TU Delft

Intro to Spatio-Temporal Modelling

Yanan Xin (yanan.xin@tudelft.nl)

19-3-2025

Spatio-Temporal Data

Temporal Data (Wikle et al. 2019)

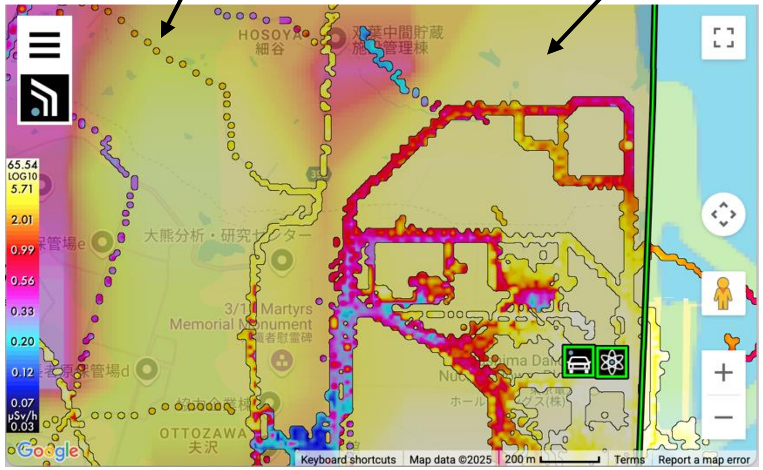
- Regular or irregular intervals
- Continuous or discrete time
- Random events – point process

Spatial Data

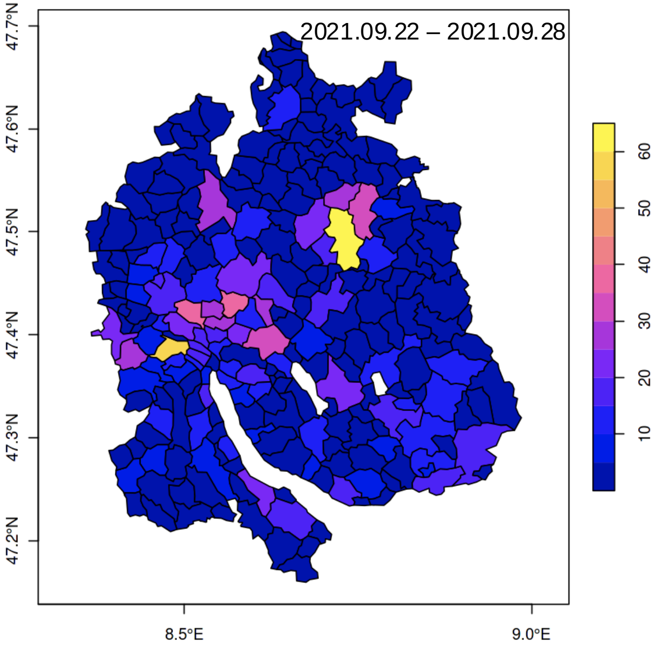
- Continuous spatial data
- Lattice data
- Spatial point process (point pattern)
- Network data
- Objects (e.g., trajectories)

Point measurements along trajectories

Interpolated continuous spatial data



Safecast radiation map (https://map.safecast.org/)



Lattice data

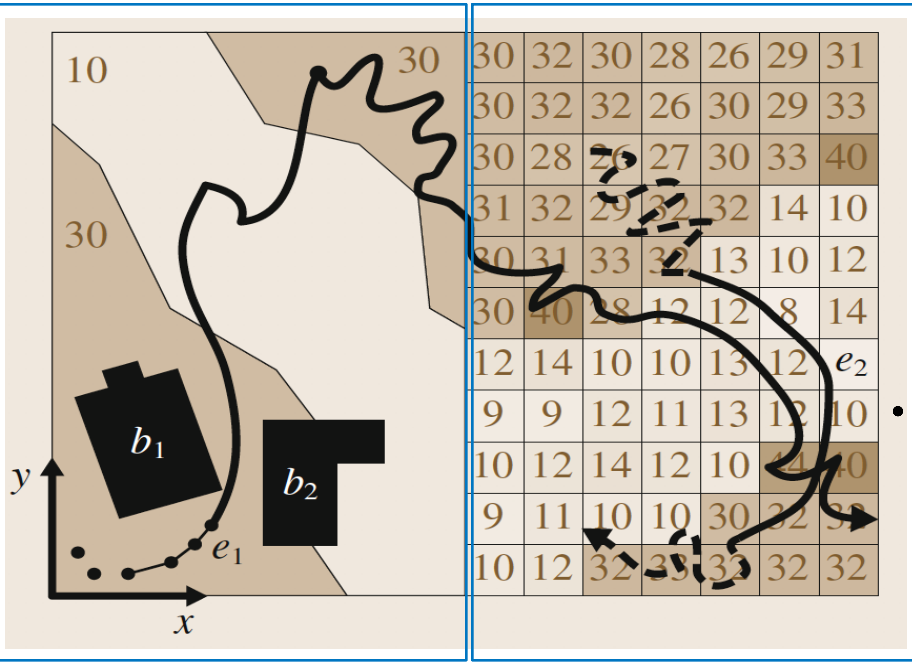
7-day new COVID cases in Zurich by zip code

TU Delft

Spatial Data Representations

Vector Data

- More accurate shape delineation
- Irregularly-shaped spatial objects



Raster Data

- Faster computations
- Regularly-shaped spatial data (pixels)

Image from (Gudmundsson et al. 2011)

Implications:

Modifiable Areal Unit Problem (MAUP), Precision and Uncertainty

Special Effects of Spatial Data

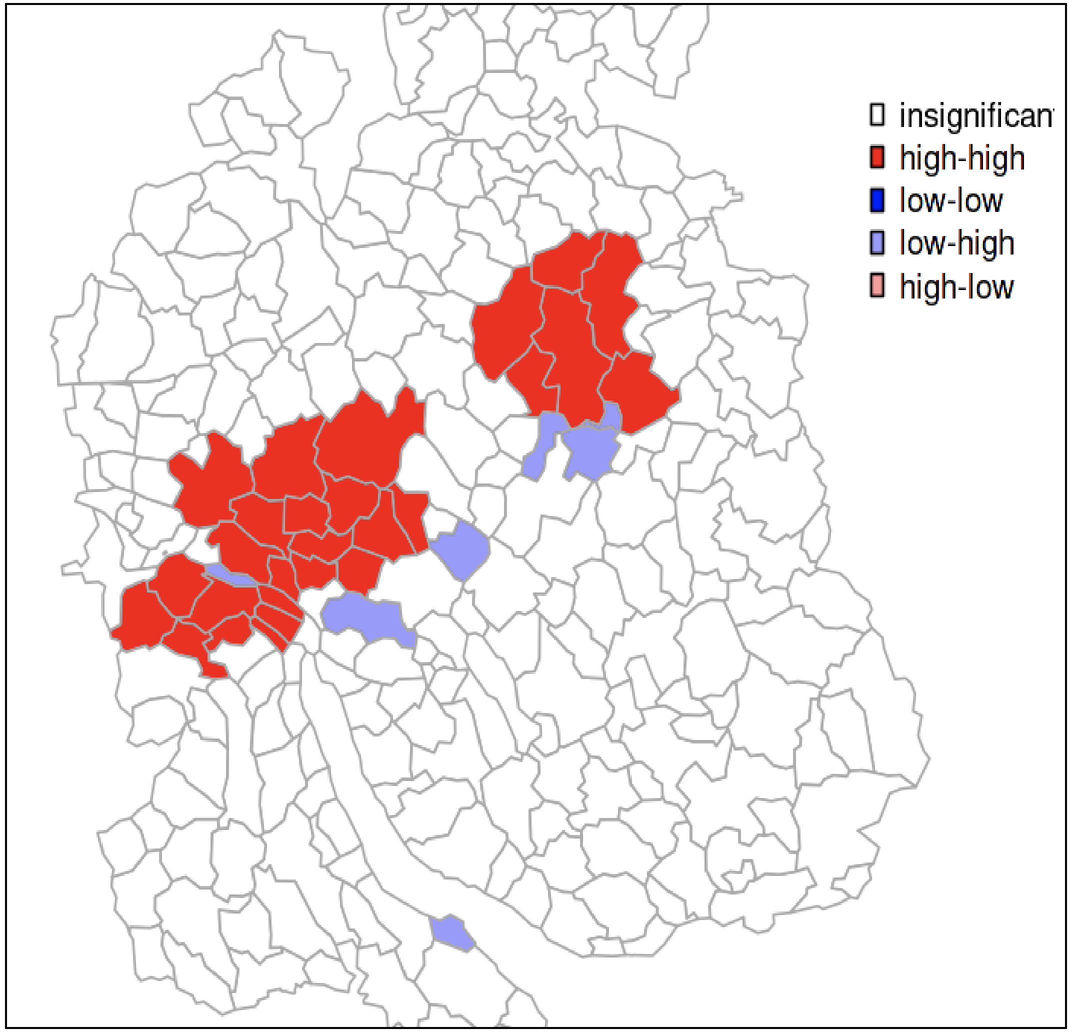
violate i.i.d assumption

Spatial dependence – interactions among spatial units

Spatial heterogeneity – spatially varying processes

Spatial Dependence or Autocorrelation

Object values are spatially clustered (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation)

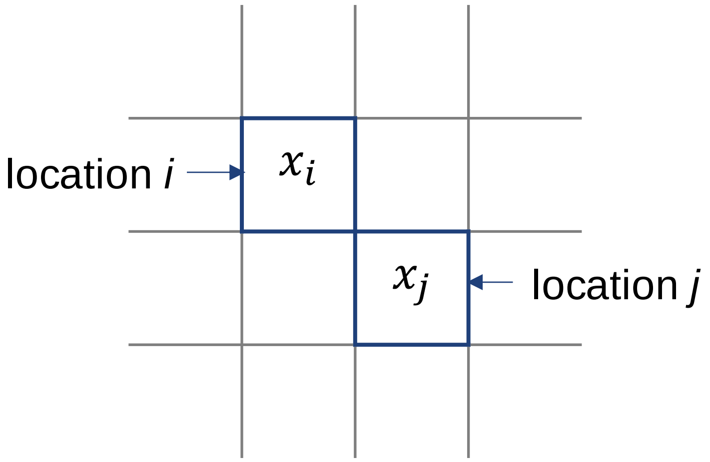


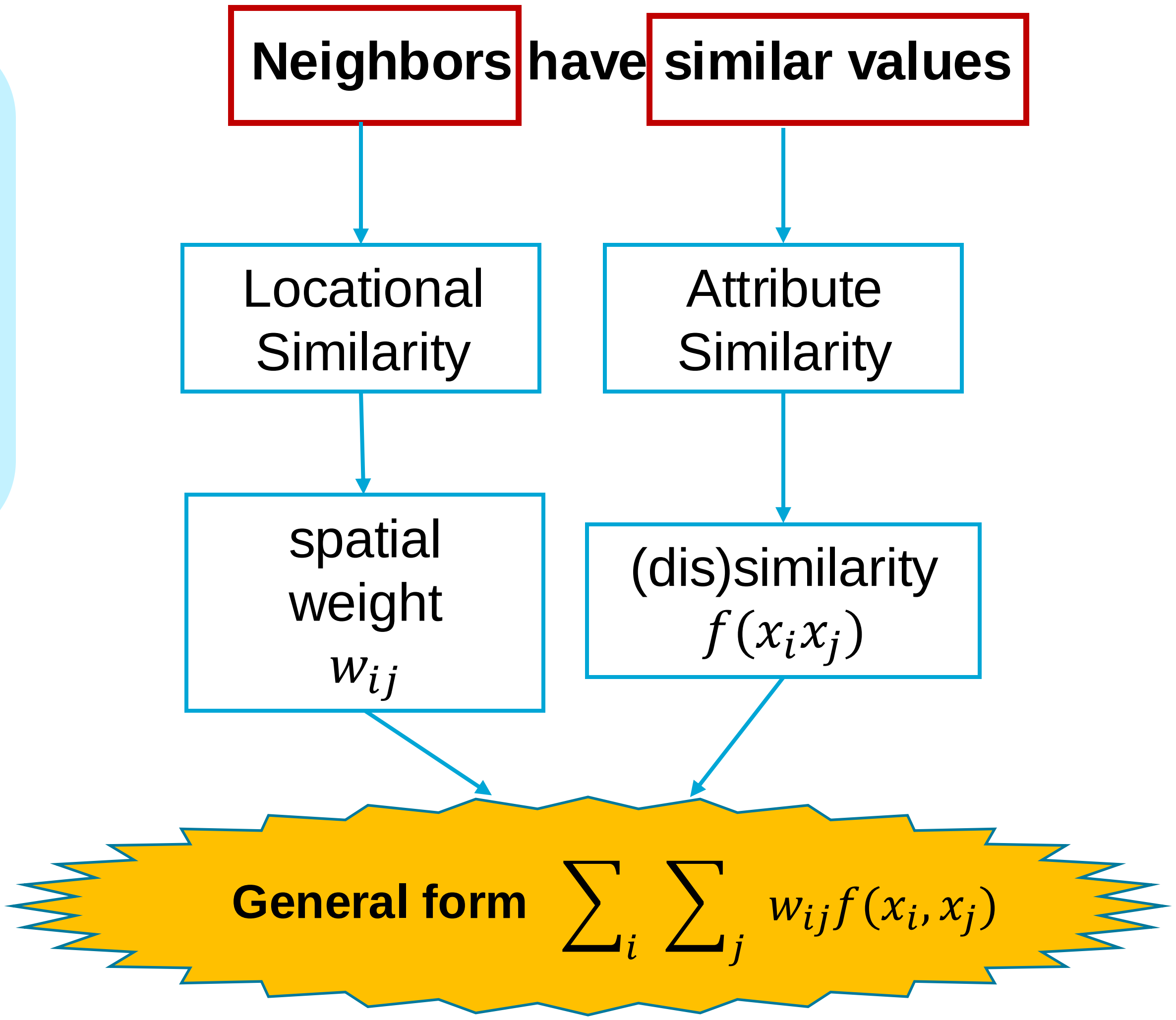
Spatial Autocorrelation Statistic

First law of geography

“everything is related to everything else, but near things are more related than distant things.”

-- Waldo Tobler (1970)

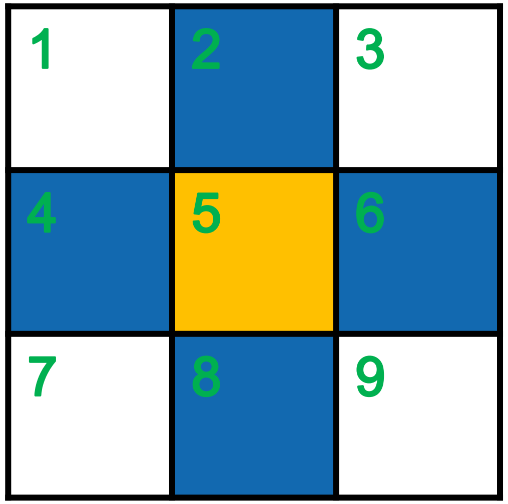




Attribute Similarity

- Measure the similarity (or dissimilarity) of a variable *x* at different locations *i* and *j*
- Calculate $f(x_i, x_j)$
  - *Covariance* $(x_i - \bar{x}).(x_j - \bar{x})$ *(e.g., Moran's I)*
  - *Squared distance* $(x_i - x_j)^2$ *(e.g.,* Geary's C *)*

Locational Similarity - Spatial Weights



Rook Contiguity

$$W_{ij} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Row standardization

$$W'_{ij} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Spatial Weights

- Contiguity-Based Weights
  - Queen Contiguity Weights
  - Rook Contiguity Weights
- Distance-Based Weights
  - Distance-Band Weights
  - K-Nearest Neighbor Weights
  - Kernel Distance Weights
- Mixed Weights
  - Contiguity + KNN
- Interaction-Based Weights
  - Number of commuters between places

Global Moran's I

- Test whether observations with high (or low) values are surrounded by similar high (or low) values.
- $f\left(x_i, x_j\right) = (x_i - \bar{x})(x_j - \bar{x})$

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})\left(x_j - \bar{x}\right) \; /W}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \; /n}$$

x_i : feature value of observation at location i
$\bar{x}$: mean of the feature *x*
n : number of samples
w_{ij}: spatial weight between observation at *i* and observation at *j*
$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$: sum of all spatial weights of all pairs (*i, j*)

The Moran Scatter Plot

- If we row-standardize the weights and have zero diagonal, $W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} = n$

weighted average of neighboring values => "spatially lagged value"

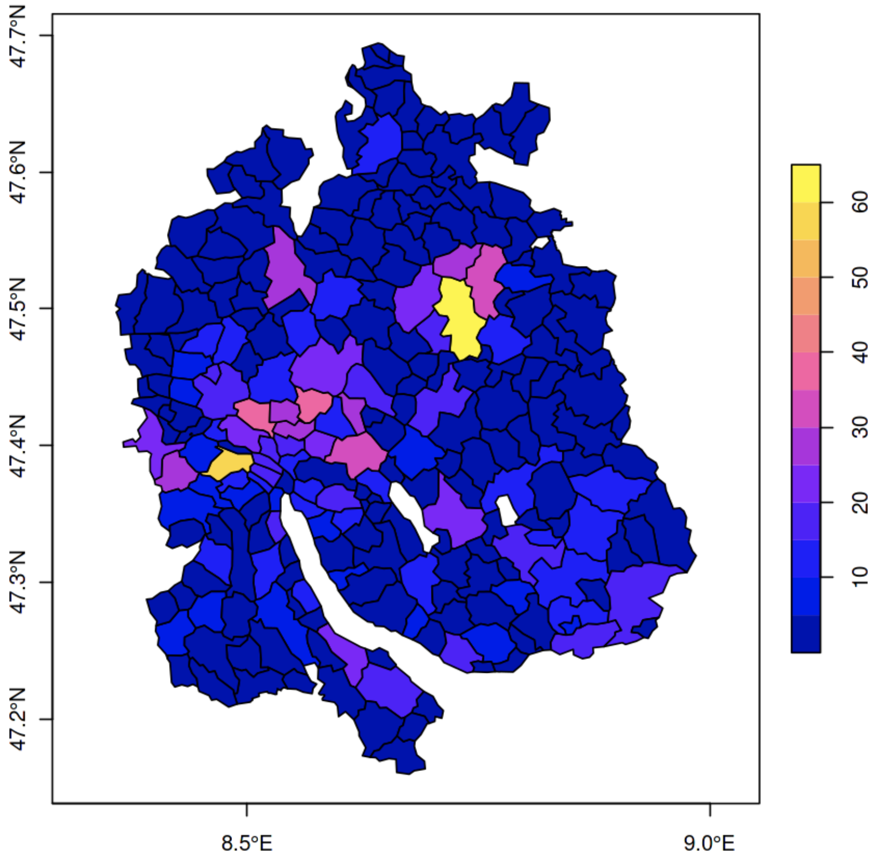
$$z_i = x_i - \bar{x}$$

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x}) / W}{\sum_i (x_i - \bar{x})^2 / n} = \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} = \frac{\sum_i (z_i \times \sum_j w_{ij} z_j)}{\sum_i z_i^2}$$

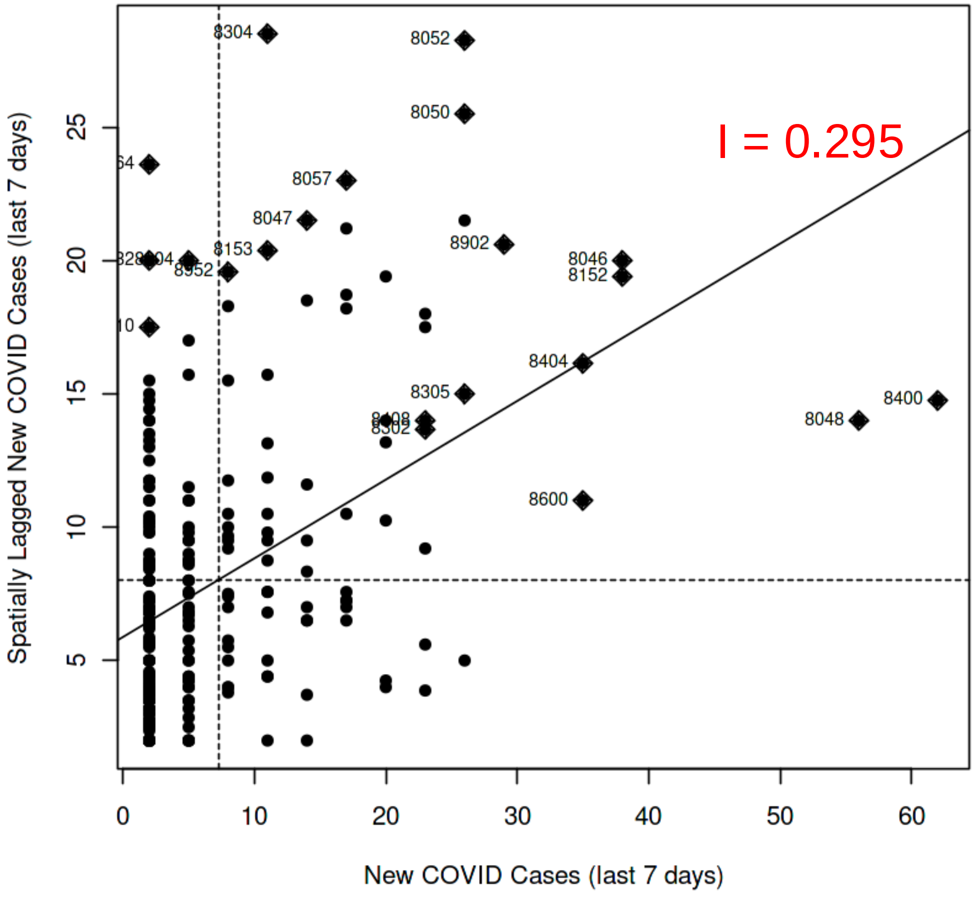
$y = \alpha + \beta x$, the least squares estimate for β is $\frac{\sum_i (x_i \times y_i)}{\sum_i x_i^2}$

- The Moran's I value can be calculated by fitting a linear regression line between spatially lagged values and the original value. The slope of the linear fit equals Moran's I.

The Moran Scatter Plot



Seven-Day Average of New COVID Cases in Zurich by Zip Code from 2021-09-22 to 2021-09-28.



Moran Scatter Plot of the Seven-Day New COVID Cases.

Significance Test of Moran's I

- Inference for spatial autocorrelation is based on a null hypothesis of spatial randomness.
- Computational approach to calculate a pseudo p-value -> **permutation**

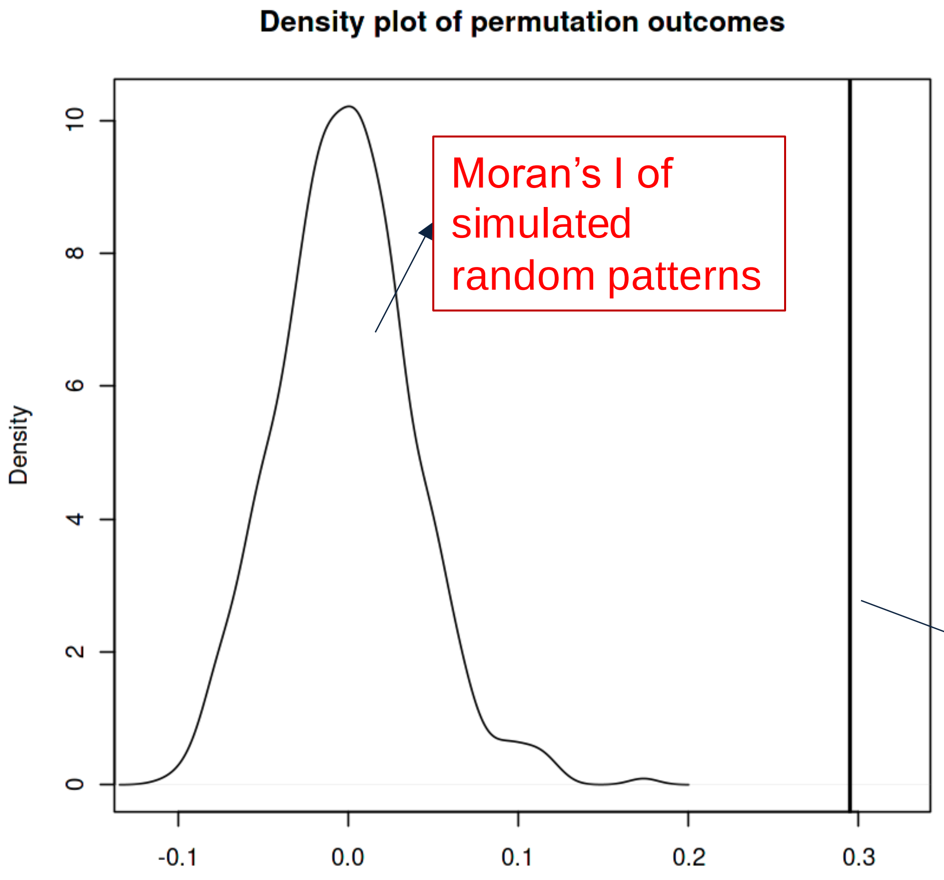
$$p = \frac{R+1}{M+1}$$

R is the number of times the computed Moran's I is more extreme than the one calculated with our observed pattern.

M is the total number of permutations (e.g., 99, 999, or 9999)

- **Alert**:
  - 1) It is a pseudo p-value and should not be interpreted as an analytical p-value.
  - 2) The significance level depends on the number of permutations M.

Significance Test of Moran’s I



Monte Carlo Simulation of Moran’s I

```
        Monte-Carlo simulation of Moran I

data:  plz_cases$new_case_7days
weights: lw
number of simulations + 1: 1000

statistic = 0.295, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater
```

Moran’s I of the observed pattern

Spatial Regression

- Spatial Lag Model
- Spatial Error Model

Spatial Lag Model

The value of the **dependent variable** at one location is associated with the value of its neighbours.

=> The model uses **spatially lagged dependent variable** as a predictor.

$$OLS: \quad y = \underbrace{\beta_0}_{\substack{Intercept \\ (Constant)}} + \underbrace{\beta_1 X_1 + \dots + \beta_n X_n}_{Predictors} + \underbrace{\varepsilon}_{Residuals}$$

$$Spatial\ Lag: \quad y = \underbrace{\rho W y}_{Lag\ of\ y} + \underbrace{\beta_0}_{\substack{Intercept \\ (Constant)}} + \underbrace{\beta_1 X_1 + \dots + \beta_n X_n}_{Predictors} + \underbrace{\varepsilon}_{Residuals}$$

Spatial Lag Model

$y = \rho W y + X\beta + \varepsilon$

Reduced form:

(I−ρW)y =$X\beta + \varepsilon$

$y = (I - \rho W)^{-1} X\beta + \varepsilon$

The expected change in y given the change in x is:

$E[y|\Delta X] = (I - \rho W)^{-1}(\Delta X)\beta = [1 + \underline{\rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots}](\Delta X)\beta$

direct effect   indirect effect

Spatial multiplier effect

Spatial Error Model

Residuals at one location is associated with residuals of its neighbours. => The model uses **spatially lagged residuals** as a predictor.

$$\text{Spatial Error}: \begin{cases} y = \underbrace{\beta_0}_{\substack{\text{Intercept} \\ \text{(Constant)}}} + \underbrace{\beta_1 X_1 + \ldots + \beta_n X_n}_{\text{Predictors}} + \varepsilon \\ \varepsilon = \underbrace{\lambda W \varepsilon}_{\text{Spatially Lagged Residuals}} + \underbrace{u}_{\text{Random Noise}} \end{cases}$$

$$y = \underbrace{\beta_0}_{\substack{\text{Intercept} \\ \text{(Constant)}}} + \underbrace{\beta_1 X_1 + \ldots + \beta_n X_n}_{\text{Predictors}} + \underbrace{\lambda W \varepsilon}_{\text{Spatially Lagged Residuals}} + \underbrace{u}_{\text{Random Noise}}$$

Spatial Error Model

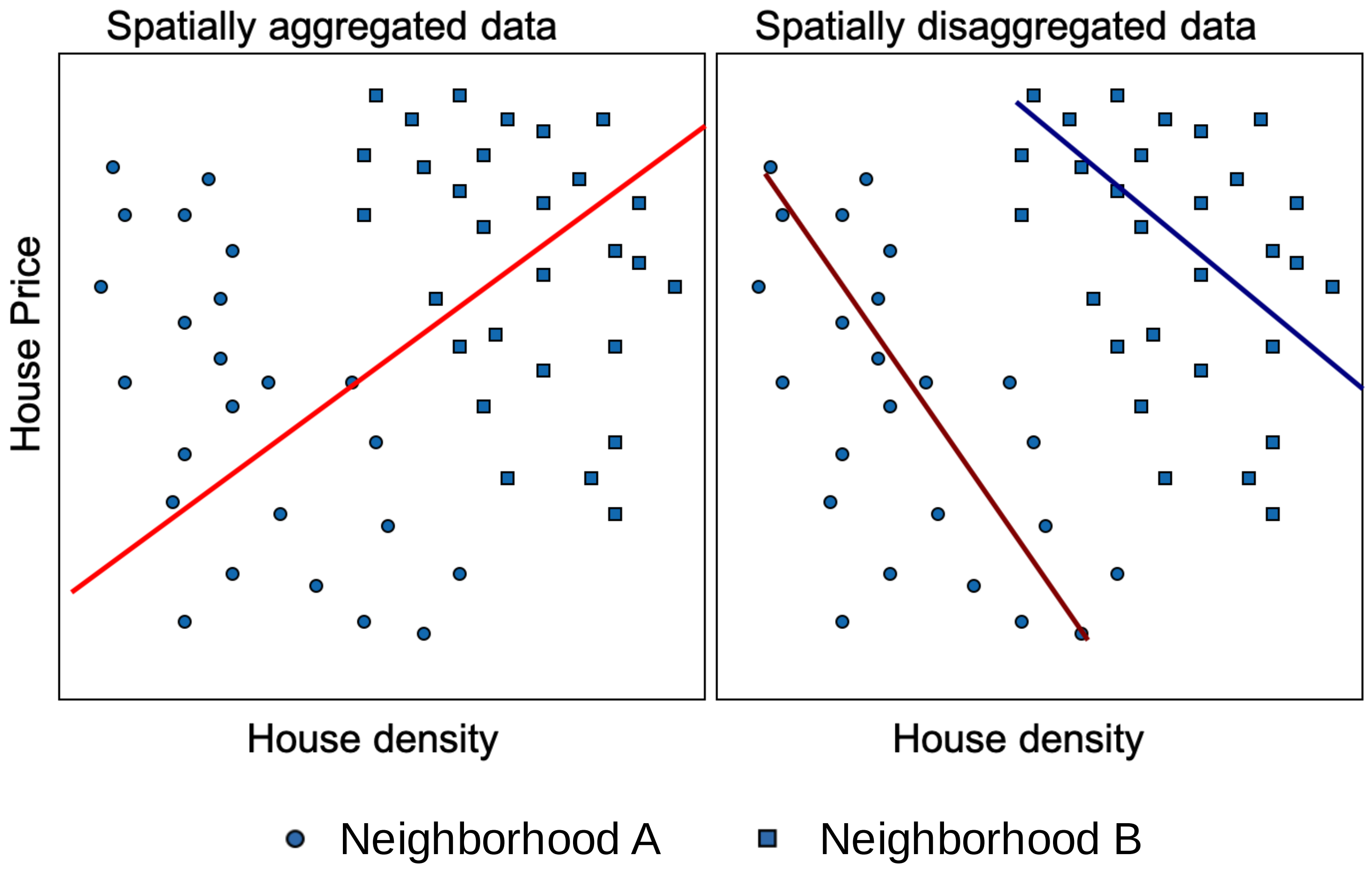
Reduced form:

$$y = X\beta + (I - \lambda W)^{-1}e$$

The spatial autocorrelation only impacts the error term

The effect is relatively local

Spatial Heterogeneity



- Spatial processes vary across locations

Geographically Weighted Regression (GWR)

- A typical linear regression model OLS assumes a spatial stationary process, β *is the same for all observations* $i \in \{1, 2, \dots, n\}$

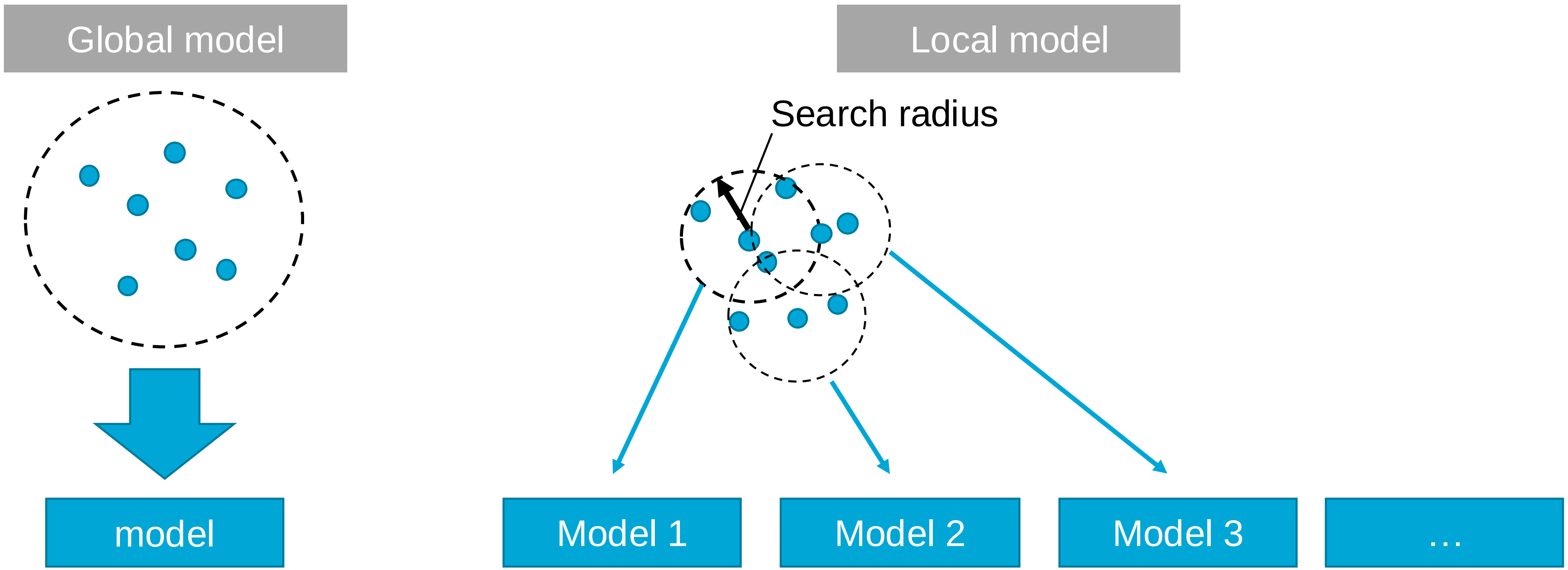
$$y_i = \sum_{j=0}^{k} \beta_j x_{ij} + \varepsilon_i \qquad (1)$$

- GWR allows the β to vary over locations (u_i, v_i) by fitting a local regression using a subset of observations

$$y_i = \sum_{j=0}^{k} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \qquad (2)$$

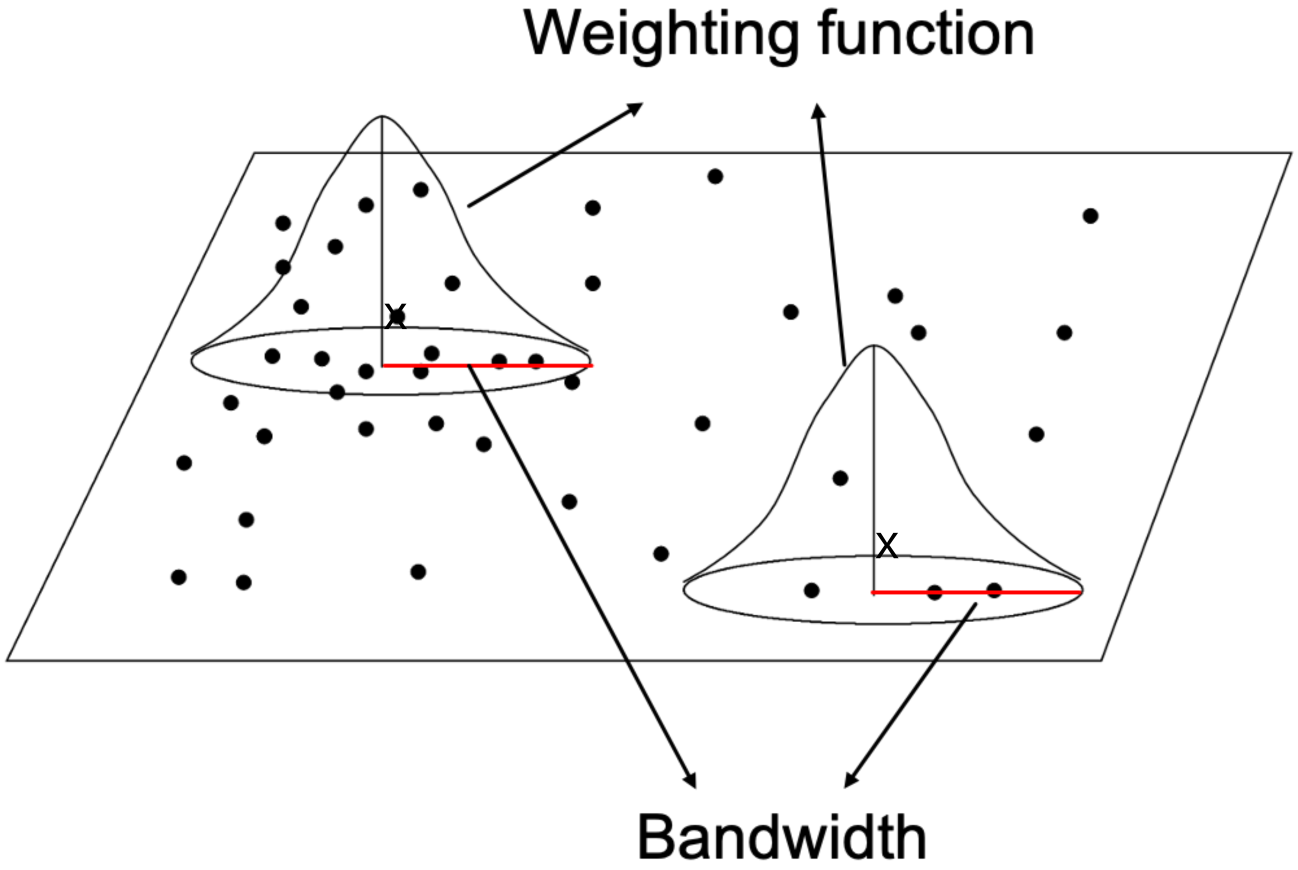
Brunsdon, C., Fotheringham, S. and Charlton, M., 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(3), pp.431-443.

Global Model vs. Local Model



Important Parameters of GWR

- Local search radius -> bandwidth
- The importance of selected points on the target regression point is weighted using kernel functions

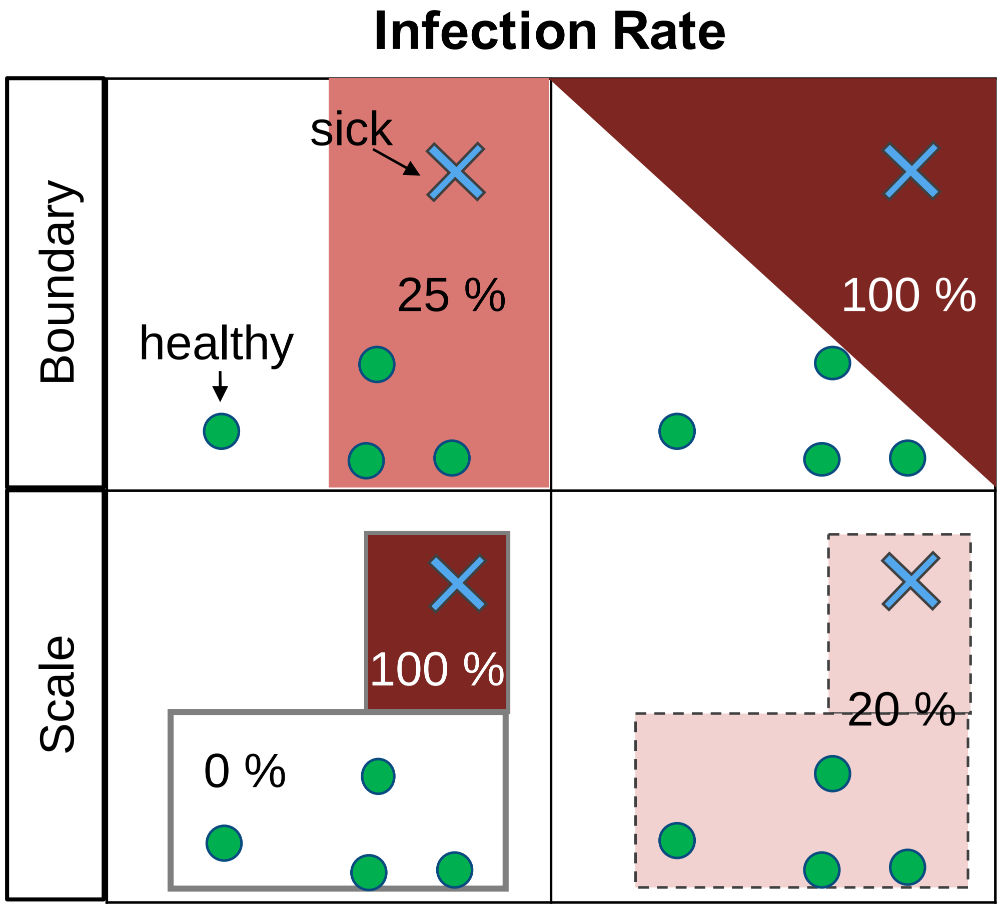


References

- Anselin, L., 1988a. *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.
- Brunsdon, C., Fotheringham, S. and Charlton, M., 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(3), pp.431-443.
- Gudmundsson, J., Laube, P. and Wolle, T., 2011. Computational movement analysis. In Springer handbook of geographic information (pp. 423-438). Springer, Berlin, Heidelberg.
- Wikle, C.K., Zammit-Mangion, A. and Cressie, N., 2019. *Spatio-temporal statistics with R*. Chapman and Hall/CRC.

U

Modifiable Areal Unit Problem

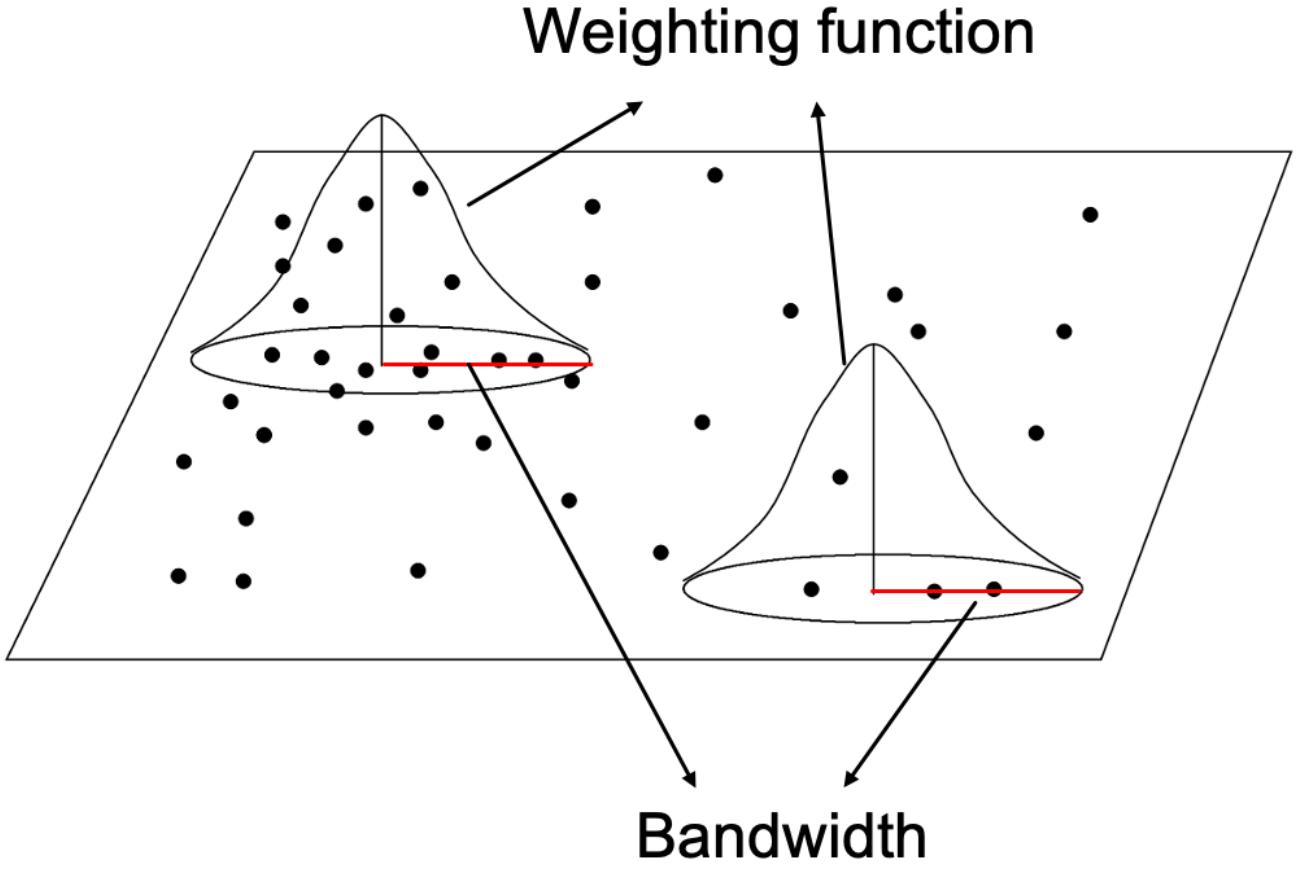


Openshaw, S. 1984. The modifiable areal unit problem. Norwich, UK: Geo Books.

GWR - Bandwidth

Fixed Bandwidth:

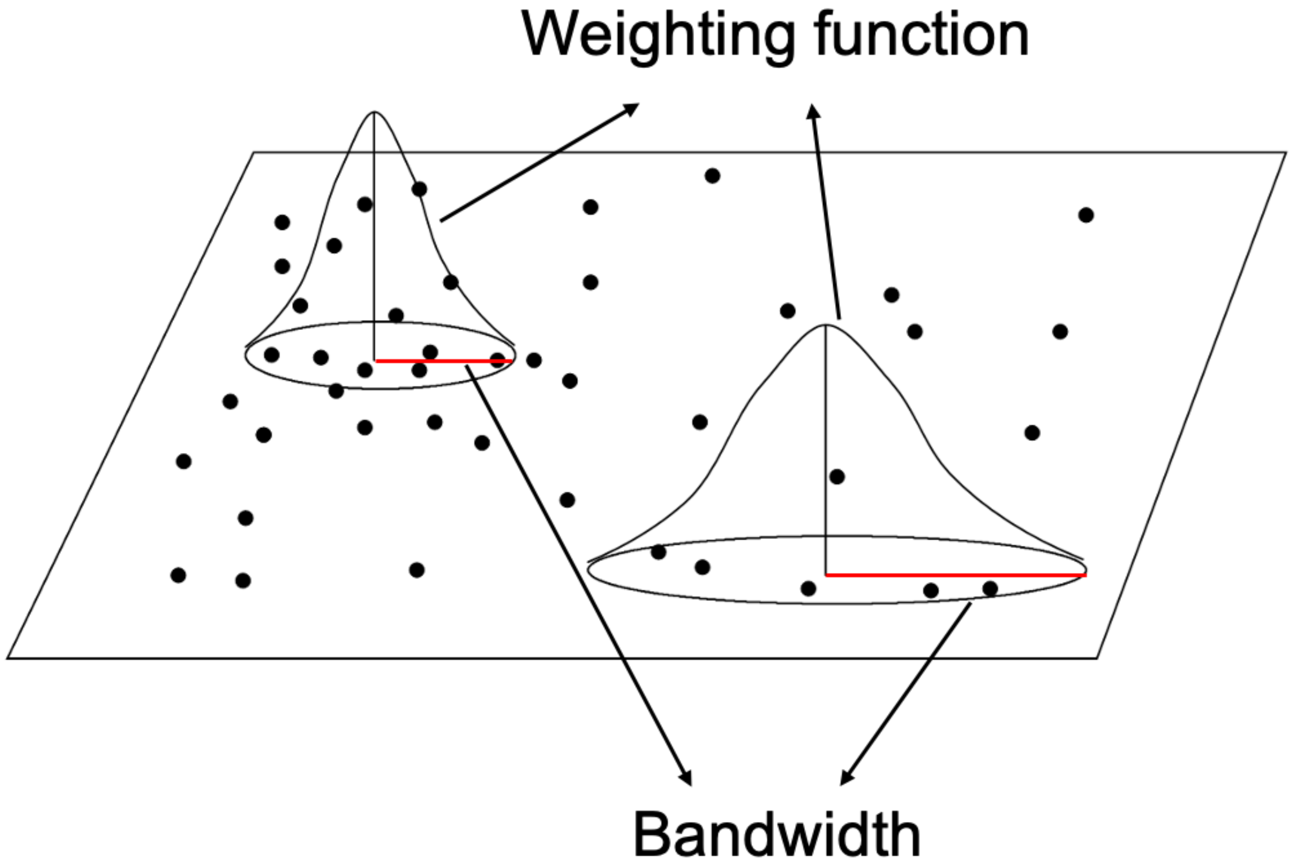
- the same distance bandwidth is used everywhere
- might produce large estimate variances where data are sparse



GWR - Bandwidth

Adaptive Bandwidth:

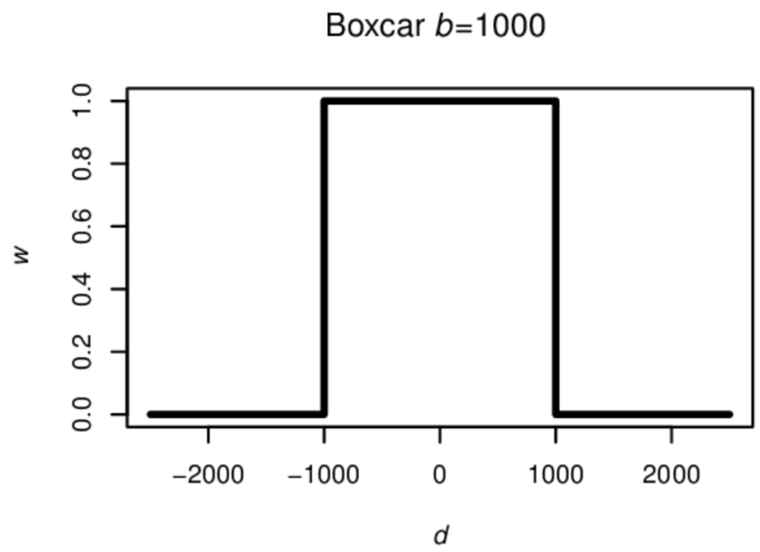
- keep the number of neighbors constant by adjusting the distance bandwidth
- ‘localness’ changes according to the density of data



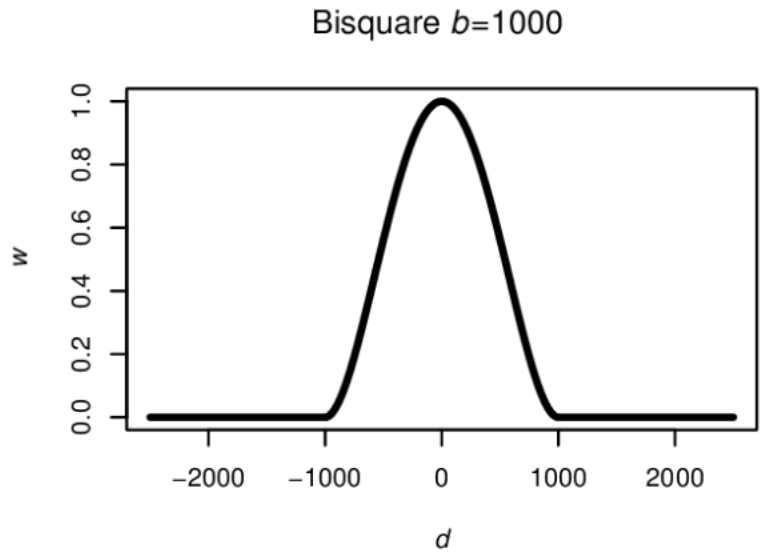
GWR – Kernel Functions

Points within a bandwidth are weighted based on spatial proximity

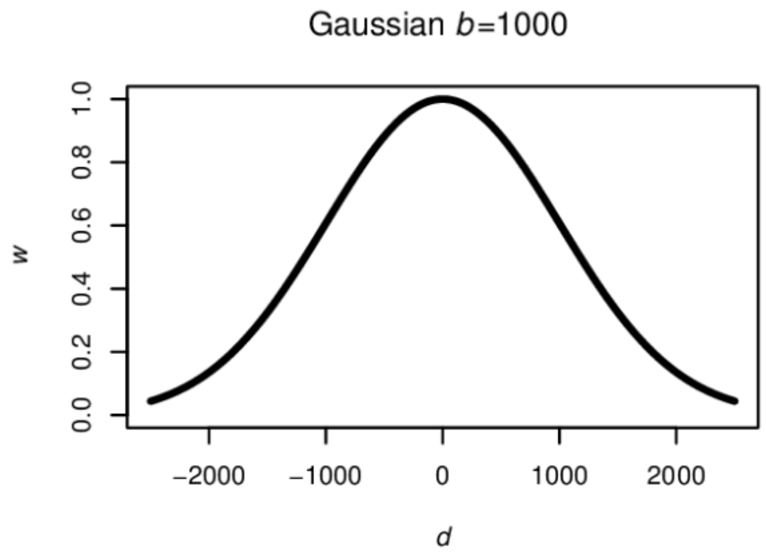
- Kernel weighting schemes:
  - Gaussian, Boxcar, Bisquare



$$w_{ij} = \begin{cases} 1, & |d_{ij}| < b \\ 0, & otherwise \end{cases}$$



$$w_{ij} = \begin{cases} [1-(d_{ij}/b)^2]^2, & |d_{ij}| < b \\ 0, & otherwise \end{cases}$$



$$w_{ij} = \exp[-(d_{ij}/b)^2/2]$$

(Gollini et al. 2015)

GWR - Bandwidth

How to choose the bandwidth?

- Prior knowledge of the scale at which the spatial process operates
- Leave-one-out cross validation (prediction accuracy)
- AICc, corrected Akaike Information Criterion (trade-off between accuracy and complexity)

Leave-one-out cross validation

$$\sum_{i=1}^{n}[y_i - y^*_{\neq i}(h)]^2$$

$y^*_{\neq i}(h)$ is the fitted value of y_i with the ith observation omitted

$$AICc = 2nln(\hat{\sigma}) + nln(2\pi) + n\frac{n + tr(S)}{n - 2 - tr(S)}$$

n: local sample size
$\hat{\sigma}$: estimated standard deviation of the error term;
$tr(S)$: trace of the hat matrix S.