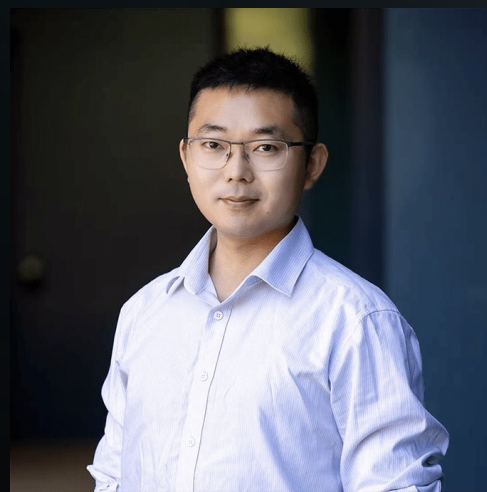


Geospatial Data Analysis, Prediction, and GeoAI: New Theories, Methods, and Software

Topic 2. Spatial modelling for heterogeneity and determinant analysis

Lecture website: <https://yongzesong.com/lecture-20260104>



Yongze Song, Associate Professor

School of Design and the Built Environment, Curtin University, Australia
RGS Fellow (UK), DAAD Alnet Fellow (Germany), SDL Fellow (Harvard University, US)

Email: Yongze.song@curtin.edu.au

Associate Editor: *International Journal of Earth Observation and Geoinformation* (IF 8.6, Q1)

Associate Editor: *GIScience & Remote Sensing* (IF 6.9, Q1)

Topic Editor: *Geoscientific Model Development* (IF 4.9, Q1)

Lecture outline

Lecture website: <https://yongzesong.com/lecture-20260104>

Password: 20260104

Contents

Part 1. Lecture (1st class)

1. Concepts of heterogeneity
2. Review of OPGD model
3. A spatial learning model: GOZH
4. Local driver analysis: LISP

Part 2. Practice (2nd to 4th classes)

1. Feedback session for Assignment 1 (20 min)
2. Practice on GOZH and LISP (40 min)
3. Assignment 2 (40 min)
4. Submit your assignment (10 min)

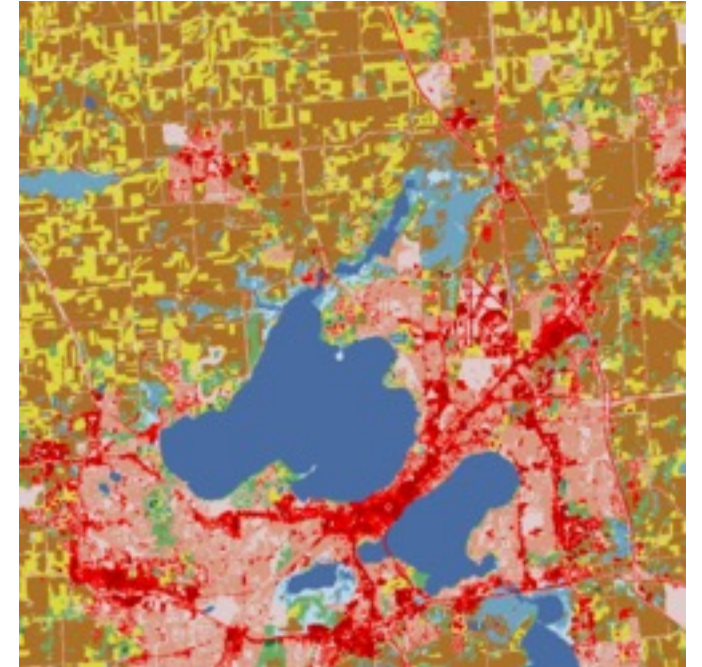
Tips:

1. Working in a group of 3 or 4
2. Collecting feedback at each day
3. Individual assignments and working in a group
4. Complete assignments in the class
5. Assignment feedback sections after day 2
6. Install R and Rstudio
7. Sign Up for Google Earth Engine

Concepts of heterogeneity

Spatial heterogeneity is a property generally ascribed to a landscape or to a population. It refers to the uneven distribution of various concentrations of each species within an area. A landscape with spatial heterogeneity has a mix of concentrations of multiple species of plants or animals (biological), or of terrain formations (geological), or environmental characteristics (e.g. rainfall, temperature, wind) filling its area. A population showing spatial heterogeneity is one where various concentrations of individuals of this species are unevenly distributed across an area; nearly synonymous with "patchily distributed."

Source: https://en.wikipedia.org/wiki/Spatial_heterogeneity



Stratified heterogeneity

There exist two main types of spatial heterogeneity.

- The **spatial local heterogeneity** categorises the geographic phenomena whose attributes' values are significantly similar within a directly local neighbourhood, but which significantly differ in the nearby surrounding-areas beyond this directly local neighbourhood (e.g. hot spots, cold spots).
- The **spatial stratified heterogeneity** categorises the geographic phenomena where the within-strata variance of its attributes' values is significantly lower than its between-strata variance, such as collections of ecological zones or land-use classes within a given geographic area depict.

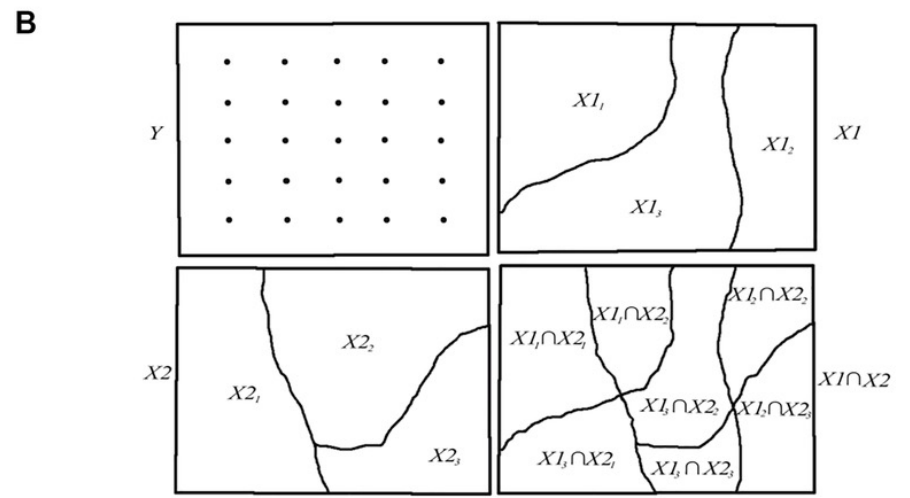
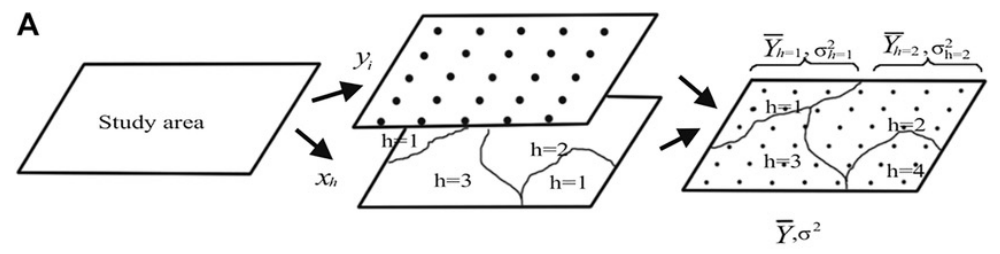
Source:

https://en.wikipedia.org/wiki/Spatial_heterogeneity

y

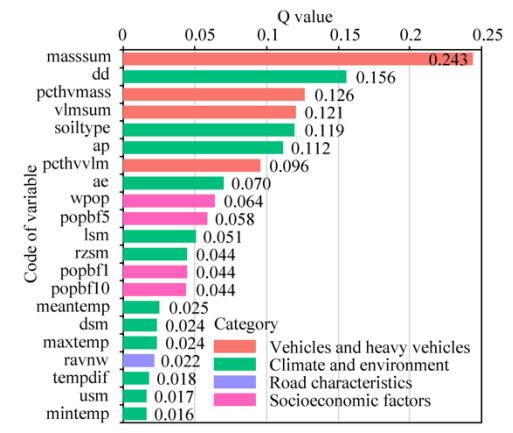
Spatial stratified heterogeneity

$$q = 1 - \frac{1}{N\sigma^2} \sum_{h=1}^L N_h \sigma_h^2$$



Optimal parameters-based geographical detectors (OPGD) model

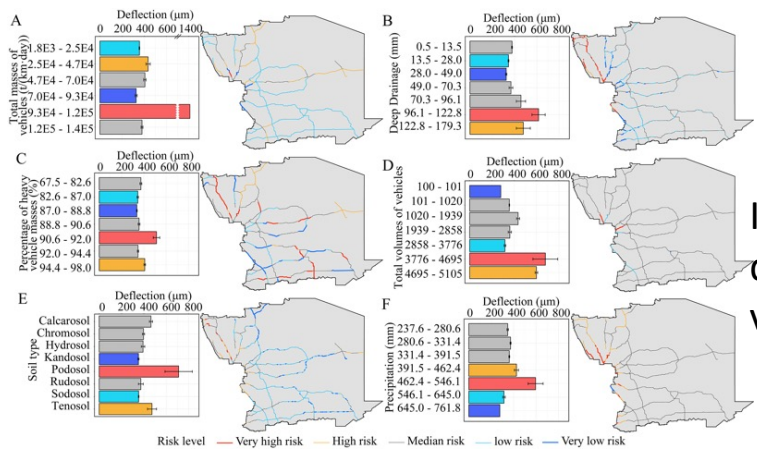
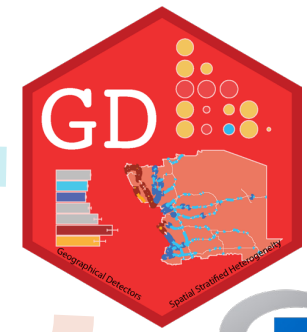
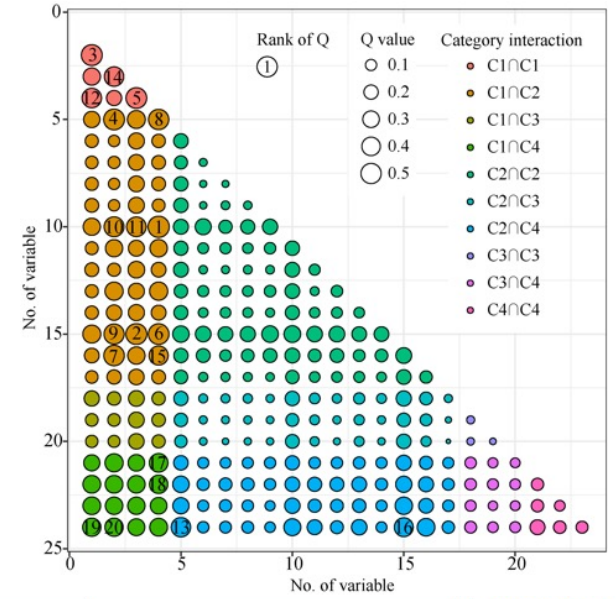
Identifying geographical, environmental, traffic, and pavement factors of road performance.



Impacts of individual variables

$$Q_v = 1 - \frac{\sum_{j=1}^M N_{v,j} \sigma_{v,j}^2}{N_v \sigma_v^2}$$

Impacts of variable interactions



Impacts in different data intervals of variables

A "GD" R package is developed for the OPGD model. The latest version in CRAN is **Version 10**.

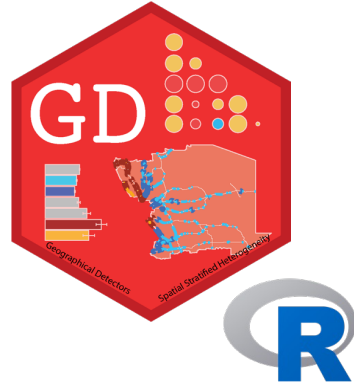
Category	C1: Vehicles and heavy vehicles				C2: Climate and environment		
	Variable	vlmsum	pcthvml	masssum	pcthvms	ap	dd
C1: Vehicles and heavy vehicles	vlmsum						
	pcthvml	0.543*					
	masssum		0.451				
	pcthvms	0.457		0.532			
C2: Climate and environment	ap		0.535		0.505		
	soiltype		0.481	0.464	0.566*		
	dd		0.496	0.544	0.525		
	ae		0.516		0.435		
C4: Socioeconomic factors	popbf1				0.407*		
	popbf5				0.393		
	wpop	0.385	0.384			0.453*	0.423

* The largest interaction within each category, including C1∩C1, C1∩C2, C1∩C4, and C2∩C4

Song, Yongze, et al. 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*, 2020. 57(5): 593-610.

Song, Yongze, et al. Segment-Based Spatial Analysis for Assessing Road Infrastructure Performance Using Monitoring Observations and Remote Sensing Data. *Remote Sensing*, 2018. 10(11): 1696.

"GD" R package (Version 10)



GD v10.8 downloads 102K downloads 3102/month

<https://cran.r-project.org/web/packages/GD/vignettes/GD.html>

Start from the *one-step function*: `gdm` (*Highly Recommended*)

The package provides a one-step function for performing optimal discretization and geographical detectors at the same time.

The output contains all data and visualization results.

```
## install and library the package
install.packages("GD")
library("GD")

## Example 1
## NDVI: ndvi_40
## set optional parameters of optimal discretization
## optional methods: equal, natural, quantile, geometric, sd and manual
discmethod <- c("equal", "natural", "quantile")
discitv <- c(4:6)
## "gdm" function
## In this case, Climatezone and Mining are categorical variables,
## and Tempchange and GDP are continuous variables.
ndvigdm <- gdm(NDVIchange ~ Climatezone + Mining + Tempchange + GDP,
              continuous_variable = c("Tempchange", "GDP"),
              data = ndvi_40,
              discmethod = discmethod, discitv = discitv) # ~3s

ndvigdm
plot(ndvigdm)
```

Song, Yongze, et al. 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*, 2020. 57(5): 593-610.

Song, Yongze, et al. Segment-Based Spatial Analysis for Assessing Road Infrastructure Performance Using Monitoring Observations and Remote Sensing Data. *Remote Sensing*, 2018. 10(11): 1696.

Geographical Optimal Zones-based Heterogeneity (GOZH) model

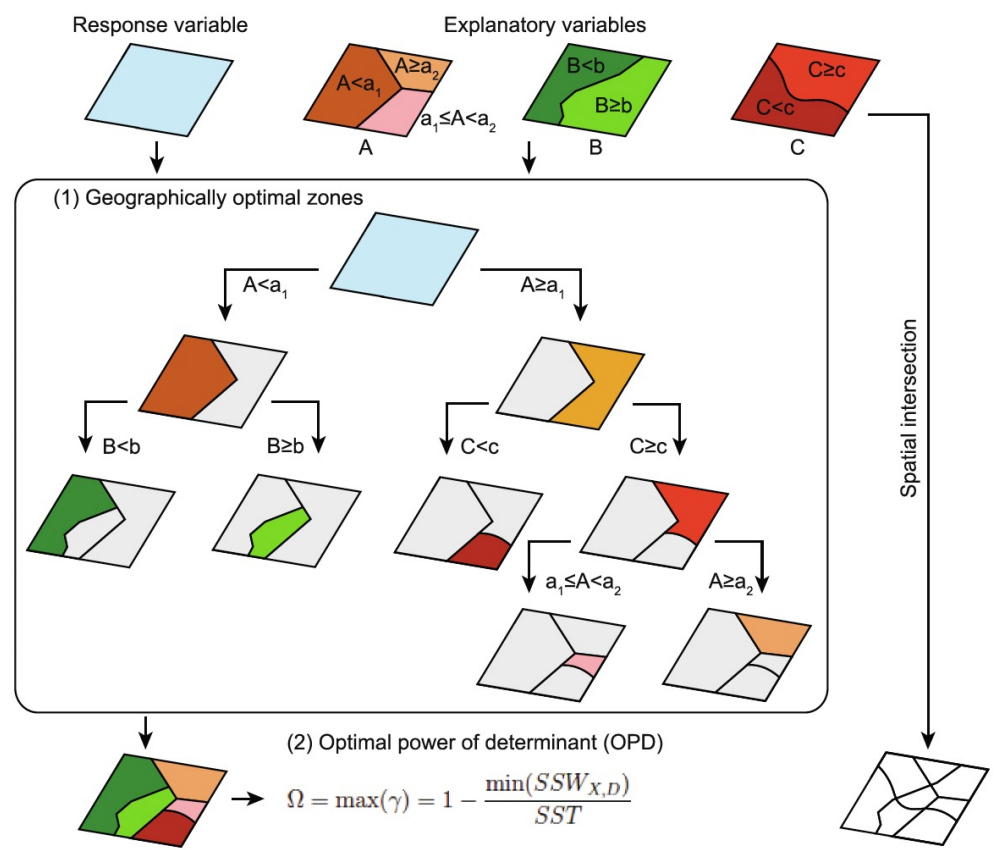


Fig. 1. Process and principle of the geographically optimal zones-based heterogeneity (GOZH) model.

$$q = 1 - \frac{1}{N\sigma^2} \sum_{h=1}^L N_h \sigma_h^2$$

Luo, P., Song, Y.*, Huang, X., Ma, H., et al., 2022. Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS Journal of Photogrammetry and Remote Sensing. 185, 111-128.

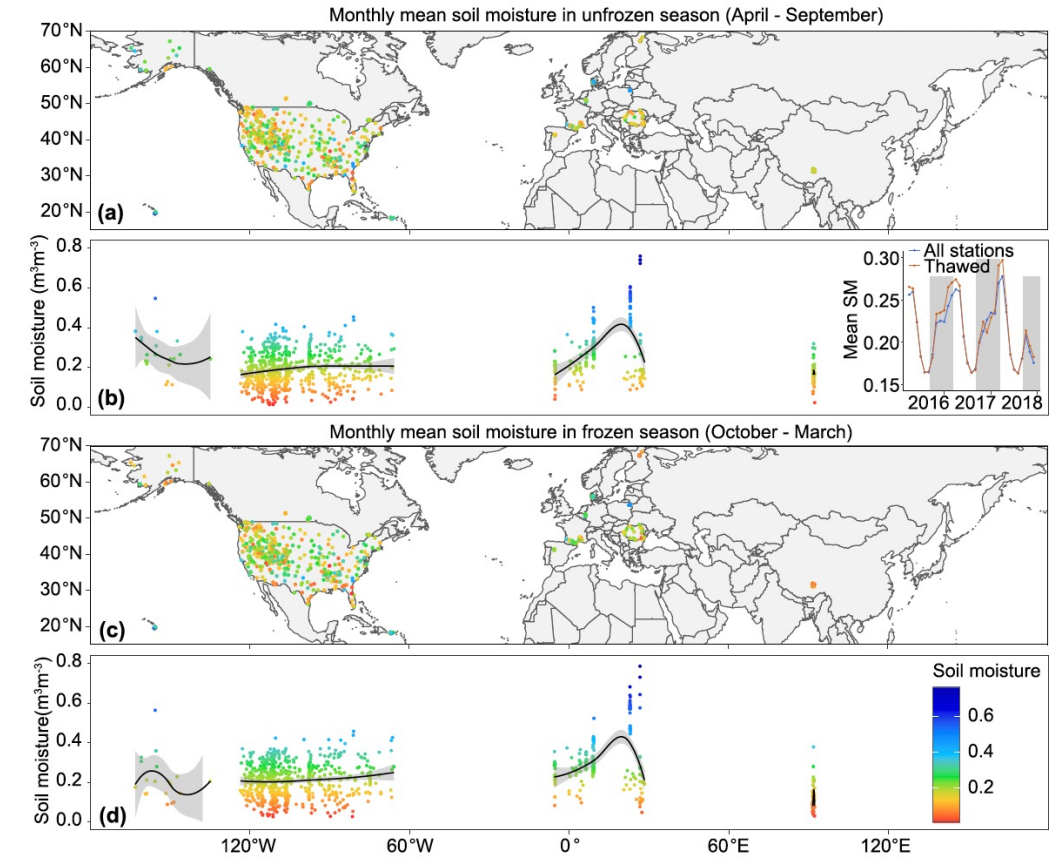


Fig. 3. Spatial distributions of monthly mean in situ soil moisture in the Northern Hemisphere in unfrozen (a and b) and frozen (c and d) seasons. In the small figure inside B, white background shows unfrozen seasons, and gray background shows frozen season.

Geographical Optimal Zones-based Heterogeneity (GOZH) model

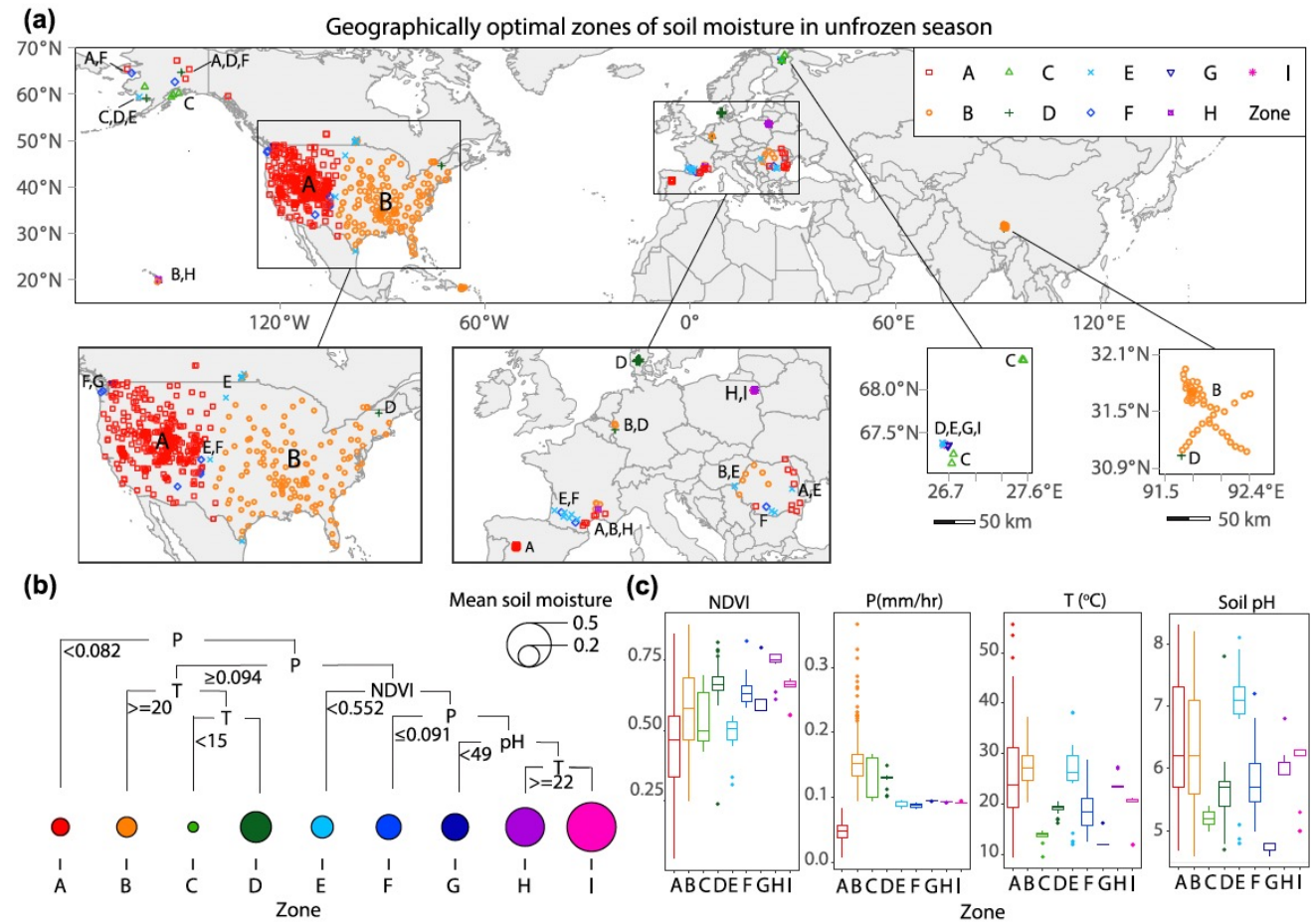


Fig. 6. Geographically optimal zones of soil moisture in the unfrozen season identified using the GOZH model (a), the process of identifying optimal zones (b), and statistical summaries of explanatory variables within zones for explaining characteristics of zones (c).

Geographical Optimal Zones-based Heterogeneity (GOZH) model

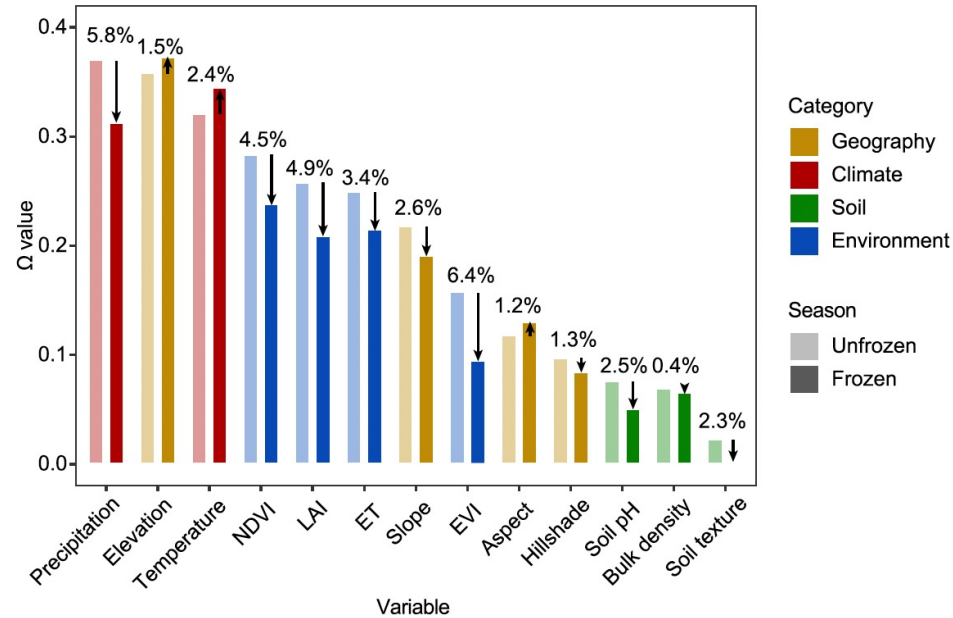


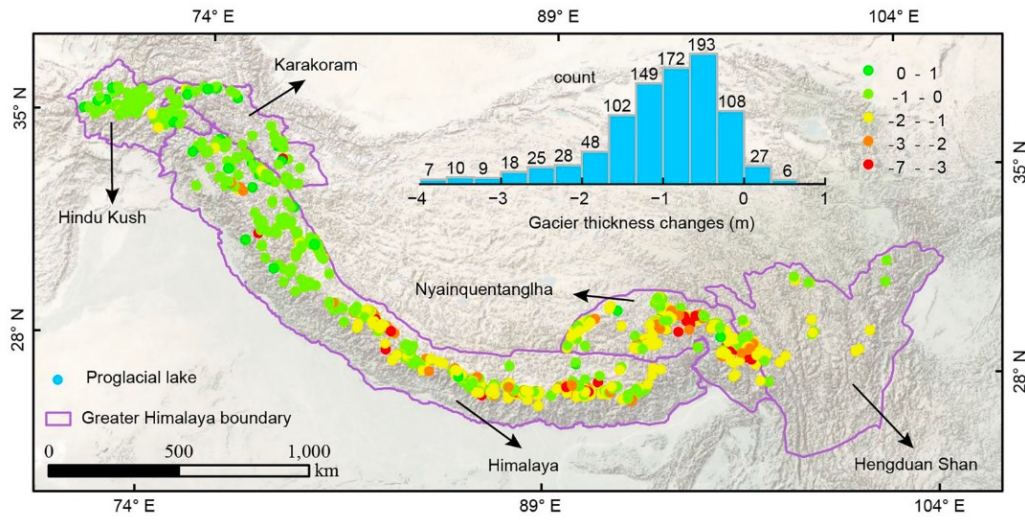
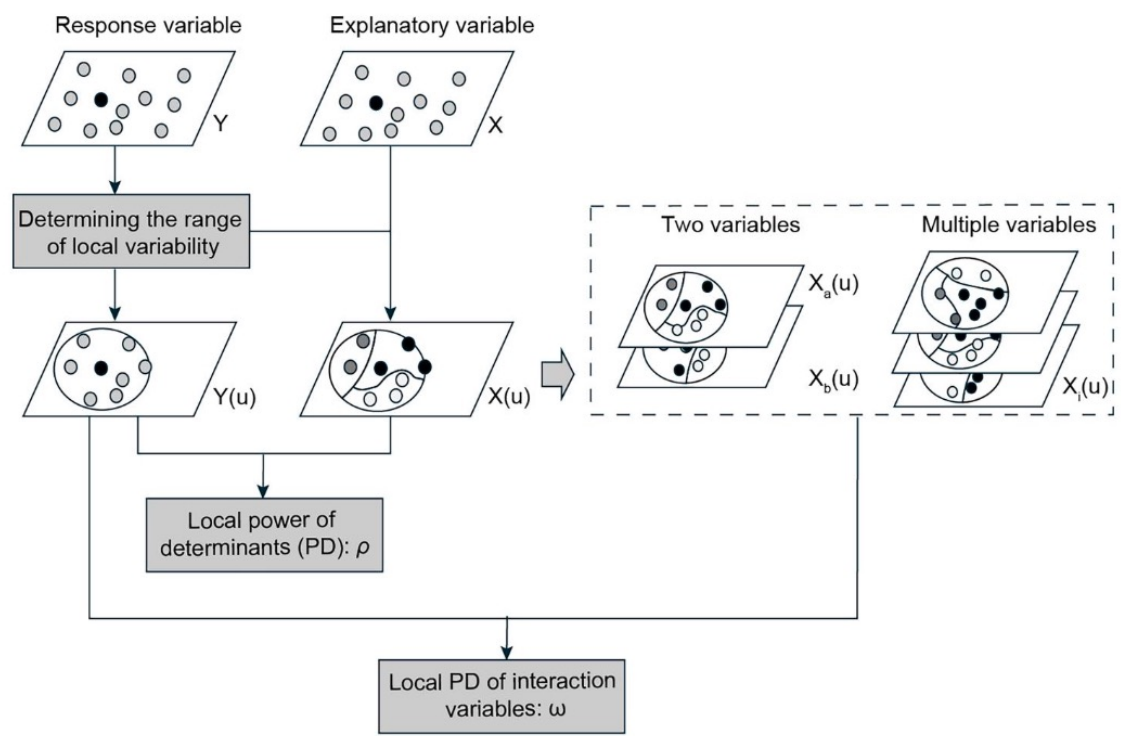
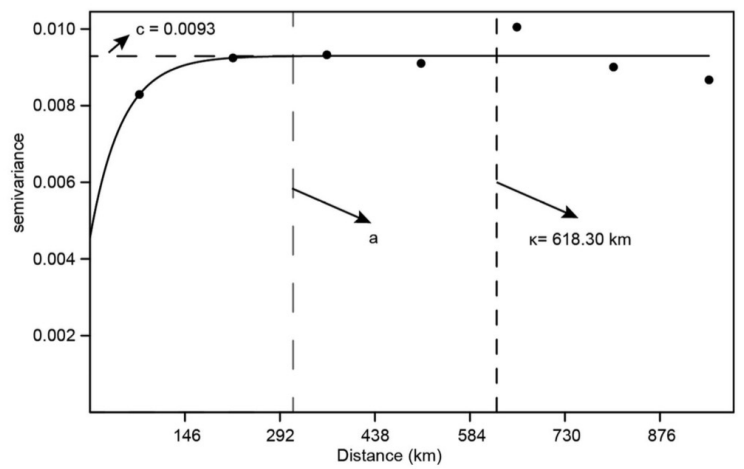
Fig. 5. Ω values of explanatory variables on spatial patterns of seasonal soil moisture and the comparison between unfrozen and frozen seasons.

Table 3

Contributions of explanatory variables on spatial patterns of soil moisture and contributions to dividing optimal zones in unfrozen and frozen seasons.

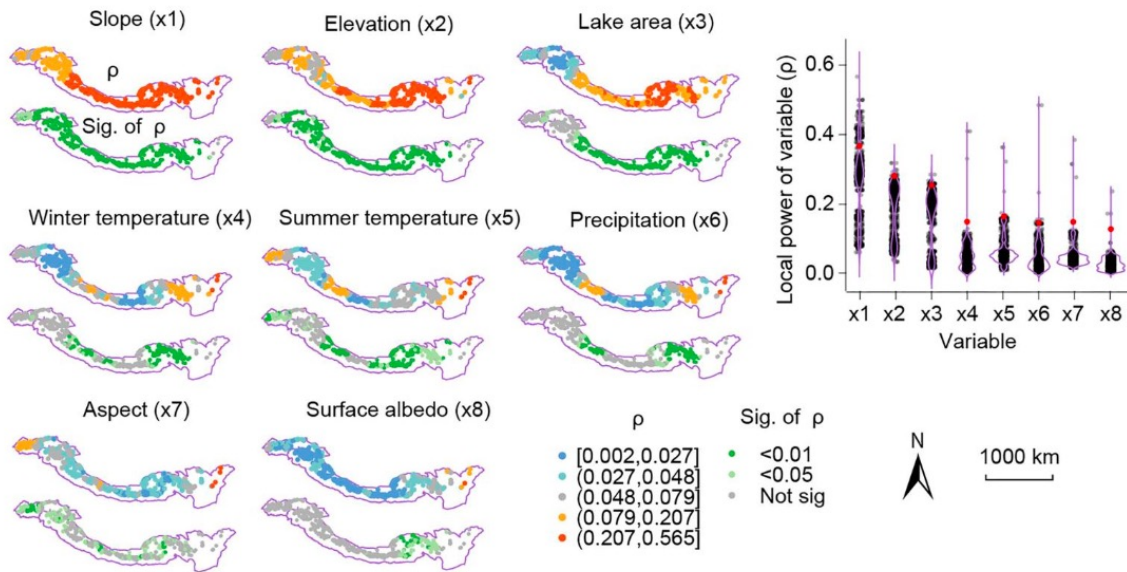
Unfrozen season			Frozen season		
Variable	Contribution to spatial patterns	Contribution to dividing zones	Variable	Contribution to spatial patterns	Contribution to dividing zones
Precipitation	20.99%	72.22%	Slope	13.72%	54.55%
NDVI	11.26%	11.11%	Soil bulk density	12.62%	12.73%
Temperature	7.90%	11.11%	ET	6.58%	14.55%
Soil pH	7.47%	5.56%	EVI	6.31%	10.91%
Overall Ω	47.62%	/	Precipitation	4.04%	3.64%
			Soil pH	4.01%	3.64%
			Overall Ω	47.69%	/

Local indicator of stratified power (LISP)

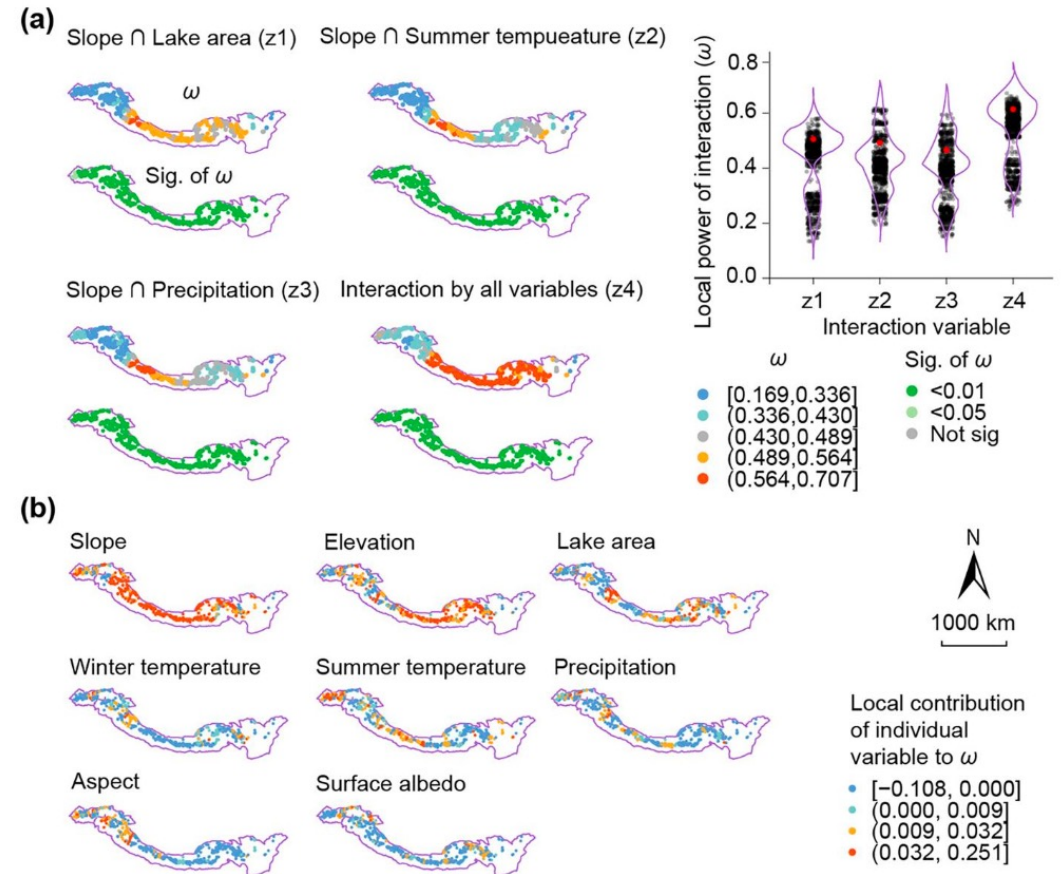


Spatial distribution of the thickness variations of lake-terminating glaciers in the Greater Himalayas.

Local indicator of stratified power (LISP)



The local power of determinants of individual variables and their significance tests.



The local power of interaction determinants. (a) The interaction effects with two and all factors and their significance tests.

Local indicator of stratified power (LISP)

Table 2. Comparison of the LISP-derived local PD values and global PD values in the existing stratified heterogeneity models.

Variable		Local PD of LISP mean [min, max]	Percentage of locations with significant LISP ($p < 0.05$)	PD of global geographical detector
Individual variables	Slope	0.266 [0.062, 0.565]	99.34%	0.207**
	Elevation	0.181 [0.033, 0.318]	95.70%	0.062**
	Lake area	0.154 [0.014, 0.285]	75.88%	0.155**
	Winter temperature	0.064 [0.009, 0.363]	73.12%	0.008*
	Summer temperature	0.049 [0.002, 0.409]	46.15%	0.056**
	Precipitation	0.046 [0.002, 0.484]	41.62%	0.066**
	Aspect	0.048 [0.016, 0.384]	51.87%	0.021**
Interaction variables	Surface albedo	0.028 [0.002, 0.236]	25.55%	0.053**
	Slope \cap Lake area	0.441 [0.169, 0.615]	98.67%	0.361**
	Slope \cap Summer temperature	0.427 [0.230, 0.652]	100%	0.289**
	Slope \cap Precipitation	0.401 [0.189, 0.632]	100%	0.372**
	Interaction of all variables	0.549 [0.284, 0.707]	100%	0.547**

Significance levels: ** $p < 0.01$; * $p < 0.05$.

Local impacts of factors

Global impacts of factors

Table 3. Illustrative examples explaining the importance of employing local models in spatial analysis.

Category	Global models and indicators	Local models and indicators
Spatial autocorrelation	Moran statistic (MORAN 1950); Geary c statistic (Geary 1954)	Local Indicator of Spatial Association (Anselin 1995); Local Geary c statistic (Zhang <i>et al.</i> 2024b); Heterogeneous spatial autocorrelation (Zhang <i>et al.</i> 2024b)
Spatial heterogeneity	Head-tail index (Jiang and Yin 2014)	Geographically weighted regression (Brunsdon <i>et al.</i> 1998); Multiscale geographically weighted regression (Fotheringham <i>et al.</i> 2017)
Spatial heteroscedasticity	Spatial error model with heteroscedasticity (Tolozza <i>et al.</i> 2024)	Local spatial heteroscedasticity (Ord and Getis 2012); Spatiotemporal autoregressive conditional heteroscedasticity (Otto 2024)
Spatial stratified heterogeneity	Geographical detectors (Wang <i>et al.</i> 2010), optimal parameters-based geographical detectors (Song and Wu 2021)	Local indicator of stratified power (LISP) (This study)

Feedback session for Assignment 1

Write a 200 word essay on using Optimal Parameters-based Geographical Detectors (OPGD) model and “GD” R package to analyze the power of determinant of climatic drivers on bushfire or your own data. Please try to use figures and maps to explain your analysis. (refer IGARSS 2025 part 3.2)

Send to Email before 10 am next day:

Yongze.song@outlook.com

Document name: A1_YourName.docx

Geographical Optimal Zones-based Heterogeneity (GOZH) model

Practice

R Code:

<https://github.com/yongzesong/gozh/#r-codes-for-the-geographical-optimal-zones-based-heterogeneity-goZh-model>

Assignment 2

You will have 40 min for working on Assignment 2

Write a 200 word essay on the analysis of potential spatial factors of vegetation using GOZH model with the data **ndvi_30** in GD package.

Send to Email before 10 am next day:

Yongze.song@outlook.com

Document name: A2_YourName.docx

Any questions?