Geographically Weighted Regression GWR

Fifth Session

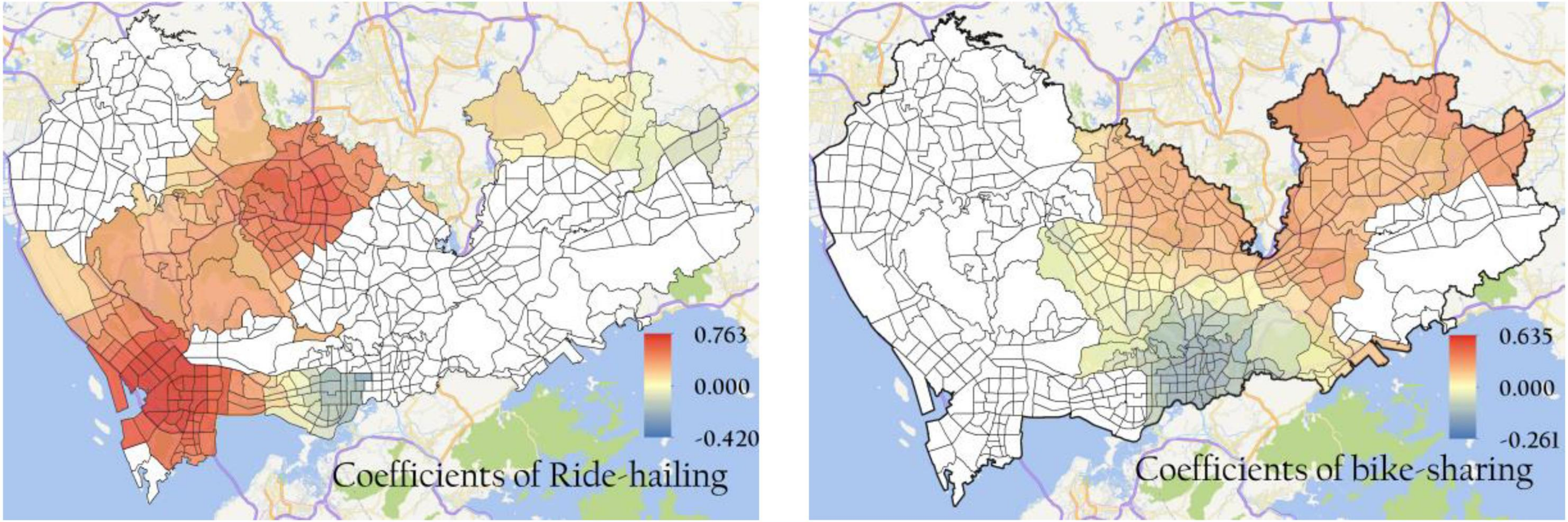
Orlando Sabogal-Cardona
PhD researcher
University College London UCL

Do you remember spatial heterogeneity?

Photo by NASA on Unsplash

Intuition

Outcome variable: taxi ridership

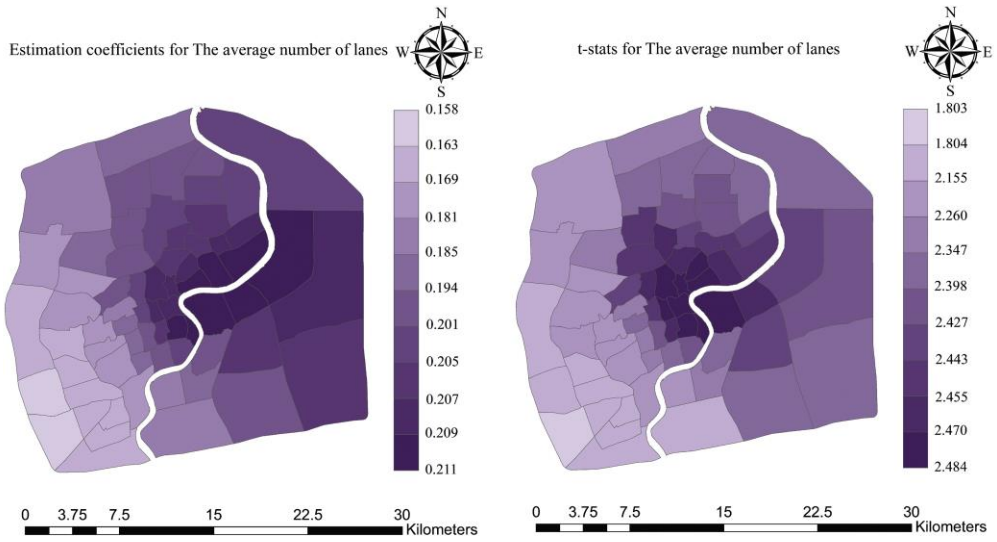


(a) Coefficients of ride-hailing

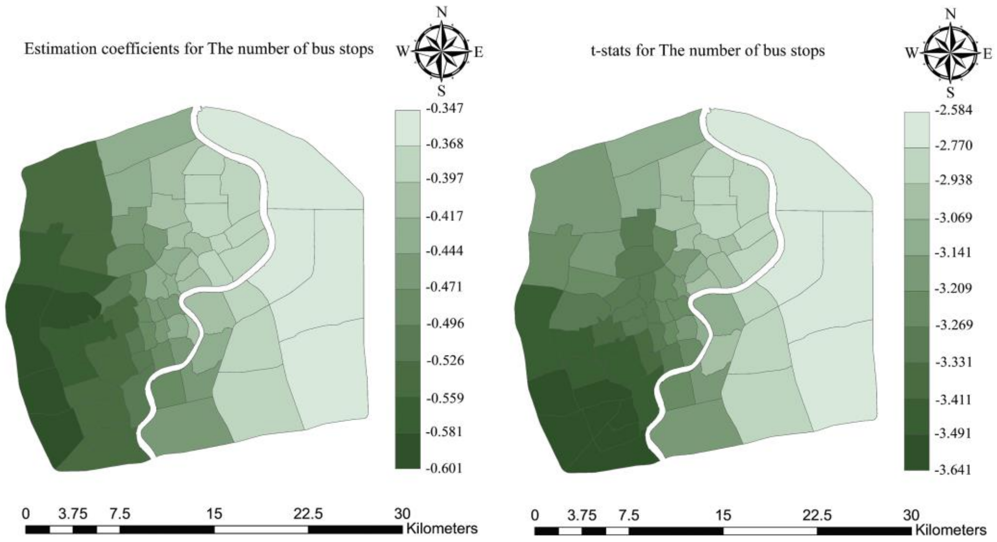
(b) Coefficients of bike-sharing

Source: Tang et al., (2021)
Available at: https://www.sciencedirect.com/science/article/pii/S0966692321001873?casa_token=pM_KHhZBHGMAAAAA:khXqhNNg4gW3Cnxr3NHp9fPLCzmqvPARG7ucCIgn93lwNYtz7IF0tTSr7_aaaEc6Liwm5c_h7cY#s0030

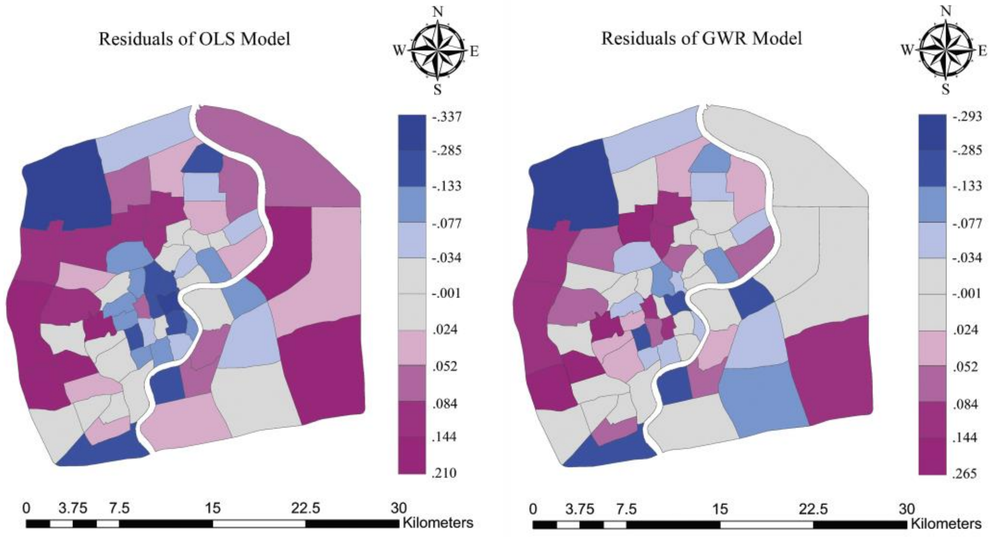
Intuition



(a) Estimated coefficients (b) t-statistics



(a) Estimated coefficients (b) t-statistics



(a) Residuals of OLS (b) Residuals of GWR

Outcome variable: TSI (traffic state index) – road traffic congestion

Source: Pan et al., (2020)
Available at: https://www.sciencedirect.com/science/article/pii/S0966692319305988?casa_token=fLGE-kWbo2UAAAAA:pdUV4Vmvdhz8gcoFGl_rGohvtN1nFKvfX2dGQnqFHACKJnrXU3xYvqWv3VchDWjw0dVkZXJ4C0s#f0030

Intuition

- Geographically Weighted Regression (GWR) is a spatial regression technique used to analyze and model spatially varying relationships between variables. It extends traditional regression analysis by allowing the relationships between the dependent variable and independent variables to vary across different locations in space.
- In standard regression models, the relationships between variables are assumed to be constant across the entire study area. However, in many real-world scenarios, relationships may exhibit spatial non-stationarity, meaning that the relationship strength or direction can vary across different locations.

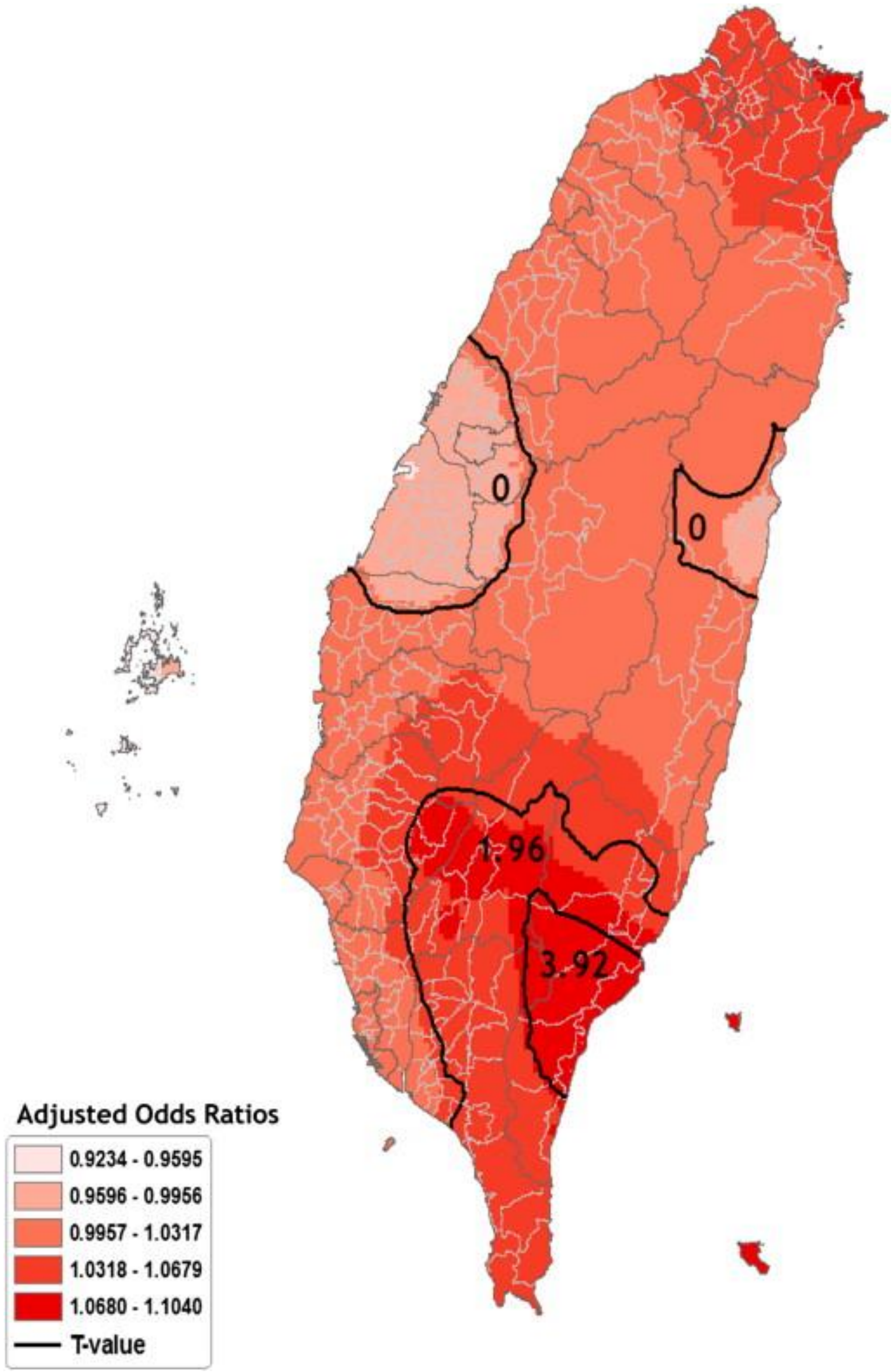
- GWR addresses this spatial non-stationarity by estimating local regression models at each location, considering only the observations within a specific neighborhood or bandwidth. The GWR model calculates separate coefficients for each independent variable at each location, taking into account the neighboring observations

Intuition

Smoothed varying association between obesity & township disadvantages

"··· presents the local variability in the adjusted OR estimates for the township disadvantage index. As shown, the estimated ORs ranged from 0.923 to 1.104, supporting the hypothesis of spatial variation. Fig. 2 also presents t-statistics, which indicate the level of statistical significance. In one particular area of southeastern Taiwan, the relationship between the township disadvantage index and obesity is highly significant (t-statistic = 3.92), while this relationship is less significant or even insignificant in other regions. Further statistical analysis shows that one particular area of southeastern Taiwan has significantly lower average wages, shorter work days, and higher unemployment rates than the rest of Taiwan (results not shown in the Tables)"

Source: Chen and Truong, 2012
Available at:
https://www.sciencedirect.com/science/article/pii/S0143622811001597?casa_token=bCWHmiWVZ9sAAAAA:RVjCIWGLZ6dCJc11klOmzYJ6YF5GUiakSwsbhDBxXq-IlmeWZhAuZLKIUMLF5soUnAojyHTHDqY#fig2



Methodology

“To estimate the model, the following hypothesis is used**: the closer two observations are geographically, the more similar the influence of the explanatory variables on the dependent variable, i.e. the closer the coefficients of the explanatory parameters of the regression.”** (Handbook of Spatial Analysis - INSEE Eurostat 2018, page 234)

“The solution is therefore to **reduce the importance of the most remote observations by giving each observation a decreasing weight with the distance to the point of interest.**” (Handbook of Spatial Analysis - INSEE Eurostat 2018, page 235)

Methodology

We need "something like" a W weighted matrix. In this case: "Matrix $W_{(u_i,v_i)}$ contains the weight of each observation according to its distance to the point i of coordinates (u_i, v_i)" (page 236)

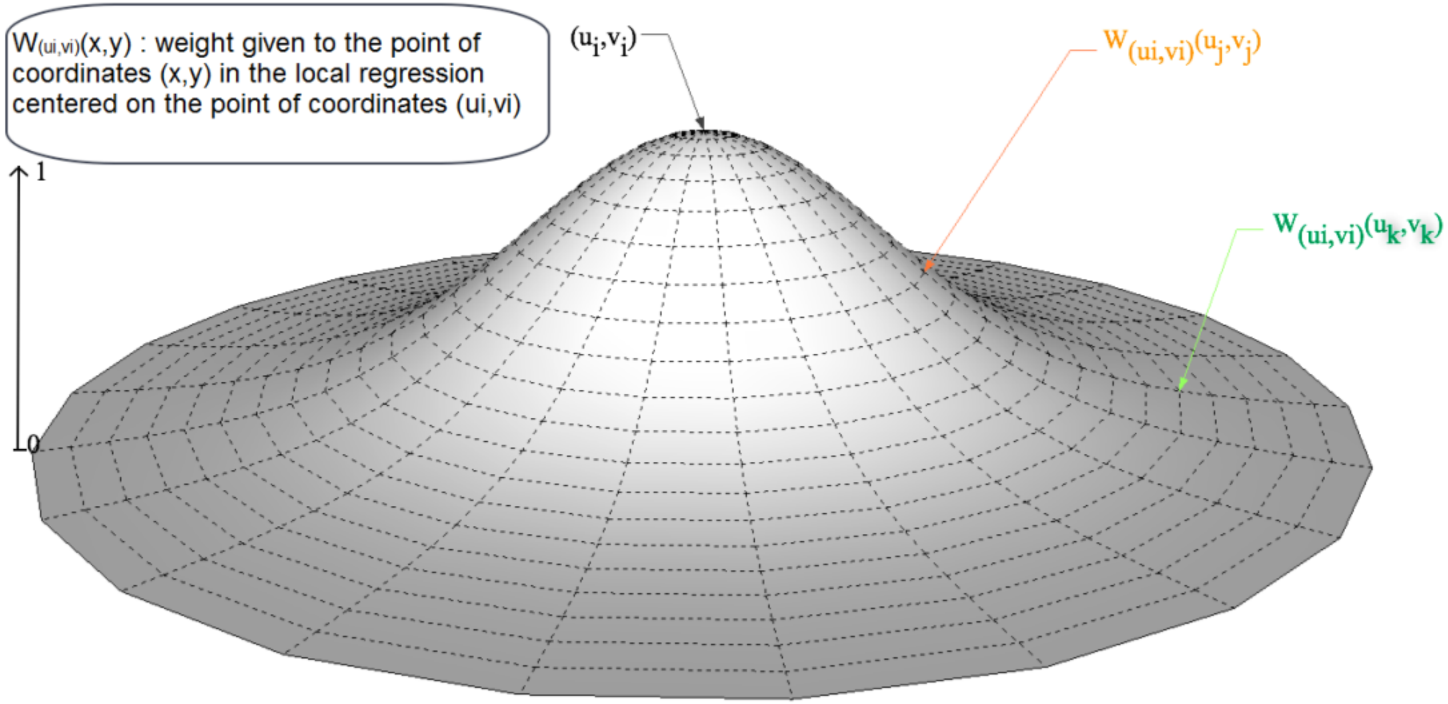


Figure 9.3 – Graphical representation of matrix W

Source: Handbook of Spatial Analysis - INSEE Eurostat 2018, page 236

Kernel

A **Kernel function** determines the weight of each observation with respect to any specific point. There are three key parameters:

- Shape: continuous (considers all observations) and compact support (some observations are weighted to zero after a distance)
- Fixed or Adaptive
- Bandwidth size (strongest influence)

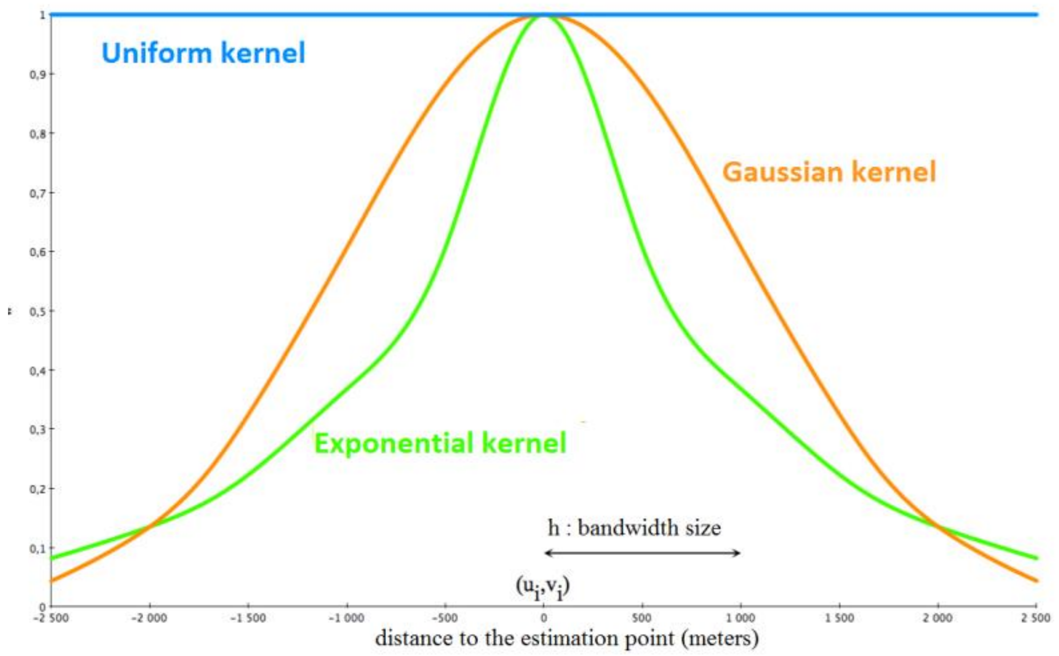
Shape of the Kernel

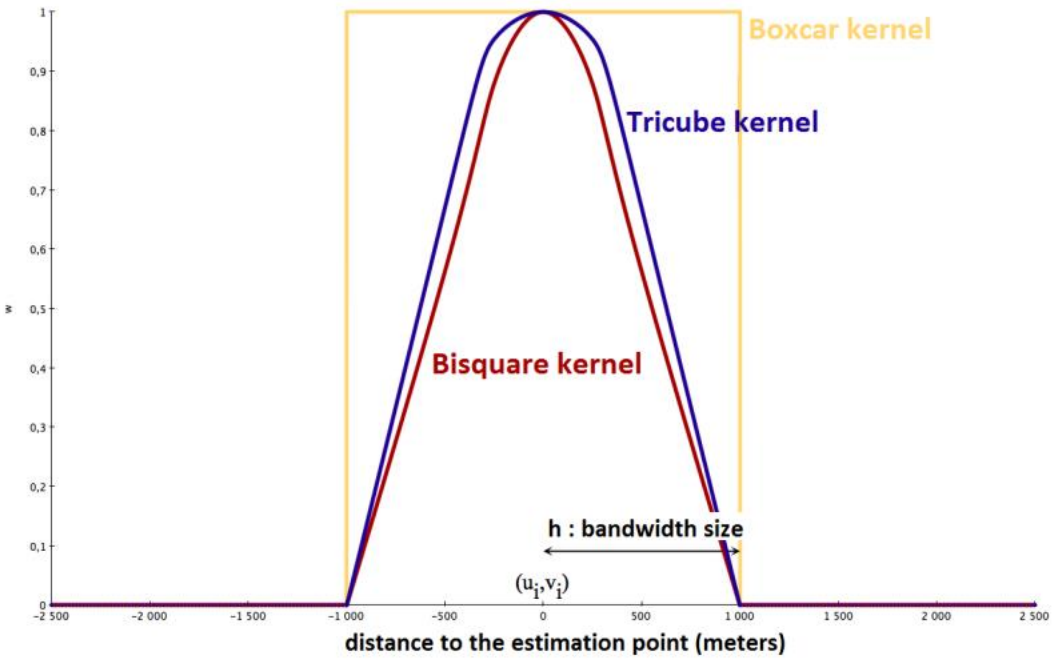
Continuous: considers all observations

Compact support: some observations are weighted to zero after a distance (Bi-square is often prefered)

Uniform kernel	$w(d_{ij}) = 1$
Gaussian kernel	$w(d_{ij}) = exp(-\frac{1}{2}(\frac{d_{ij}}{h})^2)$
Exponential kernel	$w(d_{ij}) = exp(-\frac{1}{2}(\frac{\lvert d_{ij}\rvert}{h}))$

Box-Car Kernel	$w(d_{ij}) = 1$ if $\lvert d_{ij}\rvert < h$, 0 otherwise
Bi-Square	$w(d_{ij}) = (1-(\frac{d_{ij}}{h})^2)^2$ if $\lvert d_{ij}\rvert < h$, 0 otherwise
Tri-Cube Kernel	$w(d_{ij}) = (1-(\frac{\lvert d_{ij}\rvert}{h})^3)^3$ if $\lvert d_{ij}\rvert < h$, 0 otherwise





Fixed Kernel or Adaptive Kernel?

Fixed Kernel:
- The extent is determined by the distance to the point
- Suitable for uniform spatial distributions, problematic in non-uniform spatial distributions
- Number of observations included in the regression could vary considerably. Few observations in low density areas

Adaptive Kernel:
- the extent is determined by the number of neighbors

Bandwidth of the Kernel

- Finding the bandwidth is a calibration strategy
- What would happen if is too small? Too large?

Two criteria based on prediction:

- Cross-validation criteria
- Adjusted Akaike Criterion

- Done automatically by statistical software

Bandwidth of the Kernel

Cross-validation criteria

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

h: Distance, bandwidth

$\hat{y}_{\neq i}(h)$: Prediction based on a model estimated without "i".

Run several times for many bandwidths "h" and select bandwidth that minimizes CV (that would be the distance with higher predictive power)

Adjusted Akaike Criterion

$$AIC_c(h) = 2nln(\bar{\sigma}) + nln(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\}$$

$tr(S)$: Trace of the hat matrix

Trade-off between predictive power and complexity

Results in larger bandwidths than the CV

Thank you

Orlando Sabogal-Cardona
PhD researcher
University College London UCL