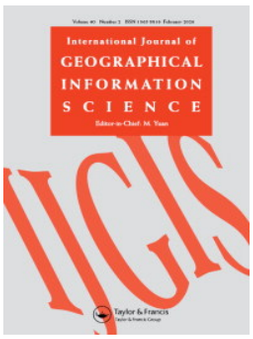

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Degree of spatial interpretability

Haiyang Liu, Yongze Song & Wen Yi

To cite this article: Haiyang Liu, Yongze Song & Wen Yi (26 Jan 2026): Degree of spatial interpretability, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2026.2614335

To link to this article: https://doi.org/10.1080/13658816.2026.2614335

Published online: 26 Jan 2026.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2026.2614335

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Degree of spatial interpretability

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

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

ABSTRACT

Model validation ensures that mathematical models accurately represent real-world phenomena and meet scientific requirements in the Earth sciences. Current validation approaches for spatial modelling mainly use accuracy metrics such as goodness of fit and prediction error to validate regression, machine learning, spatial models and geospatial intelligence models. However, existing spatial validation still faces challenges in assessing models' capacity for spatial interpretability. This study develops a degree of spatial interpretability (DSI) indicator to evaluate the effectiveness and ability of spatial models to capture spatial characteristics such as spatial autocorrelation and heterogeneity and to support model selection. DSI is implemented to evaluate ten common models for spatial prediction of Australian vascular plant species diversity using nineteen abiotic explanatory variables. The results reveal that strong goodness of fit and low prediction error do not necessarily correspond to high spatial interpretability and demonstrate the need to employ DSI in model evaluation. The experiments confirm the importance of assessing a model's ability to explain spatial characteristics. The DSI indicator enhances model evaluation by providing a spatially informed assessment perspective and has strong potential to advance model evaluation systems.

ARTICLE HISTORY
Received 10 December 2024
Accepted 5 January 2026

KEYWORDS
Spatial interpretability; model validation; spatial autocorrelation; spatial heterogeneity; species diversity

1. Introduction

Earth science research usually concerns large-scale systems that evolve slowly so observations and models are both required to infer system behavior and predict beyond available measurements (Tredennick *et al.* 2021). Spatial prediction models support practical and efficient inference and prediction by characterizing complex ecosystem responses and predicting ecosystem responses to key drivers (Jørgensen 2008). Model validation is critical to ensure that models accurately reflect real-world phenomena and meet scientific standards (Planque *et al.* 2022). Current validation of spatial modelling still relies predominantly on accuracy metrics such as goodness of fit and prediction error across regression, machine learning, spatial models (e.g. geographically weighted

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

regression, GWR) and geospatial intelligence models (e.g. convolutional neural network, CNN) in addressing Earth and urban applications (Song *et al*. 2026). Methods for assessing model accuracy include direct value comparisons (e.g. mean absolute error (MAE), root mean square error (RMSE)), coupling between observed and modeled values (e.g. Nash–Sutcliffe efficiency (NSE), coefficient of determination (R^2)) and indirect parameter-based indicators (Bennett *et al*. 2013, Augusiak *et al*. 2014).

Spatial analysis helps us better understand the application of data in spatial distribution, relationships and patterns of change by explaining processes and spatial structures occurring on the Earth's surface (Jiang 2007, Longley *et al*. 2015). Understanding these spatial characteristics – which fundamentally encompass attributes such as distribution patterns, spatial dependence (autocorrelation) and spatial heterogeneity (Longley *et al*. 2015) – is a higher-level requirement for assessing model performance. Previous studies have shown that considering spatial autocorrelation in ecological modeling is crucial; otherwise, it may lead to biased estimates and increased error rates (Beale *et al*. 2010). The spatial characteristics of data can also influence the dependence of model residuals, affecting statistical bias and model accuracy (Dormann *et al*. 2007, Elith and Leathwick 2009). Ignoring spatial characteristics can also result in changes to regression coefficients, leading to incorrect interpretations of relationships between data (Legendre 1993, Bini *et al*. 2009). There is a close relationship between data, models and spatial characteristics (Guisan and Thuiller 2005). However, relying only on traditional accuracy metrics does not fully capture the complexity of Earth sciences, especially when dealing with data with significant spatial characteristics.

The existing model evaluation system needs further development. First, accuracy metrics only reflect model performance from the perspectives of fitness and errors, leading to a homogeneous assessment of models. Second, spatial attributes are among the most essential properties of geoscience models, yet current evaluations lack assessments of a model's ability to capture spatial characteristics. Third, quantitative metrics are inadequate for evaluating a model's performance in capturing spatial information and fail to provide comprehensive information. While spatially explicit assessment methods exist, such as Smulders *et al*. (2010) using conditional randomization to identify regions of inaccurate SDM prediction, these primarily focus on the spatial distribution of predictive accuracy. However, focusing on the location of prediction errors neglects whether the model captures and explains the inherent spatial characteristics of the data.

This study develops the degree of spatial interpretability (DSI) to assess the effectiveness of a model in capturing spatial characteristics, such as spatial autocorrelation and heterogeneity, through its computational processes. DSI provides an objective evaluation metric for models requiring spatial analysis. The study used the relationship between vascular plant diversity distribution and abiotic factors in Australia as a case study to validate the feasibility of DSI while also providing a reference for biodiversity conservation and research in Australia.

2. Degree of spatial interpretability

2.1. Concept of DSI

Current model evaluation in Earth sciences prioritizes predictive accuracy (e.g. R^2 and RMSE) while ignoring a model's ability to capture and explain inherent spatial

characteristics such as autocorrelation and heterogeneity. To address this gap, we introduce the concept of spatial interpretability, defined here as a model's capability, via its computational process, to effectively represent and explain intrinsic spatial patterns within data. We propose the DSI as a quantitative indicator of spatial interpretability. A spatial pattern in a dataset is a phenomenon to be explained, and a highly interpretable model can elucidate the causes of this pattern through its variables and algorithmic structure. The DSI directly quantifies the success of this 'explanation' by measuring the attenuation of the spatial pattern from the original data to the model's residuals. The rationale for DSI is grounded in the established principle of using residual diagnostics to assess model misspecification in spatial analysis. This principle, foundational to spatial statistics (Anselin 1988), posits that any systematic spatial patterns in the dependent variable (such as spatial autocorrelation or non-stationarity) that are not accounted for by a model will be transferred to its residuals. For instance, the spatially local complexity or geocomplexity pattern explains a significant portion of spatial model errors (Zhang *et al.* 2023). Meanwhile, models that effectively capture spatial characteristics produce residuals with significantly less spatial structure (e.g. Song *et al.* 2020). Therefore, DSI essentially quantifies the model's spatial explanatory power by measuring the attenuation of original spatial characteristics, aligning with the principle that unexplained variance often manifests as spatially structured residuals. Consequently, DSI complements traditional accuracy metrics, aiding in selecting models suitable for specific spatial analysis tasks across fields like ecological modeling and geography, and is adaptable to various spatial characteristics and model types (e.g. linear regression, random forests (rf) and support vector machines (SVMs)).

2.2. Calculation process of DSI

The process of using DSI to validate spatial models includes two primary steps as shown in Figure 1. The first step is to obtain the DSI for modeling individual spatial characteristics (η) by calculating the spatial characteristics of the dependent variable and the model residuals. The second step is to compute the DSI for modeling multiple spatial characteristics (θ), including the least explanatory power, the most probable interpretability and the maximum capability to represent spatial characteristics.

2.2.1. DSI for individual spatial characteristics (η)

Various indicators are used to describe spatial characteristics, such as spatial autocorrelation (the correlation among values at nearby locations, reflecting Tobler's first law; Tobler 1970) and spatial heterogeneity (the variation in characteristics or relationships across space, also known as non-stationarity; Anselin 1995). Here, we define the index used to measure a model's ability to explain an individual spatial characteristic as the DSI for individual spatial characteristics (η). The calculation process is shown in the following equation:

$$\eta = \frac{\delta_0 - \delta_r}{\delta_0} \tag{1}$$

where δ_0 is the spatial characteristic measured on the dependent variable data, while δ_r is the same spatial characteristic measured on the model residuals. Residuals, in this

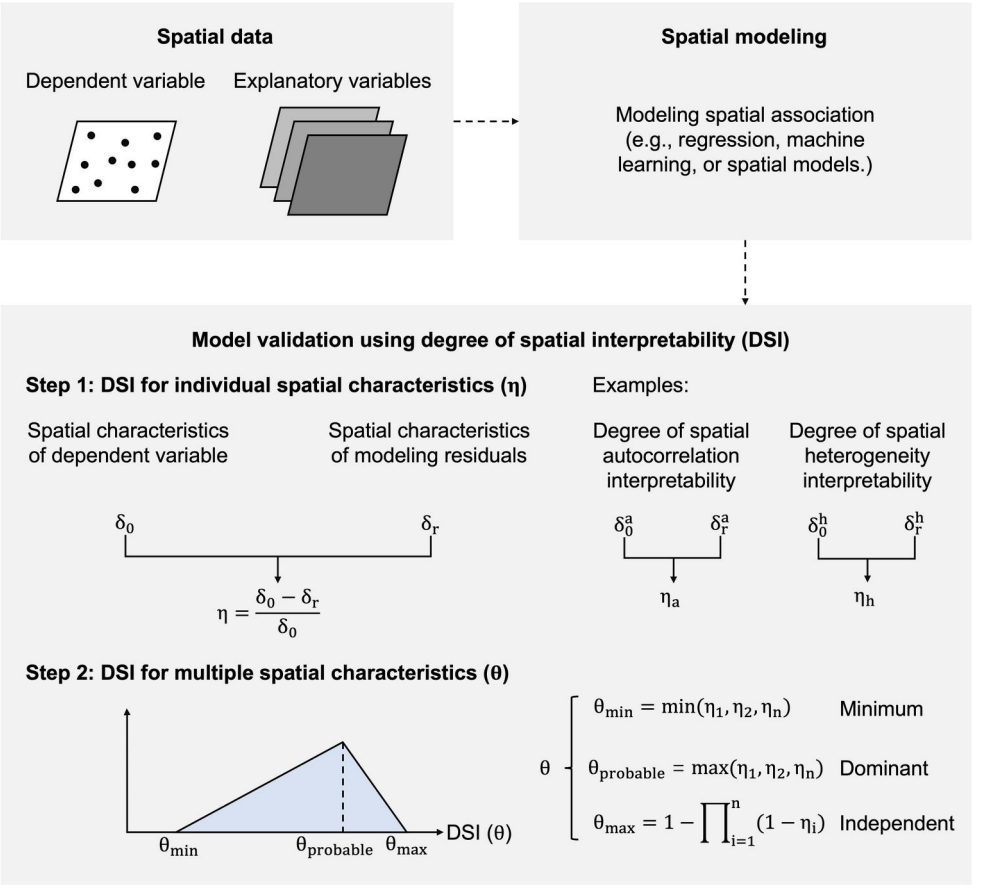

Figure 1. The calculation process of the degree of spatial interpretability (DSI).

context, refer to the differences between the actual observations and the model's predicted values (Anscombe 1961). The comparison between δ_0 and δ_r is central to this index, grounded in residual diagnostics for detecting spatial model misspecification, particularly unaccounted spatial autocorrelation and non-stationarity (Heuvelink *et al.* 1989, Zhang *et al.* 2023). This theory assumes that spatial structures (such as autocorrelation or heterogeneity) present in the dependent variable (δ_0) that are not adequately captured by the model will persist in the residuals (δ_r). Therefore, examining the spatial characteristics of residuals reveals how well the model captures spatial patterns and potential sources of spatial error. The difference, $\delta_0 - \delta_r$, consequently reflects the extent to which the model has accounted for the original spatial characteristic. Dividing this difference by the initial characteristic of the response variable, δ_0, yields $\eta = \frac{\delta_0 - \delta_r}{\delta_0}$. This normalized value, η, quantifies the model's specific spatial characteristic interpretability for the feature being assessed. The closer the value of η is to 1, the better the model performs in explaining that individual spatial characteristic.

For example, to obtain a model's DSI for spatial autocorrelation, by calculating the spatial autocorrelation value (Moran's *I*) of the dependent variable, denoted as δ_0^a, and the spatial autocorrelation value of the model residuals, δ_r^a, if the residuals' Moran's *I*

value is not significant, this indicates that the residuals are randomly distributed, and the model has explained most of the spatial autocorrelation in the data. Conversely, if the residuals' Moran's *I* value is significant, it suggests that spatial autocorrelation characteristics still exist within the residuals, meaning the model has not fully captured this spatial structure. By applying Equation (1), the model's DSI for individual spatial characteristics η_a can be obtained: for example, if η_a equals 49%, it indicates that the model explains 49% of the spatial autocorrelation. Similarly, the spatial heterogeneity value (*Q* value) of the dependent variable, δ_0^h, and the corresponding residual *Q* value of the model, δ_r^h, can be determined, and η_h can be calculated accordingly.

2.2.2. DSI for multiple spatial characteristics (θ)

To comprehensively interpret complex geographical phenomena, which are often defined by multiple, co-existing spatial characteristics, relying on a single spatial metric is often inadequate. Therefore, the concept of DSI for multiple spatial characteristics (θ) is introduced, which integrates the results from multiple individual DSI measures. Crucially, this framework does not simply 'mix' or average these values. Instead, it establishes a diagnostic framework that provides a performance range (e.g. minimum and maximum contributions), thereby revealing a model's distinct strengths and weaknesses without making prior assumptions about the interrelationship between the chosen spatial metrics. This multi-faceted approach offers a more holistic perspective on model performance, aiding researchers in understanding complex model behaviors and providing a more robust scientific basis for model selection and optimization.

The DSI for multiple spatial characteristics (θ) comprises three components: minimum DSI (θ_{min}), which quantifies the least shared explanatory power when multiple spatial characteristics contain overlapping information; most probable DSI (θ_{probable}), which captures the interpretability when a dominant characteristic exists among the multiple spatial characteristics; and maximum DSI (θ_{max}), which measures the interpretability when spatial characteristics are independent. The DSI (θ) for multiple spatial characteristics is calculated using the following equation:

$$\theta = \begin{cases} \theta_{\text{min}} = \min(\eta_1, \eta_2, \ldots, \eta_n) & \text{minimum} \\ \theta_{\text{probable}} = \max(\eta_1, \eta_2, \ldots, \eta_n) & \text{dominant} \\ \theta_{\text{max}} = 1 - \prod_{i=1}^{n}(1 - \eta_i) & \text{independent} \end{cases} \tag{2}$$

where $\eta_i (i = 1, \ldots, n)$ represents the *i*th DSI for individual spatial characteristics. In the equation of θ_{max}, $(1 - \eta_i)$ represents the portion of the spatial characteristic the model cannot explain and $\prod$ represents the product symbol indicates the multiplication of all $(1 - \eta_i)$ terms. $\prod_{i=1}^{n}(1 - \eta_i)$ reflects the cumulative effect of the geographical spatial characteristics that the model has not explained, and using a product form avoids excessive accumulation that could result from simple addition. Finally, $1 - \prod_{i=1}^{n}(1 - \eta_i)$ represents the model's overall explanatory power for all spatial characteristics. In DSI, the closer the value of θ is to 1, the better the overall explanatory power of the model for multiple geographic spatial characteristics.

In this study, only spatial autocorrelation (η_a) and spatial heterogeneity (η_h) are considered. The explanatory degree of the model for spatial autocorrelation and

spatial heterogeneity is calculated as follows:

$$\theta = \begin{cases} \theta_1 = \min(\eta_a, \eta_h) & (\text{minimum shared contribution}) \\ \theta_2 = \max(\eta_a, \eta_h) & (\text{dominant characteristic}) \\ \theta_3 = 1 - (1 - \eta_a)(1 - \eta_h) & (\text{independent contributions}) \end{cases} \tag{3}$$

where θ_1 is the minimum shared contribution, θ_2 is the interpretability of the dominant characteristic and θ_3 is the interpretability when the spatial characteristics are independent.

3. Case study: assessing spatial prediction models of vascular plant species diversity in Australia using DSI

Vascular plants possess lignified tissues, accounting for about 80% of all plant species (Govaerts *et al*. 2021). Australia hosts over 21,000 species of vascular plants, of which 84% are endemic – a result of prolonged independent evolution (Chapman 2009). However, human disturbance, invasive alien species, inappropriate fire management and climate change are now severely affecting the survival and diversity of Australian plant populations (Broadhurst and Coates 2017). Conducting large-scale research on the driving factors of plant species diversity is essential for maintaining Australia's highly diverse natural ecological environment and socio-economic culture (Broadhurst and Coates 2017). Climate conditions, topographical heterogeneity and hydrology influence plant species diversity, while abiotic factors such as soil and human activities also have significant impacts (Kreft and Jetz 2007, Keil and Chase 2019). This study selects data from over 110,000 plant diversity plots with 19 abiotic explanatory variables and employs 10 representative statistical models for modeling; finally, spatial autocorrelation and spatial heterogeneity are selected to analyze the DSI.

3.1. Plant diversity data

The data on Australian vascular plant species diversity used in this study were sourced from publicly available data compiled by Mokany *et al*. (2022). Except for several small islands distant from the Australian mainland, the dataset encompasses most of Australia's territory and includes 115,083 plant community survey plots from 1900 to 2019. Each plot recorded the complete assemblage of above-ground vascular plant species. In each plot, at least 70% of the recorded plant species are native to Australia, and at least 90% have been identified to the species level; this ensures that the plots are representative. To standardize data from plots of varying sizes, the species–area power relationship model was used ($S = C \times A^z$, with z set to 0.25 as validated) to adjust each plot's data to represent plant species richness within an area of 400 m^2 (Rosenzweig 1995). Figure 2(a) illustrates the distribution and species richness values of all plots used in the experiment. We generated a $0.5^\circ \times 0.5^\circ$ grid in latitude and longitude for computational simplicity. The average species richness of the plots within each grid cell was calculated to represent the vascular plant species richness of that cell, resulting in a total of 2031 grids (1958 grids were used for the experiment, as grids with missing variable values were removed), as shown in Figure 2(b). The

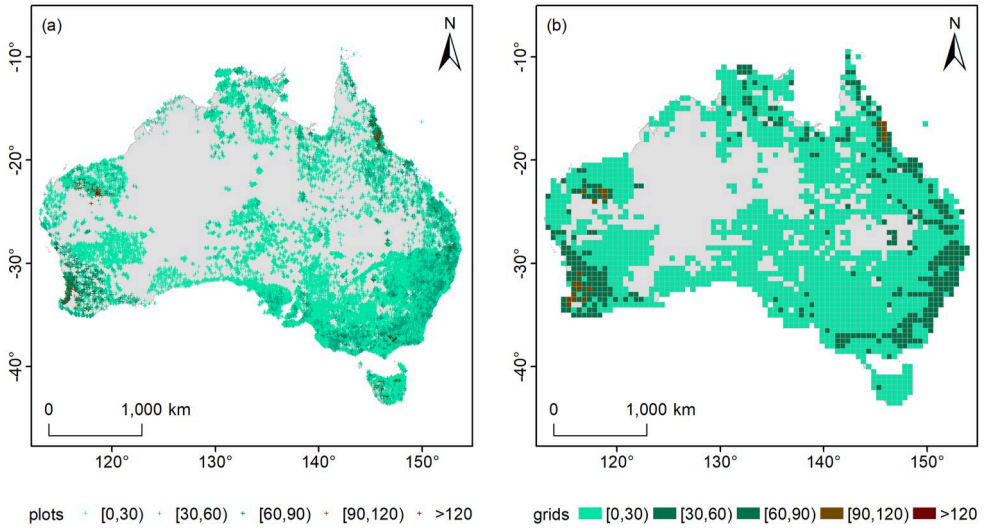

Figure 2. Distribution of vascular plant species diversity in Australia: (a) survey plots and (b) plots aggregated to a 0.5° × 0.5° latitude–longitude grid.

Australian boundary data were obtained from the GDA2020 standard boundary data released by the Australian Bureau of Statistics.

3.2. Explanatory data

This study encompasses five categories of explanatory variables, including climate, soil, topography, hydrology and human activities, with a total of 19 explanatory variables used for modeling, as shown in Figure 3. The climate data comprise four variables: temperature, precipitation, shortwave radiation and wind speed, with a spatial resolution of 5000 m, sourced from the ERA5-Land, CHIRPS and FLDAS datasets, and averaged over the period 2013–2023 (temperature is reported as a mean). The soil data comprise nine variables, including soil depth, bulk density, nutrient element content (total organic carbon, total nitrogen and total phosphorus), pH value, cation exchange capacity and soil type (clay and sand content), derived from the 0 to 5 cm mean soil data of the TERN Landscapes Phase 2 project released in 2024 by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), with a resolution of 90 m, and averaged over the period 1950–2021 (Malone 2022, 2023). The topography data consist of elevation, slope and aspect based on the SRTM dataset with a resolution of 5000 m; the slope and aspect data are calculated from the digital elevation model. The aspect explanatory variable is decomposed into two independent explanatory variables after taking the sine and cosine values for model fitting to offset the problem that the aspect's end-to-end connection leads to discontinuous predictions at the boundary (0° and 360°), which affects the accuracy of the model (Stage 1976, Brenning and Trombotto 2006, Brenning *et al*. 2015). Hydrological and human activity data are based on Australia's 2015 land cover type data, with a resolution of 250 m, released by Digital Earth Australia (DEA) (Lymburner 2015). The variable 'distance to water bodies' considers inland water bodies, salt lakes and oceans as water bodies in

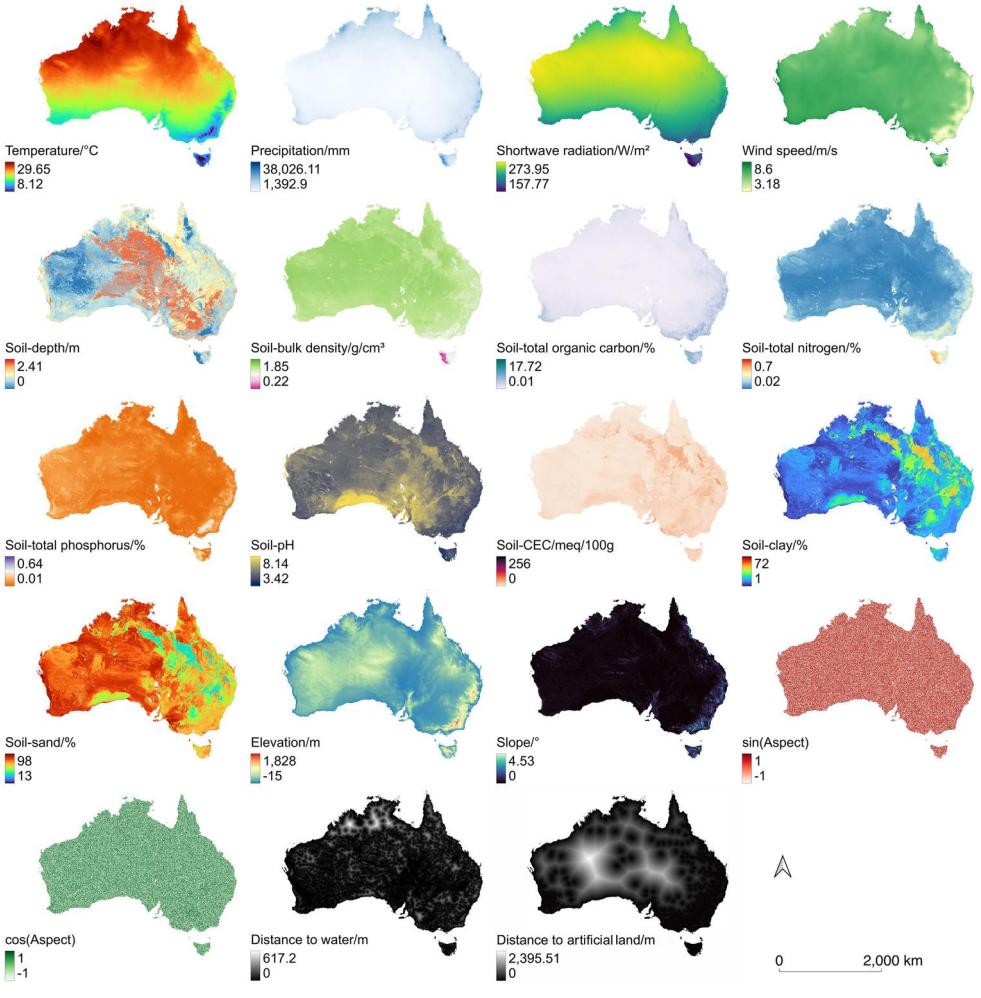

Figure 3. Spatial distributions of explanatory variables for the spatial prediction of species diversity in Australia.

the land cover classification, calculating the distance from each area to these water bodies. Since ocean currents are important factors affecting plant seeds' dispersal and spread and significantly impact plant species diversity, oceans are also considered water bodies (Miryeganeh *et al.* 2014). Urban areas in the classification are regarded as 'artificial land,' and the distance from each location to artificial land is calculated. Distance calculations are based on the GDA94/Australian Albers projection (EPSG:3577). Specific information on each variable is shown in Table 1.

3.3. Experiment design

3.3.1. Data preprocessing

This experiment verifies the feasibility of DSI by modeling the relationship between the diversity of Australian vascular plants and abiotic factors, and the whole

Table 1. A summary of explanatory variables used for the species diversity prediction.

Categories	Variable name	Data description	Spatial resolution	Unit	Source
Climate	Temperature	10-Year average air temperature at 2-m height.	5000 m	°C	ERA5-Land
	Precipitation	10-Year total precipitation.	5000 m	mm	CHIRPS
	Shortwave radiation	10-Year average shortwave radiation.	5000 m	W/m^2	FLDAS
	Wind speed	Average wind speed at 10 m height.	5000 m	m/s	FLDAS
Soil	Depth	Vertical distance from surface to impermeable layer.	90 m	m	CSIRO
	Bulk density	0–5 cm average soil bulk density.	90 m	g/cm^3	CSIRO
	Organic carbon	0–5 cm soil organic carbon content.	90 m	%	CSIRO
	Total nitrogen	0–5 cm soil total nitrogen content.	90 m	%	CSIRO
	Total phosphorus	0–5 cm soil total phosphorus content.	90 m	%	CSIRO
	pH	0–5 cm soil average pH.	90 m		CSIRO
	Cation exchange capacity	0–5 cm soil average cation exchange capacity.	90 m	mequiv./100 g	CSIRO
	Clay content	0–5 cm soil average clay content.	90 m	%	CSIRO
	Sand content	0–5 cm soil average sand content.	90 m	%	CSIRO
Topography	Elevation	Vertical height above sea level.	5000 m	m	SRTM
	Slope	The degree of inclination of the earth's surface relative to the horizontal.	5000 m	°	SRTM
	sin (aspect)	Take the sine of the slope direction relative to north.	5000 m	°	SRTM
	cos (aspect)	Take the cosine of the slope direction relative to north.	5000 m	°	SRTM
Hydrology	Distance to water	Average distance to water.	250 m	m	DEA
Human activities	Distance to artificial land	Average distance to artificial land.	250 m	m	DEA

experiment contains four main steps, as shown in Figure 4. First, the data were prepared; the specific data are shown in Sections 3.1 and 3.2; the final dataset consists of the mean vascular plant diversity value as the dependent variable and the mean values of 19 explanatory variables for regions covered by 1958 $0.5^\circ \times 0.5^\circ$ latitude-longitude grids. The following from Sections 3.3.2–3.3.4 correspond to steps 2–4 in Figure 4.

3.3.2. Modeling of biodiversity

This study employed seven categories, encompassing 10 standard data training and testing models, as shown in Table 2. The caret package in R was used to invoke various functions for computation and to manage the modeling workflow (Kutner *et al.* 2005, Kuhn 2008). Specifically, stratified random partitioning was applied to split the data to ensure a consistent distribution of the dependent variable (richness) in both the training and testing sets to avoid sampling bias (Kuhn 2008). Seventy percent of the data (1372 grids) were allocated for training and 30% (586 grids) for testing. Model performance during the training phase was evaluated using fivefold cross-

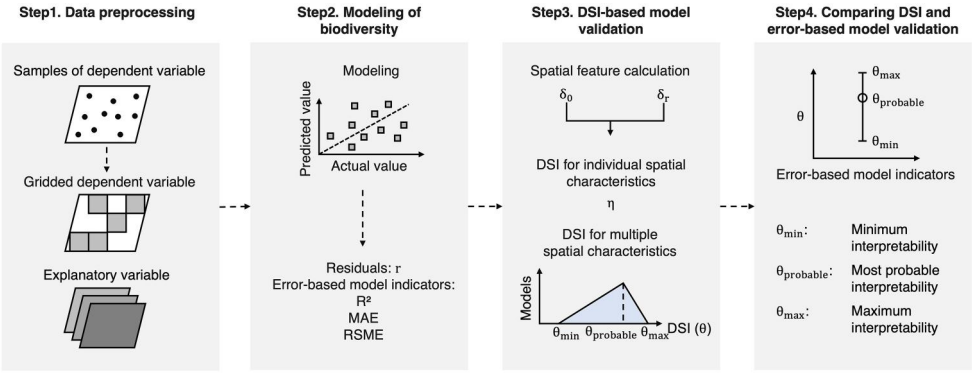

Figure 4. Flowchart of DSI calculation and indicator comparison.

Table 2. Information on models used to fit Australian vascular plant diversity to abiotic factors.

Type	Model name	Model code
Linear models	Linear regression	lm
Tree-based models	Random forest	rf
	Cubist model	cubist
	Extreme gradient boosting (XGBoost)	xgbTree
	Gradient boosting machine (GBM)	gbm
Neighborhood-based models	k-Nearest neighbors (k-NNs)	knn
Non-linear and adaptive regression models	Generalized additive model (GAM)	gam
	Multivariate adaptive regression splines (MARS)	earth
Dimensionality reduction regression models	Partial least squares regression (PLS)	pls
Support vector machine-based models	Support vector machine (SVM)	svmRadial

validation (fivefold CV). The built-in default hyperparameter settings provided by the caret package for each model type were utilized. The function codes referenced within the caret package for these models are listed in the third column of Table 2, and these codes are used in the text to denote the corresponding models. The models' fitting performance was subsequently measured on both the training and final test sets using R^2 (coefficient of determination), RMSE and MAE (Nussbaum *et al*. 2018, Karunasingha 2022). Each model produces two sets of residuals (for the training set and the test set).

3.3.3. DSI-based model validation

This study uses spatial autocorrelation and spatial heterogeneity as metrics for spatial analysis of DSI. Moran's *I* index is a commonly used spatial autocorrelation metric that determines whether the values of a variable exhibit clustering in geographic space. Moran's *I* index ranges from −1 to 1, with larger absolute values indicating stronger spatial autocorrelation (Moran 1950, Miller 2004). In this study, to compute Moran's *I*, we used the spdep package in R. A spatial weights matrix was constructed based on the K-nearest neighbors (KNNs) method, with $k = 8$. The analysis was performed on the data's projected coordinates (GDA94/Australian Albers, EPSG:3577) to ensure accurate distance calculations. Here, we will obtain the spatial autocorrelation values of the dependent variables Moran's *I* (δ_0^a), the spatial autocorrelation value of the residuals of each model (δ_r^a), and η_a of each model.

Spatial heterogeneity refers to local differences or uneven distributions of geographic phenomena in space and is typically measured using the geographical detector model (Anselin 1995). The optimal parameters-based geographical detector is a parameter-optimized geographical detector model that adds a parameter optimization process to the traditional model, providing more reliable and accurate spatial heterogeneity analysis results in a broader range of application scenarios (Song et al. 2021). The geographical detector value (Q value) ranges from 0 to 1; a larger Q value indicates more robust spatial heterogeneity of the variable (Hengl et al. 2018). In this study, we use the explanatory variables as the basis of stratification to calculate the Q value of the corresponding samples and residuals, and obtain δ_0^h, δ_r^h and η_h.

According to Equation (3), based on η_a and η_h, the DSI for multiple spatial characteristics (θ) can be calculated for each model to obtain three metrics (θ_1, θ_2 and θ_3). To compare the difference of θ between different models in terms of the trend of change between θ_1, θ_2 and θ_3 in one model, the θ_1, θ_2 and θ_3 of each model were arranged in order of size on the axes, and the different models were arranged in order of size from top to bottom according to θ_3, to construct the θ comparison figure.

3.3.4. Comparing DSI-based and error-based model validation

Common error metrics, such as R^2, RMSE and MAE, primarily assess model fit and predictive accuracy. While R^2 is interpretable and standardized, it's sensitive to the number of variables (risking overfitting) and its squaring obscures error sign, meaning high R^2 does not guarantee unbiased predictions (James *et al.* 2013). RMSE effectively highlights large errors, but its squaring term also amplifies outlier impacts, potentially distorting perceived performance (Montgomery *et al.* 2021, Ren *et al.* 2025). Conversely, MAE, using absolute errors, is robust to outliers and directly measures average error, though it gives less weight to large errors (Hyndman and Koehler 2006).

To make a visual comparison between the DSI-based indicators and the error-based indicators, a coordinate system with R^2, RMSE or MAE as the horizontal coordinate and θ-value as the vertical coordinate was created, and the θ_1, θ_2 and θ_3 values of each model were connected, thus comparing the trends and differences of the different models in the two indicator systems.

4. Results

4.1. Accuracy metrics-based model validation

Figure 5 shows the fitting scatter plots and accuracy metrics results of 10 models on the training and testing sets. The results indicate that the rf and cubist models achieved the best overall performance on the training and testing datasets, with high R^2 values and well-controlled prediction errors. The svmRadial, gradient boosting machine (gbm), gam and xgbTree models also demonstrated strong generalization ability, with prediction errors within an acceptable range. In contrast, the earth, lm, knn and pls models exhibited poor R^2 and error metrics, particularly predictive performance on the testing set.

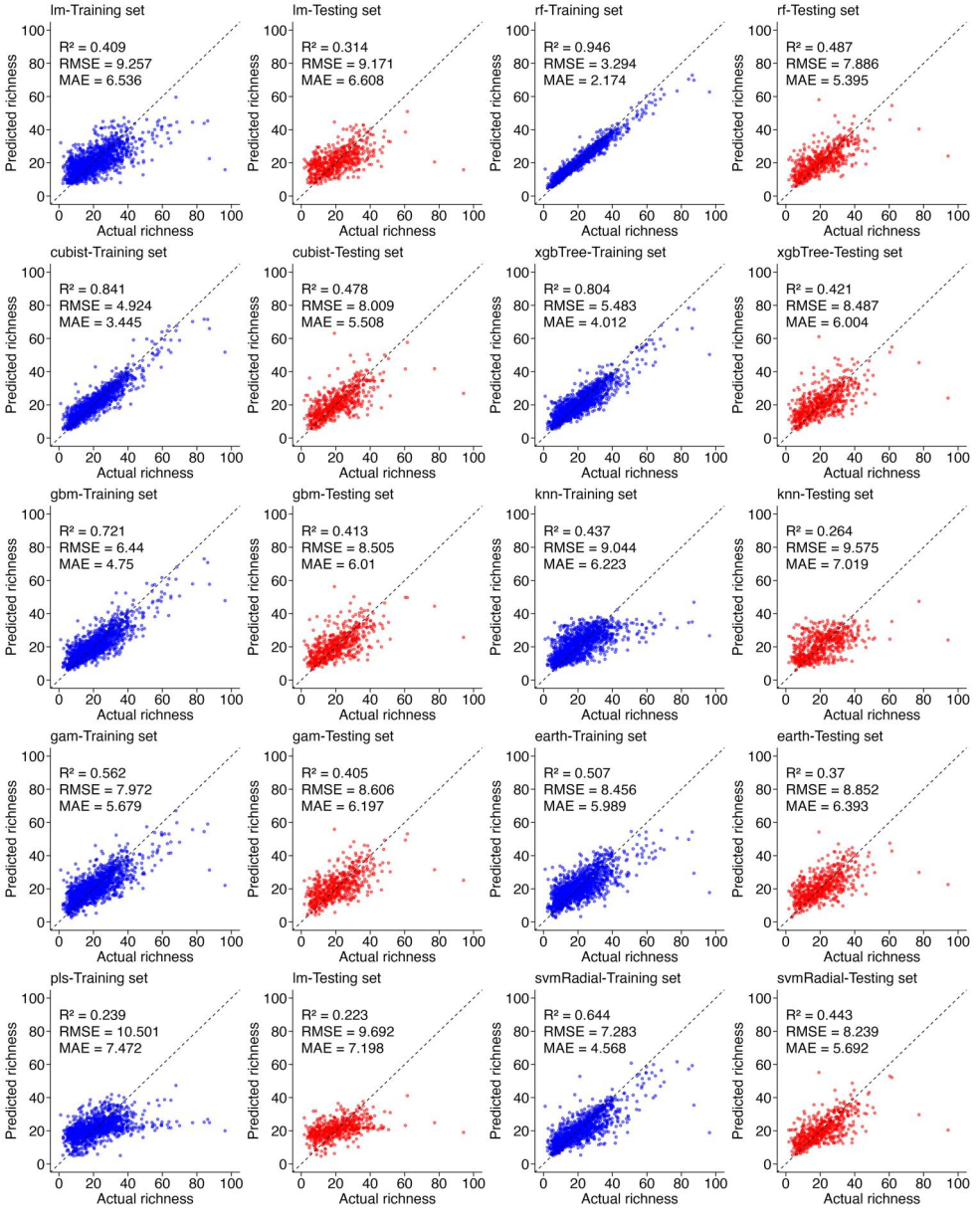

Figure 5. Scatterplots of training set and testing set fits to different models of vascular plant richness and abiotic factors in Australia.

4.2. DSI-based model validation

4.2.1. Spatial characteristics of data and modeling residuals

Table 3 lists the measured spatial characteristics (δ_0 and δ_r) and the resulting DSI values for individual spatial features (η_a and η_h) for each model. These metrics were calculated using the test set to better reflect generalization ability. The measured spatial characteristics (δ_0 and δ_r) are statistically significant ($p < 0.01$).

Table 3. Geospatial indicator values and DSI for individual spatial characteristics (η) values for each model.

Model	Spatial autocorrelation measured by Moran's I		Degree of spatial autocorrelation interpretability: η_a	Spatial heterogeneity measured by Q value		Degree of spatial heterogeneity interpretability: η_h
	δ_0^a	δ_r^a		δ_0^h	δ_r^h	
lm	0.409**	0.189**	0.538	0.513**	0.203**	0.604
rf		0.039**	0.905		0.127**	0.752
cubist		0.027**	0.934		0.103**	0.799
xgbTree		0.048**	0.883		0.095**	0.815
gbm		0.078**	0.809		0.122**	0.762
knn		0.143**	0.650		0.186**	0.637
gam		0.095**	0.768		0.133**	0.741
earth		0.142**	0.653		0.164**	0.680
pls		0.272**	0.335		0.274**	0.466
svmRadial		0.059**	0.856		0.157**	0.694

**$p < 0.01$ (statistically significant).

The results show significant spatial autocorrelation within the test set samples (Moran's $I = 0.409$). The cubist model performs best, exhibiting the lowest δ_r^a value (0.027), indicating minimal remaining autocorrelation. Models like rf, svmRadial, xgbTree, gbm and gam also perform well, showing relatively low δ_r^a. Conversely, the earth, knn, lm and pls models struggle to account for spatial autocorrelation, exhibiting the highest levels of δ_r^a.

Regarding spatial heterogeneity, the test set samples exhibit a high degree ($Q = 0.513$). Evaluating residual spatial heterogeneity (δ_r^h) shows the xgbTree model performs best, achieving the lowest δ_r^h value (0.095). The cubist, gbm, rf and gam models also have relatively low δ_r^h. In contrast, the svmRadial, earth, knn, lm and pls models perform poorly.

4.2.2. DSI for individual spatial characteristics (η)

DSI for individual spatial characteristics (η) quantifies the degree to which a model explains geospatial indicators, reflecting the model's ability to capture geospatial characteristics. A high η value indicates that the model better captures the spatial characteristics of the data and has higher reliability and predictive capability.

According to the test results (shown in Table 3), the cubist model performs the best, with high explanatory ability under different metrics, with an explanatory degree ranging from 0.799 to 0.934 and an average value of 0.875. The rf, xgbTree, gbm, svmRadial and gam models showed high and relatively stable explanatory power in the test set, with values that did not change much under different metrics, showing good generalization. Their explanatory power ranges from 0.694 to 0.910, with an average value of 0.794. The performance of the rf and svmRadial models varies more under different indicators, with more pronounced fluctuations in the degree of explanation. The earth, knn, lm and pls models perform poorly under all indicators and have lower explanatory power. Their explanatory degrees range from 0.343 to 0.680, with a mean value of 0.576.

4.2.3. DSI for multiple spatial characteristics (θ)

DSI for multiple spatial characteristics (θ) synthesizes DSI for individual spatial characteristics (η) to provide an overall evaluation of geospatial interpretability. Figure 6

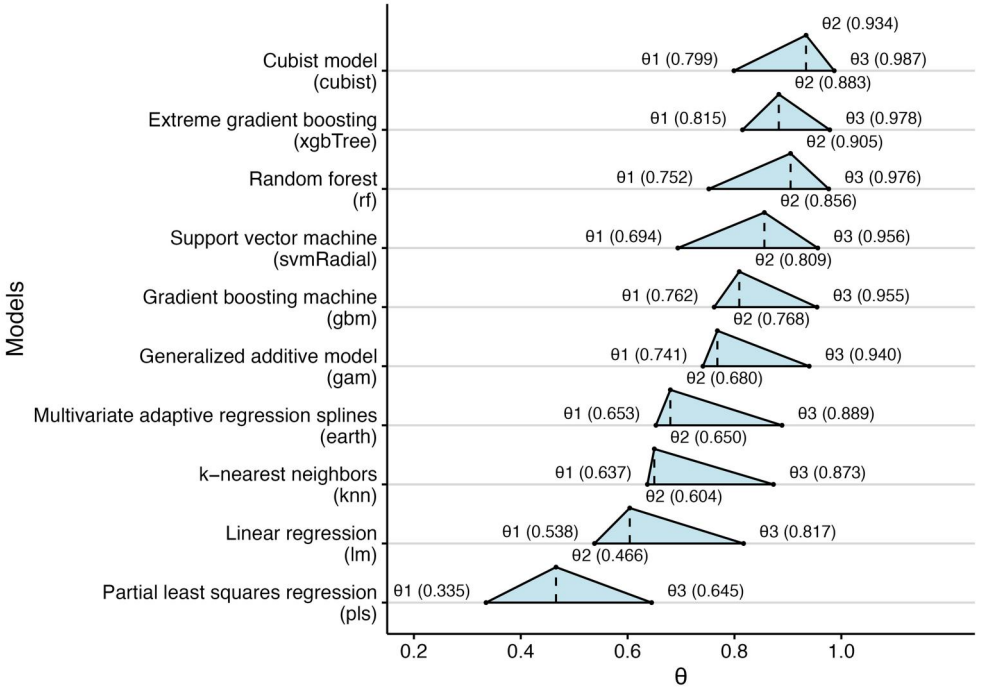

Figure 6. Distribution of DSI for multiple spatial characteristics (θ) for different models.

displays the distribution of θ for different models and specific values of the θ_1, θ_2 and θ_3 in the figure. The dashed line indicates the projected horizontal axis position of the θ_2 value, and the vertical displacement is only for visual distinction purposes.

Results demonstrate that DSI effectively identifies the interpretability of different machine learning models used in geospatial modelling with respect to multiple spatial characteristics. For instance, the cubist model performs best among all models in terms of θ, with DSI values ranging from 0.799 to 0.987 and a most probable DSI of 0.934, indicating a strong capability in explaining spatial characteristics. The xgbTree, rf, svmRadial, gbm and gam models exhibit relatively high overall interpretabilities, with θ_1 between 0.694 (svmRadial) and 0.815 (xgbTree), θ_2 between 0.768 (gam) and 0.905 (rf), and θ_3 between 0.940 (gam) and 0.978 (rf), suggesting robustness in capturing both spatial autocorrelation and heterogeneity. In contrast, the earth, knn, lm and pls models show lower overall interpretability, with θ_1 between 0.335 (pls) and 0.653 (earth), θ_2 between 0.466 (pls) and 0.680 (earth), and θ_3 between 0.645 (pls) and 0.889 (earth), lagging considerably across all θ metrics. This highlights the effectiveness of the DSI framework in differentiating models based on their ability to account for intrinsic spatial structures.

4.3. Comparison between DSI and accuracy metrics-based model validation

Most current spatial modelling assumes that models with high goodness of fit and low prediction errors, as assessed by traditional performance evaluation metrics, also possess high spatial interpretability. However, models that achieve high prediction

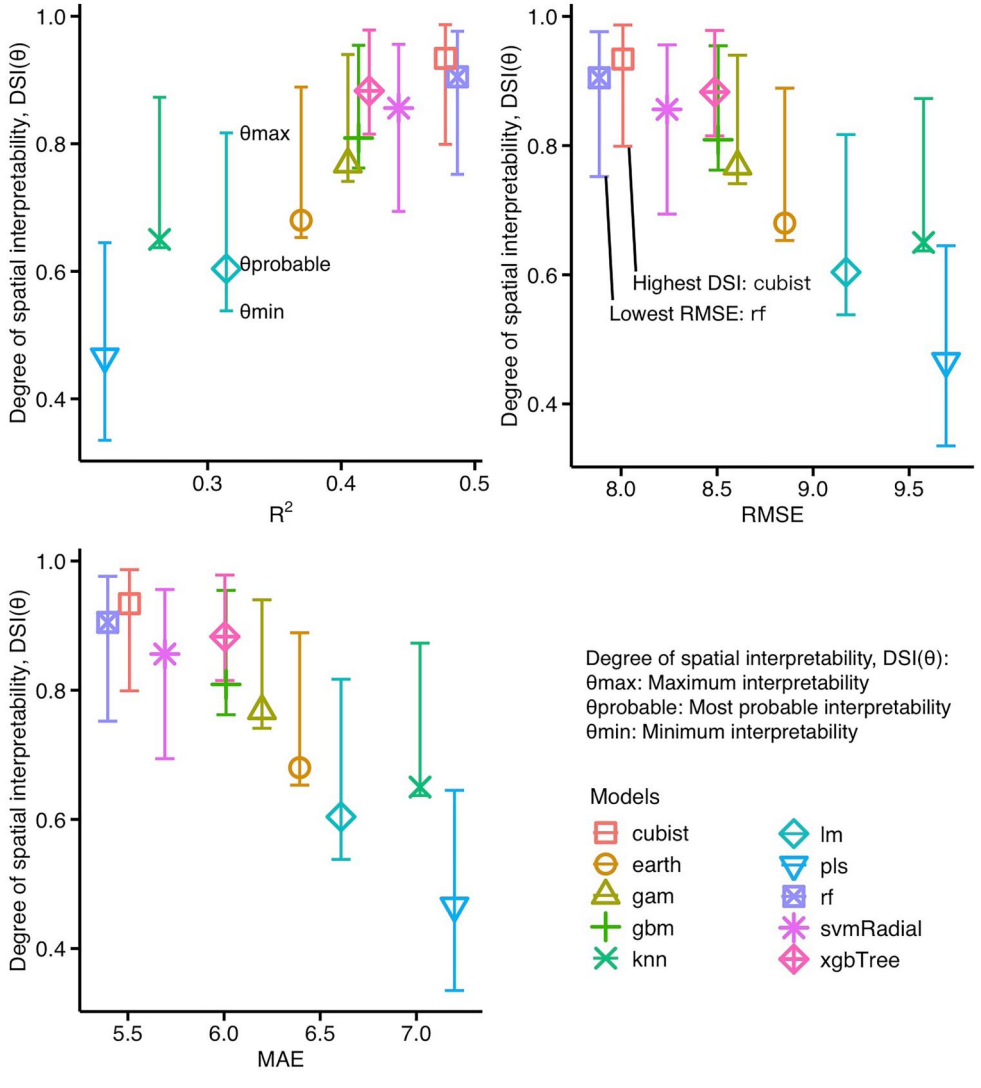

Figure 7. Comparison of DSI and accuracy metrics-based model validation for different models.

accuracy may still have limited ability to explain underlying spatial characteristics. The inconsistency between accuracy and spatial interpretability represents a key blind spot of traditional metrics and demonstrates the growing need for DSI in practice. The distinct contribution of DSI is to seperate a model's capacity for meaningful spatial explanation from the prediction accuracy by identifying cases where prediction accuracy and spatial explanation power differ, although DSI is generally correlated with traditional accuracy metrics. Figure 7 shows the distributions of traditional metrics and DSI for the models in testing set of model validation, where the upper, middle and lower nodes of each model in the plot represent θ_3, θ_2 and θ_1, respectively. The validation results show that the rf model ($R^2 = 0.487$, RMSE $= 7.886$, MAE $= 5.395$, $\theta_1 = 0.752$, $\theta_2 = 0.905$, $\theta_3 = 0.976$) has a better fit than the cubist model ($R^2 = 0.478$, RMSE $= 8.009$, MAE $= 5.508$, $\theta_1 = 0.799$, $\theta_2 = 0.934$, $\theta_3 = 0.987$), but its DSI is lower.

Comparing the knn model ($R^2 = 0.264$, RMSE = 9.575, MAE = 7.019, $\theta_1 = 0.637$, $\theta_2 = 0.650$, $\theta_3 = 0.873$) and the lm model ($R^2 = 0.314$, RMSE = 9.171, MAE = 6.608, $\theta_1 = 0.538$, $\theta_2 = 0.604$, $\theta_3 = 0.817$) also confirms that the fit of a model does not always represent its DSI.

Therefore, introducing DSI in model evaluation is increasingly required. In this study, some models perform well according to traditional accuracy metrics such as R^2, RMSE and MAE, but show relatively low DSI. This type of accuracy-DSI inconsistency may arise because the model does not fully account for or explore spatial characteristics, such as spatial autocorrelation and heterogeneity (Hengl 2018). In contrast, some models achieve higher DSI, indicating a stronger ability to capture spatial characteristics. Therefore, in practical applications, model selection should be guided by specific spatial analysis objectives through a joint consideration of prediction accuracy and spatial interpretability.

5. Discussion

5.1. Validation of DSI effectiveness using simulation data

To evaluate the diagnostic efficacy of the DSI metric, we conducted a simulation experiment under controlled conditions. This approach avoids the uncertainty of underlying spatial processes in real-world ecological data. Specifically, we generated a synthetic dataset on a 30 × 30 regular grid comprising 900 observation points. The dependent variable was constructed via a predefined linear equation containing variables with strong spatial autocorrelation and non-linear interactions, where the error term was intentionally designed to retain spatial dependence (Zurell *et al*. 2010).

The results highlight a distinct contrast between models with similar predictive accuracy. Both the rf and gbm models achieved high goodness-of-fit (R^2 of 0.873 and 0.877, respectively). However, their spatial interpretability differed significantly: rf achieved a higher DSI ($\theta_{probable} = 0.948$) compared to gbm ($\theta_{probable} = 0.906$). This analysis reveals a subtle difference undetectable by R^2 alone. Although the gbm model achieved high predictive accuracy, it failed to capture one of the spatial characteristics ($\theta_{min} = 0.023$), resulting in spatially clustered residuals. In contrast, rf demonstrated a balanced capability in explaining both variance and spatial structures ($\theta_{min} = 0.151$). Consequently, DSI serves as an independent validation tool alerting researchers when models achieve high accuracy but fail to resolve intrinsic spatial processes (Bahn and McGill 2013, Fourcade *et al*. 2018) (Figure 8, Table 4).

5.2. Impact of validation strategy on DSI

To investigate the impact of validation strategies on DSI assessment, we compared standard cross-validation (CV) with spatial blocking methods. Standard random sampling usually makes spatial models have relatively high accuracy due to the impacts of spatial autocorrelation of data, but it has challenges in telling if the model has learned the true patterns or is just relying on nearby data (Legendre 1993, Roberts *et al*. 2017). Specifically, we evaluated the rf model using two validation strategies: standard

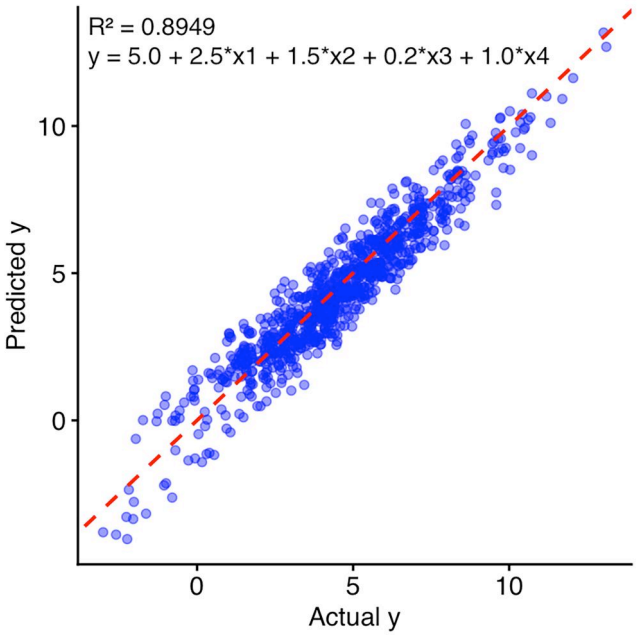

Figure 8. Scatter plot of predicted and actual values for the simulated dataset used to validate the DSI metric.

Table 4. Calculation of DSI for spatial autocorrelation and spatial heterogeneity of rf and gbm models based on simulated data.

Models	δ_0^a	δ_r^a	δ_0^h	δ_r^h	DSI ($\theta_{probable}[\theta_{min}, \theta_{max}]$)
rf	0.523**	0.444**	0.827**	0.043**	0.948 [0.151, 0.956]
gbm	0.523**	0.511**	0.827**	0.078**	0.906 [0.023, 0.908]

**$p < 0.01$ (statistically significant).

Table 5. Comparison of DSI performance under cross-validation (CV) and spatial block cross-validation (BCV) to assess sensitivity to validation strategy.

Validation strategy	η_a	η_h	DSI ($\theta_{probable}[\theta_{min}, \theta_{max}]$)
CV	0.905	0.752	0.905 [0.752, 0.976]
BCV	0.874	0.716	0.874 [0.716, 0.964]
Difference (Δ)[a]	3.43%	4.79%	$\Delta(\theta_{probable}) = 3.43\%$, $\Delta(\theta_{min}) = 4.79\%$, $\Delta(\theta_{max}) = 1.25\%$

[a]The difference used to assess DSI's sensitivity to validation strategy is calculated as: $\Delta = (\eta(BCV) - \eta(CV))/\eta(CV)$ or $\Delta = (\theta(BCV) - \theta(CV))/\theta(CV)$.

K-fold CV and spatial block cross-validation (BCV) via the blockCV package (the block size was set to 20 km) (Valavi *et al.* 2018).

The results in Table 5 show that the DSI of rf model maintains robust spatial interpretability under both CV and BCV validation strategies. Under standard CV, the DSI is 0.905 [0.752, 0.976], comprising a spatial autocorrelation interpretability of 0.905 and a spatial heterogeneity interpretability of 0.752. Under BCV, the DSI is 0.874 [0.716, 0.964], with spatial autocorrelation and heterogeneity interpretability values of 0.874 and 0.716. The slightly lower BCV-based DSI indicates that spatial autocorrelation

and local pattern continuity among neighboring samples can influence validation outcomes. All sensitivity indicators, including the differences in overall DSI and in its spatial dependence and heterogeneity components, remain below 5%, showing that DSI provides a robust and reliable measure of spatial interpretability when different spatial validation strategies are applied.

6. Conclusions

This study developed the DSI to address a methodological gap in current model evaluation systems regarding the assessment of spatial characteristics. Traditional evaluation metrics focus primarily on fit and predictive accuracy, overlooking a model's ability to interpret spatial characteristics. The introduction of the DSI provides a more comprehensive evaluation tool for assessing models in geographic information science, enhancing the scientific rigor of model selection. This advancement is particularly beneficial for geospatial analysis and ecological modeling, enabling more accurate handling of complex geographic phenomena. It is essential to recognize that relying solely on traditional metrics when dealing with data exhibiting significant spatial characteristics can lead to misleading conclusions. Incorporating the spatial interpretability metric, DSI, improves research findings' reliability and practical applicability. However, the interactions between multiple spatial characteristics are very complex, and how to define and accurately calculate these interactions remains an area for further exploration. Understanding these intricate relationships can deepen our understanding of geographic phenomena more precisely, contributing to the further development of spatial model assessment systems and is an essential direction for future research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Acknowledgements

We thank Karel Mokany for providing the data support, and the reviewers for their constructive suggestions.

Notes on contributors

Haiyang Liu is a former Research Assistant at The Hong Kong Polytechnic University and is currently a PhD student at Hokkaido University. His research focuses on ecology and spatial methods. He was responsible for data processing, manuscript drafting and revisions for this paper (haiyang.liu.q6@elms.hokudai.ac.jp).

Yongze Song is an Associate Professor at Curtin University. His research interests include spatial methods, geospatial intelligence and sustainable infrastructure. He serves as an Associate Editor for *International Journal of Applied Earth Observation and Geoinformation*, and *GIScience & Remote Sensing*. He led the study design and methodology development (Yongze.Song@curtin.edu.au).

Wen Yi is an Associate Professor at The Hong Kong Polytechnic University. Her research interests include optimization in construction management, construction safety and health, and

construction engineering and management. She contributed to manuscript revisions and data processing (wen.yi@polyu.edu.hk).

ORCID

Yongze Song http://orcid.org/0000-0003-3420-9622

Data and codes availability statement

Data and codes supporting the findings of this study are available at https://doi.org/10.6084/m9.figshare.27959079.

References

Anscombe, F.J., 1961. Examination of residuals. *In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, Volume 1: contributions to the theory of statistics*. Berkeley and Los Angeles, CA: University of California Press, 1–36.

Anselin, L., 1988. *Spatial econometrics: methods and models*. Dordrecht, The Netherlands: Springer.

Anselin, L., 1995. Local indicators of spatial association—LISA. *Geographical Analysis*, 27 (2), 93–115.

Augusiak, J., Van den Brink, P.J., and Grimm, V., 2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecological Modelling*, 280, 117–128.

Bahn, V. and McGill, B.J., 2013. Testing the predictive performance of distribution models. *Oikos*, 122 (3), 321–331.

Beale, C.M., *et al.*, 2010. Regression analysis of spatial data. *Ecology Letters*, 13 (2), 246–264.

Bennett, N.D., *et al.*, 2013. Characterising performance of environmental models. *Environmental Modelling & Software*, 40, 1–20.

Bini, L.M., *et al.*, 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, 32 (2), 193–204.

Brenning, A. and Trombotto, D., 2006. Logistic regression modeling of rock glacier and glacier distribution: topographic and climatic controls in the semi-arid Andes. *Geomorphology*, 81 (1–2), 141–154.

Brenning, A., *et al.*, 2015. Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Natural Hazards and Earth System Sciences*, 15 (1), 45–57.

Broadhurst, L. and Coates, D., 2017. Plant conservation in Australia: current directions and future challenges. *Plant Diversity*, 39 (6), 348–356.

Chapman, A.D., 2009, September. *Numbers of living species in Australia and the world*. 2nd ed. A report for the Australian Biological Resources Study. Canberra, Australia: Australian Biological Resources Study (ABRS).

Dormann, C.F., *et al.*, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30 (5), 609–628.

Elith, J. and Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40 (1), 677–697.

Fourcade, Y., Besnard, A.G., and Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27 (2), 245–256.

Govaerts, R., *et al.*, 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, 8 (1), 215.

Guisan, A. and Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8 (9), 993–1009.
Hengl, T., *et al.*, 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518.
Heuvelink, G.B., Burrough, P.A., and Stein, A., 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, 3 (4), 303–322.
Hyndman, R.J. and Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22 (4), 679–688.
James, G., *et al.*, 2013. *An introduction to statistical learning*. New York: Springer.
Jiang, B., 2007. Some thoughts on geospatial analysis and modeling. *Computers, Environment and Urban Systems*, 31 (5), 477–480.
Jørgensen, S.E., 2008. Overview of the model types available for development of ecological models. *Ecological Modelling*, 215 (1–3), 3–9.
Karunasingha, D.S.K., 2022. Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609–629.
Keil, P. and Chase, J.M., 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, 3 (3), 390–399.
Kreft, H. and Jetz, W., 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (14), 5925–5930.
Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28 (5), 1–26.
Kutner, M.H., *et al.*, 2005. *Applied linear statistical models*. 5th ed. Boston, MA: McGraw-Hill/Irwin.
Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74 (6), 1659–1673.
Longley, P.A., *et al.*, 2015. *Geographic information science and systems*. 4th ed. Hoboken, NJ: John Wiley & Sons.
Lymburner, L., *et al.*, 2015. *Dynamic land cover dataset version 2.1*. Canberra: Geoscience Australia. Available from: https://pid.geoscience.gov.au/dataset/ga/83868.
Malone, B., 2022. *Soil and landscape grid national soil attribute maps – cation exchange capacity (3″ resolution) – release 1. v2* [Data set]. Canberra, ACT, Australia: CSIRO. https://doi.org/10.25919/pkva-gf85.
Malone, B., 2023. *Soil and landscape grid national soil attribute maps – bulk density – whole earth – release 2. v3* [Data set]. Canberra, ACT, Australia: CSIRO. https://doi.org/10.25919/gxyn-pd07.
Miller, H.J., 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94 (2), 284–289.
Miryeganeh, M., *et al.*, 2014. Long-distance dispersal by sea-drifted seeds has maintained the global distribution of *Ipomoea pes-caprae* subsp. *brasiliensis* (Convolvulaceae). *PLOS One*, 9 (4), e91836.
Mokany, K., *et al.*, 2022. Patterns and drivers of plant diversity across Australia. *Ecography*, 2022 (11), e06426.
Montgomery, D.C., Peck, E.A., and Vining, G.G., 2021. *Introduction to linear regression analysis*. 6th ed. Hoboken, NJ: John Wiley & Sons.
Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37 (1–2), 17–23.
Nussbaum, M., *et al.*, 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. *Soil*, 4 (1), 1–22.
Planque, B., *et al.*, 2022. A standard protocol for describing the evaluation of ecological models. *Ecological Modelling*, 471, 110059.
Ren, K., Song, Y., and Yu, Q., 2025. Second-dimension outliers for spatial prediction. *International Journal of Geographical Information Science*, 39, 1–28.
Roberts, D.R., *et al.*, 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40 (8), 913–929.
Rosenzweig, M.L., 1995. *Species diversity in space and time*. Cambridge: Cambridge University Press.

Smulders, M., *et al.*, 2010. A spatially explicit method for evaluating accuracy of species distribution models. *Diversity and Distributions*, 16 (6), 996–1008.

Song, Y., *et al.*, 2020. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. *GIScience & Remote Sensing*, 57 (5), 593–610.

Song, Y., *et al.*, 2021. A spatial heterogeneity-based segmentation model for analyzing road deterioration network data in multi-scale infrastructure systems. *IEEE Transactions on Intelligent Transportation Systems*, 22 (11), 7073–7083.

Song, Y., *et al.*, 2026. GeoAI: beyond mapping earth and cities through explainability, adaptability, and sustainability. *iScience*, 29 (1), 114407.

Stage, A.R., 1976. An expression for the effect of aspect, slope, and habitat type on tree growth. *Forest Science*, 22 (4), 457–460.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (Suppl. 1), 234–240.

Tredennick, A.T., *et al.*, 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 102 (6), e03336.

Valavi, R., *et al.*, 2018. blockCV: an r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10 (2), 225–232.

Zhang, Z., *et al.*, 2023. Geocomplexity explains spatial errors. *International Journal of Geographical Information Science*, 37 (7), 1449–1469.

Zurell, D., *et al.*, 2010. The virtual ecologist approach: simulating data and observers. *Oikos*, 119 (4), 622–635. https://doi.org/10.1111/j.1600-0706.2009.18284.x.