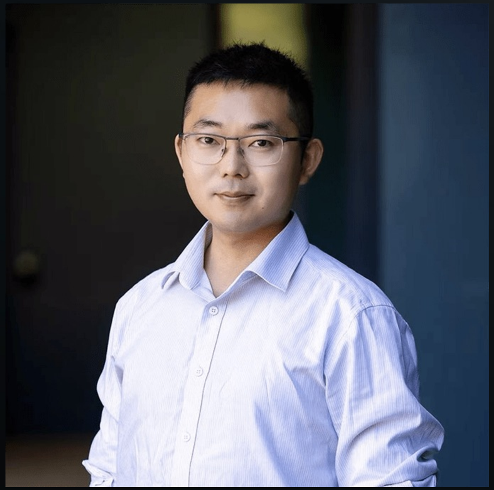
Geospatial Data Analysis, Prediction, and GeoAI: New Theories, Methods, and Software

Topic 3. Spatial modelling for prediction

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 AInet 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 interpolation and prediction
2. Kriging (geostatistics)
3. SDA model
4. GOS model

Part 2. Practice (2^{nd} to 4^{th} classes)

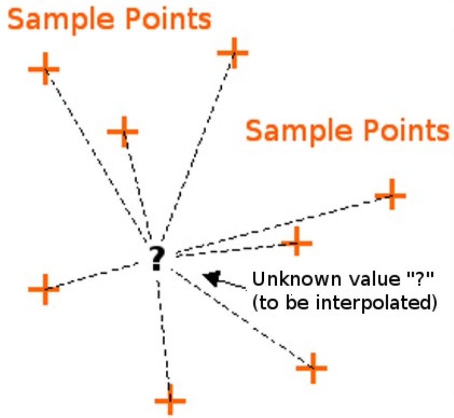
1. Feedback session for Assignment 2 (20 min)
2. Practice on Kriging (40 min)
3. Assignment 3 (40 min)
4. Submit your assignment (10 min)

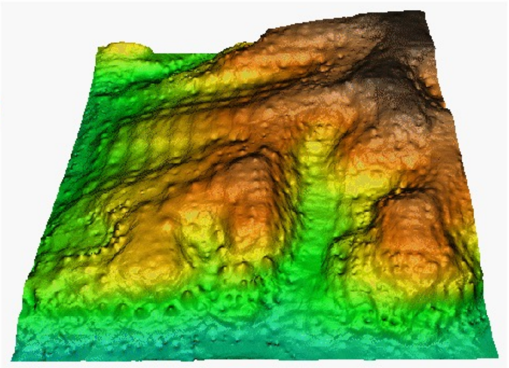
Tips:

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

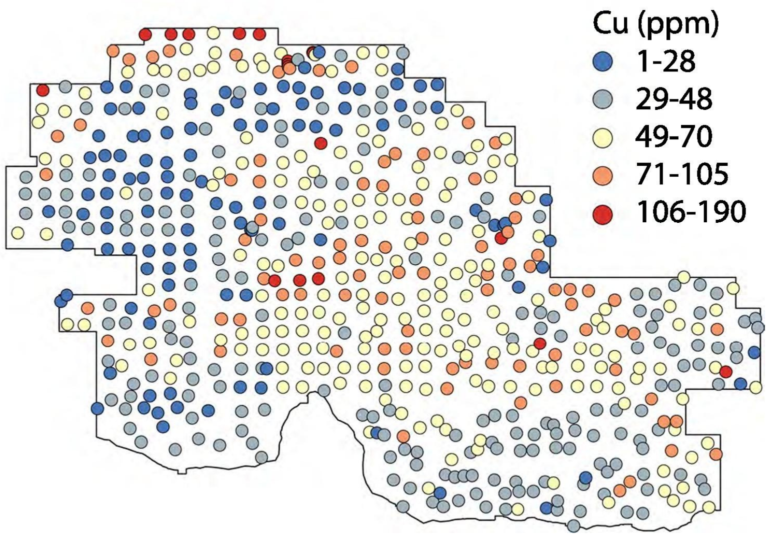
Concepts of interpolation and prediction

Spatial interpolation and spatial prediction estimate values at unsampled locations by using observed data and their spatial relationships to infer a continuous surface or to predict a target variable across space.

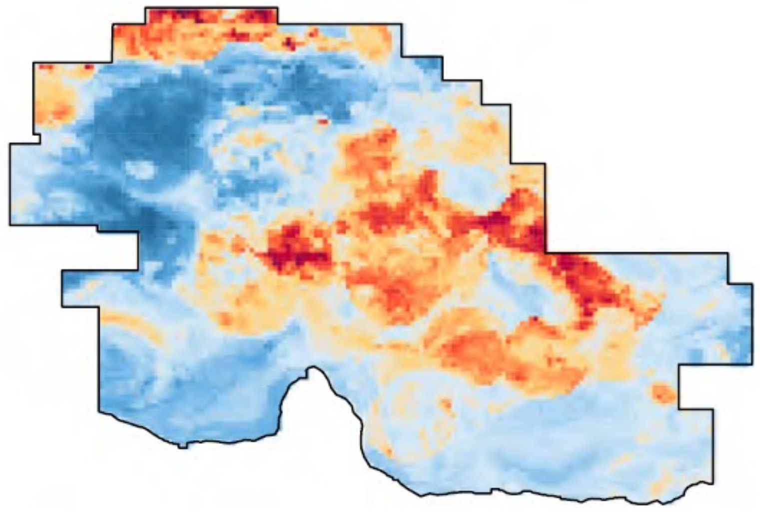




https://www.neonscience.org/resources/learning-hub/tutorials/spatial-interpolation-basics



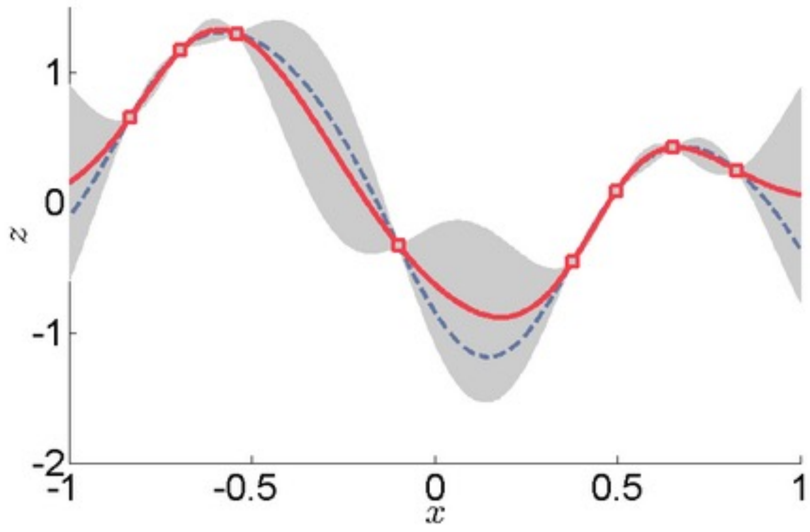
Spatial prediction



Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Kriging (Geostatistics)

- Kriging, also known as Gaussian process regression, is a geostatistical interpolation method that uses a Gaussian process with prior covariance structure to predict values at unsampled locations and under suitable assumptions provides the best linear unbiased prediction.
- Kriging was formalized by Georges Matheron in 1960 based on earlier work by Danie G Krige who used sample data to estimate gold grade distributions in South Africa.
- Kriging is widely used in spatial analysis and computer experiments and although it can be computationally intensive it can be scaled using approximation methods.



Example of one-dimensional data interpolation by kriging, with credible intervals.

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

Kriging (Geostatistics)

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

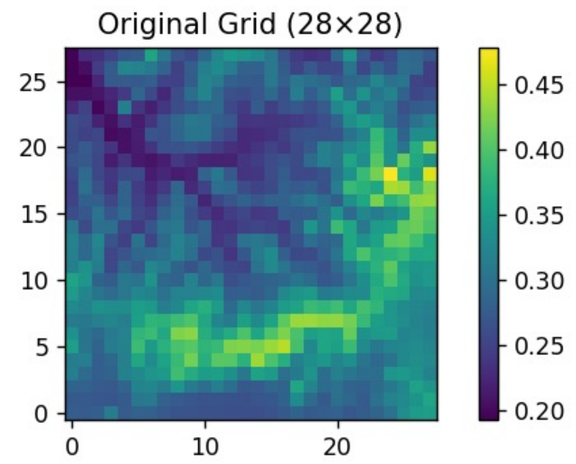
- where:

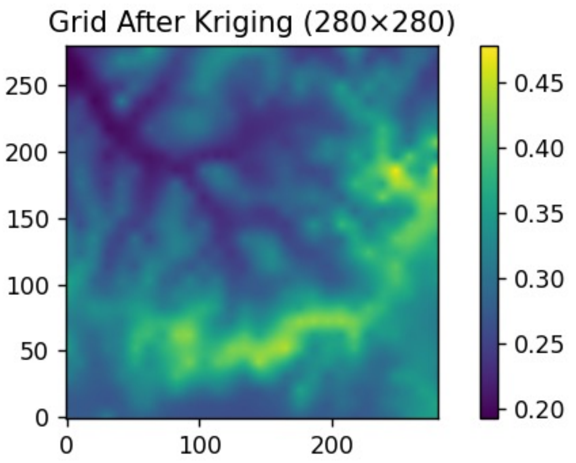
  $Z(s_i)$ = the measured value at the ith location

  λ_i = an unknown weight for the measured value at the ith location

  s_0 = the prediction location

  N = the number of measured values

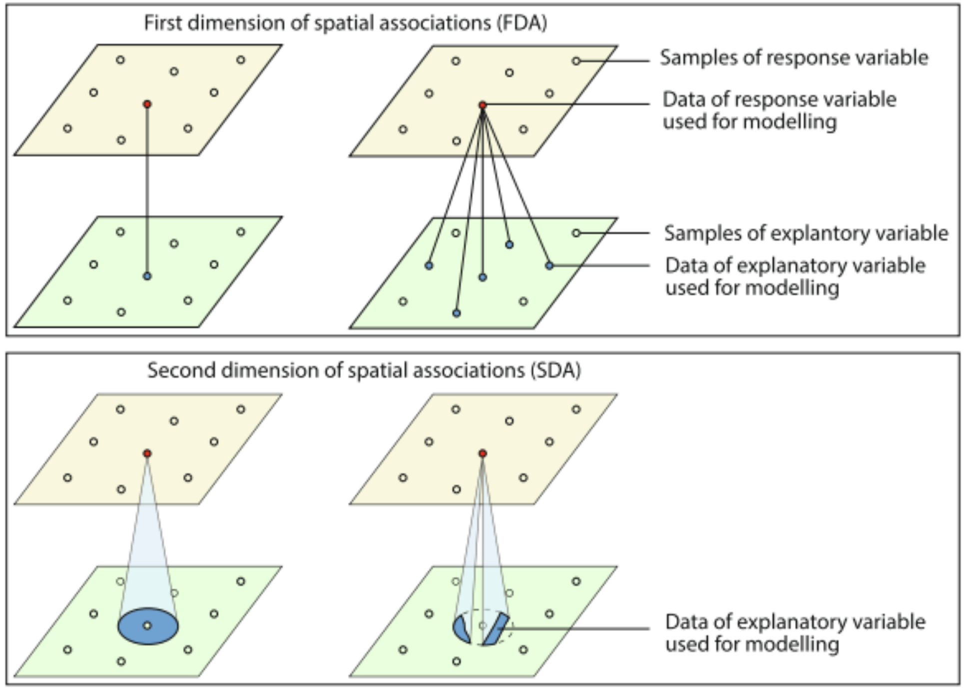




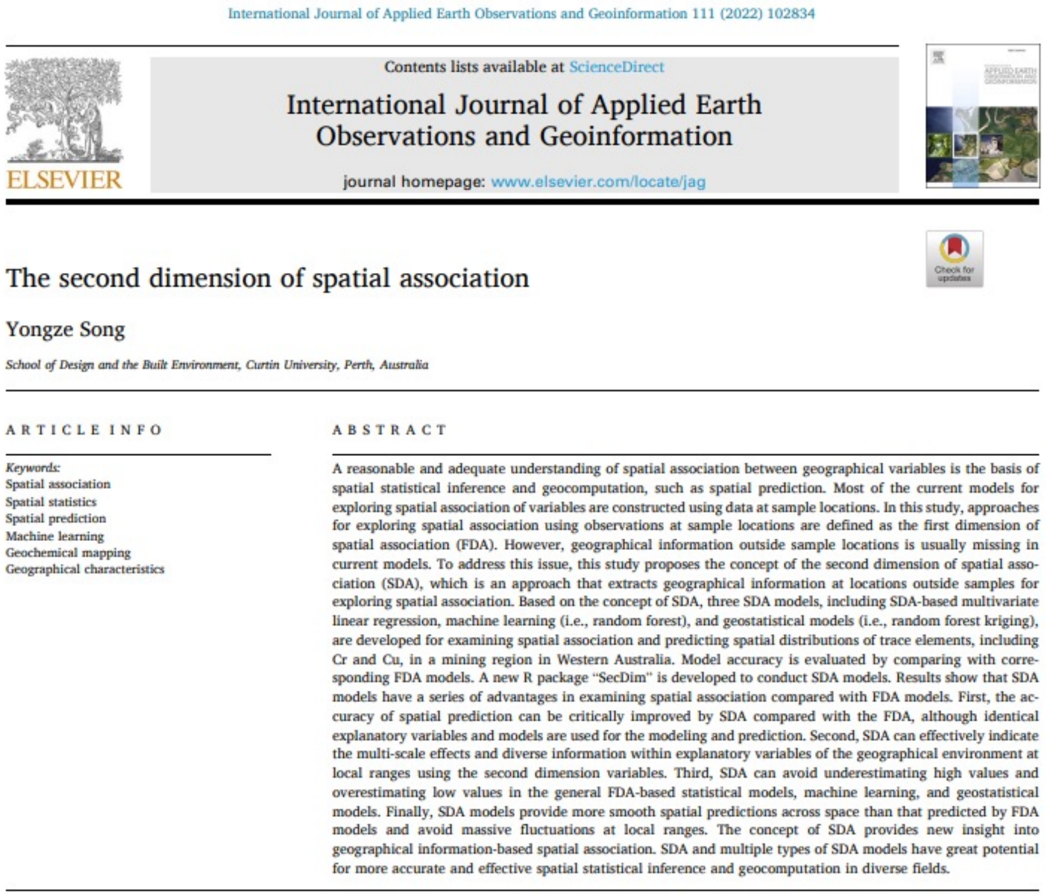
https://pro.arcgis.com/en/pro-app/3.4/tool-reference/3d-analyst/how-kriging-works.htm
https://en.wikipedia.org/wiki/Kriging

Second-Dimension Spatial Association

SDA examines spatial association by extracting more information about the geographical environment outside sampling locations



Comparison of concepts of the first and the second dimensions of spatial association (FDA and SDA).



International Journal of Applied Earth Observations and Geoinformation 111 (2022) 102834

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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

The second dimension of spatial association

Yongze Song

School of Design and the Built Environment, Curtin University, Perth, Australia

ARTICLE INFO

Keywords:
Spatial association
Spatial statistics
Spatial prediction
Machine learning
Geochemical mapping
Geographical characteristics

ABSTRACT

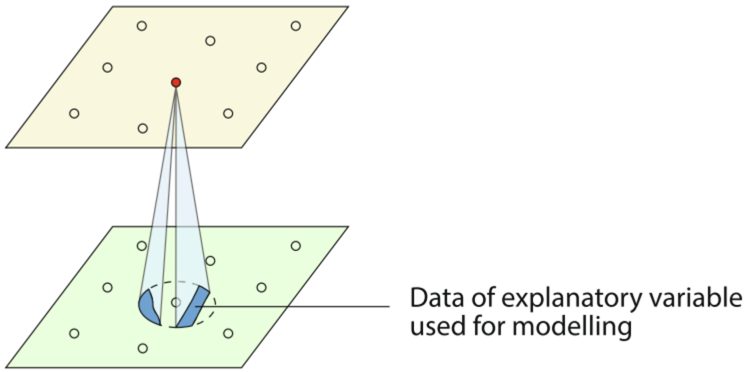
A reasonable and adequate understanding of spatial association between geographical variables is the basis of spatial statistical inference and geocomputation, such as spatial prediction. Most of the current models for exploring spatial association of variables are constructed using data at sample locations. In this study, approaches for exploring spatial association using observations at sample locations are defined as the first dimension of spatial association (FDA). However, geographical information outside sample locations is usually missing in current models. To address this issue, this study proposes the concept of the second dimension of spatial association (SDA), which is an approach that extracts geographical information at locations outside samples for exploring spatial association. Based on the concept of SDA, three SDA models, including SDA-based multivariate linear regression, machine learning (i.e., random forest), and geostatistical models (i.e., random forest kriging), are developed for examining spatial association and predicting spatial distributions of trace elements, including Cr and Cu, in a mining region in Western Australia. Model accuracy is evaluated by comparing with corresponding FDA models. A new R package "SecDim" is developed to conduct SDA models. Results show that SDA models have a series of advantages in examining spatial association compared with FDA models. First, the accuracy of spatial prediction can be critically improved by SDA compared with the FDA, although identical explanatory variables and models are used for the modeling and prediction. Second, SDA can effectively indicate the multi-scale effects and diverse information within explanatory variables of the geographical environment at local ranges using the second dimension variables. Third, SDA can avoid underestimating high values and overestimating low values in the general FDA-based statistical models, machine learning, and geostatistical models. Finally, SDA models provide more smooth spatial predictions across space than that predicted by FDA models and avoid massive fluctuations at local ranges. The concept of SDA provides new insight into geographical information-based spatial association. SDA and multiple types of SDA models have great potential for more accurate and effective spatial statistical inference and geocomputation in diverse fields.

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association

Background:

Geospatial data contains essential information on geographical characteristics, but current spatial prediction models are limited in extracting diverse geographical information outside sample locations.



Observations at sample locations **u**

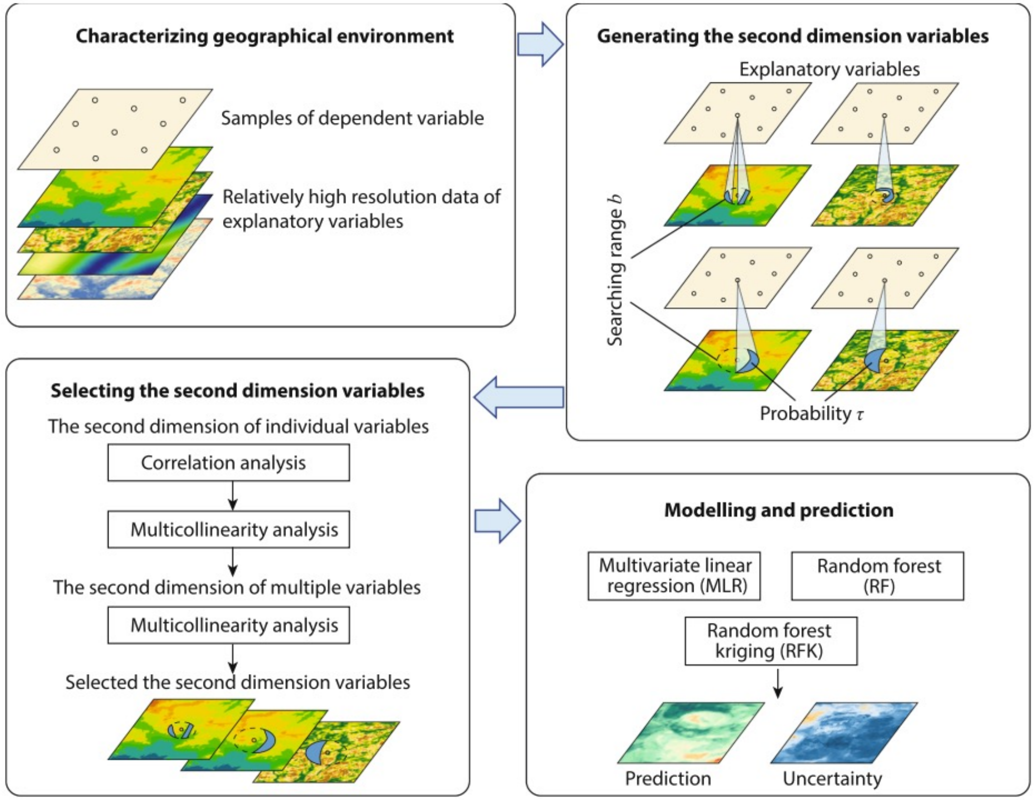
$$Y(\mathbf{u}) = f(\mathbf{X}'(\mathbf{v}, \mathbf{b}_{\mathbf{p}}, \boldsymbol{\tau}_q))$$

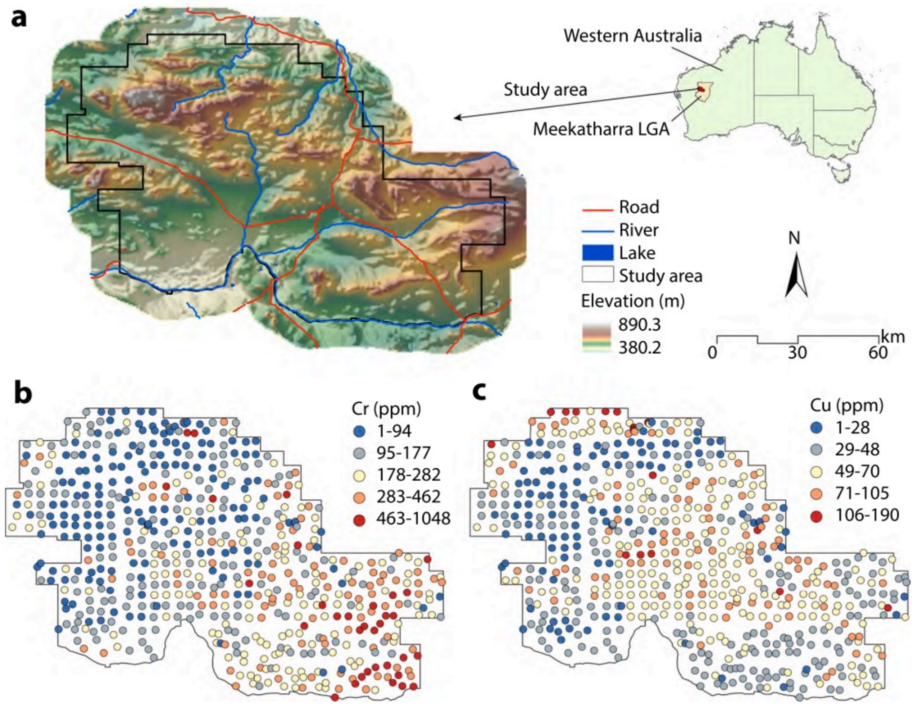
The selected second dimension variables at locations **v**, which are outside sample locations **u**, and determined by the searching range parameter $\boldsymbol{b_p}$ and probability parameter $\boldsymbol{\tau_q}$

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association

Schematic overview of SDA models for spatial prediction



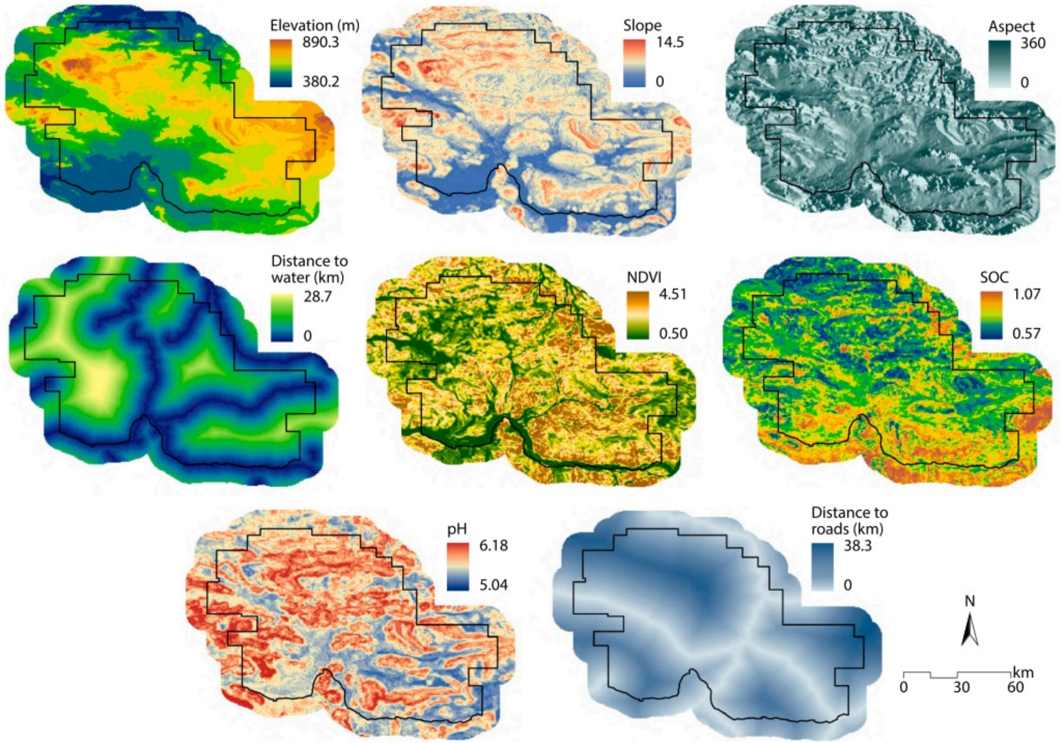


Spatial distributions of trace element data, Cr (b) and Cu (c).

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association

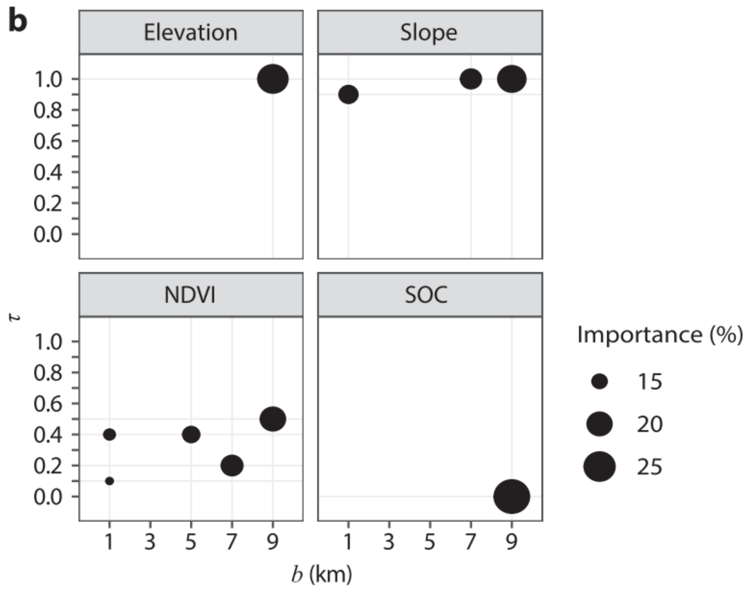
Generating second-dimension spatial variables



Explanatory variables

The second-dimension variables for predicting Cu:

$$Y = f\left(X_{SOC}^{b=9,\tau=0}, X_{Elevation}^{b=9,\tau=1}, X_{Slope}^{b=9,\tau=1}, X_{NDVI}^{b=9,\tau=0.5}, X_{NDVI}^{b=7,\tau=0.2}, X_{Slope}^{b=7,\tau=1}, X_{Slope}^{b=1,\tau=0.9}, X_{NDVI}^{b=5,\tau=0.4}, X_{NDVI}^{b=1,\tau=0.4}, X_{NDVI}^{b=1,\tau=0.1} \right)$$



Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association

SDA improves prediction accuracy compared with Random Forest Kriging, one of the best-performed spatial machine learning models.

Summary of cross-validation of FDA and SDA models for predicting trace elements.

Trace element	Cross validation	FDA-MLR	FDA-RF	FDA-RFK	SDA-MLR	SDA-RF	SDA-RFK
Cr	R^2	0.184	0.239	0.286	0.325	0.370	0.379
	RMSE	0.686	0.655	0.637	0.620	0.606	0.603
	MAE	0.525	0.499	0.486	0.492	0.480	0.476
Cu	R^2	0.065	0.070	0.142	0.270	0.358	0.346
	RMSE	0.501	0.502	0.480	0.440	0.412	0.414
	MAE	0.393	0.384	0.365	0.336	0.309	0.312

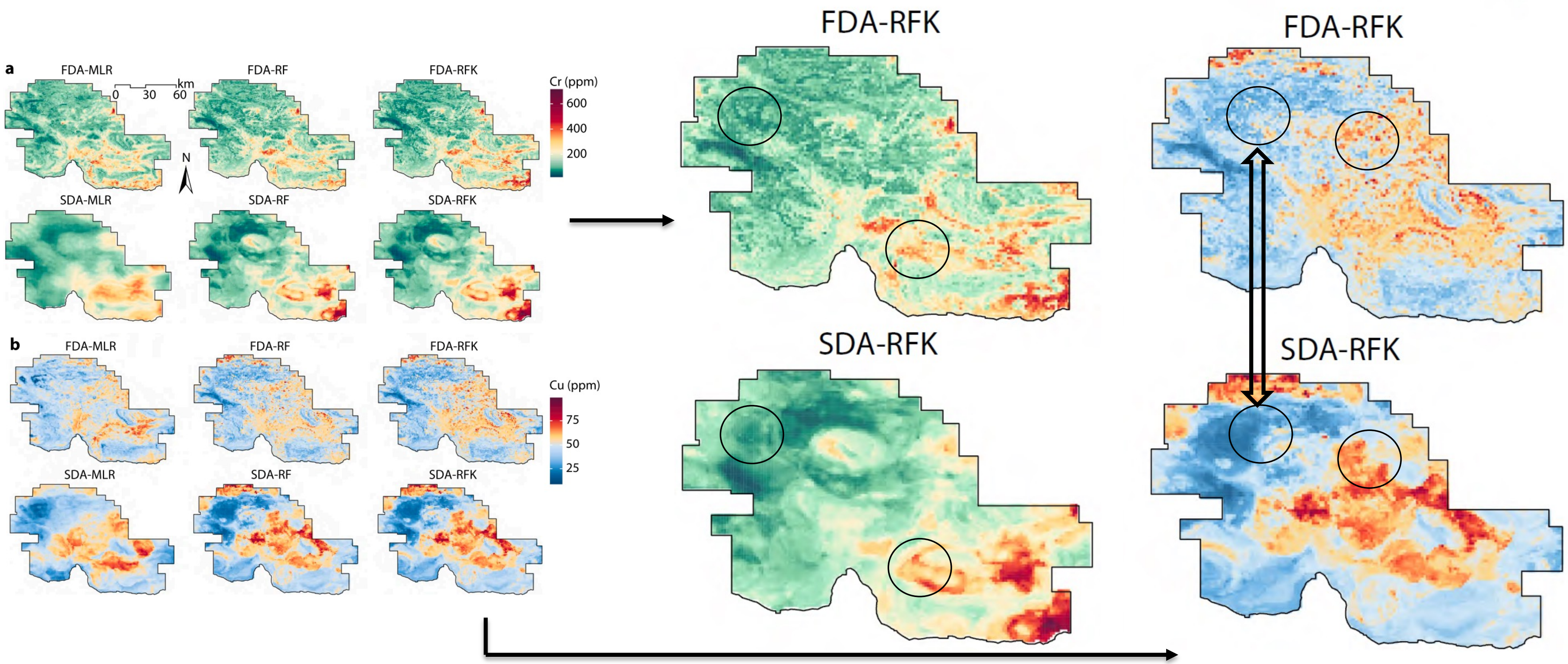
Improvements of model accuracy by SDA models compared with FDA models.

Trace element	Improvement by SDA	MLR	RF	RFK
Cr	R^2 improvement	77.0%	54.9%	32.8%
	RMSE reduction	9.7%	7.6%	5.4%
	MAE reduction	6.3%	3.9%	2.0%
Cu	R^2 improvement	316.6%	411.3%	144.4%
	RMSE reduction	12.1%	18.0%	13.8%
	MAE reduction	14.6%	19.6%	14.5%

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

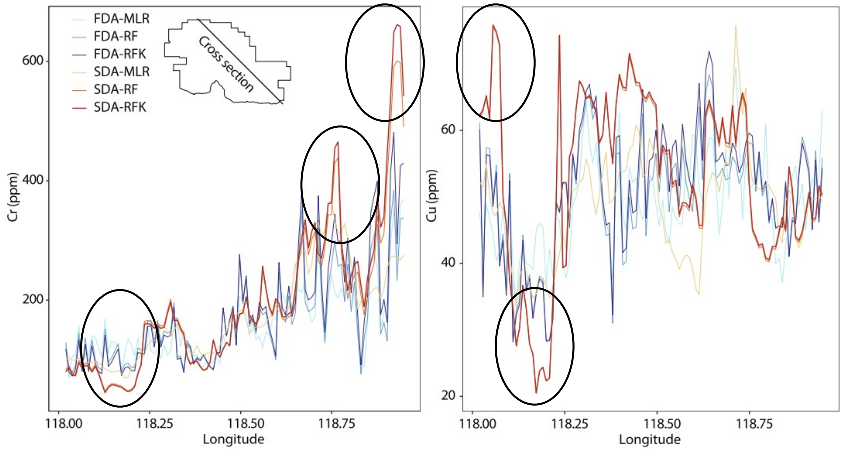
Second-Dimension Spatial Association

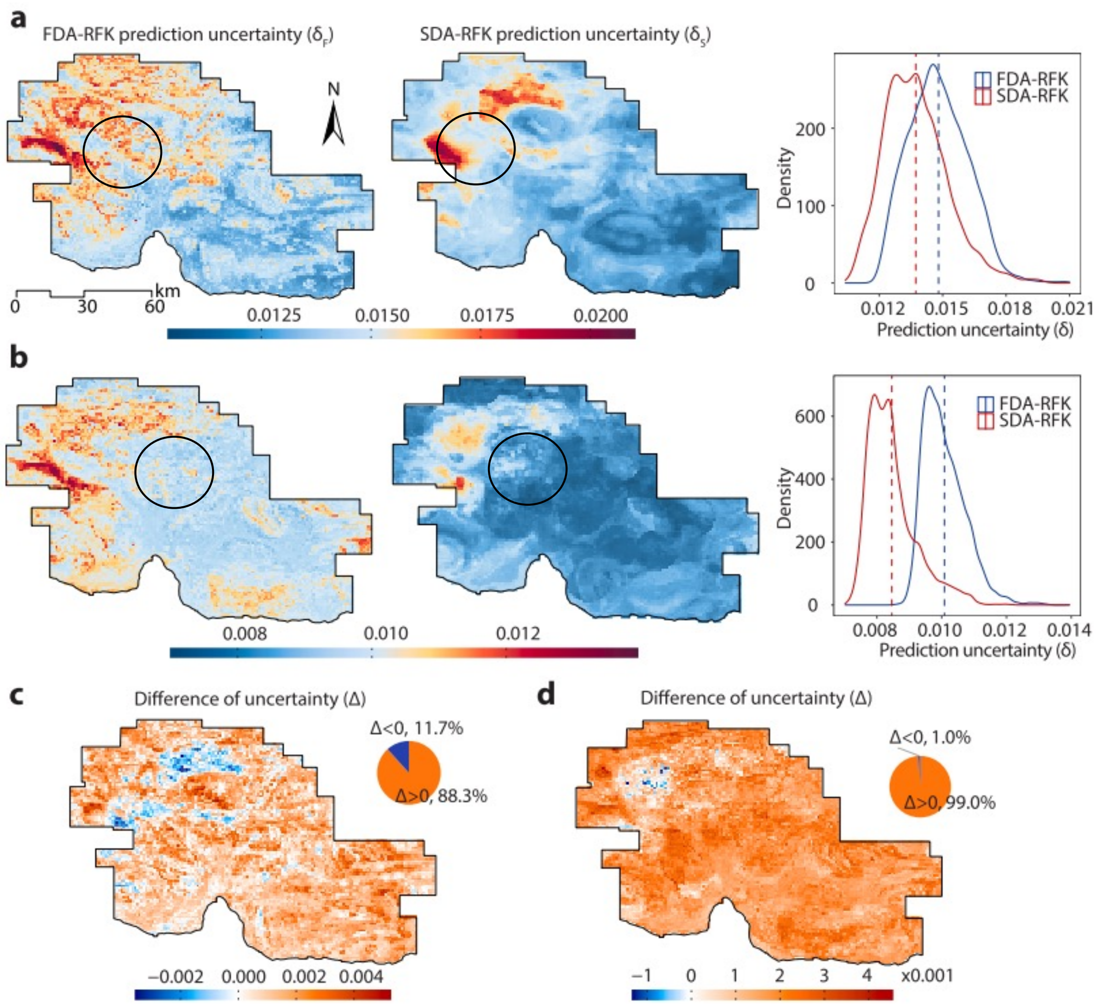
SDA models provide smooth, effective, and low-uncertainty spatial predictions.



Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association





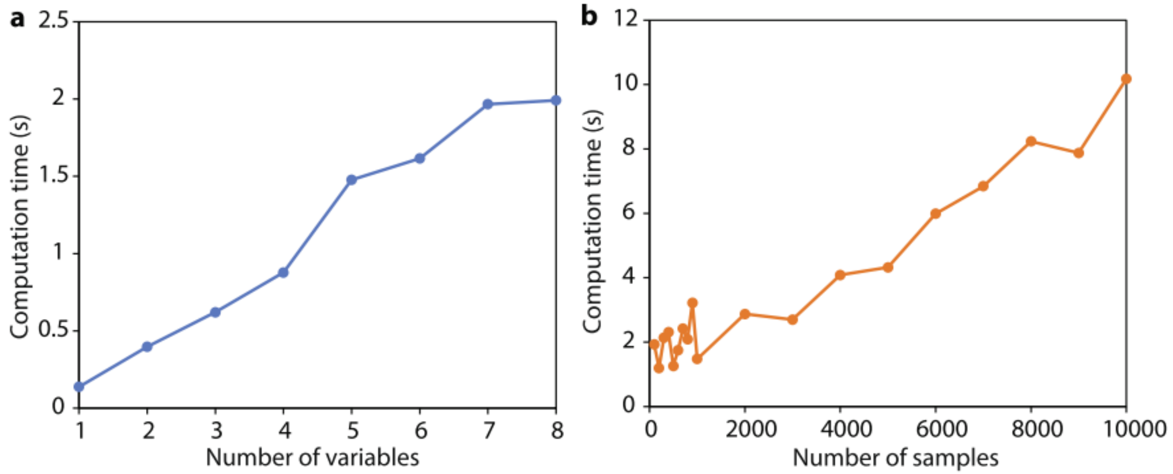
Uncertainties of spatial predictions of trace element Cr (a) and Cu (b) derived using the FDA and SDA-based random forest kriging (RFK) models, and the corresponding difference of uncertainties (c-d).

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-Dimension Spatial Association

SecDim Package for The Second Dimension of Spatial Association

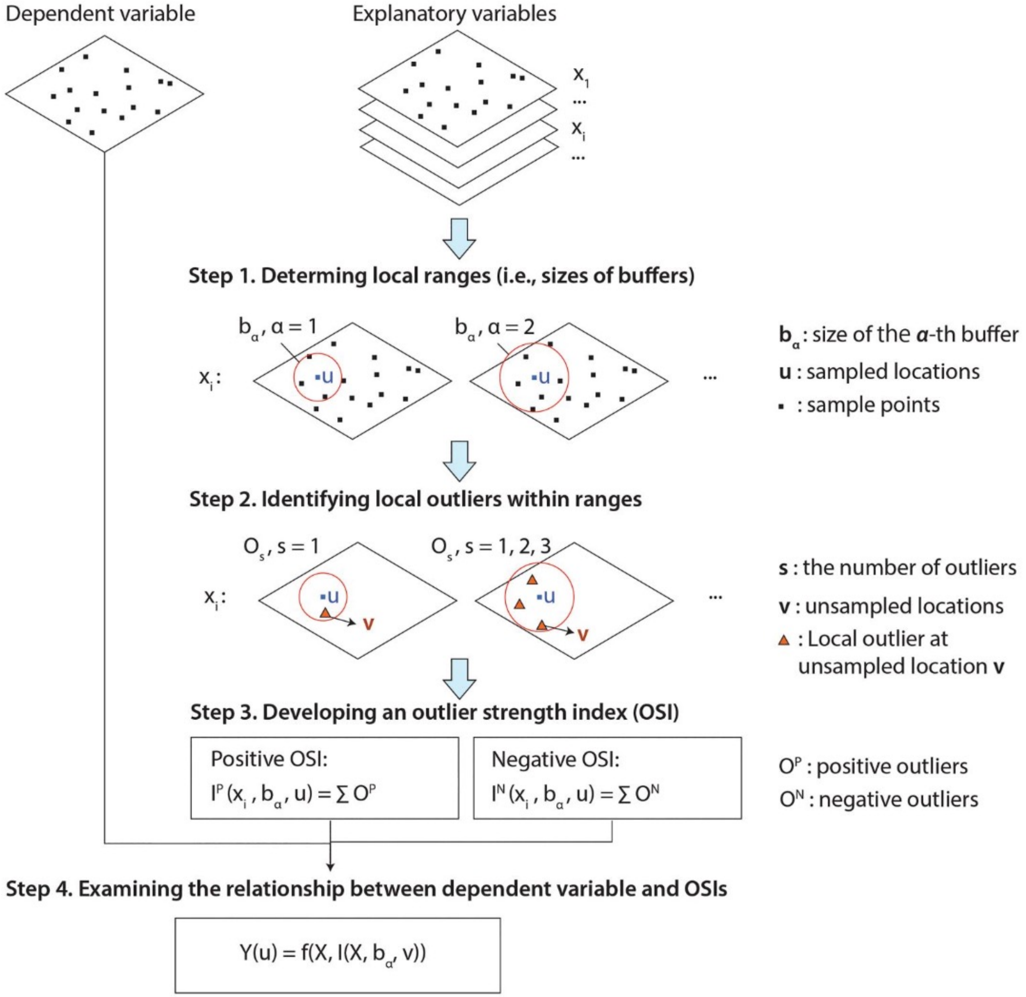
https://cran.r-project.org/web/packages/SecDim/vignettes/SecDim.html



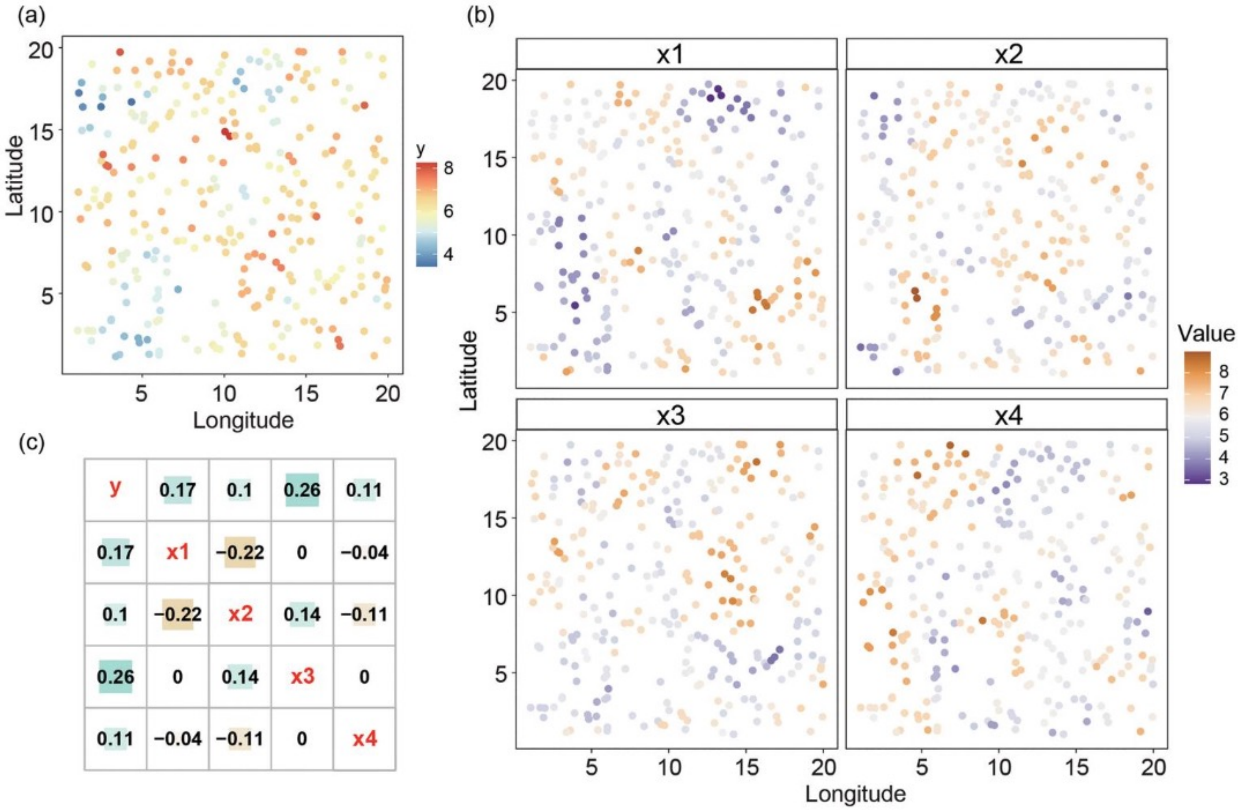
Comparison of time consumption in the second dimension variable selection stage in SDA-based models. (a) Computation time with different numbers of variables; and (b) computation time with different numbers of samples.

Song, Y.* The second dimension of spatial association. International Journal of Applied Earth Observation and Geoinformation. 2022. 111, 102834.

Second-dimension outliers for prediction



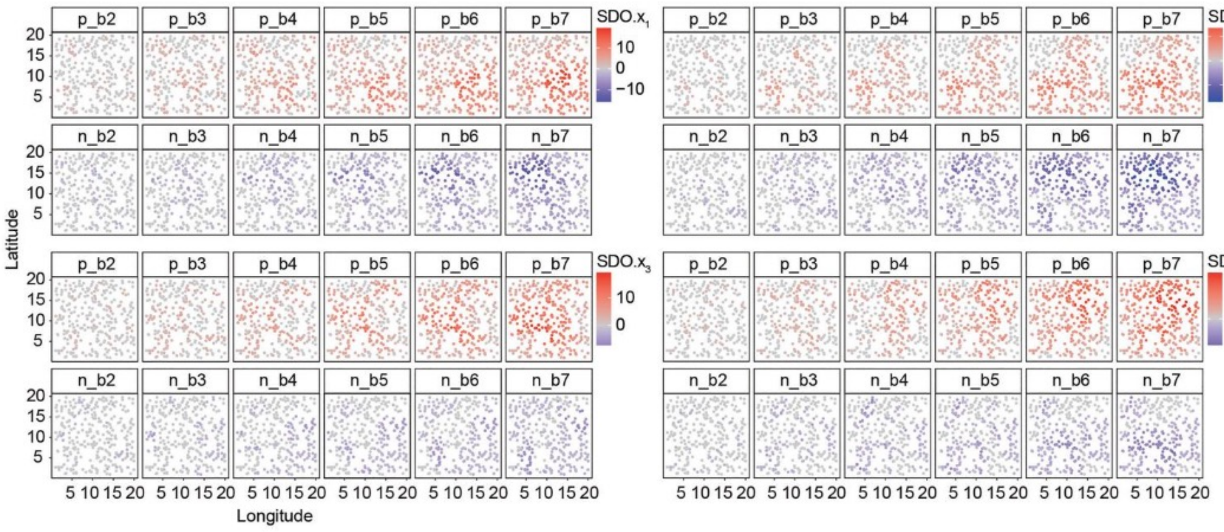
Simulation data



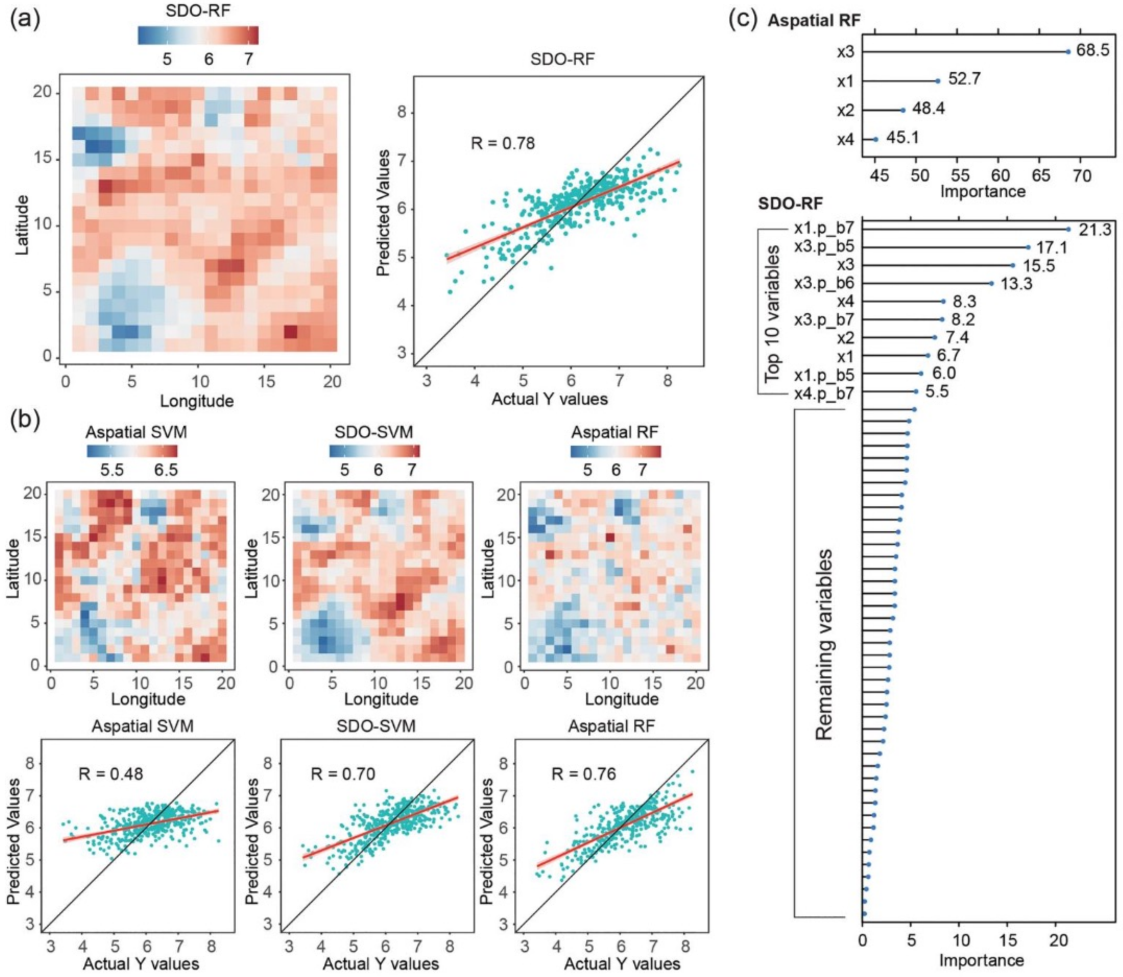
Ren, K., Song, Y.* and Yu, Q., 2025. Second-dimension outliers for spatial prediction. International Journal of Geographical Information Science.

Second-dimension outliers for prediction

Second-dimension outliers

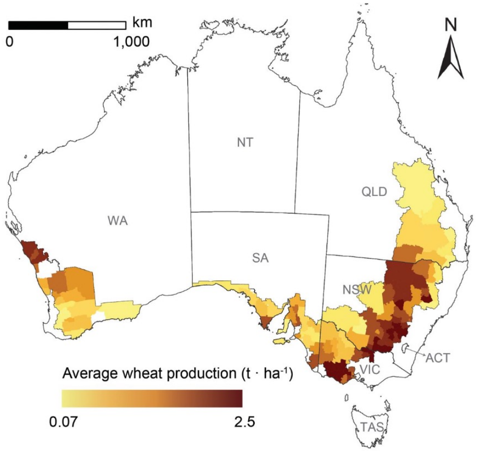


SDO model vs Random forest (RF)

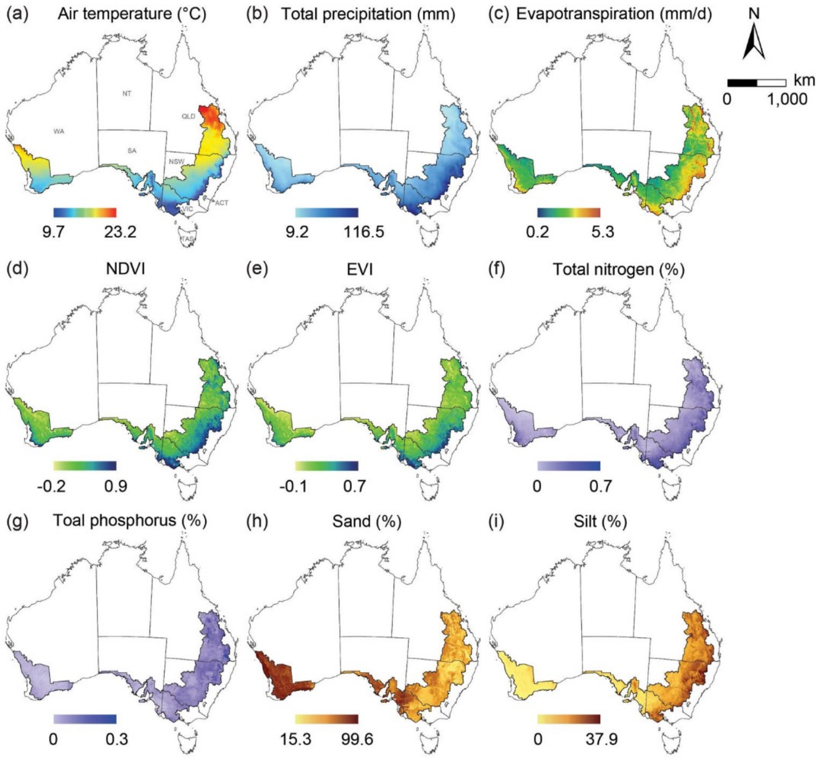


Ren, K., Song, Y.* and Yu, Q., 2025. Second-dimension outliers for spatial prediction. International Journal of Geographical Information Science.

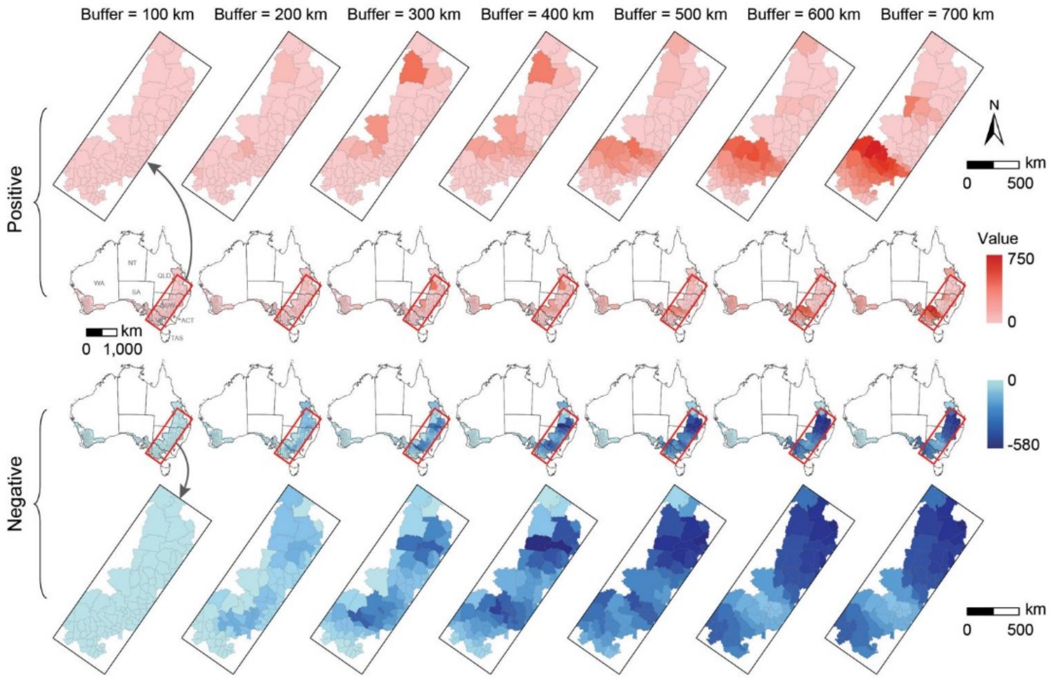
Second-dimension outliers for prediction



Wheat production



Climate and environment



SDO variables

Ren, K., Song, Y.* and Yu, Q., 2025. Second-dimension outliers for spatial prediction. International Journal of Geographical Information Science.

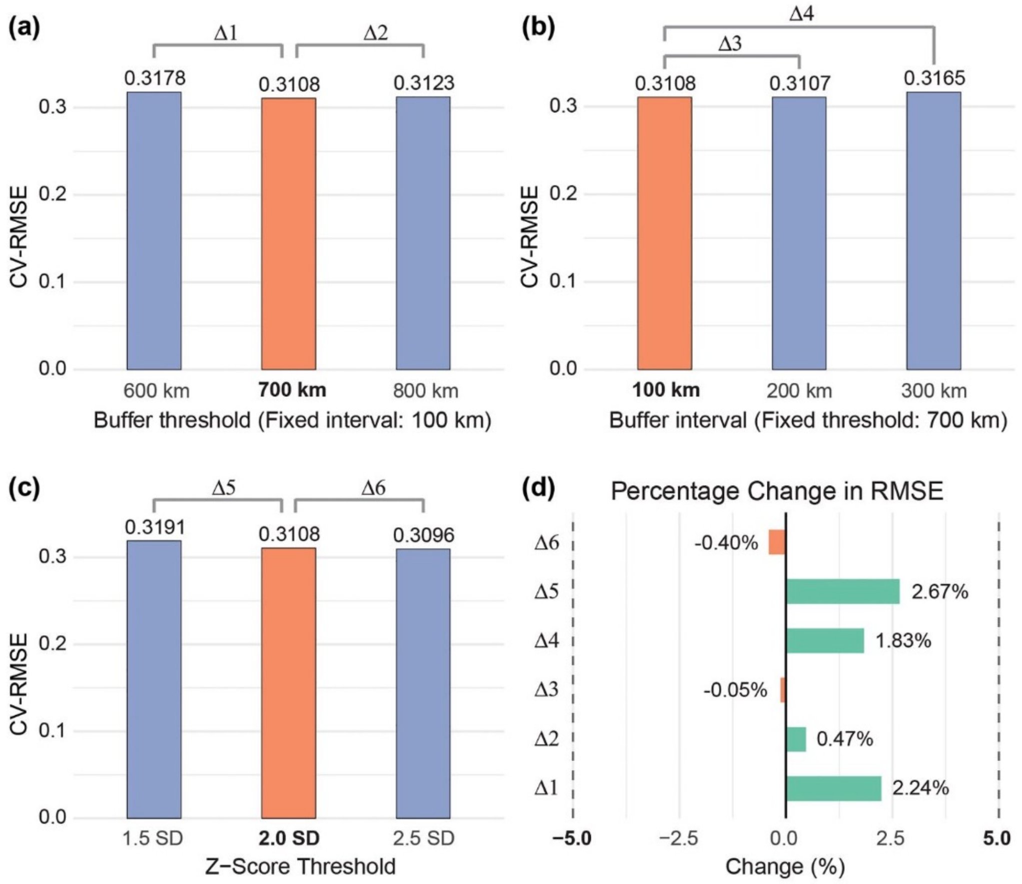
Second-dimension outliers for prediction

Accuracy

Table 2. Improvements of model accuracy in machine learning by SDO models compared with aspatial models (bold values indicating the best-performing SDO results).

Model		RF	CRM	XGB	SVM	GBM	KNN	ENR
R^2	Aspatial	0.527	0.634	0.548	0.555	0.496	0.455	0.580
	SDO	0.615	0.590	0.628	**0.671**	0.596	0.637	0.585
	Improvement	16.7%	−6.9%	14.6%	20.9%	20.2%	**40.0%**	0.9%
RMSE	Aspatial	0.368	0.321	0.370	0.355	0.377	0.400	0.343
	SDO	0.338	0.350	0.330	**0.311**	0.337	0.321	0.360
	Reduction	8.2%	−9.0%	10.8%	12.4%	10.6%	**19.8%**	−5.0%
MAE	Aspatial	0.279	0.241	0.286	0.267	0.283	0.310	0.272
	SDO	0.256	0.254	0.243	**0.231**	0.248	0.241	0.265
	Reduction	8.2%	−5.4%	15.0%	13.5%	12.4%	**22.3%**	2.6%

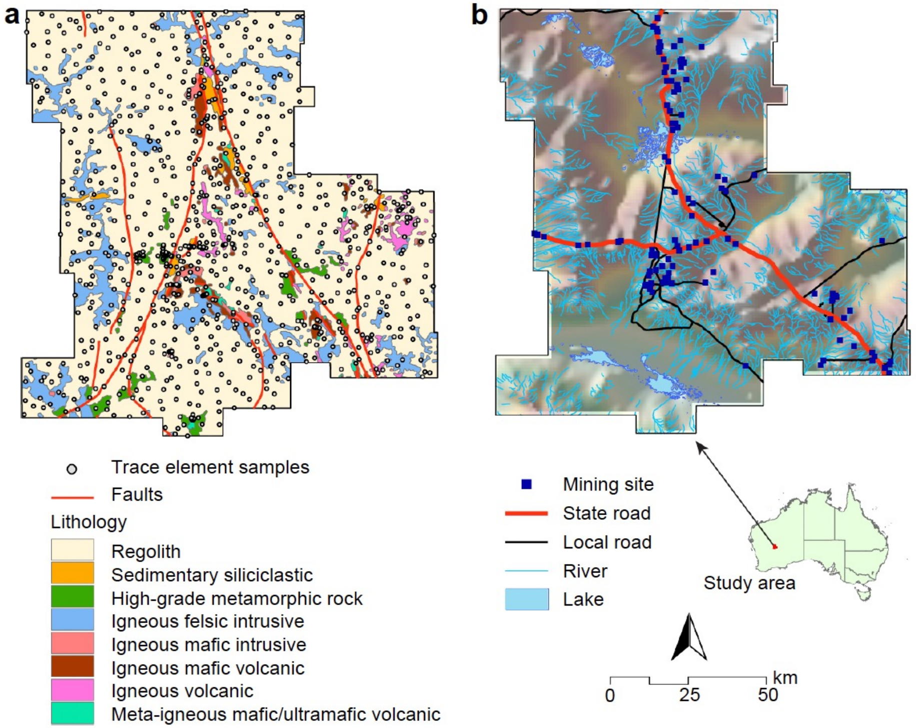
Sensitivity



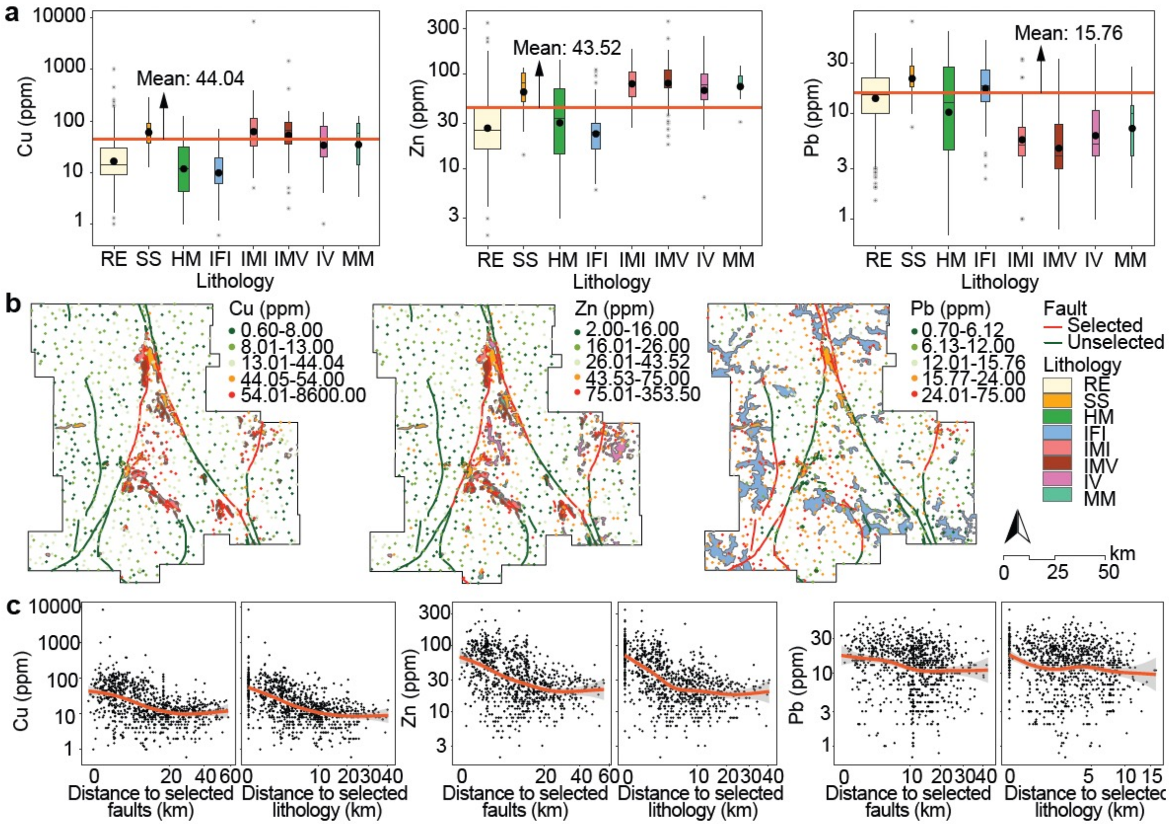
Ren, K., Song, Y.* and Yu, Q., 2025. Second-dimension outliers for spatial prediction. International Journal of Geographical Information Science.

Geographically optimal similarity

Trace element samples: Cu, Zn, Pb



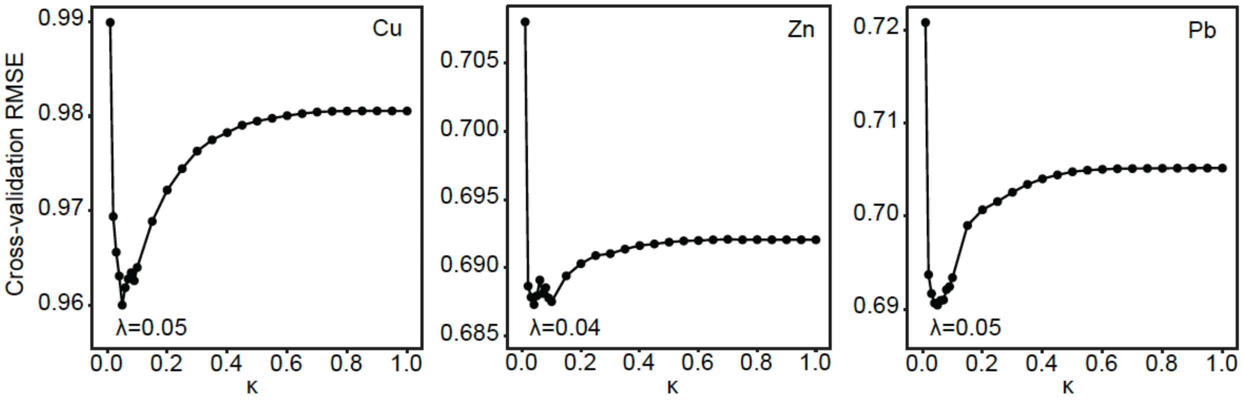
Fault and lithology-related variables



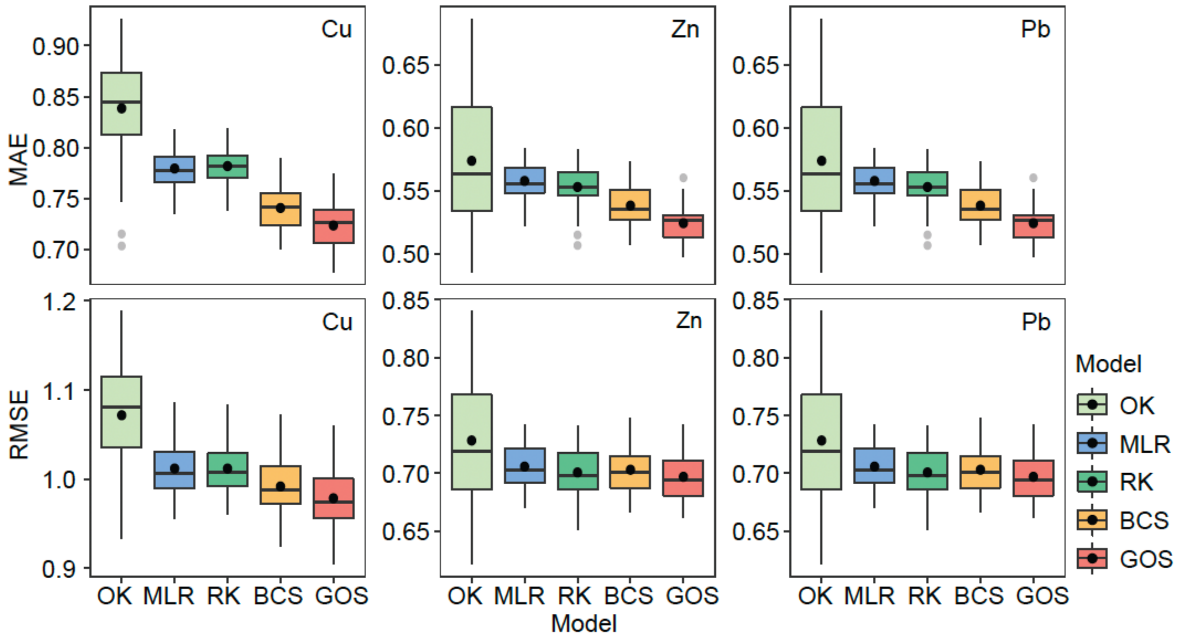
Song, Yongze*. Geographically optimal similarity. Mathematical Geosciences, 2023.

Geographically optimal similarity

Optimal parameters: Using 4% - 5% data with more similar geographical configurations for spatial prediction.



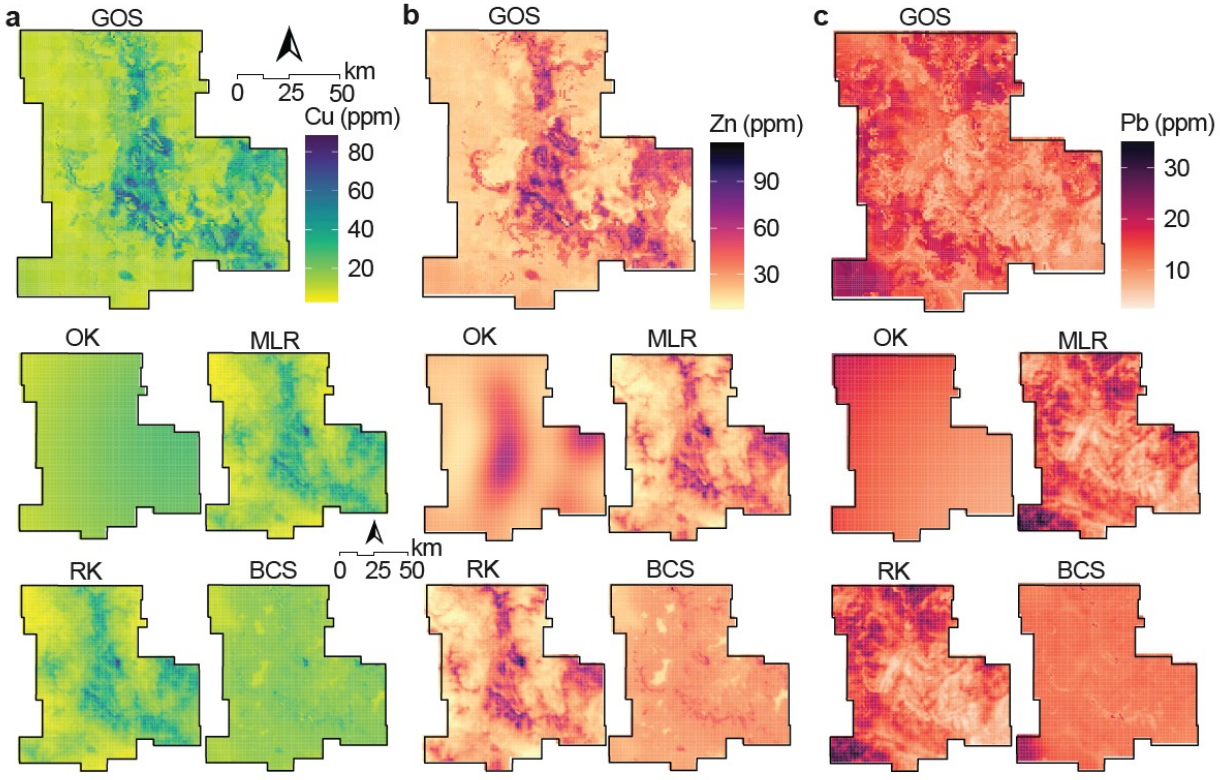
Cross validation



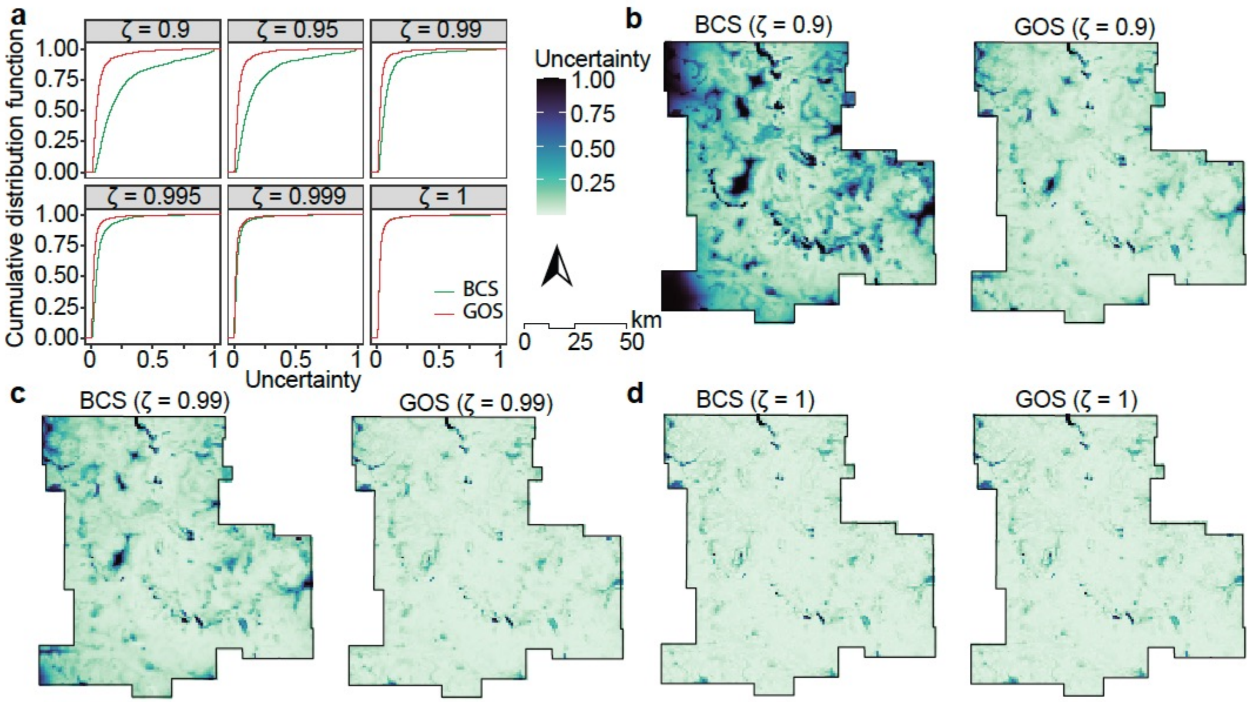
Song, Yongze*. Geographically optimal similarity. Mathematical Geosciences, 2023.

Geographically optimal similarity

Spatial predictions



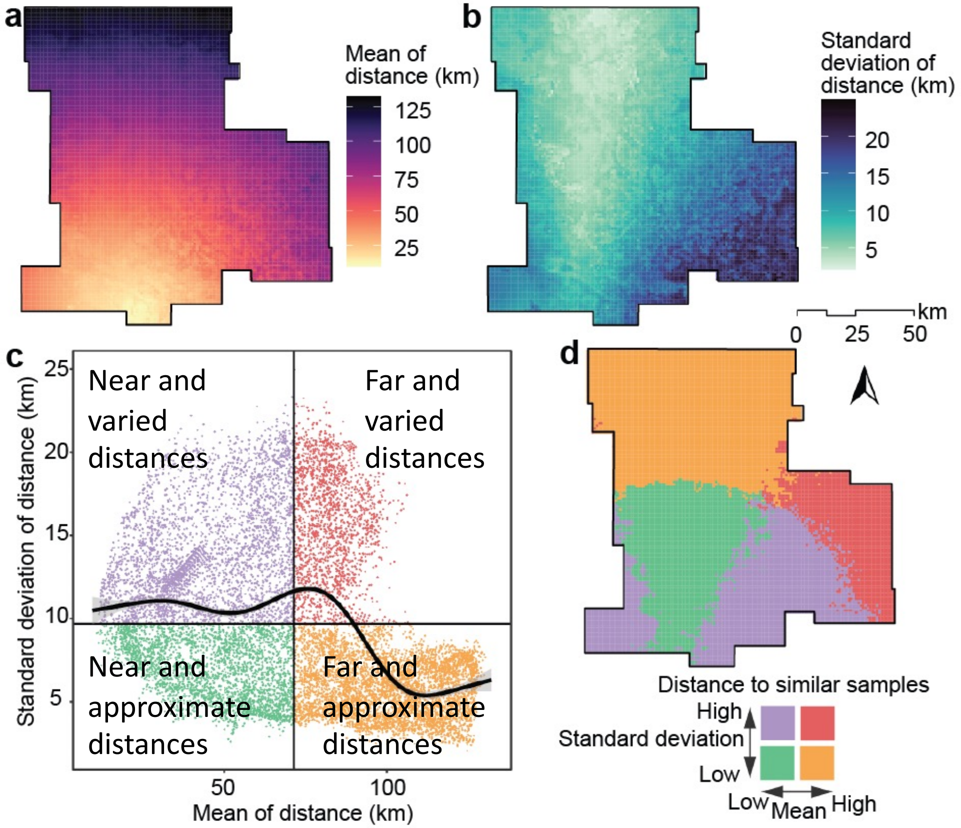
Uncertainties of spatial predictions



Song, Yongze*. Geographically optimal similarity. Mathematical Geosciences, 2023.

Geographically optimal similarity

Data with similar configurations may be located at significantly varied distances.



Song, Yongze*. Geographically optimal similarity. Mathematical Geosciences, 2023.

Feedback session for 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.

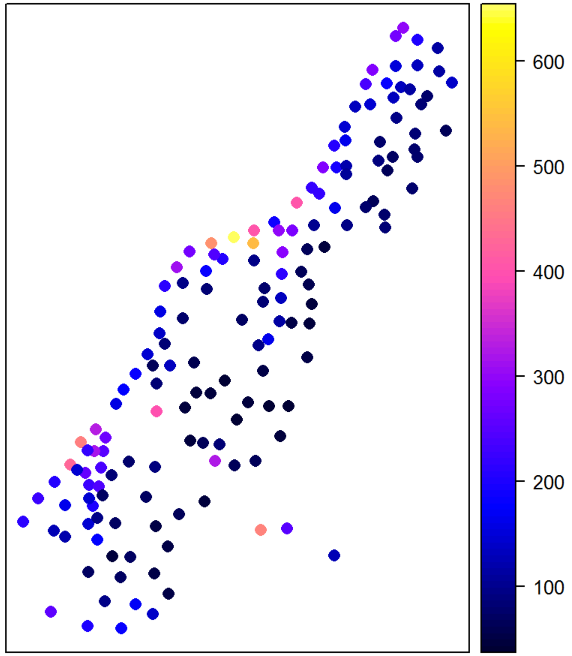
Send to Email before 10 am next day:
Yongze.song@outlook.com

Document name: A2_YourName.docx

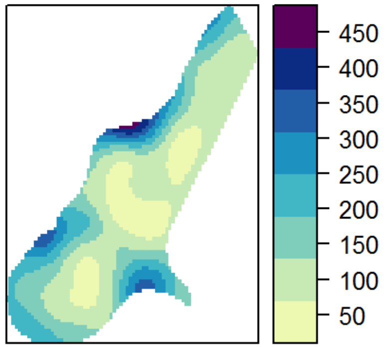
Kriging for spatial prediction

Practice

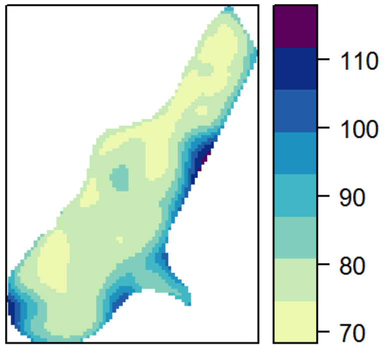
R Code: https://rpubs.com/rez/80464



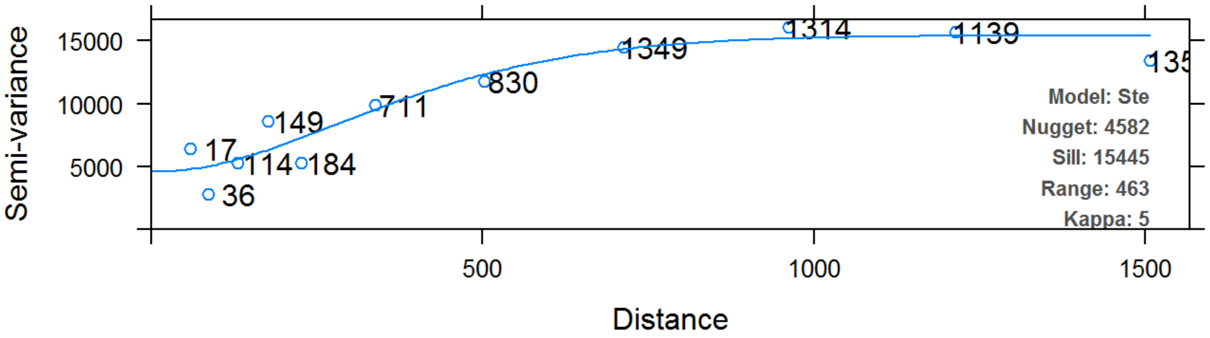
Kriging prediction



Kriging standard error



Experimental variogram and fitted variogram model



Assignment 3

You will have 40 min for working on Assignment 3

Write a 200 word essay on the analysis of spatial prediction using Kriging model (refer to R code user guide).

Send to Email before 10 am next day:
Yongze.song@outlook.com

Document name: A3_YourName.docx

Any questions?