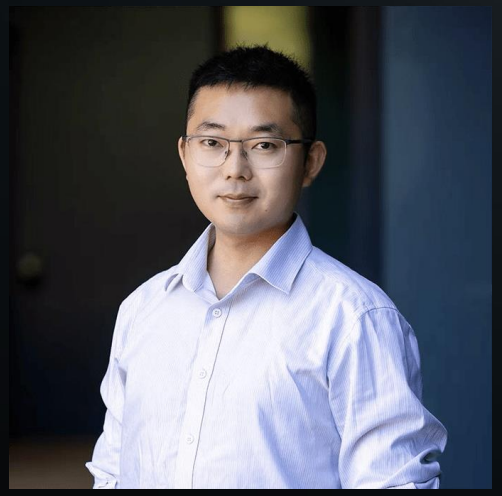
Curtin University | Geospatial Intelligence Lab | International Society for Spatial Methods | HARVARD UNIVERSITY | Spatial Data Lab @Center for Geographic Analysis, IQSS

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

Topic 6. Spatial validation methods

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)

Curtin Geospatial Intelligence Lab
https://yongzesong.com

International Society of Spatial Methods
https://spatialmethods.org

Harvard Spatial Data Lab
https://sdl.gis.harvard.edu

Lecture outline

Lecture website: https://yongzesong.com/lecture-20260104
Password: 20260104

Contents

Part 1. Lecture (1st class)

1. Concepts of spatial validation
2. Spatial validation strategies
3. Spatial validation indicators
4. Spatial block cross-validation

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

1. Practice (40 min)
2. Assignment (40 min)
3. Submit your assignment (10 min)

Tips:

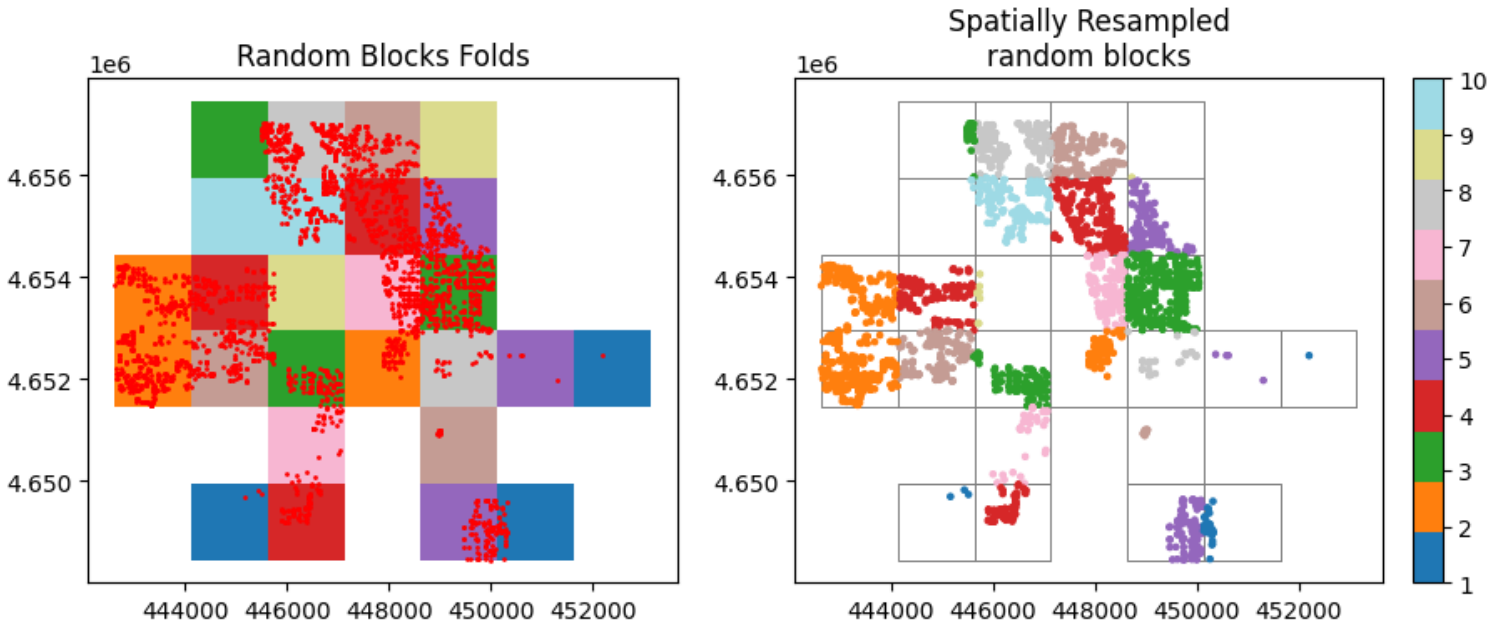
1. Working in a group of 3 or 4
2. Collecting feedback at each day
3. Individual assignments and working in a group
4. Complete assignments in the class
5. Assignment feedback sections after day 2
6. Install R and Rstudio
7. Sign Up for Google Earth Engine

Why Spatial Validation Is Needed

Motivation and problem definition

- Spatial data violate the independence assumption due to spatial autocorrelation.
- Random cross-validation can lead to over-optimistic accuracy.
- Spatial validation aims to evaluate model generalisation in space, not just in sample space.

Good prediction accuracy does not necessarily mean good spatial prediction.

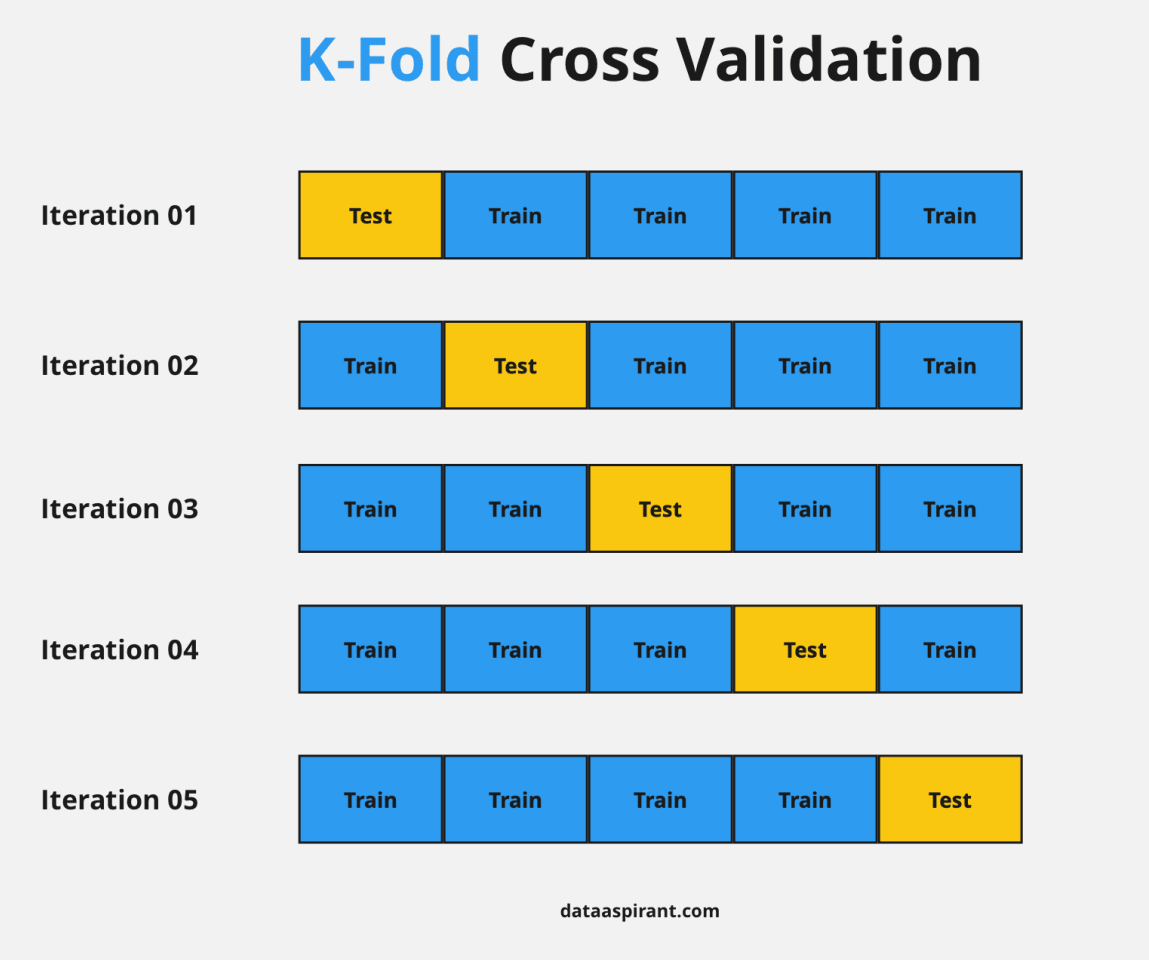


https://pypi.org/project/spatial-kfold/

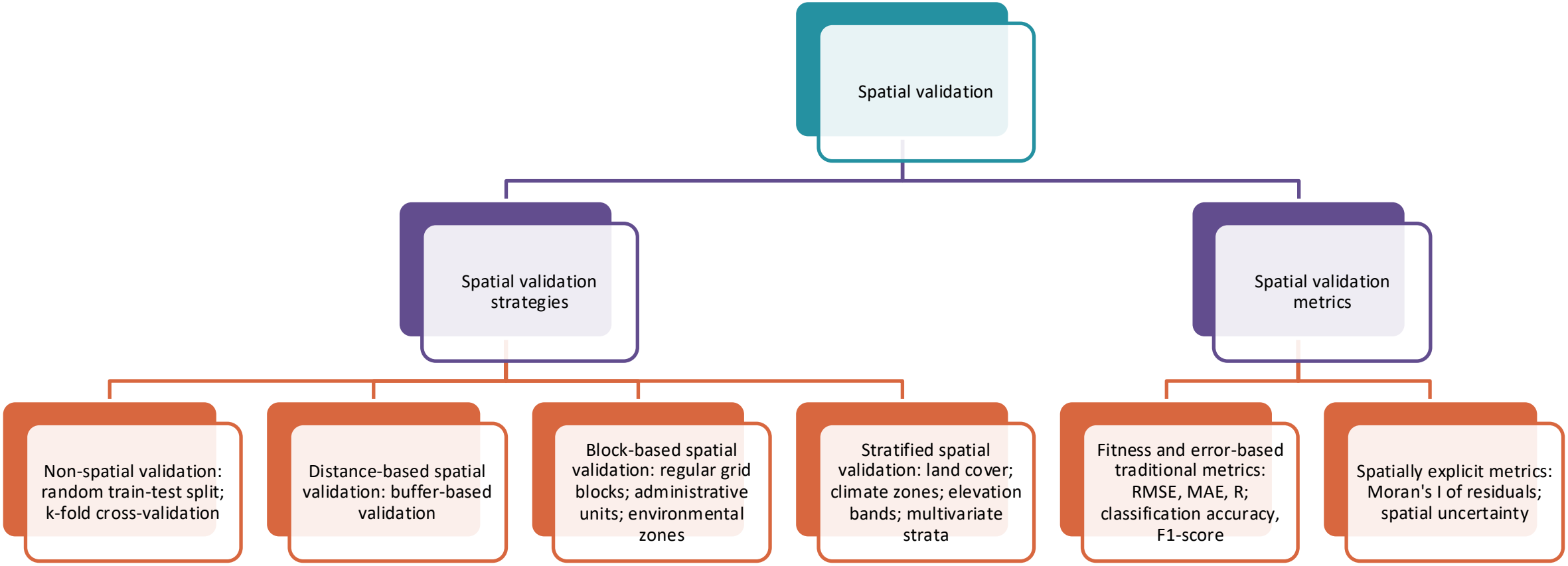
Cross validation

Cross validation evaluates model performance by repeatedly splitting data into training and testing sets, fitting the model on the training data, and assessing predictions on the withheld data.

By aggregating performance across multiple splits or folds, it provides a robust estimate of how well the model generalises to unseen data.

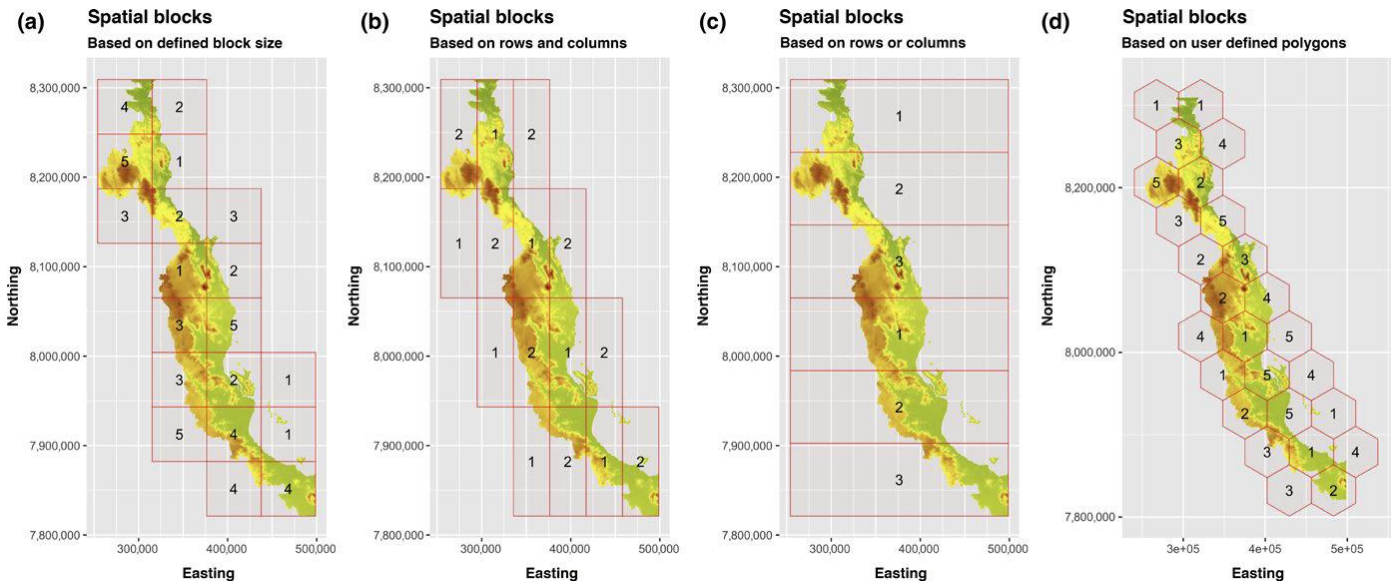


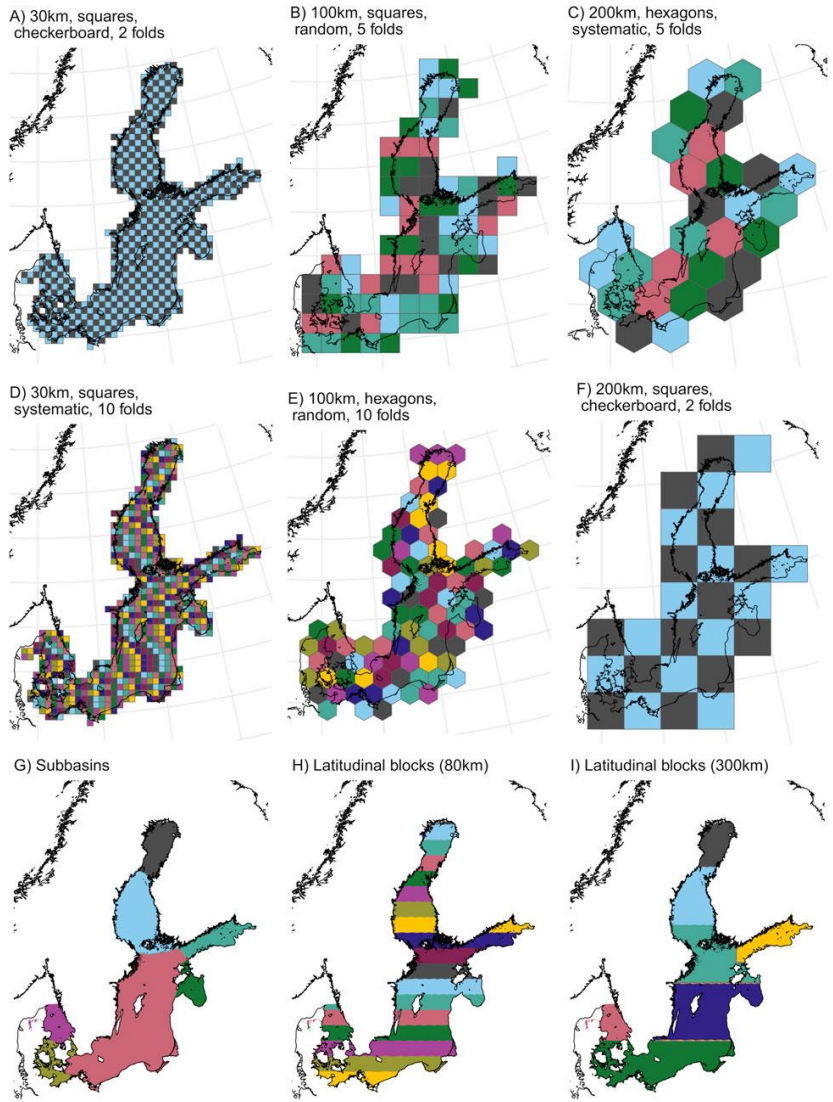
Spatial validation methods



Block cross validation

Block cross validation assesses model performance by partitioning the study area into spatial blocks, training the model on data from some blocks, and testing it on data from held-out blocks to account for spatial dependence.





[blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models](https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2025.1531097/full)

Block cross validation

- In random cross-validation, test points are often located very close to training points, leading to strong spatial autocorrelation leakage and overly optimistic accuracy estimates.

- In block cross-validation, test points are separated into large spatial blocks far from the training data, which can be overly pessimistic and highly sensitive to the chosen block size.

- In nearest neighbour distance matching (NNDM), test points are spatially separated from training points while preserving realistic distance relationships, so each fold respects the natural spatial structure of the data without excessive leakage or over-separation.
    - NNDM (Nearest Neighbour Distance Matching) = a spatially aware CV strategy that selects test points so that their distance relationships with training points mirror the spatial structure of the entire dataset.

https://cran.r-project.org/web/packages/blockCV/vignettes/tutorial_2.html

Block cross validation

Practice

R Code:

https://github.com/rvalavi/blockCV

https://cran.r-project.org/web/packages/blockCV/vignettes/tutorial_1.html

https://cran.r-project.org/web/packages/blockCV/vignettes/tutorial_2.html

https://geocompx.org/post/2025/sml-bp6/

Assignment 6

You will have 40 min for working on Assignment 6

Write a 200 word essay on the analysis of spatial block cross validation

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

Document name: A6_YourName.docx

Any questions?