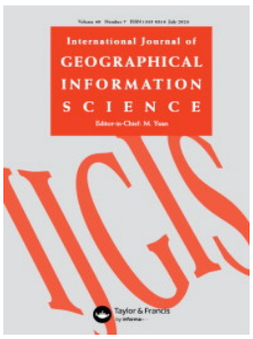

International Journal of Geographical Information Science

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

Spatial outliers as a pattern determinant for explaining heterogeneity

Kai Ren, Yongze Song, Xinyue Yang, Xi Wang, Min Chen & Qiang Yu

To cite this article: Kai Ren, Yongze Song, Xinyue Yang, Xi Wang, Min Chen & Qiang Yu (10 Jun 2026): Spatial outliers as a pattern determinant for explaining heterogeneity, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2026.2682957

To link to this article: https://doi.org/10.1080/13658816.2026.2682957

Published online: 10 Jun 2026.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2026.2682957

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Spatial outliers as a pattern determinant for explaining heterogeneity

Kai Ren[a,b,c,d,e], Yongze Song[e], Xinyue Yang[e], Xi Wang[f,g], Min Chen[a,b,c,d] and Qiang Yu[h]

[a]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, Nanjing, China; [b]Ministry of Education, Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Nanjing, China; [c]School of Geography, Nanjing Normal University, Nanjing, Jiangsu, China; [d]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China; [e]School of Design and the Built Environment, Curtin University, Perth, Australia; [f]Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China; [g]Department of Architecture, National University of Singapore, Singapore, Singapore; [h]State Key Laboratory of Soil and Water Conservation and Desertification Control, Northwest A&F University, Yangling, Shaanxi, China

ABSTRACT

Explaining spatial heterogeneity typically relies on first-dimension covariates that describe spatial gradients. However, many geographic phenomena also exhibit irregular and locally extreme structures that are difficult to capture using smooth relationships. This study proposed a second-dimension outlier-driven heterogeneity (SOH) model, in which the first dimension refers to covariate variation across space, while the second dimension represents spatial pattern information captured from local outlier configurations. SOH derives multi-scale spatial outlier patterns (SOPs) using a second-dimension outlier model and embeds them in a stratification-detection workflow, where decision tree-based stratification defines strata and the geographical detector evaluates explanatory power via the power of determinant (PD). The model supports evaluation of individual effects, SOP interactions, and SOP-variable interactions, with assessment of scale dependence across neighbourhood buffers. Application of this model to spatial heterogeneity in Australian barley production showed that SOPs strengthened heterogeneity explanation relative to original variables, and that SOP interactions and SOP-variable interactions yielded synergistic gains in PD. A scale threshold around 200 km was identified, beyond which SOP-only models approached the explanatory performance of combined models, indicating that multi-scale SOPs captured broad spatial context. Overall, SOH provides a unified approach for incorporating outlier-driven spatial patterns into spatial heterogeneity analysis.

ARTICLE HISTORY
Received 3 February 2026
Accepted 27 May 2026

KEYWORDS
Spatial heterogeneity; spatial outlier patterns; second-dimension outliers; geographical detector; stratification-based analysis

CONTACT Kai Ren kai@njnu.edu.cn; Yongze Song yongze.song@curtin.edu.au

1. Introduction

Spatial heterogeneity is a fundamental characteristic of geographical phenomena, describing the non-uniform distribution of attributes and processes across space (Pickett and Cadenasso 1995, De Marsily *et al.* 2005). Such heterogeneity reflects the combined influences of natural conditions, human activities, and their interactions, and has long been regarded as a central concept in geographic information science (GIS) and spatial analysis (Roy *et al.* 2024). Robust characterisation and explanation of spatial heterogeneity are critical for advancing the understanding of spatial processes, enhancing spatial prediction, and informing evidence-based decision-making in domains such as environmental management, land-use planning, and resource allocation (Young *et al.* 2001; Vinatier *et al.* 2011). As spatial datasets continue to increase in volume, spatial resolution, and thematic richness, the interpretation of complex and heterogeneous spatial patterns has become a key methodological challenge (Shu *et al.* 2019).

Existing approaches to explaining spatial heterogeneity predominantly rely on first-dimension explanatory variables, namely the original covariates and their systematic variation in the attribute domain, typically representing broad environmental, biophysical, or socio-economic conditions across space (Liu *et al.* 2026; Turner and Chapin III, 2005). These variables are typically incorporated into statistical or machine learning models to explain spatial variation in a dependent variable through global or locally varying relationships (Wagner and Fortin 2005, Zhang *et al.* 2023). Numerous studies have demonstrated the importance of such factors in shaping spatial heterogeneity across a wide range of applications, including agricultural productivity, ecosystem functioning, urban dynamics, and public health outcomes (Eberhardt and Teal 2013, Collins *et al.* 2018). In this framework, spatial heterogeneity is largely understood as a consequence of systematic variation in explanatory variables across space (Downing 1986, Ren *et al.* 2025).

However, spatial systems often exhibit irregular, discontinuous, and locally extreme patterns that cannot be adequately captured by smooth gradients or average relationships alone (Zhang 2008, Cortes 2008). These patterns may arise from abrupt environmental transitions, localised disturbances, management interventions, or complex interactions among multiple processes (Zurlini *et al.* 2006). In practice, such irregularities frequently appear as spatial outliers that deviate from their surrounding neighborhoods to some extent (Harris *et al.* 2014, Sun *et al.* 2026). While spatial outliers are commonly treated as noise or anomalies to be filtered out during data preprocessing, growing evidence suggests that they may encode meaningful information about underlying spatial structures and processes (Shekhar *et al.* 2011).

Despite their potential importance, spatial outlier patterns have received limited attention in the context of explaining spatial heterogeneity (Wachowicz and Liu 2016). Most existing studies focus on detecting spatial outliers for quality control, anomaly identification, or hotspot analysis, rather than explicitly integrating outlier information into heterogeneity explanation frameworks (Liu *et al.* 2001, Wang *et al.* 2025). As a result, many components of spatial structure that are neither fully random nor well explained by first-dimension variables remain underexplored (Song 2022). This gap is particularly evident in studies that rely on global or locally weighted regression

frameworks, where heterogeneity is modelled primarily through coefficient variation rather than through representations of spatial pattern irregularities (Lu *et al.* 2014, Zhang *et al.* 2025).

To address this limitation, it is necessary to develop a conceptual and methodological approach that treats spatial outliers not simply as model residuals or noise, but as informative representations of underlying spatial patterns (Cao *et al.* 2026). While the second-dimension outlier (SDO) model proposed in Ren *et al.* (2026) provides a means of identifying multi-scale spatial outlier patterns, that study was primarily concerned with pattern detection rather than heterogeneity explanation. Building on that foundation, the present study extends the role of second-dimension outlier information from spatial pattern identification to spatial heterogeneity explanation (Yang *et al.* 2025). From this perspective, spatial heterogeneity can be understood as arising from both variation in attribute values and structural complexity in spatial configurations (Walter *et al.* 2023). Therefore, spatial outlier patterns provide a complementary way to capture local discontinuities, extreme behaviors, and unusual spatial configurations, which are difficult to represent using covariates alone.

In this study, we proposed the second-dimension outlier-driven heterogeneity (SOH) model to incorporate spatial outlier patterns into spatial heterogeneity analysis. SOH derives multi-scale spatial outlier patterns from the second-dimension outlier model and uses them as second-dimension explanatory components that complement conventional covariates. By combining CART-based stratification with the geographical detector, SOH quantifies the explanatory power of original variables, spatial outlier patterns, and their interactions through the power of determinant (PD). This design provides an operational pipeline for evaluating individual effects, interaction effects, and scale-dependent behavior of spatial patterns within a unified heterogeneity-explanation model. Distinct from the earlier SDO model, which primarily focuses on spatial outlier identification, SOH extends second-dimension outlier information towards spatial heterogeneity explanation. We demonstrated the model through an application to Australian barley production, showing how spatial outlier patterns and their interactions with environmental variables can reveal additional explanatory structure across multiple spatial scales.

2. Spatial outlier pattern for explaining heterogeneity

2.1. Concept of the second-dimension outlier-driven heterogeneity (SOH) model

The second-dimension outlier-driven heterogeneity (SOH) model provides a general approach for explaining and quantifying spatial heterogeneity by incorporating localized spatial outliers into heterogeneity analysis. Previous approaches typically characterize heterogeneity through first-dimension spatial variation, in which spatial differences in a dependent variable are attributed to corresponding variations in explanatory factors. In many geospatial contexts, however, heterogeneity is also shaped by irregular and locally anomalous structures that are not adequately represented by smooth gradients or global relationships.

SOH addresses this limitation by introducing a second dimension of spatial information derived from a second-dimension outlier (SDO) model (Ren *et al.* 2026). In SOH,

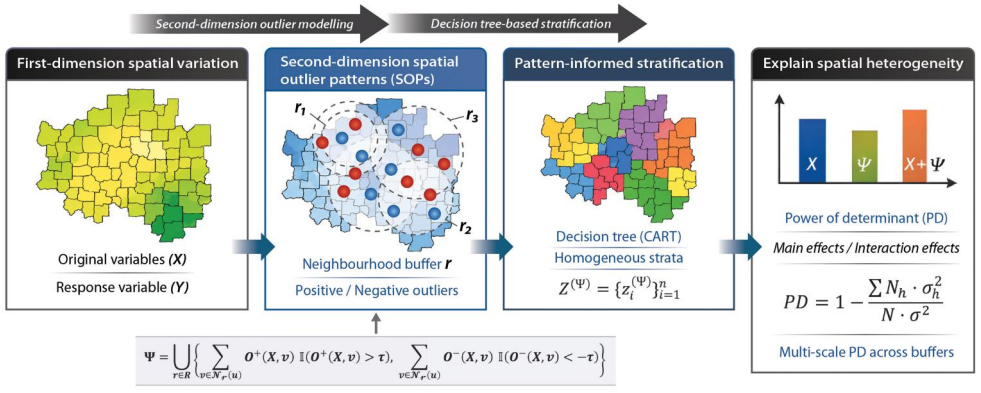

Figure 1. Conceptual framework of the second-dimension outlier-driven heterogeneity (SOH) model.

SDO outputs are regarded as spatial outlier patterns (SOPs) that describe the configuration and intensity of local anomalies relative to surrounding neighborhoods. These SOPs are then integrated into a heterogeneity-explaining workflow to compute a power of determinant (PD) statistic, which quantifies the explanatory strength of predictors in accounting for the spatial heterogeneity of the dependent variable. SOH adopts a two-stage strategy that couples decision tree-based stratification with the geographical detector (GD) (Wang *et al.* 2010, Song *et al.* 2020). A CART regression tree is first used to divide observations into data-adaptive strata based on SOP-informed predictors, and the resulting strata are subsequently evaluated within GD to estimate PD (Lewis *et al.* 2000). By considering both factor-driven variation and outlier-driven spatial structure, SOH can better represent anomaly-related spatial heterogeneity in different geospatial contexts.

2.2. SOH model

Figure 1 illustrates the conceptual workflow of the SOH model. Building on the conceptual discussion in Section 2.1, this subsection formalizes the model by defining the original variables, constructing spatial outlier patterns (SOPs) across multiple neighborhood scales, and linking these patterns to heterogeneity explanation through CART-based stratification and the geographical detector.

Let the original spatial variable be defined as

$$X = \{x_i\}_{i=1}^{n} \tag{1}$$

where x_i denotes the value of X observed at spatial unit i, and n is the total number of spatial units in the study domain.

Similarly, the response variable is represented as

$$Y = \{y_i\}_{i=1}^{n} \tag{2}$$

where y_i denotes the corresponding outcome measured at spatial unit i. All variables are defined on a common spatial support to ensure consistency in subsequent heterogeneity analysis.

To characterize spatial patterns beyond first-dimension covariate variation, the SOH model defines a spatial pattern variable Ψ as a collection of multi-scale second-dimension outlier components derived from X (Ren *et al.* 2026):

$$\Psi = \underset{r\in R}{\cup}\left\{\sum_{v\in\mathcal{N}_r(u)} O^+(X,v)\mathbb{I}\big(O^+(X,v)>\tau\big),\ \sum_{v\in\mathcal{N}_r(u)} O^-(X,v)\mathbb{I}(O^-(X,v)<-\tau)\right\} \tag{3}$$

where Ψ is the spatial pattern variable composed of second-dimension outlier components across multiple spatial scales; $R=\{r_1,r_2,\ldots,r_k\}$ represents a predefined set of neighbourhood buffer radii used to capture spatial patterns at different scales; $\mathcal{N}_r(u)$ is the spatial neighbourhood centred at target spatial unit u with buffer radius r, and v indexes spatial units located within $\mathcal{N}_r(u)$. $O^+(X,v)$ and $O^-(X,v)$ represent the positive and negative outlier components of the explanatory variable X at location v, respectively, which quantify upward and downward deviations from the local neighborhood distribution derived using the second-dimension outlier model. $\mathbb{I}(\cdot)$ is an indicator function that equals 1 when the condition is satisfied and 0 otherwise, and τ denotes the outlier threshold used to distinguish significant local deviations from the neighborhood distribution.

In practice, the positive and negative outlier components in Equation (3) are computed using the SDO model (Ren *et al.* 2026). For a given explanatory variable and neighborhood buffer, local outliers are identified by comparing values at surrounding unsampled locations with the neighborhood distribution around the target location. Positive and negative outliers are defined as values greater than $\bar{x}+2\sigma$ and smaller than $\bar{x}-2\sigma$, respectively, which corresponds to using $\tau=2\sigma$ as the outlier threshold in Equation (3). This threshold follows the original SDO formulation and provides a relatively conservative criterion for retaining pronounced local deviations while avoiding the inclusion of minor neighborhood fluctuations as outlier signals. More generally, τ can be treated as an adjustable parameter that controls the strictness of local outlier identification; depending on the spatial structure, noise level, and scale of a given dataset, alternative values such as 1.5σ, 2σ, or 2.5σ may be used to define different levels of local outlier intensity. For each buffer radius, the values of positive and negative outliers are accumulated separately to derive positive and negative outlier intensities. Repeating this procedure across multiple neighborhood buffers yields a set of multi-scale positive and negative outlier variables, which are then used to construct the spatial outlier pattern (SOP) representation in the SOH model.

For each spatial unit i, the corresponding spatial pattern set Ψ_i is expressed as

$$\Psi_i = \left\{O_{+,i}^{(r_1)},\ O_{-,i}^{(r_1)},\ O_{+,i}^{(r_2)},\ O_{-,i}^{(r_2)},\ \ldots,\ O_{+,i}^{(r_k)},\ O_{-,i}^{(r_k)}\right\} \tag{4}$$

where $O_{+,i}^{(r)}$ and $O_{-,i}^{(r)}$ represent the positive and negative outlier summaries for spatial unit i at neighborhood scale r, respectively. Collectively, Ψ_i represents the set of multi-scale second-dimension outlier summaries associated with X for spatial unit i.

To incorporate spatial pattern information into heterogeneity explanation, SOH adopts a stratification–detection strategy. Let

$$Z^{(\Psi)} = \{z_i^{(\Psi)}\}_{i=1}^{n} \tag{5}$$

denote a categorical stratification variable induced by the spatial pattern set Ψ, where $z_i^{(\Psi)}\in\{1,2,\ldots,L\}$ indicates the stratum to which spatial unit i is assigned, and L is the

total number of strata. In practice, $Z^{(\Psi)}$ is obtained via a data-driven stratification procedure using a CART regression tree implemented with the `rpart` package in R (Quinlan 1996). In this study, the tree was fitted with a complexity parameter of $cp = 0.01$, while the remaining tree-growing parameters followed the default `rpart` settings. The number of strata L was not fixed in advance; instead, it was determined adaptively by the number of terminal nodes produced by the fitted tree for a given predictor configuration. This procedure was applied consistently throughout the PD calculations for individual variables, SOPs, and their interaction-based combinations. Here, the CART model is used as a data-adaptive stratification tool for constructing categorical partitions in the heterogeneity analysis, rather than as a predictive model for out-of-sample generalization.

Based on the stratification variable $Z^{(\Psi)}$ induced by Ψ, the GD model is employed to quantify the explanatory power of the resulting strata (Wang *et al.* 2010, Song *et al.* 2020), yielding the power of determinant (PD) defined as

$$PD = \Omega_{\Psi} = 1 - \frac{\sum_{h=1}^{L} N_h^{(\Psi)} \cdot \sigma_h^{2(\Psi)}}{N^{(\Psi)} \cdot \sigma^{2(\Psi)}} \tag{6}$$

where Ω_{Ψ} is the proportion of the total spatial variance explained by the stratification induced by Ψ, $N_h^{(\Psi)}$ is the number of spatial units in the h-th stratum, $N^{(\Psi)} = \sum_{h=1}^{L} N_h^{(\Psi)}$ is the total number of spatial units, $\sigma_h^{2(\Psi)}$ is the within-stratum variance, and $\sigma^{2(\Psi)}$ is the overall variance across the entire study area.

A larger PD value indicates that the stratification induced by the spatial pattern variable Ψ yields stronger between-stratum differentiation relative to within-stratum variability, and thus provides greater explanatory power for spatial heterogeneity in the dependent variable Y.

Within the SOH model, PD can be consistently evaluated under different predictor configurations, including original covariates (X), spatial pattern variables (Ψ), and their combinations ($X \cup \Psi$). Moreover, PD values can be computed for joint predictor sets to assess interaction effects between multiple spatial patterns or between spatial patterns and original variables, enabling a unified and extensible model for explaining complex spatial heterogeneity.

3. SOH-based explanation of spatial heterogeneity in Australian barley production

3.1. Study area and data

Australia is a major contributor to global cereal production and ranked second worldwide in barley output in 2023, with a total production of 13,491,375 tonnes (Food and Agriculture Organization of the United Nations (FAO) 2025). As an economically important cereal crop, barley is extensively cultivated across Australia's temperate broadacre farming zones and exhibits pronounced spatial heterogeneity driven by climatic variability, soil properties, topographic conditions, and vegetation dynamics. These characteristics make Australian barley production a representative case for investigating spatial heterogeneity in agricultural systems.

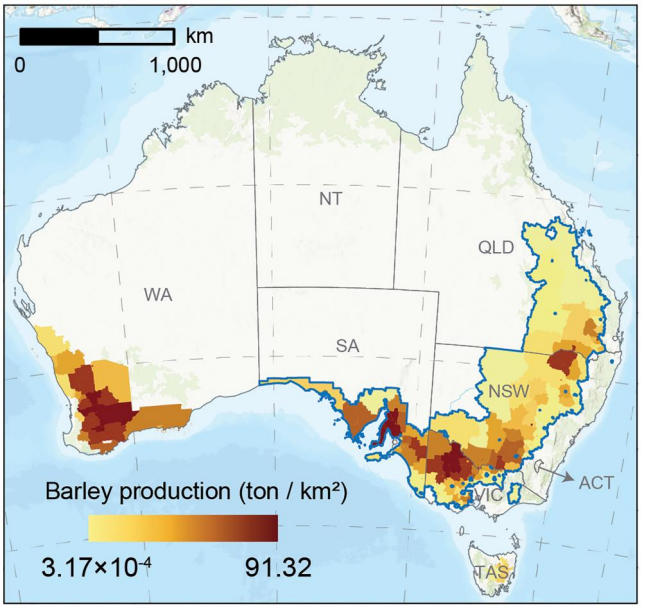

Figure 2. Spatial distribution of barley production normalised by SA2 polygon area (ton/km^2) across Australia and the study area. The blue boundary delineates the study area analysed in this study.

Barley cultivation in Australia is spatially concentrated within a contiguous agricultural belt spanning eastern and central regions, where cropping systems are well established and yield gradients are clearly structured at regional scales. Accordingly, this study focuses on the major barley-producing areas of New South Wales (NSW), Victoria (VIC), Queensland (QLD), and South Australia (SA), where barley production is more spatially concentrated and regionally connected. Barley production data for the 2023–2024 growing season were obtained from the Australian Bureau of Statistics (ABS) at the Statistical Area Level 2 (SA2) (Australian Bureau of Statistics, 2025). SA2 was adopted as the analysis unit because the response variable was reported at that scale, allowing consistent integration with the explanatory variables. We acknowledge that the use of SA2-level areal units may introduce MAUP-related effects, which are further discussed in Section 5. Figure 2 presents the spatial distribution of barley production across Australia during 2023–2024. In this study, barley production refers to production normalized by the projected area of each SA2 polygon and was used as the response variable in subsequent analyses. It is reported in ton/km^2 to remain consistent with the GIS-based areal analysis and the terminology used in the original ABS dataset.

The explanatory variables considered in this study represent key environmental determinants of barley production and are grouped into four categories: climate (C1), topography (C2), soil properties (C3), and vegetation dynamics (C4) (Table 1). These category labels (C1-C4) are used throughout the analysis to facilitate category-level stratification, interaction assessment, and interpretation of spatial heterogeneity. Climate variables include air temperature (AT), total precipitation (TP), and vapour pressure deficit (VPD), derived from the ERA5-Land reanalysis dataset released by the

Table 1. A summary of explanatory variables that potentially affect spatial disparities of barley production.

Category	Variable	Code	Product	Resolution
Climate (C1)	Air temperature	AT	ERA5_Land	0.1°
	Total precipitation	TP	ERA5_Land	0.1°
	Vapour pressure deficit	VPD	ERA5_Land	0.1°
Topography (C2)	Elevation	ELE	DEM-S (Geoscience Australia)	30 m
Soil properties (C3)	Soil organic carbon	SOC	CSIRO/SLGA	92.77 m
	Total nitrogen	NTO	CSIRO/SLGA	92.77 m
	Total phosphorus	PTO	CSIRO/SLGA	92.77 m
	Soil pH ($CaCl_2$)	pHc	CSIRO/SLGA	92.77 m
	Available water capacity	AWC	CSIRO/SLGA	92.77 m
Vegetation (C4)	Normalized difference vegetation index	NDVI	MOD13A2 V6.1	1 km
	Net primary production	NPP	MOD17A3HGF V6.1	500 m
	Evapotranspiration	ETa	TERN/AET/CMRSET Landsat V2.2	30 m

Copernicus Climate Change Service, with a native spatial resolution of 0.1° (Muñoz Sabater 2019). Topographic structure is represented by elevation (ELE), obtained from Geoscience Australia's Smoothed Digital Elevation Model (DEM-S) at a 30 m resolution (Geoscience Australia 2015). Soil physicochemical properties, including soil organic carbon (SOC), total nitrogen (NTO), total phosphorus (PTO), soil pH ($CaCl_2$) (pHc), and available water capacity (AWC), were sourced from the Soil and Landscape Grid of Australia (SLGA). This dataset was produced by CSIRO and provided at an approximate spatial resolution of 92.77 m (Viscarra Rossel *et al.* 2014). Vegetation dynamics were characterised using the MOD13A2 Version 6.1 NDVI product (Didan 2021) and the MOD17A3HGF Version 6.1 net primary production (NPP) product (Running and Zhao 2021). Actual evapotranspiration (ETa) was obtained from the CSIRO TERN/AET/CMRSET Landsat V2.2 product at a 30 m resolution, representing a high-resolution eco-hydrological indicator of surface water balance (Guerschman *et al.* 2022).

All explanatory variables were accessed through the Google Earth Engine (GEE) data catalogue, ensuring consistent data acquisition and preprocessing workflows. To achieve strict spatial interoperability across datasets with differing native resolutions and coordinate systems, all layers were reprojected to the `GDA_1994_Australia_Albers` coordinate system (EPSG:3577). The resulting spatial patterns of the explanatory variables across the study area are shown in Figure 3. Spatial resampling and harmonisation procedures were then applied to generate a unified analytical grid, enabling coherent integration of all predictor surfaces in the subsequent spatial heterogeneity modelling.

3.2. Experimental design

Figure 4 illustrates the case-study workflow used to implement and evaluate the SOH model. Following the input of original spatial variables into the SOH model, the analysis consists of five sequential steps, including SOP distribution analysis, PD estimation for individual variables and SOPs, interaction analysis, and model validation.

Before these analytical steps, original spatial variables representing climatic, topographic, soil, and vegetation conditions were preprocessed to ensure spatial consistency. All variables were resampled to a common spatial grid and coordinate system,

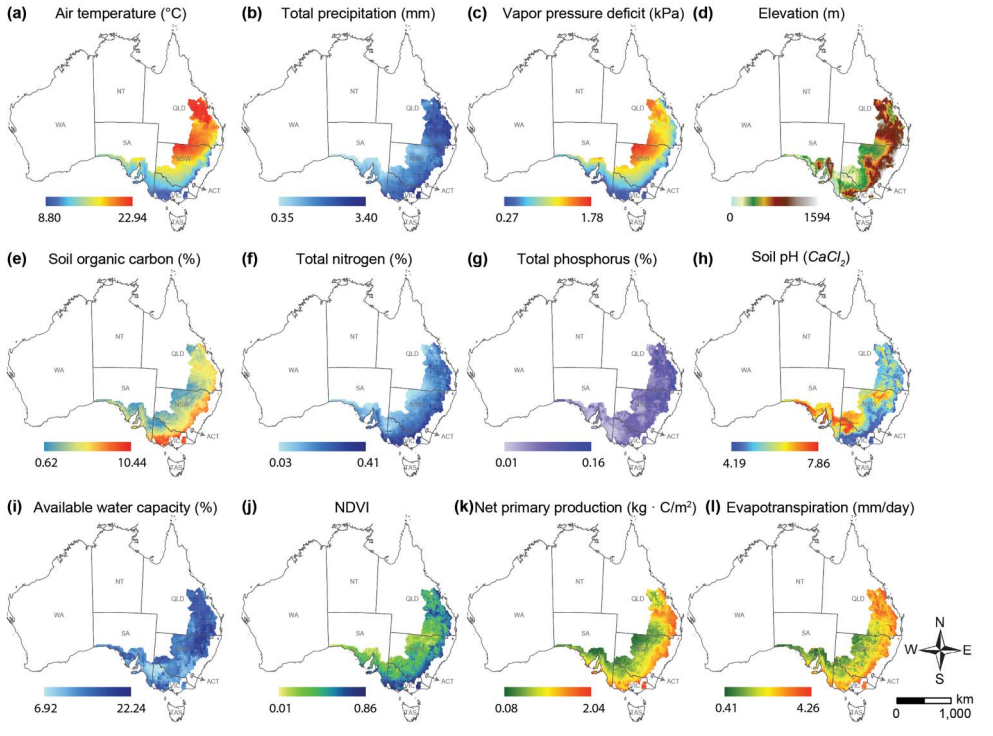

Figure 3. Spatial distribution of explanatory variables for barley production in southeastern Australia. (a)-(c) Climate: air temperature (AT), total precipitation (TP), and vapour pressure deficit (VPD); (d) Topography: elevation (ELE); (e)-(i) Soil properties: soil organic carbon (SOC), total nitrogen (NTO), total phosphorus (PTO), soil pH ($CaCl_2$) (pHc), and available water capacity (AWC); (j)-(l) Vegetation: normalized difference vegetation index (NDVI), net primary production (NPP), and evapotranspiration (ETa).

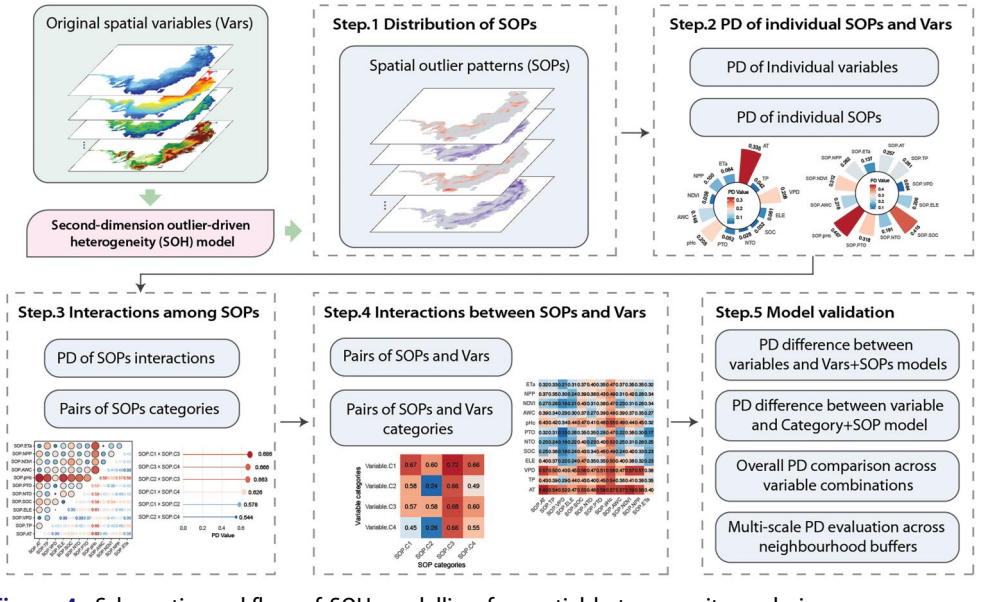

Figure 4. Schematic workflow of SOH modelling for spatial heterogeneity analysis.

and descriptive spatial distributions were examined to characterize baseline spatial variability across the study area. In addition, global spatial autocorrelation of the response variable was assessed using the univariate Moran's I statistic in GeoDa, based on a first-order queen-contiguity spatial weights matrix derived from the SA2 polygons and 999 random permutations.

First, spatial outlier patterns (SOPs) were generated for each original variable using the second-dimension outlier (SDO) model. For each neighborhood buffer, positive and negative local outliers were identified relative to the neighborhood distribution using a mean $\pm$ 2 standard deviation criterion, and their values were accumulated separately to derive positive and negative outlier intensities. SOPs were constructed across multiple neighborhood buffer radii to capture local-to-broad spatial anomaly structures. In the case study, buffer radii from 20 km to 200 km at 20 km intervals were used for the main SOP analysis, allowing systematic evaluation of SOP behavior from relatively local to broader neighborhood contexts. The spatial distributions of SOPs were then mapped to visualize how spatial outlier signals evolve with increasing neighborhood extent.

Second, PD values were calculated for individual original variables and individual SOPs using the GOZH model, which combines decision tree-based stratification and geographical detector analysis. This step enables comparison between first-dimension explanatory variables and second-dimension spatial outlier patterns in terms of their individual explanatory power.

Third, interactions among SOPs were systematically examined to assess whether combinations of spatial outlier patterns provide enhanced explanatory power for spatial heterogeneity. PD values were computed for all pairwise SOP interactions as well as for interactions among SOP categories, enabling the identification of dominant synergistic pattern combinations.

Fourth, interactions between SOPs and original spatial variables were evaluated to explore how spatial patterns complement first-dimension covariate information. PD values were calculated for pairwise combinations of SOPs and original variables, as well as for interactions between their corresponding categories. This step explicitly tests whether incorporating spatial outlier patterns improves the explanatory strength of conventional spatial variables.

Finally, model validation was conducted to assess the robustness and added value of the SOH model. Differences in PD between original variables and their SOP-augmented counterparts were examined at both the individual-variable and category levels. In addition, overall PD values were evaluated under three scenarios, namely original variables only, SOPs only, and the combination of original variables and SOPs, and their sensitivity to neighborhood buffer size was systematically analyzed. This validation step ensures that the observed improvements in explanatory power are systematic, stable, and not confined to a single predictor setting or spatial scale.

4. Results

4.1. Spatial autocorrelation of the response variable

Global Moran's I analysis indicated that the response variable exhibited strong and statistically significant positive spatial autocorrelation (Moran's I = 0.692, pseudo

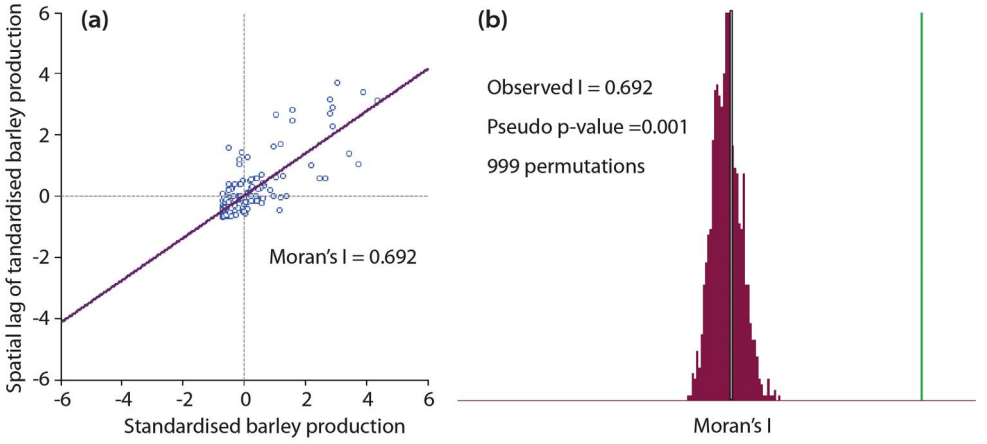

Figure 5. Global Moran's I diagnostic for SA2-level barley production. (a) Moran scatter plot showing the relationship between standardised barley production and its spatial lag. (b) Reference distribution based on 999 random permutations, indicating significant positive spatial autocorrelation in the response variable.

p-value = 0.001, z-value = 13.414; Figure 5). This result indicated that neighboring SA2 units tended to have similar barley production values, confirming pronounced spatial structure in the response variable and providing a basis for further examining how spatial outlier patterns contribute to heterogeneity explanation.

4.2. Spatial distribution of SOPs

Figure 6 presents the spatial outlier patterns (SOPs) of the three explanatory variables with the highest power of determinant (PD), namely air temperature (AT), vapor pressure deficit (VPD), and soil pH (pHc), evaluated across neighborhood buffer radii ranging from 20 km to 200 km at 20 km intervals. Within each panel, the first and second rows display positive spatial outlier patterns (SOP^{+}), while the third and fourth rows display negative spatial outlier patterns (SOP^{-}).

Across all three variables, the magnitude and spatial extent of SOPs increased progressively with larger neighbourhood buffers, indicating a strengthening of spatial outlier signals as broader contextual information is incorporated. For AT, positive spatial outlier patterns (SOP^{+}) were primarily concentrated in southeastern Queensland (QLD), southern South Australia, and western New South Wales (NSW) near the margins of the barley belt, whereas negative patterns (SOP^{-}) were mainly observed in southeastern QLD, northern Victoria (VIC), and both the northern and southern parts of NSW. The spatial configuration of SOPs derived from VPD exhibited a broadly similar pattern to that of AT, suggesting a shared climatic control on their spatial anomaly structures.

In contrast, pHc displayed a distinct spatial organization of outlier patterns. Positive SOPs were predominantly concentrated in southern VIC and southeastern New South Wales, while negative SOPs were mainly distributed across northwestern Victoria and western South Australia. The distribution of SOPs indicated that the spatial outlier patterns of different variables were distinct, reflecting their different roles in influencing the spatial heterogeneity of barley production.

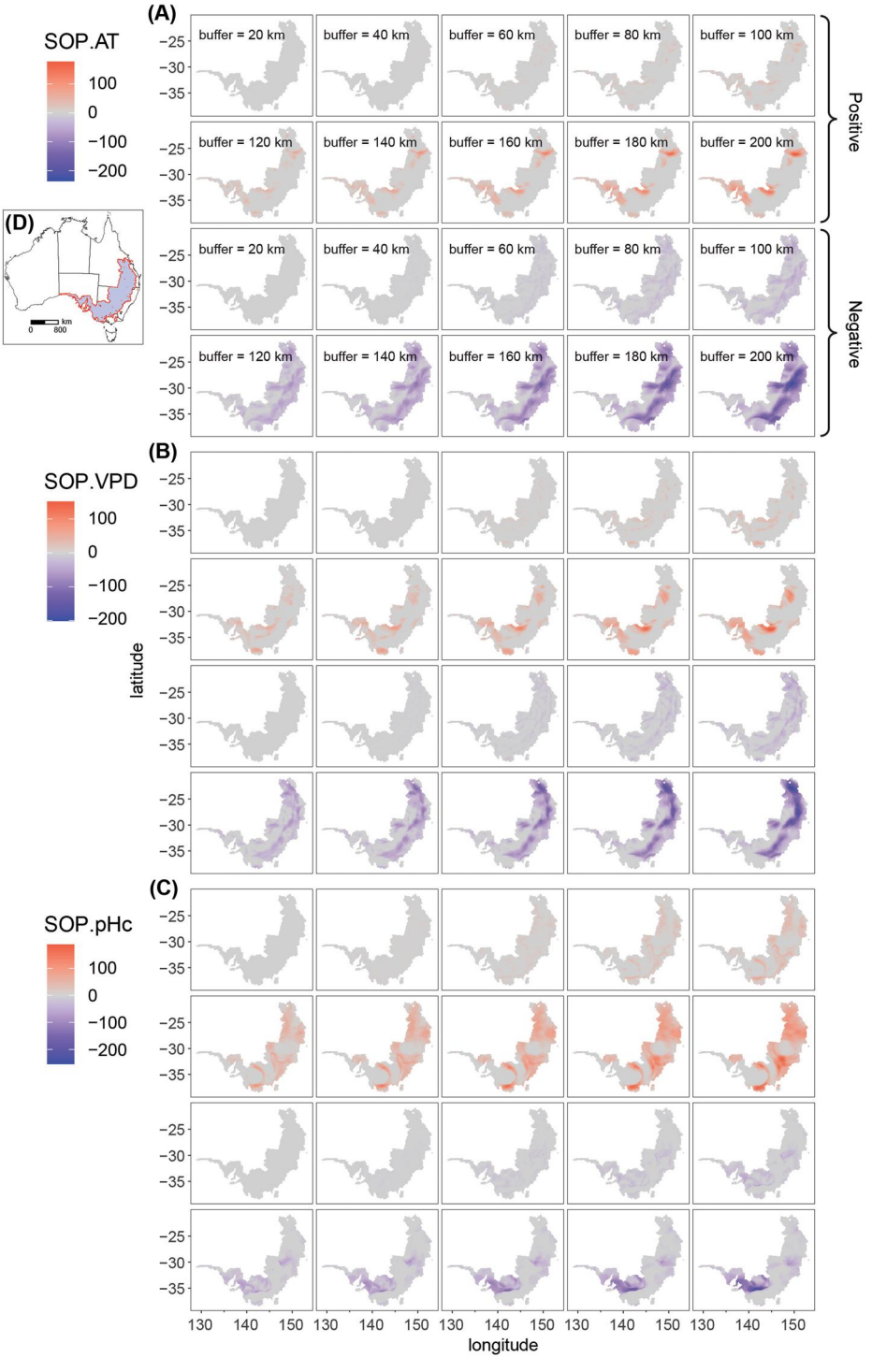

Figure 6. Spatial distribution of second-dimension outlier patterns (SOPs) for the three explanatory variables with the highest PD values in barley production spatial heterogeneity.

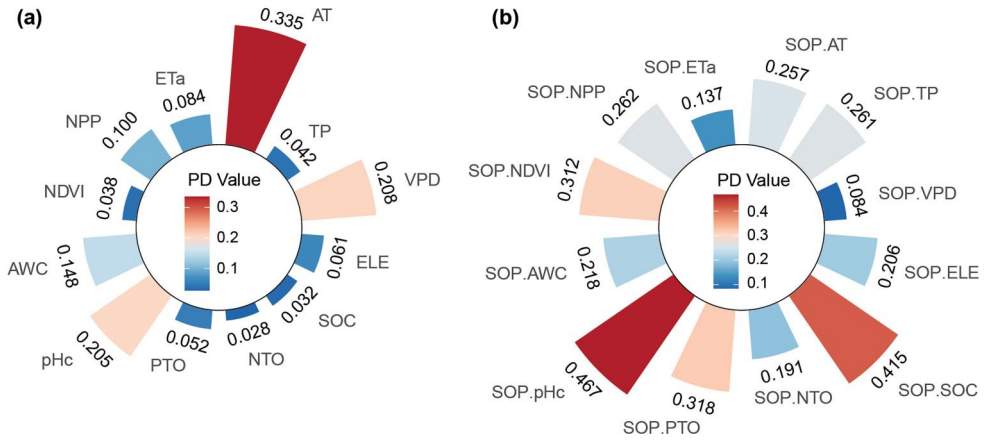

Figure 7. PD values of individual original variables and their corresponding spatial outlier patterns (SOPs) for explaining spatial heterogeneity in barley production. (a) shows PD values of individual original variables, while (b) presents PD values of the corresponding individual SOPs.

4.3. Power of determinant (PD) of individual SOPs

Figure 7 presents the PD values of individual original variables and their corresponding SOPs. Among the original variables (Figure 7 (a)), air temperature (AT) showed the highest explanatory power (PD = 0.335), followed by vapour pressure deficit (VPD; PD = 0.208) and soil pH (pHc; PD = 0.205), whereas most soil and vegetation variables exhibited relatively low PD values (generally below 0.15). In contrast, the corresponding SOPs (Figure 7 (b)) displayed higher PD values for most variables, with SOP.pHc (PD = 0.467), SOP.SOC (PD = 0.415), SOP.PTO (PD = 0.318), and SOP.NDVI (PD = 0.312) ranking as the strongest individual determinants. Only SOP.AT (PD = 0.257) and SOP.VPD (PD = 0.084) exhibited lower or comparable explanatory power relative to their original counterparts. These results indicated that spatial outlier patterns enhanced the individual explanatory strength of most environmental variables, particularly soil- and vegetation-related factors.

4.4. Interactions among SOPs

Figure 8 (a) illustrates the pairwise interactions among individual SOPs in terms of their PD values, while Table 2 summarizes the top ten SOP-SOP interactions. Overall, SOP interactions exhibited substantially higher explanatory power than individual SOPs, indicating strong synergistic effects among spatial outlier patterns. Among all pairs, interactions involving soil pH-related SOPs dominated the upper range of PD values. In particular, SOP.AT-SOP.pHc achieved the highest PD value (0.624), followed by SOP.ELE-SOP.pHc (0.610) and SOP.TP-SOP.pHc (0.599). Additional high-ranking interactions included SOP.pHc with ETa (0.591), AWC (0.588), SOC (0.583), and NTO (0.578), demonstrating that pHc-related spatial outlier patterns dominated in structuring heterogeneity through interactions.

Figure 8 (b) further aggregates SOP interactions at the category level, revealing consistent patterns across SOP groups, where SOP.C1-SOP.C4 correspond to climate,

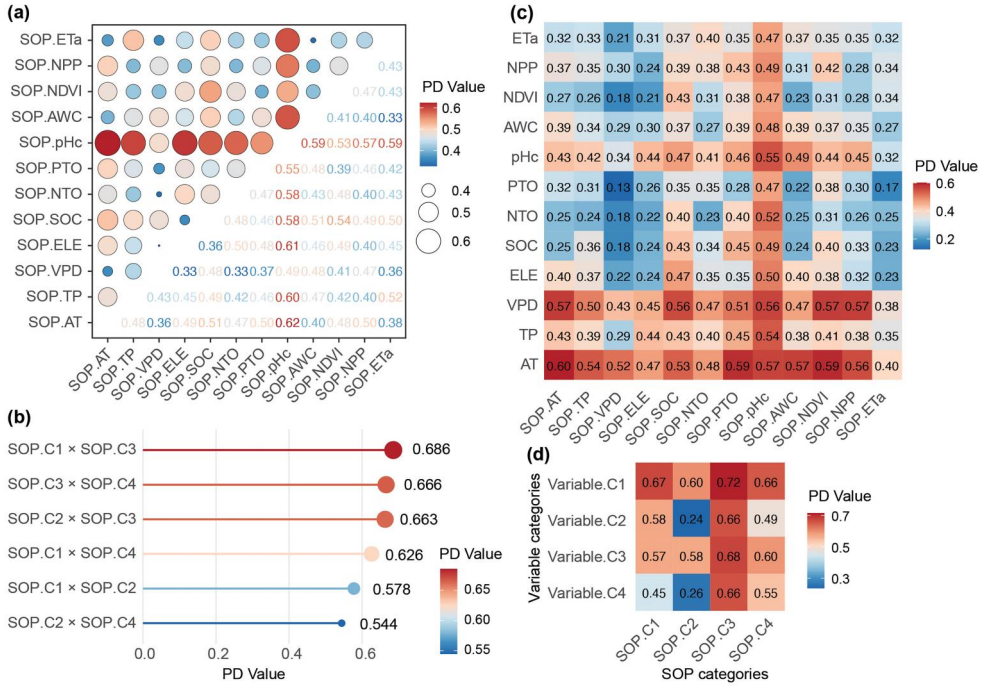

Figure 8. Power of determinant (PD) of interactions among spatial outlier patterns (SOPs) and between SOPs and original spatial variables for explaining spatial heterogeneity in barley production. (a) shows PD values of pairwise interactions among individual SOPs, (b) presents PD values of interactions among SOP categories, (c) illustrates PD values of pairwise interactions between individual original variables and SOPs, and (d) displays PD values of interactions between original variable categories and SOP categories.

topography, soil properties, and vegetation categories, respectively. Pairwise interactions between SOP categories exhibited PD values ranging from 0.544 to 0.686, all exceeding those of most individual SOPs. The strongest interaction was observed between SOP.C1 and SOP.C3 (PD = 0.686), followed by SOP.C3-SOP.C4 (0.666) and SOP.C2-SOP.C3 (0.663). In contrast, interactions involving SOP.C2 and SOP.C4 generally yielded comparatively lower PD values. These results indicated that SOP interactions not only enhanced explanatory power at the variable level, but also showed consistent interaction structures at the category level. This suggests that the explanatory role of outlier-driven spatial patterns is better understood through their joint configurations rather than through isolated effects alone.

4.5. Interactions between SOPs and original spatial variables

Figure 8 (c) presents the interaction effects between individual original variables and SOPs in terms of PD values, while Table 3 lists the top 30 variable-SOP interactions ranked by explanatory power. Overall, interactions between original variables and SOPs consistently yielded higher PD values than most individual variables or SOPs alone, highlighting the importance of jointly considering first- and second-dimension spatial information. The strongest interactions were dominated by climate-related

Table 2. Top 10 pairwise interactions among spatial outlier patterns (SOPs) ranked by power of determinant (PD).

SOP	SOP	*qv*	*Sig.*
AT	pHc	0.624	7.98×10^{-10}
ELE	pHc	0.610	6.08×10^{-10}
TP	pHc	0.599	6.59×10^{-10}
pHc	ETa	0.591	5.84×10^{-10}
pHc	AWC	0.588	6.63×10^{-10}
SOC	pHc	0.583	8.66×10^{-10}
NTO	pHc	0.578	1.00×10^{-9}
pHc	NPP	0.566	5.91×10^{-10}
PTO	pHc	0.549	8.72×10^{-10}
SOC	NDVI	0.539	9.51×10^{-10}

Table 3. Top 30 interactions between original spatial variables and spatial outlier patterns (SOPs) ranked by power of determinant (PD).

Variable	SOP	*qv*	*Sig.*	Variable	SOP	*qv*	*Sig.*
AT	AT	0.603	6.54×10^{-10}	NTO	pHc	0.520	5.28×10^{-10}
AT	NDVI	0.591	7.42×10^{-10}	AT	VPD	0.518	9.14×10^{-10}
AT	PTO	0.587	8.61×10^{-10}	VPD	PTO	0.513	6.26×10^{-10}
VPD	AT	0.574	5.64×10^{-10}	ELE	pHc	0.504	6.45×10^{-10}
VPD	NDVI	0.572	9.21×10^{-10}	VPD	TP	0.503	7.40×10^{-10}
VPD	NPP	0.570	6.60×10^{-10}	SOC	pHc	0.490	5.58×10^{-10}
AT	AWC	0.567	7.88×10^{-10}	NPP	pHc	0.489	8.92×10^{-10}
AT	pHc	0.565	4.79×10^{-10}	pHc	AWC	0.488	7.06×10^{-10}
VPD	pHc	0.564	9.24×10^{-10}	AWC	pHc	0.481	8.44×10^{-10}
VPD	SOC	0.557	5.75×10^{-10}	AT	NTO	0.476	7.61×10^{-10}
AT	NPP	0.556	6.80×10^{-10}	VPD	NTO	0.474	4.80×10^{-10}
pHc	pHc	0.548	6.74×10^{-10}	ELE	SOC	0.473	7.64×10^{-10}
TP	pHc	0.543	8.13×10^{-10}	pHc	SOC	0.472	9.20×10^{-10}
AT	TP	0.542	5.51×10^{-10}	AT	ELE	0.471	7.66×10^{-10}
AT	SOC	0.532	9.07×10^{-10}	VPD	AWC	0.468	8.29×10^{-10}

variables, particularly air temperature (AT) and vapor pressure deficit (VPD). The highest PD value was observed for AT-SOP.AT (0.603), followed by AT-SOP.NDVI (0.591) and AT-SOP.PTO (0.587). Similarly, VPD exhibited strong interactions with multiple SOPs, including SOP.AT (0.574), SOP.NDVI (0.572), and SOP.NPP (0.570). Interactions involving SOP.pHc also appeared frequently among the top-ranked combinations, such as AT-SOP.pHc (0.565), VPD-SOP.pHc (0.564), and TP-SOP.pHc (0.543).

It is worth noting that the highest-ranked variable–SOP interaction, AT–SOP.AT, should not be interpreted as a trivial self-combination. AT captures the original first-dimension magnitude of air temperature, whereas SOP.AT represents its second-dimension spatial anomaly pattern relative to the surrounding neighborhood context. Their joint effect therefore reflects the combined explanatory contribution of variable level and spatial configuration. At the same time, the fact that some SOP–SOP interactions, such as SOP.AT–SOP.pHc and SOP.ELE–SOP.pHc, yield higher PD values than AT–SOP.AT suggests that interactions among second-dimension spatial patterns can, in some cases, provide even stronger explanatory power than variable–pattern combinations.

Figure 8 (d) further summarizes interaction effects at the category level, showing variations in explanatory power across different variable-SOP category combinations. Interactions involving Variable.C1 and SOP.C3 exhibit the highest PD values, reaching up to 0.72, whereas interactions associated with SOP.C2 generally display lower

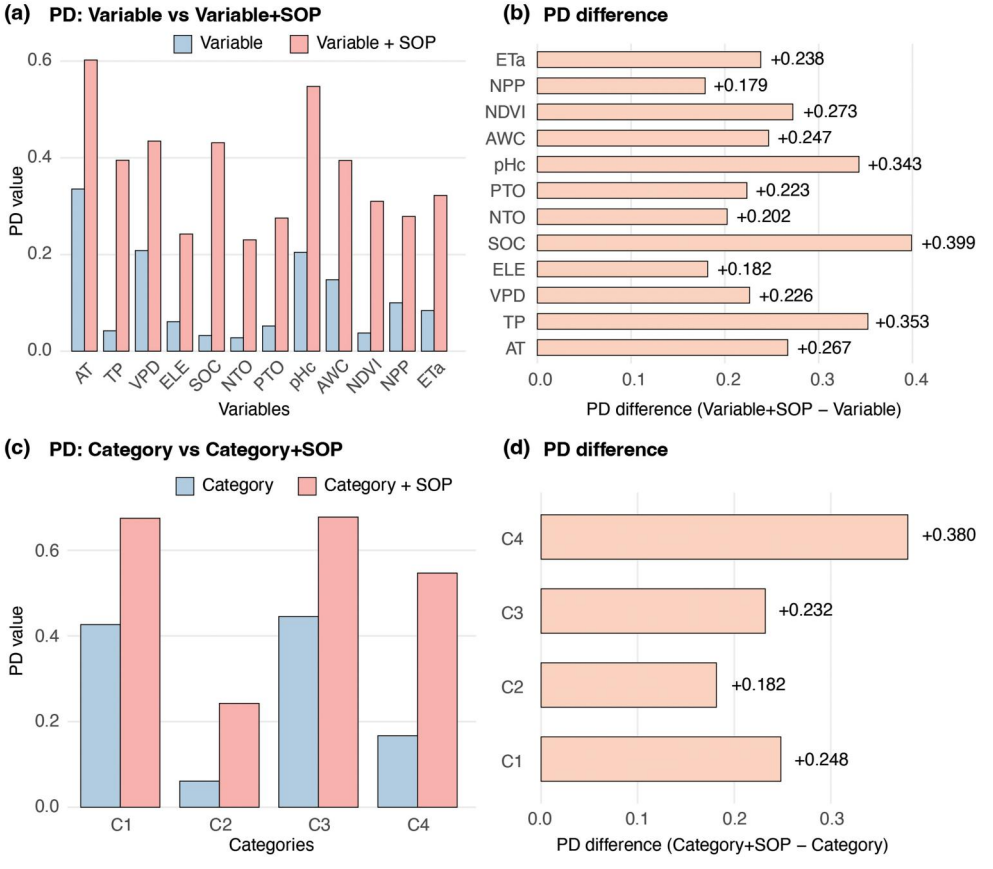

Figure 9. Comparison of power of determinant (PD) values between original variables (or categories) and their SOP-enhanced counterparts. (a) PD values of individual original variables and corresponding Variable + SOP models. (b) PD differences between Variable + SOP and original Variable models, highlighting the contribution of SOPs. (c) PD values of variable categories with and without SOP integration. (d) PD differences between Category + SOP and original Category models.

explanatory power, with PD values as low as 0.24. Across all categories, interactions between original variable groups and SOP categories consistently enhance PD values relative to individual components, indicating that second-dimension spatial outlier patterns provide complementary information that strengthens heterogeneity explanation when integrated with conventional spatial variables.

4.6. Model validation

The effectiveness of incorporating SOPs was first assessed at both the individual variable and category levels. Figure 9 (a,c) compare the PD values obtained using original variables (or categories) alone with those obtained after integrating SOPs, while (b) and (d) explicitly quantify the PD gains attributable to SOPs. For individual variables, the inclusion of SOPs led to consistent and substantial improvements in explanatory power, with PD increases generally exceeding 0.2 and reaching up to approximately 0.40 for variables such as SOC, TP, and pHc. Similar enhancement patterns were observed at the category

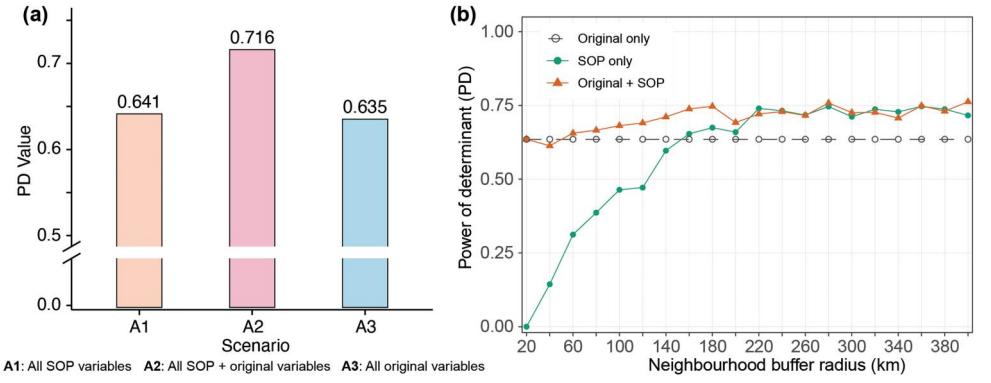

Figure 10. Overall performance and scale sensitivity of the second-dimension outlier-driven heterogeneity (SOH) model. (a) Comparison of PD values for three scenarios: all SOP variables (A1), all SOP variables combined with original variables (A2), and all original variables alone (A3). (b) Variation of PD values across neighbourhood buffer radius for original variables, SOP-only models, and combined Original + SOP models.

level. All four variable categories exhibited marked increases in PD after incorporating SOPs, with the magnitude of improvement varying across categories. In particular, PD increases of approximately 0.248 and 0.380 were observed for categories C2 and C4, respectively, indicating a substantial gain in explanatory power. Categories C1 and C3 also showed consistent improvements, although with comparatively smaller magnitudes. These results demonstrated that SOPs systematically strengthened the ability of both individual variables and aggregated categories to explain spatial heterogeneity, rather than benefiting only a limited subset of predictors.

At an overall level, comparisons among different predictor configurations further confirmed the robustness of the SOH model (Figure 10 (a)). When three scenarios were considered, including SOP-only predictors (A1), combined SOP and original predictors (A2), and original predictors alone (A3), the combined scenario consistently achieved the highest PD value (PD = 0.716). This indicated that SOPs provided complementary spatial information beyond that contained in original variables. Notably, the PD value obtained using SOPs alone is comparable to, and slightly higher than, that of original variables alone, suggesting that spatial outlier patterns by themselves already capture a substantial portion of the spatial heterogeneity structure.

Figure 10 (b) shows how PD values changed with increasing neighbourhood buffer radius used to construct SOPs. For buffer sizes below approximately 200 km, the PD of the combined model using original variables and SOPs remained consistently high and increased gradually, while the PD of the SOP-only model rose rapidly with increasing buffer size. In contrast, the PD of original variables remained constant because it was independent of spatial scale. Beyond a buffer radius of around 200 km, both the SOP-only model and the model combining original variables and SOPs exhibited stabilised PD values with only minor fluctuations, indicating the presence of a scale threshold. This transition suggested that SOPs progressively integrated broader spatial context as buffer size increased, and that beyond this threshold the SOPs already encapsulated sufficient spatial information, including information implicitly related to original variables, to explain spatial heterogeneity effectively on their own. Together, these results confirmed the

robustness of the SOH model and highlighted its capacity to bridge local and global spatial heterogeneity through multi-scale spatial outlier patterns.

5. Discussion

This study advanced spatial heterogeneity analysis by treating spatial outlier patterns as explanatory descriptors rather than as noise or post-hoc diagnostic signals. Conventional approaches often explain heterogeneity through first-dimension covariates that describe smooth gradients or systematic differences across space (De Marsily *et al.* 2005, Zhang *et al.* 2026). The SOH model extends this view by incorporating second-dimension spatial information, where SOPs encode the direction, intensity, and neighbourhood context of local deviations. Unlike LISA-type indicators, which are mainly used to diagnose local spatial autocorrelation (Li *et al.* 2021), SOPs are constructed as stratification-ready explanatory variables that can be directly incorporated into heterogeneity detection. Methodologically, SOH therefore provides a stratification–detection model that links spatial pattern construction with geographical detector-based explanation, while avoiding the need to specify a regression response model. At the same time, PD should not be interpreted as predictive performance or causal effect. It measures the extent to which a given spatial partition accounts for variance in the response variable, and its value is influenced by the constructed strata and the spatial structure of the data.

The barley production case illustrated why such pattern-based information can add explanatory value. Original variables represent the magnitude of environmental conditions, whereas SOPs capture local deviations and neighborhood contrasts. This distinction helps explain why SOPs often produced higher PD values than the corresponding original variables. In agricultural systems, yield heterogeneity is rarely governed only by smooth gradients in climate, soil, or vegetation conditions; it is also shaped by local anomalies, abrupt transitions, and clustered contrasts that affect crop responses to surrounding environmental conditions (Shi and Xingguo 2011). In this sense, the performance of SOPs suggests that barley production heterogeneity in the study area is structured not only by factor intensity, but also by the spatial organisation and contextual contrast of those factors.

A particularly important empirical finding was the dominant role of pHc-related SOP interactions. In Table 2, nearly all of the highest-ranked SOP–SOP interactions involve SOP.pHc, suggesting that soil pH anomalies act as an organising component in the broader heterogeneity structure of the barley belt. Agronomically, soil pH regulates nutrient availability, root development, and microbial processes, and therefore affects crop growth through multiple pathways (Barrow and Hartemink 2023). Spatially anomalous pH conditions may also reflect broader contrasts in soil condition, weathering status, hydrological processes, and agronomic suitability. Their interactions with climatic, topographic, soil, and vegetation-related SOPs likely capture coupled anomaly structures among soil chemical constraints, water–energy conditions, and crop response (Xing *et al.* 2025). The scale threshold around 200 km further indicates that SOPs are scale-dependent rather than arbitrary. Smaller buffers mainly capture local anomaly structures, whereas increasing buffer size progressively incorporates broader

regional context. The stabilisation of PD values beyond approximately 200 km suggests that, in this study area, the main meso-scale spatial structure relevant to barley heterogeneity has largely been captured by that point.

Several limitations should be noted. First, the response variable shows significant positive spatial autocorrelation, indicating that neighbouring SA2 units are not fully independent. This may increase apparent between-stratum differentiation and affect the magnitude of PD values. Therefore, PD should be interpreted as stratification-based explanatory differentiation under spatially structured data, rather than as a spatially independent or causal estimate. Second, the use of SA2-level areal units may introduce MAUP-related effects. Larger rural SA2 units may smooth within-unit variability and weaken fine-scale anomaly signals, so the results should be understood as SA2-level heterogeneity patterns rather than field-scale agronomic relationships. Finally, the SOH model is most suitable for systems with meaningful local anomaly structures, while its added value may be more limited in smoother spatial systems. The threshold parameter τ can also be adjusted according to the spatial structure, noise level, and analytical scale of the data.

6. Conclusions

This study proposed a second-dimension outlier-driven heterogeneity (SOH) model that incorporates spatial outlier patterns as informative descriptors for explaining spatial heterogeneity. SOH extends heterogeneity analysis beyond first-dimension covariate variation by deriving multi-scale spatial outlier patterns (SOPs) from a second-dimension outlier model and integrating them into a stratification-detection pipeline. The case study on Australian barley production demonstrated that SOPs enhanced explanatory power at both the individual-variable and interaction levels, and that their contributions were robust across spatial scales, with a clear transition from local to global heterogeneity around a neighborhood threshold. These findings highlighted the importance of spatial pattern irregularities in shaping heterogeneity and showed that spatial outliers can be transformed from anomalies into meaningful explanatory features. While the current model focuses on environmental drivers and a single agricultural application, future work could extend SOH to other domains, incorporate temporal dynamics, and explore adaptive scale selection strategies to further improve the interpretation of complex spatiotemporal heterogeneity.

Acknowledgements

The authors sincerely acknowledge the Editor-in-Chief, the Associate Editor, and the anonymous reviewers for their constructive comments and valuable suggestions, which have greatly contributed to improving the quality of this manuscript.

Author contributions

CRediT: **Kai Ren**: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing; **Yongze Song**: Conceptualization, Formal analysis, Supervision, Validation, Writing – original draft, Writing –

review & editing; **Xinyue Yang**: Data curation, Software, Visualization, Writing – original draft, Writing – review & editing; **Xi Wang**: Methodology, Validation, Visualization, Writing – review & editing; **Min Chen**: Formal analysis, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing; **Qiang Yu**: Resources, Supervision, Writing – original draft, Writing – review & editing.

AI disclosure statement

During the preparation of this manuscript, generative AI tools (ChatGPT 5.4) were used to assist with language polishing and code debugging. All concepts, methodological design, model construction, analyses, interpretations, and conclusions presented in this study were developed, verified, and approved by the authors.

Disclosure statement

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

Funding

This research was supported by the National Natural Science Foundation of China (Outstanding Young Scholars Program, Grant No. 42325107).

Notes on contributors

Kai Ren is a postdoctoral researcher at the State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, and was a visiting researcher at Curtin University, Australia. His research interests are geographic information science, spatial statistics, geospatial artificial intelligence, spatial heterogeneity and spatial prediction. His contributions to this paper include conceptualisation, methodology, software, formal analysis, data curation, visualisation, writing - original draft, and writing - review & editing.

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. His contributions to this paper include conceptualisation, formal analysis, supervision, validation, writing - original draft, and writing - review & editing.

Xinyue Yang is a PhD candidate in Construction Management at Curtin University, with research interests in geospatial modelling and urban thermal environment analysis. She contributed to the data curation, software development, visualisation, writing - original draft, and writing - review & editing.

Xi Wang is a joint PhD candidate with Lanzhou Jiaotong University and the National University of Singapore. His research interests include the application of geographic information and remote sensing technologies and spatial statistical modelling in urban applications. He participated in methodology, validation, visualisation, and writing - review & editing.

Min Chen is currently a professor at Nanjing Normal University. His research interests include geographic modelling, open geographic modelling and simulation, and virtual geographic environments. He contributed to resources, formal analysis, supervision, funding acquisition, writing - original draft, and writing - review & editing.

Qiang Yu is a Professor at the State Key Laboratory of Soil and Water Conservation and Desertification Control, Northwest A&F University, China. His research focuses on agricultural

resources and environment, climate change, and agro-ecosystem studies. He contributed to resources, supervision, writing - original draft, and writing - review & editing.

ORCID

Kai Ren http://orcid.org/0009-0007-1079-902X
Yongze Song http://orcid.org/0000-0003-3420-9622
Xinyue Yang http://orcid.org/0009-0000-1216-4334
Xi Wang http://orcid.org/0009-0009-9510-7786
Min Chen http://orcid.org/0000-0001-8922-8789
Qiang Yu http://orcid.org/0000-0001-6950-1821

Data and codes availability statement

The data and codes that support the findings of this study are available on Figshare at: https://doi.org/10.6084/m9.figshare.31169701.

References

Australian Bureau of Statistics 2025., Australian Agriculture: Broadacre Crops, 2023–24. Accessed: 27 Febrary 2026. URL https://www.abs.gov.au/statistics/industry/agriculture/australian-agriculture-broadacre-crops/latest-release

Barrow, N., and Hartemink, A.E., 2023. The effects of ph on nutrient availability depend on both soils and plants. *Plant and Soil*, 487 (1-2), 21–37.

Cao, X., *et al.*, 2026. Identifying determinants of groundwater storage using a geocomplexity-heterogeneity model. *Geo-Spatial Information Science*, 1–19.

Collins, S.L., *et al.*, 2018. Temporal heterogeneity increases with spatial heterogeneity in ecological communities. *Ecology*, 99 (4), 858–865.

Cortes, J., 2008. Discontinuous dynamical systems. *IEEE Control Systems Magazine*, 28 (3), 36–73.

De Marsily, G., *et al.*, 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13 (1), 161–183.

Didan, K., 2021. MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V061. NASA Land Processes Distributed Active Archive Center (LP DAAC).

Downing, J.A., 1986. Spatial heterogeneity: evolved behaviour or mathematical artefact? *Nature*, 323 (6085), 255–257.

Eberhardt, M., and Teal, F., 2013. No mangoes in the tundra: spatial heterogeneity in agricultural productivity analysis. *Oxford Bulletin of Economics and Statistics*, 75 (6), 914–939.

Food and Agriculture Organization of the United Nations (FAO) 2025. Production – crops and livestock products. Last updated: 31 January 2025; Accessed: 27 Febrary 2026. URL https://www.fao.org/faostat/en/#data/QCL

Geoscience Australia 2015. *Digital Elevation Model (DEM) of Australia derived from LiDAR 5 metre grid*. Geoscience Australia.

Guerschman, J.P., *et al.*, 2022. Estimating actual evapotranspiration at field-to-continent scales by calibrating the CMRSET algorithm with MODIS, VIIRS, Landsat and Sentinel-2 data. *Journal of Hydrology*, 605, 127318.

Harris, P., *et al.*, 2014. Multivariate spatial outlier detection using robust geographically weighted methods. *Mathematical Geosciences*, 46 (1), 1–31.

Lewis, R.J …, p., *et al.*, 2000. An introduction to classification and regression tree (cart) analysis. In: *Annual meeting of the society for academic emergency medicine in San Francisco, California*. Vol. 14. p. 106.

Li, D., *et al.*, 2021. Spatial patterns of vegetation coverage change in giant panda habitat based on modis time-series observations and local indicators of spatial association. *Ecological Indicators*, 124, 107418.
Liu, H., Jezek, K.C., and O'Kelly, M.E., 2001. Detecting outliers in irregularly distributed spatial data sets by locally adaptive and robust statistical analysis and gis. *International Journal of Geographical Information Science*, 15 (8), 721–741.
Liu, H., Song, Y., and Yi, W., 2026. Degree of spatial interpretability. *International Journal of Geographical Information Science*, 0 (0), 1–21.
Lu, B., *et al.*, 2014. The gwmodel r package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17 (2), 85–101.
Muñoz Sabater, J., 2019. *ERA5-Land monthly averaged data from 1981 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).
Pickett, S.T., and Cadenasso, M.L., 1995. Landscape ecology: spatial heterogeneity in ecological systems. *Science (New York, N.Y.)*, 269 (5222), 331–334.
Quinlan, J.R., 1996. Learning decision tree classifiers. *ACM Computing Surveys*, 28 (1), 71–72.
Ren, K., *et al.*, 2025. Identifying climate and environmental determinants of spatial disparities in wheat production using a geospatial machine learning model. *GIScience & Remote Sensing*, 62 (1), 2533487.
Ren, K., Song, Y., and Yu, Q., 2026. Second-dimension outliers for spatial prediction. *International Journal of Geographical Information Science*, 40 (6), 1915–1942.
Roy, S., *et al.*, 2024. Spatial heterogeneity in the urban household living conditions: A-gis-based spatial analysis. *Annals of GIS*, 30 (1), 81–104.
Running, S., and Zhao, M., 2021. *MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061*. NASA Land Processes Distributed Active Archive Center (LP DAAC).
Shekhar, S., *et al.*, 2011. Identifying patterns in spatial information: A survey of methods. *WIREs Data Mining and Knowledge Discovery*, 1 (3), 193–214.
Shi, H., and Xingguo, M., 2011. Interpreting spatial heterogeneity of crop yield with a process model and remote sensing. *Ecological Modelling*, 222 (14), 2530–2541.
Shu, H., *et al.*, 2019. Quantifying the spatial heterogeneity of points. *International Journal of Geographical Information Science*, 33 (7), 1355–1376.
Song, Y., 2022. The second dimension of spatial association. *International Journal of Applied Earth Observation and Geoinformation*, 111, 102834.
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.
Sun, Y., *et al.*, 2026. Local effects of pattern interactions in driving urbanization. *International Journal of Applied Earth Observation and Geoinformation*, 146, 105072.
Turner, M.G. and Chapin, III, F.S., 2005. Causes and consequences of spatial heterogeneity in ecosystem function. In Ecosystem function in heterogeneous landscapes. Springer New York, pp. 9–30.
Vinatier, F., *et al.*, 2011. Factors and mechanisms explaining spatial heterogeneity: a review of methods for insect populations. *Methods in Ecology and Evolution*, 2 (1), 11–22.
Viscarra Rossel, R., *et al.*, 2014. *Soil and landscape grid national soil attribute maps – soil attribute release 1, version 2*. CSIRO.
Wachowicz, M., and Liu, T., 2016. Finding spatial outliers in collective mobility patterns coupled with social ties. *International Journal of Geographical Information Science*, 30 (9), 1806–1831.
Wagner, H.H., and Fortin, M.-J., 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology*, 86 (8), 1975–1987.
Walter, R.J., Tillyer, M.S., and Acolin, A., 2023. Spatiotemporal crime patterns across six us cities: Analyzing stability and change in clusters and outliers. *Journal of Quantitative Criminology*, 39 (4), 951–974.
Wang, J.-F., *et al.*, 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the heshun region, china. *International Journal of Geographical Information Science*, 24 (1), 107–127.

Wang, Y., *et al.*, 2025. A novel potential outlier recognition approach considering local heterogeneity enhancement to improve the quality of soil datasets. *Geoderma*, 454, 117200.

Xing, Y., Wang, X., and Mustafa, A., 2025. Exploring the link between soil health and crop productivity. *Ecotoxicology and Environmental Safety*, 289, 117703.

Yang, X., *et al.*, 2025. Irregular anisotropy in surface urban heat island footprint. *Sustainable Cities and Society*, 131, 106779.

Young, I.M., Crawford, J.W., and Rappoldt, C., 2001. New methods and models for characterising structural heterogeneity of soil. *Soil and Tillage Research*, 61 (1-2), 33–45.

Zhang, C., 2008. An analysis of urban spatial structure using comprehensive prominence of irregular areas. *International Journal of Geographical Information Science*, 22 (6), 675–686.

Zhang, P., *et al.*, 2025. Gaussian mixture segmentation for managing deterioration of large-scale road networks. *IEEE Transactions on Intelligent Transportation Systems*, 26 (8), 12461–12473.

Zhang, X., *et al.*, 2026. Quantifying the effects of local determinants and geocomplexity on pm2. 5 spatial patterns. *GIScience & Remote Sensing*, 63 (1), 2665912.

Zhang, Y., *et al.*, 2023. A framework for estimating actual evapotranspiration through spatial heterogeneity-based machine learning approaches. *Agricultural Water Management*, 289, 108499.

Zurlini, G., *et al.*, 2006. Disturbance patterns in a socio-ecological system at multiple scales. *Ecological Complexity*, 3 (2), 119–128.