武漢大學
WUHAN UNIVERSITY

遥感信息工程学院
School of Remote Sensing and Information Engineering

笃志 敦行 和协 拓新

Geographically weighted models and tools: GWmodel and GWmodelS

Binbin Lu
School of Remote Sensing and Information Engineering,
Wuhan University

Aug 27, 2024

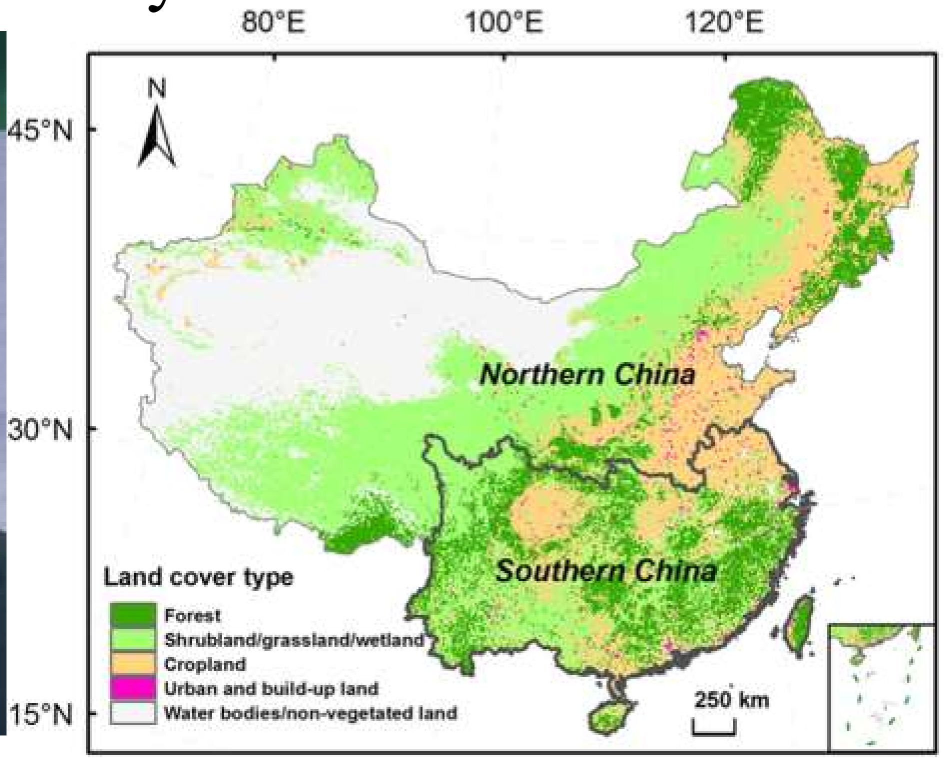
BEGIN seminar @ the University of St Andrews

Contents

- Background: spatial heterogeneity
- Geographically weighted (GW) models
  - ✓GW Regression – standard and multiscale
  - ✓GW Summary Statistics
  - ✓GW principal components analysis
- GWmodel and GWmodelS
- Ending remarks

Background: spatial heterogeneity

- Spatial heterogeneity/spatial non-stationarity



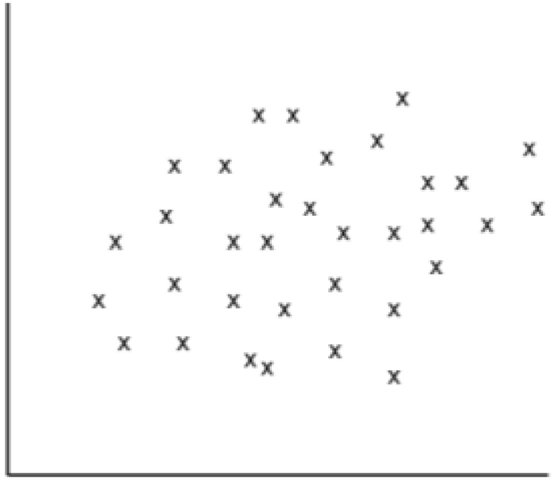
When orange trees grow south of the Huai River, they produce sweet oranges, but when they are grown north of the Huai River, they produce bitter trifoliate oranges. Though the leaves appear similar, the taste of the fruit is different, due to the differences in soil and water.“----(Yanzi, ~700 B.C.)

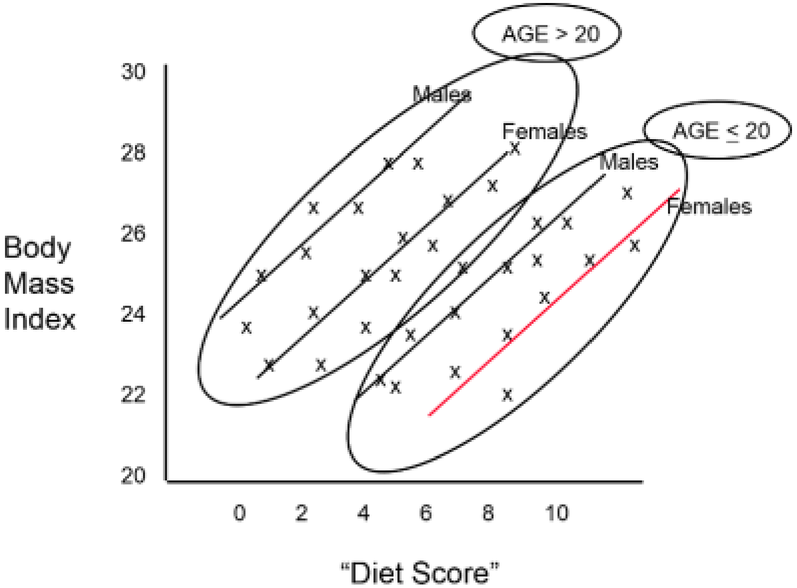
Background: spatial heterogeneity

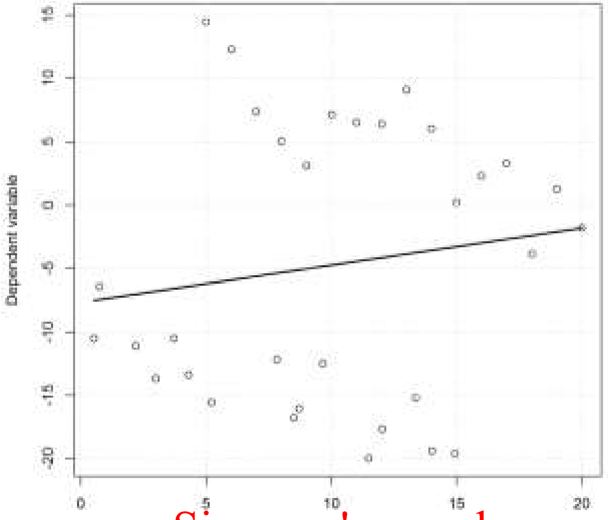
- A candidate second law of geography by Goodchild (2004)
  - **The principle of spatial heterogeneity**
  - "Spatial heterogeneity or non-stationarity in the statistical meaning of that term, implies that geographic variables exhibit uncontrolled variance."
  - "A corollary of uncontrolled variance in space is that the results of analysis depend explicitly on the bounds of the analysis: move the study area, and the results will change"

Background: spatial heterogeneity

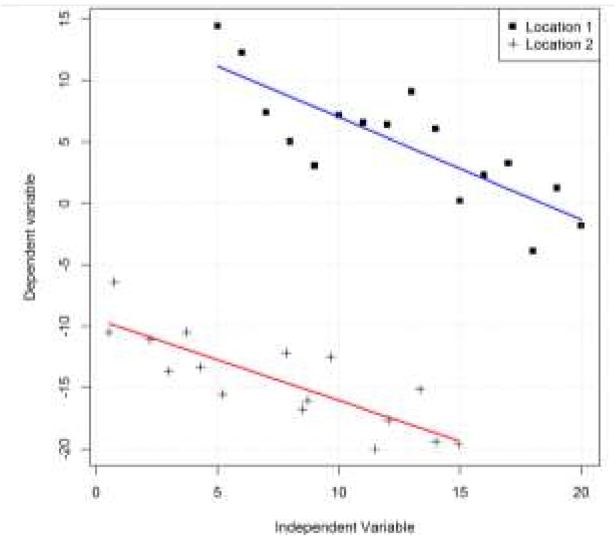
➢ Spatial heterogeneity ---- Local techniques

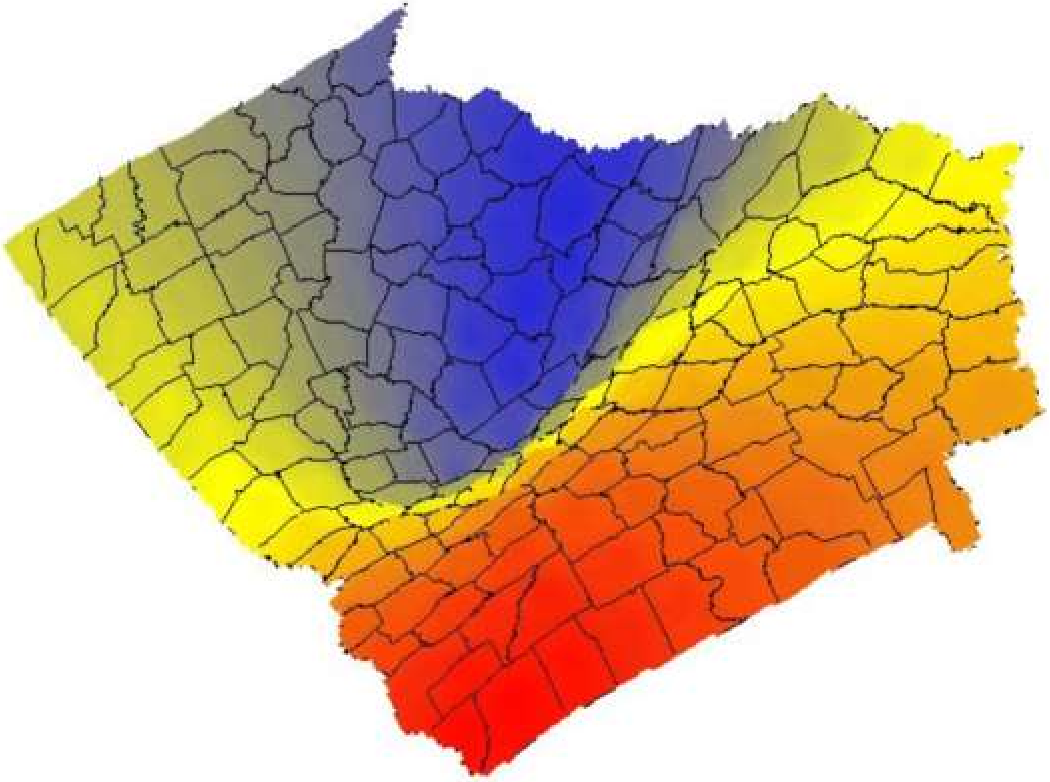






Simpson's paradox





Spatially varying relationships

Background: spatial dependence – weighting

- The first law of geography by Tobler (1970)
  - ✓ "Everything is related to everything else, but near things are more related than distant things"
  - ✓ Spatial dependence
- Modelling rule: Geographically weighting by distance or spatial proximity - decaying

- General expression:

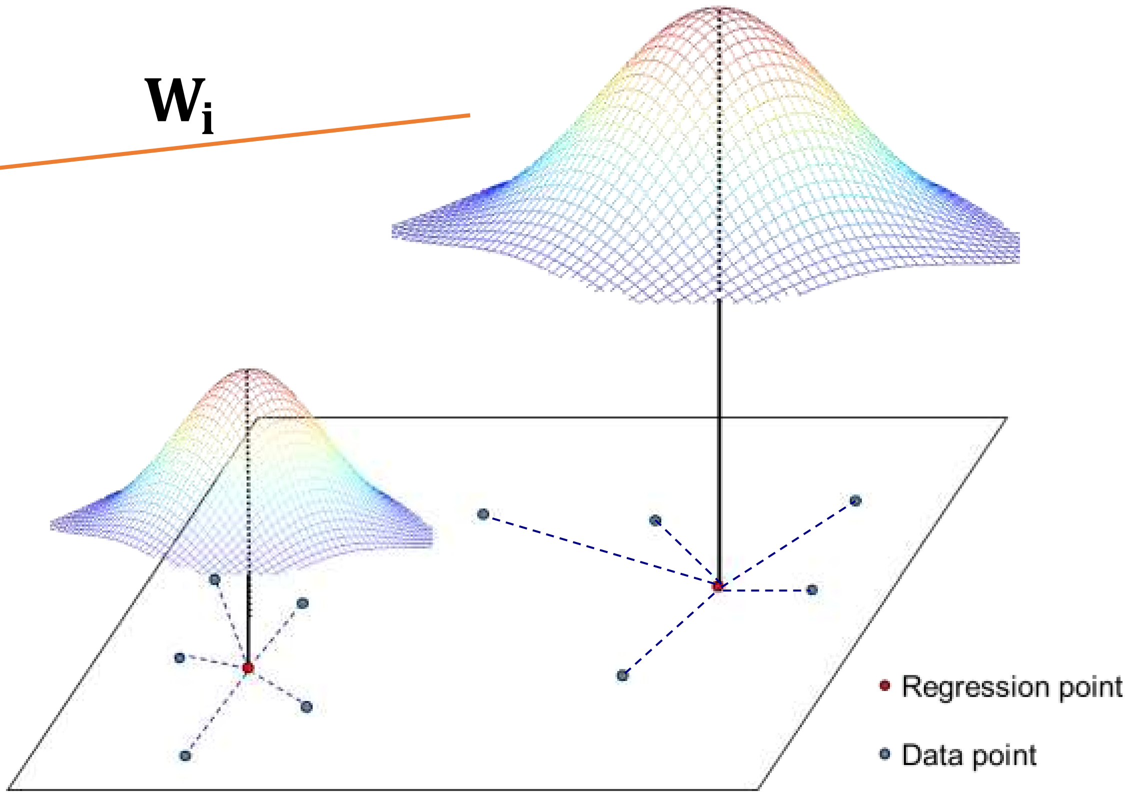
$$y_i = \beta_0(u_i,v_i) + \beta_1(u_i,v_i)x_{i1} + \cdots + \beta_m(u_i,v_i)x_{m1} + \varepsilon_i$$

- Weighted least square (WLS)

$\mathbf{W_i}$

$$\hat{\beta}_i = (\mathbf{X^t W_i X})^{-1}\mathbf{X^T W}_i y$$

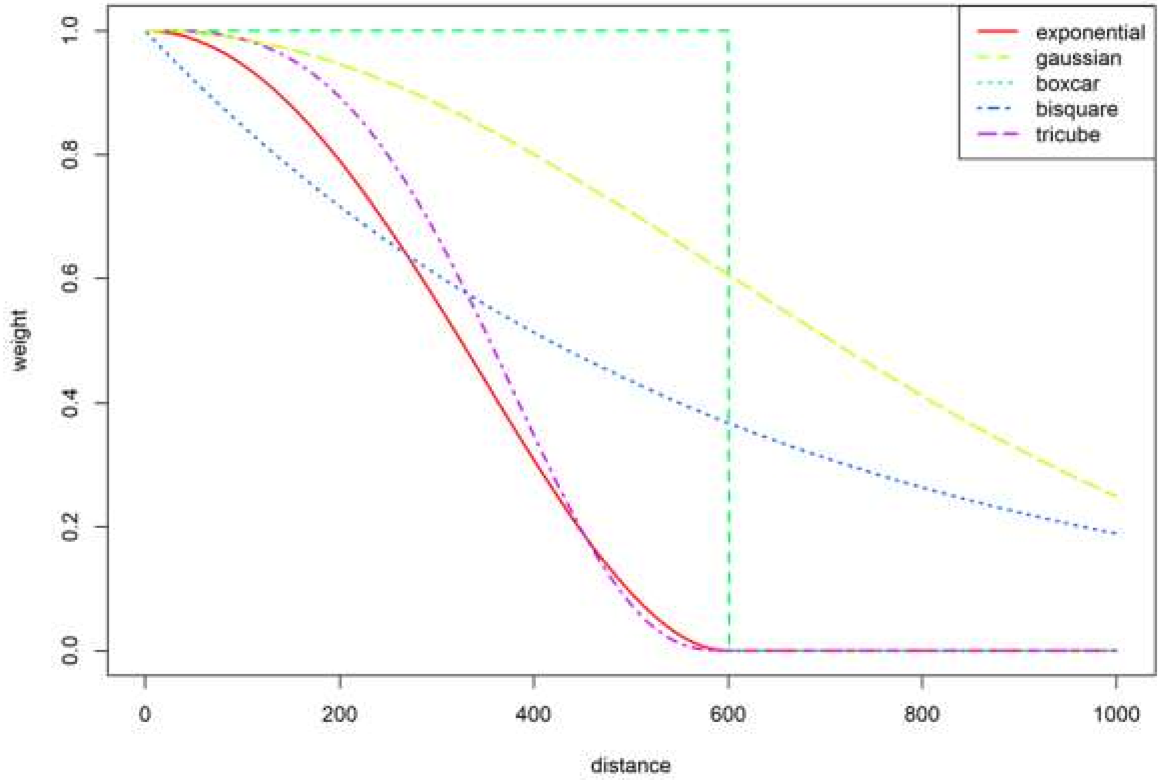
Distance - decaying



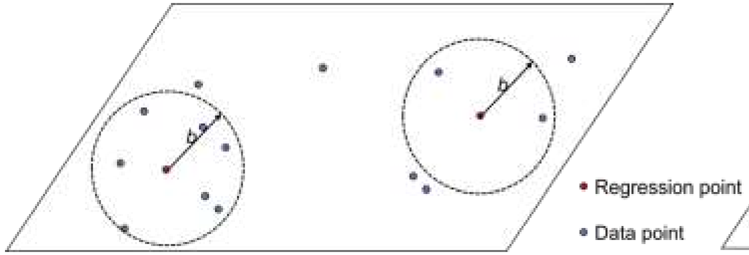
$$\mathbf{W_i} = \begin{matrix} w_{i1} & 0 & 0 & \ldots & 0 \\ 0 & w_{i2} & 0 & \ldots & 0 \\ 0 & & w_{i3} & \ldots & 0 \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & w_{in} \end{matrix}$$

Geographically Weighted Regression

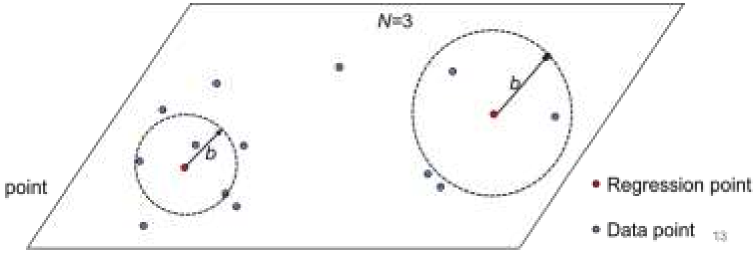
- Global model $$w_{ij} = 1, \ \forall i, j$$
- Gaussian $$w_{ij} = e^{-\frac{\left(d_{ij}/b\right)^2}{2}}$$
- Exponential $$w_{ij} = \exp\left(-\frac{|d_{ij}|}{b}\right)$$
- Box-car $$w_{ij} = \begin{cases} 1, & if\ d_{ij} \le b \\ 0, & otherwise \end{cases}$$
- Bi-square $$w_{ij} = \begin{cases} \left(1-\left(d_{ij}/b\right)^2\right)^2, & if\ d_{ij} \le b \\ 0, & otherwise \end{cases}$$
- Tri-cube $$w_{ij} = \begin{cases} \left(1-\left(d_{ij}/b\right)^3\right)^3, & if\ d_{ij} \le b \\ 0, & otherwise \end{cases}$$



Kernel functions

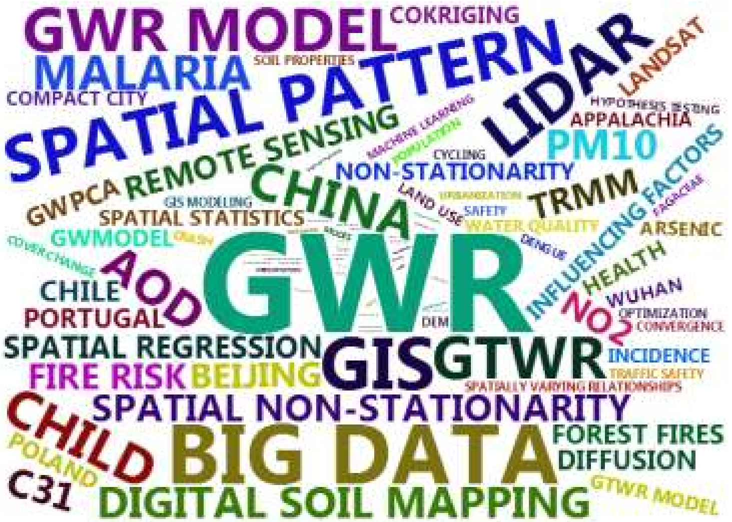


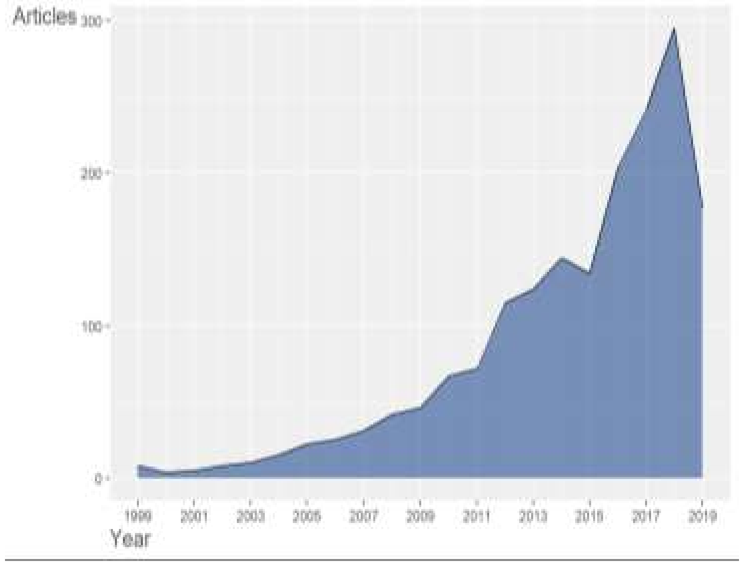
Fixed weighting scheme

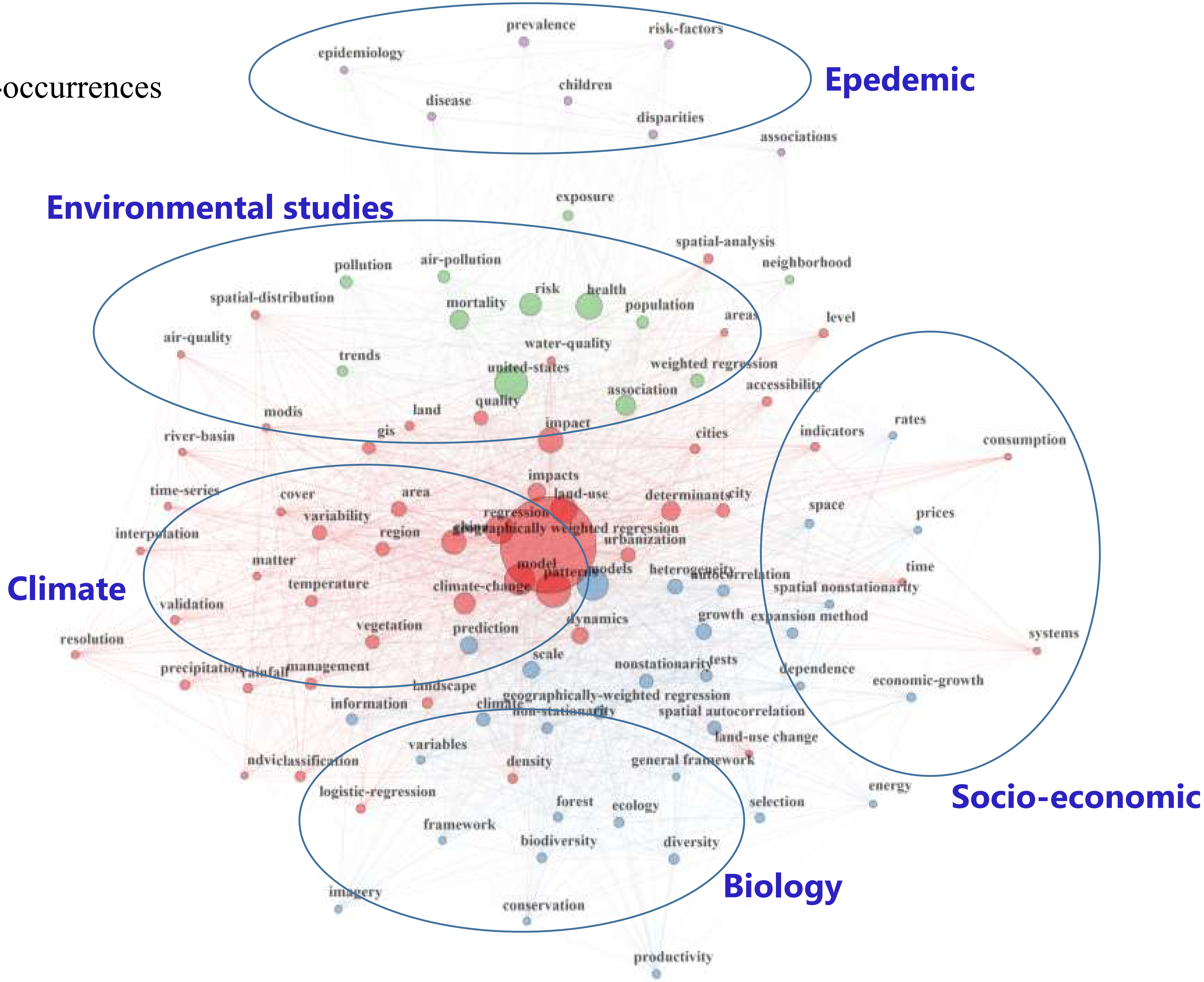


Adaptive weighting scheme

Keyword Co-occurrences







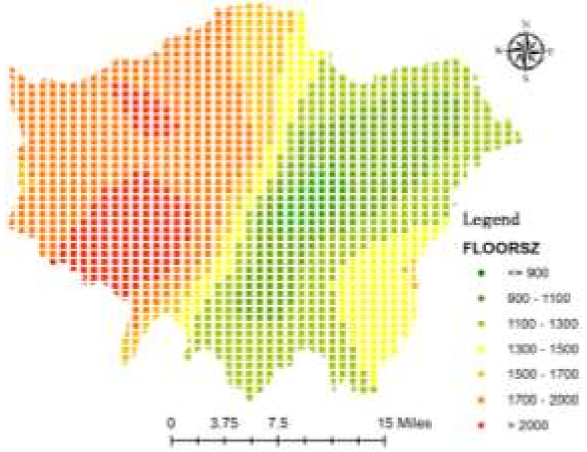
Multiscale Geographically Weighted Regression

- Mixed GWR (Brunsdon et al., 1999)

$$y_i = \sum_{j=1,k_a} a_j x_{ij}(a) + \sum_{j=1,k_b} b_l(u_i,v_i) x_{il}(b) + \varepsilon_i$$

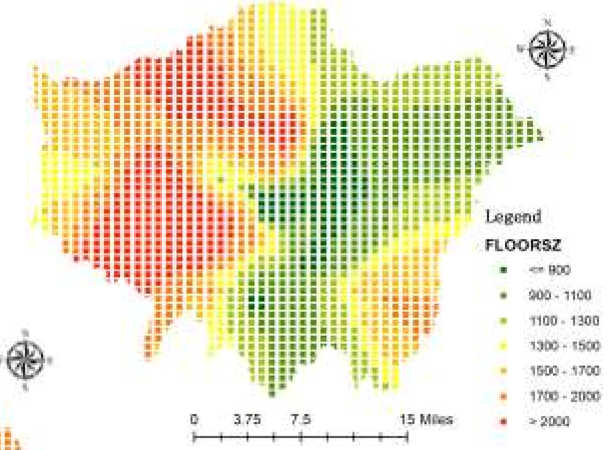
- GWR with flexible bandwidths (Yang, 2014)

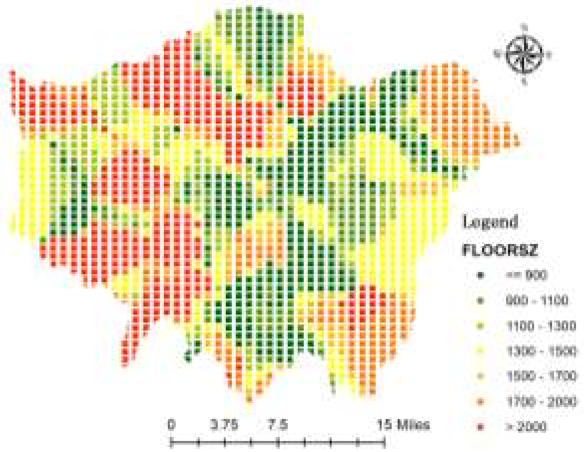
$$y_i = \beta_{bw_0}(u_i, v_i) + \sum_{k=1,n} \beta_{bw_k}(u_i, v_i) x_{ik} + \varepsilon_i$$



Bandwidth=100

Bandwidth=50





Bandwidth=10

Multiscale Geographically Weighted Regression

- GWR with parameter-specific distance metrics (PSDM GWR, Lu et al. 2017, 2018)

$$y_i = \beta_{DM_0, bw_0}(u_i, v_i) + \sum_{k=1,n} \beta_{DM_k, bw_k}(u_i, v_i) x_{ik} + \varepsilon_i$$

- Multiscale GWR (FBGWR, Fortheringham et al. 2017)

- Conditional geographically weighted regression (CGWR, Leong and Yue 2017)

- Multiscale geographically and temporally weighted regression (Wu et al. 2018)

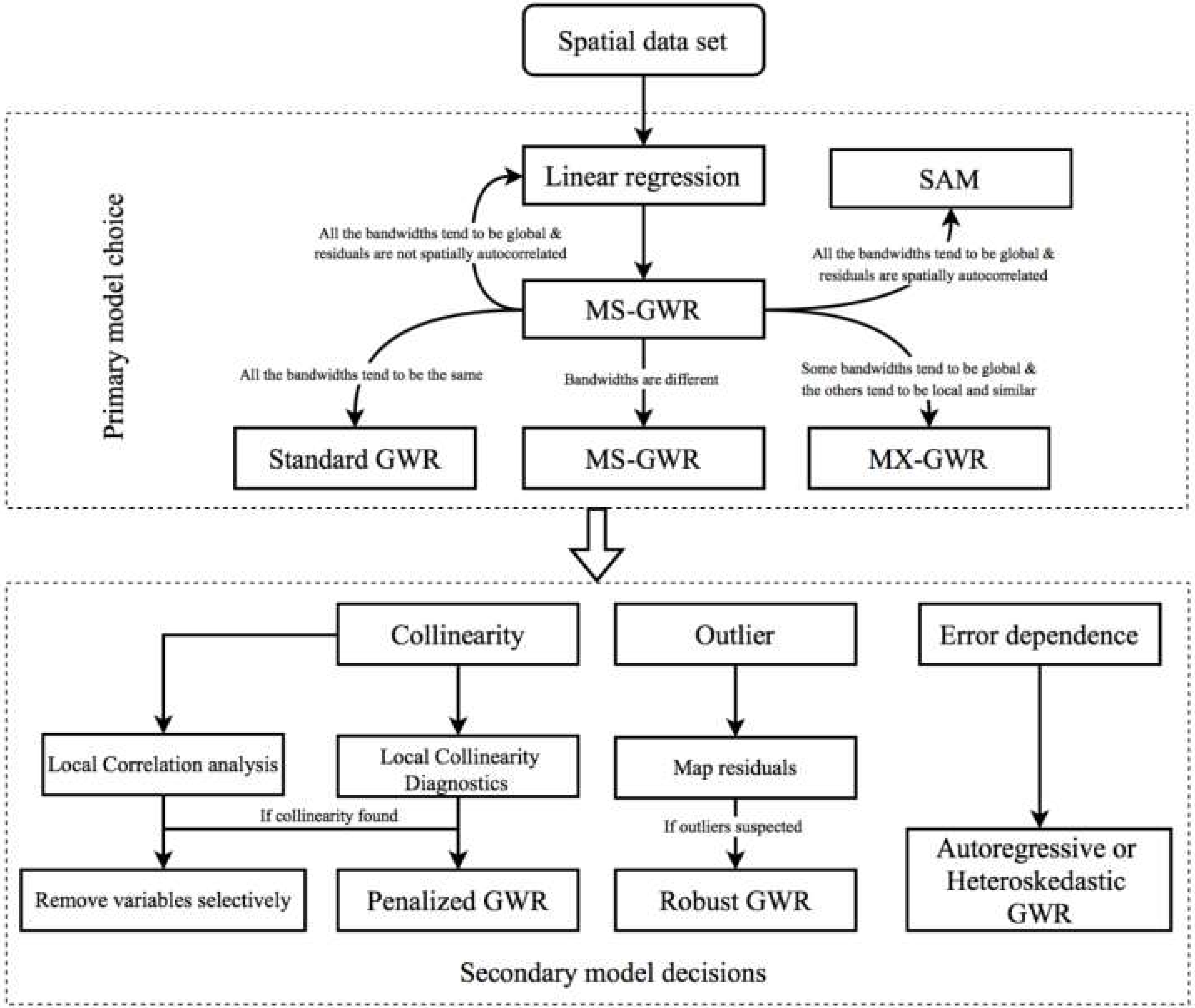
Multiscale Geographically Weighted Regression

Standard GWR

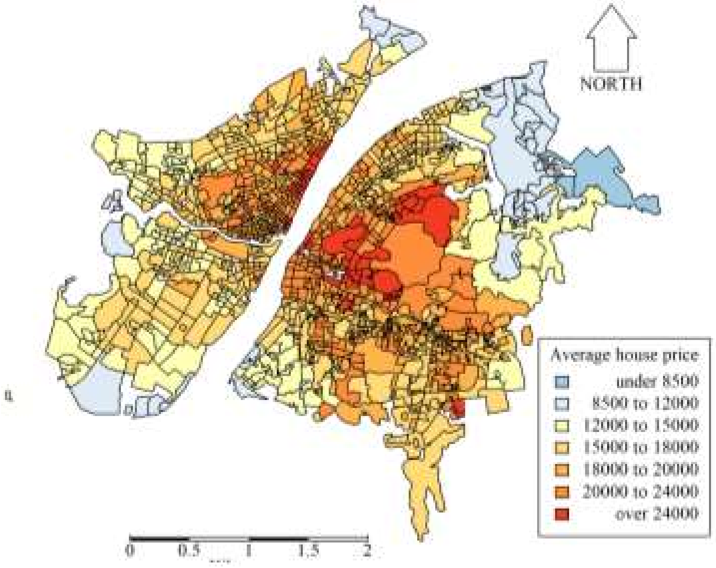
MS-GWR：multiscale GWR

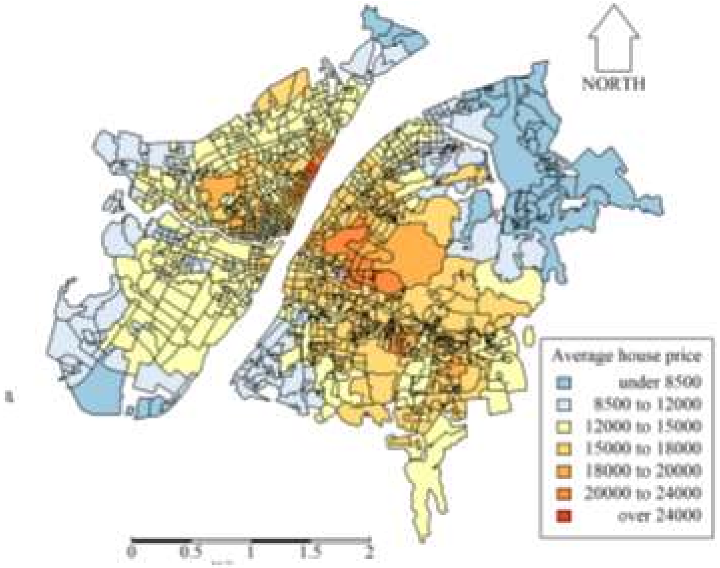
MX-GWR：mixed GWR

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., Harris, P., 2023. **A route map for successful applications of geographically weighted regression**. *Geographical Analysis* 55 (1), 155-178.



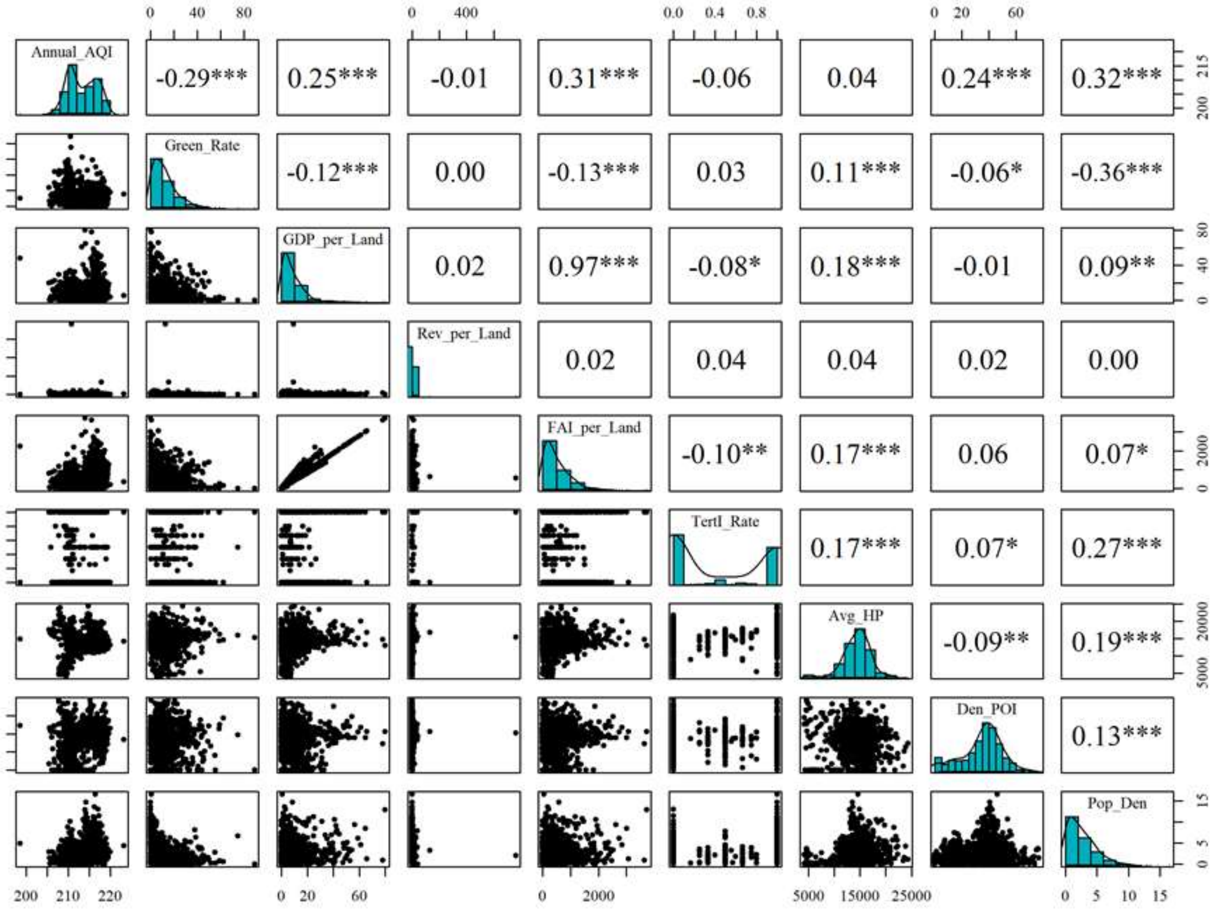
Multiscale Geographically Weighted Regression



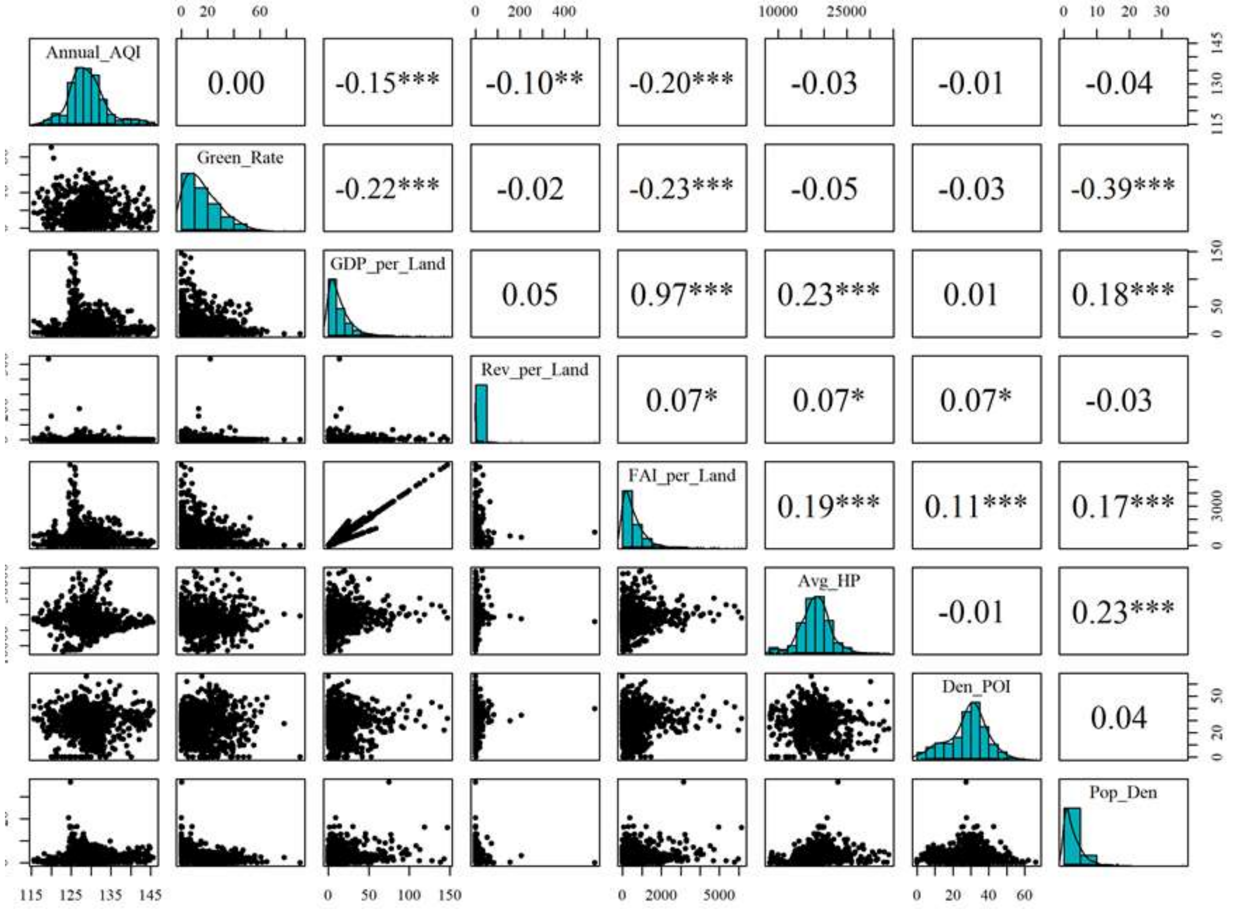


Variable	**Description**
Avg_HP	Average house price at the community-level (*Yuan/m*^{2})
Annual_AQI	Annual air quality index (AQI)
Pop_Den	Population density ($*10^5$ / km^2)
Green_Rate	Percentage of green space (%)
GDP_per_Land	GDP per land area ($*10^4$ *Yuan/* km^2)
Rev_per_Land	Revenue per land area ($*10^4$ *Yuan/* km^2)
FAI_per_Land	Fixed assets investment per land area ($*10^4$ *Yuan/* km^2)
TertI_Rate	Percentage of tertiary industry (%)
Den_POI	Density of places of interest (POI) ($*10^{-2}$/ km^2)

Multiscale Geographically Weighted Regression



2015



2019

Multiscale Geographically Weighted Regression

$$Avg\_HP = \beta_0 + \beta_1 * Green\_Rate + \beta_2 * GDP\_per\_Land + \beta_3 * Rev\_per\_Land + \beta_4 * Den\_POI + \beta_5 * Pop\_Den$$

- Spatial regression models
  - ✓ Linear regression, LR
  - ✓ Spatial lag model , SLM
  - ✓ Basic GWR
  - ✓ Multiscale GWR

Multiscale Geographically Weighted Regression

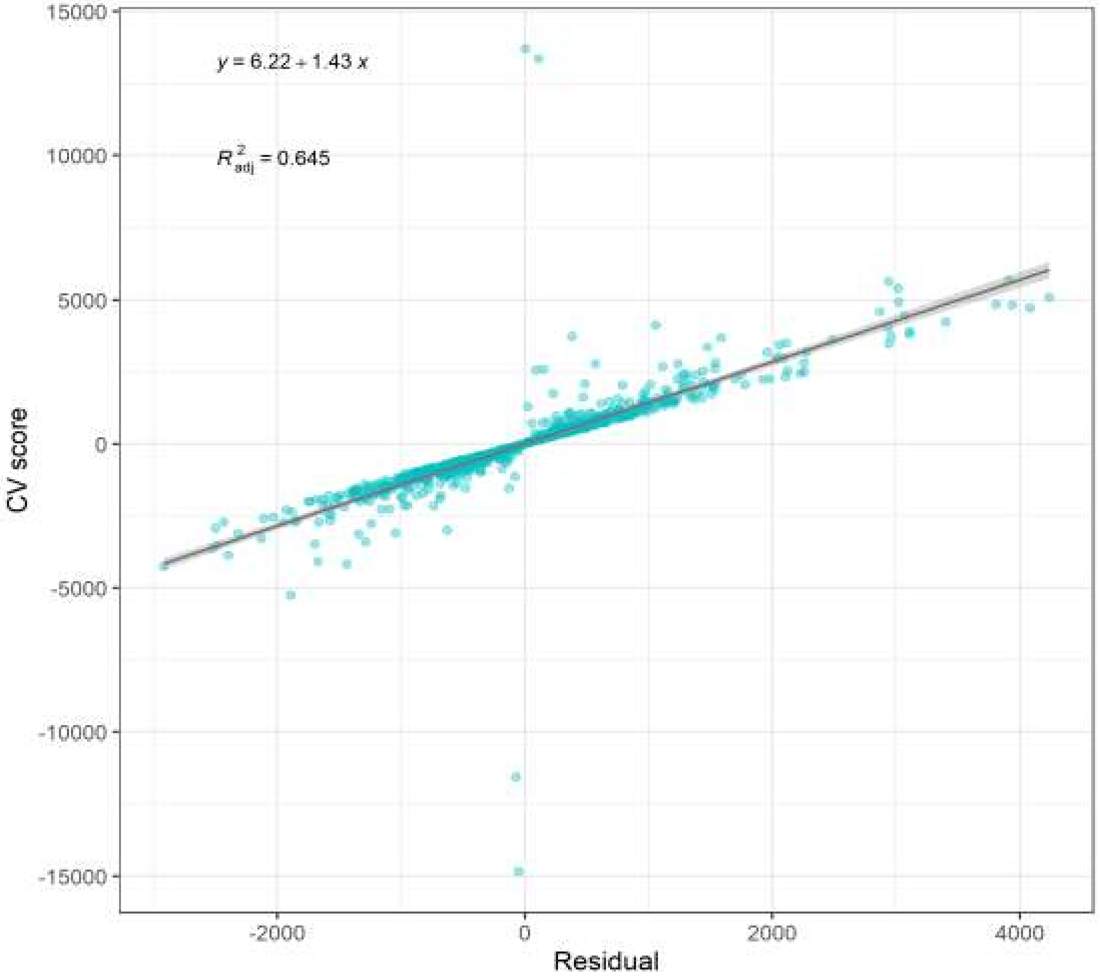
$$Avg\_HP = \beta_0 + \beta_1 * Green\_Rate + \beta_2 * GDP\_per\_Land + \beta_3 * Rev\_per\_Land + \beta_4 * Den\_POI + \beta_5 * Pop\_Den$$

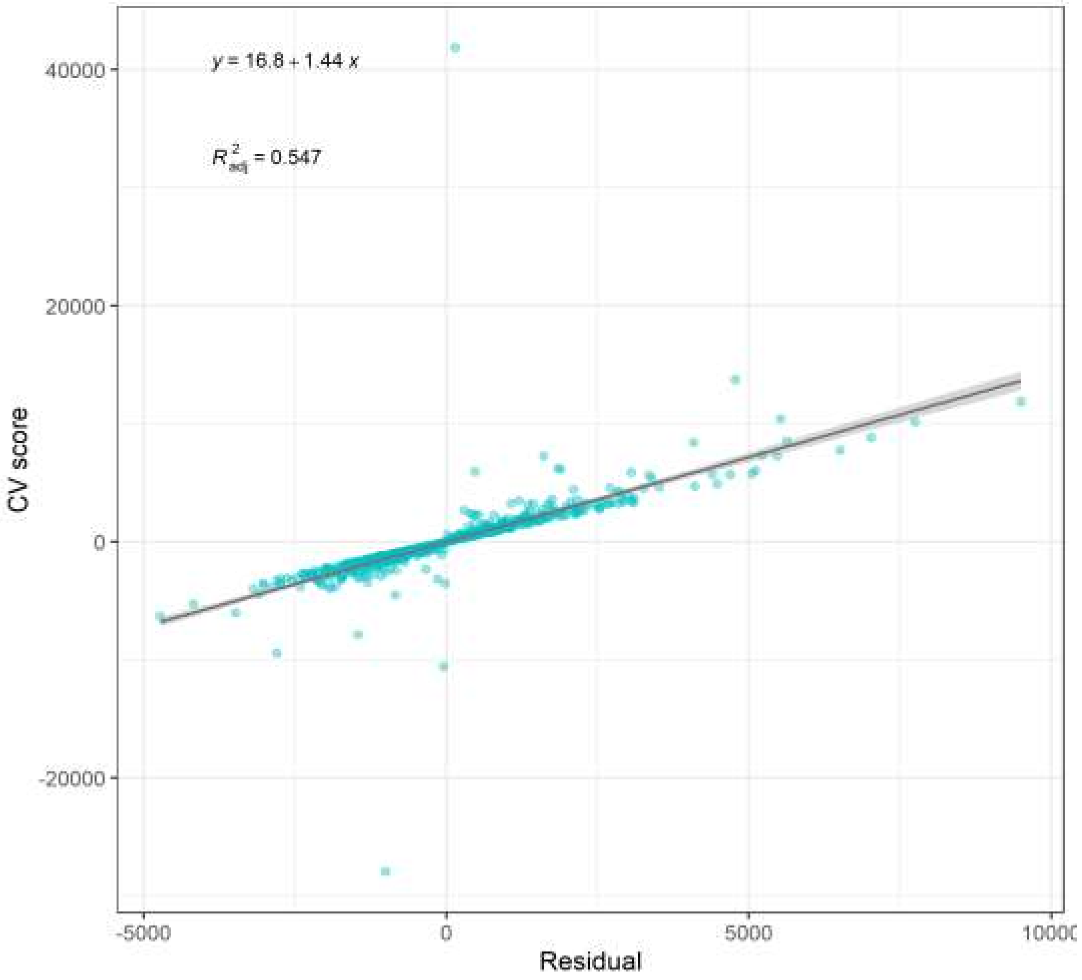
2015	LR	Adjusted R^2	0.114					
		AICc	18267.87					
		AIC	18267.75					
		RSS	7856055677					
	SLM	Adjusted R^2	0.213					
		AIC	16511					
		RSS	1005346118					
	GWR	Bandwidth	45 (number of nearest neighbors)					
		Adjusted R^2	0.868					
		AICc	16757.53					
		AIC	16240.47					
		RSS	744307992					
	MGWR	Coefficient	β_0	β_1	β_2	β_3	β_4	β_5
		Bandwidth	14	398	22	145	328	22
		Adjusted R^2	0.941					
		AICc	16108.75					
		AIC	15342.66					
		RSS	274926241					

2019	LR	Adjusted R^2	0.094					
		AICc	18715.25					
		AIC	17823.47					
		RSS	12436195146					
	SLM	Adjusted R^2	0.272					
		AIC	17162					
		RSS	1958635540					
	GWR	Bandwidth	47 (number of nearest neighbors)					
		Adjusted R^2	0.807					
		AICc	17539.42					
		AIC	17042.59					
		RSS	1707951694					
	MGWR	Coefficient	β_0	β_1	β_2	β_3	β_4	β_5
		Bandwidth	12	50	24	107	27	24
		Adjusted R^2	0.919					
		AICc	17101.01					
		AIC	15954.91					
		RSS	474525718					

Lu, B., Ge, Y., Shi, Y., Zheng, J., Harris, P., 2023. Uncovering drivers of community-level house price dynamics through multiscale geographically weighted regression: A case study of wuhan, china. *Spatial Statistics* 53, 100723.

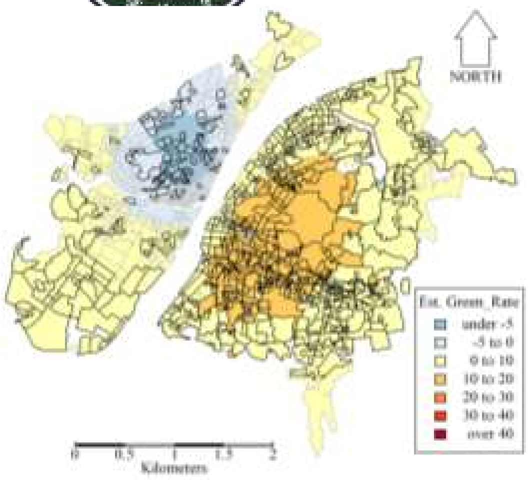
Multiscale Geographically Weighted Regression



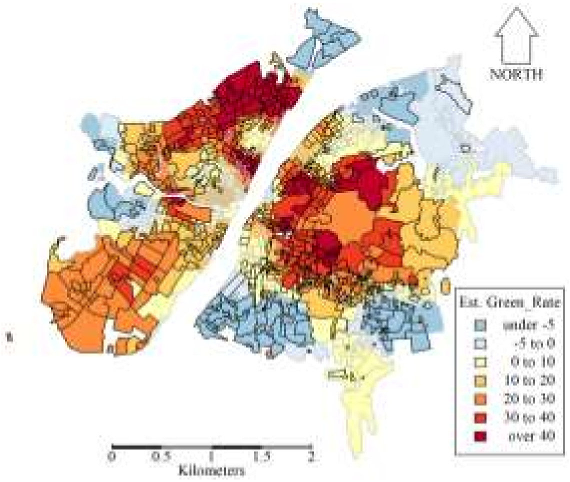


Multiscale GWR - Locally overfitting? Validated by cross-validation approach

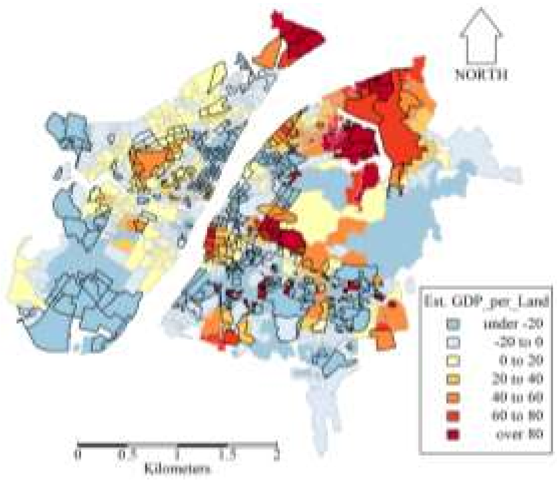
Multiscale Geographically Weighted Regression



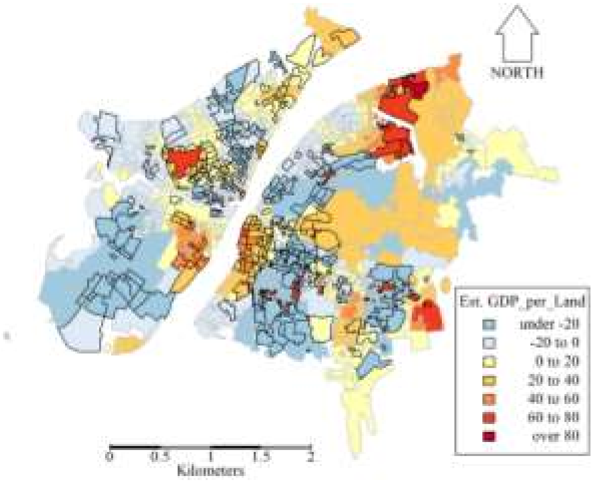
2015



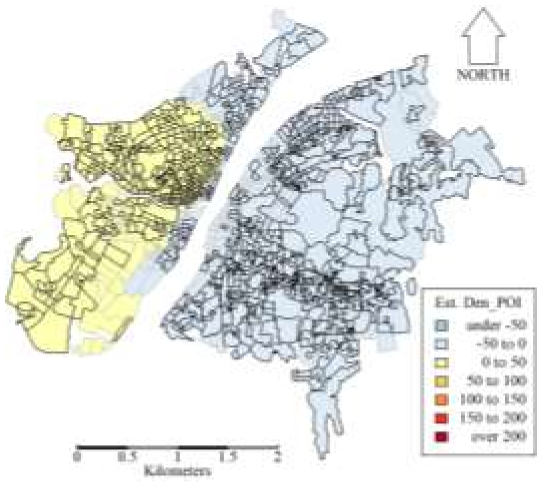
2019



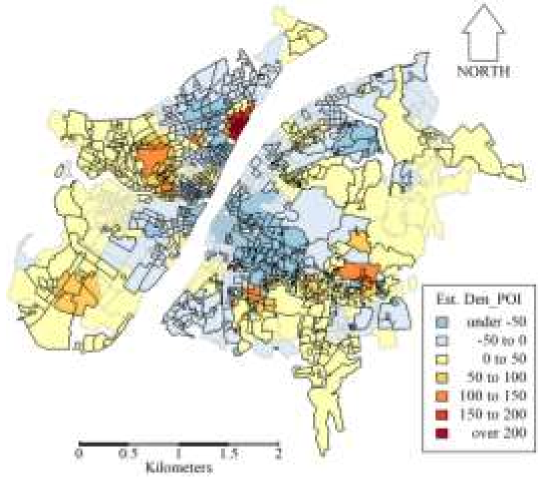
2015



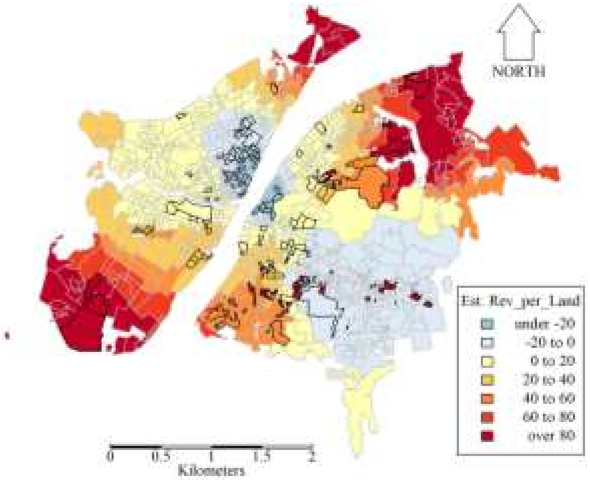
2019



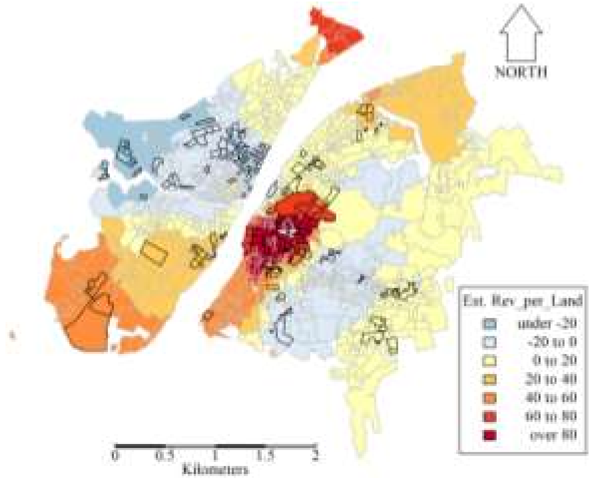
2015



2019



2015



2019

GW models – a locally modelling framework

Spatial heterogeneity

Spatial dependence

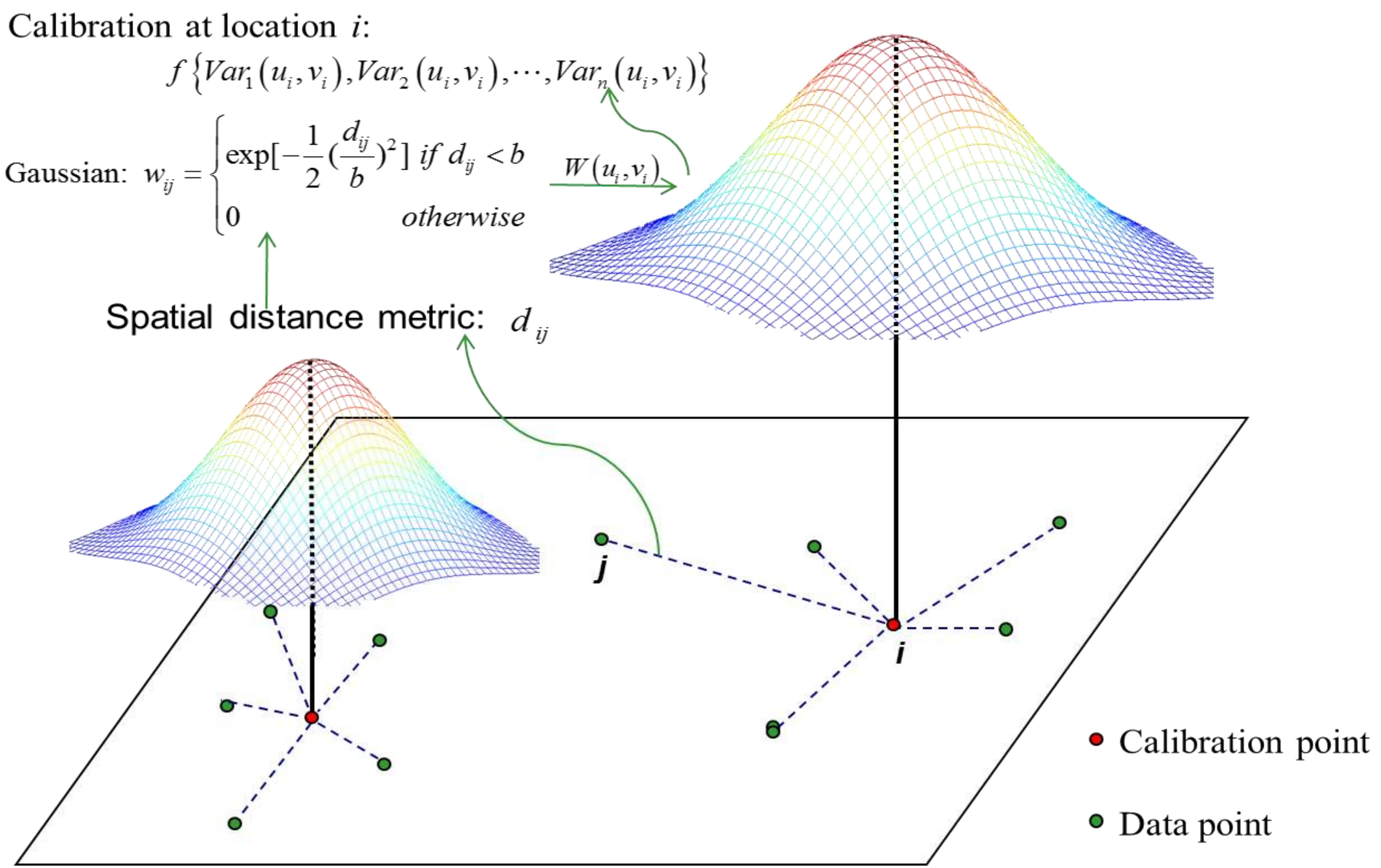
Spatial interaction

Quantitative

Distance-decaying

Location specific

Mappable-GIS friendly



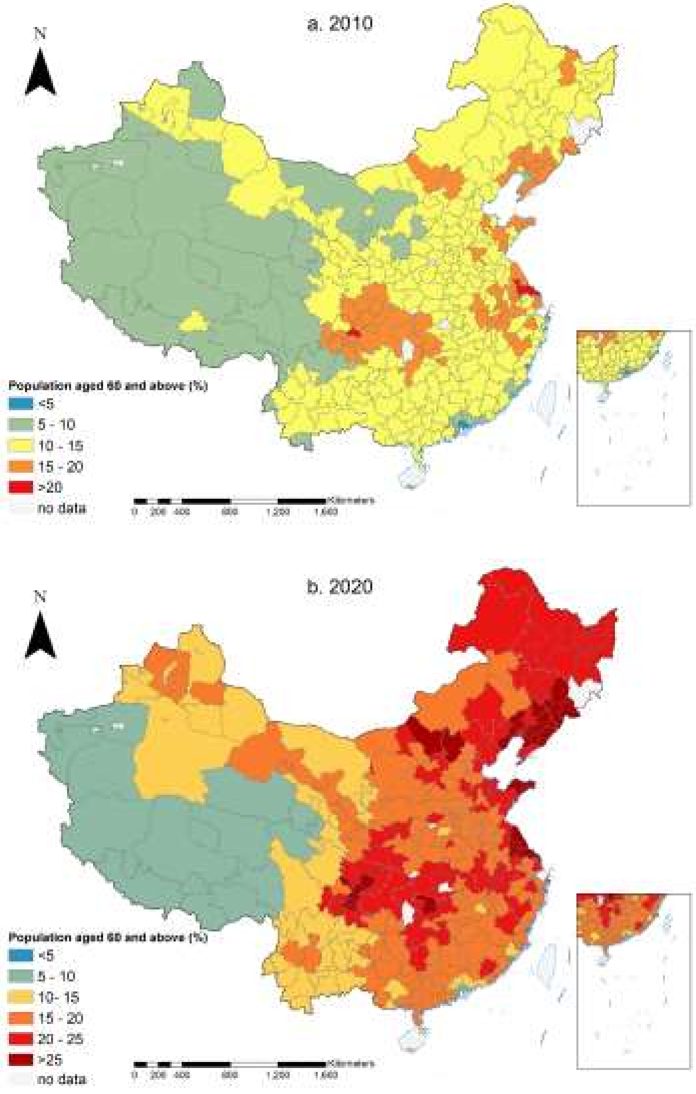
GW Summary Statistics

GW mean $$\mu(x_i)=\sum_{j=1}^{n} w_{ij}x_j \Big/ \sum_{j=1}^{n} w_{ij}$$

GW standard deviation $$s(x_i)=\sqrt{\sum_{j=1}^{n} w_{ij}\left(x_j-\mu(x_i)\right)^2 \Big/ \sum_{j=1}^{n} w_{ij}}$$

GW correlation coefficient $$\boldsymbol{\rho}(\boldsymbol{x_i},\boldsymbol{y_i})=\boldsymbol{c}(\boldsymbol{x_i},\boldsymbol{y_i})/(\boldsymbol{s}(\boldsymbol{x_i})\boldsymbol{s}(\boldsymbol{y_i}))$$

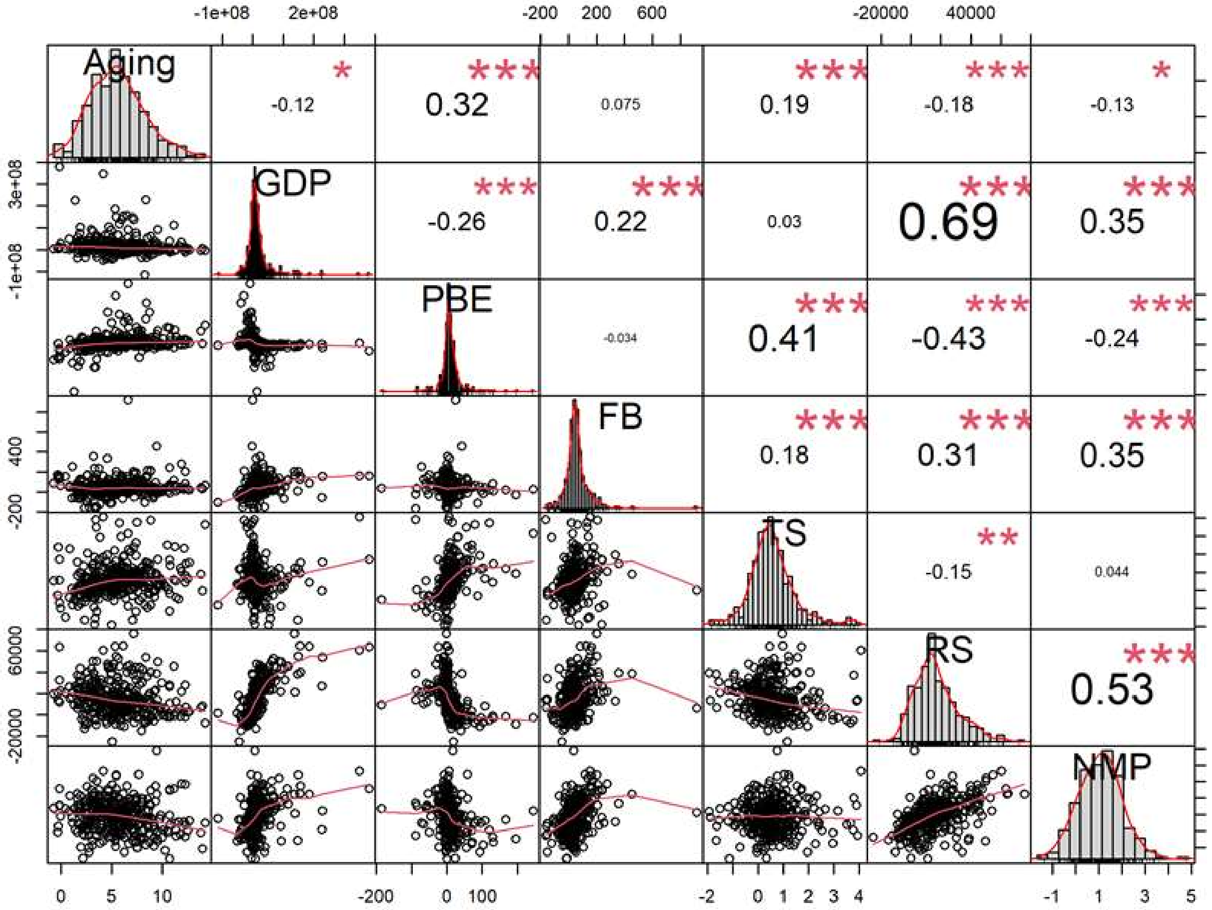
GW covariance $$c(x_i,y_i)=\sum_{j=1}^{n} w_{ij}\left\{\left(x_j-\mu(x_i)\right)\left(y_j-\mu(y_i)\right)\right\} \Big/ \sum_{j=1}^{n} w_{ij}$$

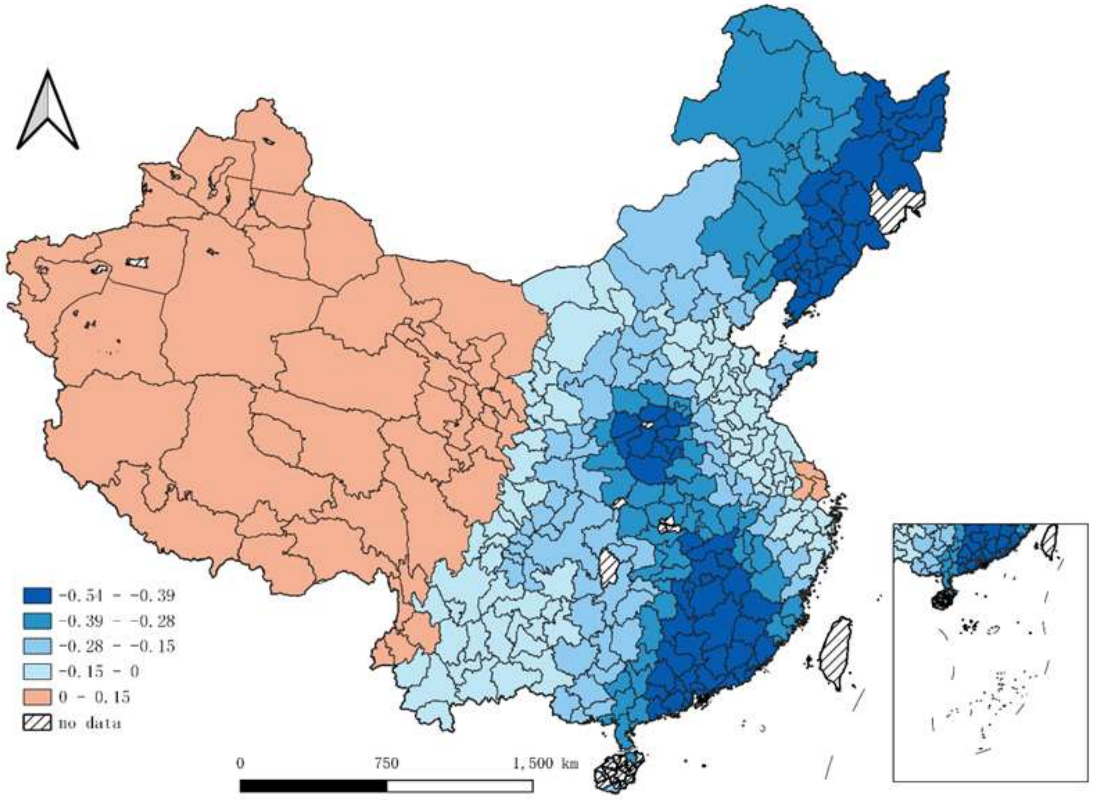


GW Summary Statistics

Variable	Description	Reference
Aging	Proportion of individuals aged over 60 in the total population (%)	Population Aging Level
GDP	Gross Domestic Product of the region (in ten thousand yuan)	Economic Development Level
PBE	Proportion of public budget expenditure in GDP (%)	Government Intervention Level
FB	Proportion of yearly financial institution loan balance in GDP (%)	Financial Development Level
TS	Ratios between the added values of the tertiary and secondary industries	Advanced Level of Industrial Structure
RS	Total retail sales per capita (Yuan)	Social Consumption Level
NMP	Number of medical practitioners per thousand people	Medical Level

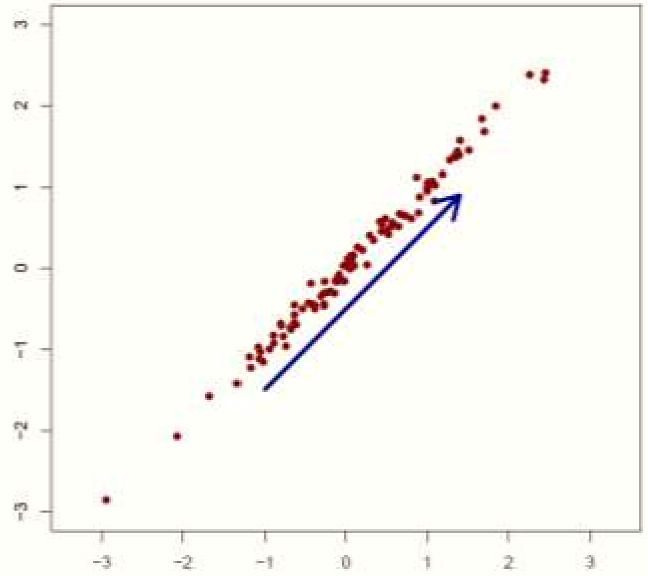
GW Summary Statistics

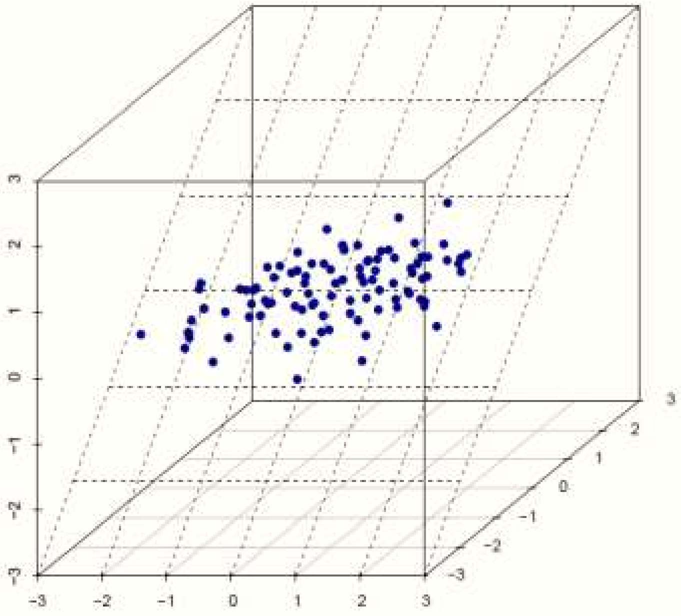




GW correlation coefficients between *Aging* and *GDP*

GW principal components analysis





- Conventional principal components analysis

$$\sum = \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \cdots & \mathrm{cov}(X_1, X_m) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X_m, X_1) & \cdots & \mathrm{cov}(X_m, X_m) \end{bmatrix}$$

- Global variance-covariance matrix $LVL^T = \sum$

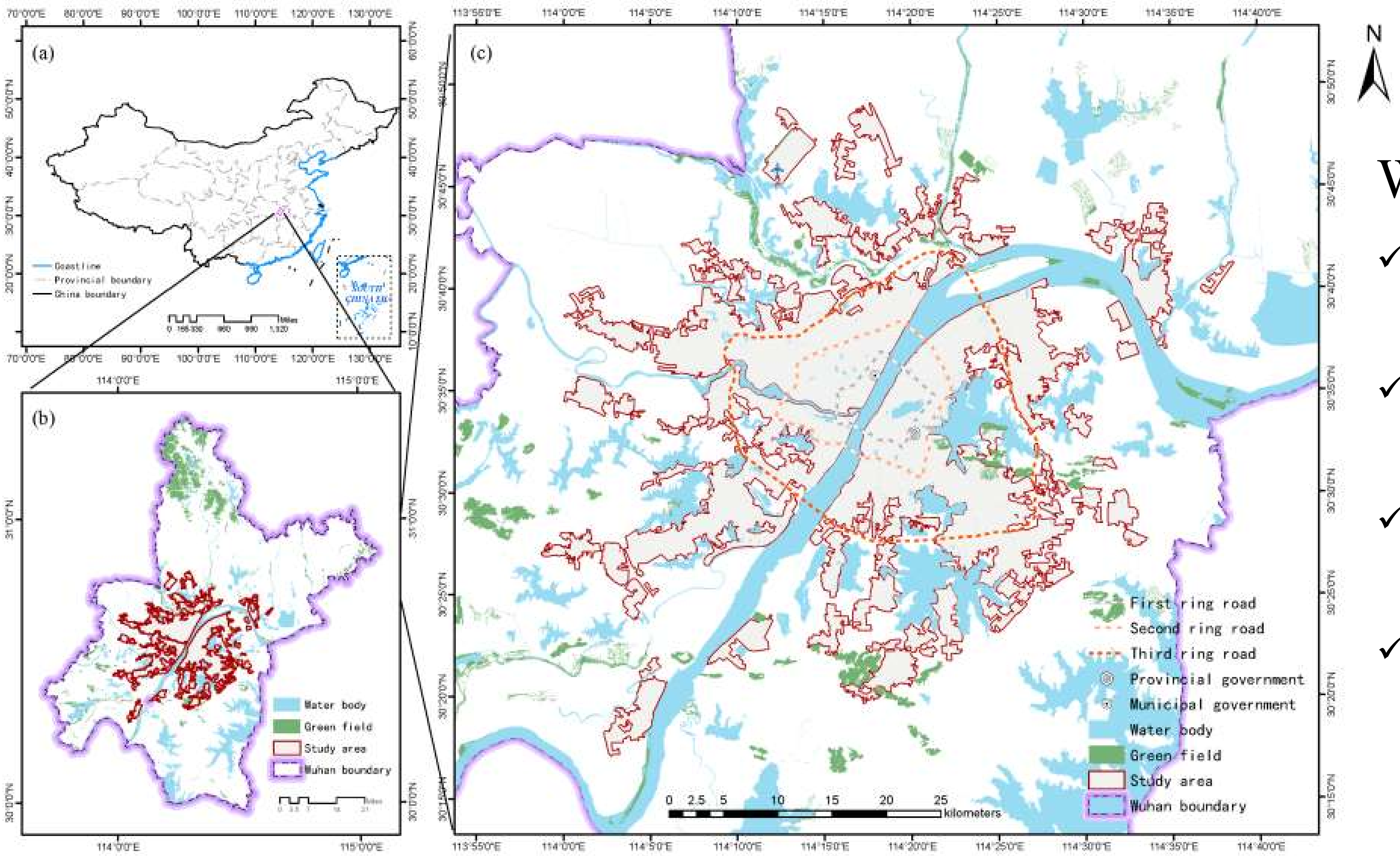
- Local variance-covariance matrix

$$\boldsymbol{\Sigma}(\boldsymbol{u},\boldsymbol{v}) = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{u},\boldsymbol{v})\mathbf{X}$$

- Local eigenvalues and local eigenvectors (or loading vectors)

$$\mathbf{L}(\boldsymbol{u}_j,\boldsymbol{v}_j)\mathbf{V}(\boldsymbol{u}_j,\boldsymbol{v}_j)\mathbf{L}(\boldsymbol{u}_j,\boldsymbol{v}_j)^{\mathrm{T}} = \boldsymbol{\Sigma}(\boldsymbol{u}_j,\boldsymbol{v}_j)$$

GW principal components analysis



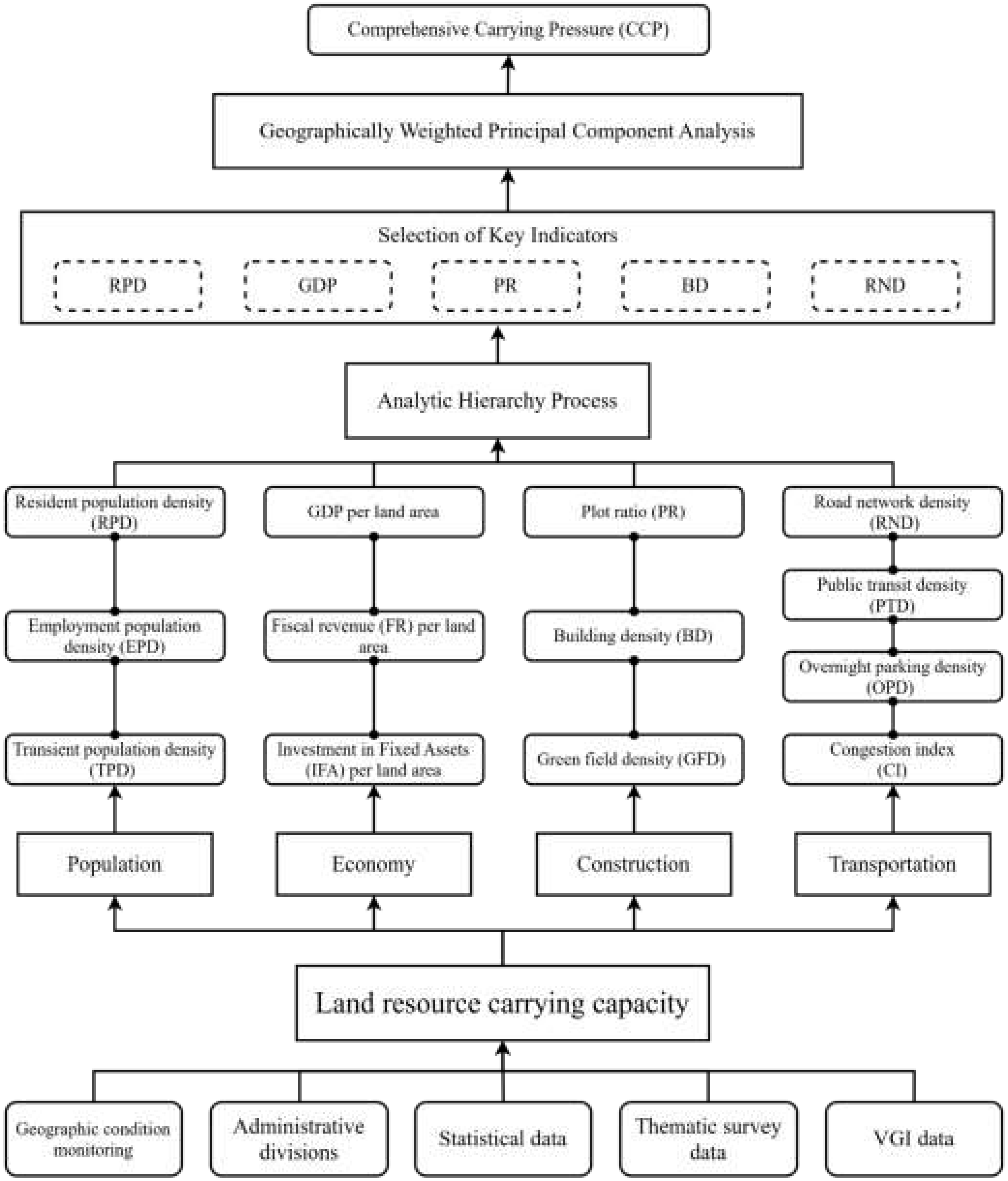
Wuhan

- ✓ the capital city of Hubei province
- ✓ the **largest city** in central China
- ✓ a total area of 8569.15 square kilometers
- ✓ a residential **population 11.12 million**

GW principal components analysis

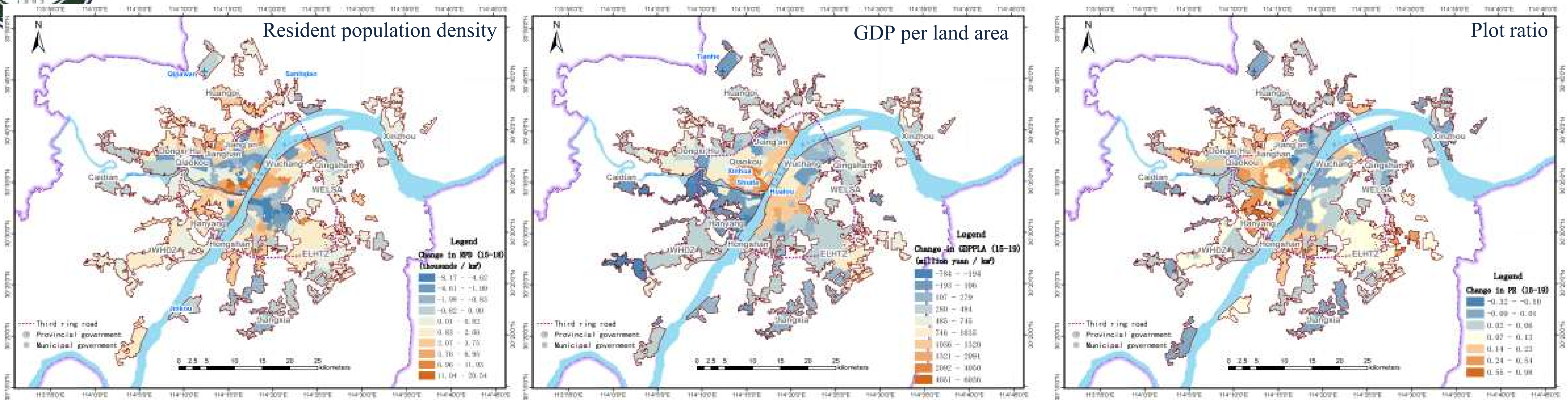
Urban land resource carrying capacity

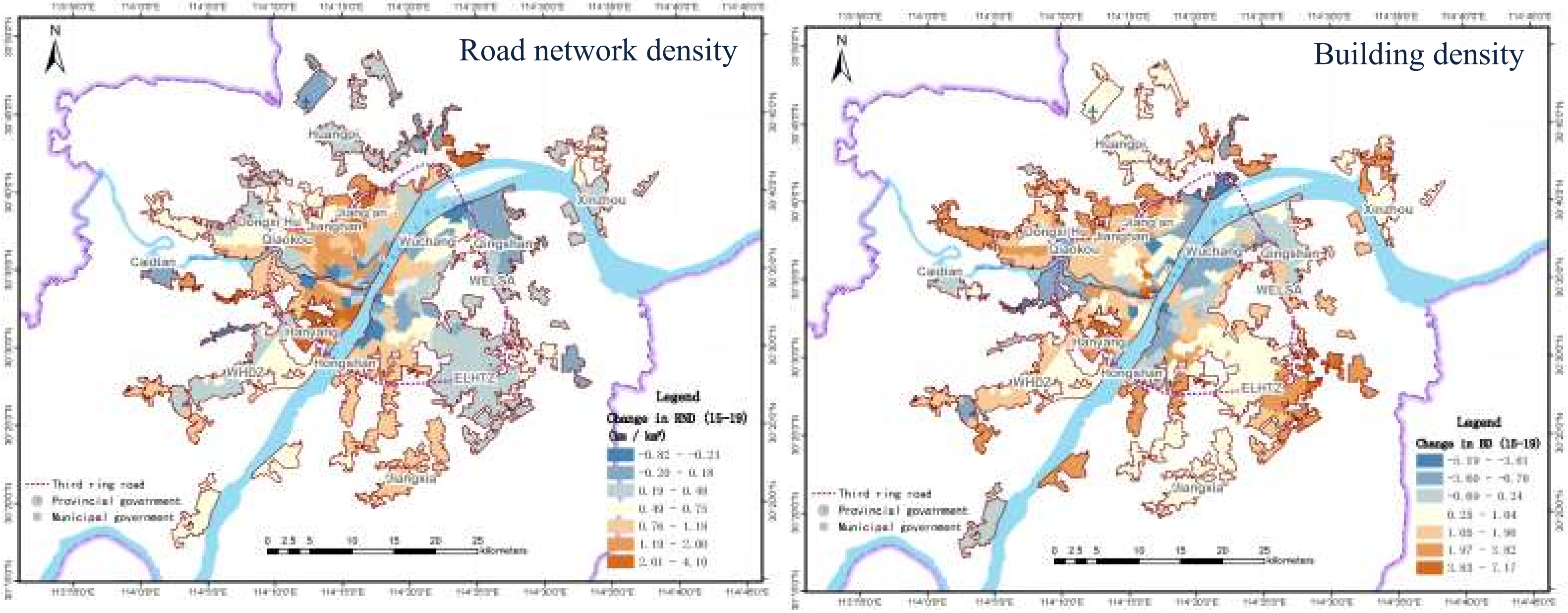
Carrying object	Variable / Indicator	Description	Unit	Start value	End value
Population	Resident population density (RPD)	RPD per square kilometer	thousands / km^2	8.676 (2015)	9.430 (2018)
	Employment population density (EPD)	EPD per square kilometer	thousands / km^2	5.920 (2017)	6.102 (2020)
	Transient population density (TPD)	TPD per square kilometer	thousands / km^2	3.31 (2017)	3.34 (2020)
Economy	GDP per land area	GDP per square kilometer	million yuan / km^2	979.74 (2015)	1452.68 (2019)
	Fiscal revenue (FR) per land area	FR per square kilometer	million yuan / km^2	198.99 (2015)	274.71 (2019)
	Investment in Fixed Assets (IFA) per land area	IFA per square kilometer	million yuan / km^2	347.55 (2015)	347.85 (2019)
Construction	Plot ratio (PR)	The proportion of the building area		0.823 (2015)	0.903 (2019)
	Building density (BD)	The proportion of the building base area	%	19.88 (2015)	20.93 (2019)
	Green field density (GFD)	Park and green area per square kilometer		0.25 (2015)	0.274 (2020)
Transportation	Road network density (RND)	Road network length per square kilometer	km/km^2	4.62 (2015)	5.29 (2020)
	Public transit density (PTD)	Bus line length and rail transit line length per square kilometer	km/km^2	13.34 (2015)	17.39 (2020)
	Overnight parking density (OPD)	Number of overnight parking per square kilometer	vehicles/ km^2	335 (2017)	433 (2020)
	Congestion index (CI)	The ratio of actual transit time to unimpeded transit time		1.317 (2019)	1.431 (2020)



GW principal components analysis

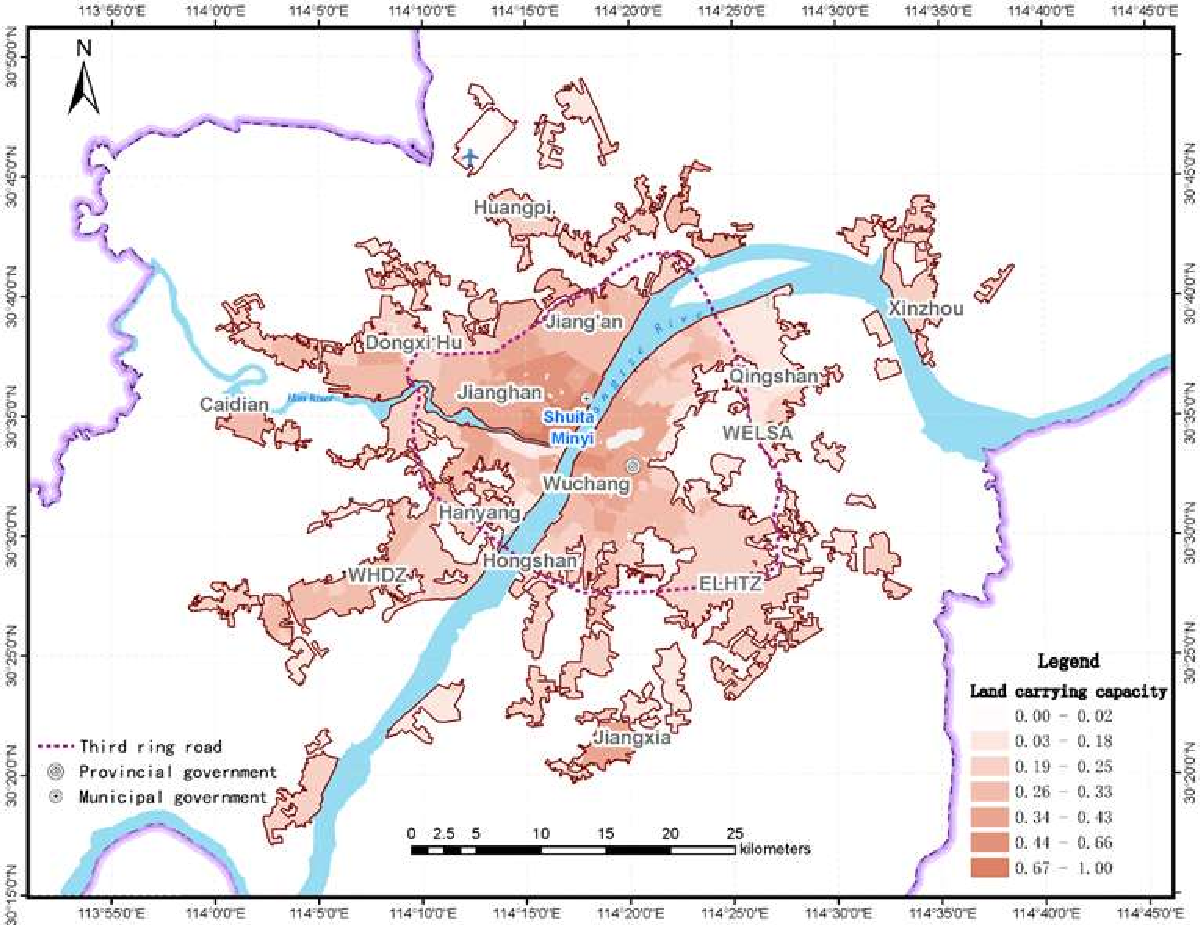
Urban land resource carrying capacity



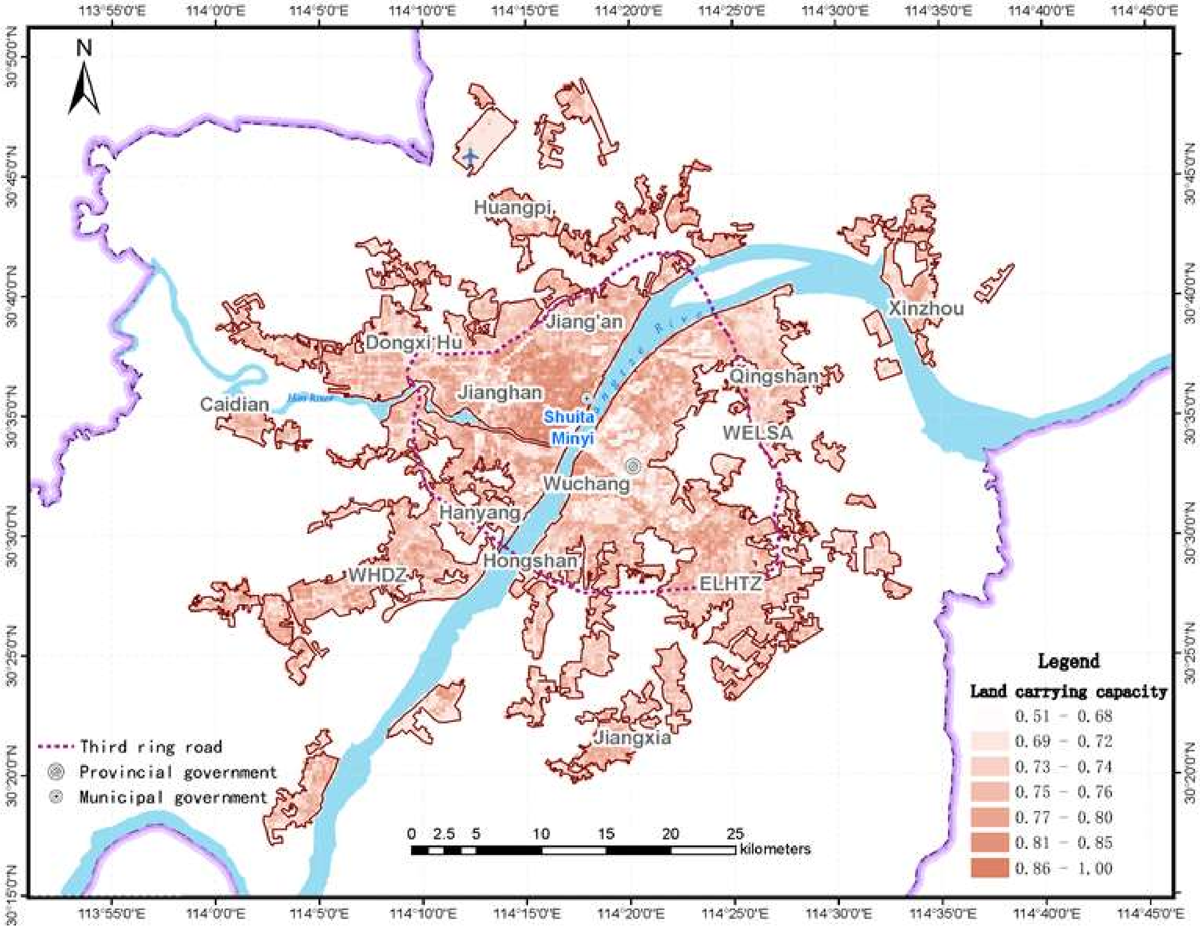


- How to evaluate a comprehensive carrying pressure (CCP):
  - ✓ traditionally **global** methods (entropy weighting)
  - ✓ urban functional heterogeneities - **Local** - **GWPCA**

GW principal components analysis



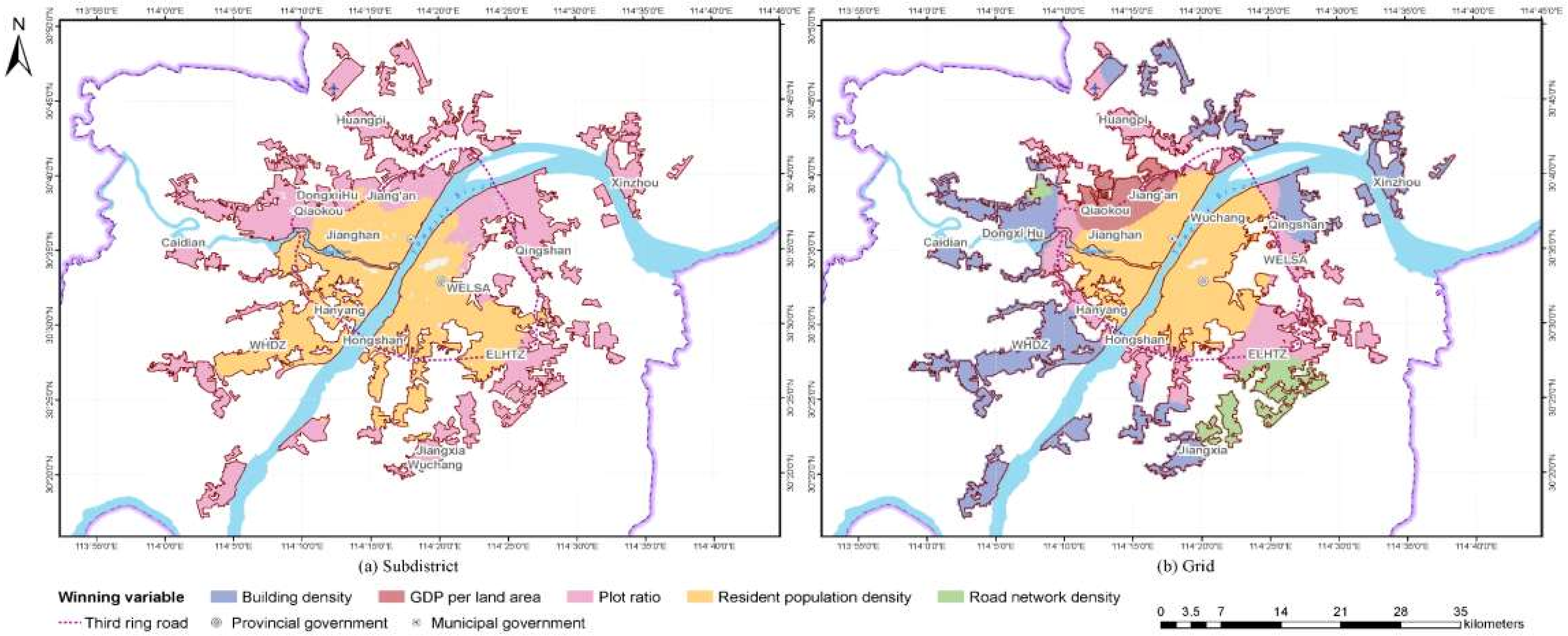
(a) Subdistrict level



(b) Grid level

Comprehensive carrying pressure (CCP) via the GWPCA scores for the first PC

GW principal components analysis



(a) Subdistrict level

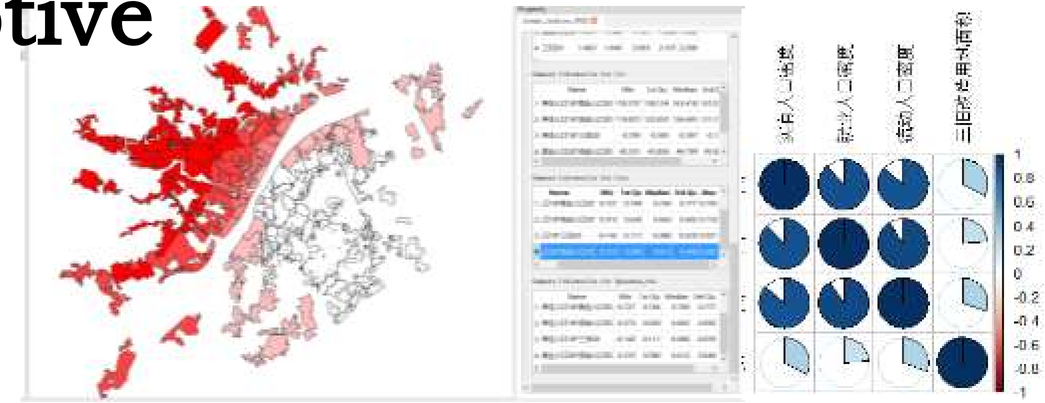
(b) Grid level

Dominant indicators of CCP via the largest eigenvalues (winning variables) from the GWPCA

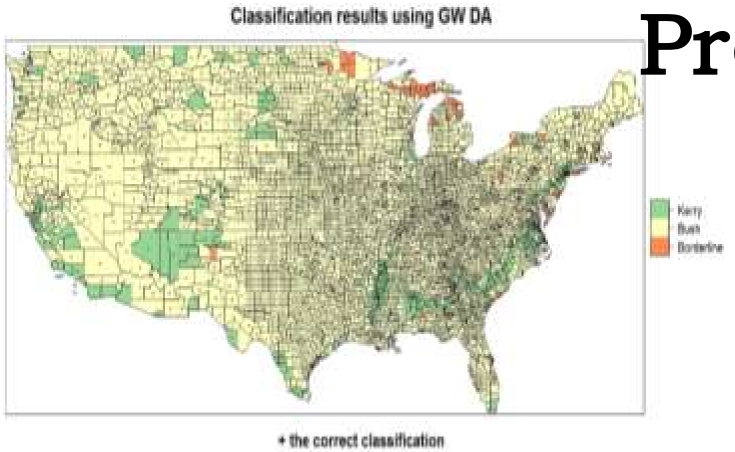
GW models: a technical framework

Descriptive

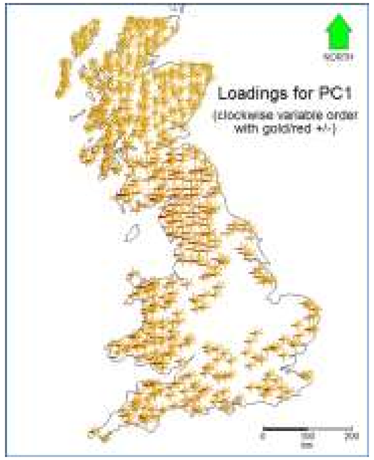
Predictive



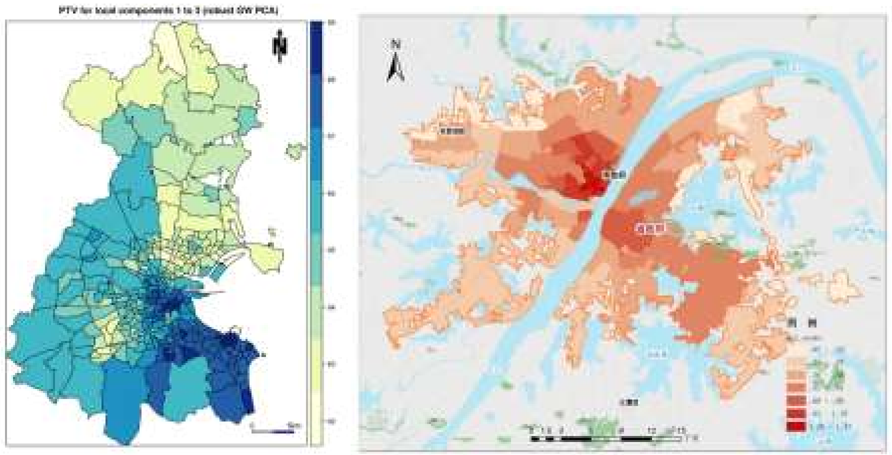
GW Summary Statistics



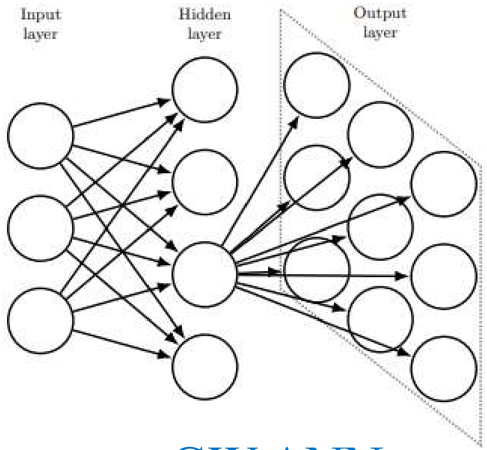
GW Discriminant Analysis



GW PCA



GWR



GW ANN

Exploratory

Interpretive

GWmodel – the R package

Package 'GWmodels'

July 13, 2012

Version 0-1

Date 2012-07-12

Title Geographically weighted models

Author Binbin Lu, Isabella Gollini, Paul Harris, Martin Charlton,Chris Brunsdon

Maintainer Binbin Lu <lubinbin220@gmail.com>

Depends R (>= 2.10.0), maptools (>= 0.5-2)

Description Functions for computing geographically weighted regressions using a flexibale distance metirc

License GPL (>= 2)

URL https://sites.google.com/site/neugwr/

R topics documented:

62 versions

Package 'GWmodel'

July 29, 2024

Type Package

Version 2.3-3

Date 2024-07-29

Title Geographically-Weighted Models

Depends R (>= 3.0.0), robustbase,sp (> 1.4-0), Rcpp (>= 1.0.12)

Imports methods, sf, grDevices, spacetime,spdep,spatialreg,FNN

LinkingTo Rcpp, RcppArmadillo, RcppEigen

Suggests mvoutlier, RColorBrewer, gstat,spData

Description Techniques from a particular branch of spatial statistics,termed geographically-weighted (GW) models. GW models suit situations when data are not de-scribed well by some global model, but where there are spatial regions where a suitably lo calised calibration provides a better description. 'GWmodel' includes functions to cali-brate: GW summary statistics (Brunsdon et al., 2002)<doi:10.1016/s0198-9715(01)00009-6>, GW principal components analysis (Har-ris et al., 2011)<doi:10.1080/13658816.2011.554838>, GW discriminant analysis (Bruns-don et al., 2007)<doi:10.1111/j.1538-4632.2007.00709.x> and various forms of GW regres-sion (Brunsdon et al., 1996)<doi:10.1111/j.1538-4632.1996.tb00936.x>; some of which a vided in basic and robust (outlier resistant) forms.

Maintainer Binbin Lu <binbinlu@whu.edu.cn>

License GPL (>= 2)

Repository CRAN

URL http://gwr.nuim.ie/

NeedsCompilation yes

SystemRequirements GNU make

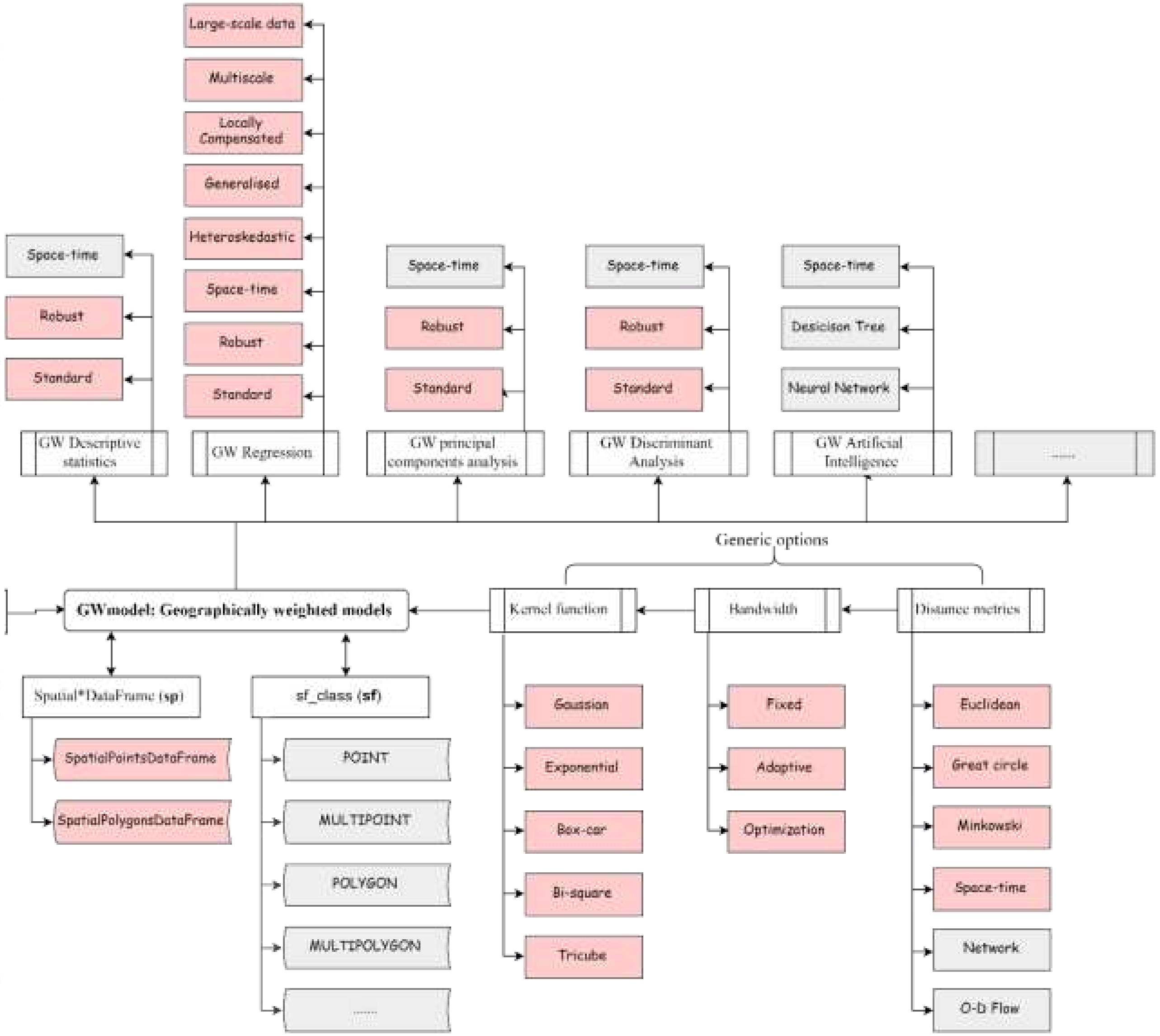
Date/Publication 2024-07-29 12:40:02 UTC

Author Binbin Lu [aut, cre],
Paul Harris [aut],
Martin Charlton [aut],
Chris Brunsdon [aut],
Tomoki Nakaya [aut],
Daisuke Murakami [ctb],

1

R topics documented:

GWmodel – the R package



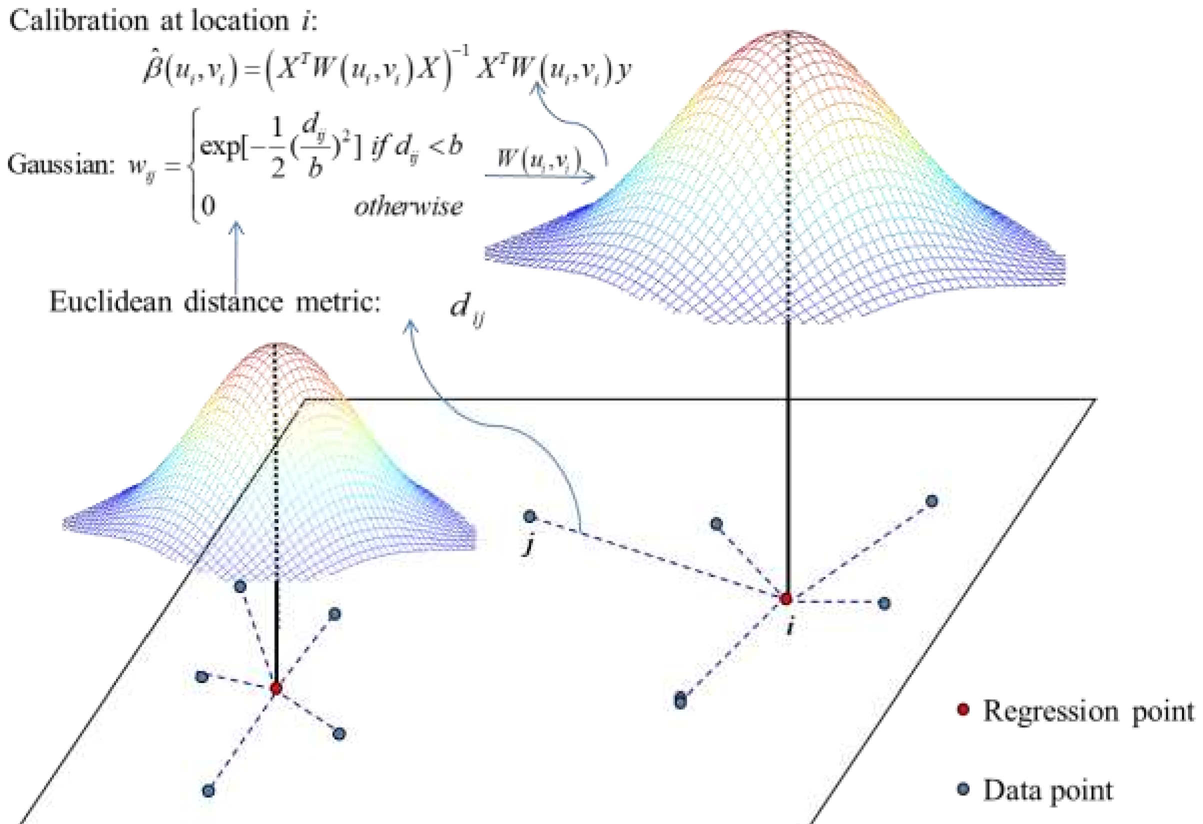
Over **160,000** downloads and installations since 2013

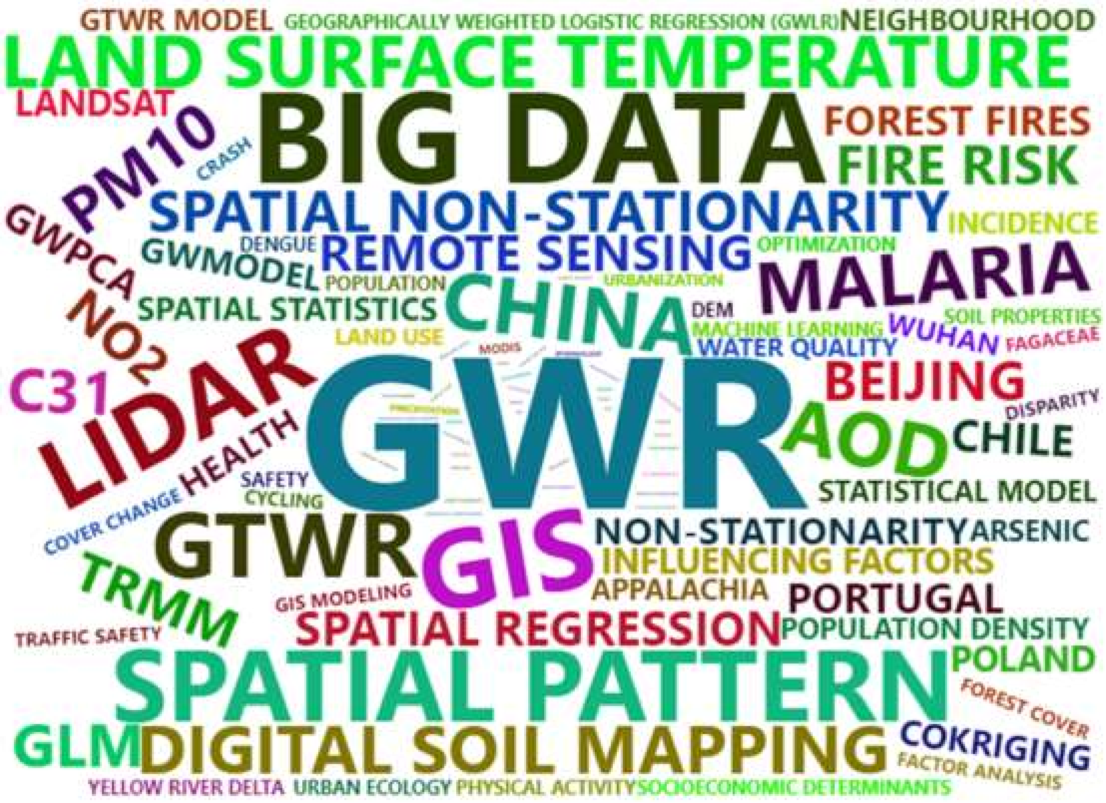
GW models-Big Models

- ✓ The original GWmodel is slow
- ✓ The storage limit in R~2GB~**16000*16000**

High complexity

O(n^2)





GWmodel: High performance GW regression

- Storage efficiency： O (n^2) -> O (nm)

$$\hat{\beta}(u_i,v_i)=\boxed{\left(X^TW(u_i,v_i)X\right)}^{-1}\boxed{X^TW(u_i,v_i)y}$$

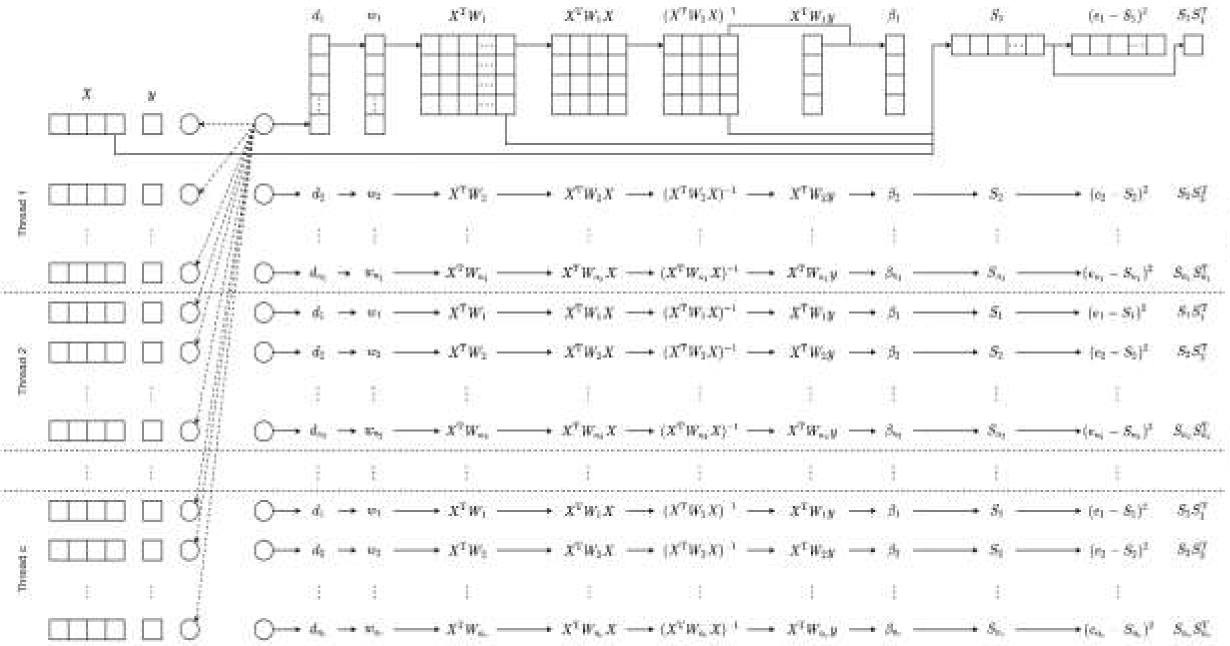
$$X^TW(u_i,v_i)X=\left(X_1^T,\cdots,X_n^T\right)\begin{pmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & w_{in} \end{pmatrix}\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}=\sum_{j=1}^{n} w_{ij}X_j^TX_j$$

$$X^TW(u_i,v_i)y=\left(X_1^T,\cdots,X_n^T\right)\begin{pmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & w_{in} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}=\sum_{j=1}^{n} w_{ij}y_jX_j^T$$

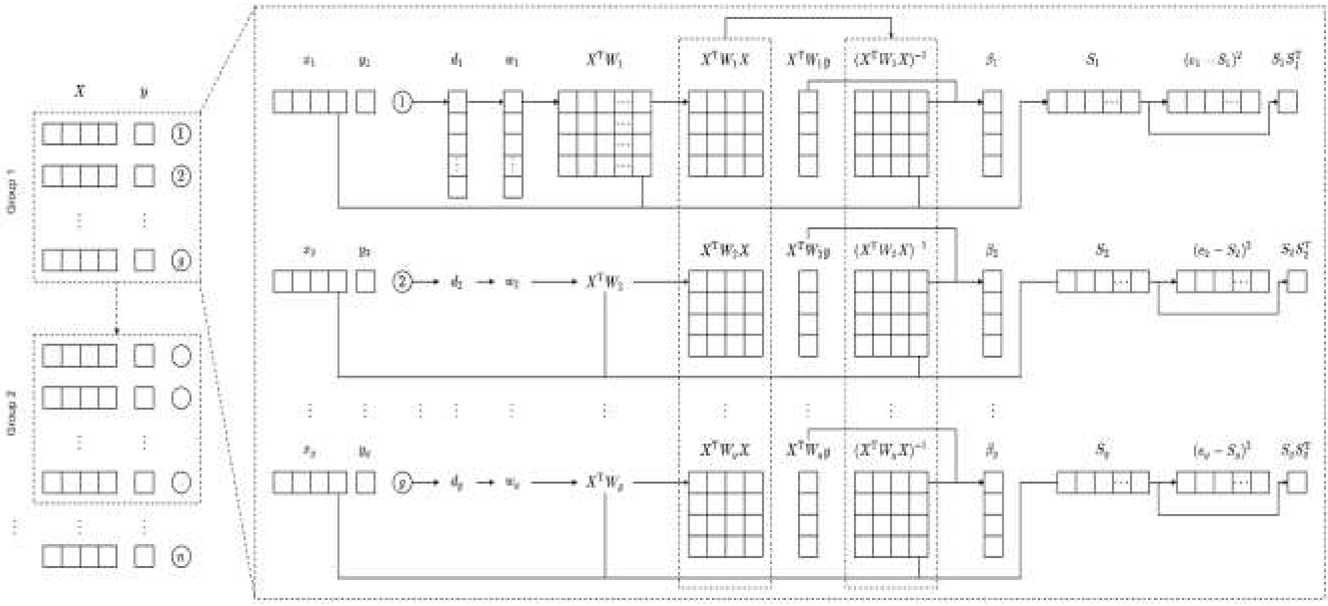
- Parallel computing：
  - ✓ **Multi-core parallel**
  - ✓ **Compute Unified Device Architecture (CUDA)**

GWmodel: High performance GW regression

➢ Parallel computing strategies



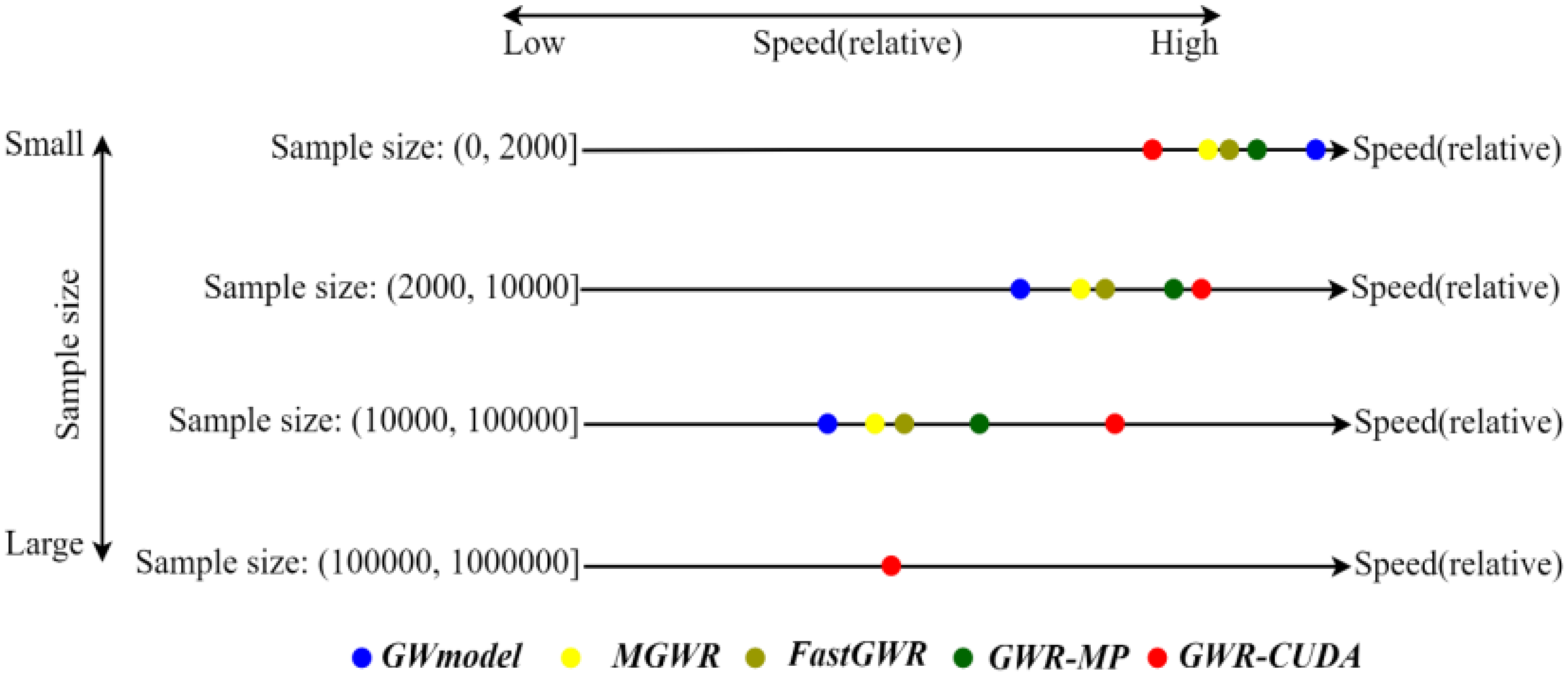
GWR-MP

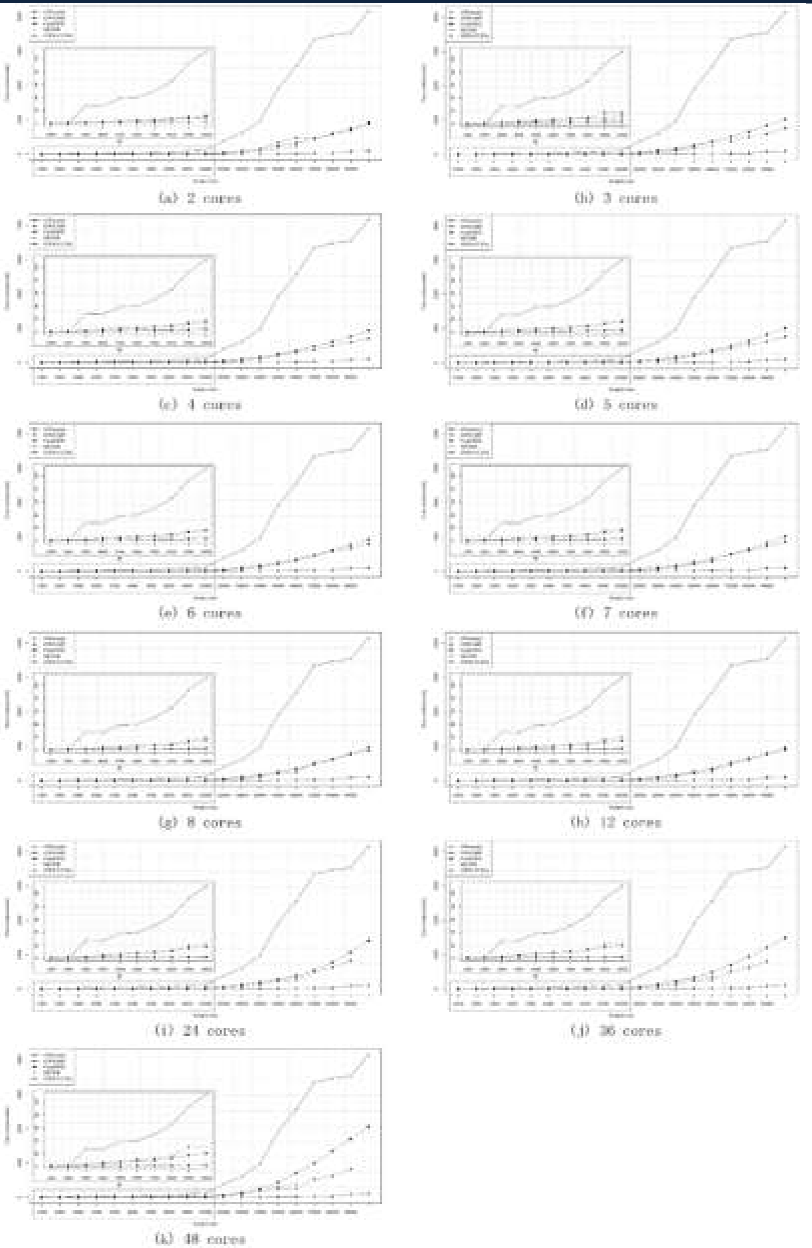


GWR-CUDA

GWmodel: High performance GW regression

- Parallel computing strategies

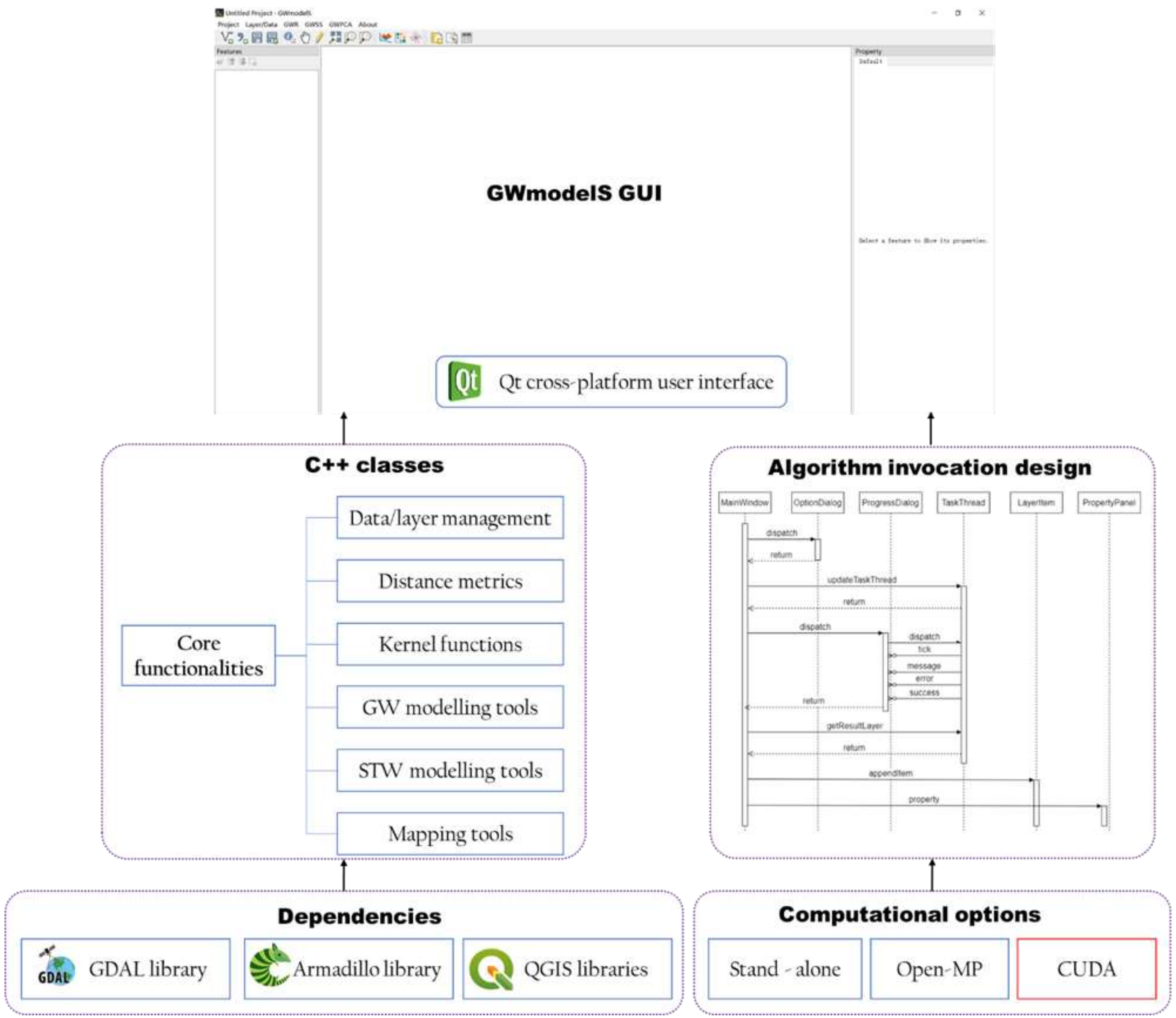


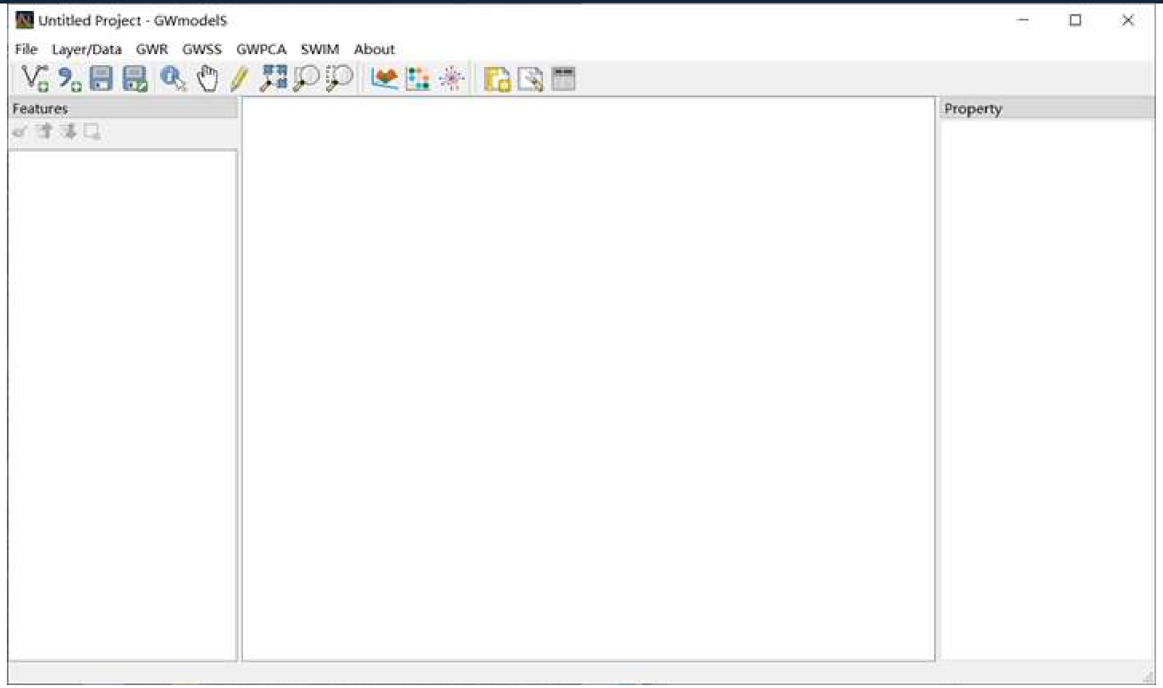


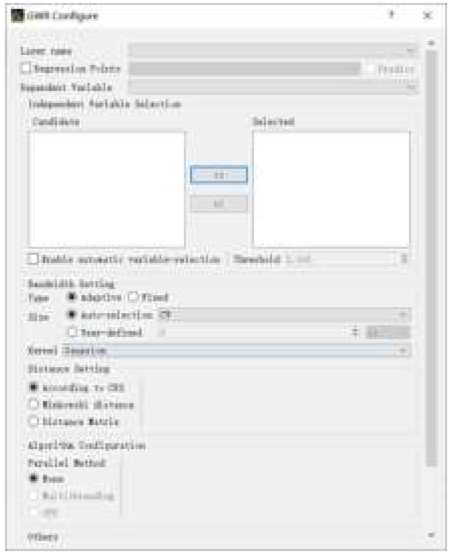
GWmodel: bottlenecks…

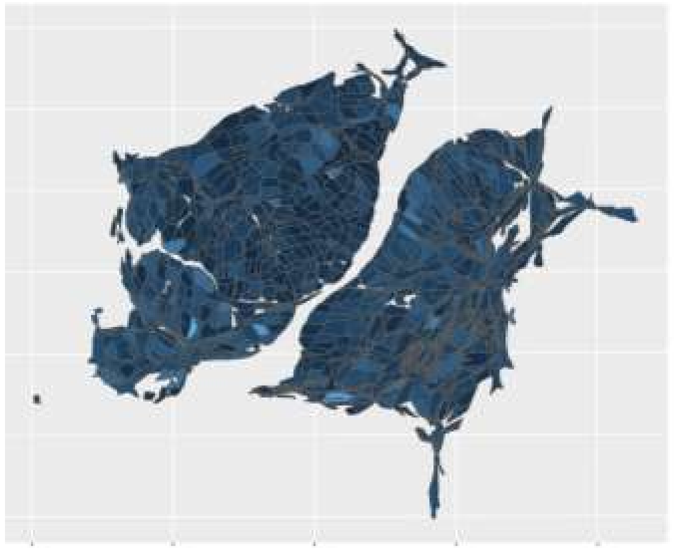
- The **GWmodel** R package is becoming popular, but,
  - ✓ **Programming skills with R are required, particualy with a manual of 85 pages in length.**
  - ✓ **Memory and computational limits explicitly exist in R, although high-performance solutions could be applied with specifically configurations.**
  - ✓ **Data processing and mapping are always required for GW models.**

GWmodelS: a standalone software

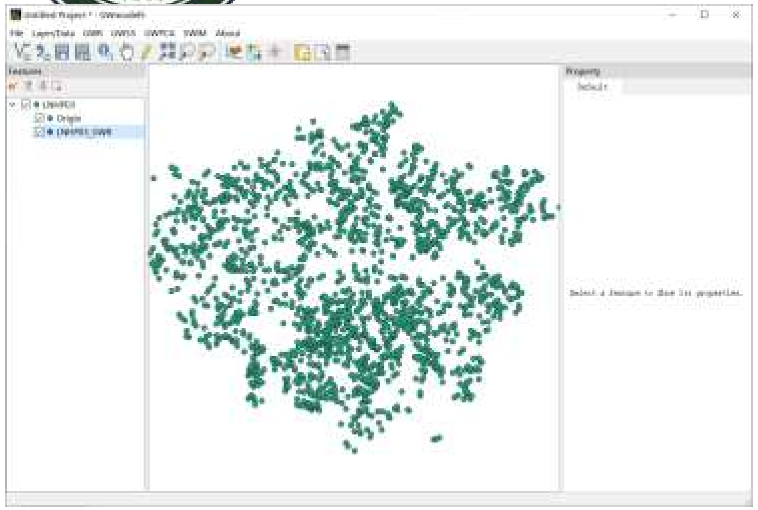




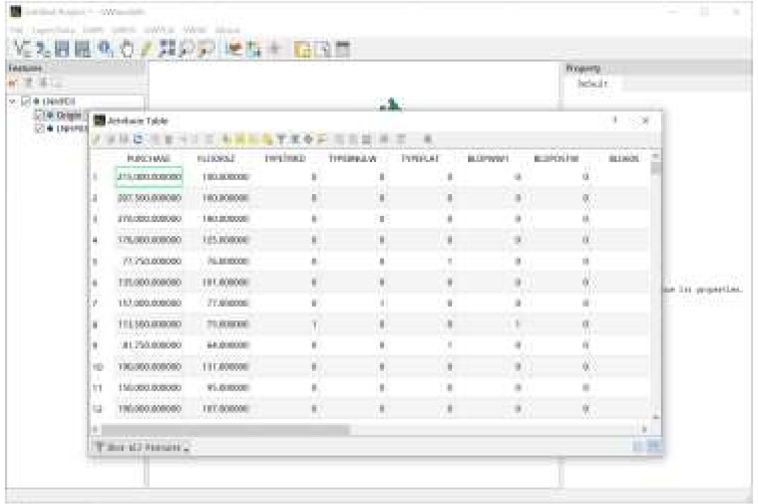




GWmodelS: a standalone software

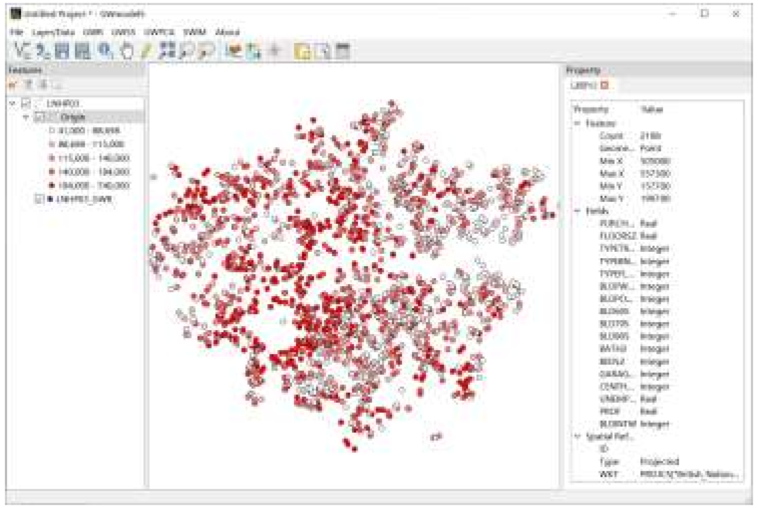


Data view

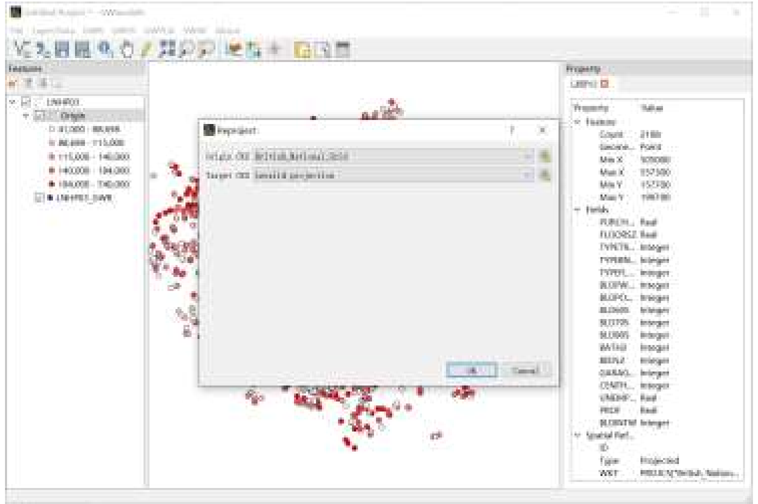


Attribute table

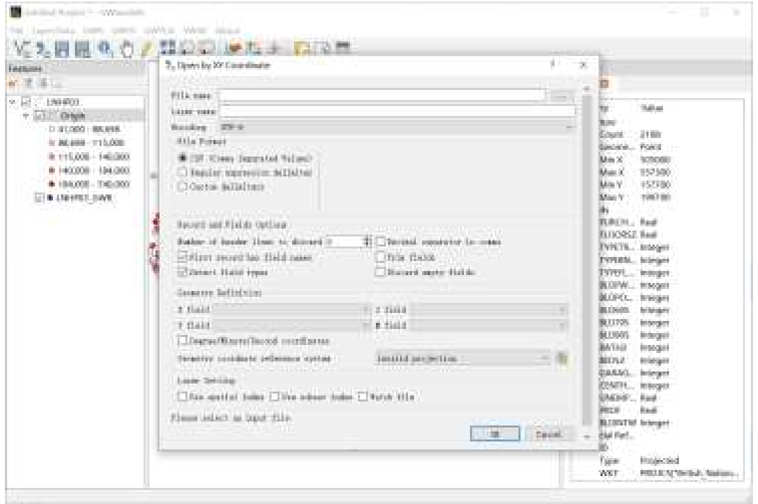
Symbols



Property of layer

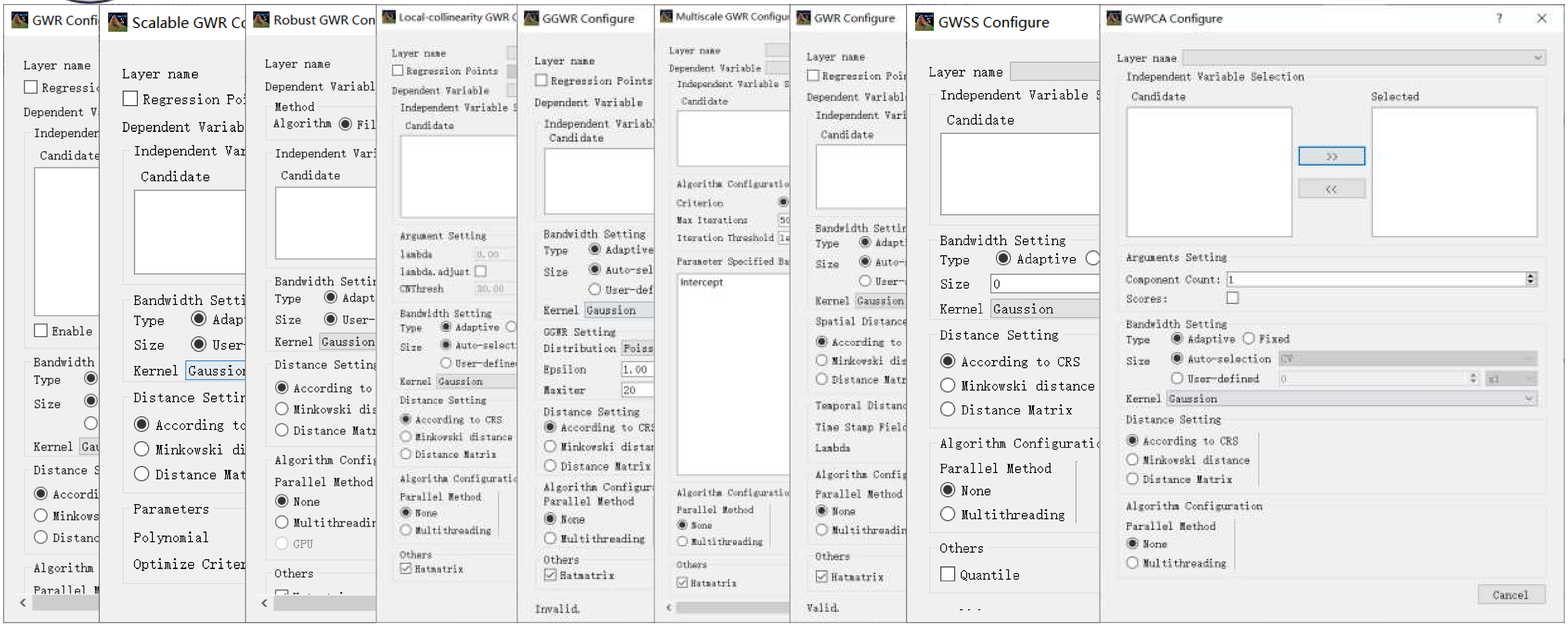


Project conversion



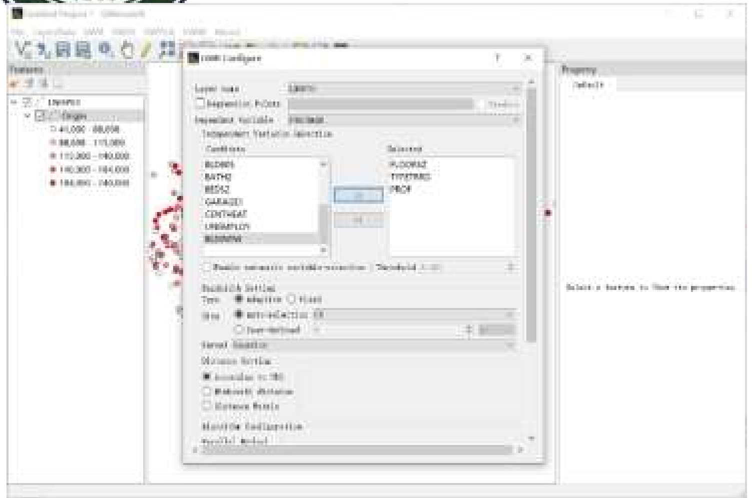
CSV file

GWmodelS: a standalone software

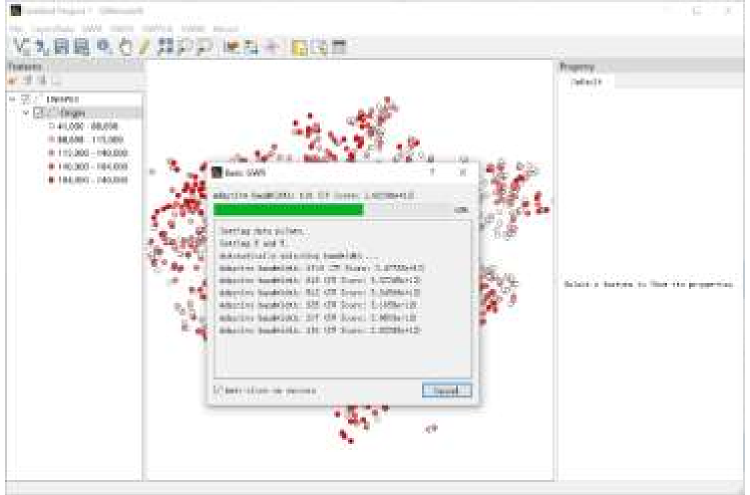


GUI for specific GW models

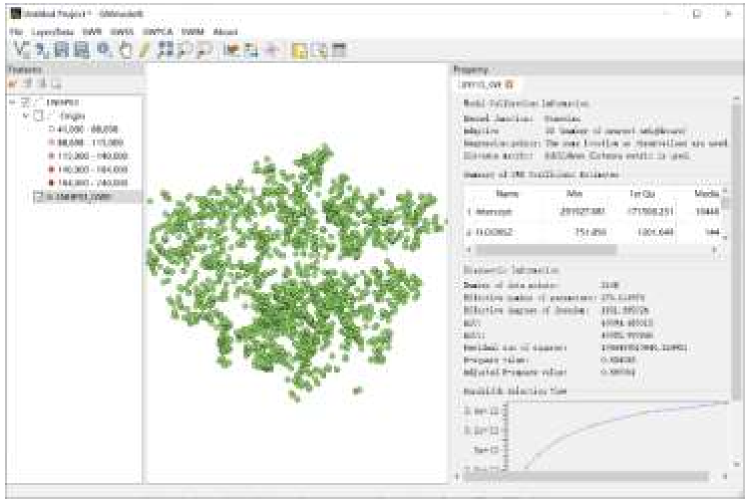
GWmodelS: a standalone software



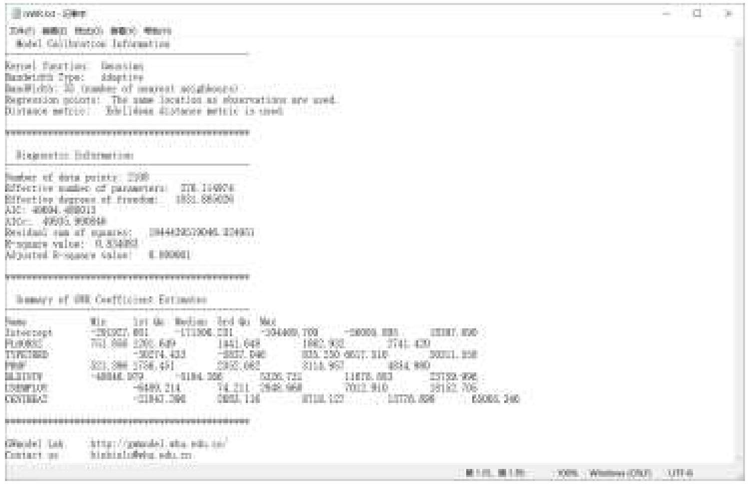
Algorithm configuration



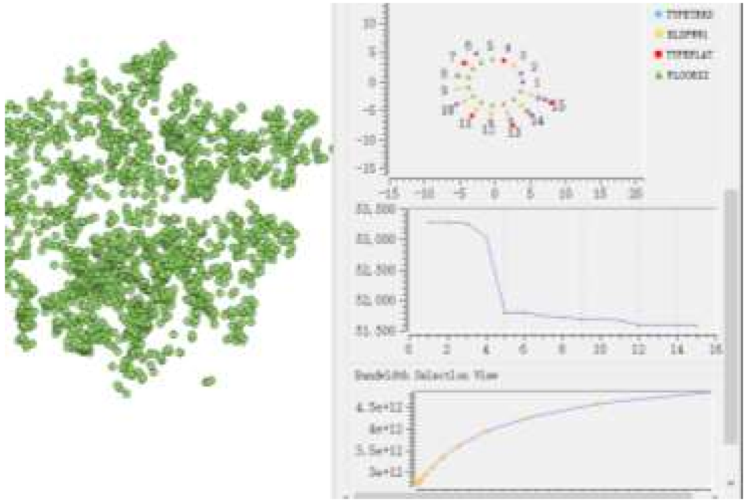
Computation



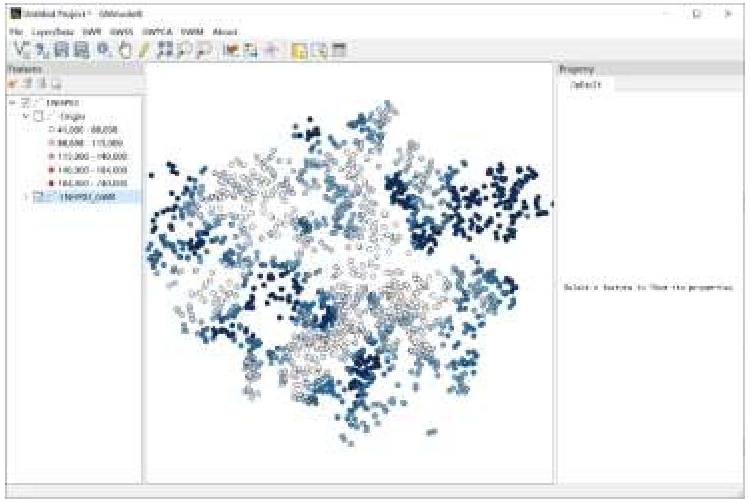
Results overview



Summary file

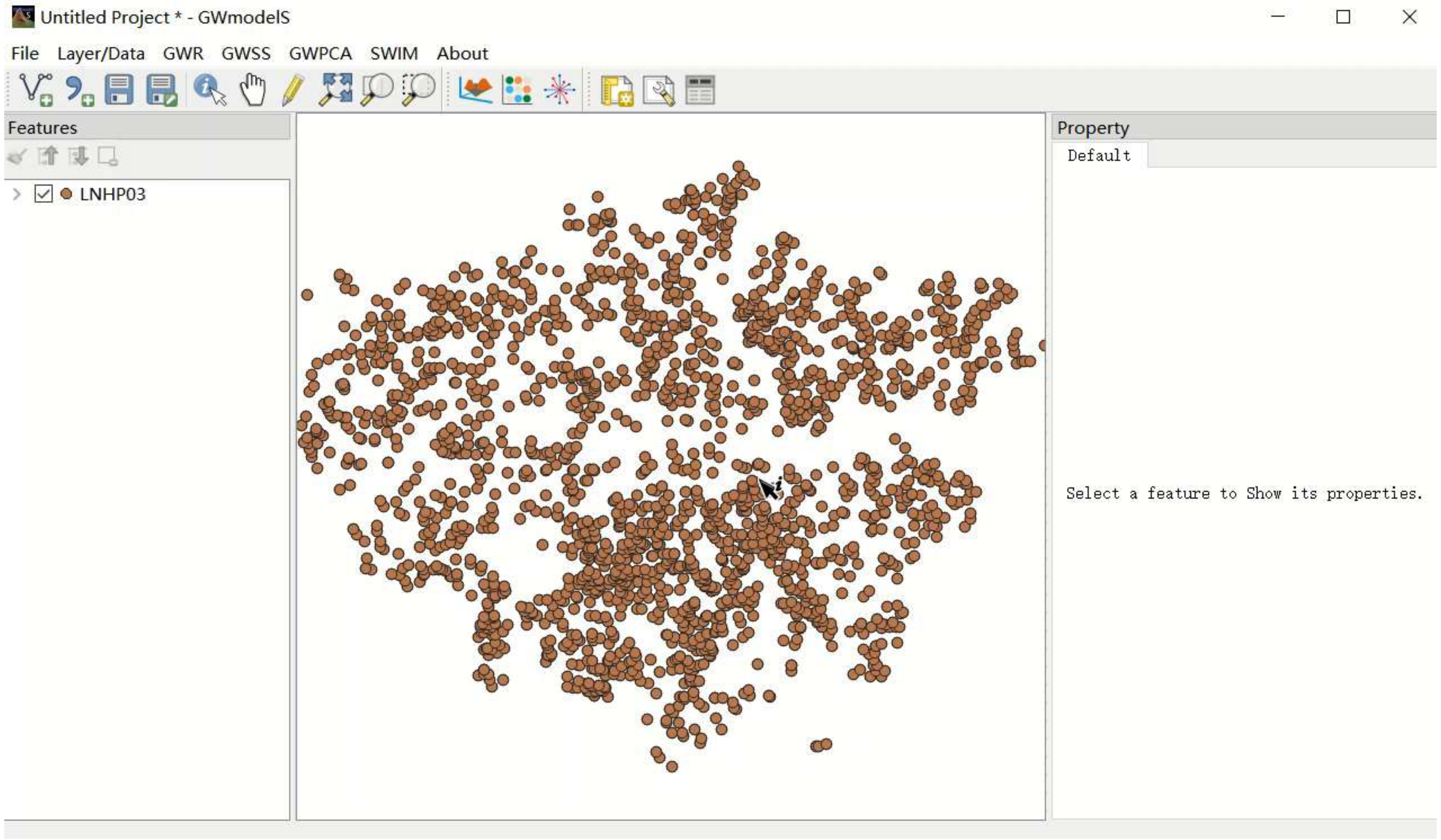


Model selection



Visualization

GWmodelS: a standalone software

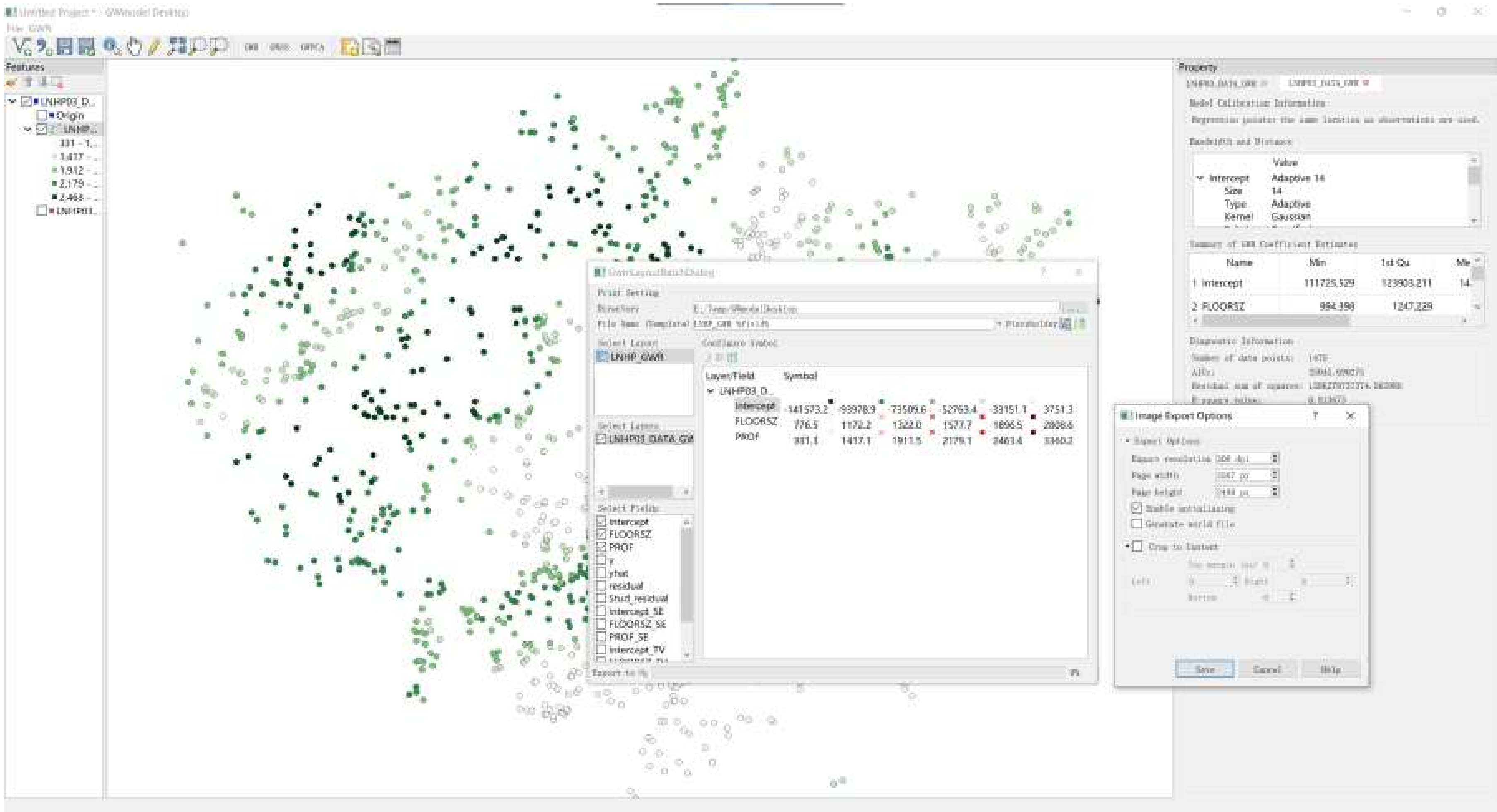


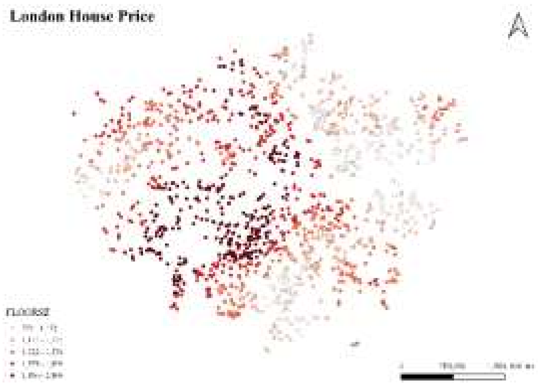
Running time comparison:

- ✓ Single thread: 15 seconds
- ✓ Multiple threads(3): 9 seconds
- ✓ How many threads could be adopted is up to the configuration of your PC

High-performance solutions are internally incorporated in **GWmodelS**, and the data volume could be up to millions

GWmodelS: a standalone software

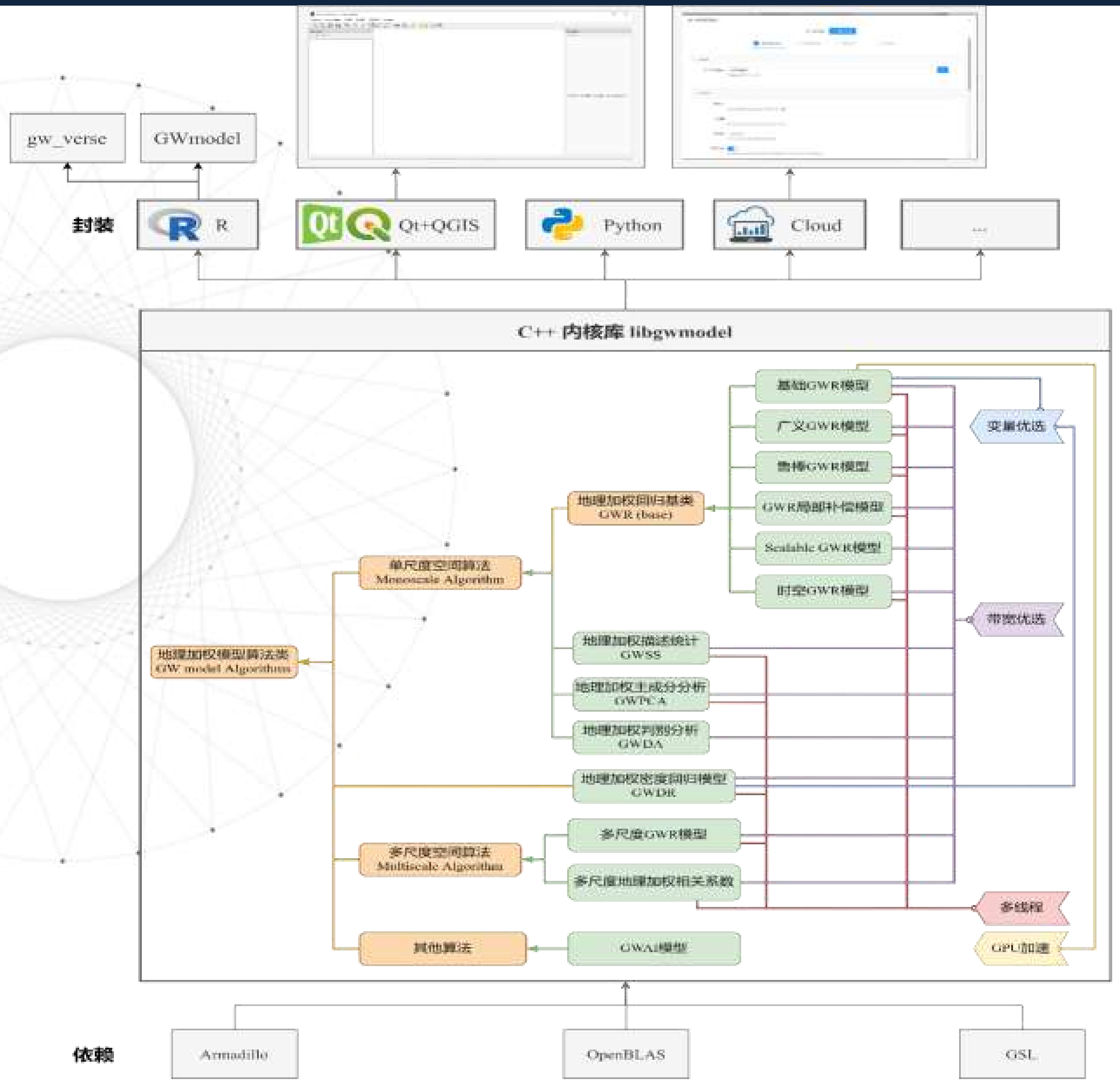




Batch mapping tool: setting all the cartographical parameters as one-off and producing all the maps accordingly

Not an end: moving forward with a new fundamental library

- **libgwmodel library**
  - ✓ **C++**
  - ✓ **Multiple usages**
  - ✓ **Update for all**
  - ✓ **Ruse of code**
  - ✓ **Cloud computing**
    - ☐ **Open Geospatial Engine**



Ending remarks

- **A family of GW models:** a comprehensively technical framework for modelling spatial heterogeneities or non-stationarities in geographic variables and their relationships
  - ✓ Descriptive, exploratory, interpretative and predictive analysis
  - ✓ Keep evolving with new GW techniques, temporally and spatio-temporally
- **GWmodel** and **GWmodelS**
  - Keep updating with new functionalities
  - New products, e.g. python package and cloud APIs with the C++ library **libgwmodel**
- More case studies…

References for this presentation

- Lu, B., Hu, Y., Yang, D., Liu, Y., Ou, G., Harris, P., Brunsdon, C., Comber, A., Dong, G., 2024. GWmodelS: A standalone software to train geographically weighted models. Geo-spatial Information Science, 1-23.
- Lu, B., Hu, Y., Yang, D., Liu, Y., Liao, L., Yin, Z., Xia, T., Dong, Z., Harris, P., Brunsdon, C., Comber, L., Dong, G., 2023. GWmodelS: A software for geographically weighted models. SoftwareX 21, 101291.
- Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., Harris, P., 2023. A route map for successful applications of geographically weighted regression. Geographical Analysis 55 (1), 155-178.
- Lu, B., Ge, Y., Shi, Y., Zheng, J., Harris, P., 2023. Uncovering drivers of community-level house price dynamics through multiscale geographically weighted regression: A case study of wuhan, china. Spatial Statistics 53, 100723.
- Hu, Y., Lu, B., Ge, Y., Dong, G., 2022. Uncovering spatial heterogeneity in real estate prices via combined hierarchical linear model and geographically weighted regression. Environment and Planning B: Urban Analytics and City Science 49 (6), 1715-1740.
- Lu, B., Hu, Y., Murakami, D., Brunsdon, C., Comber, A., Charlton, M., Harris, P., 2022. High-performance solutions of geographically weighted regression in r. Geo-spatial Information Science 25 (4), 536-549.
- Lu, B., Brunsdon, C., Charlton, M., Harris, P., 2017. Geographically weighted regression with parameter-specific distance metrics. International Journal of Geographical Information Science 31 (5), 982-998.
- Lu, B., P. Harris, M. Charlton & C. Brunsdon (2014) The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geo-spatial Information Science, 17, 85-101.
- Gollini, I., B. Lu, M. Charlton, C. Brunsdon & P. Harris (2015) GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. Journal of Statistical Software, 63, 1-50.

THANKS！ Questions?