Supplement

S1. The Simulation Framework: Linking Disturbance and Biomass

Building on our previous work (Wang et al., 2024), we use a forward-modeling framework to quantify the link between disturbance regime and the resulting biomass statistics. The core idea is to first simulate how a wide array of known, parameterized disturbance regimes manifest as synthetic, spatially explicit patterns of aboveground biomass. By systematically generating these cause-and effect pairings, we can establish a statistical link that later allows us to retrieve the problem: inferring the characteristics of a known disturbance regime from observable biomass patterns.

The workflow proceeds in four main stages: 1) we parameterize a disturbance regime using a set of key attributes that control the extent, frequency, intensity and background state of forest mortality; 2) we stochastically generate multi-year sequences of disturbance events on a simulated landscape, where the size, location, and shape of each event are governed by the defined regime parameters; 3) we apply these disturbance sequences to a simple carbon cycle model to simulate the dynamic change in AGB across the landscape over centuries. 4) once the simulated landscape reaches a dynamically stable state, that is, a shifting mosaic of different successional stages, a comprehensive suite of statistical metrics from the AGB map is extracted, which quantitatively describes the emergent biomass pattern. This forward-modeling approach creates a large synthetic dataset where known disturbance regime parameters are linked to unique biomass pattern signatures, forming the basis for training a predictive model.

S1.1 Parameterization of disturbance regime

We characterize the disturbance regime acting upon a landscape using four key parameters, three of them—probability scale (μ), clustering degree (α), and intensity slope (β)—are adapted from our previous framework (Wang et al., 2024), but with a significant modification to the parameter range for α. Specifically, we now include lower values for the clustering degree, which allows for the simulation of fewer, larger, and more contiguous events. This adjustment is crucial for representing large-scale, intense disturbances such as stand-replacing fires. Another major advancement in our parameterization is the explicit inclusion of the background mortality rate (K_b) as a fourth independent parameter to be predicted, allowing us to disentangle the effects of baseline mortality from episodic disturbances and address a potential source of equifinality in the previous framework. The four parameters are detailed in Table 1.

Table S1. Parameters Defining the Simulated Disturbance Regimes

Parameter	Symbol	Definition	Proposed Range	Intervals	Count	Role in Simulation
Probability Scale	μ	The mean fraction of the total domain area affected by disturbance events annually.	0.01 - 0.05	0.005	9	Controls the overall frequency and extent of disturbances.
Clustering Degree	α	The scaling exponent in the power-law function	0.5 - 1.8	0.05	27	Governs the spatial pattern of disturbance,

		($n \propto z^{-\alpha}$) relating the number of events (n) to their size (z).				where increasing values shift the pattern from large, contiguous events to many small, fragmented disturbances.
Intensity Slope	β	The slope of the relationship defining the fraction of AGB lost as a function of disturbance event size.	0.03 - 0.5	0.01/0.05 /0.1	14	Determines the severity of disturbance events.
Background Mortality	K_b	The constant, first-order mortality rate representing baseline carbon loss process like litterfall, root exudates, and herbivory.	0.025 - 0.2	0.025	8	Controls the background carbon turnover time ($\tau = \frac{1}{\kappa_b}$) and influences the maximum potential biomass.

S1.2 Simulating Spatially-Explicit Disturbance Events

To translate our parameterized disturbance regime into concrete spatial patterns, we developed an advanced disturbance event generator. This tool is a key advancement from our previous framework, which was limited to computationally efficient but unrealistic rectangular shapes. The new generator is capable of producing disturbance events with a wide range of morphological complexities, allowing us to enhance the realism of our simulations and test the influence of event geometry on emergent biomass patterns.

For each combination of disturbance regime parameters, the generator creates a set of 200 annual disturbance maps on a 1000 × 1000 pixel grid, with event shapes corresponding to one of the six settings detailed in Table 2. These 200 maps are temporally independent, representing a stochastic sequence of events over time. This design choice—the absence of temporal autocorrelation—is crucial as it allows the set of maps to be shuffled, enabling multiple, unique simulation runs for the same underlying disturbance regime.

Table S2. Disturbance Event Shape Settings

Category	Shape Setting	Description
Simple Regular	Rectangle	The baseline setting from (Wang et at., 2024); all events are four-sided rectangles.
	Triangle	All events are simple three-sides polygons.
	Circle	All events are uniform, non-directional circular pathes.
	Gradient	Event complexity (number of sides) is proportional to event size; larger events have more sides.
Complex Convex	Complex	All events are generated with maximum complexity (49 sides), regardless of their size.
	Random	The complexity of each event is a random integrer of sides between 3 and 49.

The complete spatiotemporal output of the generator for a single disturbance regime is stored in a disturbance reference cube. This three-dimensional array (1000 × 1000 pixels × 200 years) serves as the precise blueprint of all disturbance locations, sizes, and shapes over the entire simulation period. A unique disturbance reference cube is generated for every combination of the six shape settings and the disturbance regime parameters (μ, α and β). This systematic approach allows us to isolate the impact of event morphology on emergent biomass patterns. This methodology tests the robustness of our

framework, with the hypothesis that landscape-level biomass statistics are more sensitive to the fundamental regime parameters than to the specific geometry of individual disturbance patches.

S1.3 Generating Synthetic Biomass Statistics with a Carbon Cycle Model

The carbon modeling workflow is conceptually identical to our previous framework. Each disturbance reference cube is used to drive a simple, dynamic carbon cycle model at the pixel level. The annual change in aboveground biomass is governed by the balance between carbon gains and losses:

$$\frac{dAGB}{dt} = NPP_{AGB} - L_{total}$$

where the total loss, L_{total}, is partitioned into episodic disturbance loss (L_d) and continuous background mortality ($L_d = AGB \times K_b$).

The simulation experiment was designed to be comprehensive. Our full factorial design included every combination of the four disturbance regime parameters (9 μ values × 27 α values × 14 β values × 8 Kb values), five different levels of primary productivity (Photosynthetic capacity), and all six disturbance shape settings. To account for stochasticity, each of these scenarios was replicated 10 times using a different random shuffle of the annual disturbance maps, resulting in a total of 8,164,800 individual simulation runs.

Each simulation runs for 200 years to ensure the landscape reaches a dynamically stable equilibrium. To characterize this stable state, we extracted the Gross Primary Production (GPP) from the final simulation year and the average aboveground biomass (AGB) map over the last 10 years. It is from this 10-year average AGB map that we derived a comprehensive vector of spatial statistics to serve as the quantitative signature of the biomass pattern. This suite of metrics—encompassing first-order distribution statistics and second-order texture metrics from a Gray-Level Co-occurrence Matrix (GLCM)—is methodologically identical to the one derived from the observed biomass data described in the following section, ensuring a consistent basis for comparison.

The final output is a massive dataset linking each unique set of input parameters to a corresponding vector of output features (Table S3, 16 biomass pattern statistics plus the average GPP). This dataset provides the foundation for training a machine learning algorithm to infer disturbance regimes from biomass patterns.

Table S3. Features of Biomass and GPP used to train the Random Forest model

Type	Feature	Meaning
Histogram Features	Mean	Mean biomass value of domain
	Median	Median biomass value of domain
	Variance	Statistical Variance of biomass values in the domain
	Standard Deviation	Statistical deviation of biomass values in the domain
	Coefficient of Variation	Ratio of the standard deviation to the mean of biomass values

	Skewness	Measure of the asymmetry of the biomass value distribution
	Kurtosis	Measure of the weight of the tails or the sharpness of the central peak of the biomass value distribution
	Percentile 25%	The 25th percentile biomass value of the domain (P25)
	Percentile 75%	The 75th percentile biomass value of the domain (P75)
	Range	Distance between 90th percentile biomass and 20th percentile biomass (P90 - P20)
	Trimean	The Tukey's trimean of the biomass values, calculated as (P25 + 2*Median + P75)/4
Informative Feature	Shannon Entropy	A measure of the diversity or uncertainty in the distribution of biomass values
Texture Features	Contrast	Measures how "sharp" or "different" neighboring biomass values are. (High contrast = big jumps between neighbors)
	Correlation	Measures how related neighboring biomass values are. (High correlation = neighbors are usually very similar)
	Energy	Measures how "uniform" or "orderly" the biomass pattern is. (High energy = a very simple, repetitive pattern)
	Homogeneity	Measures the "smoothness" of the biomass pattern. (High homogeneity = most neighbors have very similar values)
Photosynthesis Feature	GPP	Mean Gross Primary Production at the end of the simulation

S2. Determining the Optimal Aggregation Scale

S2.1 Mismatch between Model Simulation and EO

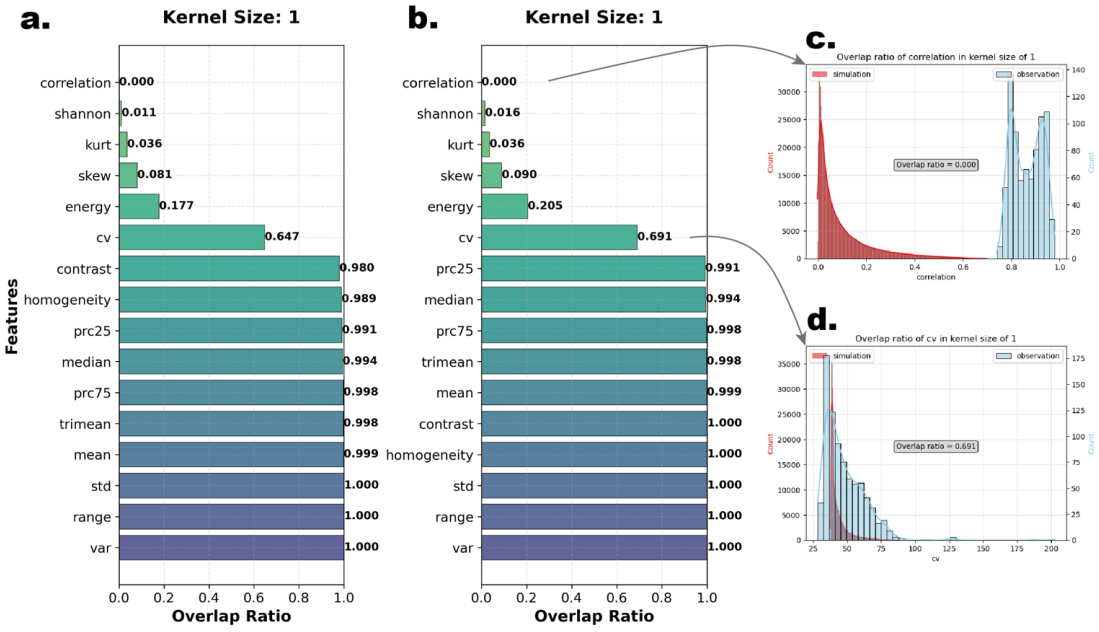

Figure S2.1 Statistical overlap between simulated and observed biomass feature distributions. The panels compare the percentage of overlap for 17 statistical features derived from (a) the original forward modeling framework and (b) the improved framework, which incorporates an expanded parameter space and non-rectangular disturbance shapes (see Supplementary S1 for details). The overlap percentage is calculated based on the intersecting range between the frequency distributions of simulated and observed values. Panels (c) and (d) show the specific distributions for the two most important predictive features identified in Wang et al. (2024): GLCM Correlation and Coefficient of Variation, respectively.

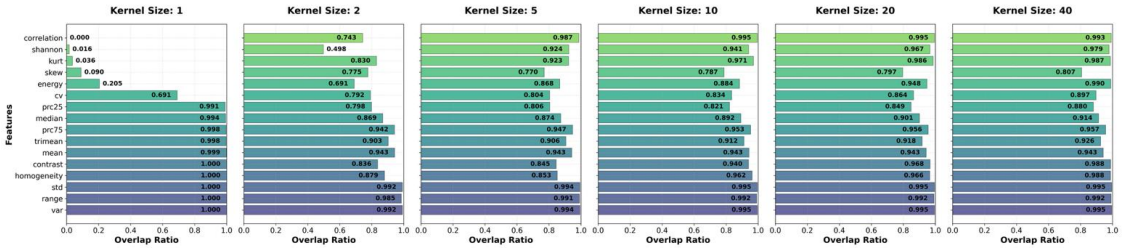

Figure S2.2 The effect of spatial aggregation on the statistical overlap between simulated and observed biomass features. The figure illustrates how statistical overlap changes as a function of aggregation scale, which is represented by the kernel size used for averaging (larger kernels correspond to coarser spatial resolutions). The results demonstrate a clear positive trend: as the level of aggregation increases, the statistical overlap between the simulated and observed datasets consistently improves. This confirms that spatial aggregation is an effective strategy for reducing the discrepancy between the two domains, an effect that is particularly pronounced for key predictive features such as GLCM Correlation and the Coefficient of Variation (CV).

S2.2 Weighted Overlap Ratio

To objectively identify the optimal aggregation scale, we developed the Feature Importance Weighted Overlap Ratio (WOR), a robust metric that quantifies the similarity between the multi-dimensional feature spaces of the simulated and observed datasets. The WOR calculation prioritizes the most influential features (Correlation, Kurtosis, Skewness, and CV) by assessing their pairwise overlap and weighting the result by their combined, scale-dependent feature importance ($FI'_{i,j}$). The pairwise overlap ratio ($OR_{i,j}$) for any two feature distributions, $\mathrm{p}(\mathrm{x},\mathrm{y})$ and $\mathrm{q}(\mathrm{x},\mathrm{y})$, is calculated as the intersecting volume of their probability density function:

$$OR_{i,j} = \iint min(p(x,y), q(x,y))dxdy$$

The final WOR score is a single value between 0 and 1 that measures statistical alignment, calculated as:

$$\mathrm{WOR} = \sum_{\mathrm{i,j}} \mathrm{FI}'_{\mathrm{i,j}} \times \mathrm{OR}_{\mathrm{i,j}}$$

A higher WOR values signifies greater similarity and thus higher confidence in the model's predictive capability at that scale.

S2.3 Optimal Aggregation Scale

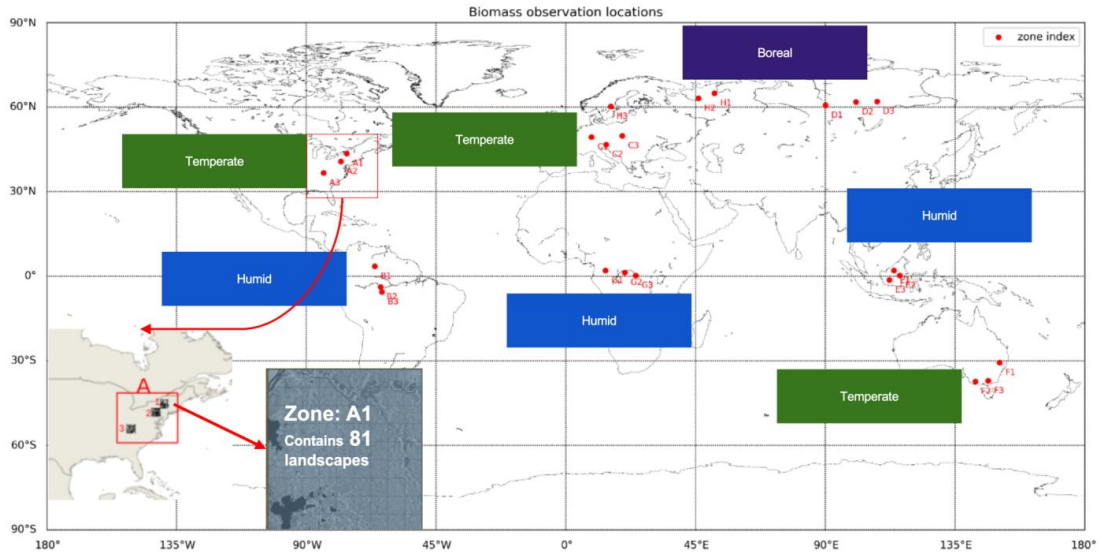

Figure S2.3 Global distribution of biomass observation locations for optimal aggregation analysis. The map displays the locations of the 24 globally distributed zones randomly selected for the multi-level overlap analysis, categorized by biome. Each zone (e.g., A1, inset) is composed of 81 individual landscapes, where each landscape consists of a 1000×1000 pixel grid corresponding to approximately 25×25 km at the equator. This design forms the hierarchical structure used for evaluating the simulation-observation gap at the zone, region, biome, and global levels.

The optimal aggregation scale for generating global product was determined through a comprehensive, multi-level sensitivity analysis using the Weighted Overlap Ratio (WOR). This was conducted across a nested hierarchy of observational levels, based on a global distribution of landscape zones sampled from Boreal, Temperate, and Humid Biomes **(Fig S2.3)**. The WOR was calculated at each level of the hierarchy by progressively expanding the observational pool: from individual zones, to regions (e.g., Region A= pool of A1, A2, A3), to biomes, and finally to the entire global dataset.

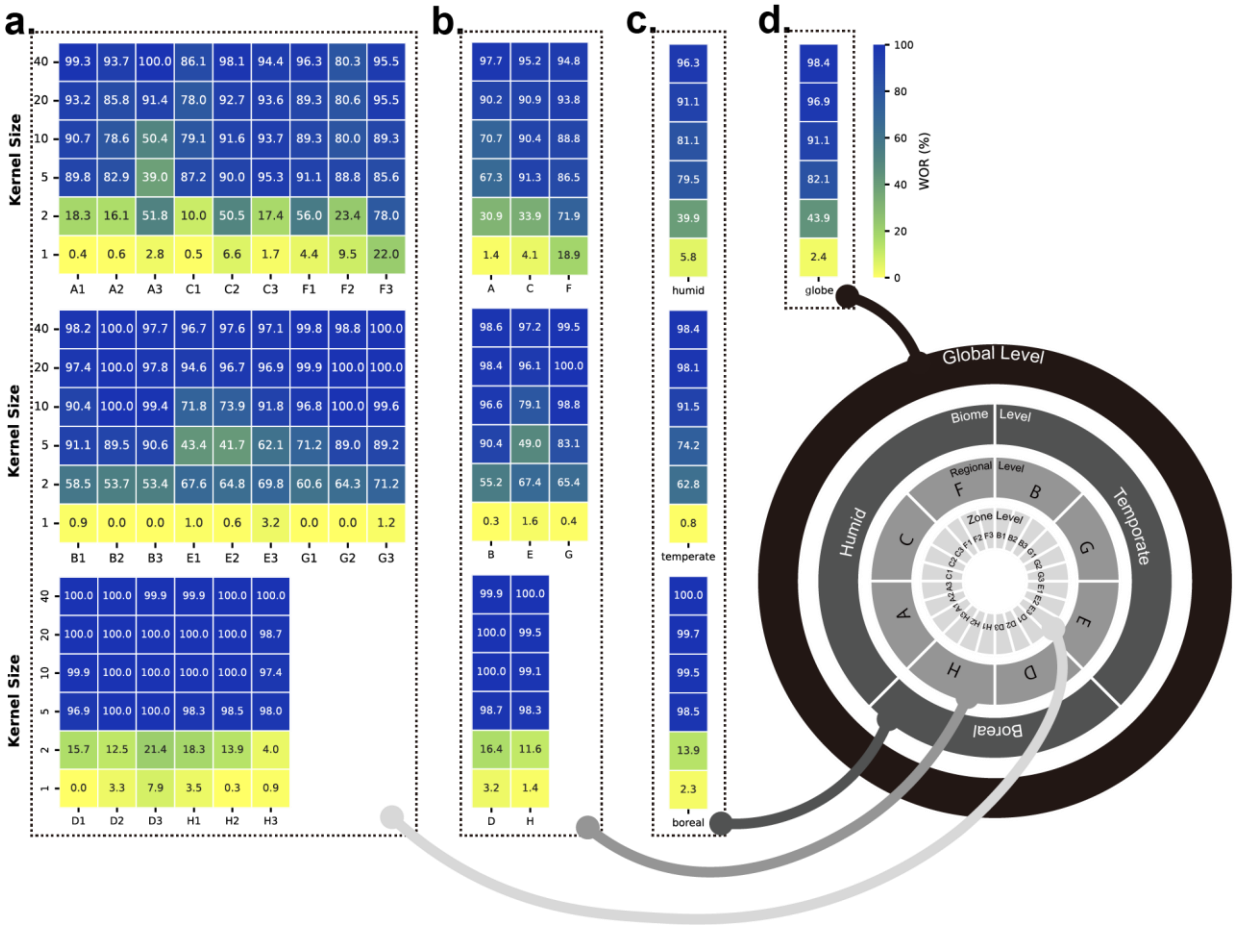

Figure S2.4 Multi-scale Weighted Overlap Ratio (WOR) analysis across hierarchical spatial domains. The figure displays WOR heatmaps for four distinct observational levels: (a) individual landscape zones (A1-H3), (b) regional groups (e.g., A, B, C), (c) biome groups (Humid, Temperate, Boreal), and (d) the global scale. For each heatmap, the rows represent different aggregation kernel sizes, and the columns represent the different spatial domains (e.g., individual zones, regions). Color intensity represents the WOR percentage (0-100%), indicating the statistical similarity between observed and simulated datasets when both are processed with the same aggregation scale. The conceptual diagram (bottom right) illustrates the hierarchical spatial organization from landscape zones through regional and biome aggregations to global scale.

The result of the hierarchical WOR analysis **(Fig S2.4)** reveals several key findings. First, a consistent trend was observed across all landscape zones: as the degree of aggregation increases, the WOR value improves, highlighting the general effectiveness of this strategy in narrowing the simulation-observation gap. For most zones, an aggregation with a kernel size of 10 is sufficient to achieve a WOR greater than 90%, indicating good statistical consistency. Second, this trend holds at higher spatial levels (region, biome, and global), confirming the value of aggregation. Specifically, boreal ecosystems consistently exhibit a relatively higher WOR compared to temperate and humid biomes at equivalent aggregation scales. Third, by setting a target threshold of 90% for the WOR and considering that prediction accuracy is highest at smaller kernel sizes, a kernel size of 10 emerges as the optimal

choice for global-scale prediction. While this represents the global optimum, the framework allows for this scale to be adjusted for specific regional or biome-level analyses.

S3. Uncertainty Assessment based on Meyer and Pebesma's framework

S3.1 Theoretical Formulation

The DI metric quantifies the relative dissimilarity of a prediction point to the training data space:

$$DI_k = \frac{D_I}{D_{AVG}}$$

Where DI_k is the minimum weighted Euclidean distance from the prediction point to any training sample, D_{AVG} is the average weighted distance between training samples. The weighting is based on feature importance derived from aggregated Random Forest models, ensuring that more informative features contribute more to the dissimilarity calculation.

For a prediction point x_p and training data $\{x_i\}_{i=1}^{n}$, the weighted distance is computed as:

$$D_I = \min_i \sqrt{\sum_{j=1}^{d} w_j \left(x_{p,j} - x_{i,j}\right)^2}$$

where w_j represents the importance weight for feature j,

The D_{AVG} (Distance Average) computes the average pairwise distances between training samples to establish a baseline for the DI metric:

$$D_{AVG} = \frac{1}{n(n-1)/2} \sum_{i<k} \sqrt{\sum_{j=1}^{d} w_j \left(x_{i,j} - x_{k,j}\right)^2}$$

S3.2 Scalable Implementation Workflow

Directly applying the DI formulation described in Section 2.5.1 to large-scale datasets is computationally prohibitive due to the quadratic complexity of calculating the pairwise distance matrix for the full training dataset (T). To address this challenge, we develop a scalable workflow that decouples the calculation into an efficient pre-calculation for each prediction point. This process involves three key steps: data standardization, baseline dissimilarity pre-calculation, and the final DI computation.

To ensure that all features contribute proportionally to the distance metric and to mitigate the influence of extreme outliers, each feature f is independently standardized, we use a robust percentile-based normalization where each value $x_{i,j}$ of a sample i for a feature f is scaled to the range defined by the 1st and 99th percentiles of that feature in the training data:

$$x'_{i,f} = \frac{x_{i,f} - P_1(f)}{P_{99}(f) - P_1(f)}$$

Where $x'_{i,f}$ is the normalized feature value, and $P_{99}(f)$ and $P_1(f)$ are the 99th and 1st percentiles of feature f across the training set T, respectively. All subsequent calculations are performed on these normalized features.

The average pairwise distance between all samples in the training data, D_{avg}, serves as a stable baseline for the DI metric. To compute this efficiently, we pre-calculate it on a representative and computationally tractable subset of the training data. First, a small subset $T_s \subset T$ is sampled (typically 0.1% of the full dataset) to maintain statistical representativeness while ensuring computational feasibility. For each feature f, the unweighted average distance, $D_{avg}(f)$, is computed across all pairs of points within this subset:

$$D_{avg}(f) = \frac{1}{\binom{n_s}{2}} \sum_{i=1}^{n_s-1} \sum_{j=i+1}^{n_s} \left| x'_{i,f} - x'_{j,f} \right|$$

Where n_s is the number of samples in the subset T_s. These single-feature average distances are pre-calculated and stored for each cross-validation fold, forming the building blocks for the final weighted baseline dissimilarity, D_{avg}. With the component values pre-calculated, the final DI for a new prediction point (observed statistical features from a realistic landscape), x_{pred}, is computed efficiently.

First, the weighted average distance $\bar{d}$ is assembled by combining the pre-calculated single-feature distances with their corresponding feature importance weights, w_f, derived from the trained Random Forest model:

$$\bar{d} = \sum_{j=1}^{d} w_f \cdot D_{avg}(f)$$

Where d is the total number of features, 17 in this study.

Next, the minimum weighted Euclidean distance, d_k, from the new prediction point x_{pred} to any sample in the full training set T is calculated:

$$d_k(x_{pred}, \mathrm{T}) = \min_{i \in T} \left(\sqrt{\sum_{f=1}^{d} w_f (x'_{pred,f} - x'_{i,f})^2} \right)$$

Finally, the Dissimilarity Index for the prediction point is computed as the ratio of these two values:

$$DI(x_{pred}) = \frac{d_k(x_{pred})}{\bar{d}}$$

This scalable workflow enables the practical application of the DI framework to extensive datasets by strategically avoiding the most computationally expensive operation while preserving the theoretical integrity of the uncertainty metric.

S4. Predictions

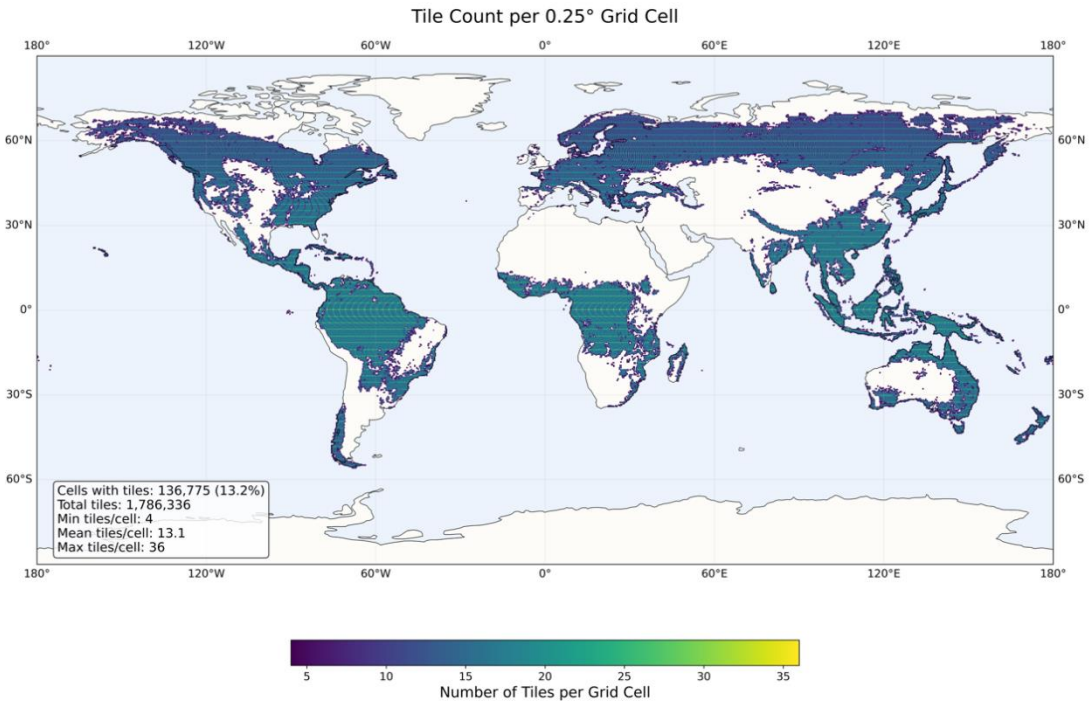

Figure S4.1 Global map of the number of 25 km x 25 km landscape tiles aggregated into each 0.25° grid cell. This tile_count layer serves as an indicator of sampling density, showing the number of underlying tile-level predictions used to calculate each grid cell value. Higher values indicate that the gridded cell value is derived from a more robust spatial sample.