Spatial Analysis and Modeling

Modeling Spatial Heterogeneity

Corso di formazione su ML e DL
Fondazione LINKS
25/09/2025

ISI Foundation

What is spatial heterogeneity?

- **Relationships between predictors** X **and outcome** Y **can vary across space**: **local contexts** modulate effect sizes.
- Formal view: model parameters depend on location, e.g. $y(\mathbf{s}) \approx \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}(\mathbf{s})$, where $\mathbf{s}$ denotes spatial location and $\boldsymbol{\beta}(\mathbf{s})$ are spatially varying coefficients.
- Examples:
  - **Housing markets**: the marginal effect of floor area, $\beta_{\text{area}}(\mathbf{s})$, differs by neighborhood.
  - **Environmental response**: pollutant–health coefficient $\beta_{\text{pollutant}}(\mathbf{s})$ may be larger near industrial corridors.
  - **Service access**: the effect of distance to a hospital, $\beta_{\text{dist}}(\mathbf{s})$, depends on local transport infrastructure.
- Implication: **assuming a global constant** $\boldsymbol{\beta}$ **can bias estimates and obscure important local patterns (keeping residual spatially correlated)**.

Types of Spatial Nonstationarity

1. First-Order Nonstationarity
   - **Mean varies spatially**: $\mathbb{E}[Y(s)] = \mu(s)$
   - **Example**: House prices vary by neighborhood characteristics
2. Second-Order Nonstationarity
   - **Covariance structure varies**: $\mathrm{Cov}[Y(s_i), Y(s_j)] = C(s_i, s_j)$
   - **Example**: Spatial dependence stronger in urban vs. rural areas
3. Parameter Nonstationarity
   - **Regression coefficients vary**: $\beta_k = \beta_k(s)$
   - **Example**: Effect of income on education varies by region

Detecting spatial heterogeneity — practical diagnostics

- **Visual inspection**
  - Map raw outcome, key predictors, and model residuals (standardized). Spatial patterns in residuals are an immediate hint.
- **Global spatial autocorrelation**
  - Compute Moran's I (or Geary's C) on outcome and on residuals. Significant positive autocorrelation in residuals suggests missing spatially varying effects.
- **Local indicators**
  - Run LISA / local Moran and Getis-Ord Gi* to find hot/cold spots and localized clustering of residuals or errors.

- **Model comparisons (global vs local)**
  - Fit a global model and a local/partitioned model (GWR, spatially varying coefficients, or regionwise models). Large improvements in CV/AICc or predictive error point to heterogeneity.
- **Coefficient diagnostics**
  - Estimate location-wise coefficients (GWR, local fits, hierarchical SVC). Map estimates and their uncertainties; systematic spatial variation in coefficients is direct evidence.
- **Spatial cross-validation tests**
  - Use spatially blocked/buffered CV. If predictive performance drops under spatial CV vs random CV, spatial dependence/heterogeneity is important and may bias naive evaluation.

Quick checklist to run:

1. Map outcome & residuals.
2. Compute Moran's I on residuals.
3. Inspect variogram (range, sill, anisotropy).
4. Run LISA / Getis-Ord for local clusters.
5. Compare global vs local model CV/AICc.
6. Map local coefficients and their SEs; test significance/permute.

Use multiple diagnostics together — visual + global + local + predictive checks — to build robust evidence of spatial heterogeneity.

1. Spatial coordinate features

Include (x,y) or basis functions (splines, RBF) **as predictors**.

- **Pros**: simple, fast; good for smooth global trends.
- **Cons**: misses complex, non-isotropic/local boundaries.
- Use when change is gradual and you need a quick baseline.

2. Geographically Weighted Regression (GWR)

- **Pros**: interpretable local coefficients; diagnostic maps.
- **Cons**: computationally heavy, sensitive to bandwidth and collinearity, assumes isotropy in kernel.

3. Decomposition-based ensembles (partition & local models)

Partition space (clustering, administrative units, road network), **fit local models**, **combine**.

- **Pros**: flexible region shapes and models per region.
- **Cons**: partitioning is nontrivial; boundary discontinuities; maintenance overhead.
- Good for clearly segmented regimes (urban/rural, climate zones).

4. Multi-Task Learning (MTL) with spatial regularization

- **Tasks** = **locations/zones**; learn parameters $\Theta = [\theta_1, \ldots, \theta_n]$ with a graph-Laplacian penalty:
$$\min_{\Theta} \sum_i \mathcal{L}_i(\theta_i) \; + \; \lambda \operatorname{tr}\big(\Theta \, L \, \Theta^\top\big)$$
- **Pros**: encourages smoothness across neighbors (penalizes $\|\theta_i - \theta_j\|$ for connected pairs), borrows strength where labels are scarce.
- **Cons**: requires differentiable losses and careful task definition; heavier optimization and tuning of λ and the Laplacian L.
- Best when many related local prediction tasks with limited per-location data; common variants include shared feature layers, low-rank coupling, and semi-supervised graph regularization.

5. Hierarchical and multi-scale models

What they are

- Hierarchical spatial models (aka multi-level spatial models) embed **spatial structure as latent layers in a probabilistic model**.

Why they are useful (intuition + examples)

- **Separate sources of variation**: covariate effects ($\mathbf{X}\boldsymbol{\beta}$), structured spatial residuals (ϕ), and uncorrelated noise (ε), clarifying interpretation and reducing confounding.
- **Provide principled uncertainty quantification** (posterior and predictive intervals for $\boldsymbol{\beta}$, $\phi(\cdot)$, and predictions).
- Support **multi-scale** modeling and **data fusion** (e.g. combine coarse areal counts and point observations by sharing latent fields).

Main modeling approaches (short catalog)

- **CAR/GMRF** (areal data)
  - Discrete-area priors (ICAR / proper CAR). Efficient: sparse precision (Q).
  - Use when data are aggregated over polygons (districts, counties).
- **Gaussian Processes** (GP; point-referenced)
  - Continuous covariance models (Matérn, exponential). Estimate range/scale parameters.
  - Use when locations are points and smooth continuous dependence is plausible.
- **SPDE** → **GMRF** (mesh-based approximation)
  - Represent Matérn GP as a sparse GMRF via FEM mesh (Lindgren et al., 2011).
  - Scales to larger datasets with INLA-style workflows.
- **Hierarchical GLMMs** with spatial random effects
  - Combine link functions (Poisson, binomial) with CAR/GP latent effects.

Python tooling & workflows

- **PyMC** (pymc): flexible Bayesian models; user-defined CAR/GP priors; NUTS or MAP inference for moderate sizes.
- **Stan** / **cmdstanpy**: specify CAR/GP as multivariate normals with precision/covariance matrices; powerful but can be slow for large n.
- **GPyTorch** / **GPflow**: scalable GP toolkits (sparse/approximate GPs) for large point-referenced data and deep GP variants.
- **R-INLA** (recommended for heavy SPDE/GMRF use): fast approximate Bayesian inference for SPDE/GMRF; callable from Python when needed.

Geographic Weighted Regression

The Core Problem

- **Global models** assume spatial stationarity: $\beta_k = \text{constant}$
- **Reality**: Relationships vary across space due to:
  - **Contextual effects**: Local institutions, culture, policies
  - **Scale effects**: Processes operate at different spatial scales
  - **Heterogeneity**: Different causal mechanisms in different regions

GWR Solution

- **Local regressions** at each location s_0
- **Distance-weighted** observations: closer points get higher weights
- **Spatial adaptation**: Coefficients vary smoothly across space

GWR Mathematical Formulation

Basic Model

For location $s_0 = (u_0, v_0)$, the GWR model is:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

- This writes the response y_i at observation location $s_i = (u_i, v_i)$ as a linear function whose coefficients $\beta_k(u_i, v_i)$ vary with space.
- Intuition: instead of one global slope for each predictor, GWR fits a different slope at (or around) each location so local relationships can differ across the study area.
- ε_i is the local residual (assumed mean zero); it captures variation not explained by the spatially varying linear predictor.

Local Estimation

At each location s_0, coefficients are estimated by:

$$\hat{\beta}(s_0) = \arg\min_{\beta} \sum_{i=1}^{n} w(s_i, s_0) \left[y_i - \sum_{k=0}^{p} \beta_k x_{ik} \right]^2$$

- This is a weighted least-squares objective centered at the target location s_0.
- The weight function $w\left(s_i, s_0\right)$ assigns **larger weights to observations near** s_0 and **smaller weights to distant observations** (common choices: Gaussian kernel, bi-square, or nearest-neighbor).
- Conceptually: **we fit a regression using nearby data only**, but rather than a hard cutoff we **smoothly downweight farther points**.

Interpretation

- **What is computed at each location s_0**:
  - Assign each observation a weight based on its distance to s_0.
  - Solve weighted least squares to estimate local coefficients $\hat{\beta}(s_0)$.
  - Weights $W(s_0)$ are determined by the chosen **kernel** and **bandwidth**; closer points have more influence.
- **Meaning of symbols**:
  - y: observed outcomes.
  - X: predictor matrix (including intercept).
  - $W(s_0)$: diagonal matrix of spatial weights for s_0.
  - $\hat{\beta}(s_0)$: locally estimated intercept and slopes, varying by location.

- **How to interpret results**:
  - Map each $\hat{\beta}_k(s_0)$ to visualize spatial variation in predictor effects.
  - Local predictions: $X(s_0) \cdot \hat{\beta}(s_0)$.
  - Local R^2 and residuals indicate model fit at each location.
- **Practical notes**:
  - **Bandwidth controls smoothness**: small $\rightarrow$ more local detail (**higher variance**); large $\rightarrow$ smoother, closer to global (**higher bias**).
  - Check local standard errors, t-values, and use multiple testing corrections.
  - Watch for **local multicollinearity** and edge effects—these can destabilize $\hat{\beta}(s_0)$.
  - If predictors operate at different scales, use MGWR (variable-specific bandwidths) for more robust inference.

Kernel Functions in GWR

1. Gaussian Kernel

$$w(s_i, s_0) = \exp\left(-\frac{d_{i0}^2}{2b^2}\right)$$

2. Bisquare Kernel

$$w(s_i, s_0) = \begin{cases} \left(1 - \frac{d_{i0}^2}{b^2}\right)^2 & \text{if } d_{i0} \leq b \\ 0 & \text{if } d_{i0} > b \end{cases}$$

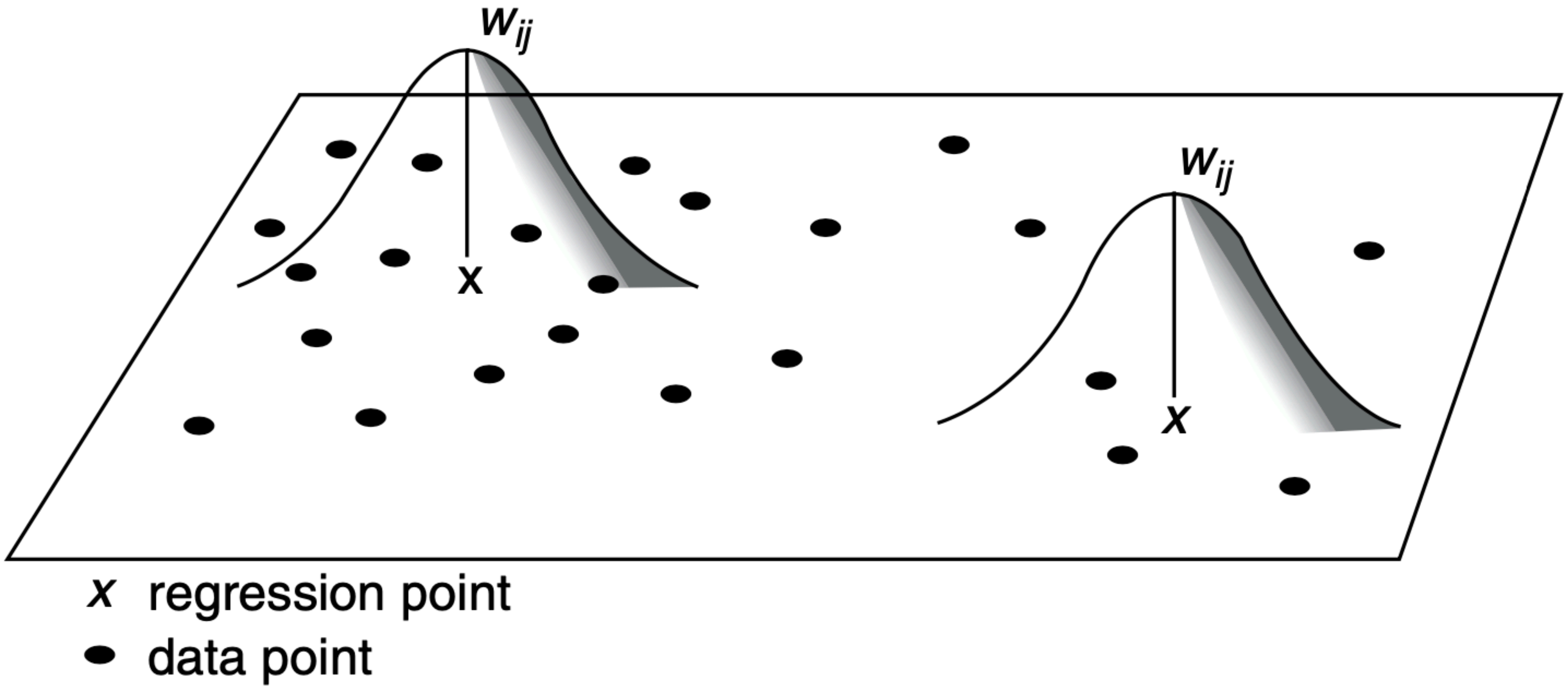
3. Exponential Kernel

$$w(s_i, s_0) = \exp\left(-\frac{d_{i0}}{b}\right)$$

Where $d_{i0} = ||s_i - s_0||$ and b is the bandwidth parameter.

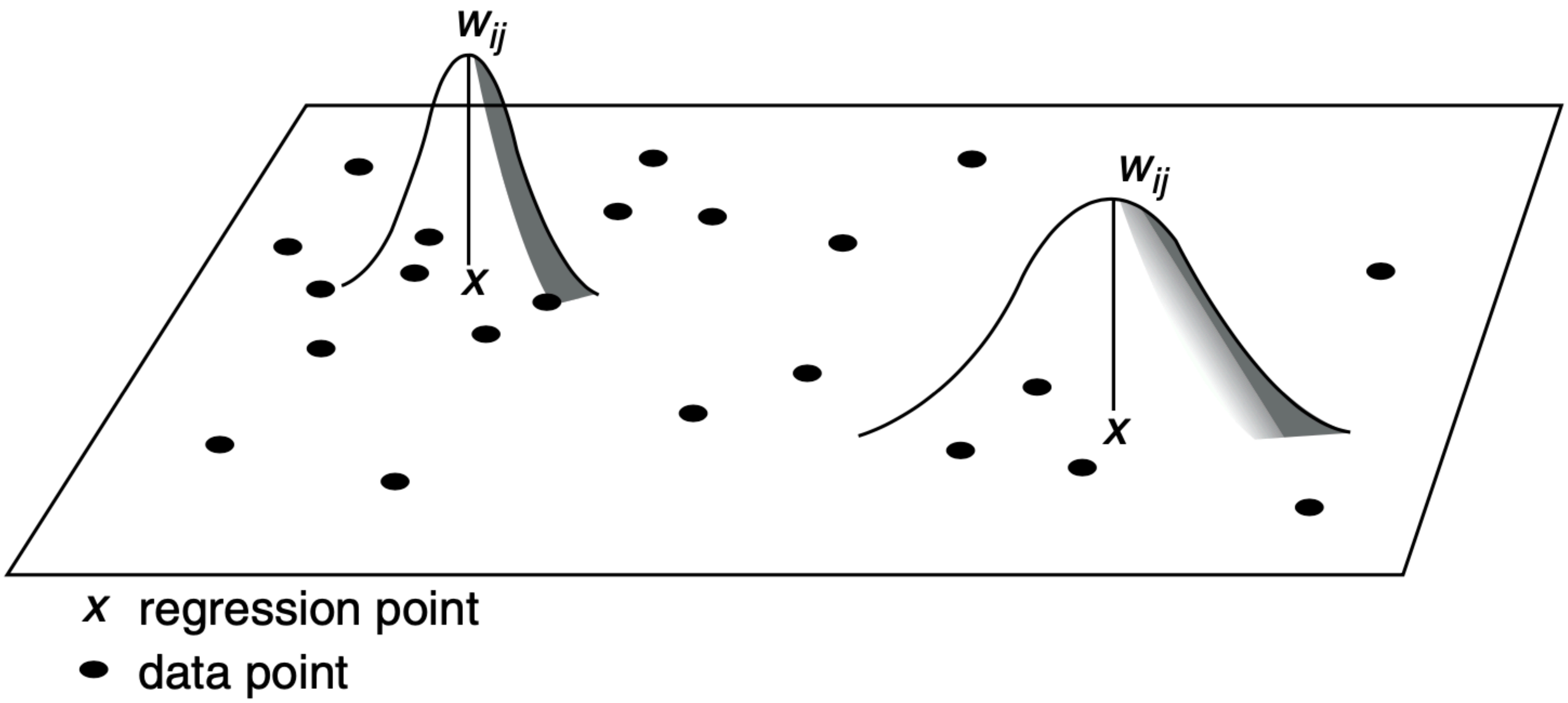
GWR — Intuition (fixed kernel)

- For each location x, weight observations by distance via a **kernel**.
- Fit a local regression; move across space; map **local coefficients** and **local R²**.



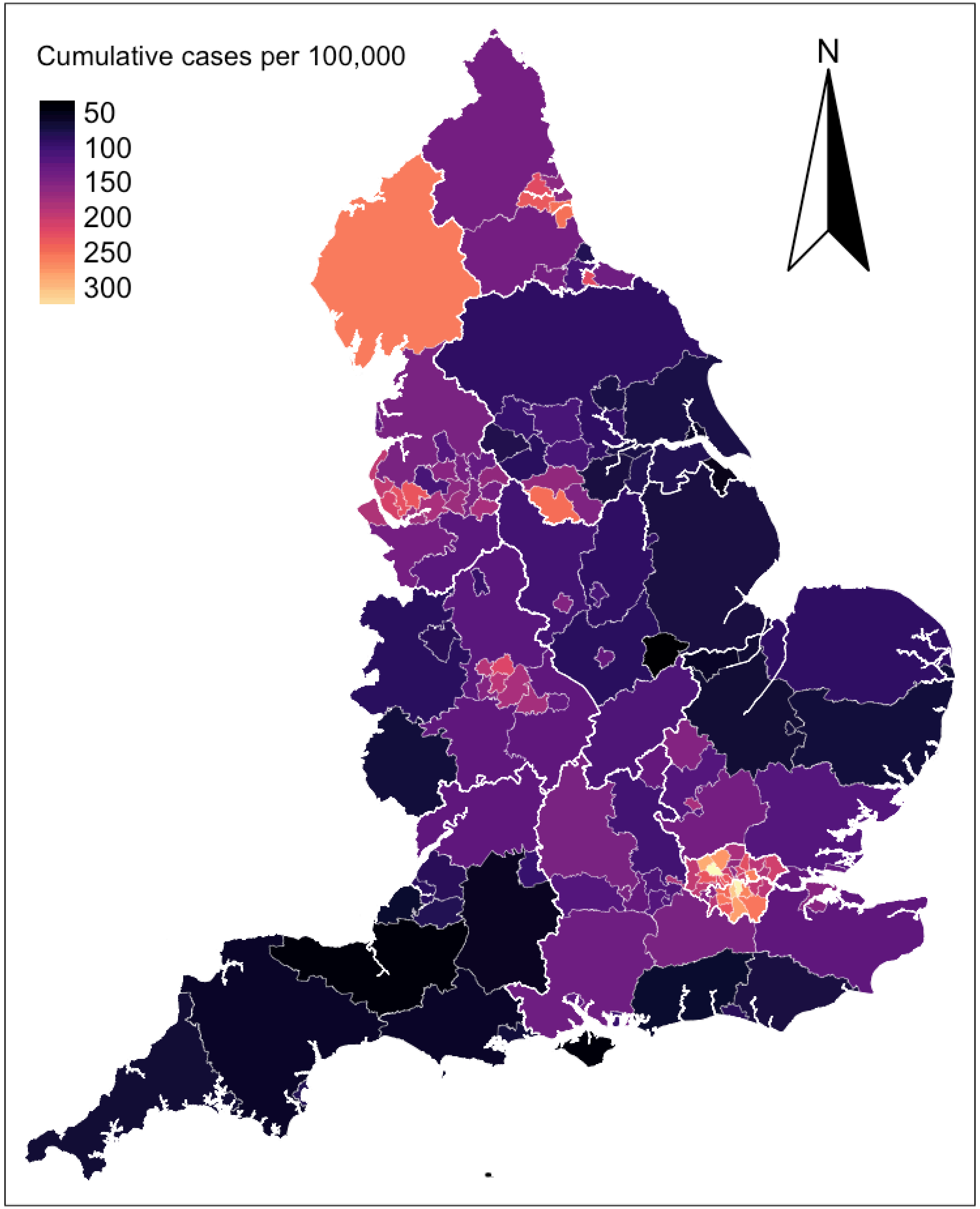
GWR — Intuition (adaptive kernel)

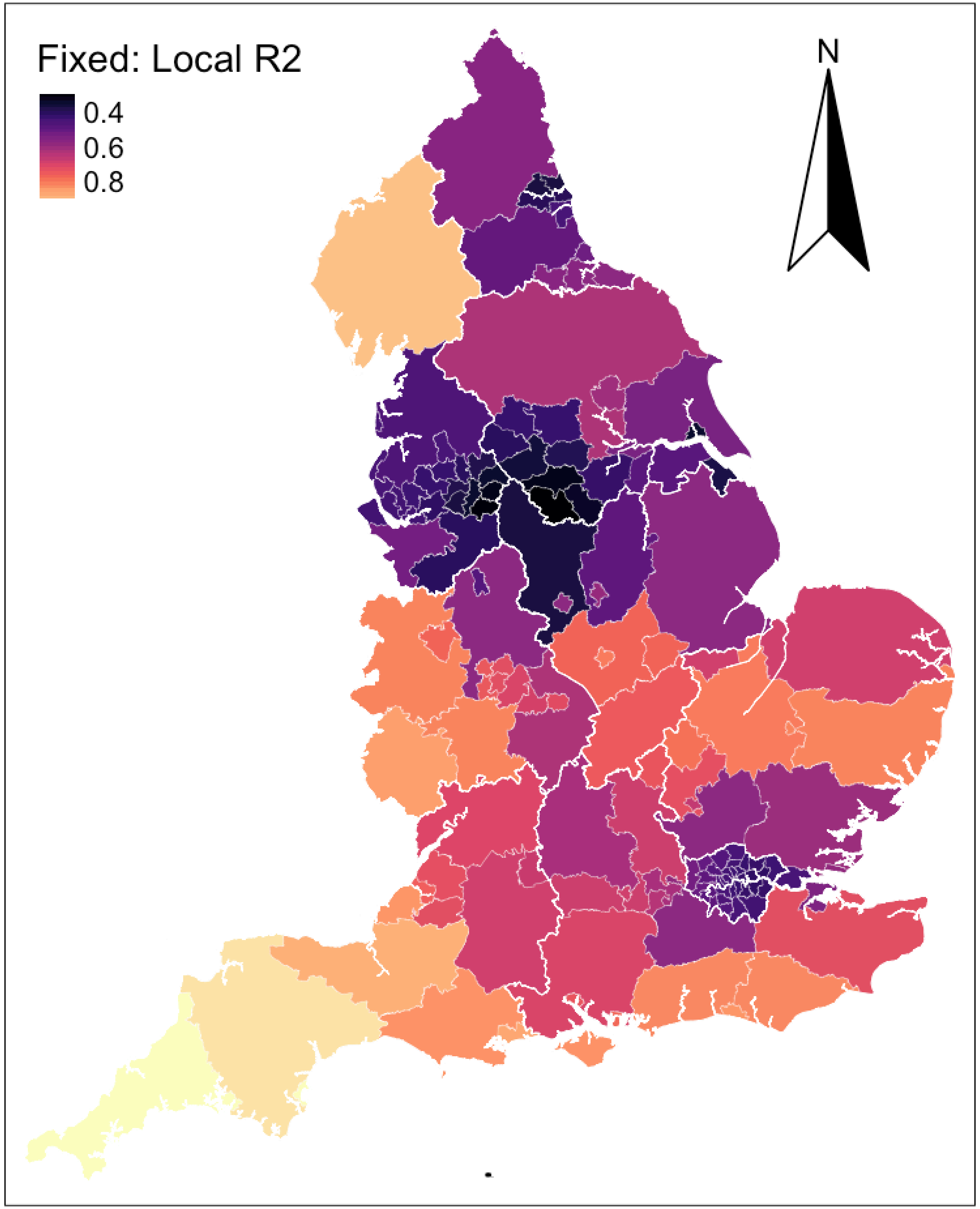
- **Adaptive** kernels vary their bandwidth: **larger** where data are sparse, **smaller** where dense.
- Often yields more stable local fits when polygon/point densities vary.

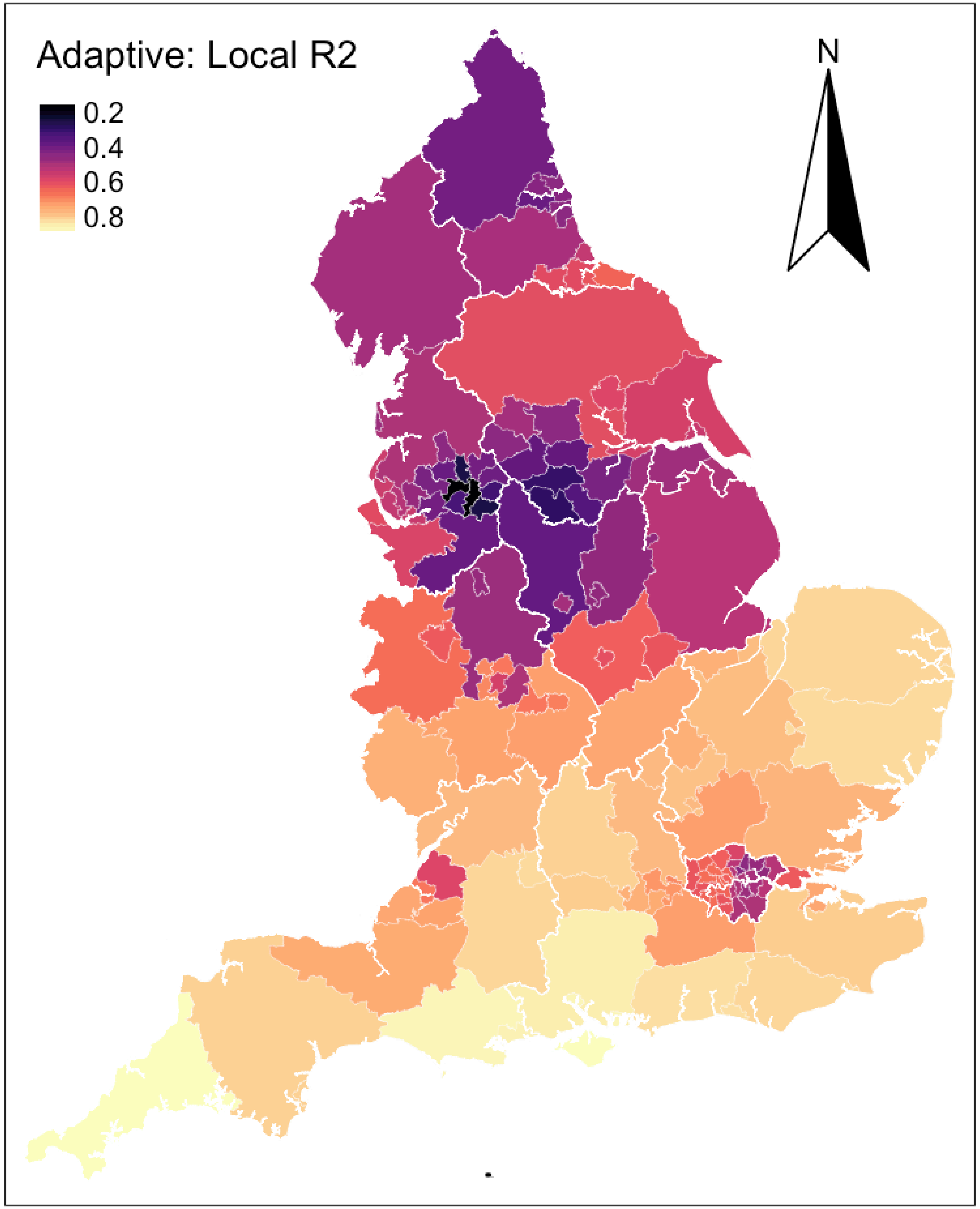


Example

(data distribution)




Fixed: Local R2
0.4
0.6
0.8
N


Adaptive: Local R2
0.2
0.4
0.6
0.8
N

Model Diagnostics and Validation

1. Spatial Stationarity Tests

Monte Carlo Test

Test H_0: $\beta_k(s) = \beta_k$ (constant) vs H_1: $\beta_k(s) \neq \beta_k$

Procedure:

- Calculate observed variance: $V_k = \mathrm{Var}[\hat{\beta}_k(s)]$
- Generate random permutations of coordinates
- Recalculate variance for each permutation
- Compare observed variance to permutation distribution

Interpretation:

- The p-value for each coefficient is the proportion of permutations where the permuted variability ≥ the observed variability.
- If p-value is small (say p-value < 0.05) you reject the null hypothesis of parameter stationarity for that coefficient and conclude its local estimates vary more than expected by chance.

2. Local Multicollinearity Diagnostics

Local Condition Number

$$\mathrm{CN}_i = \frac{\lambda_{\max}(X_i^T W_i X_i)}{\lambda_{\min}(X_i^T W_i X_i)}$$

Rule of thumb: $\mathrm{CN}_i > 30$ indicates multicollinearity. It measures how much your local predictor variables are linearly dependent (collinear) when you weight observations around that location.

Local Variance Inflation Factor

$$\mathrm{VIF}_{ki} = \frac{1}{1 - R_{ki}^2}$$

Where R_{ki}^2 is the local R^2 from regressing x_k on other predictors.

3. Significance Testing

Local t-statistics

$$t_{ki} = \frac{\hat{\beta}_{ki}}{\mathrm{SE}(\hat{\beta}_{ki})}$$

Multiple Testing Corrections

- **Why corrections are needed**: Many location × variable tests increase false positives (declaring an effect significant when it is actually due to chance).
  - **Bonferroni**: $\alpha_{adj} = \alpha / n$
  - **FDR**: Benjamini-Hochberg procedure
  - **Permutation**: Spatial permutation tests
- **Practical tips**:
  - Report adjusted significance levels and corrected p-values.
  - Use permutation tests to account for spatial dependence.
  - Consider the effective number of tests when interpreting results.

What do we interpret from GWR?

Coefficient Maps

- **Sign patterns**: Positive/negative effects across space
- **Magnitude variation**: Strength of relationships
- **Spatial clustering**: Similar effects in nearby areas

Local R^2 Maps

- **Model fit variation**: Where the model works well/poorly
- **Spatial patterns**: Regions of high/low predictability
- **Diagnostic tool**: Identify problematic areas

Statistical Significance

- **Significant coefficients**: Reliable local estimates
- **Spatial patterns**: Clusters of significant effects
- **Multiple testing**: Corrected significance levels

Critical Limitations and Methodological Challenges

1. Statistical Issues

Multiple Testing Problem

- **Issue**: n locations × p variables = np hypothesis tests
- **Consequence**: Inflated Type I error rate
- **Solutions**:
  - Bonferroni correction: $\alpha_{adj} = \alpha/(np)$
  - False Discovery Rate (FDR) control
  - Spatial permutation tests
  - Focus on spatial patterns, not individual tests

Local Multicollinearity

- **Cause**: Overlapping kernels create correlated local regressors
- **Diagnosis**: Local condition numbers, VIF
- **Thresholds**: CN > 30, VIF > 10
- **Solutions**:
  - Variable selection/regularization
  - Increase bandwidth
  - Remove highly correlated variables

2. Methodological Limitations

Bandwidth Sensitivity

- **Problem**: Results highly dependent on bandwidth choice
- **Impact**: Different bandwidths $\longrightarrow$ different conclusions
- **Mitigation**:
  - Sensitivity analysis across bandwidth range
  - Cross-validation for bandwidth selection
  - Report bandwidth selection criteria

Kernel Choice Effects

- **Gaussian**: Smooth, continuous weights
- **Bisquare**: Sharp cutoff, may create discontinuities
- **Impact**: Different kernels $\longrightarrow$ different results
- **Recommendation**: Test multiple kernels, report sensitivity

3. Spatial and Computational Issues

Edge Effects

- **Problem**: Unstable estimates at study area boundaries
- **Cause**: Asymmetric local samples
- **Solutions**:
  - Buffer zones around study area
  - Adaptive bandwidths
  - Report edge effect diagnostics

Small Sample Problems

- **Issue**: Local regressions with few observations
- **Consequence**: Unstable, unreliable estimates
- **Detection**: Local effective sample size
- **Solutions**:
  - Minimum sample size requirements
  - Adaptive bandwidths
  - Spatial aggregation

4. Interpretation Challenges

Causality vs. Correlation

- **Warning**: GWR coefficients are not necessarily causal
- **Reality**: Spatial correlation ≠ spatial causation
- **Best Practice**:
  - Use for exploratory analysis
  - Combine with theoretical knowledge
  - Consider omitted variable bias

Scale Dependence

- **Problem**: Results depend on spatial units (MAUP)
- **Solution**: Test sensitivity to different aggregations
- **MGWR Advantage**: Explicitly models scale differences

5. Computational Considerations

Computational Complexity

- **GWR**: $O(n^2)$ for each location
- **MGWR**: $O(n^2 \times p)$ for bandwidth selection
- **Large datasets**: May require sampling or parallel processing
- **Memory**: Store $n \times n$ weight matrices

Software Limitations

- **mgwr**: Limited to Gaussian kernels
- **Memory**: Large datasets may exceed RAM
- **Convergence**: MGWR may not converge with poor starting values

MGWR: Multiscale Geographically Weighted Regression

The Multiscale Problem

Standard GWR: **All variables/predictors use the same bandwidth** b **(that can be adaptive)**

- Reality: **Different processes operate at different spatial scales**
- **Example**:
  - Income effects: Regional scale (large bandwidth)
  - Local amenities: Neighborhood scale (small bandwidth)
  - Climate: Continental scale (very large bandwidth)

MGWR

Variable-specific bandwidths: Each predictor k has its own bandwidth b_k

$$\hat{\beta}_k(s_0) = \arg\min_{\beta_k} \sum_{i=1}^{n} w_k(s_i, s_0) \left[y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right]^2 \texttt{ where } w_k(s_i, s_0) = K\left(\frac{d_{i0}}{b_k}\right)$$

Practical reasoning & guidance

- Workflow: run a global model $\longrightarrow$ run GWR to inspect coefficient surfaces $\longrightarrow$ if some coefficients look very smooth and others very noisy, try MGWR.
- Report: per-variable bandwidths (with units or effective sample sizes), coefficient maps, and a comparison to GWR (e.g., AICc, cross-validation).
- Communication tip: explain bandwidths in plain terms (e.g., "var1 varies at neighborhood scale; var2 varies regionally") rather than only reporting numbers.
- Common pitfalls: MGWR is computationally heavier and can overfit if bandwidths are too small; standardize predictors before bandwidth search; interpret cautiously where local multicollinearity or small local samples occur.

Advantages

1. **Scale-specific modeling**: Captures processes at their natural scales
2. **Reduced overfitting**: Prevents over-smoothing of local processes
3. **Improved interpretation**: Clearer understanding of spatial scales
4. **Better prediction**: More accurate local estimates

When to use GWR vs MGWR

Decision Framework

Use GWR when:

- **Exploratory analysis**: Initial investigation of spatial patterns
- **Single scale processes**: All variables operate at similar scales
- **Computational constraints**: Limited resources for complex models
- **Simple relationships**: Linear relationships expected

Use MGWR when:

- **Multiscale processes**: Variables operate at different scales
- **Theoretical knowledge**: Different scales expected a priori
- **Improved fit**: GWR shows poor fit or unstable results
- **Policy relevance**: Scale-specific interventions needed

References

Foundational Literature

Seminal Books

- **Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002)**. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.
- **Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003)**. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Key Journal Articles

- **Fotheringham, A. S., Yang, W., & Kang, W. (2017)**. Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247-1265.
- **Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., & Harris, P. (2022)**. A roadmap for handling multiscale modelling using geographically weighted regression. *Geographical Analysis*, 54(1), 1-25.
- **Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1996)**. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281-298.

Methodological Advances

MGWR and Extensions

- **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.
- **Wolf, L. J., Anselin, L., Arribas-Bel, D., & Oshan, T. M. (2020)**. A comparison of multiscale geographically weighted regression and multiscale spatial filtering approaches. *Geographical Analysis*, 52(4), 540-560.

Spatiotemporal Extensions

- **Huang, B., Wu, B., & Barry, M. (2010)**. Geographically and temporally weighted regression for modeling spatiotemporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383-401.
- **Fotheringham, A. S., Crespo, R., & Yao, J. (2015)**. Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, 47(4), 431-452.

Software and Implementation

Python Libraries

- **mgwr**: Multiscale Geographically Weighted Regression
  - GitHub: https://github.com/pysal/mgwr
  - Documentation: https://mgwr.readthedocs.io/
- **PySAL**: Python Spatial Analysis Library
  - Website: https://pysal.org/
  - GWR module: https://pysal.org/spreg/

R Packages

- **GWmodel**: Geographically Weighted Models
  - CRAN: https://cran.r-project.org/package=GWmodel
- **spgwr**: Geographically Weighted Regression
  - CRAN: https://cran.r-project.org/package=spgwr

Online Resources and Tutorials

Comprehensive Guides

- **GDSL-UL SAN — 09: Geographically Weighted Regression** https://gdsl-ul.github.io/san/09-gwr.html
- **Deepnote — [PYTHON] GWR and MGWR** https://deepnote.com/app/carlos-mendez/PYTHON-GWR-and-MGWR-71dd8ba9-a3ea-4d28-9b20-41cc8a282b7a

Tutorials and Examples

- **Spatial Analysis and Modeling Course Materials**
  - https://gdsl-ul.github.io/san/
- **PySAL Tutorials**
  - https://pysal.org/notebooks/
- **GWmodel Tutorial**
  - https://cran.r-project.org/web/packages/GWmodel/vignettes/GWmodel.html