Geographically Weighted Models

D G Rossiter

Cornell University, Soil & Crop Sciences Section

Nanjing Normal University, Geographic Sciences Department
南京师范大学地理学学院

April 7, 2022

Local vs. global

When considering **spatially-distributed attributes**, we can view these in two ways:

Global **all** spatial units are considered together

- aim: to characterize the entire population with one model (statistical summaries, regressions, . . .)

Local a **geographically-compact subset** of spatial units are considered together

- aim: to see if there is **spatial heterogeneity** within the model . . .
- . . . and if so, at which **scale**
- general term: **Geographically-weighted** (GW) models

Global vs. local – example

- Closely related to the Modifiable Area Unit Problem (MAUP)
- Example: **Summary statistics** at different resolutions
  - MAUP: nation, state, county, town, ward . . . proportion of votes per candidate
  - GW models: proportion of different soil types over the entire map vs. sub-maps; e.g., northern vs. southern Tompkins County
- Example: Empirical-statistical models example: **regression on covariates**
  - MAUP: regression model of votes vs. demography
  - GW models: relation of soil properties to covariates (elevation, slope, . . .)

Main purpose of local models

Why build **local** models?

- **Detect** whether there is **spatial heterogeneity** in what is being studied
- **Detect** the **spatial scale** of this heterogeneity
- From these, **explain** why

Approaches to local models

Strata Divide area into (multi-)polygons according to some *a priori* stratifying factor

- soil mapping example: pre-defined Major Land Resource Areas

Moving window re-compute summaries, regressions etc. for the observations within some *window*, i.e., restricted neighbourhood

- this neighbourhood moves across the study area

Weighted moving window same, but *weight* the observations

- closer to the window centre receive more weight than further
- requires a **kernel** function defining the weight
- function of **distance** from the centre of the window

Locations of moving-window centres

Several possibilities:

1. **regular tessellation**: centres of pre-defined grids
   - e.g., 10 x 10 km grid
   - result is a model, statistics etc. for each pre-defined grid
2. at **observation points**; may be irregular
   - result is a model, statistics etc. for each observation point and its neighbourhood

Kernel functions – concept

- These define the weights to be given to observations within a window
- Model **form**: various forms of **distance** *d* **decay**, see next slide
- Parameter: **bandwidth** *h*, relation to *d*
- Can choose between model forms and select bandwidth by **cross-validation**, see next section
    - But often the model form is set by the knowledge of the target variable

Kernel functions – model forms

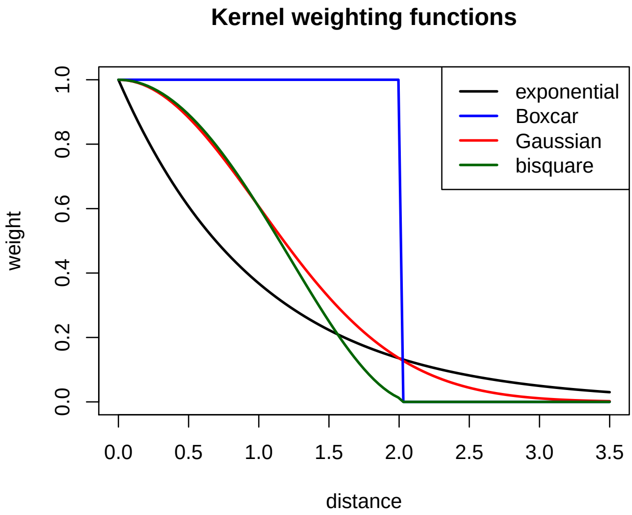
boxcar $w_{ij} = 1$ if $d_{ij} <= h$, else $w_{ij} = 0$: unweighted within a neighbourhood

bisquare $w_{ij} = (1 - (d_{ij}^2/h^2))^2$ if $d_{ij} <= h$, else $w_{ij} = 0$; inverse square within some neighbourhood

exponential $w_{ij} = e^{-d_{ij}/h}$; considers all the points, with exponentially decaying weight; reaches a weight of 0.5 at $d = -\log(0.5) \approx 0.693h$

Gaussian $w_{ij} = e^{-d_{ij}^2/2h^2}$; considers all the points, with exponentially decaying weight; reaches a weight of 0.5 at $d = h\sqrt{-2\log(0.5)} \approx 1.117h$

Kernel functions compared



How “local” is local?

- Obviously, we do not want to fit too **narrowly**, because:
    - not enough sample points to reliably calibrate a model;
    - artificial local variability, *not* corresponding to the **process**.
- But we do not want to fit too **broadly**, because this would miss “true” local variability

This is the **bandwidth problem** – it should correspond to the **process** which varies locally.

Bandwidth vs. weights

- the **bandwidth** h parameter in the kernel functions determines the range of influence of points in the regression . . .
- . . . their **relative weights** is determined by the kernel function

Fixed vs. adaptive bandwidths

The bandwidth can vary across the map or not:

fixed
: as the **distance parameter** *h* in the above formulations
  - This corresponds to a process with a fixed dependence on distance

adaptive
: a **proportion** of the points to use for each local fit
  - This is appropriate if points are irregularly spread – it ensures that there are enough points to calibrate the regression.
  - It also mitigates edge effects with fewer points

Figure 13.2 A fixed spatial weighting function.

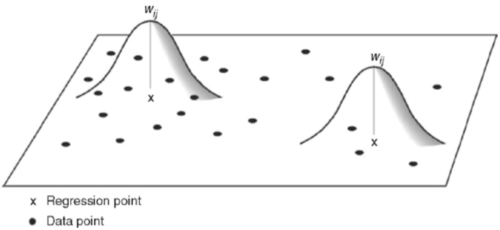
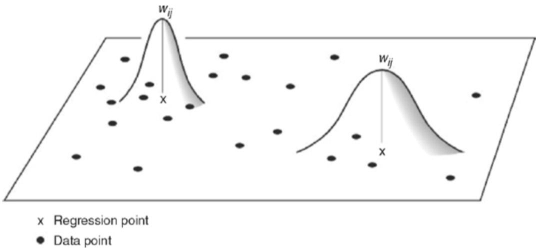


Figure 13.3 A spatially adaptive weighting function.



source: [2]

Geographically-weighted models

These have:

- any statistical **model form**;
- use a **weighted moving window**;
- a **kernel function** to define the neighbourhood;
- defined **centres**, either on on each observation point *or* a set of prediction points

Geographically-weighted regression (GWR)

- developed by Fotheringham et al. [3];
- an extension of linear or generalized linear regression;
- GWR **fits** the regression equation at **each data point . . .**
  - . . . based on some **neighbourhood** and . . .
  - . . . a **weighting scheme** (kernel function).

Why use GWR?

- GWR is appropriate if the process being modelled is **spatially non-stationary.**
  - i.e., *the relation is not the same over the whole map*.
- A single global model, although representing the **overall** relation, would miss important **local variations.**
- There should be a **physical/social basis**, i.e., some reason to think there might be non-stationarity.
  - why?, and over what spatial extent? (see “bandwidth problem”)
  - GWR can detect if this is the case . . .
  - . . . but careful for **artefacts** of the method: apparent variability not corresponding to the process, just to random noise.

GWR outputs

GWR gives explicit values of:

1. the **bandwidth** within which a local regression should be fit;
   - this is determined by cross-validation
2. the **regression coefficients** at each point
3. the **variability and spatial pattern** of these.

GWR example application

Voting choices:

- e.g., percent for each political party) explained by demographic factors (income, home ownership, age . . .)

Possible model forms:

- **global** model, probably with an spatial autoregressive (SAR) model to account for local correlation
- **GWR** model: different coefficients of each predictor; different importance of predictors in different areas

Improper use of GWR

- **Prediction**
  - It is possible to predict with GWR by evaluating the local formula at each prediction point (not necessarily observation points)
  - "Please also be aware that using GWR for prediction has no good basis anywhere for anything - and the standard errors should not be given any credibility. This is not what GWR is for at all." – Roger Bivand
- Modelling
  - GWR does *not* account for **local spatial correlation** within each window
  - compare with GLS and SAR models, which do

Spatial prediction *without* GWR

- **Spatial Autoregressive** (SAR) regression models
  - account for local correlations to adjust global model coefficients, but still one model
- **Regression Kriging** (RK): the global trend is fit (multiple regression, SAR, random forests . . .) and then adjusted locally by kriging the **residuals** and adding them to the trend prediction.
  - Assumes that the global trend is correct, but affected by local factors.
- **Kriging with External Drift** (KED) in a restricted neighbourhood
  - the trend is re-fit at each prediction point according to some restricted radius;
  - the residuals from this local trend, in the same neighbourhood are at the same time kriged;
  - uses a **model of spatial dependence** (variogram of the residuals)

Global linear regression

- GWR uses the normal OLS formulation:
  - model: $y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$
  - fit from sets of known $(y_i, \mathbf{X_i})$
  - the errors ε_i are I.I.D. and not spatially-correlated
  - solution:

$$\widehat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

  - GWR does *not* use Generalized Least Squares (GLS), **no accounting for eventual spatial correlation of residuals**.
- In a **global** model, all observations participate equally in a single model.
- GWR builds a set of **local** models, one per data point
- *All* observations participate in *each* model, but **un-equally** and differently for each model

GWR

OLS but in a **moving window**:

- the model is **separately fit at each data point** with coordinates (u_i, v_i) and known values $(y_i, \mathbf{X_i})$
- $\mathbf{W}_{(u_i,v_i)}$ is a **matrix** of the **weights** of the known points to be used to fit the model for observation i
  - $\mathbf{W}_{(u_i,v_i)}$ is a **diagonal** matrix, *no correlation* between weights (compare GLS)
  - All observations are considered but some may have 0 weight
  - Weights determined by a **kernel function** (see below)
- **Solution** by OLS:

$$\widehat{\beta}_{(u_i,v_i)} = (\mathbf{X}^T\mathbf{W}_{(u_i,v_i)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_{(u_i,v_i)}\mathbf{y}$$

GWR as a special case of WLS

- GWR is a **weighted least-squares** regression (WLS);
  - WLS: weight some observations more than others in computing the regression coefficients
  - example: inverse weight by measurement variance, gives more weight to more reliable observations
- the weights are chosen to represent the neighbourhood;
- the weights change at each point

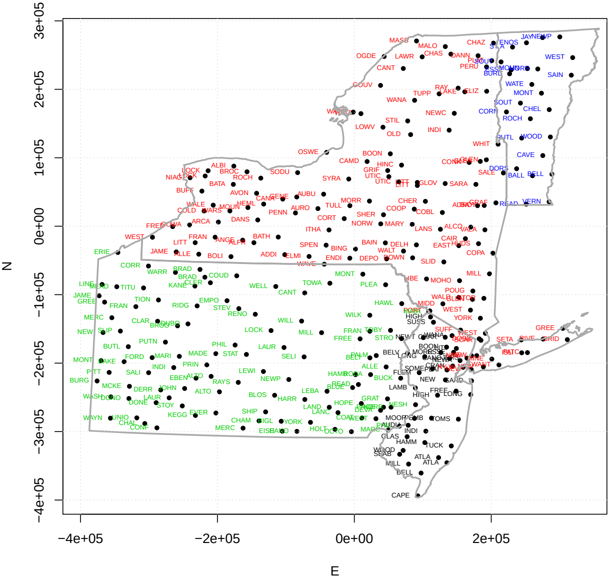
R packages

spgwr	Bivand [1]; one of the authors of the sp package
GWmodel	Gollini et al. [5]; Lu et al. [7]

GWR example – 4-state climate

- Four US States: VT, NY, NJ, PA
- 305 climate stations
- target variable: Growing Degree Days base-50° C (accumulated heat units for crop growth)
- predictors: North, East, elevation (square root)

Climate stations

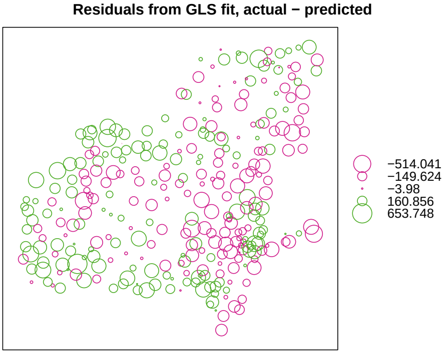


Global model

- GLS: ANN_GDD50 `~sqrt(ELEVATION_) + N`
- Fitted coefficients:

  (Intercept) 3136.37 (GDD50)
  sqrt(ELEVATION_) -3.00 (per $\sqrt{m}$)
  N -1.91 (per km)
  spatial correlation of residuals effective range $\approx$ 52 km
- adjusted $R^2 \approx 0.86$, RMSE 217 GDD_50
- Interpretation: strong **regional** effect of elevation and Northing on the annual heat units
  Easting not significant in the *global* (regional) model

GLS model residuals



Model was not equally good everywhere! And there are clear clusters of +/- residuals.

What to do about this model?

- The model is successful over the region ...
  ... but there are important local variations.
  What to do?
  1. **Krige the residuals** and add to the GLS prediction (GLS-RK)
     - This accounts for a **local** process, **within** the **regional** process
     - e.g., presence of large water bodies
  2. **GWR** to fit the model **locally**
     - Will **miss the regional variation**
     - Assumes the process is **local**
     - Maybe will better fit locally, and reveal the local importance of the three predictors
     - Does **not** account for spatial correlation of the residuals
- Question: which seems more appropriate in *this* case?

GWR model – select a bandwidth

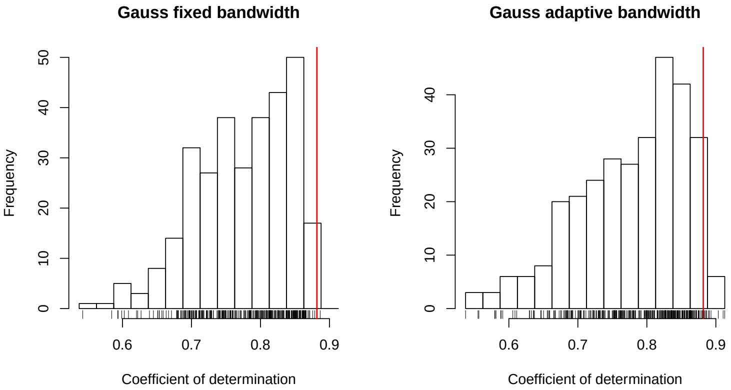
Use a Gaussian kernel; optimize by cross-validation

fixed 72.4 km

- at this radius a point receives $e^{1/2} = 0.6065$ weight.
- all points will be considered

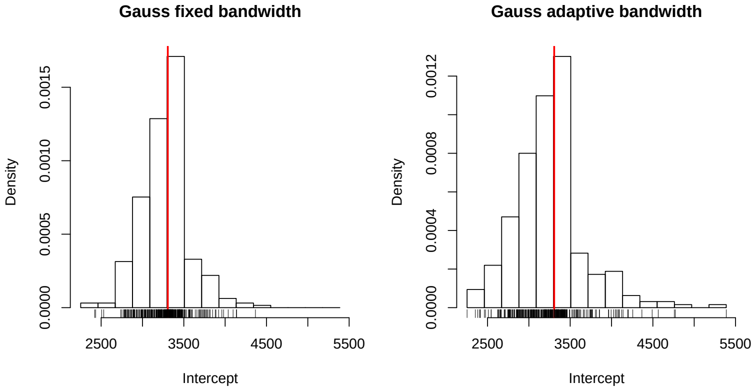
adaptive 3.35% of the stations in each window, i.e., about 10 stations for each regression

GWR model – R^2

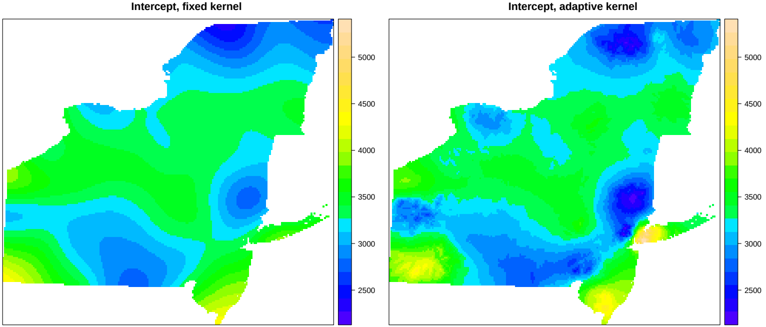


Regional value shown with red vertical line
Most **local** models have a poorer fit
Because of the restricted range of predictors in a local window

GWR model – intercepts - feature-space distribution

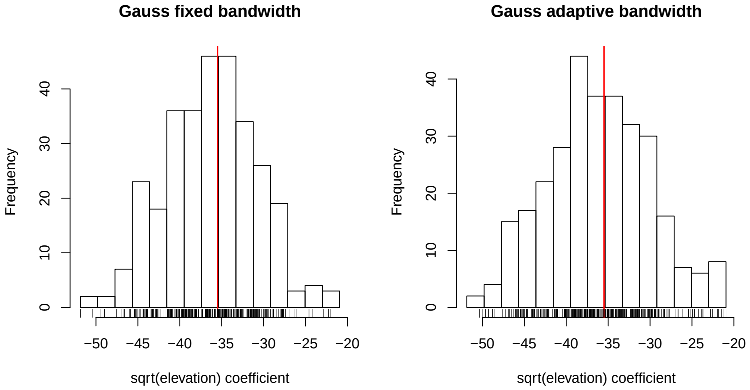


GWR model – intercepts - spatial distribution

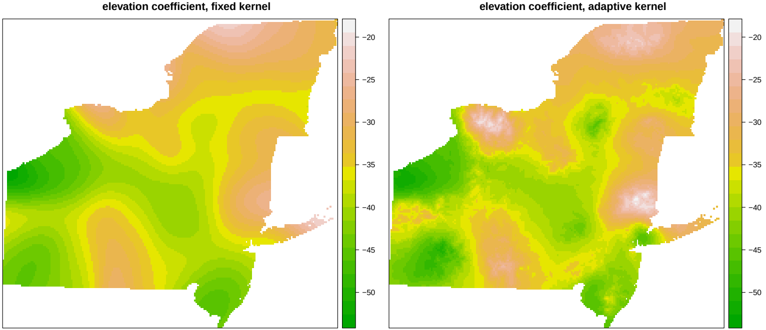


Not the average! A centering constant. Note low values in southcentral PA & the Taconics as well as northern NY/VT

GWR model – elevation - feature-space distribution

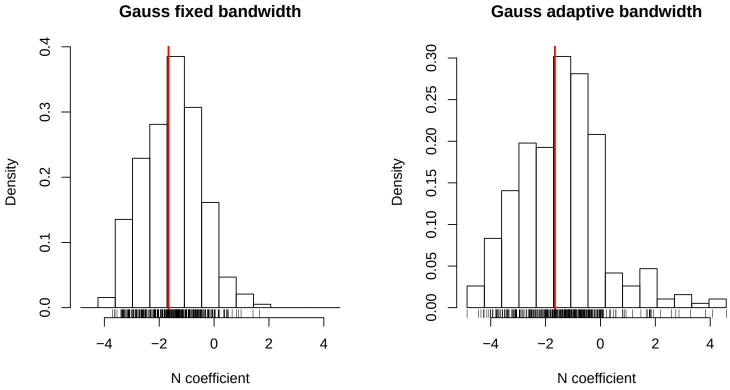


GWR model – elevation - spatial distribution

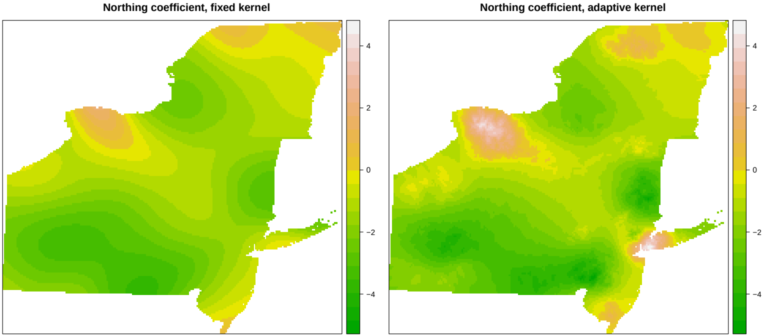


Much of this pattern seems to be an artefact of GWR
Stronger vertical GDD gradient on Lake Erie plain than Lake Ontario plain?

GWR model – Northing - feature-space distribution

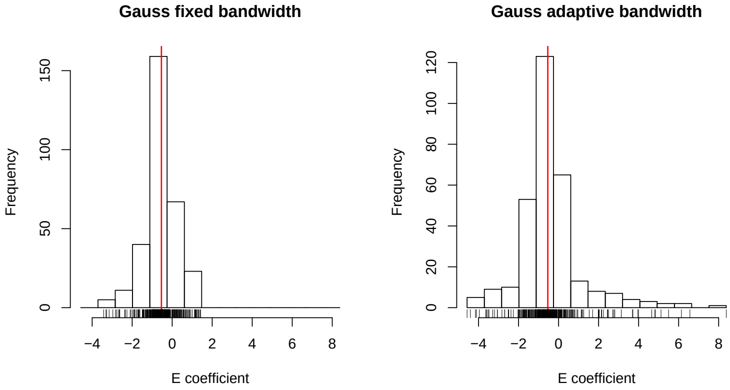


GWR model – Northing - spatial distribution

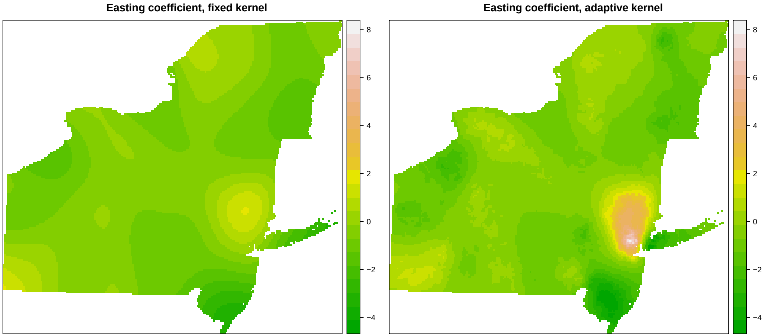


Can be locally **positive**, disagrees with physical principles

GWR model – Easting - feature-space distribution

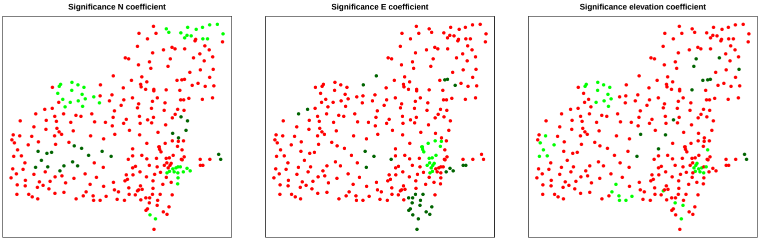


GWR model – Easting - spatial distribution



Local effect in lower Hudson valley

Significance of coefficients



red: non-significant; dark green: negative; light green: positive

Intercepts are always highly significant, i.e., $\neq 0$; they centre the local regression

Interpretation: most local models are fit only with the local average (intercept)!

Global vs. GWR model

- **Global** model finds the average effect, over the *entire region*, of the predictors
  - the physically-plausible Northing and elevation are highly significant
  - these have a **wide range** of values over the region
  - good fit, over 85% of variance explained
- **GWR** model:
  - local models with an effective radius $\approx$ 100 km
  - wide range intercepts (averages) $\rightarrow$ local means
    - this takes out most of the effect of Northing
    - some effect of Northing, Easting near water bodies
    - elevation only important in windows with significant relief
  - usually much lower R^2, less of each window is explained by factors other than the local mean
- **In this case** the GWR model is *not* justified.

Example – Georgia (USA) poverty

- Georgia (USA) counties 1990 census; originally used in [3]
- Problem: how to explain the proportion of the population in **poverty**?
- Possible **predictors**: percent of population which is:
  1. rural
  2. has a bachelor’s degreee or higher
  3. elderly
  4. foreign-born
  5. of African descent
- Practical application: if we know what is correlated with poverty (positive or negative), we can think of interventions

Global model – computation (OLS)

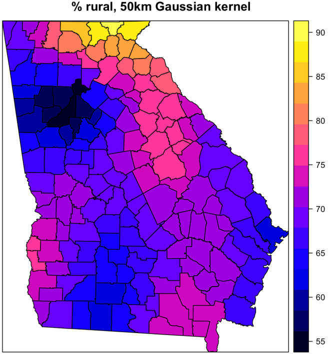
```
## lm(formula = lm.formula, data = educ.spdf@data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8282 -2.8418 -0.2404  2.6184 17.4764
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.506033   2.325226   3.228 0.001525 **
## PctRural    -0.007883   0.015780  -0.500 0.618121
## PctBach     -0.293767   0.083418  -3.522 0.000566 ***
## PctEld       0.709494   0.126583   5.605 9.46e-08 ***
## PctFB        0.148516   0.366098   0.406 0.685549
## PctBlack     0.259411   0.019638  13.210  < 2e-16 ***
##
## Multiple R-squared:  0.7078, Adjusted R-squared:  0.6982
```

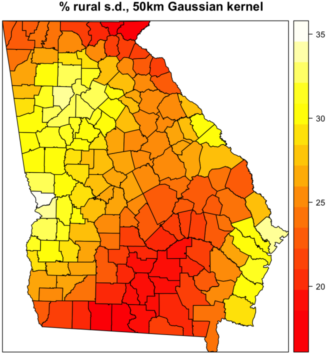
Global model – interpretation

- about 70% of the variability in poverty is explained by these factors
- The strongest predictors are **education** (moderately negative), **elderly** (strongly postive), **racial group** (moderately positive).
- Proportion of **rural** residents has almost no effect
  - but is this because we are mixing urban (Atlanta, Savannah) and rural areas?
- Proportion of **foreign-born** residents has almost no effect

Local statistics

A null model can be used to find locally-weighted statistics of a target variable; e.g., % rural





global mean 70.18

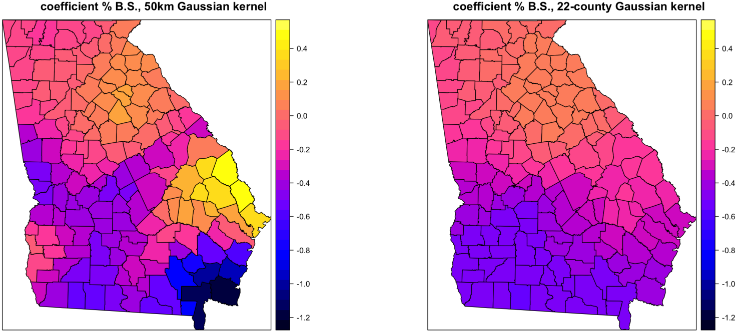
global s.d. 27.1

Note: bounding box about 443 x 514 km

Comparing kernels

- GWR depends on the choice of kernel
  1. functional form
  2. bandwidth
  3. fixed vs. adaptive
- Next slides show the difference between kernels

GWR coefficients – education

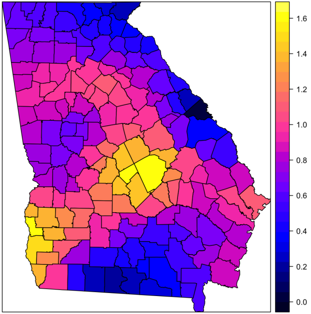


global coefficient -0.29

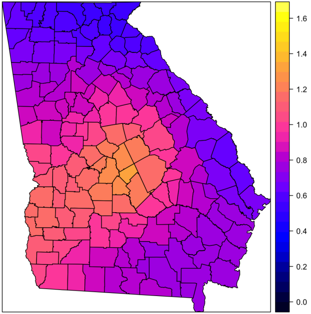
Note: education is associated with *increased* poverty in E central (Athens – University of Georgia)

GWR coefficients – % elderly

coefficient % Elderly, 50km Gaussian kernel

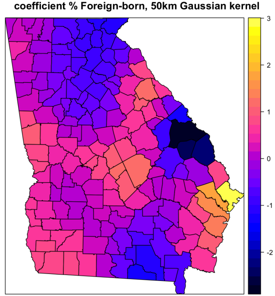


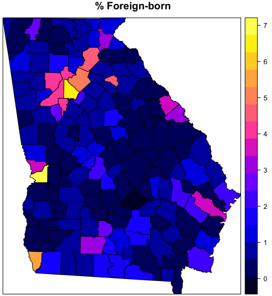
coefficient % Elderly, 22-county Gaussian kernel

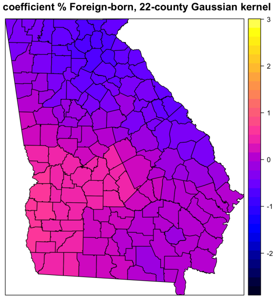


Note the increased noise with the narrower kernel.

Artefacts – foreign-born







Interpretation - 1

Substantial differences in **regression coefficients** across map

- In some cases even the **sign** changes – this may be a true effect
- Suggests **different causes/correlations** in different areas

Interpretation - 2

Substantial differences with **choice of kernel**

- So what is a “local” effect?
- **Question**: is 50 km with Gaussian weights an appropriate fixed bandwidth?
- **Question**: are 22 counties with Gaussian weights an appropriate adaptive bandwidth?

Mixed GWR

- GWR model with some coefficients **global**, i.e., *not* varying with the moving window
- Allows global/regional effects
  - Example: soil organic matter: affected by **regional** climate, but by **local** topographic effects [9]
- Mixed GWR tests which predictors are fixed and which can vary (and at which bandwidth) [8]

Multiscale Geographically Weighted Regression (MGWR)

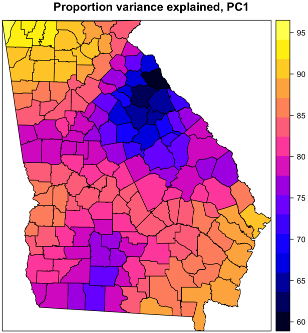
- Developed by Fotheringham et al. [4]
- GWR with **different bandwidths** for **different processes** (represented by predictors)
- computes an **optimal bandwidth vector** in which each element indicates the **spatial scale** at which a particular process takes place
- can interpret the various bandwidths to infer the spatial processes

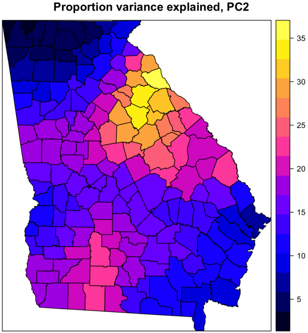
Geographically-weighted PCA

- As with OLS regression, but now Principal Components [6]
- Look for the multivariate correlations among predictors in a moving window
- Interpret the PC loadings, per window
- Can use the PC scores to create new, uncorrelated variables

GW PCA

Georgia poverty predictors, 50 km Gaussian bandwidth





PC1 much more explanatory in NW GA, i.e., predictors are much more correlated there

Conclusion

- A useful tool to **investigate** spatial heterogeneity in regression models
  - changing coefficients, changing variable importance, changing R^2
  - the **bandwidth** reveals the **spatial scale** of the heterogeneity
- This should be **interpretable** in terms of the physical/social setting

References I

[1] Roger Bivand. *Geographically Weighted Regression*. Oct 2017. URL https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf.

[2] A. Stewart Fotheringham. *Geographically Weighted Regression*, pages 242–253. SAGE Publications, Ltd, 2009. ISBN 978-1-4129-1082-8. doi: 10.4135/9780857020130.n13.

[3] A. Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, 2002. ISBN 0-471-49616-2. doi: 10.4135/9781849209755.

[4] A. Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, 107(6):1247–1265, Nov 2017. doi: 10.1080/24694452.2017.1352480.

[5] Isabella Gollini, Binbin Lu, Martin Charlon, Christopher Brundson, and Paul Harris. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 17, Feb 2015. doi: 10.18637/jss.v063.i17.

[6] Paul Harris, Chris Brunsdon, and Martin Charlton. Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25(10):1717–1736, Oct 2011. doi: 10.1080/13658816.2011.554838.

[7] Binbin Lu, Paul Harris, Martin Charlton, and Chris Brunsdon. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, 17(2):85–101, Apr 2014. doi: 10.1080/10095020.2014.917453.

References II

[8] Chang-Lin Mei, Ning Wang, and Wen-Xiu Zhang. Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A: Economy and Space*, 38(3):587–598, Mar 2006. doi: 10.1068/a3768.

[9] Canying Zeng, Lin Yang, A-Xing Zhu, David G. Rossiter, Jing Liu, Junzhi Liu, Chengzhi Qin, and Desheng Wang. Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. *Geoderma*, 281:69–82, Nov 2016. doi: 10.1016/j.geoderma.2016.06.033.