International Journal of Applied Earth Observation and Geoinformation 139 (2025) 104531

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Local pathways of association

Jiao Hu [a,e,1], Rui Qu [b,c,d,1], Yongze Song [e,*], Peng Wu [e]

[a] College of Earth and Planet Science, Chengdu University of Technology, Chengdu 610059, China
[b] State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China
[c] College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China
[d] School of Earth and Planetary Sciences, Discipline of Spatial Sciences, Curtin University, Perth 6845 WA, Australia
[e] School of Design and the Built Environment, Curtin University, Perth 6845, Australia

ARTICLE INFO

Keywords:
Spatial association
Local indicators
Path analysis
Spatial variation

ABSTRACT

Spatial association reveals the interconnected nature of geographical phenomena, describing the interactions and influences of environmental variables across geographic space. Path analysis can explore complex causal relationships between variables by analyzing path coefficients. However, in large-scale studies, path analysis methods are often affected by local effects, which can influence the accuracy and reliability of the results. This study develops a local pathway association (LPA) model to analyze local effects of pathways among variables that integrates path analysis and local pathway coefficient estimations. The LPA model was employed to investigate the spatial heterogeneity of spatial associations between factors such as climate, soil, and vegetation on the Tibetan Plateau. Results indicate that the LPA model effectively reveals the spatial variation characteristics of local path coefficients between geographic variables, avoiding the underestimation or overestimation of global path coefficients in traditional path coefficient studies. The developed LPA model provides an effective technical tool for revealing spatial differences in path associations of large-scale spatial studies. The strong data compatibility of the LPA model allows for broad applicability across various disciplines and a deeper understanding of localized interactions and variations in complex geospatial and Earth systems.

1. Introduction

Accurately identifying the spatial distribution patterns between geographic variables is crucial for exploring the spatial association of geographical factors (Getis, 2001; Goodchild et al., 1992). The spatial association is typically constructed based on spatial statistical inference, which analyzes the relationships between geographic locations and their attributes (Getis and Ord, 1992; Song, 2023). In recent years, spatial association models have been widely applied in various fields, including ecosystem management (Chen et al., 2019), biodiversity (Stein et al., 2014), public health (Yao et al., 2020), energy and carbon emissions (Bai et al., 2020; Xu et al., 2022), urban planning (Majumder et al., 2023), and disaster risk assessment (Coker et al., 2020).

Methods for analyzing spatial association can be generally categorized into three types. The first type involves spatial overlay analysis is a technique within Geographic Information Systems (GIS) that superimposes multiple spatial data layers, such as maps, geographical features, and attribute information, to identify and evaluate the spatial relationships and interactions among these layers (Hasanloo et al., 2019; Unwin, 2019; Zhang et al., 2024a). The spatial overlay analysis can utilize a unified spatial grid index to overlay lines and polygons (Wang et al., 2015), allowing for the determination of positional errors and their propagation patterns through the overlay error propagation process (Shi et al., 2004). The second type of approaches characterizes the spatial dependency of variables, operating on the assumption that values of attributes in closer proximity exhibit stronger correlations than those farther apart. Spatial dependency is typically quantified through spatial neighborhood relations, lag effects, or spatial weighting matrices (Crawford 2009, Varouchakis 2019, Song and Wu 2021). The third type quantifies spatial association based on spatial heterogeneity. Spatial heterogeneity refers to the variation or uneven distribution of spatial data, phenomena and patterns across different locations (Pickett and Cadenasso, 1995; Zhang et al., 2024b), reflecting the distinct characteristics of different regions (De Marsily et al., 2005; Luo et al., 2023).

Among these approaches, path analysis effectively quantifies spatial associations, while the Structural Equation Model (SEM) and its spatial

* Corresponding author.
E-mail addresses: hujiao@stu.cdut.edu.cn (J. Hu), qurui@stu.cdut.edu.cn (R. Qu), yongze.song@curtin.edu.au (Y. Song), peng.wu@curtin.edu.au (P. Wu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jag.2025.104531
Received 15 November 2024; Received in revised form 27 March 2025; Accepted 6 April 2025
Available online 13 April 2025

extensions based on spatial dependency enhance analysis by incorporating latent variables and causal relationships to improve the understanding of spatial pathways (Perry and Dixon 2002, Caldas de Castro and Singer 2006). The SEM employs path analysis to elucidate the internal structures among variables (Jöreskog and Sörbom 1982) and quantifies the associations between explanatory and response variables through path coefficients (Yuan and Bentler 2006). In practical applications, SEM serves as a powerful tool for investigating causal relationships and identifying coefficients among variables (Mueller and Hancock 2018). Methodologically, the incorporation of Bayesian estimation has enhanced SEM's capacity to handle small sample sizes, complex models, and prior information (Raftery 1993).

Although SEM is an effective tool for exploring complex spatial associations between geographic variables through path coefficient calculations, it often does not account for the impact of local effects on the accuracy of global path coefficients. In geospatial data studies, spatial associations between geographic variables can vary significantly across regions, especially in large-scale analyses at national, continental, or global levels (Hu et al., 2025). Local geographical feature analysis, a core component of geospatial analysis, involves dividing the study area into subregions to compute local values, thereby identifying hotspots and delineating their boundaries, as well as revealing the spatial variability in relationships among geographic variables (Song, 2022; Ord, 2024). For instance, the concept of Local Indicators of Spatial Association (LISA) was introduced to address the limitations of traditional global spatial analysis methods (Anselin, 1995). By defining local spatial relationships and computing local statistics, the LISA approach effectively identifies and analyzes hotspots, cold spots, local anomalies, and spatial clustering phenomena within the study area (Bivand and Wong, 2018).

This study develops a local pathway association (LPA) model to analyze local paths of spatial association and explain spatial variations in these associations. The LPA model is applied in a case study to determine local paths of spatial association among vegetation, climate, and soil explanatory variables on the Tibetan Plateau. The structure of this paper is as follows: Section 2 outlines the development steps of the

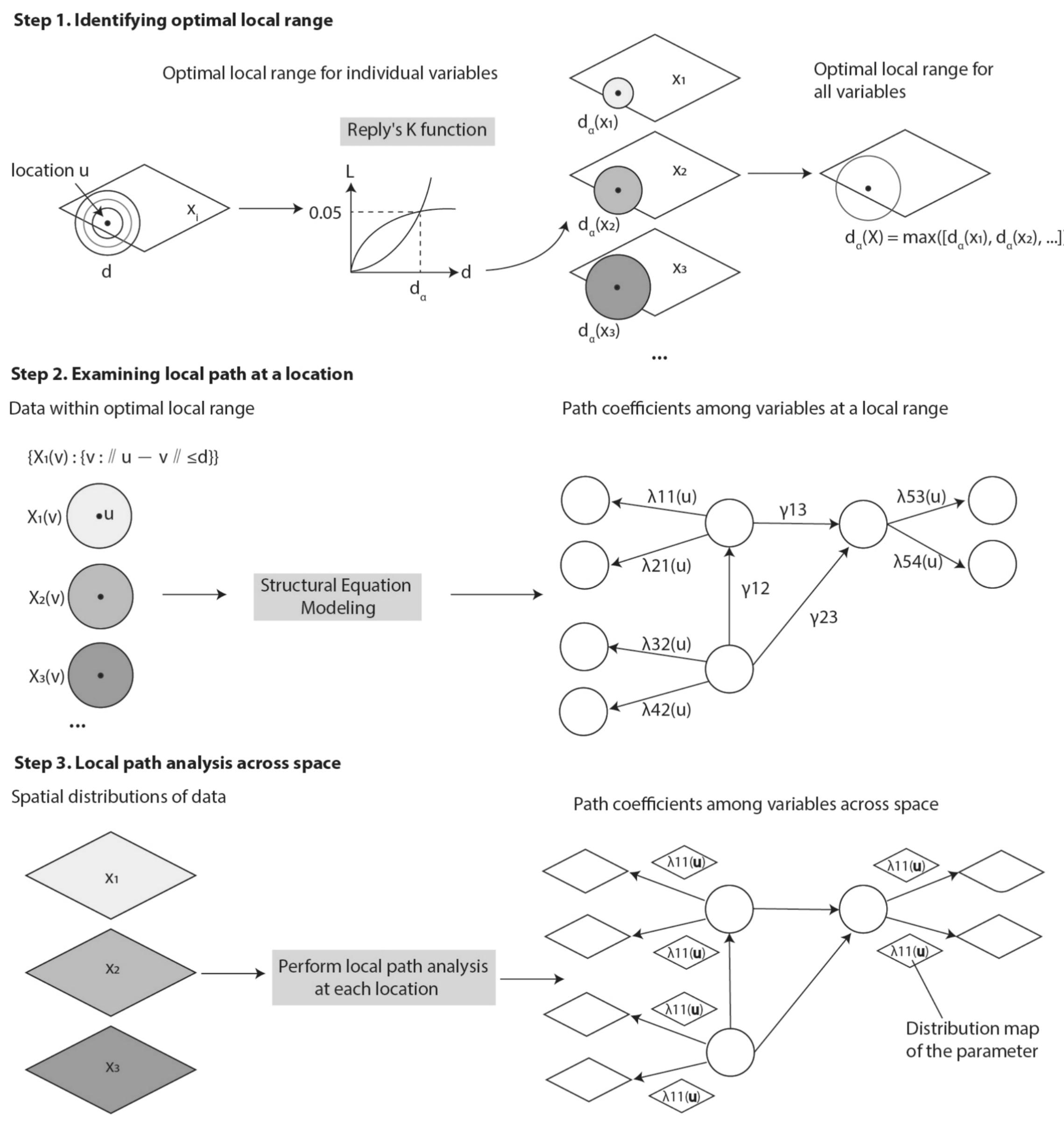

Fig. 1. The technical flowchart of the local pathway association (LPA) model.

LPA model, Section 3 presents the case study employing the LPA, Section 4 showcases the results, Section 5 discusses the advantages and limitations of the LPA, and finally, Section 6 concludes with the main findings of the research.

2. Local pathways of association

In this study, a local pathway association (LPA) is proposed to analyze the local relationships of variables in spatial data and their changes in different spatial locations and to reveal the complex relationships between variables in spatial data through local analysis and global spatial extension. Fig. 1 shows the overall process of LPA, which includes three steps. First, the local optimal range of each variable is determined using Ripley's K function, and the integrated local range is calculated by combining multiple local optimal ranges. Second, local data are extracted within the integrated local ranges, and local path coefficients are estimated using SEM to reveal causal relationships between variables. Finally, local path analysis is applied across all spatial locations across the study area, with path coefficients calculated point by point to examine variations in relationships within the optimal spatial ranges.

2.1. Identifying the optimal local range

The first step of the LPA model is to identify the optimal local ranges of spatial variables by analyzing the aggregation patterns of spatial data at varying distances. In this step, a Ripley's K function is used to examine the distribution characteristics of data in space, and the optimal local range for each variable is identified using the normalization with the L function. The process consists of four steps: (1) calculating the distances between pairs of spatial data, (2) fitting Ripley's K function to assess spatial clustering, (3) normalizing Ripley's K function using the L function, and (4) determining the optimal local range based on the normalized values.

In spatial analysis, the distance between locations of spatial data is an important factor in determining spatial correlation. First, we need to calculate the Euclidean distance between each pair of data. For two locations i and j in space, the distance between them can be calculated by the following formula:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + \left(y_i - y_j\right)^2} \tag{1}$$

where (x_i, y_i) and (x_j, y_j) are the coordinates of locations i and j, respectively. d_{ij} is the Euclidean distance between the two locations.

Ripley's K function is a spatial statistical method for measuring the degree of aggregation or dispersion of point distributions at different scales. It calculates the number of pairs of data within a given distance r and normalizes it to the area of the study area and the total number of data to quantify the aggregation of data at different distances. The method can characterize the aggregation pattern of the point distribution with distance, with higher K values when data are highly aggregated at small scales. In the study, a range of distances r is selected, and the number of point pairs within each distance range is calculated. The formula is:

$$K(r) = \frac{A}{n^2} + \sum\nolimits_{i=1}^{n} \sum\nolimits_{j \neq i} I\left(d_{ij} \leq r\right) \tag{2}$$

where A is the area of the study area, which is a known constant, and n is the number of spatial data in the study area. d_{ij} is the distance between point i and point j. $I\left(d_{ij} \leq r\right)$ is an indicator function that takes the value of 1 when $d_{ij} \leq r$, indicating that the distance of the point pair is within r; otherwise, it equals 0.

Although the results of Ripley's K function provide information about the distribution of data at different distances, to facilitate the interpretation of the results, it is common to standardize the K function and use the L function (or Besag's L function) to express the degree of aggregation. Through the L function, we can intuitively determine the spatial distribution pattern at different distances. When $L(r) > 0$, it means that the distribution of data within a distance r is aggregated, and the data are more concentrated than in the case of a random distribution; while when $L(r) < 0$, it means that the distribution of data is more dispersed at that distance. The formula for the L function is:

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r \tag{3}$$

where $L(r)$ is the standardized distance function used to measure the degree of clustering of data within a distance r. The term $\sqrt{\frac{K(r)}{\pi}}$ converts the value of the K function into an average distance based on the circumference of an area, making it easier to compare with the actual distance r. r represents the current distance value.

Based on the calculation of Ripley's K function and L function, we can determine the optimal local range of a variable, i.e., the range where the variable has the most significant effect in space. By calculating the values of the L function corresponding to different distances r, we find the maximum value of the function $L(r_{opt})$ corresponding to the distance r_{opt}. the formula is as follows:

$$r_{opt} = argmax \tag{4}$$

where the distance r_{opt} is the optimal local range, indicating that the aggregation of data is strongest within this distance range.

2.2. Examining the local path at a location

The local variable path coefficients are calculated based on the structural equation model (SEM). The key step involves examining local paths within the optimal local ranges identified in the last step, which aims to reveal the causal relationships between spatial variables and their path coefficients through SEM in the defined local range. The main objective is to understand the interactions as well as the causal chain of different variables in the local range. First, for each spatial location u, we extract the values of the variables at all data locations v around that location at a distance not exceeding $d_a(X_i)$.

$$\{X_i(v) : \{v : ||u - v|| \leq d_a(X_i)\}\} \tag{5}$$

where $X_i(v)$ denotes the value of the variable X_i at position v, and $||u - v|| \leq d_a(X_i)$ denotes that these positions v should lie within the local range of the center position u.

Next, we constructed SEM with the extracted data within the local range to reveal the local causal relationships between the variables. The corresponding path model equations are constructed. For X_2, the equation is:

$$X_2 = \lambda_{21}(u)X_1 + \epsilon_1 \tag{6}$$

For X_3, the equation is:

$$X_3 = \lambda_{32}(u)X_2 + \lambda_{31}(u)X_1 + \epsilon_2 \tag{7}$$

where the path coefficient $\lambda_{21}(u)$ denotes the effect of variable X_1 on X_2 at position u; $\lambda_{32}(u)$ and $\lambda_{31}(u)$ denote the effect of X_2 and X_1 on X_3, respectively, and the error terms ϵ_1 and ϵ_2 are used to capture the random noise. The above equations can describe the causal chain between the variables on a local scale and thus construct a complete path analysis model.

The path coefficients in the model are then estimated, which reflect the causality and strength of influence between the variables. A commonly used estimation method is maximum likelihood estimation, which obtains the best path coefficients by optimizing the model parameters to minimize the difference between predicted values and observed data. The estimation formula for the path coefficient is:

$$\widehat{\lambda} = \underset{\lambda}{argmin}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{8}$$

where $\widehat{\lambda}$ denotes the estimated path coefficient, y_i is the observed value, $\widehat{y}_i$ is the model predicted value, and n the sample size. The value of the calculated path coefficients can be positive or negative, indicating positive or negative influence, with larger absolute values indicating stronger influence, while revealing the relationship between the direct and indirect effects of the variables on a local scale.

2.3. Local pathway analysis across space

The last stage is to extend the local path analysis to the entire spatial extent, conducting the local path analysis not only at a certain location u but also at multiple locations throughout the study area. This stage can reveal the variations in the relationships between variables across the entire spatial extent. Therefore, the study area is divided into multiple spatial locations $u_1, u_2, \ldots, u_n$, and the same local path analysis as in step 2 is performed at each location. For each location u_i, the data within its local range are extracted, the path model is constructed, and the path coefficients $\lambda_{ij}(u_i)$ are estimated. At each location, the path coefficient $\lambda_{ij}(u_i)$ represents the effect of the variable X_i on X_j, so that the local path coefficients can be obtained at multiple locations over the entire spatial extent. The formula is:

$$X_j = \lambda_{ij}(u_i)X_i + \epsilon_j \tag{9}$$

where $\lambda_{ij}(u_i)$ is the path coefficient of the variable X_i on X_j at location u_i and ϵ_j is the error term.

The path coefficients $\lambda_{ij}(u_i)$ for each spatial location u_i can be obtained after local path analysis for each location. These path coefficients vary with geographic location, reflecting spatial heterogeneity. The spatial distribution of the path coefficients can be analyzed by visualizing how the relationship of the variables changes at different locations through images. In this way, researchers can capture spatial trends and local differences between variables by observing these distribution patterns. The equations represent the path coefficients $\lambda_{ij}(u_i)$ for each location and can be further analyzed for trends in these coefficients:

$$\lambda_{ij}(u_i) = f(u_i) \tag{10}$$

where $f(u_i)$ denotes the path coefficients as a function of spatial location u_i, demonstrating the spatial distribution of path coefficients at different locations.

3. Case study: Exploring local path associations among vegetation, climate, and soil variables on the Tibetan Plateau using the LPA model

3.1. Background and study area

The Tibetan Plateau, as the world's highest and largest plateau-referred to as the "Roof of the World" and the "Third Pole"-holds significant importance in global climate change and ecological environment research (Favre et al., 2015; Yao et al., 2000). Fig. 2 shows that Vegetation indices (such as NDVI) are crucial indicators for assessing vegetation growth conditions and ecosystem health (Liu et al., 2017). Understanding the relationships between vegetation and climate factors (surface temperature, precipitation) and soil factors (soil moisture) is essential for predicting the impacts of climate change on ecosystems and formulating environmental protection strategies (Liu et al., 2022; Wang et al., 2007). However, due to the complex terrain and diverse ecosystems of the Tibetan Plateau, traditional global models struggle to capture the spatial heterogeneity among variables (Jia et al., 2023; Wang et al., 2024). Therefore, this study employs a local pathway association (LPA) model to explore the local association pathways between these variables, aiding in a deeper understanding of ecological processes and spatial variation characteristics within the region.

The LPA model is implemented to explore the local pathways of spatial association between Explanatory variables of the Tibetan Plateau, located in southwestern China. The plateau features a complex and diverse topography, with an average elevation of 4500 m and a total area of approximately 2.5 million km^2 (Pan et al., 2012; Xuan-xue, 2010), encompassing various ecosystems such as mountains, grasslands, wetlands, and glaciers (Xia et al., 2021). Climate types range from temperate monsoon climates in the southeast to cold arid plateau

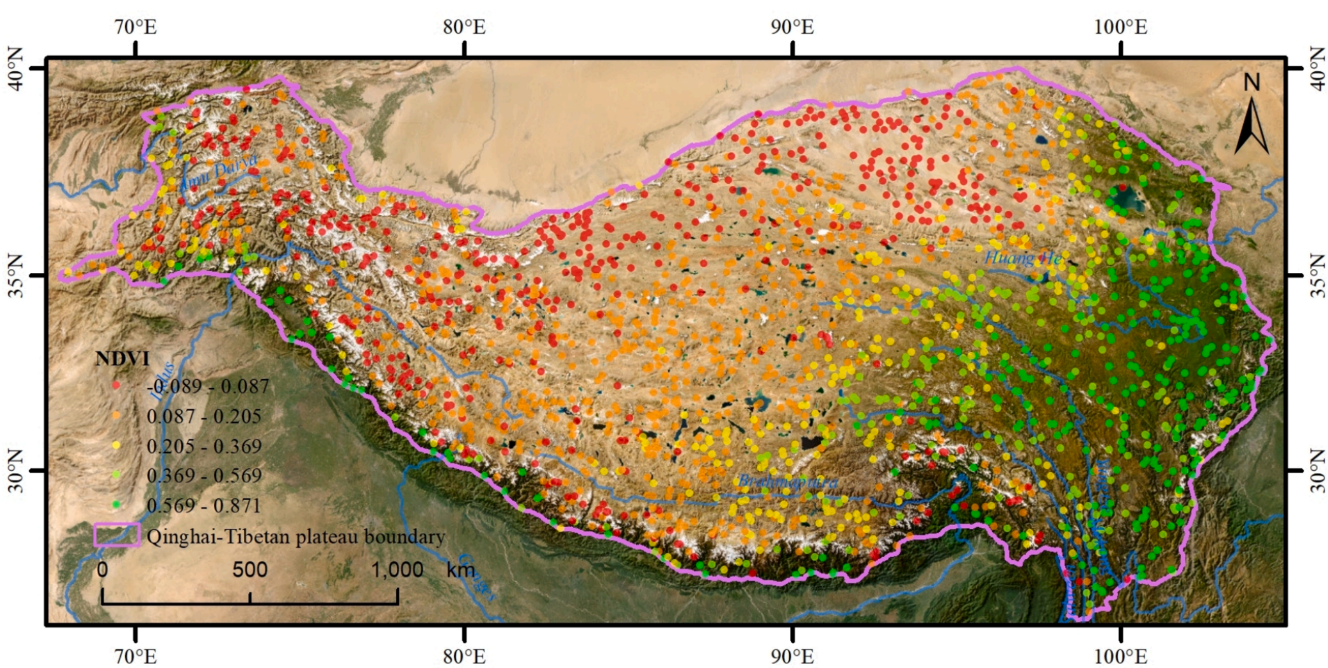

Fig. 2. Geographic location of the Tibetan Plateau for analyzing path associations among local variables. Points in the figure indicate the distribution of the normalized difference vegetation index (NDVI).

climates in the northwest, exhibiting significant spatial differences (Xu and Liu, 2007). The Tibetan Plateau is the source of several major rivers in Asia and has a significant impact on water resource supply and ecological environments in downstream regions (Jianyun et al., 2019). Selecting this area for research helps reveal the complex relationships among vegetation, climate, and soil factors, deepens the understanding of plateau ecosystems, and provides a scientific basis for regional ecological environment protection and sustainable development.

3.2. Datasets

Potential explanatory variable data were collected to explore the local correlations among vegetation, climate, and soil factors on the Tibetan Plateau. Table 1 provides detailed information on these datasets. The normalized difference vegetation index (NDVI) data for the Tibetan Plateau from 2000 to 2020 were obtained from the MODIS MOD13A1 product (https://doi.org/10.5067/MODIS/MOD13A1.061), with a spatial resolution of 500 m. Climate factors include LST and annual cumulative precipitation, all sourced from the ERA5-Land dataset (https://doi.org/10.24381/cds.68d2bb30). Using Google Earth Engine (GEE), we performed linear regression analyses on LST and annual cumulative precipitation to determine their trends from 2000 to 2020. Soil characteristics are represented by soil moisture content (volumetric percentage) at 33 kPa suction (Hengl et al., 2017), predicted at a standard depth of 10 cm with a resolution of 250 m, with data obtained from https://doi.org/10.5281/zenodo.2629589.

3.3. Experiment design

Fig. 3 depicts the flowchart of local association path analysis among vegetation, climate, and soil explanatory variables in the Tibetan Plateau based on the LPA model, including four main steps. The first step was data preprocessing of the explanatory variables. The second step was to conduct a local determinant analysis using the LPA model. The third step is to evaluate the results of the path analysis, including checking the statistical characteristics, spatial distribution pattern, and significance of the path coefficients. Finally, the model was validated to assess its accuracy and performance.

3.3.1. Data pre-processing

Data preprocessing consists of two steps. First, 2,000 sampling data points were randomly generated within the Tibetan Plateau, with more than 2 km between each pair of sampling data points. Second, observations of precipitation, surface temperature, soil moisture, and NDVI were collected for each sampling point using Google Earth Engine. Data with different temporal or spatial resolutions were synchronized to the same time step and spatial unit through interpolation or aggregation. To minimize the impact of outliers, a 2 km buffer zone was established around each sampling point, and the median value within the buffer zone was used as the observation for the sampling point, ensuring that the dataset accurately represents real-world observations. The SEM can automatically compute the values of latent variables (unobservable variables) based on the observable variables mentioned above, using a subjectively designed model structure. Therefore, manual computation is not required.

Table 1
Explanatory variables for the construction of local structural relationships in the case study.

Latent variable	Variable	Data	Period	Spatial resolution
Plant	NDVI	MOD13A1	2000–2020	500 *m*
Climate	LST	ERA5-Land	2000–2020	11 km
	Precipitation	ERA5-Land	2000–2020	11 km
Soil	Water content	OpenLandMap	average value from1950 to 2017	250 *m*

3.3.2. Implementing the LPA model in exploring local pathways of association

This section employs the local pathway association (LPA) model to investigate the local association pathways among variables, comprising three main steps. First, the LPA model is applied in conjunction with a spatial decay function to calculate the local ranges of each variable, selecting the maximum among them as the optimal overall range. Second, after extracting data within this optimal local range, a Structural Equation Model (SEM) is constructed by specifying observed variables and hypothesized path relationships, thereby determining the dominant factors and variable association pathways within each region. Finally, the path coefficients λ_k between variables are calculated within the local range, and their spatial heterogeneity is analyzed.

3.3.3. Pathways assessment and significance test

At the pathway evaluation stage, we analyze the results of the LPA model, including the statistical characteristics of path coefficients, spatial patterns, and significance testing. First, we perform statistical analysis on the local path coefficients calculated for each location, examining their spatial distribution characteristics and computing the statistical density of the path coefficients to assess the overall strength of relationships among variables. Second, we visualize the spatial distribution of the local path coefficients to reveal the spatial heterogeneity of variable association pathways across different regions. Finally, through significance testing, we determine whether the local path coefficients are statistically significant.

3.3.4. Model evaluation

To assess the proposed LPA model's performance, we developed a series of metrics to compare with the traditional structural equation model (SEM). These metrics include the λ values of each explanatory variable, their interactions, and a composite λ value applicable to all variables. The SEM is employed to analyze and validate the complex relationships among variables, particularly suitable for investigating the mutual influences between multiple latent and observed variables (Hoyle, 1995).

4. Results

4.1. Identifying the optimal local range

In this study, Ripley's K function is used to analyze the aggregation of spatial data and determine the optimal local range of point distribution in the study area. The comparison between the actual observed K value and the expected K value is shown in Fig. 4, which indicates that the distribution of data is close to a random distribution in a smaller distance range, and the clustering of data increases gradually with the increase of distance. The L function reaches its maximum value at a distance of 707.29 km, which indicates that the clustering of data is most significant in this distance range. This distance value reflects the most significant spatial interactions between data within a localized area to determine the optimal local sphere of influence.

4.2. Determine the optimal local range

In LPA, the λ values are used to quantify the path coefficients of causal relationships between variables. Fig. 5 displays the influence coefficients (λ values) among various observable variables. The results indicate that the LPA model effectively elucidates the spatial variations of λ values for each explanatory variable, exhibiting significant spatial heterogeneity. Among these variables, NDVI shows the strongest expression of influence, with higher λ values in the northern and southwestern regions of the Tibetan Plateau and lower λ values in some southeastern and central areas. The influence expression of land surface temperature (LST) on climate is also strong, with the highest values distributed in the central and southern regions. The influence of climate

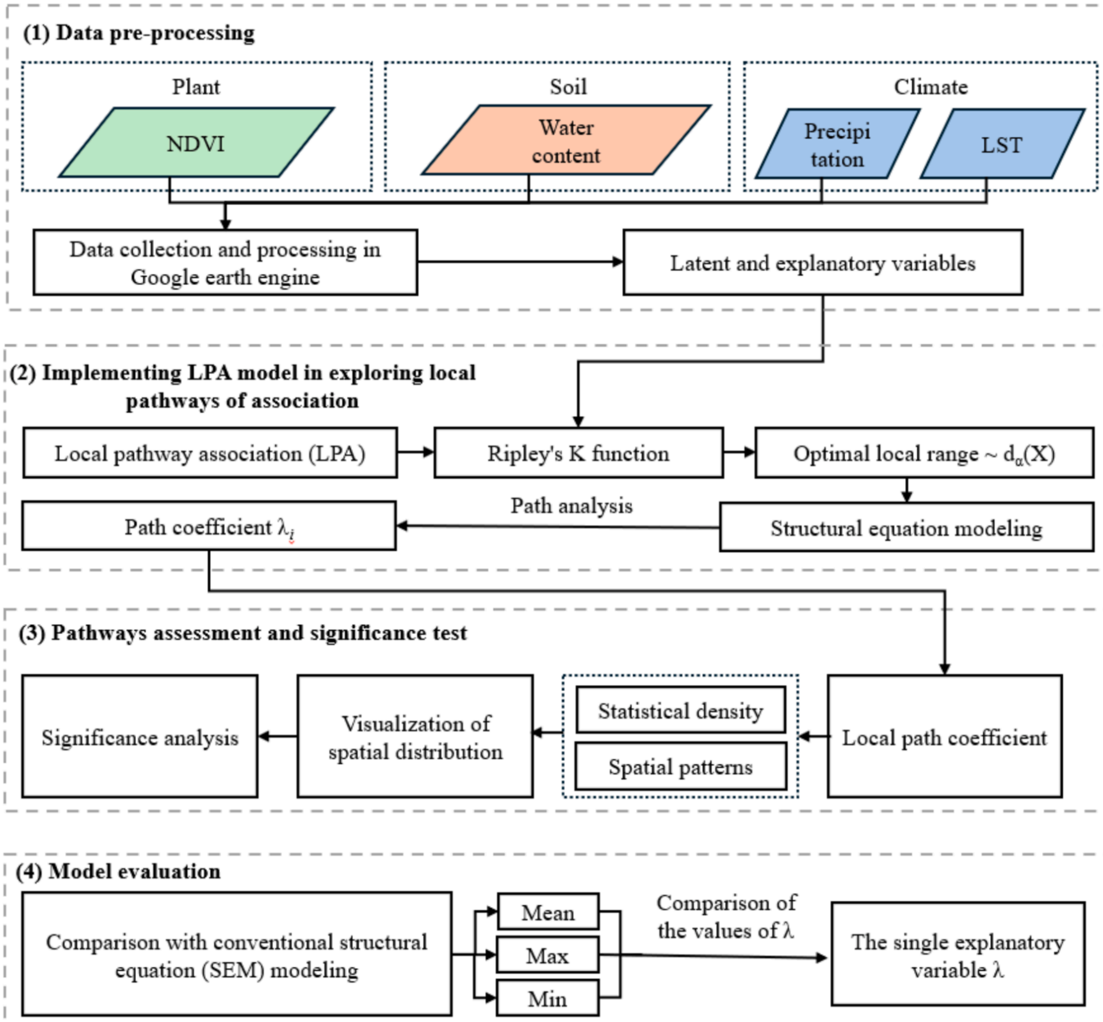

Fig. 3. Schematic diagram exploring localized pathways between vegetation, climate, and soil factors on the Tibetan Plateau using the localized path analysis model.

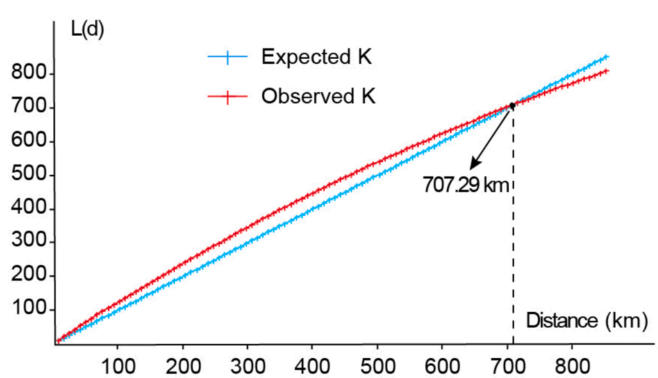

Fig. 4. Comparison of observed and expected K functions for determining the optimal local range. Note: normalized L-function curves with actual observed K-values in red and desired K-values in blue.

on vegetation is relatively strong, with the highest λ values in the western regions, lower values in the north, and significant differences in the southern regions. The influence of climate on soil is also strong, with notable differences in most central areas and lower λ values in the western regions. The influence of soil moisture content on soil changes is the weakest, which can be attributed to the low and uneven precipitation in the Tibetan Plateau region, along with the alpine environment causing frequent freeze–thaw processes in the soil. These freeze–thaw cycles lead to the easy loss of soil water, reducing the influence of soil moisture content on soil factors.

Fig. 5 presents the significance test results of the λ values between various variables. The causal effect of NDVI on vegetation is significant in most regions; except for parts of the Qiangtang Nature Reserve and the Kunlun Mountains, over 98.21 % of the areas have significance levels below 0.05, mainly because the vegetation in these areas is relatively sparse. In addition, the significance levels of LST on climate, water content on soil, and precipitation on climate are also below 0.05 in most regions, with only some areas in the southeastern part of the study area showing non-significance. The regions where climate affects plants and soil affects vegetation with higher significance are similar, mainly concentrated in certain central areas. The overall importance of climate on soil is relatively low, possibly because the impact of climate on soil is often indirect or may act through other intermediary variables. For example, climate influences precipitation and temperature, which in turn indirectly affect the physical and chemical properties of the soil by regulating vegetation, water cycles, and surface temperature.

4.3. Model evaluation

A comprehensive analysis of interactions among variables and the path coefficients of causal relationships demonstrates the effectiveness of the LPA model in revealing local path relationships among climate, vegetation, and soil variables on the Tibetan Plateau. Table 2 presents

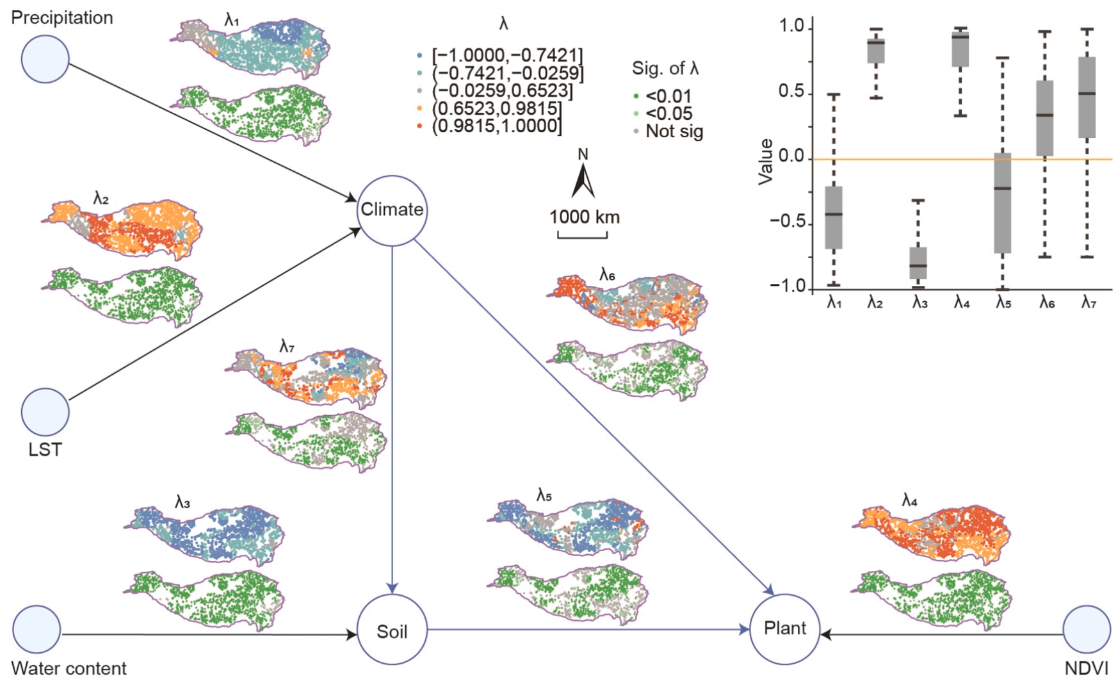

Fig. 5. Local path causality with significance test and spatial distributions among variables. Black arrows represent the regression of variable and latent variables, blue arrows represent the regression of two latent variables. The first map of the local path coefficients (λ) represents the spatial distribution map, while the second map illustrates the significance test results.

Table 2
Model validation through the comparison of local pathways of spatial associations identified by the LPA model and global λ values.

	Paths	Local λ of LPA mean [min, max]	Percentage of locations with significant LPA ($p < 0.05$)	λ of global structural equations
Variable relationship	NDVI =~ Plant	0.773 [0.004, 1.000]	98.21 %	0.995**
	LST =~ Climate	0.798 [-0.605, 0.999]	98.29 %	0.858**
	Precipitation =~ Climate	−0.370 [-0.967, 0.747]	94.39 %	−0.332**
	Water Content =~ Soil	−0.763 [-0.987, 0.430]	96.80 %	−0.895**
Latent variable relationships	Climate ~ Soil	0.427 [-0.969, 0.997]	62.35 %	0.677**
	Climate ~ Plant	0.335 [-0.994, 0.998]	63.52 %	0.267**
	Soil ~ Plant	−0.269 [-0.998, 0.774]	70.07 %	−0.528**

The regression between exogenous variable and endogenous variable in the structural equation model (SEM): =~, the regression between two latent variables. in SEM.: ~.
Significance levels: ** $p < 0.01$, * $p < 0.05$.

the statistical results of local λ values calculated by the LPA model. The statistical summaries of local λ values derived from the LPA model reveal the spatial heterogeneity of path associations and the local explanatory power of variables.

The LPA model effectively captures spatial variability and localized heterogeneity in the path coefficients of causal relationships. For instance, the local λ values representing NDVI's influence on Plant range from 0.004 to 1.000, indicating substantial spatial heterogeneity in this interaction. The identified variability shows the model's ability to examine regional differences in variable relationships and understand complex ecological and environmental systems, where interactions can vary significantly based on local conditions.

In addition, the LPA model effectively identifies significant relationships across diverse spatial locations. For example, more than 98 % of locations show significant associations for pathways such as NDVI to Plant and LST to Climate, demonstrating the model's robustness in detecting geographically significant interactions. LPA model enhances the explanatory power and reveals essential spatial variations by capturing localized influences that traditional global models often ignore. LPA model also adapts well to examining interrelationships among latent variables, such as climate-to-soil, climate-to-plant, and soil-to-plant interactions. For example, regional variations in vegetation and soil factors require models that can include these differences, which can be properly addressed by the LPA model.

Finally, the LPA model also has an advantage in capturing localized path relationships. By providing refined estimates of path coefficients and revealing significantly varied spatial interactions, LPA model overcomes the limitations of traditional global models that are often ignore spatially varied and complex variable interactions. In summary, the LPA model has strengths in addressing spatial heterogeneity, adapting to localized conditions, and identifying significant variable interactions across diverse environmental contexts. These strengths establish it as a powerful e tool for advancing the understanding of complex geospatial and Earth systems.

5. Discussion

5.1. Addressing local variability in spatial causal relationships

This study develops a local pathway association (LPA) model for analyzing local causal relationships among variables. The model demonstrates significant advantages in examining the influence coefficients of local variations and visualizing the spatial distribution of causal relationships, and it holds potential for applications across multiple fields. Path analysis, based on directed acyclic graph models, represents causal relationships between variables through path diagrams and quantifies the strength of these causal relationships using λ coefficients (Land, 1969; Lleras, 2005). In subsequent research, path analysis has been extended into the structural equation model (SEM), which not only handles multiple dependent variables but also simultaneously addresses measurement errors, latent variables, and complex causal structures (Grapentine, 2000; Hoyle, 2012; Mitchell, 1992). However, path coefficients in spatial associations may be influenced by local regions, leading to significant errors in λ values between variables, and the same set of variable paths may exhibit different λ values in different regions. Therefore, the LPA model constructed in this study can effectively address the impact of local regions on spatial association path coefficients.

5.2. Advantages of the LPA model in spatial association analysis

The LPA model has numerous advantages in spatial association path analysis. First, the LPA model possesses dynamic adaptability to local ranges. Compared to traditional models that typically use global path coefficients and cannot accommodate differences of variables in different spatial locations, the LPA model captures local heterogeneity, allowing it to more accurately reflect the influence relationships of variables within local areas and avoid biases caused by the global model's neglect of spatial differences. Second, the LPA model enables more refined spatial analysis. Traditional models cannot handle local variations in spatial data and usually yield only an averaged global conclusion (Bivand and Wong, 2018; Seya, 2020). In contrast, the LPA model can provide precise path coefficients for each location, revealing spatial heterogeneity and local features, thereby making the analysis more interpretable. Third, the path coefficients in the LPA model are calculated individually at each location, allowing the model to handle spatial nonlinearity and heterogeneity. Traditional SEM models assume that relationships between variables are linear and fixed across the entire study area, making it difficult to address variable relationships with significant spatial variation (Asparouhov and Muthén, 2009; Rose et al., 2017; Tomarken and Waller, 2005). The LPA model developed in this study demonstrates the effectiveness of local approaches in capturing complex spatial relationships, as its localized processing enables it to address both the spatial dependencies and the local complexity within these relationships (Zhang et al., 2023). Fourth, the LPA model not only provides path coefficients for each local range but also visualizes the changes in these local causal relationships through spatial distribution maps. The model results can more intuitively help understand the spatial dependency and local effects among variables. Finally, by introducing local ranges, the LPA model makes the path coefficients more locally interpretable, which has significant advantages in explaining spatial heterogeneity and handling variable relationships in different geographical regions. In addition, the prediction accuracy of the LPA model is higher than that of traditional models as it can adaptively adjust based on local data. Finally, the LPA model generates path coefficient maps for visualizing spatially varying associations, enabling the analysis of spatial associations and co-location patterns, such as those between POIs and road networks, while effectively capturing spatial heterogeneity.

5.3. Limitations and future directions of the LPA model

However, this study still has certain limitations. First, the selection of local ranges in the LPA model often relies on some prior knowledge or subjective judgment. Consequently, choosing different local ranges may lead to variations in the results. Second, is the issue of interdependence among variables. Although the LPA model can capture local causal relationships, when dealing with multiple variables, the interdependence among variables and potential multicollinearity problems may affect the accuracy of the path coefficients. Particularly in local regions, changes in certain variables may be influenced by the combined effects of multiple other variables, which reduces the applicability of the LPA model due to this complexity. Therefore, future research should further consider the determination of local ranges, dynamically establishing the optimal local ranges based on the spatial distribution of data and the variation characteristics of variables to reduce the influence of subjective judgment on the results. Meanwhile, methods that can better handle multivariate dependencies and multicollinearity issues should be adopted to enhance the accuracy of the LPA model in dealing with complex variable dependencies, thereby reducing path coefficient biases caused by collinearity. Finally, the LPA assumes spatial homogeneity in factor pathways across local regions; however, varying SEM structures within local regions are likely to exist in real-world scenarios and need further investigation in future studies.

6. Conclusions

This study develops a local pathway association (LPA) model to explore localized causal relationships between variables. By utilizing Ripley's K function, the LPA model effectively determines the optimal local range for each variable, enabling precise capture of spatial heterogeneity and local interactions. Compared to traditional global models, the LPA model addresses the limitations in handling spatial heterogeneity and complex interactions. Although the model has some limitations in dealing with high-dimensional data and irregular terrains, its applicability and robustness can be enhanced in the future by integrating machine learning techniques and extending it to temporal dynamic analysis. Overall, the LPA model provides a powerful framework for understanding localized causal mechanisms and spatial dependencies. Its adaptability and broad applicability make it particularly suitable for large-scale spatial studies, aiding in uncovering local drivers of change and assessing spatial relationships.

CRediT authorship contribution statement

Jiao Hu: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rui Qu:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Yongze Song:** Writing – review & editing, Supervision, Conceptualization. **Peng Wu:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Data and codes supporting the findings of this study are available at https://figshare.com/articles/dataset/Local_pathways_of_association/27187623.

References

Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27, 93–115.
Asparouhov, T., Muthén, B., 2009. Exploratory structural equation modeling. Struct. Equ. Model. Multidiscip. J. 16, 397–438.
Bai, C.Q., Zhou, L., Xia, M.L., Feng, C., 2020. Analysis of the spatial association network structure of China's transportation carbon emissions and its driving factors. J. Environ. Manage. 253.

Bivand, R.S., Wong, D.W., 2018. Comparing implementations of global and local indicators of spatial association. TEST 27, 716–748.
Caldas de Castro, M., Singer, B.H., 2006. Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. Geogr. Anal. 38, 180–208.
Chen, W.X., Chi, G.Q., Li, J.F., 2019. The spatial association of ecosystem services with land use and land cover change at the county level in China, 1995-2015. Sci. Total Environ. 669, 459–470.
Coker, E.S., Cavalli, L., Fabrizi, E., Guastella, G., Lippo, E., Parisi, M.L., et al., 2020. The Effects of Air Pollution on COVID-19 Related Mortality in Northern Italy. Environ. Resour. Econ. 76, 611–634.
Crawford, T.W., 2009. Scale Analytical. In: Kitchin, R., Thrift, N. (Eds.), International Encyclopedia of Human Geography. Elsevier, Oxford, pp. 29–36.
De Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. Hydrgeol. J. 13, 161–183.
Favre, A., Päckert, M., Pauls, S.U., Jähnig, S.C., Uhl, D., Michalak, I., et al., 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. Biol. Rev. 90, 236–253.
Getis, A., 2001. Spatial Association, Measures of. In: Smelser, N.J., Baltes, P.B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Pergamon, Oxford, pp. 14758–14763.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206.
Goodchild, M., Haining, R., Wise, S., 1992. Integrating GIS and spatial data analysis: problems and possibilities. Int. J. Geogr. Inf. Syst. 6, 407–423.
Grapentine, T., 2000. Path analysis vs. structural equation modeling. Mark. Res. 12.
Hasanloo, M., Pahlavani, P., Bigdeli, B., 2019. Flood risk zonation using a multi-criteria spatial group fuzzy-AHP decision making and fuzzy overlay analysis. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 42, 455–460.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12, e0169748.
Hoyle, R.H., 1995. Structural equation modeling: Concepts, issues, and applications. Sage.
Hoyle RH. Path analysis and structural equation modeling with latent variables. 2012.
Hu, J., Song, Y., Zhang, T., 2025 Apr 3. A local indicator of stratified power. Int. J. Geogr. Inf. Sci. 39 (4), 925–943.
Jia, Z., Wang, X., Feng, X., Ma, J., Wang, X., Zhang, X., et al., 2023. Exploring the spatial heterogeneity of ecosystem services and influencing factors on the Qinghai Tibet Plateau. Ecol. Ind. 154, 110521.
Jianyun, Z., Jiufu, L., Junliang, J., Tao, M., Guoqing, W., Hongwei, L., et al., 2019. Evolution and trend of water resources in Qinghai-Tibet Plateau. Bulletin of Chinese Academy of Sciences (chinese Version) 34, 1264–1273.
Jöreskog, K.G., Sörbom, D., 1982. Recent developments in structural equation modeling. J. Mark. Res. 19, 404–416.
Land, K.C., 1969. Principles of path analysis. Sociol. Methodol. 1, 3–37.
Liu, J., Xin, Z., Huang, Y., Yu, J., 2022. Climate suitability assessment on the Qinghai-Tibet Plateau. Sci. Total Environ. 816, 151653.
Liu, S., Cheng, F., Dong, S., Zhao, H., Hou, X., Wu, X., 2017. Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI. Sci. Rep. 7, 4182.
Lleras, C., 2005. Path analysis. Encyclopedia of Social. Measurement 3, 25–30.
Luo, P., Song, Y., Zhu, D., Cheng, J., Meng, L., 2023 Mar 4. A generalized heterogeneity model for spatial interpolation. Int. J. Geogr. Inf. Sci. 37 (3), 634–659.
Majumder, S., Roy, S., Bose, A., Chowdhury, I.R., 2023. Multiscale GIS based-model to assess urban social vulnerability and associated risk: Evidence from 146 urban centers of Eastern India. Sustain. Cities Soc. 96.
Mitchell, R.J., 1992. Testing evolutionary and ecological hypotheses using path analysis and structural equation modelling. Funct. Ecol. 123–129.
Mueller, R.O., Hancock, G.R., 2018. Structural equation modeling. The reviewer's guide to quantitative methods in the social sciences. Routledge 445–456.
Ord, J.K., 2024. Art Getis and local spatial statistics. J. Geogr. Syst. 26, 191–200.
Pan, G., Wang, L., Li, R., Yuan, S., Ji, W., Yin, F., et al., 2012. Tectonic evolution of the Qinghai-Tibet plateau. J. Asian Earth Sci. 53, 3–14.
Perry, J.N., Dixon, P.M., 2002. A new method to measure spatial association for ecological count data. Ecosci. 9, 133–141.
Pickett, S.T., Cadenasso, M.L., 1995. Landscape ecology: spatial heterogeneity in ecological systems. Science 269, 331–334.
Raftery, A.E., 1993. Bayesian model selection in structural equation models. Sage Focus Editions 154, 163.
Rose, S.A., Markman, B., Sawilowsky, S., 2017. Limitations in the systematic analysis of structural equation model fit indices. J. Mod. Appl. Stat. Methods 16, 5.
Seya, H., 2020. Global and local indicators of spatial associations. Spatial analysis using big data. Elsevier 33–56.
Shi, W., Cheung, C.-K., Tong, X., 2004. Modelling error propagation in vector-based overlay analysis. ISPRS J. Photogramm. Remote Sens. 59, 47–59.
Song, Y., Wu, P., 2021. An interactive detector for spatial associations. Int. J. Geogr. Inf. Sci. 35, 1676–1701.
Song, Y., 2022 Jul. The second dimension of spatial association. Int. J. Appl. Earth Obs. Geoinf. 1 (111), 102834.
Song, Y., 2023 Apr. Geographically optimal similarity. Math. Geosci. 55 (3), 295–320.
Stein, A., Gerstner, K., Kreft, H., 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecol. Lett. 17, 866–880.
Tomarken, A.J., Waller, N.G., 2005. Structural equation modeling: Strengths, limitations, and misconceptions. Annu. Rev. Clin. Psychol. 1, 31–65.
Unwin, D., 2019. Integration through overlay analysis. Spatial analytical perspectives on GIS. Routledge 127–138.
Varouchakis, E.A., 2019. 1 - Geostatistics: Mathematical and Statistical Basis. In: Corzo, G., Varouchakis, E.A. (Eds.), Spatiotemporal Analysis of Extreme Hydrological Events. Elsevier, pp. 1–38.
Wang, G., Wang, Y., Li, Y., Cheng, H., 2007. Influences of alpine ecosystem responses to climatic change on soil properties on the Qinghai–Tibet Plateau. China. Catena 70, 506–514.
Wang, S., Zhong, E., Lu, H., Guo, H., Long, L., 2015. An effective algorithm for lines and polygons overlay analysis using uniform spatial grid indexing. In: 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), pp. 175–179.
Wang, Y., Lü, Y., Lü, D., Yin, L., Wang, X., 2024. Climate change and its ecological risks are spatially heterogeneous in high-altitude region: The case of Qinghai-Tibet plateau. Catena 243, 108140.
Xia, M., Jia, K., Zhao, W., Liu, S., Wei, X., Wang, B., 2021. Spatio-temporal changes of ecological vulnerability across the Qinghai-Tibetan Plateau. Ecol. Ind. 123, 107274.
Xu, H.C., Li, Y.L., Zheng, Y.J., Xu, X.B., 2022. Analysis of spatial associations in the energy-carbon emission efficiency of the transportation industry and its influencing factors: Evidence from China. Environ. Impact Assess. Rev. 97.
Xu, W., Liu, X., 2007. Response of vegetation in the Qinghai-Tibet Plateau to global warming. Chin. Geogr. Sci. 17, 151–159.
Xuan-xue, M., 2010. A review and prospect of geological researches on the Qinghai-Tibet Plateau. Geol. China 37, 841–853.
Yao, T., Liu, X., Wang, N., Shi, Y., 2000. Amplitude of climatic changes in Qinghai-Tibetan Plateau. Chin. Sci. Bull. 45, 1236–1243.
Yao, Y., Pan, J.H., Wang, W.D., Liu, Z.X., Kan, H.D., Qiu, Y., et al., 2020. Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. Sci. Total Environ. 741.
Yuan, K.-H., Bentler, P.M., 2006. 10 structural equation modeling. Handbook of Statist. 26, 297–358.
Zhang, Z., Song, Y., Luo, P., Wu, P., 2023 Jul 3. Geocomplexity explains spatial errors. Int. J. Geogr. Inf. Sci. 37 (7), 1449–1469.
Zhang, Z., Song, Y., Karunaratne, L., Wu, P., 2024a. Robust interaction detector: a case of road life expectancy analysis. Spatial Stat. 59, 100814.
Zhang, Z., Li, Z., Song, Y., 2024b. On ignoring the heterogeneity in spatial autocorrelation: consequences and solutions. Int. J. Geogr. Inf. Sci. 38 (12), 2545–2571.