University of Glasgow

Multiscale Geographically Weighted Regression (MGWR)

Ziqi Li
(Ziqi.Li@glasgow.ac.uk)
School of Geographical and Earth Sciences
University of Glasgow

FOSS4G:UK LOCAL 2022

Nov 17, 2022

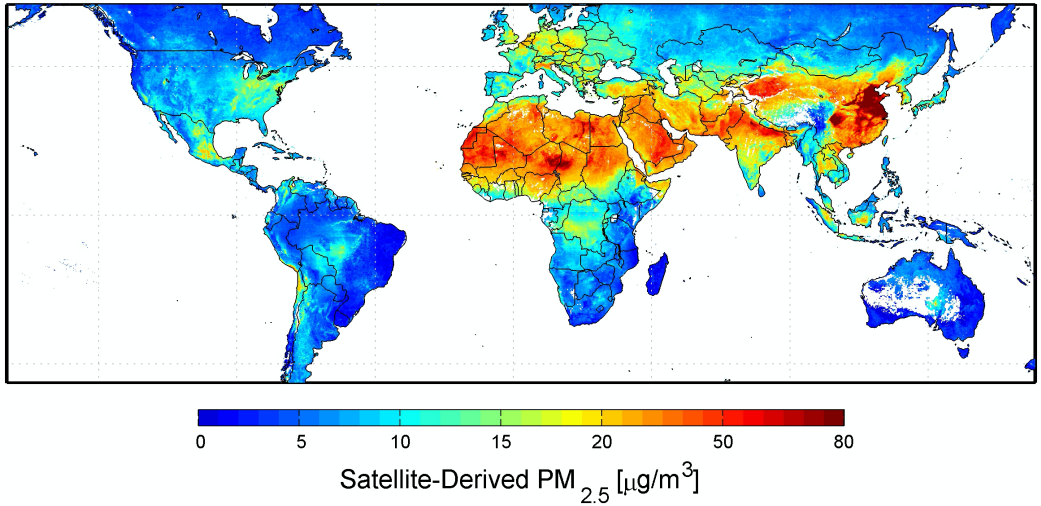
Structure for today

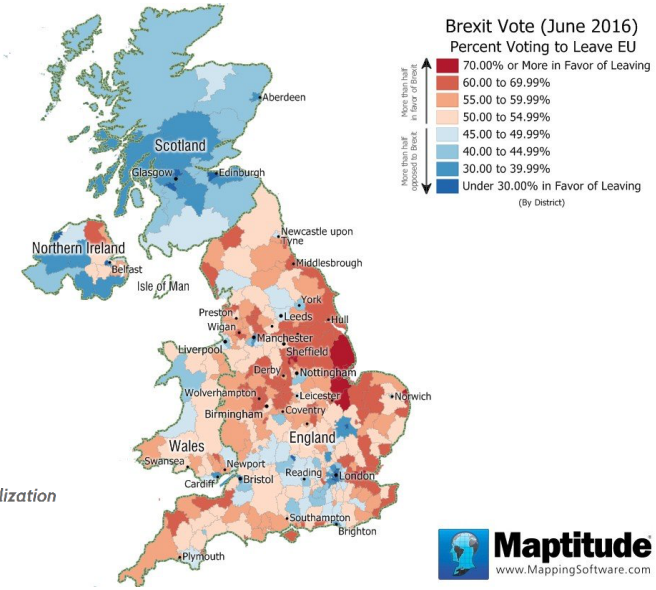
- The fundamentals
  - Why local model and MGWR
  - Inference
  - Software
- Hands-on examples in python
  - SIMD Glasgow
  - Airbnb data

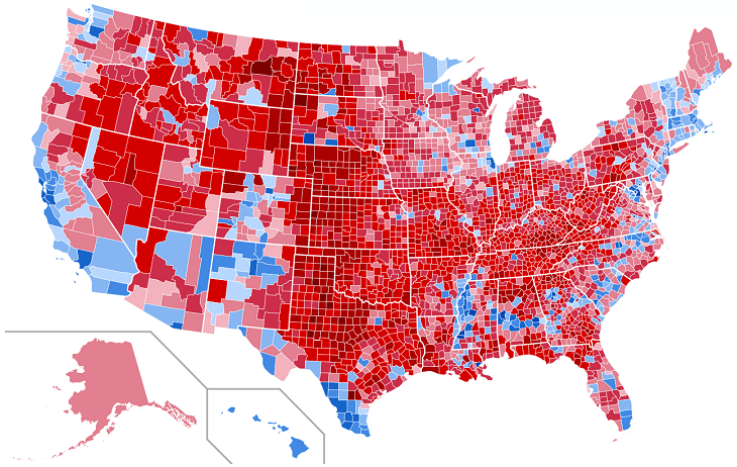
Structure for today

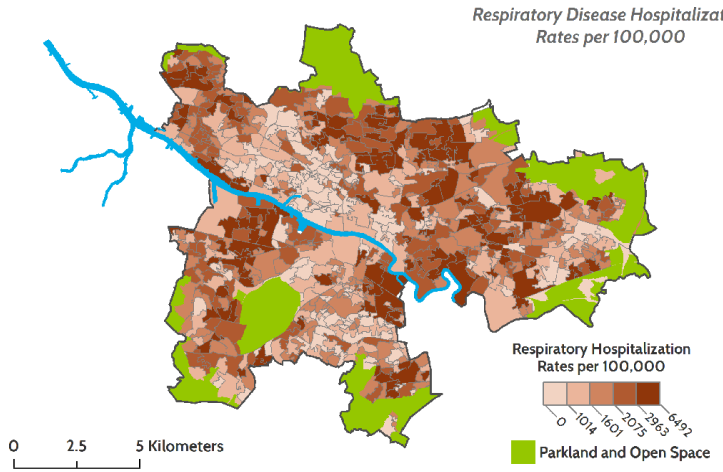
- The fundamentals
  - **Why local model and MGWR**
  - Inference
  - Software
- Hands-on examples in python
  - SIMD Glasgow
  - Airbnb data

Data are things that we can directly observe and measure

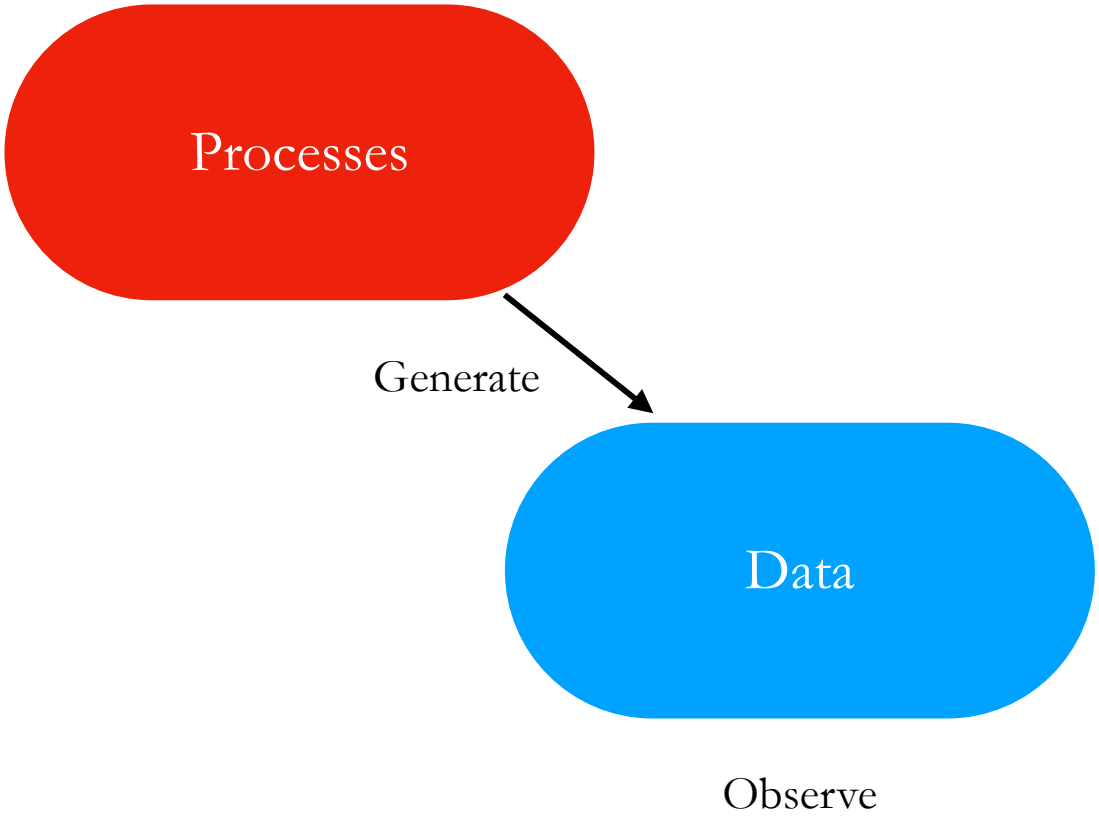




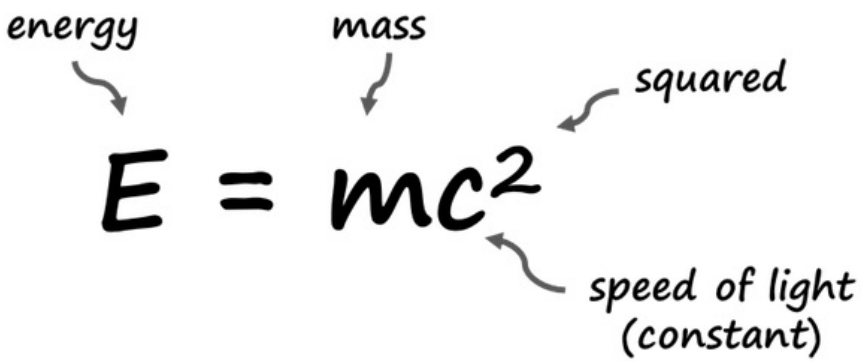




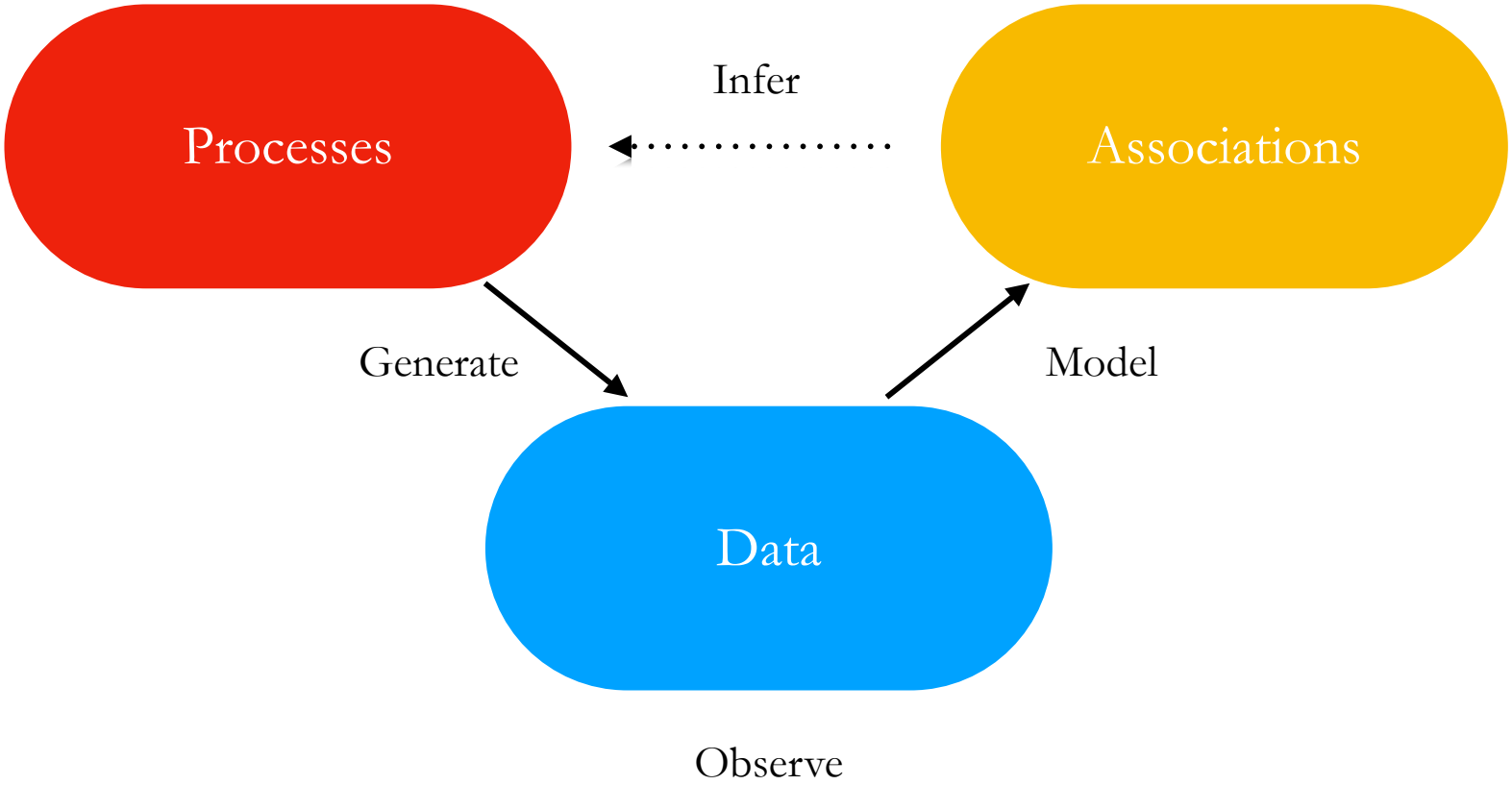
Data are generated by some processes



Processes are often complex and hidden



Data, processes and associations



Why do we want to understand data generating processes

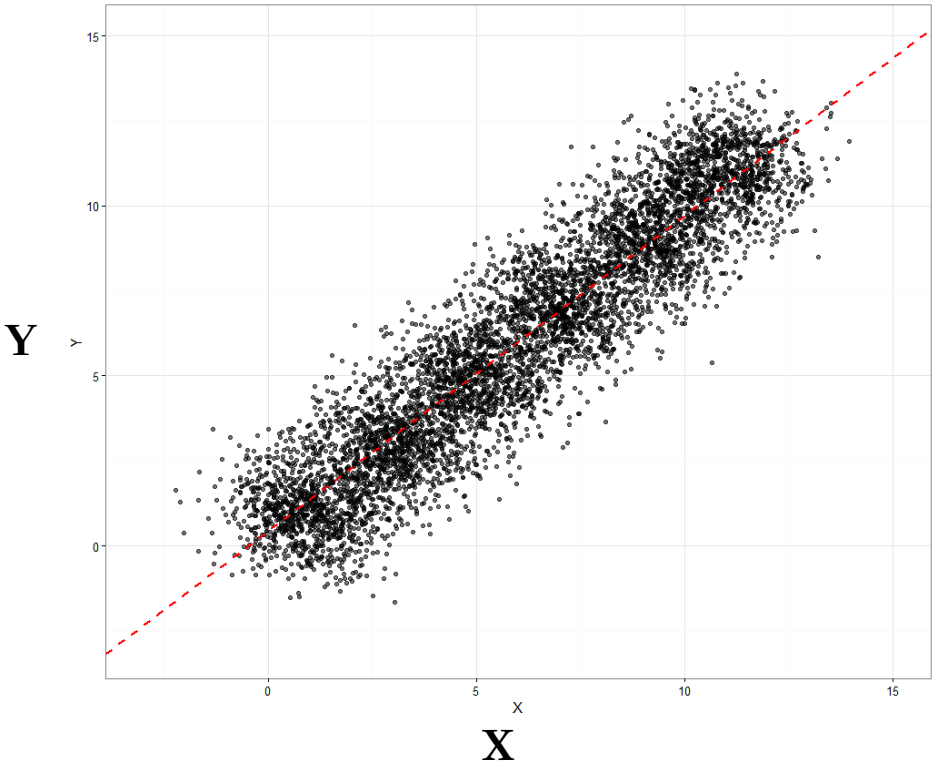
- To understand **why** things are the way they are?
    - **where** they are
- To predict/change the future
- How?
    - Through data and models

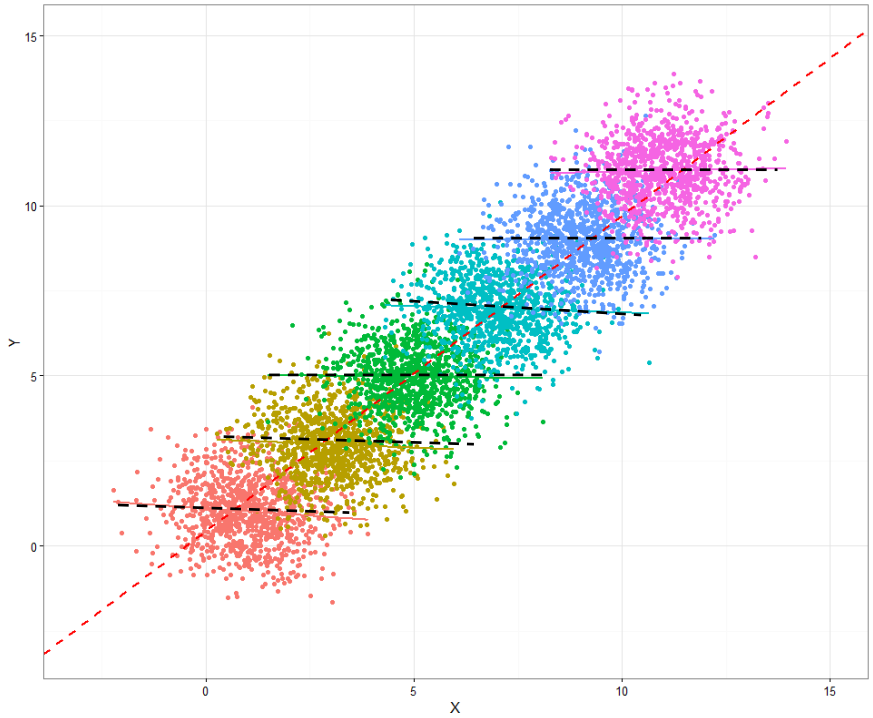
Global vs. local

- Global process: same process that is applicable everywhere
  - $E = mc^2$, $2H_2 + O_2 = 2H_2O$, etc.
- Local process: varies locally, location/context matters:
  - Local policy, local culture, etc.

Global vs. local

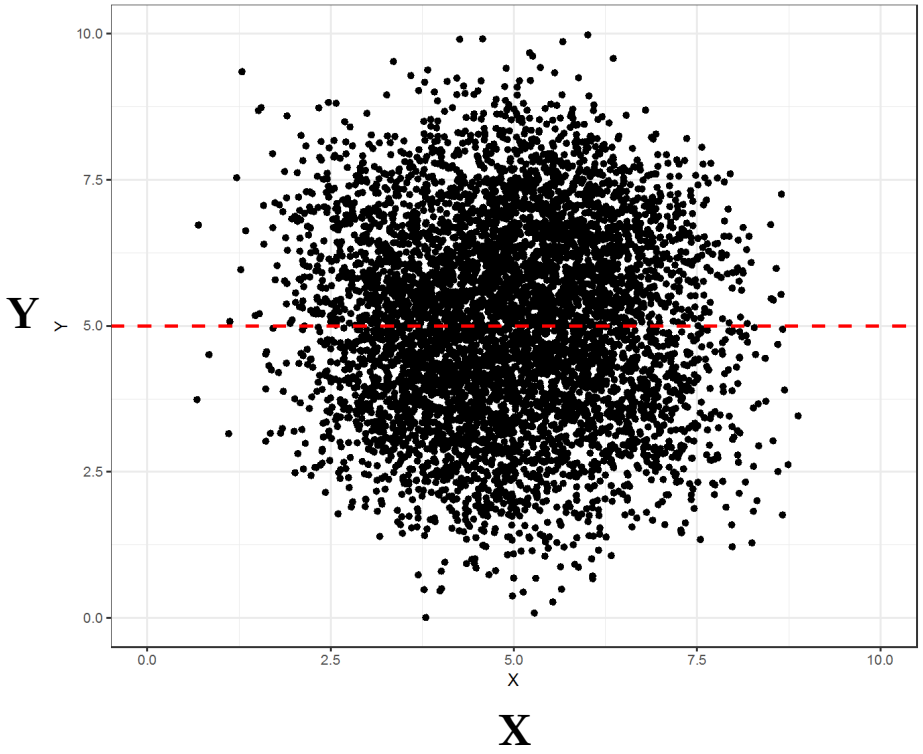
There is a trend when groups are pooled together, but no trend is found within each group.

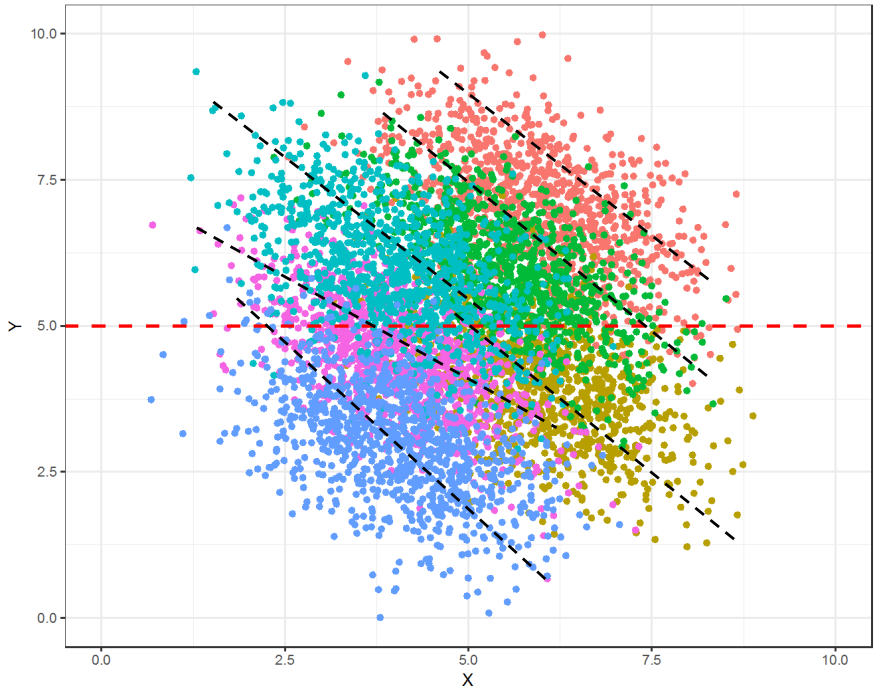




Global vs. local

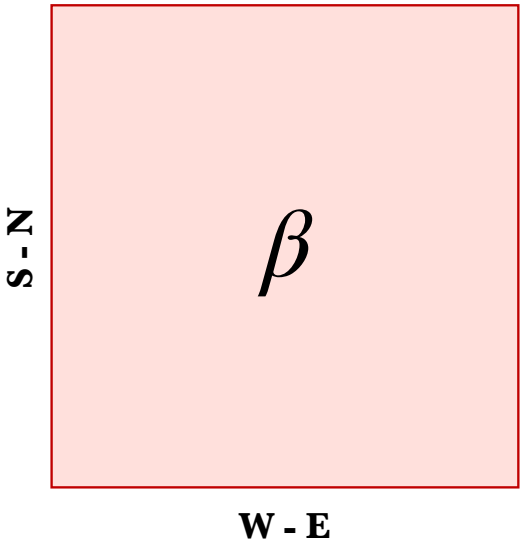
A trend appears in several different groups of data but disappears or reverses when these groups are combined.



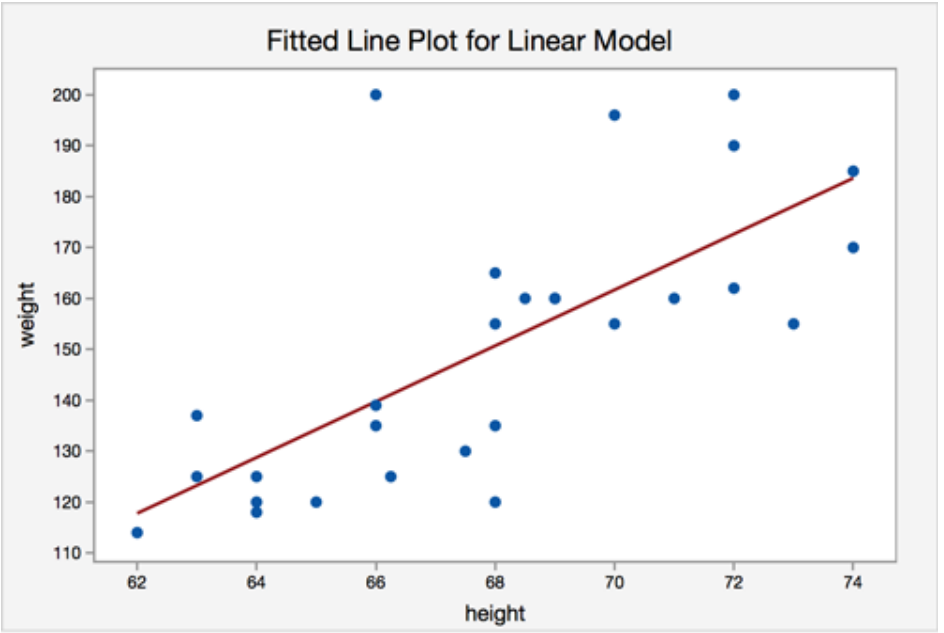


Modelling global relationship

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$



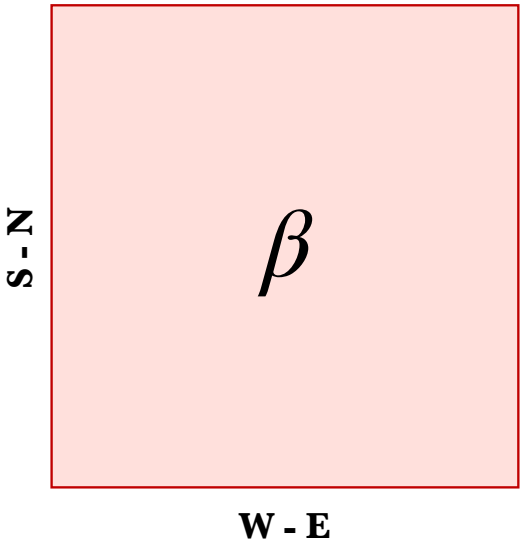
$$\hat{\beta} = (X^T X)^{-1} X^T Y$$



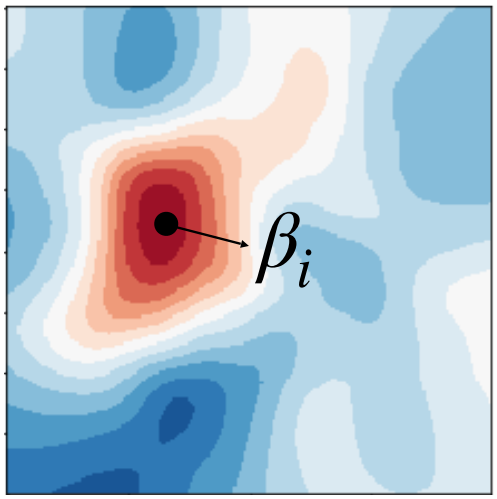
Only one linear regression equation is estimated globally, assuming the height-weight relationship is the same everywhere.

Modelling local relationship

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$



$$y_i = \beta_{0i} + \beta_{1i} x_i + \epsilon_i$$

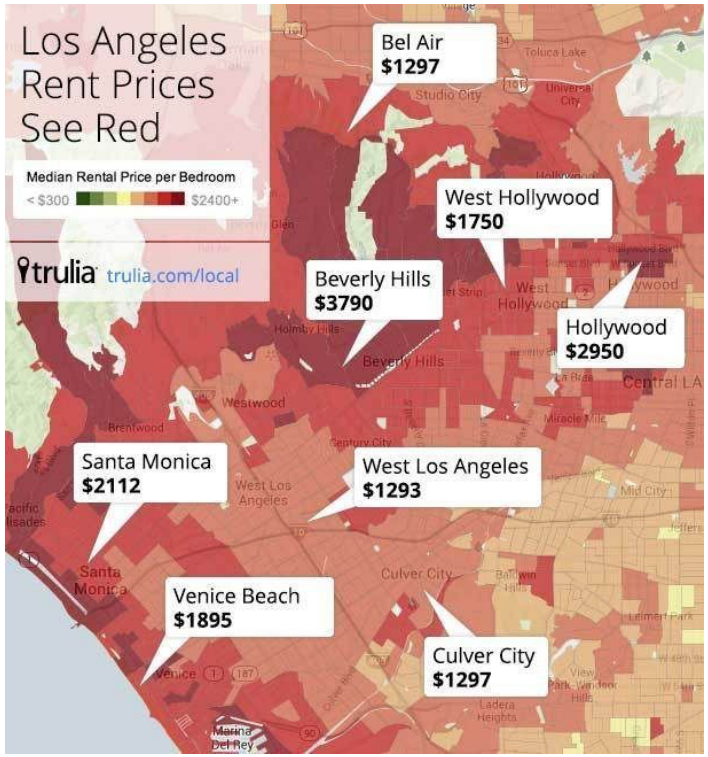


Spatial Heterogeneity

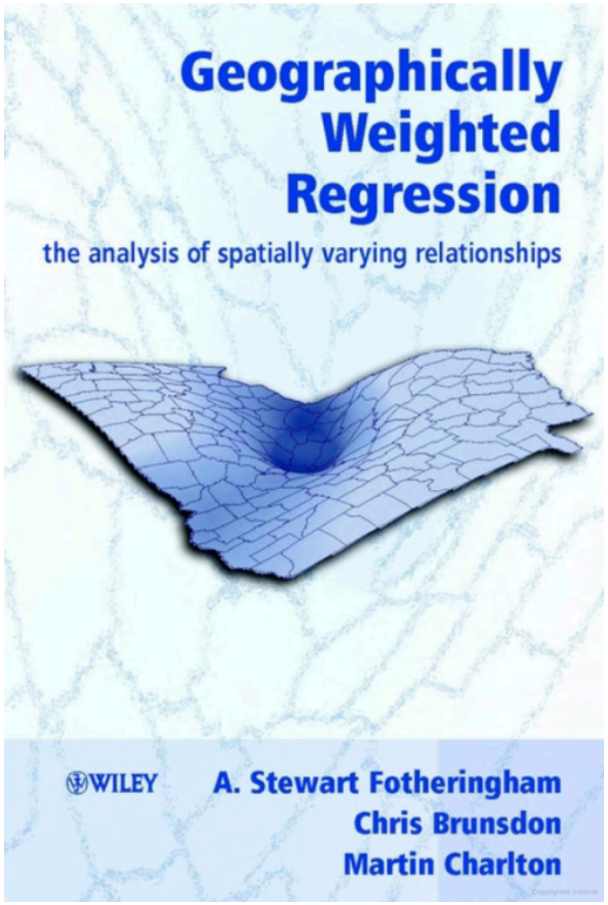
What if relationship varies from location to location?

Why might relationships vary spatially?

- Relationships intrinsically different across space:
  - For example, differences in attitudes, preferences or different administrative, political or other contextual effects produce different responses to the same stimuli.
  - Location matters. For example, the rent for a bedroom in West End is much more expensive that that in East End.



Geographically Weighted Regression (GWR)



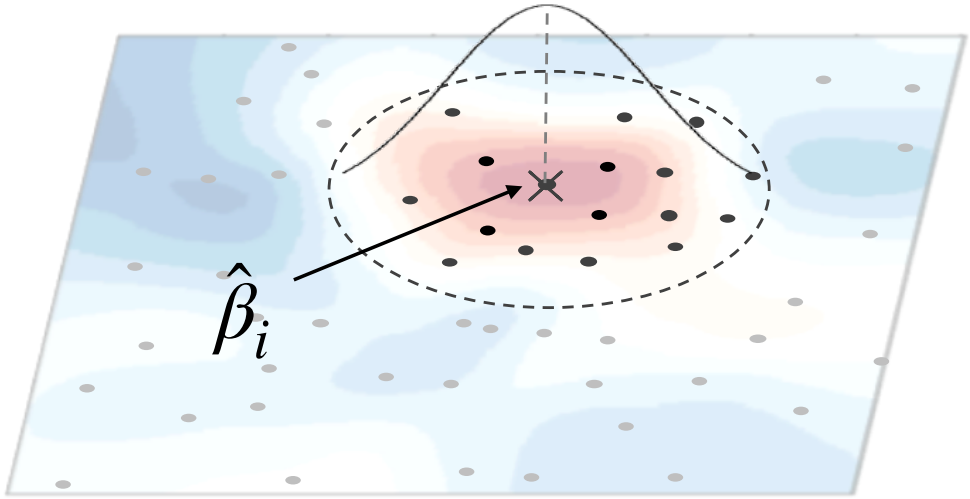
2002

GWR is one of the local models that can estimate spatially varying relationships.

The essence of GWR is to fit a local regression at each location by only using the data borrowed from nearby.

Weighted by the proximity of the location from which the data are being borrowed to the location for which the local regression is fitted.

Geographically Weighted Regression (GWR)



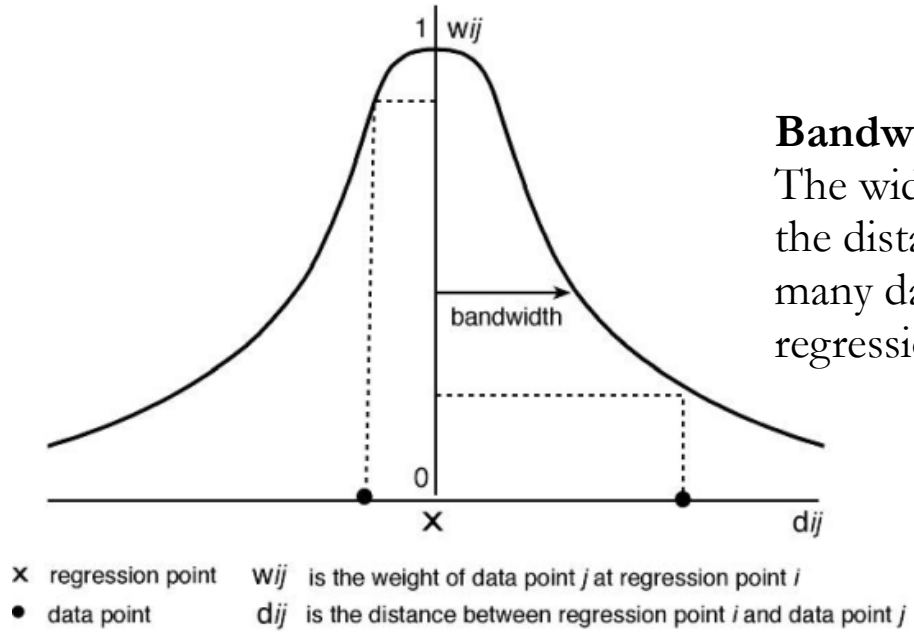
$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T Y$$

GWR is one of the local models that can estimate spatially varying relationships.

The essence of GWR is to fit a local regression at each location by only using the data borrowed from nearby.

Weighted by the proximity of the location from which the data are being borrowed to the location for which the local regression is fitted.

A typical spatial weighting function



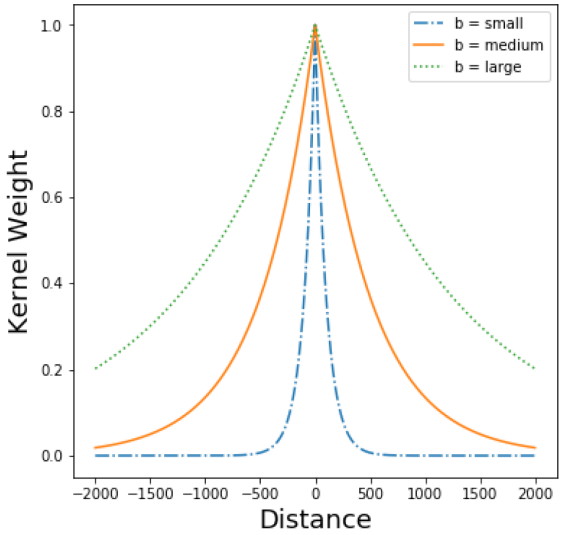
Bandwidth:
The width of the kernel which controls the distance-decay, also decides how many data points to include in each local regression.

Closer locations receive larger weights.

Far locations receive smaller weights

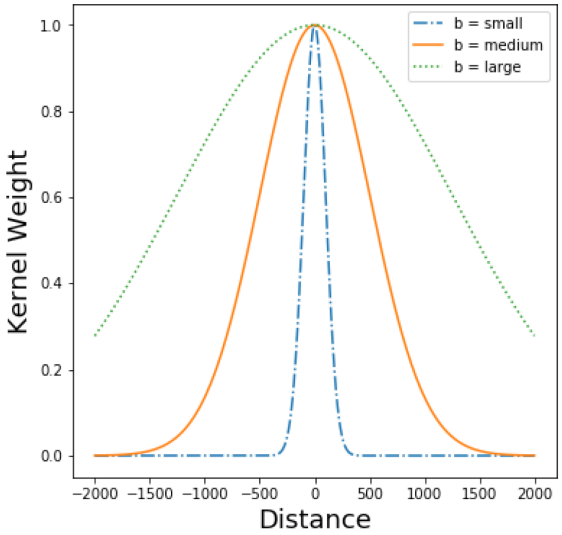
GWR weighting function choices

Exponential



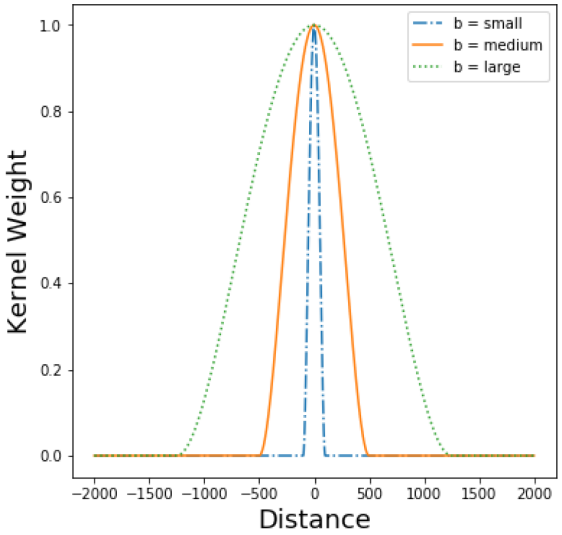
$$w_{ij} = \exp\left(-\frac{d_{ij}}{\gamma}\right)$$

Gaussian



$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{\gamma}\right)^2\right)$$

Bi-square

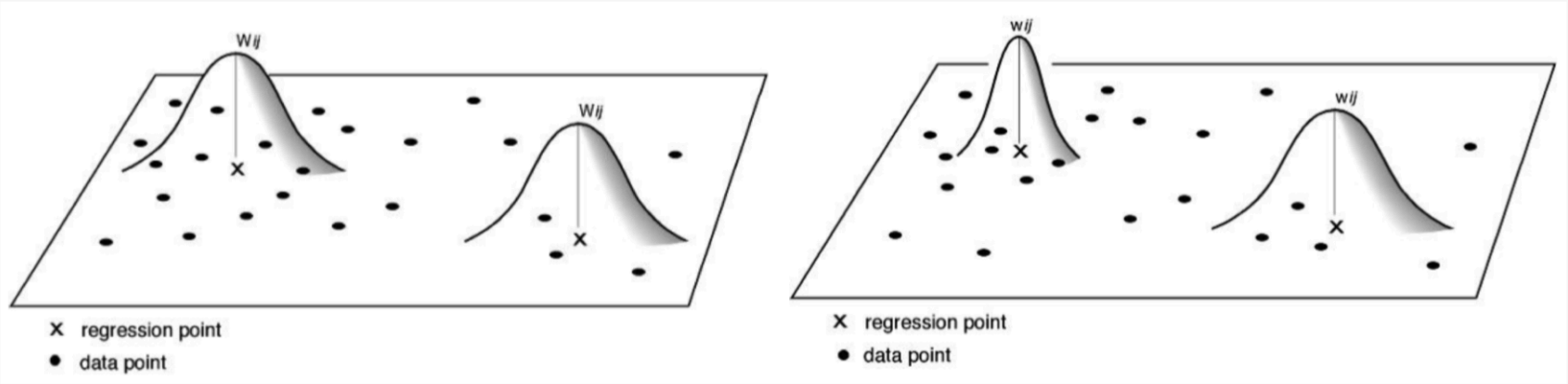


$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij}^2/\gamma^2\right)\right]^2 & \text{if } d_{ij} \leq \gamma \\ 0 & \text{if } d_{ij} > \gamma \end{cases}$$

GWR kernel

Fixed
The bandwidth is defined as:
a fixed distance.

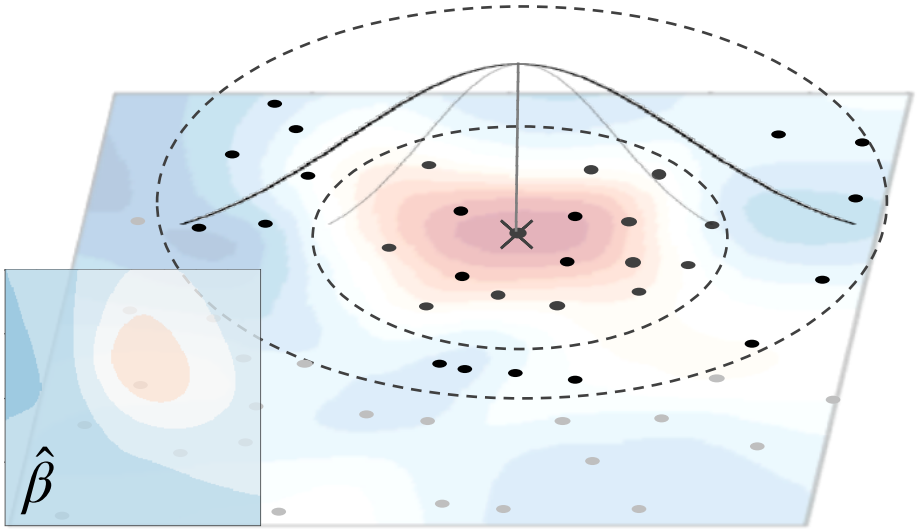
Adaptive
The bandwidth is defined as:
the number of nearest neighbours.



Best kernel combinations:

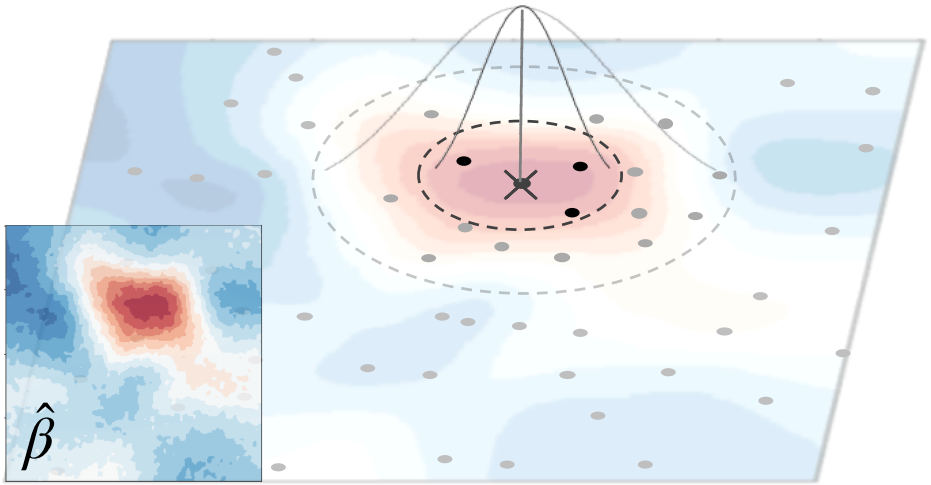
1) **Adaptive + Bisquare**
2) Fixed + Gaussian

Bandwidth is a key parameter in GWR



Large bandwidth:

- Low variance (more data in the local estimation)
- High bias (processes are more dissimilar)
- Potentially under-fitting



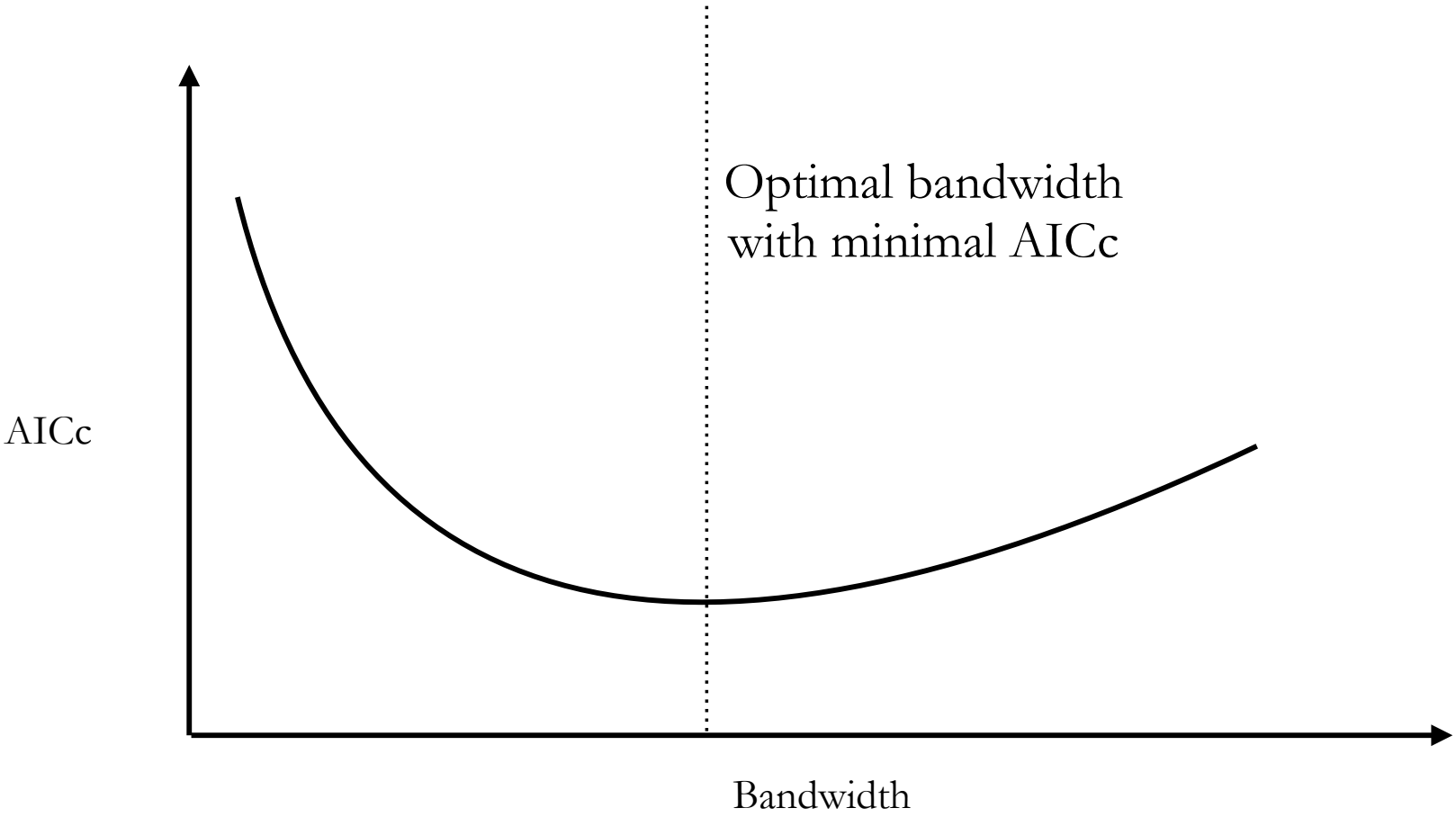
Small bandwidth:

- High variance (fewer data in the local estimation)
- Low bias (processes are more similar)
- Potentially over-fitting

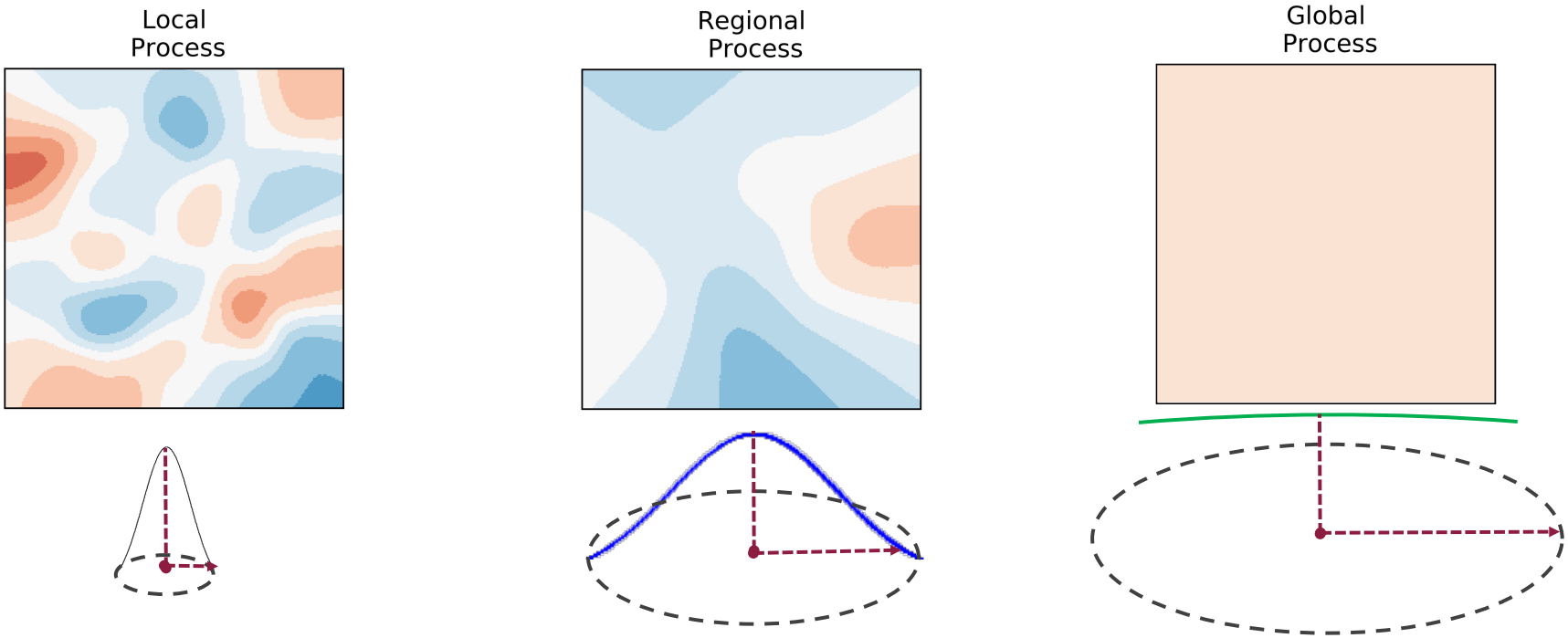
Bandwidth selection criteria

- Cross validation: Leave one out, k-fold
- Likelihood-based:
  - Akaike Information Criterion (AIC): A penalised score that balances model accuracy and complexity.
  - Corrected Akaike Information Criterion (AICc): Asymptotically equals to AIC, corrected for small samples.
  - Bayesian Information Criterion: Similar to AIC but having a larger penalty on model complexity.
- In general, **AICc** is preferred (default in the software).

Bandwidth vs. AICc



When processes operate at different spatial scales



Small bandwidth

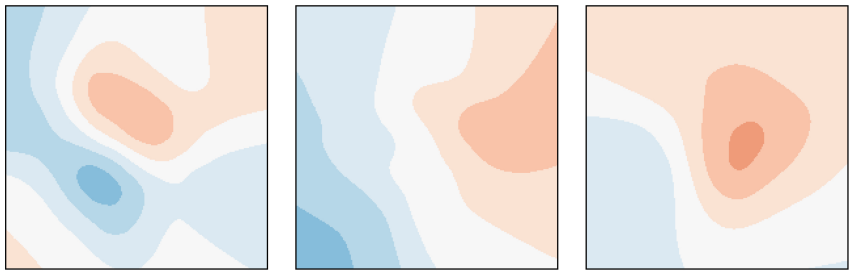
Moderate bandwidth

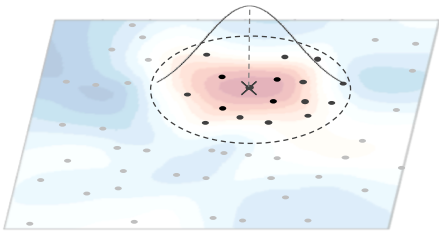
Large bandwidth

When we model them simultaneously

GWR
Single optimal bandwidth

$$y_i = \beta_{0(bw)i} + \beta_{1(bw)i}\, x_{1i} + \beta_{2(bw)i}\, x_{2i} + \beta_{3(bw)i}\, x_{3i} + \epsilon_i$$



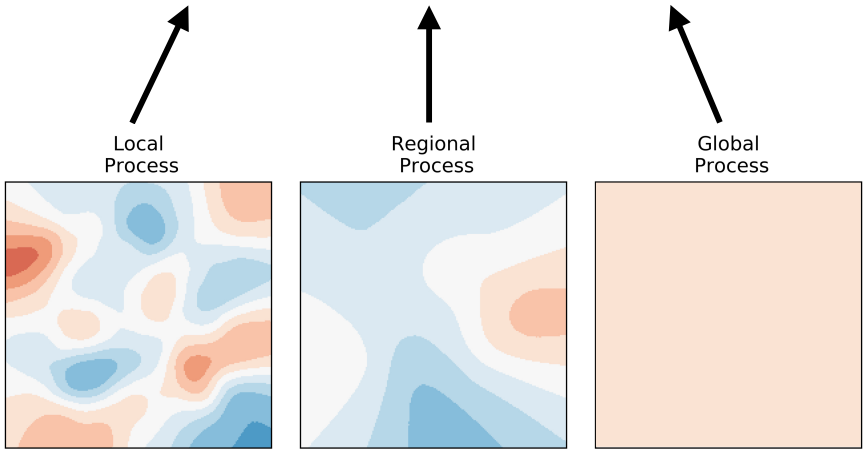


The use of single bandwidth assumes processes operate at the **same** spatial scale.

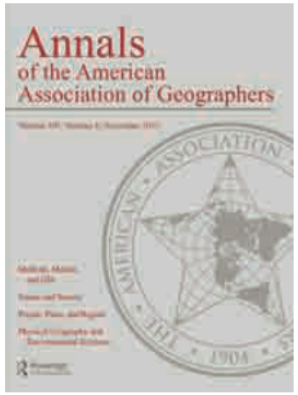
MGWR with covariate-specific bandwidths

MGWR
Covariate-specific bandwidths

$$y_i = \beta_{0(\boldsymbol{bw_0})i} + \beta_{1(\boldsymbol{bw_1})i}\, x_{1i} + \beta_{2(\boldsymbol{bw_2})i}\, x_{2i} + \beta_{3(\boldsymbol{bw_3})i}\, x_{3i} + \epsilon_i$$



By using covariate-specific bandwidths, MGWR allows processes vary at **<u>different</u>** spatial scales.



Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: http://www.tandfonline.com/loi/raag21

Multiscale Geographically Weighted Regression (MGWR)

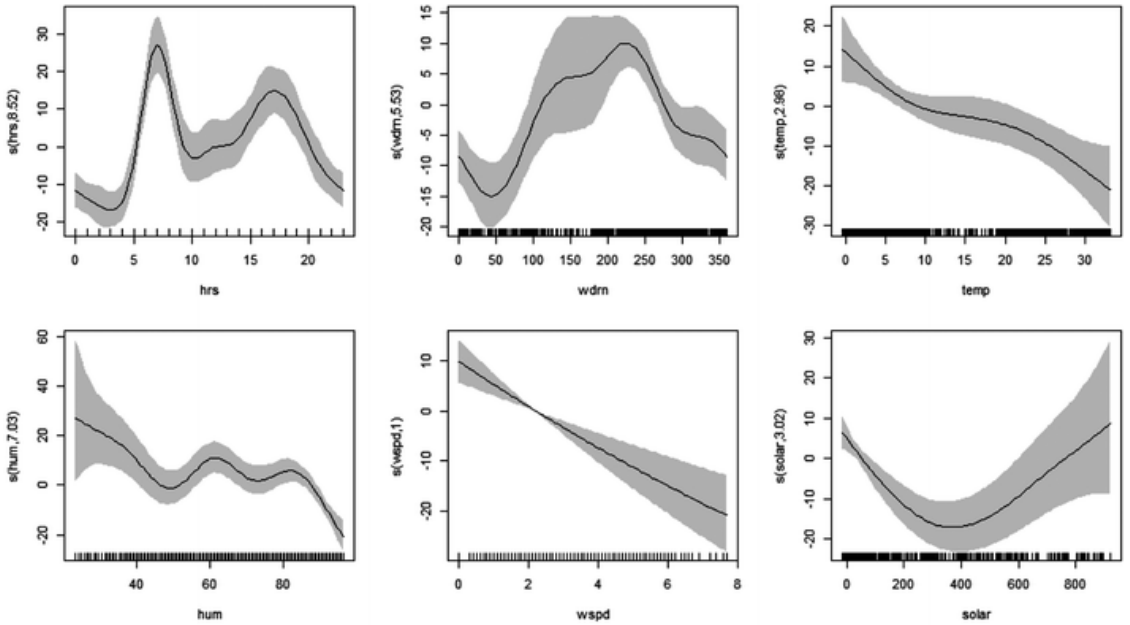
A. Stewart Fotheringham, Wenbai Yang & Wei Kang

To cite this article: A. Stewart Fotheringham, Wenbai Yang & Wei Kang (2017) Multiscale Geographically Weighted Regression (MGWR), Annals of the American Association of Geographers, 107:6, 1247-1265, DOI: 10.1080/24694452.2017.1352480

To link to this article: http://dx.doi.org/10.1080/24694452.2017.1352480

Generalised Additive Model (GAM)

$$y = f_0 + f_1(X_1) + f_2(X_2) + \ldots + f_j(X_j) + e$$



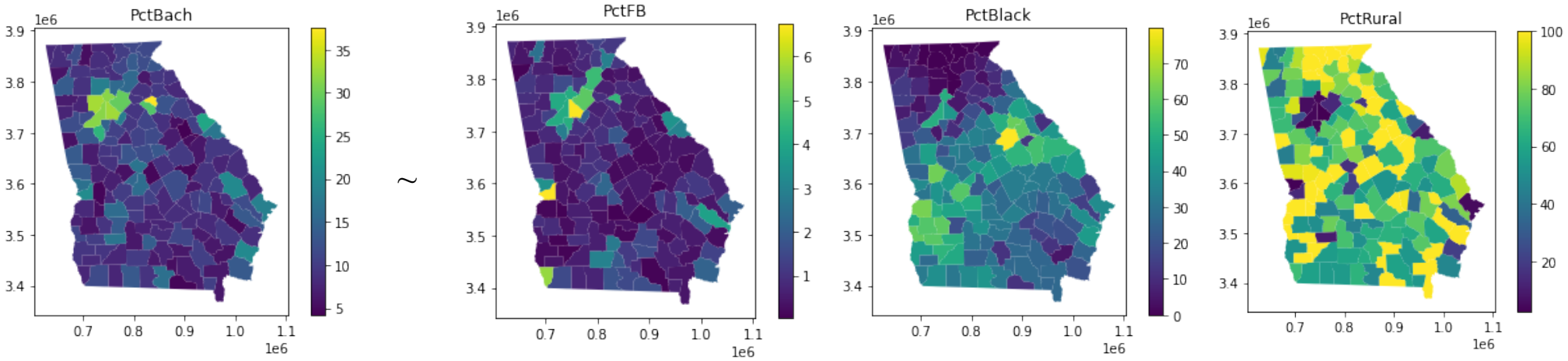
MGWR is estimated using GAM (Hastie and Tibshirani, 1990) using iterative backfitting

What questions can be answered by (M)GWR?

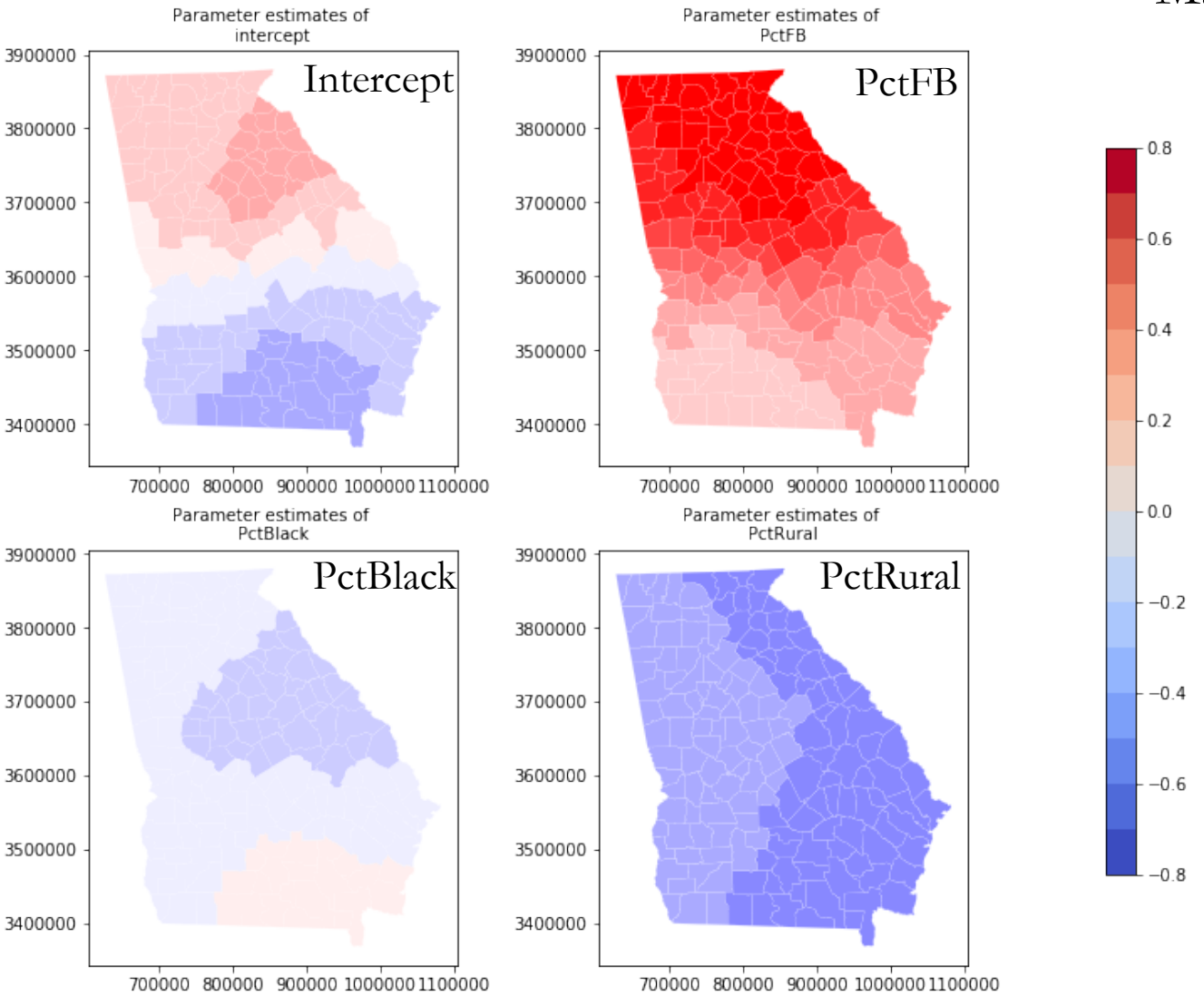
- Is the relationship between educational attainment and income consistent across the study area?
- Do certain illness or disease occurrences increase with proximity to water features?
- What are the key variables that explain high forest fire frequency?
- Which habitats should be protected to encourage the reintroduction of an endangered species?
- Where are the districts in which children are achieving high test scores? What characteristics seem to be associated? Where is each characteristic most important?
- Are the factors influencing higher cancer rates consistent across the study area?

A simple model

- Education attainment ~ Intercept + %Foreign Born + % Black + % Rural



Map of standardised coefficients



Plan for today

- The fundamentals
  - Why local model and MGWR
  - **Inference**
  - Software
- Hands-on examples in python
  - SIMD Glasgow
  - Airbnb data

There are three inferential issues in MGWR

- The inference about the ***individual*** local estimates
- The inference about the ***set*** of local estimates
- The inference about the ***bandwidth***

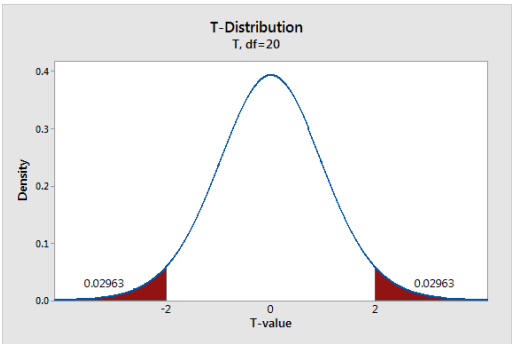
t-test in OLS

Model: Science score ~ math score + gender + social science score + reading score

science	Coef.	Std. Err.	t	P>|t|
math	.3893102	.0741243	5.25	0.000
female	-2.009765	1.022717	-1.97	0.051
socst	.0498443	.062232	0.80	0.424
read	.3352998	.0727788	4.61	0.000
_cons	12.32529	3.193557	3.86	0.000

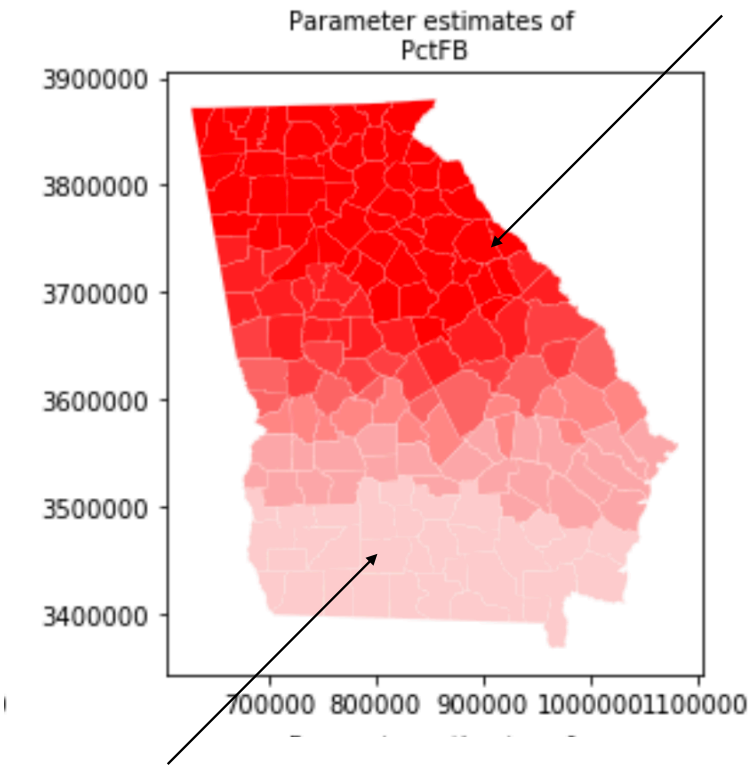
If $p<0.05$: significant
If $p>0.05$: insignificant
or

If $|t| > 1.96$: significant



t-test in (M)GWR

est: 0.5; se:0.2: t-values: 2.5
Significant*



est: 0.1; se:0.1: t-values: 1
Non significant

- In (M)GWR, each location generates 1 parameter estimate, 1 standard errors, 1 t-value and 1 p-value.
- We need to conduct t-test at each location.
- This will introduce multiple testing issues

Multiple testing issues in (M)GWR

- Multiple testing issue occurs when we consider a ***set*** of statistical inferences ***simultaneously***.
- In this case of (M)GWR, we are conducting multiple local t-tests altogether.
- The more inferences are made, the more likely erroneous inferences are to occur.

Multiple testing issues example

- Consider if our at home COVID-19 tests have a false-positive rate of 0.05.
- This is saying that there is a 5% chance one may test positive even if he doesn't have the virus. p(error) = 0.05
- What if we test twice and the probability of having at least 1 positive: $1 - 0.95^2 = 0.095$.
- What if we test 10 times and the probability of having at least 1 positive: $1 - 0.95^2 = 0.40$.
- What if we test 100 times and the probability of having at least 1 positive: $1 - 0.95^{100} = 0.99$
- **The probability of observing one false positive is inflated by the number of tests.**

Corrections to multiple testing issues

- Bonferroni correction
- Sidak
- False Discovery Rate
- …

Bonferroni correction

- If the significance level is set to α, the corrected significance level to reject the null is α/m, where m is the number of tests performed.
- Example: suppose m(number of tests) = 20 $\alpha = 0.05$ so that $\alpha/m = 0.0025$
- p(of at least one significant result) = 1- p(no significant results) = $1-(1-.0025)^{20} = 0.0488$
- suppose m = 100 $\alpha = 0.05$ so that $\alpha/m = 0.0005$
- p(of at least one significant result) = 1- p(no significant results) = $1-(1-0.0005)^{20} = 0.0488$
- So the p(of at least one significant result) being controlled to have an error rate ~0.05.

Bonferroni correction in GWR

- The correction method assumes the tests are independent. This is not the case in GWR since the tests are spatially dependent.
- An adjustment to this is to approximate the number of independent tests in GWR using the effective number of parameters (ENP).
- The ENP = tr(**S**)
- where tr(.) is a trace operator taking the sum of the diagonal elements in **S**
- **S** is the hat matrix with a dimension of n by n where **y_hat** = **Sy**.
- We replace the ENP/k (k = #of covariates) as the number of tests in GWR for adjusting the significance level α.

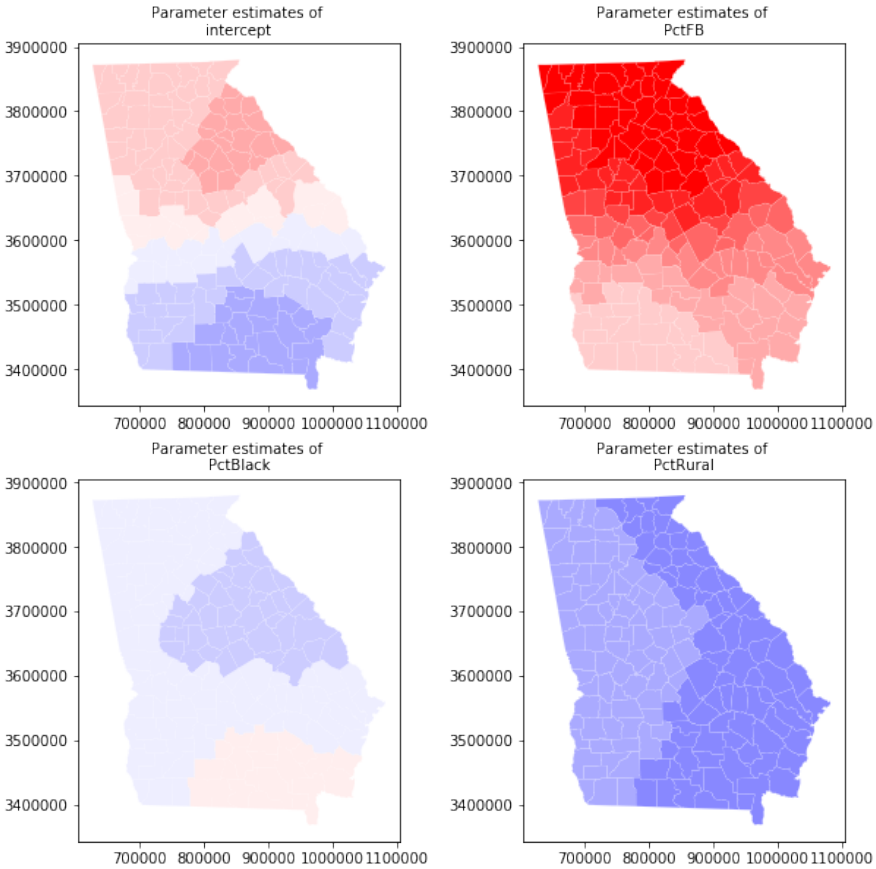
Diagnostic information

Residual sum of squares:	51.186
Effective number of parameters (trace(S)):	11.805
Degree of freedom (n - trace(S)):	147.195
Sigma estimate:	0.590
Log-likelihood:	-135.503
AIC:	296.616
AICc:	299.051
BIC:	335.913
R2:	0.678
Adjusted R2:	0.652
Adj. alpha (95%):	0.017
Adj. critical t value (95%):	2.414

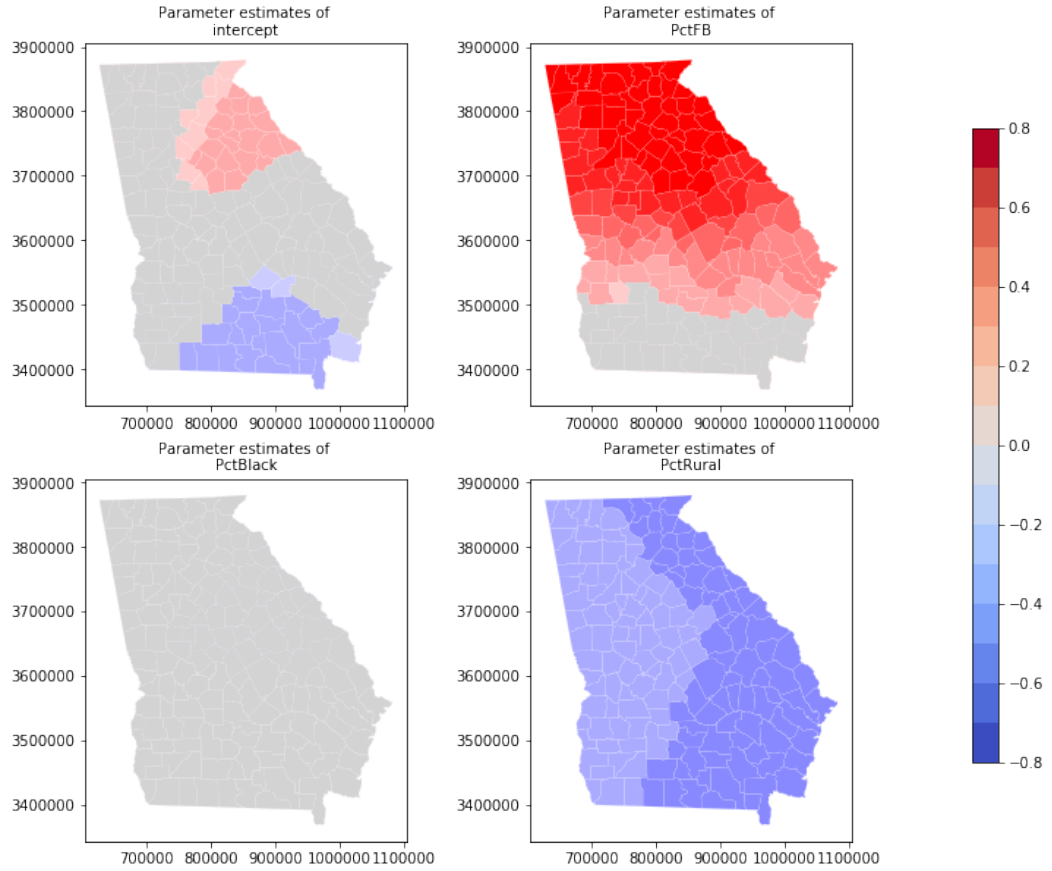
$0.017 = 0.05/(ENP/k) = 0.05/(11.8/4)$

This is output in the summary file in the mgwr software
So instead of using 0.05 as the critical p value, we should use 0.017.
If local p-value < 0.017: Significant
It also outputs the corresponding adjusted t-values:
If local |t| > 2.414: Significant

Before significance test



After significance test (with adjustment)



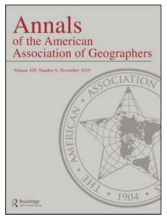
Insignificant (p>0.05) estimates are masked out in grey colour.

The inference about the ***set*** of local estimates

- Whether a parameter estimate pattern has significant spatial variability
- Provides justification to use (M)GWR
- Monte Carlo test: H_0: local estimates are the same regardless of location
- Available in the software
- Less needed in MGWR because of a global relationship will obtain a global bandwidth.

The inference about the ***bandwidth***

- Understanding covariate-specific bandwidth uncertainty is crucial to being able to make inferences about the different spatial scales over which processes operate.



Taylor & Francis
Taylor & Francis Group

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights

Ziqi Li, A. Stewart Fotheringham, Taylor M. Oshan & Levi John Wolf

To cite this article: Ziqi Li, A. Stewart Fotheringham, Taylor M. Oshan & Levi John Wolf (2020): Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights, Annals of the American Association of Geographers, DOI: 10.1080/24694452.2019.1704680

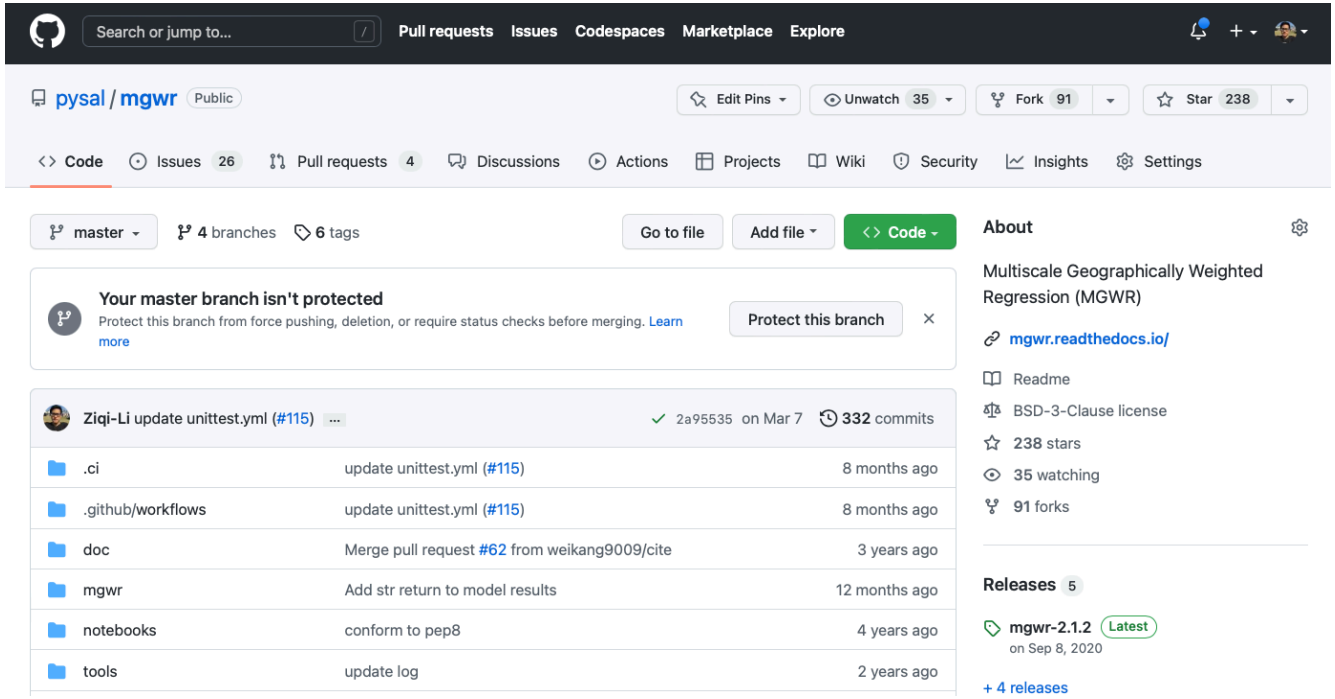
To link to this article: https://doi.org/10.1080/24694452.2019.1704680

https://www.tandfonline.com/doi/abs/10.1080/24694452.2019.1704680

Plan for today

- The fundamentals
  - Why local model and MGWR
  - Inference
  - **Software**
- Hands-on examples in python
  - SIMD Glasgow
  - Airbnb data

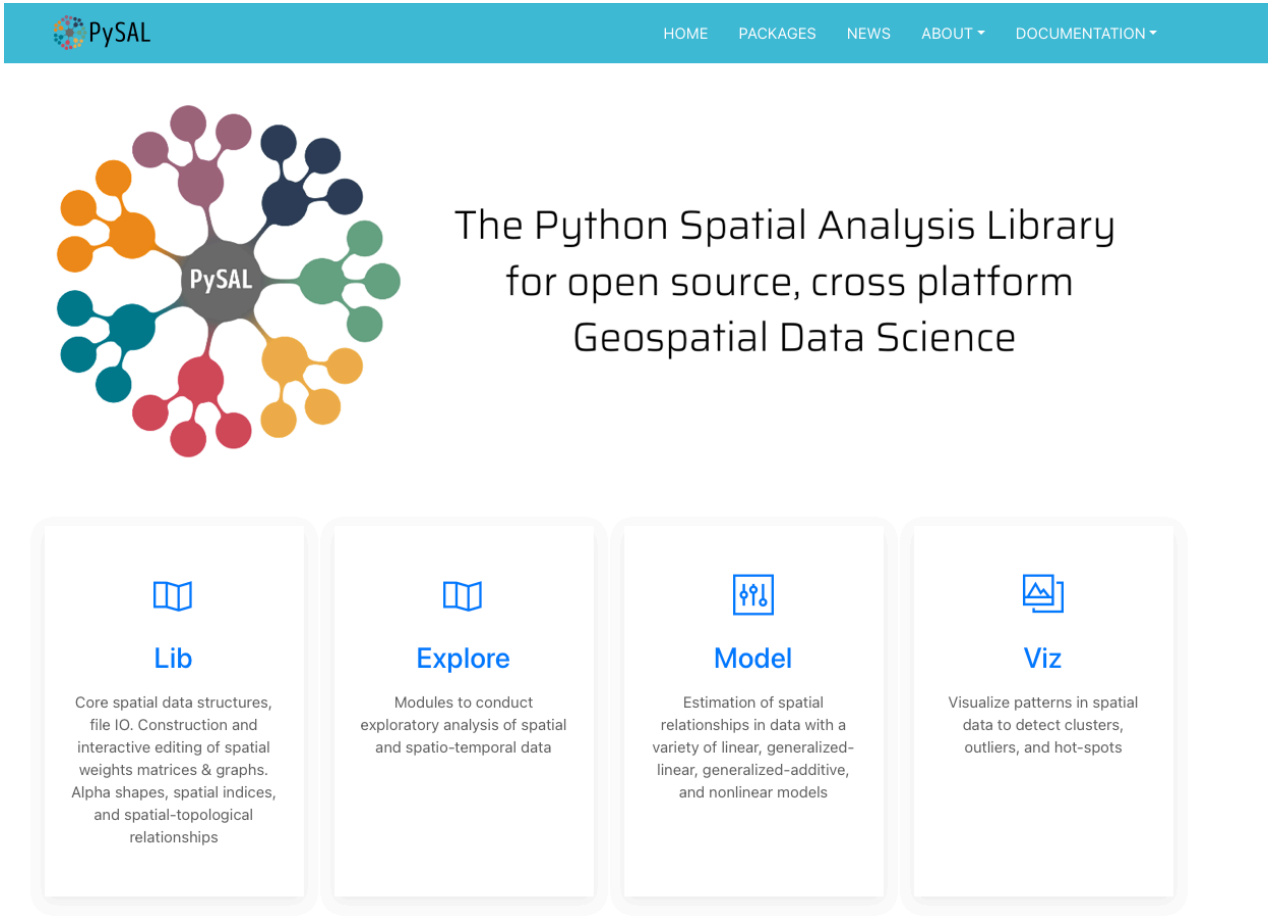
mgwr Python package



https://github.com/pysal/mgwr

PySAL
Python Spatial Analysis Library

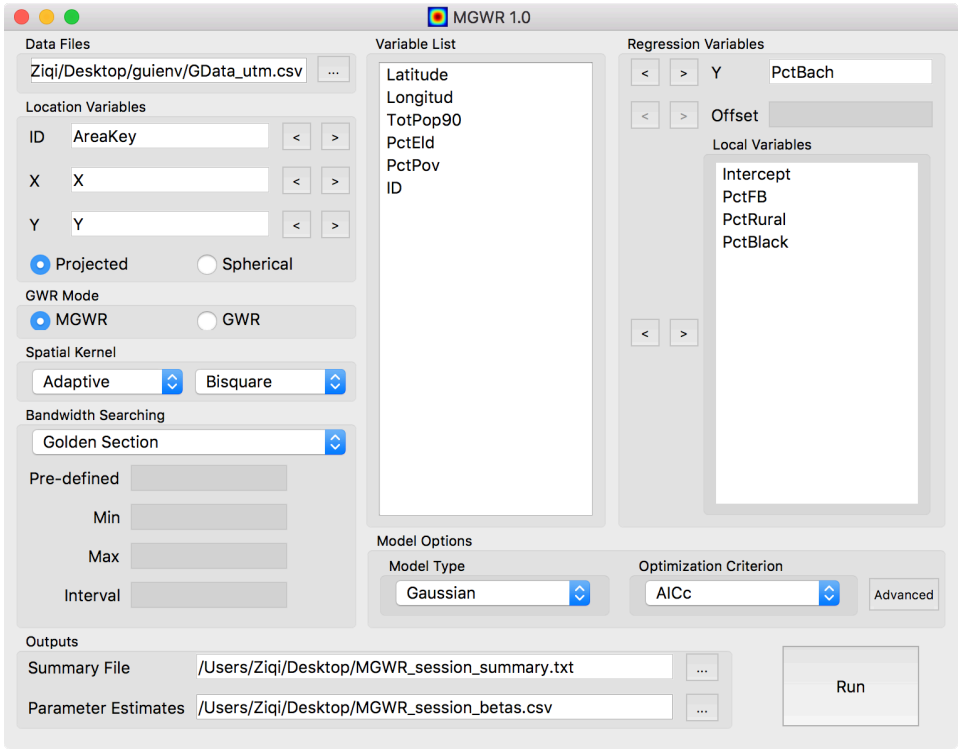
PySAL



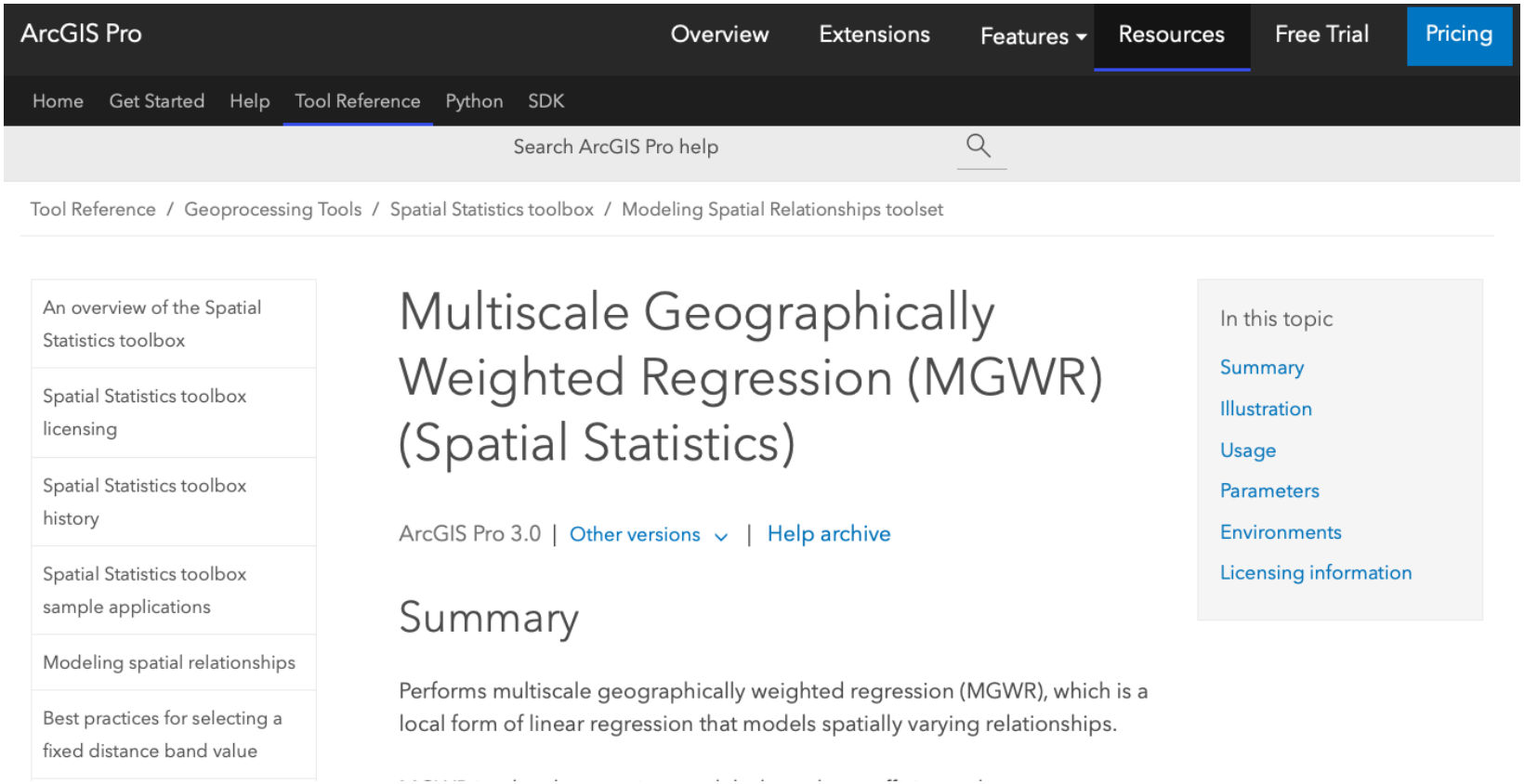
https://pysal.org

MGWR Desktop

https://sgsup.asu.edu/sparc/mgwr



ArcPro 3.0



https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/multiscale-geographically-weighted-regression.htm

Caveats

- Not all data/problems are appropriate for MGWR.
- We need a good baseline (e.g. OLS) model first.
- MGWR needs > 200 data points.
- Above 10k data points will be computationally challenging.
- Inference is important.

Key references

- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247-1265.

- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269.

- Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science*, *34*(7), 1378-1397.

- Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2020). Inference in multiscale geographically weighted regression. *Geographical Analysis*, *52*(1), 87-106.

Plan for today

- The fundamentals
  - Why local model and MGWR
  - Inference
  - Software
- **Hands-on examples in python**
  - **SIMD Glasgow**
  - **Airbnb data**

https://github.com/Ziqi-Li/foss4g_22_glasgow