GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: www.tandfonline.com/journals/tgrs20

Degree of geocomplexity for diagnosing spatial pattern loss beyond prediction accuracy

Haiyang Liu , Yongze Song, Wen Yi & Pengcheng Zhang

To cite this article: Haiyang Liu , Yongze Song, Wen Yi & Pengcheng Zhang (2026) Degree of geocomplexity for diagnosing spatial pattern loss beyond prediction accuracy, GIScience & Remote Sensing, 63:1, 2657087, DOI: 10.1080/15481603.2026.2657087

To link to this article: https://doi.org/10.1080/15481603.2026.2657087

Published online: 08 Apr 2026.

GISCIENCE & REMOTE SENSING
2026, VOL. 63, NO. 1, 2657087
https://doi.org/10.1080/15481603.2026.2657087

Taylor & Francis
by informa

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Degree of geocomplexity for diagnosing spatial pattern loss beyond prediction accuracy

Haiyang Liu[a,b,c], Yongze Song[c], Wen Yi[a] and Pengcheng Zhang[a]

[a]Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China; [b]The Hong Kong Polytechnic University, Shenzhen Research Institute, Shenzhen, People's Republic of China; [c]School of Design and the Built Environment, Curtin University, Perth, Australia

ABSTRACT

Reliable validation is critical for spatial prediction models, yet conventional approaches often neglect a model's ability to interpret geospatial characteristics. Traditional accuracy metrics such as coefficient of determination (R^2) and root mean squared error (RMSE) provide global error estimates but fail to detect spatially structured errors, potentially supports models that mislead underlying geographic patterns. This study develops a degree of geocomplexity (DG) metric to examine the ability of spatial model to capture the spatial characteristics of local patterns by comparing the local complexity pattern of the original data with those of the model residuals. We applied this DG metric to tree biomass prediction in eastern Australia, evaluating ten competing spatial models. Findings from this study demonstrate that high prediction accuracy does not guarantee strong geocomplexity interpretability, as models with high R^2 values do not necessarily achieve high DG scores. Furthermore, we demonstrate that no single model performs optimally everywhere, with the DG-based optimal model map revealing a complex pattern of spatial interpretability. The developed DG metric provides an effective validation strategy for assessing and selecting spatial models based on localized performance within a geographical context.

ARTICLE HISTORY
Received 30 October 2025
Accepted 3 April 2026

KEYWORDS
Geospatial characteristics; model validation; geocomplexity; tree biomass

1 Introduction

Spatial models are essential for understanding complex environmental phenomena. However, to ensure the reliability of their predictive results, rigorous validation is indispensable, and this remains a key challenge (Anselin 1995; Rykiel 1996). The most common current validation methods are goodness-of-fit and error metrics, such as the R^2 and RMSE (Fisher 1970; Hyndman and Koehler 2006). These metrics provide an overall assessment of a model's predictive accuracy by quantifying the overall discrepancy between observed and predicted values. Other validation dimensions include uncertainty and sensitivity analyses, which explore the robustness of model outputs to changes in input data and parameters, providing deeper insights into model stability and confidence (Bennett et al. 2013; Brenning 2005; Meyer and Pebesma 2022).

Effective spatial models must be able to capture a series of intrinsic geographic features. These features cover spatial autocorrelation (the tendency for nearby locations to have similar values, which can be measured globally by Moran's I or identified locally by LISA), spatial heterogeneity (where relationships and processes vary with spatial location), and scale dependency (where observed patterns change with the scale of analysis) (Anselin 1995; Fotheringham, Yang, and Kang 2017; Openshaw 1984). In recent years, the concept of geocomplexity has further deepened this understanding, as it describes complex local structures that go beyond simple pairwise similarity by examining the relationships between a point and its neighbours, and even among those neighbours (Goodchild 2009; Zhang et al. 2023). Therefore, the effectiveness of a spatial model is ultimately reflected in its ability to faithfully reproduce this series of complex geographic patterns.

CONTACT Yongze Song Yongze.song@curtin.edu.au

Although tools for describing spatial patterns exist, a significant gap remains in integrating them into a unified validation metric. The main deficiency lies in the reliance on non-spatial metrics such as R^2 and RMSE. These metrics are effective for assessing global predictive performance, but they are fundamentally blind to a model's ability to interpret spatial characteristics (Goodchild 2009). Here, we define spatial interpretability as a model's capability, through its computational process, to effectively represent and explain intrinsic spatial patterns in the data. Consequently, even a model with high statistical accuracy (e.g. random forest) may still produce spatially distorted outputs, a critical flaw that traditional validation metrics fail to detect (Ripley 2001). A good model should be able to fully explain the inherent spatial patterns of the data, thus producing residuals with no spatial structure (Heuvelink, Burrough, and Stein 1989). However, the application of this principle is often limited in practice to testing global spatial autocorrelation, ignoring more granular local variations. We lack a standardised metric capable of quantifying a model's ability to capture local geocomplexity. In ecology, this absence can directly lead to inaccurate carbon assessments and ineffective climate strategies (Keith, Mackey, and Lindenmayer 2009; Mitchard 2018).

This study proposes and validates a degree of geocomplexity (DG) metric to examine the ability of spatial models in capturing the spatial geocomplexity of local patterns by comparing the local complexity pattern of the original data with those of the model residuals, bridging the gap between model's statistical performance and spatial interpretability. We apply this metric to the prediction of tree biomass in eastern Australia to introduce DG as a spatial diagnostic metric—a critical task for formulating effective climate policies and sustainable forest management.

2 Degree of geocomplexity for model validation (DG)

2.1 Concept of DG

DG is a novel metric for model validation. It is derived from an principle of spatial validation: a good model should fully explain the inherent spatial structure of the data, thus leaving spatially random residuals (Liu, Song, and Yi 2026). DG applies this principle to evaluate a model's ability to interpret local geocomplexity. Geocomplexity is designed to quantify the complexity of local spatial patterns. Its core premise is that true spatial complexity arises not just from how a location relates to its neighbours, but also from how those neighbours relate to each other (Luo et al. 2022; Zhang et al. 2023). DG works by measuring how geocomplexity decreases from the original data to the model residuals, which reflects how much of this feature the model can explain. DG does not assume that geocomplexity can be split in an additive way. Instead, it computes geocomplexity for the data field and the residual field under the same setting, and treats their relative decrease as the extent to which local spatial structure is captured.

The DG metric is a tool specifically designed to evaluate a model's ability to interpret local geocomplexity. By applying the principle of residual-based assessment to local geocomplexity, DG quantifies how much of the complex intrinsic spatial structure of the original data is successfully captured by a model. Unlike metrics based on global characteristics such as Moran's I, DG is inherently local and produces spatially accurate maps that can tell mode's performance in complex regions (Song 2023).

2.2 Calculation process of DG

The developed degree of geocomplexity (DG) metric for spatial model validation comprises the following steps, as shown in Figure 1. First, spatial data, including the dependent variable and relevant explanatory variables, are prepared for modelling. Next, a suite of spatial models is employed to generate predictions and corresponding residuals; then, DG is calculated by comparing the geocomplexity inherent in the original data to that remaining in its residuals. A suite of DG results is calculated, enabling an optimal-model map that shows which model performs best in different locations. Finally, we compare DG with traditional non-spatial metrics (e.g. R^2, RMSE) to provide a dual-criterion assessment that balances predictive accuracy with spatial interpretability.

2.2.1 Calculating geocomplexity of data and modelling residuals

Geocomplexity is calculated as shown in Equations (1) and (2) (Zhang et al. 2023):

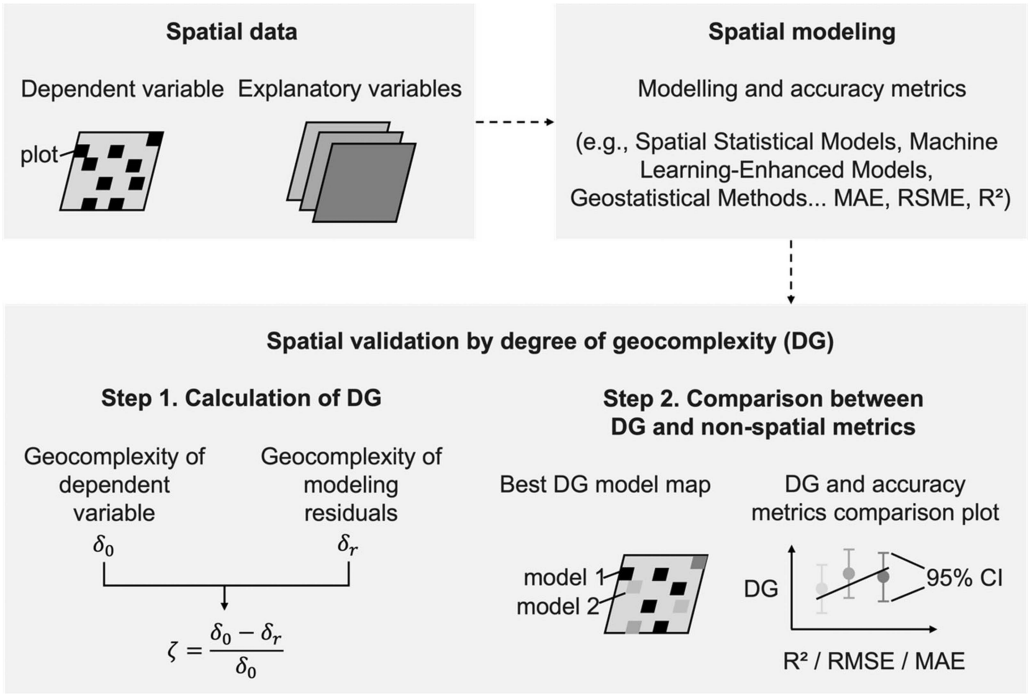

Figure 1. Flowchart of the process of calculating the degree of geocomplexity (DG) metric for spatial model validation.

$$P_i = -\frac{1}{m} Z_i \sum_{j=1}^{m} W_{ij} \cdot Z_j - \frac{1}{m} \sum_{j=1}^{m} W_{ij} \cdot Z_j \frac{1}{v_k} \sum_{k=1}^{n} W_{jk} \cdot W_{ik} \cdot Z_k \tag{1}$$

$$G_i = \frac{P_i - \min(P_i)}{\max(P_i) - \min(P_i)} \tag{2}$$

where P_i is the raw spatial local complexity value of region i; m is the number of direct neighbours of region i; Z_i is the standardised variable value; W_{ij} is the spatial weight matrix indicating whether i and j are neighbours; v_k is the number of common neighbours between i and j; G_i is the normalised complexity index, scaling P_i to the [0,1] interval; n is the total number of regions. A higher G_i value indicates greater spatial local complexity. A residual refers to the difference between the observed value (actual data) and the model-predicted value, reflecting the model's error in fitting the data (Santos and da Motta 2007). This study requires separate calculations of the geocomplexity values for both the dependent variable and the residuals of each model.

2.2.2 Degree of Geocomplexity (DG)

The DG metric quantifies a model's ability to capture local spatial patterns by comparing, under the same DG setting, the geocomplexity of the original data field with that of the residual field. Although residuals are ideally random in many models, spatial data often violate this assumption; therefore, the remaining spatial structure in residuals is informative for diagnosing what local patterns are not captured. DG is calculated as:

$$DG = \frac{\delta_0^G - \delta_r^G}{\delta_0^G} \tag{3}$$

where δ_0^G represents the geocomplexity of the dependent variable (data field), and δ_r^G denotes the geocomplexity of the corresponding residual field. This attenuation-based interpretation follows the residual-diagnostic rationale of the DSI framework (Liu, Song, and Yi 2026). Each observed point produces a corresponding residual, allowing for a point-wise calculation of DG. Note that the unnormalized geocomplexity values are used in this calculation to ensure the stability of the DG result. The value of

DG ranges from $-\infty$ to $+\infty$. The larger the DG value, the stronger the model's ability to capture spatial patterns; a value near 0 suggests the model fails to explain the spatial structure; negative values signify that the model distorts the spatial patterns. Because DG is computed point-wise, we summarise its distribution (e.g. mean and 95% CI) to compare models in terms of spatial interpretability under identical DG settings (same k and estimator). DG mainly reflects residual spatial structure (pattern attenuation), therefore, models with different residual variances or prediction biases can still be compared in terms of whether structured local spatial patterns remain in residuals.

3 Case study: assessing model performance in the spatial prediction of tree biomass in eastern Australian using DG

To rigorously test the diagnostic value of the DG metric, a case data must exhibit significant geocomplexity. The forest ecosystems of eastern Australia provide an ideal testbed for this purpose. Biomass distribution in this region—covering 21% of the land area (164.3 million hectares)—is a spatially complex phenomenon governed by diverse climatic and environmental gradients (Howell et al. 2008). Accurately quantifying tree biomass is crucial not only for assessing primary productivity but also for understanding carbon-nitrogen cycling (Scarascia-Mugnozza et al. 2000). For this validation, we used tree biomass plot data as the dependent variable and 16 environmental variables as predictors to construct ten common geospatial models. By systematically evaluating each model with both traditional non-spatial metrics and the DG, we aim to highlight the unique insights DG offers into spatial interpretability.

3.1 Study area

Eastern Australia was chosen as the study area. The study region (Figure 2a), covering the major terrestrial ecosystems of Queensland (QLD), New South Wales (NSW), and Victoria (VIC) (approx. 135°E–155°E, 10°S–40°S), is significant as a key hub for Australia's biodiversity and carbon stocks. Its diverse environmental gradients create the strong, complex spatial patterns in tree biomass necessary for a rigorous model validation scenario. Critically, the dense network of survey plots available here ensures a sufficient number of neighbours for each location, a prerequisite for the robust calculation of geocomplexity metric.

3.2 Data

3.2.1 Tree biomass data

We used a validated tree biomass dataset (1982–2015) from the Terrestrial Ecosystem Research Network (TERN), accessed via the Joint Remote Sensing Research Programme (JRSRP; https://field.jrsrp.com). The dependent variable comprises 7,823 sample plots, quantifying total above- and below-ground tree biomass (t/ha) within 1-ha circular plots. After preprocessing, the biomass values range from near 0 to 7,008.2 t/ha (National Forest Inventory Steering Committee 2019). This extreme upper limit remains ecologically plausible for eastern Australia (Keith, Mackey, and Lindenmayer 2009). The spatial distribution (WGS84 projection) is mapped using Australian Bureau of Statistics boundaries in Figure 2c. For visual clarity, point symbols are enlarged to indicate location, causing some apparent overlap.

3.2.2 Explanatory variables

We collected 19 environmental variables (Table 1) as predictors of tree biomass. These variables were grouped into three categories: soil, topography, and climate. We deliberately restricted the predictors to abiotic variables to focus on abiotic constraints. Soil variables included 11 parameters (e.g. pH, bulk density, organic carbon) sourced from the Soil and Landscape Grid of Australia (CSIRO). These datasets have a spatial resolution of 90 m and represent long-term means over 1950–2021. Topographic variables were derived via Google Earth Engine from a dataset at 5,000 m resolution and included elevation, slope, and aspect. To handle the circular nature of aspect (0°–360°), we decomposed it into two continuous variables: the sine and cosine of aspect, preventing model discontinuities (Brenning and Trombotto 2006; Brenning et al. 2015; Stage 1976). Climate variables, also obtained via Google Earth Engine at 5,000 m resolution,

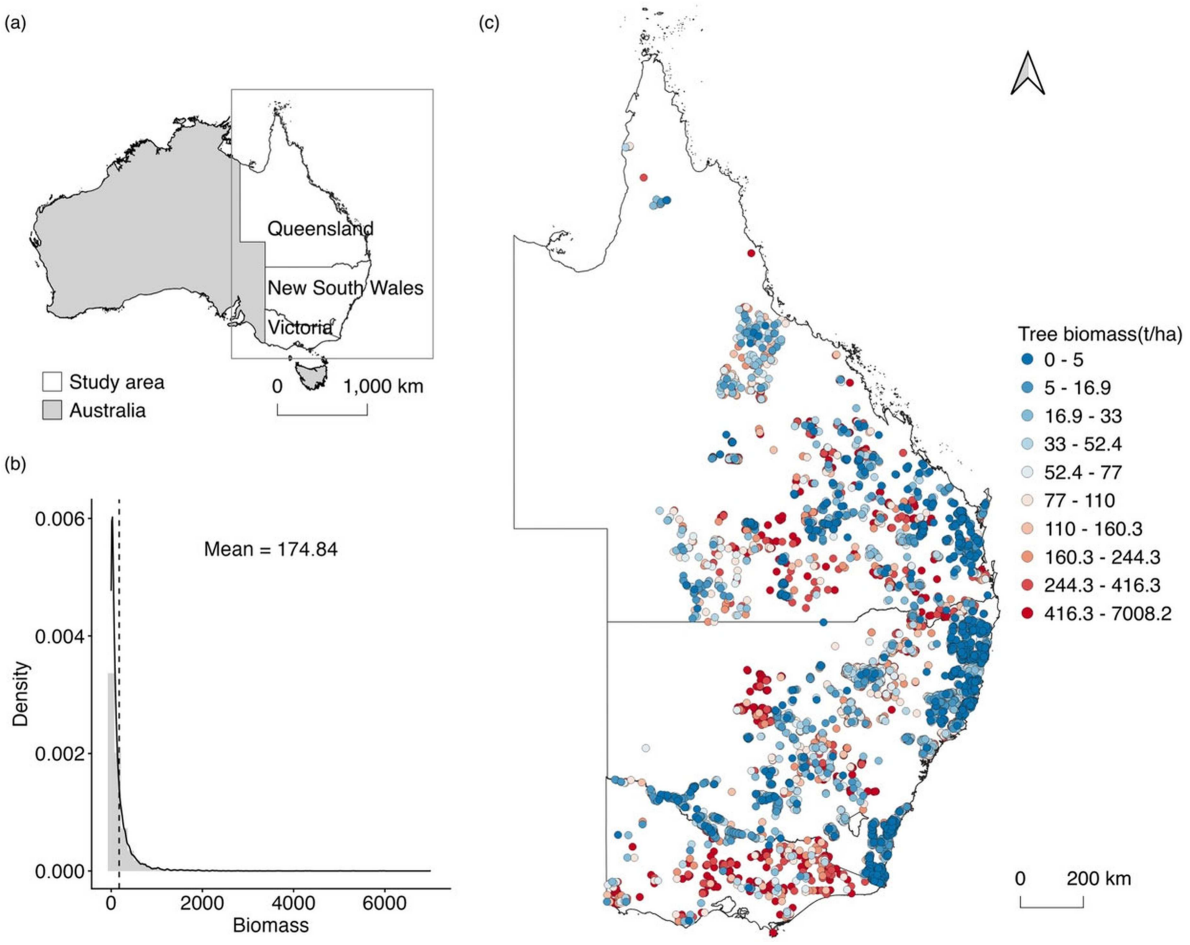

Figure 2. Study area and spatial distribution of the 7,823 tree biomass plots used in this study. (a) Location of the study area in eastern Australia, covering Queensland, New South Wales, and Victoria. (b) Distribution of observed tree biomass values across all plots, showing a strongly right-skewed distribution with a mean of 174.84 t/ha. (c) Spatial distribution of the 1-ha sampling plots in WGS84; point colors indicate tree biomass classes (t/ha).

consisted of mean precipitation, temperature, wind speed, and short-wave radiation, covering 1982–2015 (Chen et al. 2021; Michaletz, Kerkhoff, and Enquist 2018). For each sampling point, we computed the mean value of each environmental variable within the corresponding 1-ha circular plot (radius ≈ 56.4 m) and used these values as predictors. For high-resolution rasters (e.g. soil and topography), the plot mean aggregates multiple pixels within the plot; for coarse-resolution climate grids, the plot mean is effectively the value of the grid cell that contains the plot, representing background climatic constraints.

3.3 Experiment design

3.3.1 Data pre-processing

The response variable was total tree biomass, and its raw distribution was highly right-skewed (0.0004 to 7008.2 t/ha). To stabilise variance and reduce the influence of extreme values on model fitting and metrics, we applied a log transformation. Outliers were then screened in the log-transformed space as values beyond ± 3 standard deviations from the mean (Trautenmüller et al. 2021). We further excluded samples with missing predictors or duplicate coordinates. We assessed multicollinearity among the 19 predictors using a Pearson correlation matrix (Figure 3) and removed 3 variables from any pair with $|r| > 0.8$ (Fotheringham, Yang, and Kang 2017), resulting in the exclusion of short-wave radiation, soil organic carbon, and soil clay. The final dataset comprised the log-transformed biomass and 16 predictors (Figure 4).

Table 1. Explanatory variables used for tree biomass modelling.

Categories	Name	Description	Spatial resolution	Unit	Source
Soil	pH	0–5 cm soil average pH.	90m	\	CSIRO
	Depth	Vertical distance from surface to impermeable layer.	90m	m	CSIRO
	Bulk density	Dry weight of 0–5 cm soil per unit volume in natural state.	90m	g/cm^3	CSIRO
	Available water capacity	Available water capacity of 0–5 cm soil.	90m	%	CSIRO
	Cation exchange capacity (CEC)	Soil adsorption and exchange capacity of cations in solution for 0–5 cm soil.	90m	meq/100g	CSIRO
	Silt content	Mass percentage of 0.002–0.05 mm particle size in 0–5 cm soil.	90m	%	CSIRO
	Sand content	Mass percentage of 0.05–2.0 mm particle size in 0–5 cm soil.	90m	%	CSIRO
	Clay content	Mass percentage of <0.002 mm particle size in 0–5 cm soil.	90m	%	CSIRO
	Organic carbon	Soil organic carbon mass percentage of 0–5 cm soil relative to total oven-dried soil mass.	90m	%	CSIRO
	Total nitrogen	Soil nitrogen mass percentage of 0–5 cm soil relative to total oven-dried soil mass.	90m	%	CSIRO
	Total phosphorous	Soil phosphorus mass percentage of 0–5 cm soil relative to total oven-dried soil mass.	90m	%	CSIRO
Topography	Elevation	Vertical height above sea level.	5000m	m	SRTM
	Slope	The degree of inclination of the earth's surface relative to the horizontal.	5000m	°	SRTM
	sin (Aspect)	Take the sine value of Aspect.	5000m	°	SRTM
	cos (Aspect)	Take the cosine value of Aspect.	5000m	°	SRTM
Climate	Precipitation	Total precipitation from 1982 to 2015.	5000m	mm	CHIRPS
	Temperature	Annual average temperature at 2 metres height from 1982 to 2015.	5000m	°C	ERA5-Land
	Wind speed	Annual average wind speed from 1982 to 2015.	5000m	m/s	FLDAS
	Short-wave radiation	Daily average shortwave radiation from 1982 to 2015.	5000m	W/m^2	FLDAS

3.3.2 Spatial modelling

To comprehensively evaluate model performance, we selected ten distinct models spanning six major categories, from simple linear regression (lm) to complex ensembles (rf) (Table 2). All models were run using the caret package in R (Karunasingha 2022; Kuhn 2008). The pre-processed dataset was partitioned into a training set (70%) and a test set (30%) (Xu and Goodacre 2018). During training, we used 5-fold cross-validation on the training subset to tune hyperparameters and fit models, thereby reducing overfitting (Bates, Hastie, and Tibshirani 2024). The test subset was kept fully independent and used only for final performance reporting; for each of the ten models, we predicted on the test set and retained the residuals for subsequent DG calculation.

3.3.3 DG-based model validation

We implemented a validation based on DG to each model. The process began with computing the geocomplexity for the dependent variable to establish a baseline of its inherent spatial complexity. We then calculated geocomplexity for the test set residuals of the 10 models, which spatially identify the complexity that each model failed to explain. By integrating these baselines and residual geocomplexity values according to Equation 3, we got the DG result for each model, where higher values indicate a stronger ability to capture local spatial patterns. Throughout this process, the neighbourhood for each sample point was defined by its 18 nearest neighbours. Finally, by comparing the DG results across models, we generated an "Optimal DG Model Map" (Figure 8). This composite map displays the model with the highest DG value at each sample point location, thereby providing a clear visualisation of the spatial interpretability of the different models across the study area.

3.3.4 Non-spatial metrics-based model validation

We assessed the predictive accuracy of each model using conventional non-spatial metrics. This step provides a crucial baseline against which the spatial interpretability, measured by our DG metric, can be compared. We selected three complementary indicators: the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). R^2 quantifies the proportion of explained variance, serving as a measure of goodness-of-fit (Chicco et al. 2021; Fisher 1970). Concurrently, RMSE and MAE provide direct assessments of

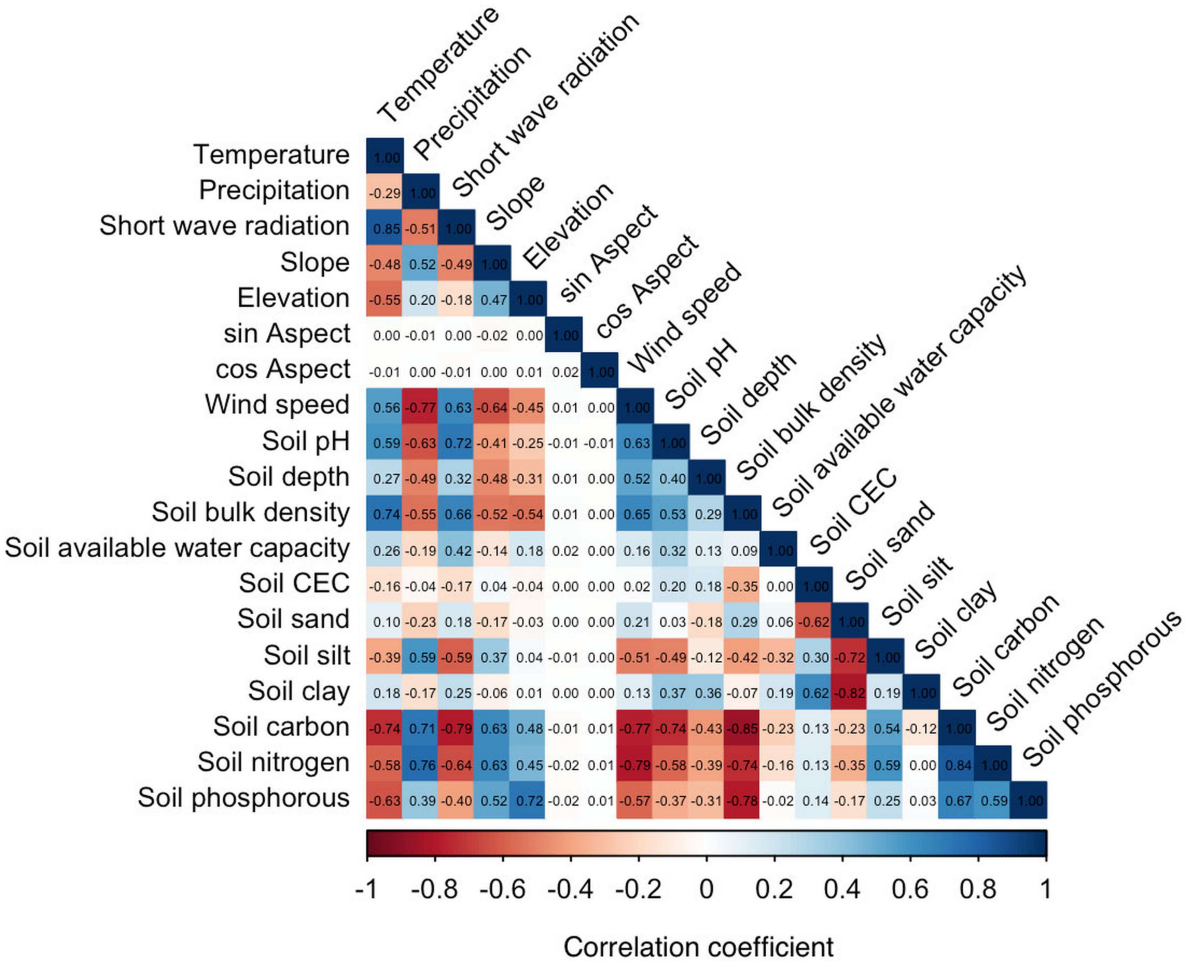

Figure 3. Pearson correlation matrix of the 19 initial environmental variables considered for biomass modelling. Cell values and colors indicate the strength and direction of pairwise correlations, ranging from −1 (red, strong negative correlation) to 1 (blue, strong positive correlation). This matrix was used for multicollinearity screening; variables involved in highly correlated pairs (| *r* |> 0.8) were excluded from subsequent modelling.

the average prediction error magnitude, reflecting model precision (Hyndman and Koehler 2006). The predictive performance of each model was visualised using scatterplots (Figure 9).

3.3.5 Comparison between DG and non-spatial metrics for model validation

To assess differences between DG and conventional accuracy metrics, we first summarised DG for each model by computing the mean DG and its 95% confidence interval (CI). This provides an overall measure of spatial interpretability in terms of geocomplexity. We then plotted these DG summaries together with the accuracy metrics in the same coordinate system for model-to-model comparison. In these plots, the y-axis shows the mean DG and its CI (the centre, upper, and lower ends of the error bars), while the x-axis shows R^2, RMSE, and MAE. This visualisation helps evaluate whether model-to-model changes show similar patterns across the two types of metrics.

4 Results

4.1 DG-based model validation

Figure 5 presents the spatial distribution and statistical summary of the geocomplexity for the dependent variable, tree biomass (denoted as δ_0^G). The histogram reveals a right-skewed distribution, with the majority of geocomplexity values (approximately 61.74%) concentrated in a narrow peak between 0.6 and 0.7, and a

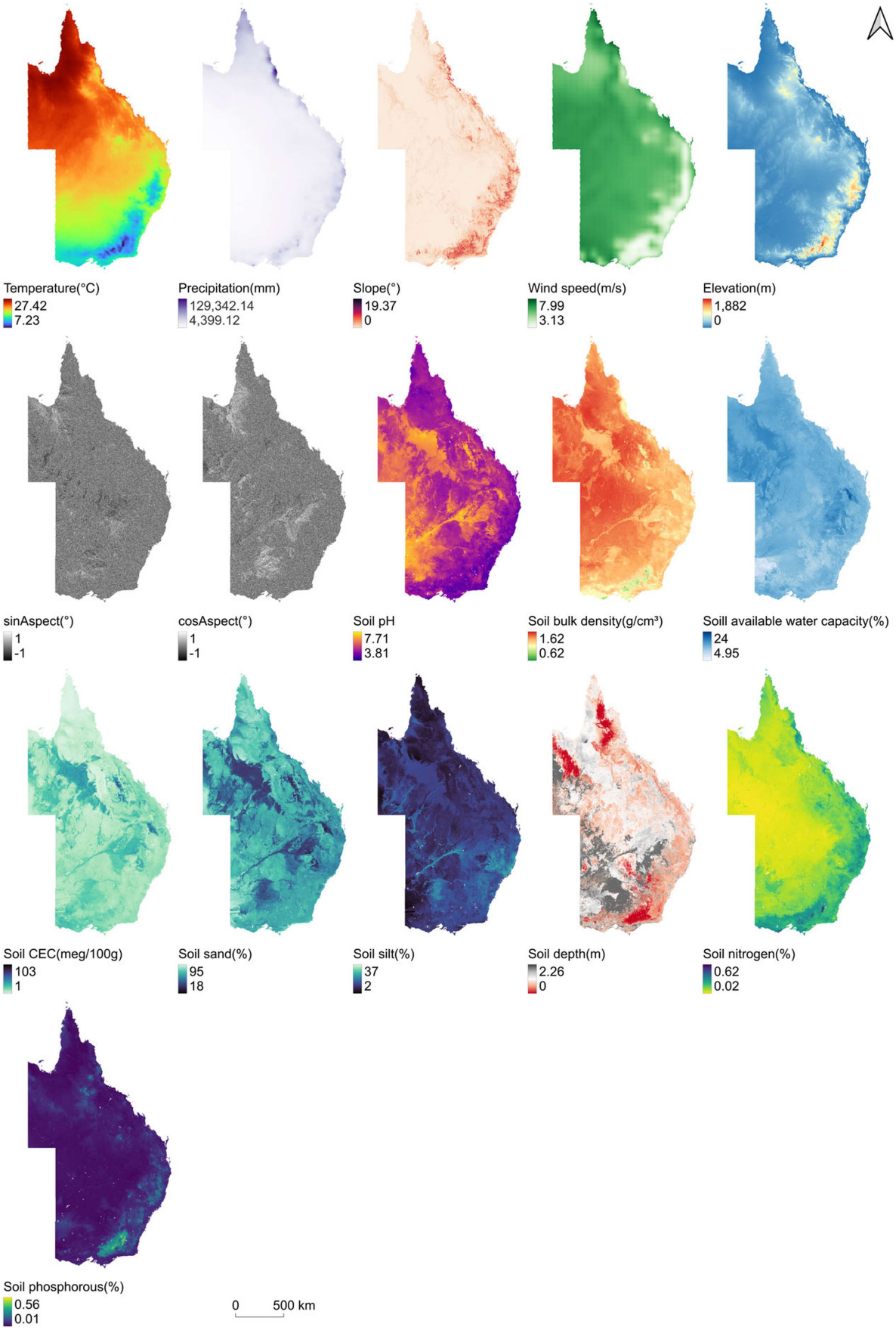
Temperature(°C)
27.42
7.23
Precipitation(mm)
129,342.14
4,399.12
Slope(°)
19.37
0
Wind speed(m/s)
7.99
3.13
Elevation(m)
1,882
0
sinAspect(°)
1
-1
cosAspect(°)
1
-1
Soil pH
7.71
3.81
Soil bulk density(g/cm³)
1.62
0.62
Soill available water capacity(%)
24
4.95
Soil CEC(meg/100g)
103
1
Soil sand(%)
95
18
Soil silt(%)
37
2
Soil depth(m)
2.26
0
Soil nitrogen(%)
0.62
0.02
Soil phosphorous(%)
0.56
0.01
0
500 km

Figure 4. Spatial distribution of the 16 environmental predictors retained for biomass modelling after correlation screening. Each panel shows the spatial pattern of one predictor across the study area, including climate variables (temperature, precipitation, wind speed), topographic variables (elevation, slope, sin aspect, cos aspect), and soil variables (pH, bulk density, available water capacity, CEC, sand, silt, depth, nitrogen, and phosphorous).

Table 2. Classification and description of the ten predictive models used for comparison.

Type	Model name	Model code
Linear models	Linear regression	lm
Tree-based models	Random forest	rf
	Cubist model	cubist
	Conditional random forest	cforest
	Extreme gradient boosting (XGBoost)	xgbTree
	Gradient boosting machine (GBM)	gbm
Support vector machine-based models	Support vector machine (SVM)	svmRadial
Neighbourhood-based models	k-Nearest neighbours (k-NN)	knn
Non-linear and adaptive regression models	Generalised additive model (GAM)	gam
Dimensionality reduction regression models	Partial least squares regression (PLS)	pls

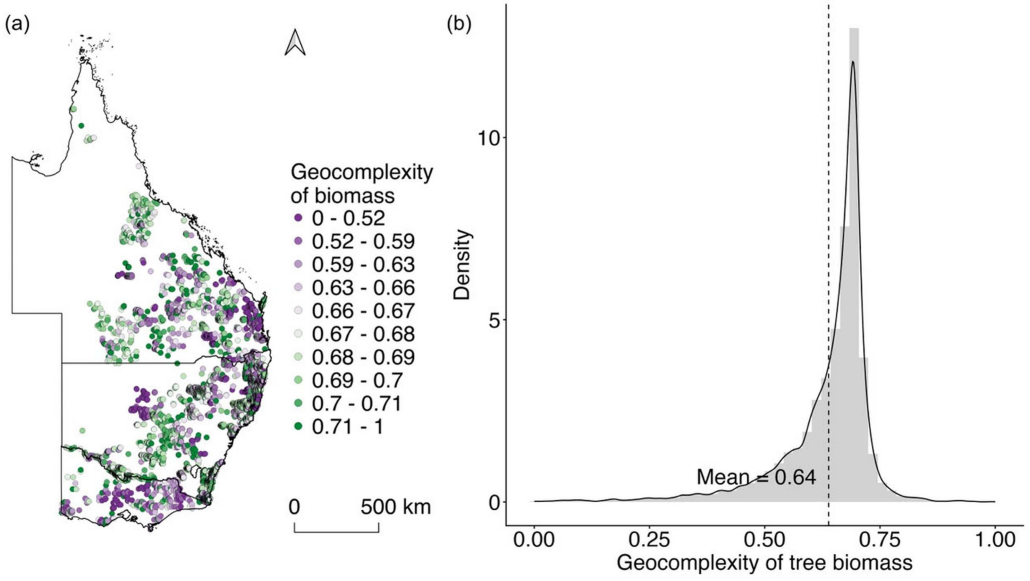

Figure 5. Spatial pattern and frequency distribution of geocomplexity for the observed tree biomass field (δ_0^G). (a) Point-wise geocomplexity values at the 7,823 sampling plots across eastern Australia. (b) Histogram and density curve of geocomplexity values, with the dashed line indicating the mean (0.64). Most values are concentrated between 0.6 and 0.7, indicating generally high local geocomplexity, with higher values more common in coastal and southeastern regions.

mean of 0.64. This indicates that the biomass patterns across eastern Australia are characterised by a predominantly high degree of geocomplexity, signifying complex local spatial structures where values often deviate from simple neighbourhood similarities. The map further illustrates a distinct spatial differentiation in these geocomplexity values. High-geocomplexity hotspots, representing the most complex spatial arrangements, are prevalent along the coastal and southeastern regions. In contrast, areas with lower geocomplexity, indicating more homogeneous or predictable local patterns, are more frequently observed in the inland regions of New South Wales and Victoria.

Figure 6 visualises the geocomplexity of the residuals (δ_r^G) for each of the ten models, offering a diagnostic map of their spatial performance. Higher residual geocomplexity indicates a larger portion of unexplained spatial complexity, signifying a model's inadequacy in capturing local patterns.

The results reveal significant differences among models. Models such as rf (mean = 0.46), cubist (0.46), and cforest (0.49) exhibit the lowest residual geocomplexity, indicating superior performance in capturing geocomplexity. Their residual geocomplexity are concentrated in the 0.4–0.5 range. In contrast, knn (0.65) and svmRadial (0.60) show the highest residual geocomplexity, with values mainly between 0.6 and 0.7,

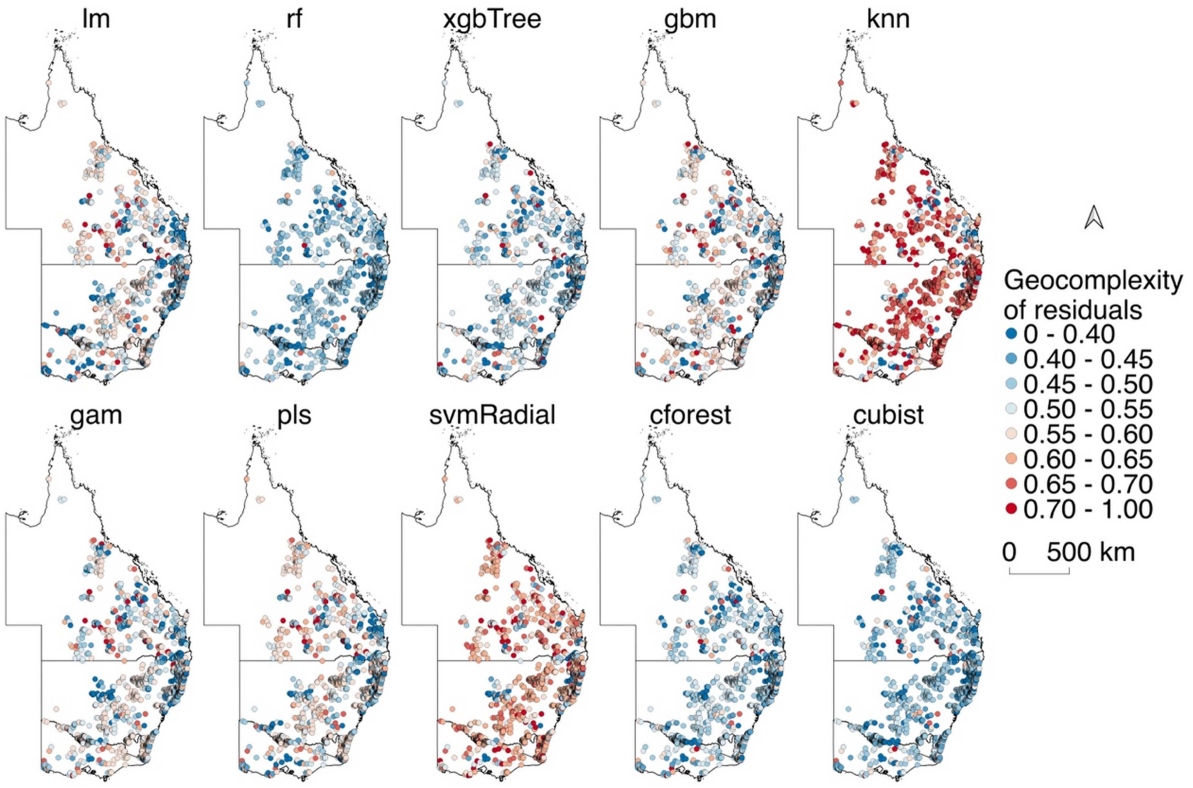

Figure 6. Spatial distribution of residual geocomplexity (δ_r^G) for the ten predictive models. Each panel shows the point-wise geocomplexity of model residuals for one model. Residual geocomplexity was computed under the same DG setting as the observed data (18 nearest neighbours). Lower residual geocomplexity indicates that less local spatial structure remains unexplained and therefore reflects better model performance in capturing geocomplexity; higher values indicate stronger uncaptured local spatial structure.

suggesting substantial uncaptured spatial structure by the models. The remaining models perform intermediately. While all models show some localised patches of low residual geocomplexity, the prevalence of high-geocomplexity residuals in models like knn clearly illustrates their weaker spatial interpretability compared to top-performers like rf.

Figure 7 displays the spatial distribution of the DG for each model, providing a direct visualisation of the model's spatial interpretability, higher DG values signify a stronger ability to explain local geocomplexity. The maps reveal stark differences in performance. Based on the average DG values, the rf (mean DG = 1.61), knn (1.33), and xgbTree (1.29) models demonstrate the highest explanatory power, showing widespread high DG values. In contrast, lm (–0.57) and pls (–0.49) perform poorly, with predominantly low or negative DG values, indicating they often distort the underlying spatial patterns. The remaining models, including svmRadial, cforest, gbm, gam, and cubist, exhibit moderate performance with average DG values ranging from 0.04 to 0.74. Within each map, significant spatial heterogeneity is evident. The spatial patterns of DG values are inversely related to the residual geocomplexity patterns shown in Figure 6, confirming that DG effectively identifies where models succeed or fail in capturing geocomplexity.

Figure 8 synthesises the DG results to illustrate the geocomplexity interpretability of each model, challenging the notion of a single best model. The map (Figure 8a) assigns to each point the model that achieved the highest local DG value, revealing a complex pattern of optimal model performance. The bar chart (Figure 8d quantifies this heterogeneity. The rf model is most frequently optimal, achieving the highest DG at 23.55% of locations ($n = 552$), followed by knn (19.28%, $n = 452$) and xgbTree (16.60%, $n = 389$). This suggests these three models have the broadest adaptability to varying geocomplexity. Spatially, the map reveals distinct patterns of model preference. For instance, while top performers like rf are widely distributed, other models exhibit clear geographic niches. The lm model shows a notable

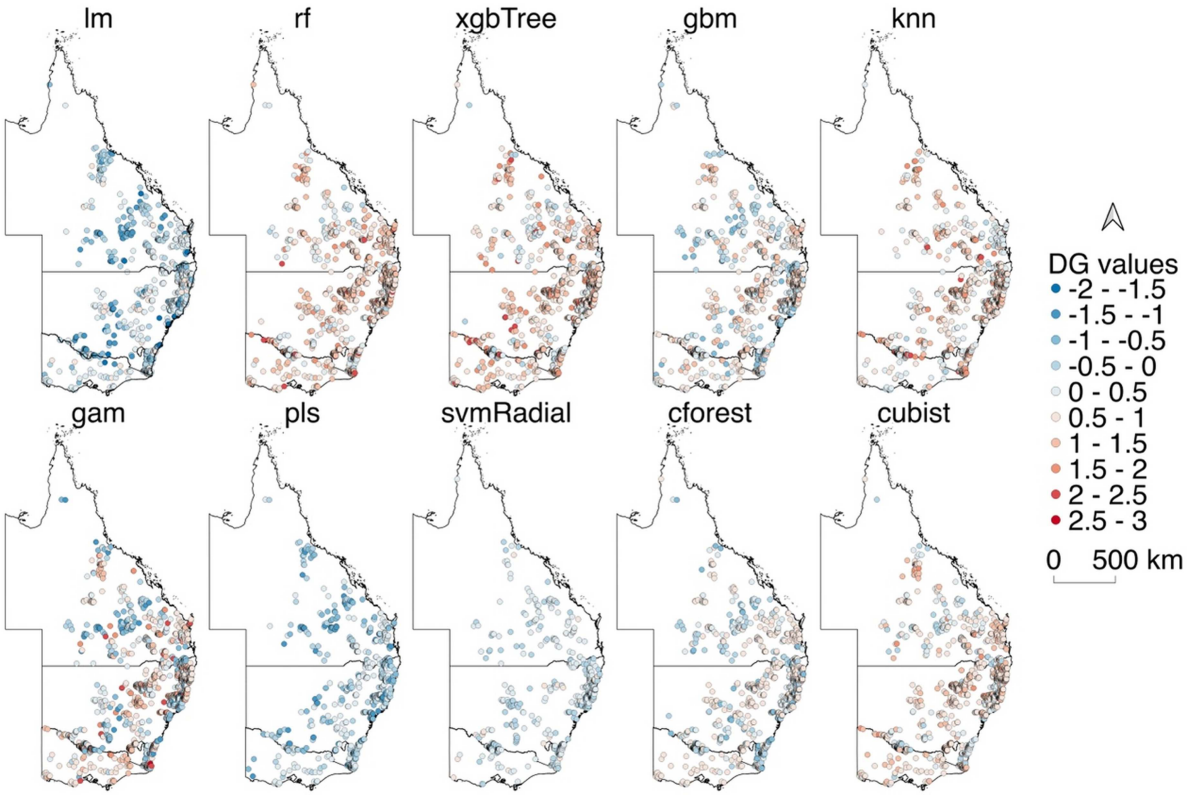

Figure 7. Spatial distribution of the DG values for ten models. Positive DG values indicate attenuation of local spatial structure by the model, values near 0 indicate little explanatory effect, and negative values indicate distortion of local spatial patterns.

concentration in southern Queensland, while knn is particularly prevalent in the southeastern coastal areas. This underscores the necessity of localised model selection, a key practical implication of the DG metric.

4.2 Comparison between DG and accuracy metrics for model validation

Figure 9 uses scatterplots to visualise each model's fit and performance in terms of non-spatial metrics. Overall, the model performance is moderate, with test set R^2 values ranging from 0.052 to 0.347. This is a reasonable outcome given that our model relies solely on environmental variables, omitting other key drivers of biomass variability like vegetation dynamics and functional traits (Cui et al. 2020; Fararoda et al. 2021).

For easier comparison, we classified model performance based on test set R^2 into three tiers: "Good" ($R^2 \geq 0.3$), "Medium" ($0.2 \leq R^2 < 0.3$), and "Weak" ($R^2 < 0.2$). Table 3 summarises these results. The models rf, cforest, and cubist achieved "Good" performance, demonstrating the highest accuracy. "Medium" performance was exhibited by xgbTree, gbm, and svmRadial. The remaining models, including gam, knn, lm, and pls, were categorised as "Weak," showing the poorest predictive power. This ranking establishes the traditional, non-spatial hierarchy of model performance, which will be contrasted with the DG-based evaluation in the following section.

Figure 10 synthesises our two dimensions evaluation, revealing the relationship between predictive accuracy and DG. The results indicate that rf ($R^2 = 0.347$, mean DG = 1.61), cubist, and xgbTree achieved a strong balance, demonstrating both high accuracy and robust DG. In contrast, some models show a clear preference for a single metric dimension, for instance, cforest is a prime example of an "accuracy-dominant" profile; despite its high R^2 (0.325), its mean DG is near zero (0.078), suggesting its predictive power comes at the cost of capturing local geocomplexity. Conversely, models such as knn and gam showed low predictive accuracy but relatively high mean DG values, suggesting stronger attenuation of local spatial structure from

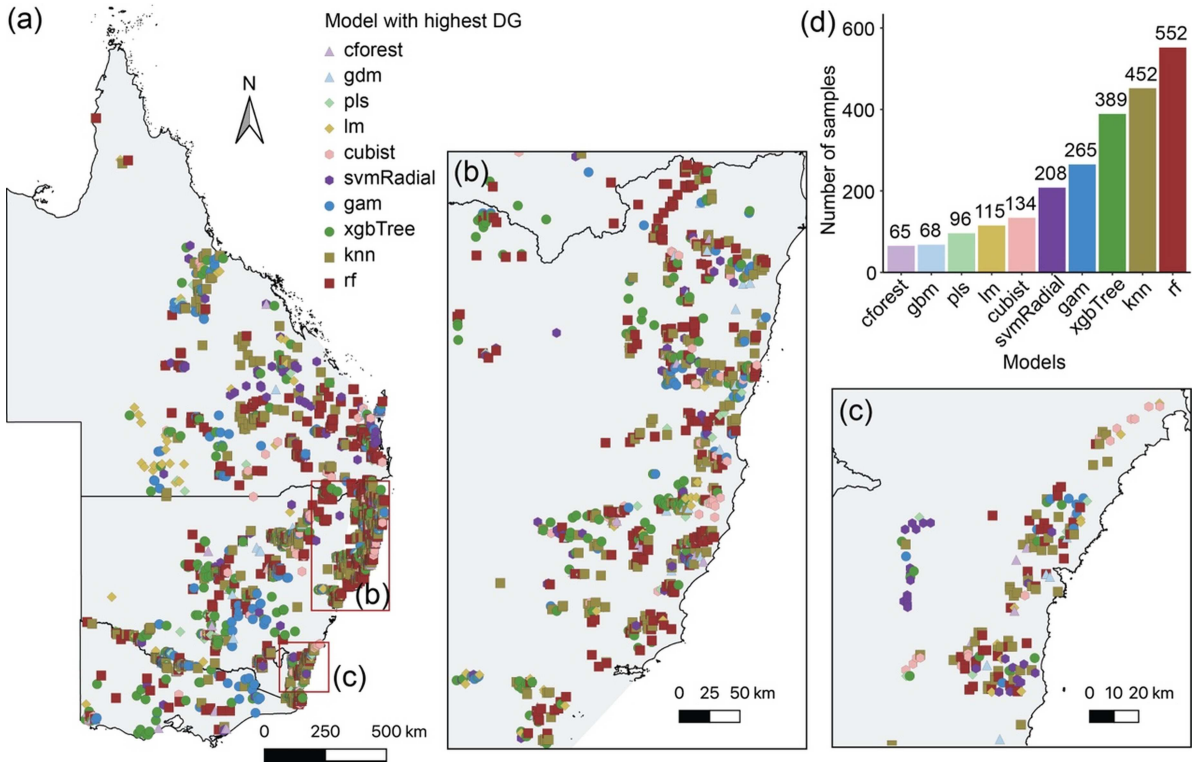

Figure 8. Locally optimal models identified using the DG metric. (a) Map showing, for each sampling location, the model with the highest local DG value. (b–c) Enlarged views of two sample-dense regions highlighted in panel (a), illustrating fine-scale spatial heterogeneity in optimal model selection. (d) Number of locations at which each model achieved the highest DG. The figure shows that no single model is optimal everywhere, although rf, knn, and xgbTree are most frequently selected as the locally best-performing models.

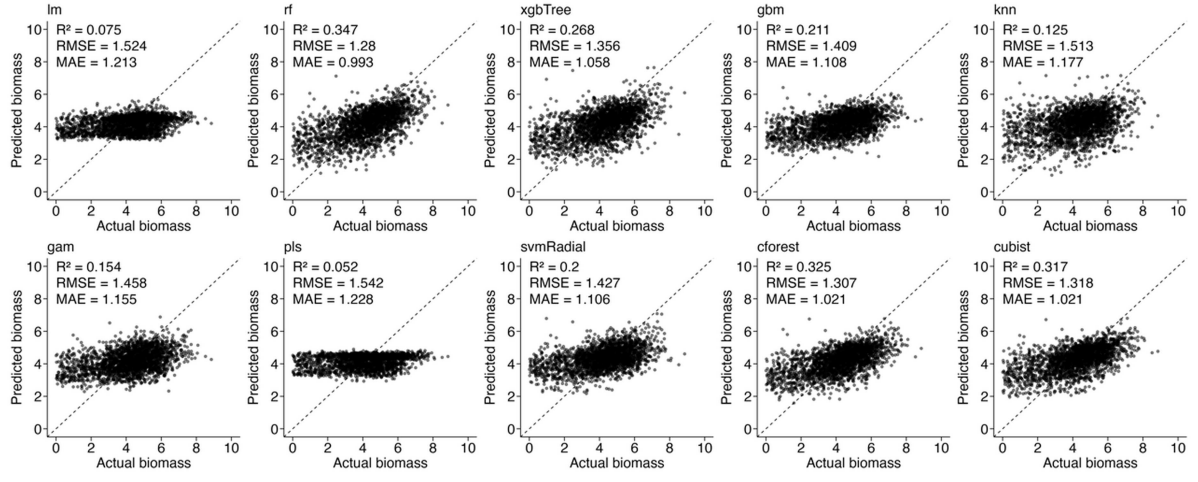

Figure 9. Comparison of non-spatial performance across models (R^2, RMSE, MAE).

Table 3. Model performance classification and evaluation results.

Quality	Model	R^2 range	RMSE range	MAE range
Good	rf, cforest, cubist	[0.317, 0.347]	[1.280, 1.318]	[0.993, 1.021]
Medium	xgbTree, gbm, svmRadial	[0.200, 0.268]	[1.356, 1.427]	[1.058, 1.108]
Weak	lm, knn, gam, pls	[0.052, 0.154]	[1.458, 1.542]	[1.155, 1.228]

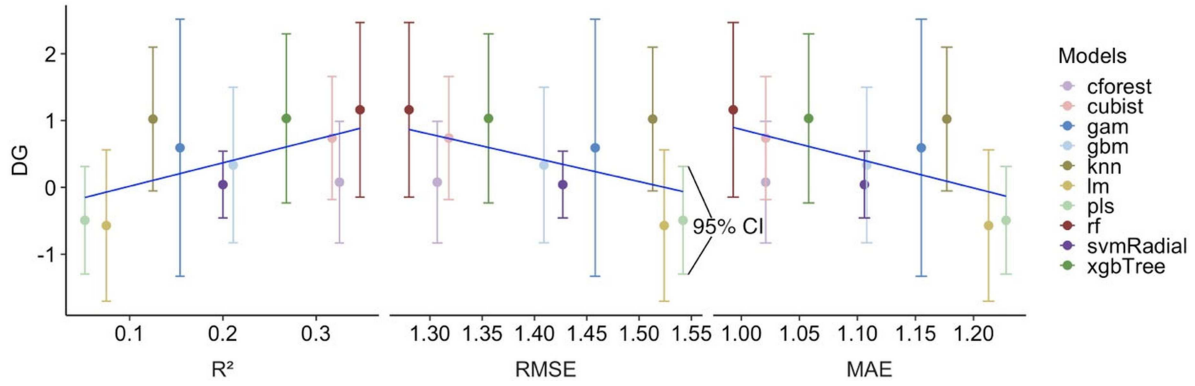

Figure 10. Dual-criterion comparison between spatial interpretability and predictive accuracy across the ten models. The y-axis shows mean DG with 95% confidence intervals, summarising each model's ability to attenuate local geocomplexity from the observed field to the residual field. The x-axis shows the corresponding conventional accuracy metrics (R^2, RMSE, and MAE). This figure highlights whether models ranked highly by predictive accuracy also perform well in terms of spatial interpretability.

data to residuals under our DG setting, even though their numerical prediction errors remain large. The fourth group, including lm, pls, and svmRadial, performed poorly on both dimensions, with low accuracy and negative mean DG values, signifying they were neither precise nor spatially insightful. This multifaceted analysis critically demonstrates that predictive accuracy and spatial interpretability are not synonymous, highlighting the diagnostic power of the DG metric in uncovering model behaviours that traditional metrics alone would mask. Accordingly, model choice should be guided by the study objective: accuracy metrics for predictive usefulness, and DG for diagnosing structured spatial residual patterns and spatial interpretability.

5 Discussion

5.1 DG as a local spatial diagnostic for model validation

The DG metric proposed in this study aims to fill a gap in model validation that is not fully addressed by existing spatial diagnostics. Conventional accuracy metrics (R^2, RMSE, MAE) summarise overall error magnitude and can therefore mask systematic local biases, whereas common spatial diagnostics based on residuals (e.g. Moran's I and local Moran's I/LISA) mainly answer whether residuals show spatial autocorrelation or clustering, focusing on the strength and significance of spatial dependence rather than whether a model has truly explained local geocomplexity (Anselin 1995). DG advances the validation target from "whether residuals are spatially correlated" to "whether the model effectively attenuates local spatial structure from the data field to the residual field under the same neighbourhood scale," thereby providing a complementary, spatially interpretable evaluation dimension alongside accuracy metrics and enabling spatially located diagnostics for identifying where a model is applicable or fails, which supports dynamic local model selection. DG can differ from R^2/RMSE/MAE because reducing the overall error does not necessarily mean that the model has removed neighbourhood-scale structure from the residuals. For instance, a model may achieve a low RMSE by smoothing/averaging, yet still leave structured residual patterns at the same neighbourhood scale, resulting in low DG. In addition, when predictors do not match the process scale, model failures can be confined to specific regions; global metrics then average these effects, whereas DG reveals them as localised areas with persistent residual structure under the same neighbourhood definition.

5.2 Future directions and robustness considerations

This study also points to several directions for further improvement. First, we intentionally used only abiotic predictors; future work could incorporate vegetation-related variables (e.g. canopy structure and functional traits) to improve predictive performance and to test DG under richer process representations (Fotheringham, Yang, and Kang 2017). Second, DG depends on the neighbourhood definition (k and

weighting scheme); to examine robustness, we selected four models representing distinct behaviours (rf: high accuracy with high DG; cforest: high accuracy but near-zero DG; knn: low accuracy but high DG; lm: weak in both), recomputed point-wise DG under the same computational framework for k = 14, 18, 22, and 26, and summarised results using the mean DG with 95% confidence intervals. The results show that the overall conclusions remain stable within this reasonable range; nevertheless, future studies could systematically compare alternative neighbourhood schemes and develop more robust summaries and inference for potentially heavy-tailed DG distributions (Anselin 1995). Third, because predictors were drawn from datasets with heterogeneous spatial resolutions, we used plot-support mean extraction rather than forcing a single pixel size; future work could further examine scale/support effects and adopt spatially explicit validation to better control spatial leakage (Gotway and Young 2002).

6 Conclusions

This study has developed and validated a DG metric that examines a spatial model's ability to capture local spatial patterns by comparing the local complexity of original data with that of model residuals, leading to three key conclusions. First, by measuring how well models explain local geocomplexity, DG fills a gap in model validation that relies mainly on accuracy metrics. Second, the DG analysis shows that no single model is optimal everywhere, highlighting the need for a dynamic, local model-selection approach. Third, these findings confirm the critical role of DG in bridging the gap between statistical accuracy and spatial interpretability in model validation. The DG metric promotes more geographically adaptive and robust modelling practices, providing researchers and decision-makers with a more reliable tool. Future research can extend the evaluation metric established in this study from geocomplexity to other key geospatial features, thereby forming a more comprehensive system for evaluating model spatial performance.

Author contributions

CRediT: **Haiyang Liu:** Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft; **Yongze Song:** Conceptualization, Formal analysis, Methodology, Resources, Supervision, Visualization, Writing – review & editing; **Wen Yi:** Supervision, Writing – review & editing; **Pengcheng Zhang:** Validation, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

No funding was received for conducting this study.

ORCID

Yongze Song 0000-0003-3420-9622

Data availability statement

The processing data generated in this study (including tree biomass and environmental variables data) are publicly available on figshare at https://doi.org/10.6084/m9.figshare.28645883 (Liu et al. 2026).

Open scholarship

This article has earned the Centre for Open Science badges for Open Data and Open Materials. The data are openly accessible at re3data.org; materials are openly accessible at re3data.org.

References

Anselin, L. 1995. "Local Indicators of Spatial Association—Lisa." *Geographical analysis* 27 (2): 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Bates, S., T. Hastie, and R. Tibshirani. 2024. "Cross-Validation: What Does It Estimate and How Well Does It Do It?" *Journal of the American Statistical Association* 119 (546): 1434–1445. https://doi.org/10.1080/01621459.2023.2197686.

Bennett, N. D., B. F. Croke, G. Guariso, J. H. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, et al. 2013. "Characterising Performance of Environmental Models." *Environmental modelling & software* 40: 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.

Brenning, A. 2005. "Spatial Prediction Models for Landslide Hazards: Review, Comparison and Evaluation." *Natural Hazards and Earth System Sciences* 5 (6): 853–862. https://doi.org/10.5194/nhess-5-853-2005.

Brenning, A., and D. Trombotto. 2006. "Logistic Regression Modeling of Rock Glacier and Glacier Distribution: Topographic and Climatic Controls in the Semi-Arid Andes." *Geomorphology* 81 (1-2): 141–54. https://doi.org/10.1016/j.geomorph.2006.04.003.

Brenning, A., M. Schwinn, A. P. Ruiz-Páez, and J. Muenchow. 2015. "Landslide Susceptibility Near Highways Is Increased By 1 Order of Magnitude in the Andes of Southern Ecuador, Loja Province." *Nat Hazards Earth Syst Sci* 15: 45–57. https://doi.org/10.5194/nhess-15-45-2015.

Chen, R., J. Ran, W. Hu, L. Dong, M. Ji, X. Jia, and J. Lu, et al. 2021. "Effects of Biotic and Abiotic Factors on Forest Biomass Fractions." *National Science Review* 8 (10): nwab025. https://doi.org/10.1093/nsr/nwab025.

Chicco, D., M. J. Warrens, and G. Jurman. 2021. "The Coefficient of Determination R-squared Is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation." *Peerj computer science* 7: e623. https://doi.org/10.7717/peerj-cs.623.

Cui, E., E. Weng, E. Yan, and J. Xia. 2020. "Robust Leaf Trait Relationships Across Species under Global Environmental Changes." *Nature communications* 11 (1): 2999. https://doi.org/10.1038/s41467-020-16839-9.

Fararoda, R., R. S. Reddy, G. Rajashekar, T. K. Chand, C. Jha, and V. Dadhwal. 2021. "Improving Forest above Ground Biomass Estimates over Indian Forests Using Multi Source Data Sets with Machine Learning Algorithm." *Ecological Informatics* 65: 101392. https://doi.org/10.1016/j.ecoinf.2021.101392.

Fisher, R. A. 1970. Statistical Methods for Research Workers In *Breakthroughs in statistics: Methodology and distribution* 66–70, New York, NY: Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_6

Fotheringham, A. S., W. Yang, and W. Kang. 2017. "Multiscale Geographically Weighted Regression (Mgwr)." *Annals of the American Association of Geographers* 107 (6): 1247–1265. https://doi.org/10.1080/24694452.2017.1352480.

Goodchild, M. F. 2009. "Geographic Information Systems and Science: Today and Tomorrow." *Annals of GIS* 15 (1): 3–9. https://doi.org/10.1080/19475680903250715.

Gotway, C. A., and L. J. Young. 2002. "Combining Incompatible Spatial Data." *Journal of the American Statistical Association* 97 (458): 632–648. https://doi.org/10.1198/016214502760047140.

Heuvelink, G. B., P. A. Burrough, and A. Stein. 1989. "Propagation of Errors in Spatial Modelling with Gis." *International Journal of Geographical Information System* 3 (4): 303–322. https://doi.org/10.1080/02693798908941518.

Howell, C. I., A. Wilson, S. Davey, and M. Eddington. 2008. "Sustainable Forest Management Reporting in Australia." *Ecological indicators* 8 (2): 123–130. https://doi.org/10.1016/j.ecolind.2006.11.004.

Hyndman, R. J., and A. B. Koehler. 2006. "Another Look at Measures of Forecast Accuracy." *International journal of forecasting* 22 (4): 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Karunasingha, D. S. K. 2022. "Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well." *Information Sciences* 585: 609–629. https://doi.org/10.1016/j.ins.2021.11.036.

Keith, H., B. G. Mackey, and D. B. Lindenmayer. 2009. "Re-Evaluation of Forest Biomass Carbon Stocks and Lessons from the world's Most Carbon-Dense Forests." *Proceedings of the National Academy of Sciences* 106 (28): 11635–11640. https://doi.org/10.1073/pnas.0901970106.

Kuhn, M. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of statistical software* 28: 1–26. https://doi.org/10.18637/jss.v028.i05.

Liu, Haiyang, Yongze Song, Wen Yi, and Pengcheng Zhang, et al. 2026. "Degree of Geocomplexity for Diagnosing Spatial Pattern Loss Beyond Prediction Accuracy." figshare. https://doi.org/10.6084/m9.figshare.28645883.v5.

Liu, H., Y. Song, and W. Yi. 2026. "Degree of Spatial Interpretability." *International Journal of Geographical Information Science.* https://doi.org/10.1080/13658816.2026.2614335.

Luo, P., Y. Song, X. Huang, H. Ma, J. Liu, Y. Yao, and L. Meng. 2022. "Identifying Determinants of Spatio-Temporal Disparities in Soil Moisture of the Northern Hemisphere Using a Geographically Optimal Zones-Based Heterogeneity Model." *Journal of Photogrammetry and Remote Sensing* 185: 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009.

Meyer, H., and E. Pebesma. 2022. "Machine Learning-Based Global Maps of Ecological Variables and the Challenge of Assessing Them." *Nature Communications* 13 (1): 2208. https://doi.org/10.1038/s41467-022-29838-9.

Michaletz, S. T., A. J. Kerkhoff, and B. J. Enquist. 2018. "Drivers of Terrestrial Plant Production Across Broad Geographical Gradients." *Global Ecology and Biogeography* 27 (2): 166–174. https://doi.org/10.1111/geb.12685.

Mitchard, E. T. 2018. "The Tropical Forest Carbon Cycle and Climate Change." *Nature (London)* 559 (7715): 527–534.

National Forest Inventory Steering Committee. 2018. National Forest Inventory Steering Committee. Australia's state of the forests report 2018.

Openshaw, S. 1984. The modifiable areal unit problem. Concepts and techniques in modern geography.
Ripley, B. D. 2001. "Spatial Statistics in R." *R news* 1 (2): 14–15.
Rykiel, E. J. 1996. "Testing Ecological Models: The Meaning of Validation." *Ecological modelling* 90 (3): 229–244. https://doi.org/10.1016/0304-3800(95)00152-2.
Santos, N. J., and S. J. da Motta. 2007. "Residual Analysis for Linear Mixed Models." *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 49 (6): 863–875. https://doi.org/10.1002/bimj.200610341.
Scarascia-Mugnozza, G., M. M. Caldwell, G. Heldmaier, O. L. Lange, H. A. Mooney, E. Schulze, U. Sommer, et al. 2000. Tree Biomass, Growth and Nutrient Pools In *Carbon and nitrogen cycling in European forest ecosystems* Vol. 142 49–62, Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-57219-7_3
Song, Y. 2023. "Geographically Optimal Similarity." *Mathematical Geosciences* 55 (3): 295–320. https://doi.org/10.1007/s11004-022-10036-8.
Stage, A. R. 1976. "An Expression for the Effect of Aspect, Slope, and Habitat Type on Tree Growth." *Forest Science* 22 (4): 457–60. https://doi.org/10.1093/forestscience/22.4.457.
Trautenmüller, J. W., S. Péllico Netto, R. Balbinot, L. F. Watzlawick, A. P. Dalla Corte, C. R. Sanquetta, and A. Behling. 2021. "Regression Estimators for Aboveground Biomass and Its Constituent Parts of Trees in Native Southern Brazilian Forests." *Ecological Indicators* 130: 108025. https://doi.org/10.1016/j.ecolind.2021.108025.
Xu, Y., and R. Goodacre. 2018. "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning." *Journal of analysis and testing* 2 (3): 249–262. https://doi.org/10.1007/s41664-018-0068-2.
Zhang, Z., Y. Song, P. Luo, and P. Wu. 2023. "Geocomplexity Explains Spatial Errors." *International Journal of Geographical Information Science* 37 (7): 1449–1469. https://doi.org/10.1080/13658816.2023.2203212.