shawn Serbin, Abigail Swann, Laura Duncanson, Joshua B. Fisher, et al. 2017. “ISS Observations Offer Insights into Plant Function.” *Nature Ecology & Evolution* 1 (7): 1–5. https://doi.org/10.1038/s41559-017-

0194.

Tatarinov, Fyodor A., and Emil Cienciala. 2006. “Application of BIOME-BGC Model to Managed Forests: 1. Sensitivity Analysis.” *Forest Ecology and Management* 237 (1): 267–79. https://doi.org/10.1016/j.foreco.2006.09.085.

Thurner, Martin, Christian Beer, Nuno Carvalhais, Matthias Forkel, Maurizio Santoro, Markus Tum, and Christiane Schmullius. 2016. “Large-Scale Variation in Boreal and Temperate Forest Carbon Turnover Rate Related to Climate.” *Geophysical Research Letters* 43 (9): 4576–85. https://doi.org/10.1002/2016GL068794.

Turner, Monica G. 2010. “Disturbance and Landscape Dynamics in a Changing World.” *Ecology* 91 (10): 2833–49. https://doi.org/10.1890/10-0097.1.

Wang, H., I. C. Prentice, and T. W. Davis. 2014. “Biophsyical Constraints on Gross Primary Production by the Terrestrial Biosphere.” *Biogeosciences* 11 (20): 5987–6001. https://doi.org/10.5194/bg-11-5987-2014.

Wang, Weile, Kazuhito Ichii, Hirofumi Hashimoto, Andrew R. Michaelis, Peter E. Thornton, Beverly E. Law, and Ramakrishna R. Nemani. 2009. “A Hierarchical Analysis of Terrestrial Ecosystem Model Biome-BGC: Equilibrium Analysis and Model Calibration.” *Ecological Modelling* 220 (17): 2009–23. https://doi.org/10.1016/j.ecolmodel.2009.04.051.

Wang, Siyuan. 2023. “Disturbance Regime: Updates for peer-reviewing (v1.2)”. *Zenodo*. https://doi.org/10.5281/zenodo.8121119.

Williams, Mathew, Timothy C. Hill, and Casey M. Ryan. 2013. “Using Biomass Distributions to Determine Probability and Intensity of Tropical Forest Disturbance.” *Plant Ecology & Diversity* 6 (1): 87–99. https://doi.org/10.1080/17550874.2012.692404.

Williams, Mathew, Paul A. Schwarz, Beverly E. Law, James Irvine, and Meredith R. Kurpius. 2005. “An Improved Analysis of Forest Carbon Dynamics Using Data Assimilation.” *Global Change Biology* 11 (1): 89–105. https://doi.org/10.1111/j.1365-2486.2004.00891.x.

Yan, Zhihong, Junan Li, Junjei Fei, Xun Mao, Peng Gao, and Yonglan Ding. 2005. “Study on the Adsorptive Catalytic Voltammetry of Emodin at a Carbon Paste Electrode.” *Analytical Letters* 38 (10): 1641–50. https://doi.org/10.1081/AL-200065808.

Yavitt, J., J. Battles, G. E. Lang, and D. H. Knight. 1995. “The Canopy Gap Regime in a Secondary Neotropical Forest in Panama.” https://doi.org/10.1017/S0266467400008853.

Tables

Table 1. Parameter Setting

Parameter	Range	Interval	Count
$\boldsymbol{\mu}$	[0.01:0.05]	0.005	9
$\boldsymbol{\alpha}$	[1.0:1.8]	0.05	17
$\boldsymbol{\beta}$	[0.03:0.5]	0.01/0.05/0.1	14
$\boldsymbol{G}$	[0.03:0.1]	0.01/0.02/0.03	5
K_b	[0.025:0.2]	0.025	8

Table 2. Statistics of the steady-state biomass map

Feature types	Statistic	Variable Names	Formula	
Histogram features	mean	AGBmean	$\frac{\sum_{i=1}^{N} A_i}{N}$	N: Total patch amount A_i: Biomass value for patch i
	median	AGBmed	$Med(A)$	A: Biomass map
	range	AGBrange	$P_{90} - P_{10}$	P90: Percentile 90% P10: Percentile 10%
	variance	AGBvar	$\frac{1}{N-1}\sum_{i=1}^{N} \lvert A_i - \mu \rvert^2$	μ: Mean biomass
	standard deviation	AGBstd	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \lvert A_i - \mu \rvert^2}$	
	coefficient of variation	AGBcv	$100 \times \frac{\sigma}{\mu}$	σ: Standard deviation
	skewness	AGBskew	$\frac{\sum_{i=1}^{N}(A_i-\mu)^3}{(N-1)\sigma^3}$	
	kurtosis	AGBkurt	$\frac{\sum_{i=1}^{N}(A_i-\mu)^4}{(N-1)\sigma^4}$	
	percentile 25%	AGBp25	P_{25}	P25: Percentile 25%
	percentile 75%	AGBp75	P_{75}	P75: Percentile 75%
	Trimean	AGBtrim	$(P_{25} + 2 \times MED + P_{75})/4$	MED: Median value
Informative feature	Shannon entropy	Shannon	$-\sum p \cdot log_2(p)$	p: Normalized histogram counts
Texture features	contrast	contrast	$\sum_{i,j} \lvert i - j \rvert^2 glc(i,j)$	i: Reference pixel value j: Neighbor pixel value $glc(i,j)$: An entry in GLCM μ_i, μ_j: Means of GLCM w.r.t. i and j σ_i, σ_j: Standard deviations of GLCM w.r.t. i and j
	correlation	correlation	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)glc(i,j)}{\sigma_i \sigma_j}$	
	energy	energy	$\sum_{i,j} glc(i,j)^2$	

	homogeneity	homogeneity	$\sum_{i,j}\frac{glc(i,j)}{1+\|i-j\|}$

Figure Captions

Figure 1. Conceptual diagram of disturbance reference cubes, (a) displays a two-dimensional disturbance reference map, which represents a snapshot composed of a 1,000 x 1,000 array from a disturbance cube array, (b) showcases a three-dimensional disturbance reference cube consisting of 200 snapshots. Each snapshot simulates a unique stochastic spatial distribution of disturbance events, and the cube features a distinct combination of μ and α. It is essential to note that the disturbance events within the cube should not (c) interface with edges or (d) overlap with one another. (e) and (f) depict two examples of disturbance reference cubes featuring different disturbance regimes.

Figure 2. The parameter α exerts control over the total number and average sizes of events in all disturbance cubes generated with the same total disturbed area (i.e., same μ), exhibiting an exponential relationship. Specifically, an increase in α results in a higher proportion of relatively small events, whereas a decrease in α tends to produce fewer, yet larger events.

Figure 3. Logistic correlation between the intensity of disturbance event and its size. The Y-axis displays the intensity value, which refers to the proportion of biomass loss attributed to the event, while the X-axis represents the gradient event size. The range of sizes is divided into two parts for ease of visualization: a linear scale for events under 32 pixels and a logarithmic scale for larger events. Notably, the curve of β-0.5 is indistinct since it saturates at an intensity of 1 from the outset of the event size.

Figure 4. The evolution of AGB and GPP is examined under varying values of Parameter G. Figures (a) and (b) represent the scenario without any disturbance events, while figures (c) and (d) demonstrate the impact of a disturbance regime with μ-0.03, α-1.0, and β-0.2.

Figure 5. The map of steady-state biomass is compared across different disturbance regimes. The first row illustrates the impact of increasing μ, which results in more areas with low biomass values. In the second row, the effect of increasing α is represented, and the low biomass areas tend to be more scattered rather clustered in big events. In the third row, the consequences of increasing β are demonstrated, with more pronounced "prints" left by disturbances.

Figure 6. Different cross validation accuracy for predicting three disturbance regime parameters. The X-axis denotes the predicted values, and the Y-axis denotes the prescribed values. (a) illustrates the prediction results of the disturbance regime parameters in the Completely Random cross-validation strategy (CR), (b) refers to the Leave One Sequence Out strategy (LOSO), and (c) refers to the Leave One Parameter Out strategy(LOPO), (d) the LOPO predictions of μ, α, and β at the boundaries were substituted with the LOSO predictions to validate the extrapolation challenge. Specifically, the trained model in LOSO predicted the values of parameter α at 1.0 and 1.8, as well as the values of 0.01 and 0.05 for μ, and 0.03 and 0.5 for β.

Figure 7. (a) shows the feature importance of multi-output disturbance regime prediction, where the assigned value denotes the degree of contribution made by each feature (see the defination in Table 2). (b) shows the prediction accuracy change for each disturbance regime parameter, ordered by ascending feature importance. The X-axis represents the feature(s) used (right) and excluded (left) during the prediction process. For instance, when X-axis is labeled as GPP, it means that the prediction process only involved three features, which are GPP, AGBcv, and correlation, and all the other features in the left are excluded. The accuracy was measured using

the NSE metric, based on the prescribed parameters and the results of multi-output random forest model's prediction.

Figure 8. Breakdown of the feature importance for three individual disturbance regime parameters, μ(a), α(b), and β(c). The corresponding feature importance is depicted through bars, while the colored lines represent the results of the cumulatively exclusive feature test. This test is similar to the one shown in Figure 7(b) but employs a single-output random forest model for three parameters individually.

Figures

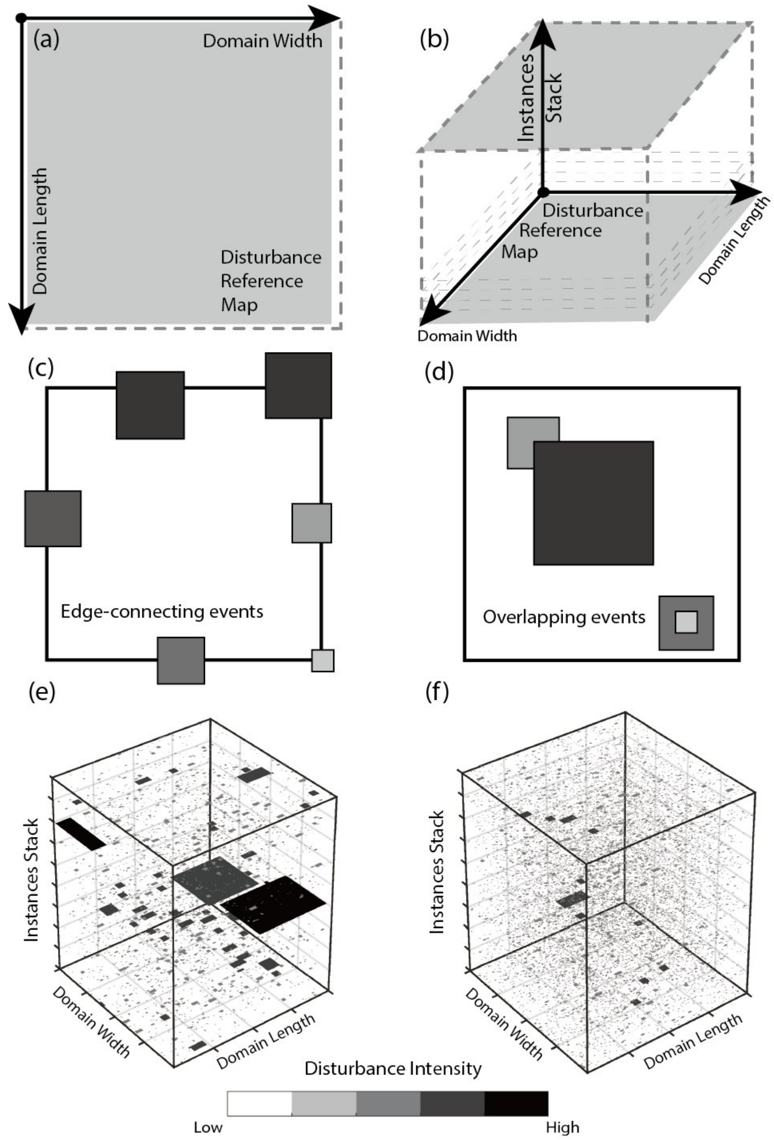


Figure 1

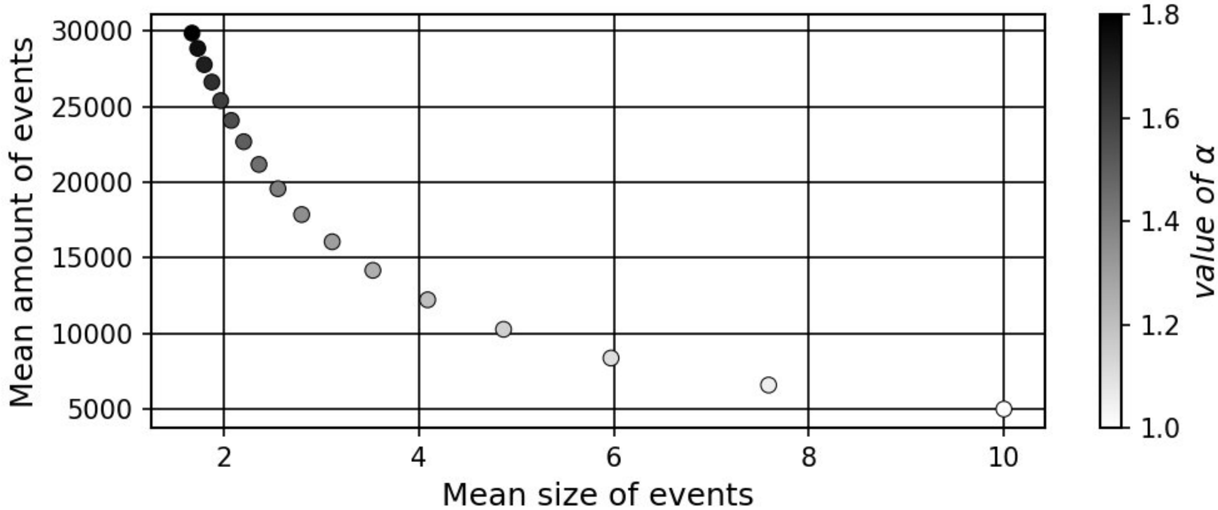

30000
25000
20000
15000
10000
5000
Mean amount of events
2
4
6
8
10
Mean size of events
1.8
1.6
1.4
1.2
1.0
value of α

Figure 2

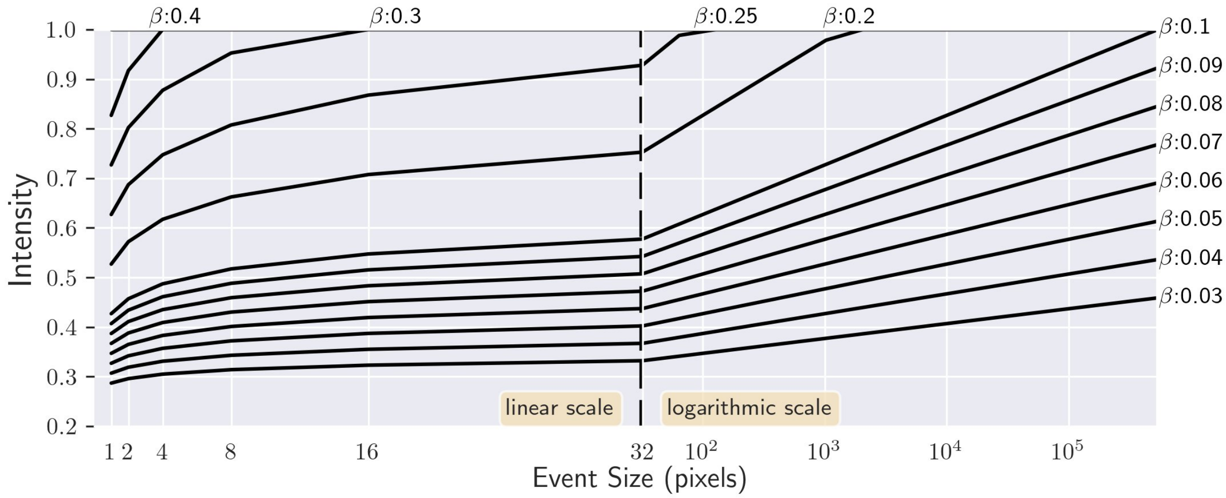

β:0.4
β:0.3
β:0.25
β:0.2
β:0.1
β:0.09
β:0.08
β:0.07
β:0.06
β:0.05
β:0.04
β:0.03
Intensity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
linear scale
logarithmic scale
1 2 4 8 16 32 10^2 10^3 10^4 10^5
Event Size (pixels)

Figure 3

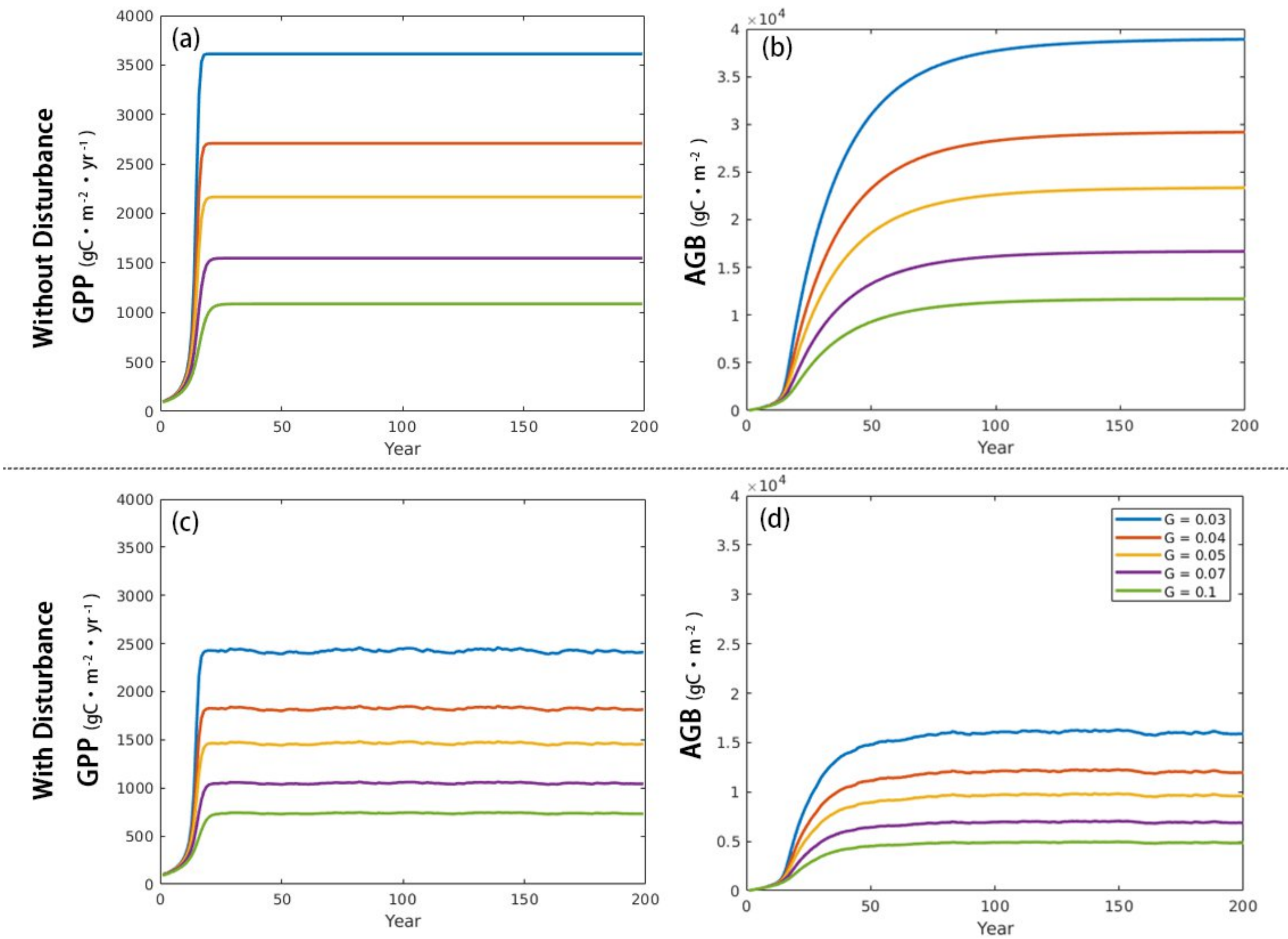
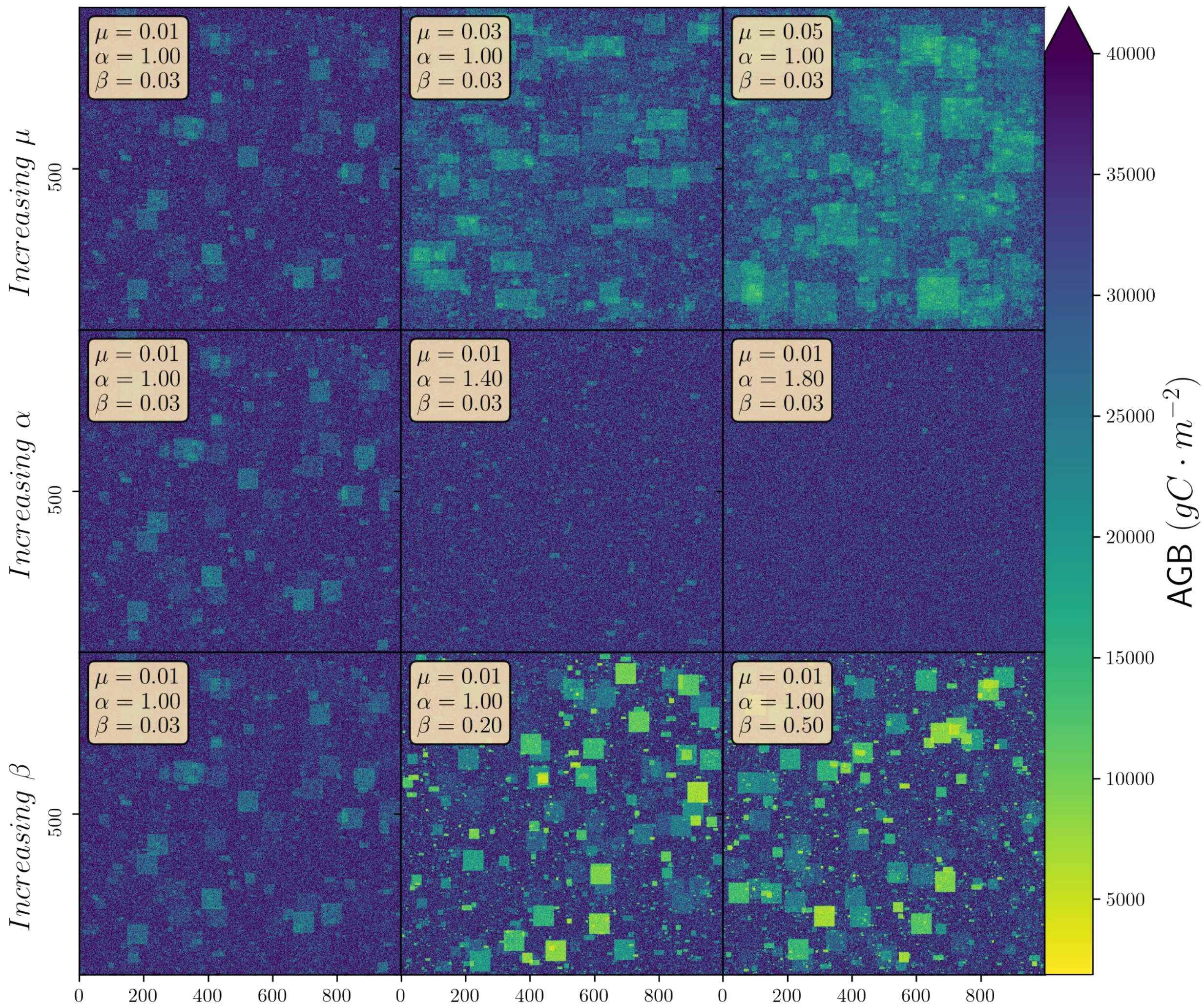
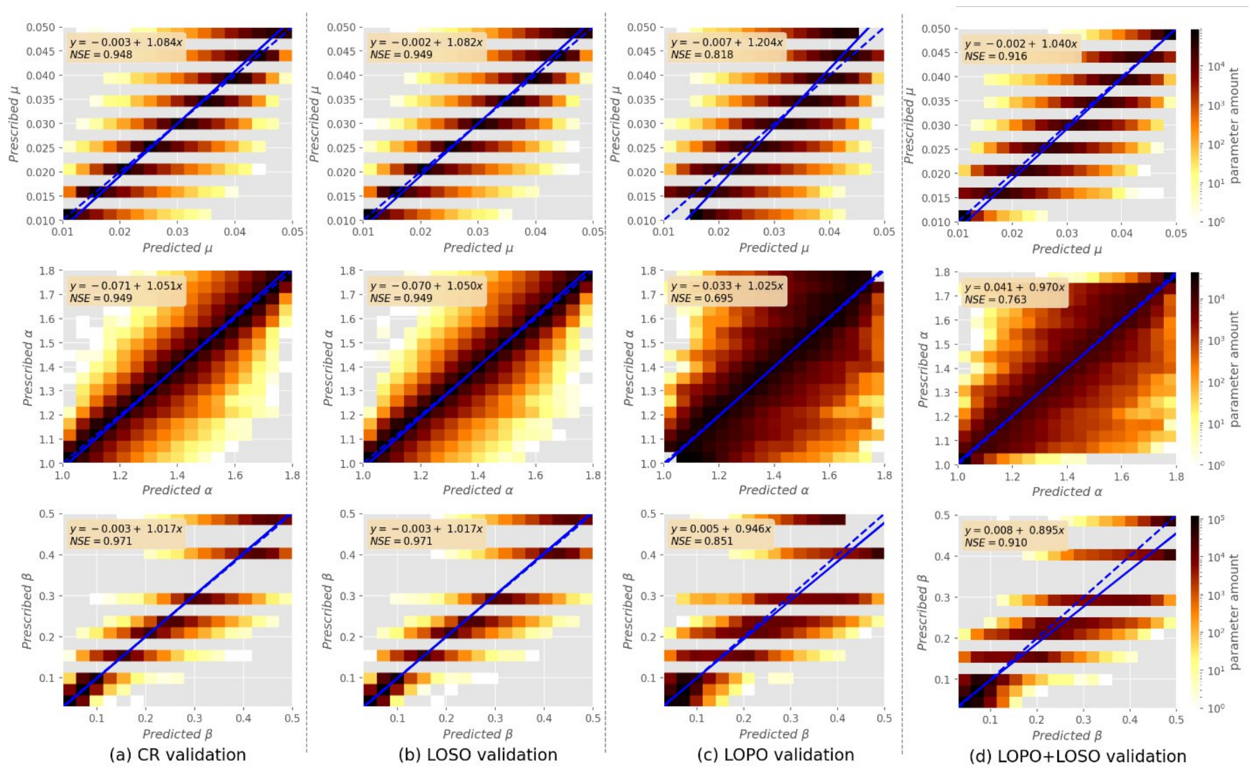

(a)
(b)
(c)
(d)
Without Disturbance
With Disturbance
GPP (gC • m-2 • yr-1)
AGB (gC • m-2)
×10^4
Year
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500
4000
0.5
1
1.5
2
2.5
3
3.5
4
G = 0.03
G = 0.04
G = 0.05
G = 0.07
G = 0.1

Figure 4



$\mu = 0.01$ $\alpha = 1.00$ $\beta = 0.03$
$\mu = 0.03$ $\alpha = 1.00$ $\beta = 0.03$
$\mu = 0.05$ $\alpha = 1.00$ $\beta = 0.03$
$\mu = 0.01$ $\alpha = 1.00$ $\beta = 0.03$
$\mu = 0.01$ $\alpha = 1.40$ $\beta = 0.03$
$\mu = 0.01$ $\alpha = 1.80$ $\beta = 0.03$
$\mu = 0.01$ $\alpha = 1.00$ $\beta = 0.03$
$\mu = 0.01$ $\alpha = 1.00$ $\beta = 0.20$
$\mu = 0.01$ $\alpha = 1.00$ $\beta = 0.50$
Increasing μ
Increasing α
Increasing β
500
0 200 400 600 800
AGB $(gC \cdot m^{-2})$
40000
35000
30000
25000
20000
15000
10000
5000

Figure 5



$y = -0.003 + 1.084x$
$NSE = 0.948$
$y = -0.002 + 1.082x$
$NSE = 0.949$
$y = -0.007 + 1.204x$
$NSE = 0.818$
$y = -0.002 + 1.040x$
$NSE = 0.916$
$y = -0.071 + 1.051x$
$NSE = 0.949$
$y = -0.070 + 1.050x$
$NSE = 0.949$
$y = -0.033 + 1.025x$
$NSE = 0.695$
$y = 0.041 + 0.970x$
$NSE = 0.763$
$y = -0.003 + 1.017x$
$NSE = 0.971$
$y = -0.003 + 1.017x$
$NSE = 0.971$
$y = 0.005 + 0.946x$
$NSE = 0.851$
$y = 0.008 + 0.895x$
$NSE = 0.910$
Prescribed μ
Predicted μ
Prescribed α
Predicted α
Prescribed β
Predicted β
parameter amount
(a) CR validation
(b) LOSO validation
(c) LOPO validation
(d) LOPO+LOSO validation

Figure 6

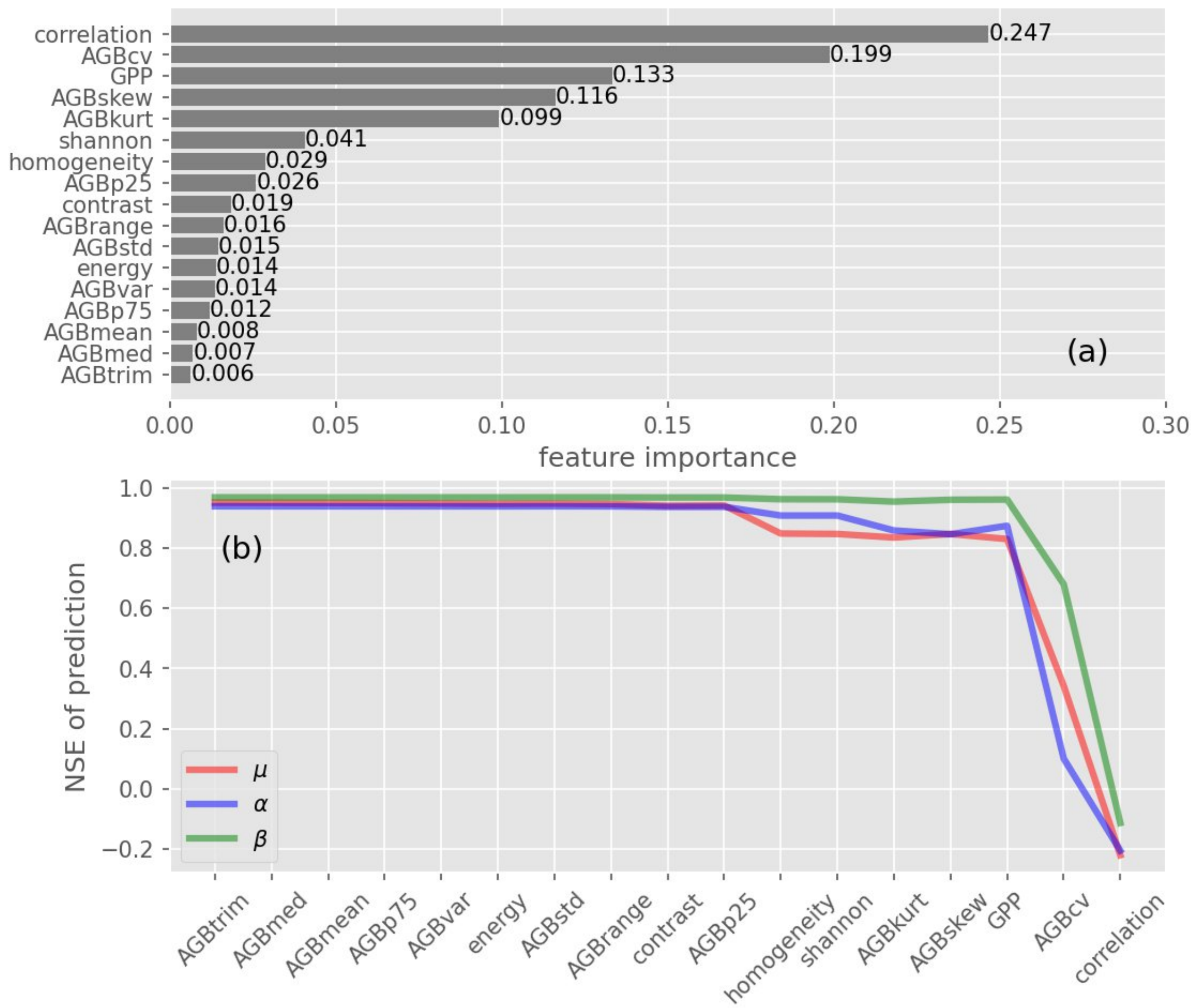

correlation
AGBcv
GPP
AGBskew
AGBkurt
shannon
homogeneity
AGBp25
contrast
AGBrange
AGBstd
energy
AGBvar
AGBp75
AGBmean
AGBmed
AGBtrim
0.247
0.199
0.133
0.116
0.099
0.041
0.029
0.026
0.019
0.016
0.015
0.014
0.014
0.012
0.008
0.007
0.006
(a)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
feature importance
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
NSE of prediction
(b)
μ
α
β
AGBtrim
AGBmed
AGBmean
AGBp75
AGBvar
energy
AGBstd
AGBrange
contrast
AGBp25
homogeneity
shannon
AGBkurt
AGBskew
GPP
AGBcv
correlation

Figure 7

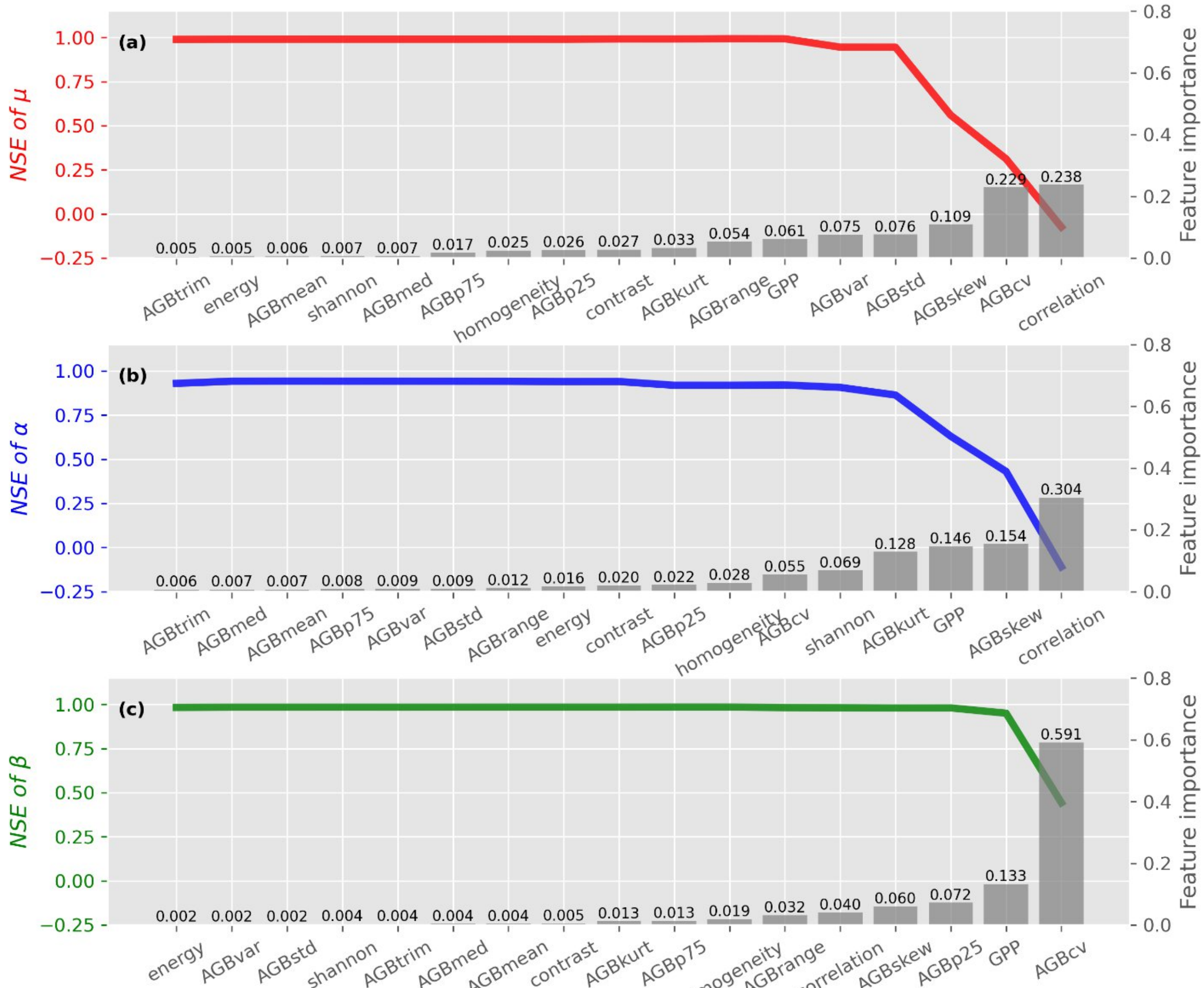

(a)
NSE of μ
Feature importance
0.005 0.005 0.006 0.007 0.007 0.017 0.025 0.026 0.027 0.033 0.054 0.061 0.075 0.076 0.109 0.229 0.238
AGBtrim energy AGBmean shannon AGBmed AGBp75 homogeneity AGBp25 contrast AGBkurt AGBrange GPP AGBvar AGBstd AGBskew AGBcv correlation
(b)
NSE of α
Feature importance
0.006 0.007 0.007 0.008 0.009 0.009 0.012 0.016 0.020 0.022 0.028 0.055 0.069 0.128 0.146 0.154 0.304
AGBtrim AGBmed AGBmean AGBp75 AGBvar AGBstd AGBrange energy contrast AGBp25 homogeneity AGBcv shannon AGBkurt GPP AGBskew correlation
(c)
NSE of β
Feature importance
0.002 0.002 0.002 0.004 0.004 0.004 0.004 0.005 0.013 0.013 0.019 0.032 0.040 0.060 0.072 0.133 0.591
energy AGBvar AGBstd shannon AGBtrim AGBmed AGBmean contrast AGBkurt AGBp75 homogeneity AGBrange correlation AGBskew AGBp25 GPP AGBcv

Figure 8

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

[Journal Name]

Ecological Applications

Appendix S1. Parameterization of disturbance regimes

In the equation 9, the number of disturbance events (n_z) at a specific event size z is following a power law mechanism, the lower α represents that the disturbance events would be more clustered, exhibiting the characteristics of large disturbing events but with rare occurrences in the domain; in contrast, the higher α will distribute the total disturbed areas more evenly and simultaneously increase the occurrences of small-sized events in the domain.

In the equation 10, the event sizes are prescribed as a numerically discrete series from 2^0 cells to half the size of the domain, capped at 2^{19} cells. Stepwise values follow the mechanism of powers of 2, leading to twenty classes of event sizes. Due to the discrete nature of the event sizes and the pseudo-random amount of the corresponding events, a limited uncertain gap remains between the total disturbed area after the generation process and the prescribed value. In an attempt to limit this gap, we performed an error threshold to regulate this randomness: the difference between generated total area and prescribed value as a percentage of total domain area should be lower than 0.001% (10 pixels to a 1000-width domain). When the gap has exceeded the threshold, the new event amount sequence will be recalculated until the condition is met. In very rare cases, it is difficult to compute an amount sequence satisfying the threshold of 0.001%, so in which circumstances, the acceptable gap is relaxed to 0.002% (20 pixels to a 1000-width domain).

In the equation 11, the parameter β controls the slopes of the logarithmic function for describing the relationship between the disturbance's intensity and its size. We descend from Chambers' description of the quantitative relationship between the average mortality rate and disturbance size (Chambers et al. 2013), inheriting a constant intercept parameter b = 0.22684 but varying interval of slope parameter β, from 0.03 to 0.5. For the same size of disturbance events, a larger β indicates a greater intensity, causing more carbon loss during the dynamic carbon cycling simulation. In practice, it is possible for the value of intensity to exceed 1, which usually happens for the big beta and large events. In those cases, all intensity exceeding 1 should be limited back to 1 as per the reality of the situation.

The disturbance generator produced 153 disturbance reference cubes (9 μ and 17 α) in total to generate spatial references for disturbance. Each cube represented a distinct combination of μ and α and comprised 200 snapshots that simulated diverse scenarios of different disturbance event spatial distribution. Notably, these snapshots are all binary (occurs or not) lacking information on intensity, which means that they only provide the spatial reference information under specific μ and α combinations.

The disturbance events, represented by independent patches with flag of True, are meant to be randomly distributed across the whole domain without overlapping or overstepping boundaries (Figure 1c-d), and intensity is then assigned according to the corresponding β values.

Reference

Chambers, Jeffrey Q., Robinson I. Negron-Juarez, Daniel Magnabosco Marra, Alan Di Vittorio, Joerg Tews, Dar Roberts, Gabriel H. P. M. Ribeiro, Susan E. Trumbore, and Niro Higuchi. 2013. "The Steady-State Mosaic of Disturbance and

Succession across an Old-Growth Central Amazon Forest Landscape.” Proceedings of the National Academy of Sciences 110 (10): 3949–54. https://doi.org/10.1073/pnas.1202894110.

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S2. Biomass dynamic simulation with disturbance cubes

With the support of the disturbance reference cube, we applied the strategy of unordered sampling with replacement to generate time series disturbance references. Ultimately, 200 maps of disturbance events were randomly extracted from cubes as a sequence of reference for simulating the occurrence of disturbances over 200 years in the domain. For each disturbance regime, we incorporated the disturbance sequence from the corresponding cube, together with the prescribed varied productivity (G), and background mortality (K_b) levels, to run the dynamic carbon model to an equilibrium state of biomass. Motivated by the consideration of randomness in the occurrence of temporal disturbances, we shuffle the sequence of 200-year disturbance reference maps up to 10 times for each run of the model.

Despite some subtle sawtooth fluctuations that can instantaneously deviate from the expectation due to the impacts of stochastic disturbances, the average biomass for the whole domain saturates and reaches a dynamic steady state over the 200-year simulation run. We averaged the biomass maps for the last decade to obtain steady-state equilibrium biomass distributions, by which three kinds of statistical features were used to characterize steady-state biomass distribution properties. In addition, we also calculated the mean value of Gross Primary Productivity from last year of simulation as an additional constraint feature to predict the varied disturbance regimes.

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S3. Cross validation strategies

For Completely Random 10-fold method (CR), we disrupted all entries in random order and equally divided them into ten parts for 10-fold cross-validation. Nine-tenths of the data is used to fit the model and the rest is for validation, and the ultimate prediction accuracy is the mean of ten cross-validation results.

For Leave-One-Sequence-Out method (LOSO), the fit and validate process was conducted according to the shuffle index, instead of randomly dividing all the data into ten sets. For instance, entries with shuffle index 1 were used for validation and the rest for training and circulated the validation shuffle index until all the data are validated and trained.

Leave-One-Parameter-Out method (LOPO) is performed to against the robustness of disturbance regime: for each of μ, α, and β, we keep each value in turn for validation and train all the remaining data to test the model's predictive capacity for the untrained parameters.

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S4. Relationship between event number and event size controlled by different α

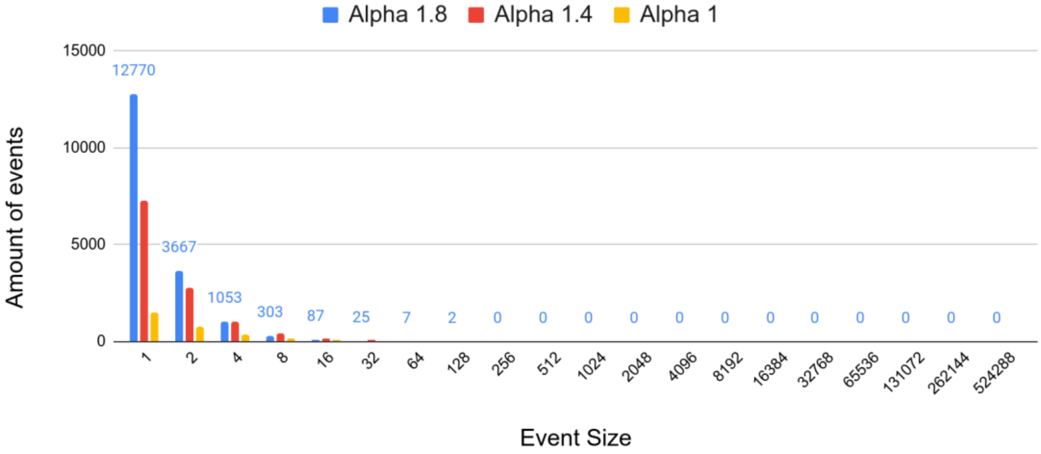


Figure S1. Relationship between the size of disturbance event and corresponding amount under different α in a domain with μ =0.03

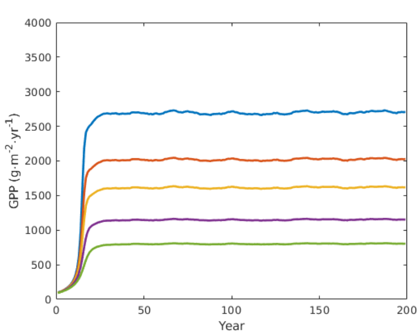
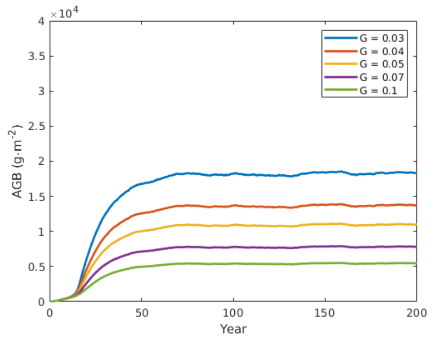
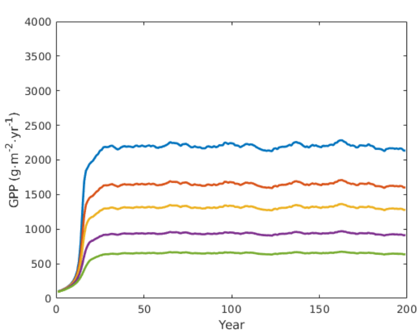
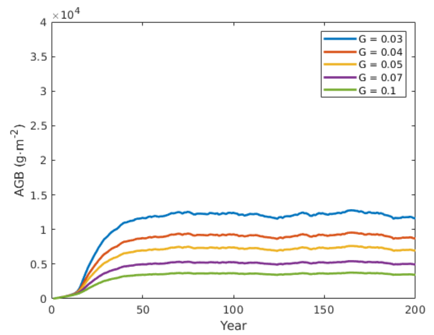
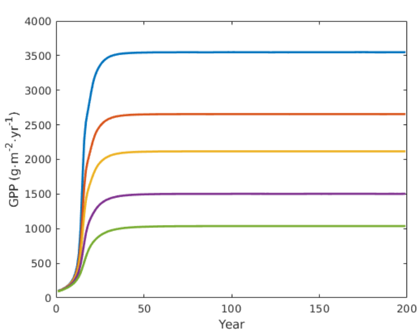
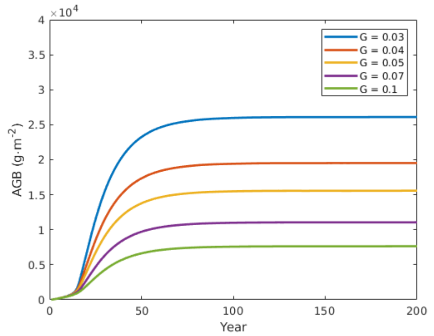
[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

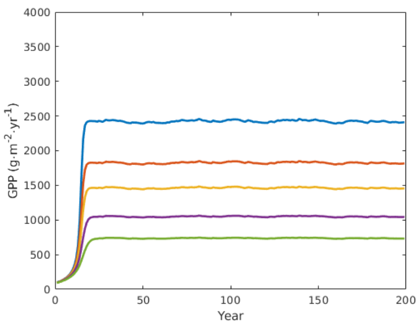
[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S5. AGB and GPP evolution in different disturbance regimes

<table>
<tr><th></th><th>GPP</th><th>AGB</th></tr>
<tr><td>μ 0.03
α 1.0
β 0.2</td><td>
</td><td>
</td></tr>
<tr><td>μ 0.05
α 1.0
β 0.2</td><td>
</td><td>
</td></tr>
<tr><td>μ 0.03
α 1.8
β 0.2</td><td>
</td><td>
</td></tr>
</table>

***μ* 0.03**
***α* 1.0**
***β* 0.5**



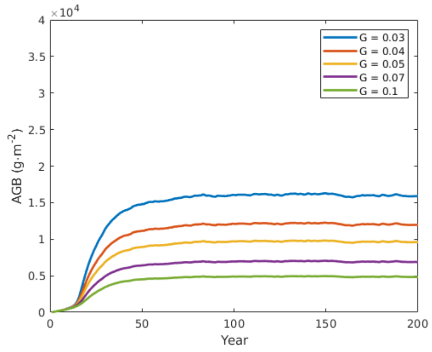


Figure S1. AGB ($gC \cdot m^{-2}$) and GPP ($gC \cdot m^{-2} \cdot yr^{-1}$) evolution trajectories against age (in year) under different disturbance regimes

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S6. Impact of disturbance shape setting

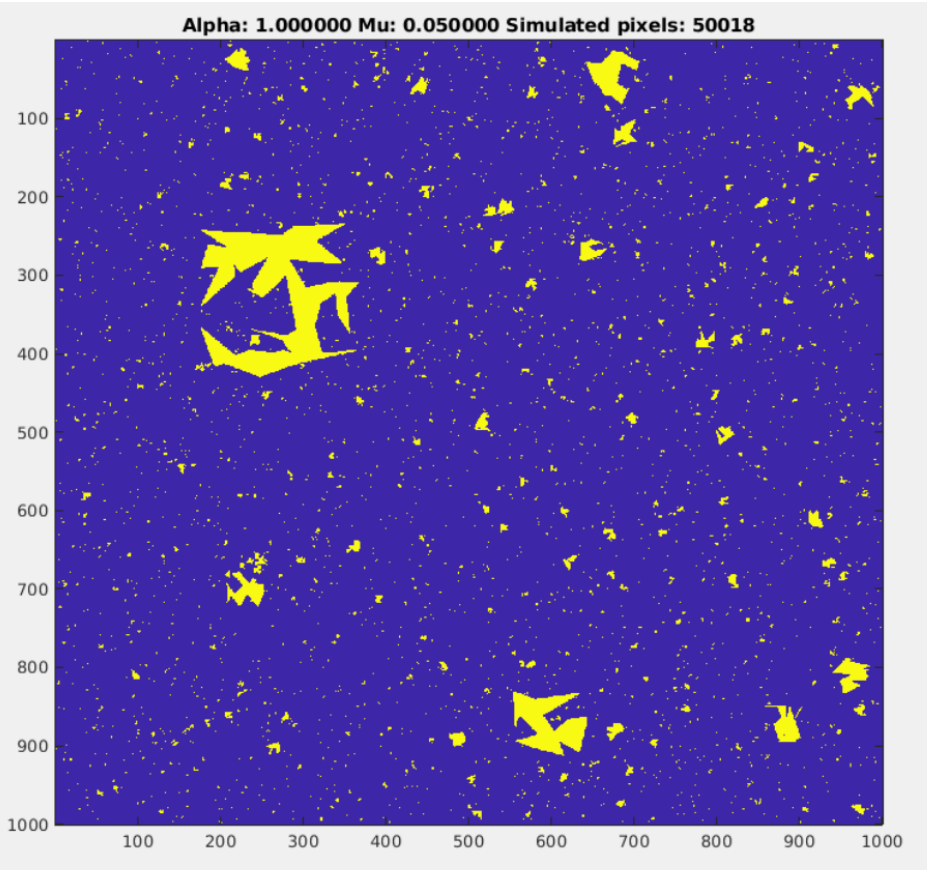


Figure S1. Disturbance reference map in irregular shapes. Compared to the original reference map, the new reference map has more complicated disturbance shape, filled with irregular convex polygons, and the number of sides for these polygons increase with the sizes of event. All properties remain the same as the previous rectangular reference map except the shape.

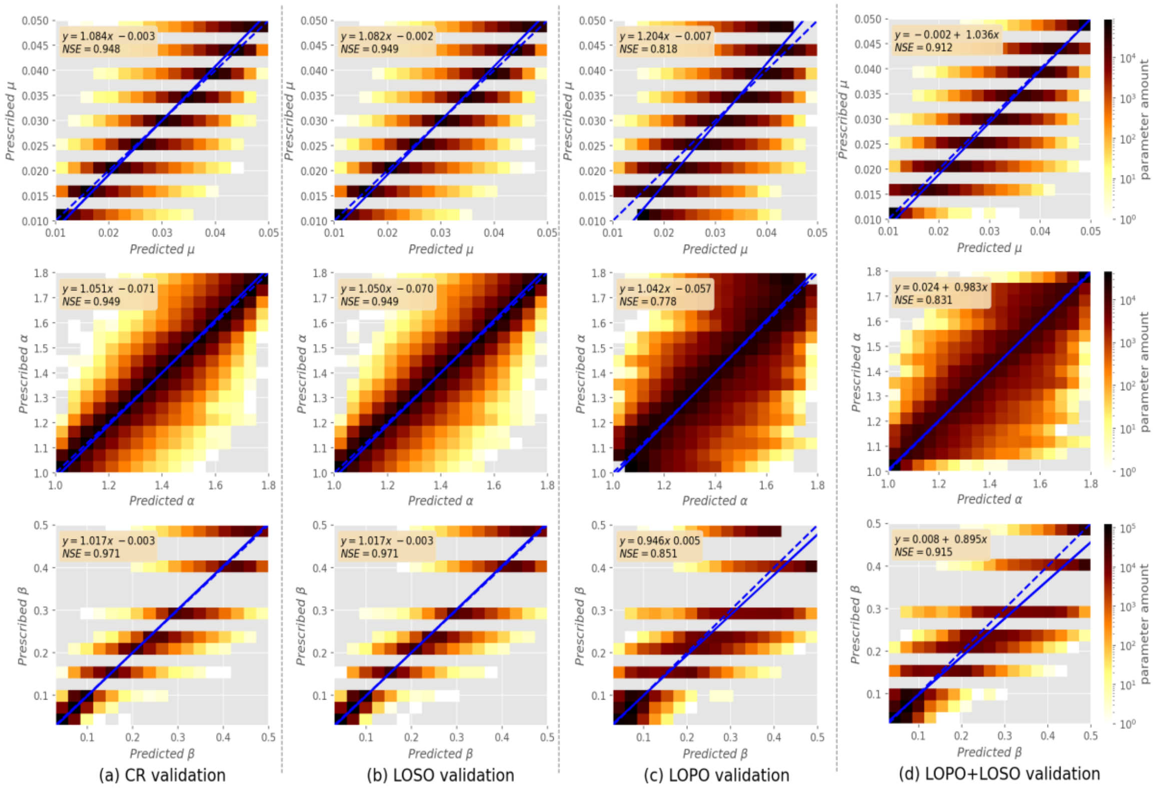


Figure S2. Cross validation result for three disturbance regime parameters by replacing the rectangular disturbance shape with irregular shapes (Figure S1). The X-axis denotes the predicted values, and the Y-axis denotes the prescribed values. (a) illustrates the prediction results of the disturbance regime parameters in the Completely Random cross-validation strategy (CR), (b) refers to the Leave One Sequence Out strategy (LOSO), and (c) refers to the Leave One Parameter Out strategy(LOPO), (d) the LOPO predictions of μ, α, and β at the boundaries were substituted with the LOSO predictions to validate the extrapolation challenge. Specifically, the trained model in LOSO predicted the values of parameter α at 1.0 and 1.8, as well as the values of 0.01 and 0.05 for μ, and 0.03 and 0.5 for β.

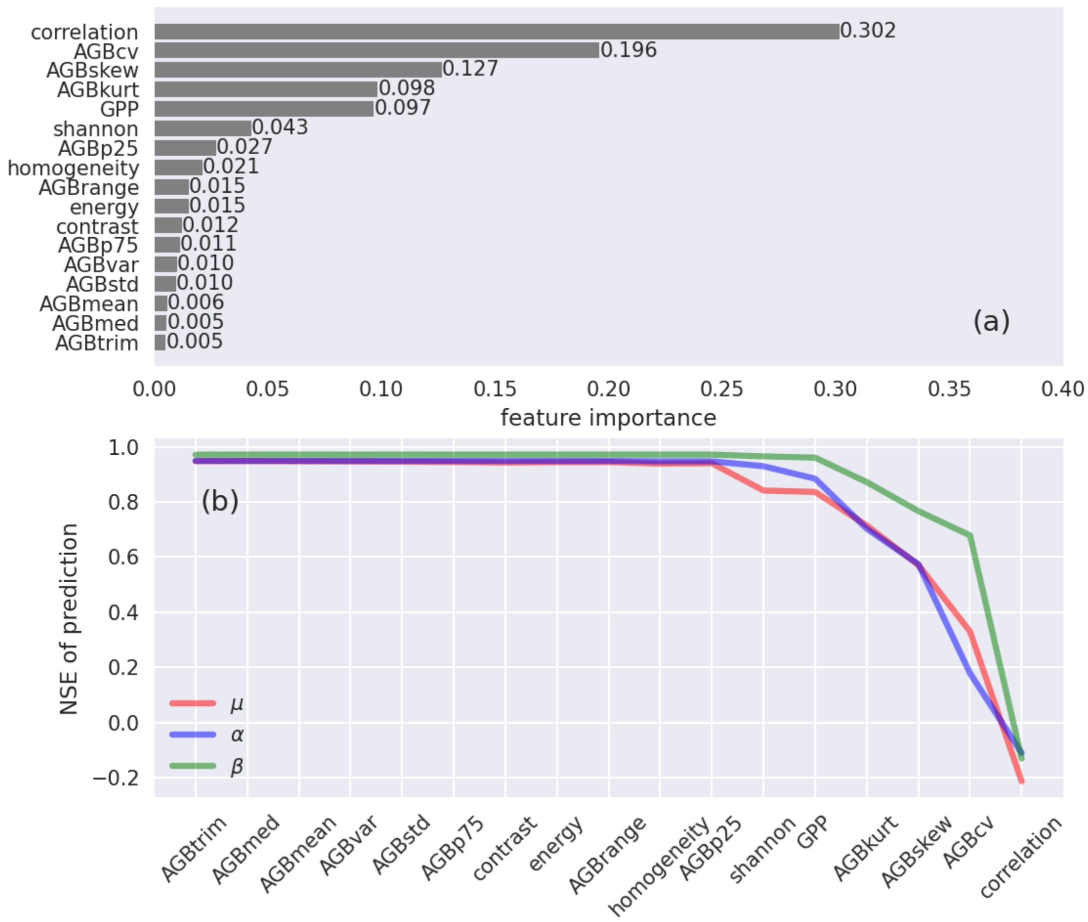


Figure S3..(a) shows the feature importance of multi-output disturbance regime prediction by the irregular disturbance shape setting, where the assigned value denotes the degree of contribution made by each feature (see the definition in Table 2). (b) shows the prediction accuracy change for each disturbance regime parameter by the irregular disturbance shape setting, ordered by ascending feature importance. The X-axis represents the feature(s) used (right) and excluded (left) during the prediction process. The accuracy was measured using the NSE metric, based on the prescribed parameters and the results of multi-output random forest model's prediction.

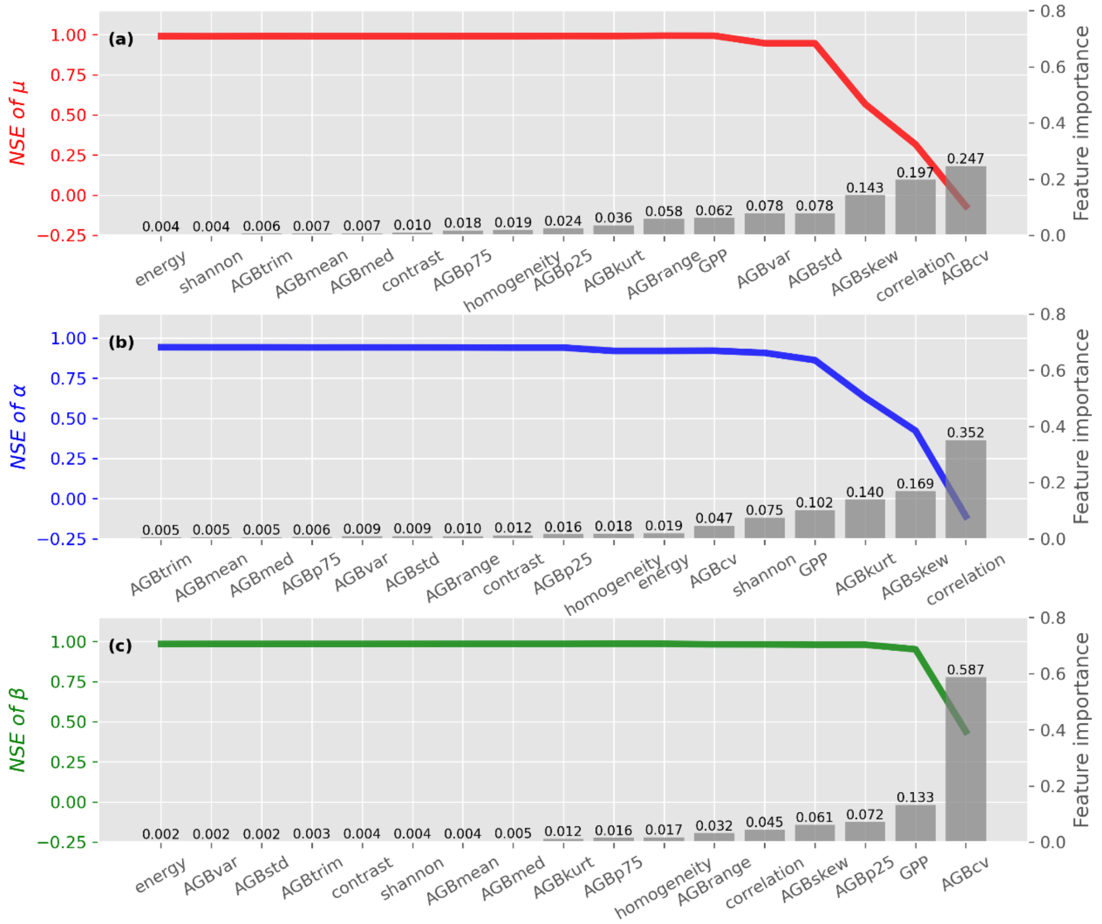


Figure S4. Breakdown of the feature importance for three individual disturbance regime parameters, μ(a), α(b), and β(c) by the irregular disturbance shape setting. The corresponding feature importance is depicted through bars, while the colored lines represent the results of the cumulatively exclusive feature test.

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S7. Outliers in three detection methods

Table S1. the number of outliers in three methods for a domain with 1 million pixels

Detection method	By column	By domain	Overlap Ratio
Median	21,471	11,054	95%
Mean	11,168	9,698	86%
Quartiles	19,898	16,122	94%

Median outliers are defined as elements more than three scaled MAD from the median; mean outliers are defined as elements more than three standard deviations from the mean; quartiles outliers are defined as elements more than 1.5 interquartile ranges above the upper quartile (75 percent) or below the lower quartile (25 percent). Two strategies were employed to determine the median, mean and quartile values. In the first approach, the detection was performed on each column of the domain matrix individually, as indicated by the first column in Table S1. The second approach involved transforming the domain matrix into a vector and then applying statistical calculations and detection, as indicated by the second column in Table S1. The overlap ratio, as shown in the third column in Table S1, represents the number of pixels labeled as outliers by the domain method that are also labeled as outliers by the column approach. The result indicated that there is a significant overlap between these two approaches.

[Authors]

Siyuan Wang, Hui Yang, Sujan Koirala, Matthias Forkel, Markus Reichstein, Nuno Carvalhais

[Manuscript title]

Understanding Disturbance Regimes from Patterns in Forest Biomass

Appendix S8. Typical disturbance regimes driven by different drivers

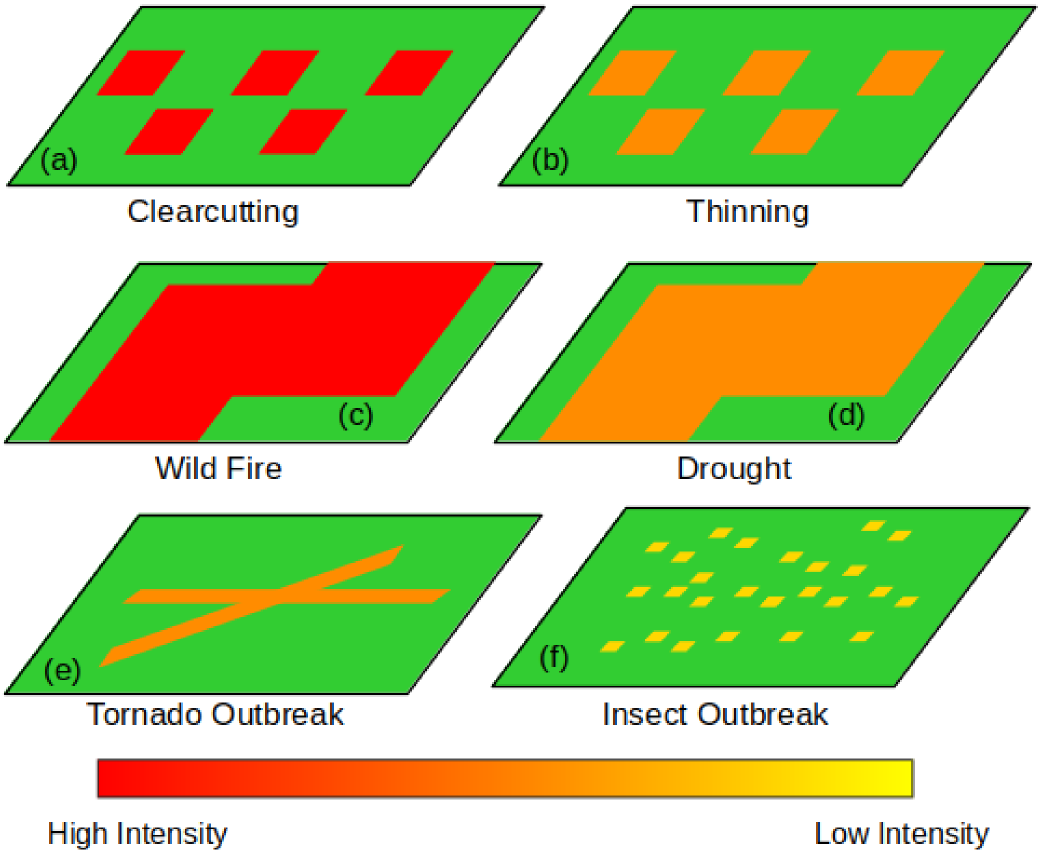


Figure S1. Conceptual diagram of various disturbance regimes driven by natural or anthropogenic drivers. Clearcutting (a) and thinning (b) may have comparable biomass spatial patterns (indicated by a moderate level of μ and α), but differ in the amount of biomass loss, with clearcutting having a higher value of β. Similarly, wild fire (c) and drought (d) can lead to the similar spatial biomass patterns (presumably characterized by a high value of μ and a low α), but differ in intensities, with wild fire having a higher β value. Tornado outbreak (e) would exhibit a unique combination of μ and α due to their distinct shapes of affected areas. And insect outbreaks (f) are more likely to result in numerous small-scale events across the landscape, exhibiting a high α value.