# Spatial Statistics in Epidemiology and Public Health
# Lecture 5: Spatial regression

**Lance Waller** and Howard Chang

# References

- Waller and Gotway (2004, Chapter 9) *Applied Spatial Statistics for Public Health Data*. New York: Wiley.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data, 2nd Ed.* Boca Raton, FL: CRC/Chapman & Hall.
- Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chichester: Wiley.

# What do we have so far?

- Tension between *statistical precision* (want large local sample sizes $\rightarrow$ big regions), and *geographic precision* (want small regions for more detail in map).
- *Disease mapping* approaches use *small area estimation* techniques to borrow information from all areas and from neighboring areas to improve local estimation in each area.
- But what about local covariates?
- Can we adjust for those (say, using regression models)?
- And still borrow information?
- With *independent* observations we know how to use *linear* and *generalized linear* models such as linear, Poisson, logistic regression.
- What happens with *dependent* observations?

# Caveat

*"...all models are wrong. The practical question is how wrong do they have to be to not be useful."*
*Box and Draper (1987, p. 74)*

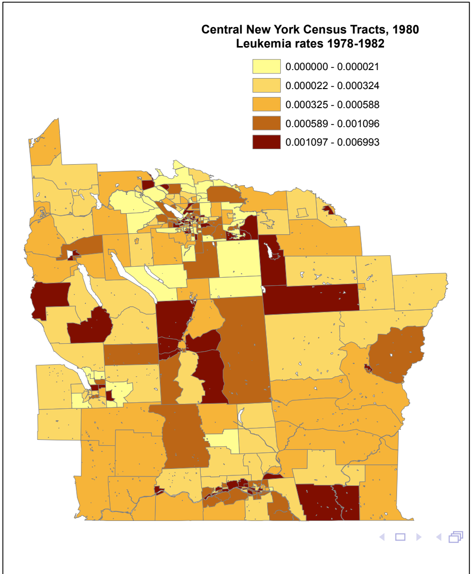
# What changes with dependence?

- In statistical modeling, we are often trying to describe the mean of the outcome as a function of covariates, assuming error terms are mutually independent.
- Where do correlated errors come from?
- Perhaps outcomes truly correlated (infectious disease).
- Perhaps we omitted an important variable that has spatial structure itself.

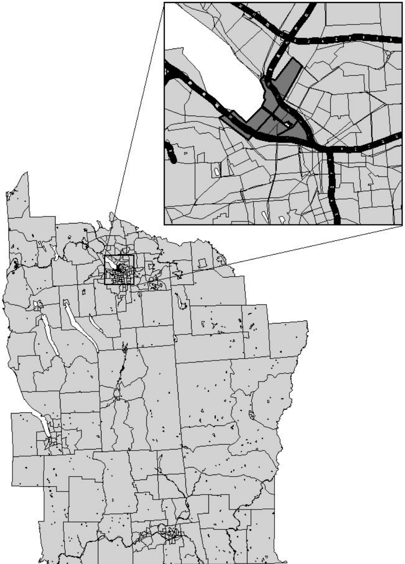
# NY leukemia data

- NY leukemia data and covariates (Waller and Gotway, 2004).
- 281 census tracts (1980 Census).
- 8 counties in central New York.
- 592 cases for 1978-1982.
- 1,057,673 people at risk.

# Crude Rates (per 100,000)



# Outliers, where are the top 3 rates?



# Building the model: Poisson regression

- Let $Y_i$ = count for region $i$.
- Let $E_i$ = *expected* count for region $i$.
- Let $(x_{i,TCE}, x_{i,65}, x_{i,home})$ be the associated covariate values.
- Poisson regression:

$$Y_i \sim Poisson(E_i \zeta_i)$$

where

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home}.$$

# Details

- Poisson distribution for counts.
- *Link function*: Natural log of mean of $Y_i$ is a linear function of covariates.
- $\beta$s represent multiplicative increases in expected counts, $e^{\beta}$ a measure of relative risk associated with one unit increase in covariate.
- $E_i$ an *offset*, what we expect if the covariates have no impact.
- Age, race, sex adjustments in either $E_i$ (standardization) or covariates.

## Adding spatial correlation: New York data

- Assume $E_i$ known, perhaps age-standardized, or based on global (external or internal) rates.
- Our model is

$$Y_i | \boldsymbol{\beta}, \psi_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(\boldsymbol{x}_i' \boldsymbol{\beta} + \psi_i)),$$

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

- The $\psi_i$ represent the *random intercepts*.
- Add *overdispersion* via $\boldsymbol{\psi}_i \overset{ind}{\sim} N(0, v_\psi)$.
- Add spatial correlation via

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

# Priors and "shrinkage"

- Overdispersion model (i.i.d. $\psi_i$) results in each estimate being a compromise between the *local* SMR and the *global average* SMR.
- "Borrows information (strength)" from other observations to improve precision of local estimate.
- "Shrinks" estimate toward global mean. (Note: "shrink" does not mean "reduce", rather means "moves toward").

## Local shrinkage

- Spatial model (correlated $\psi_i$) results in each estimate begin a compromise between the *local* SMR and the *local average* SMR.
- Shrinks each $\psi_i$ toward the average of its *neighbors*.
- Can also include *both* global and local shrinkage (Besag, York, and Mollié 1991).
- How do we fit these models?

# Bayesian inference

Bayesian inference regarding model parameters based on *posterior distribution*

$$Pr[\boldsymbol{\beta}, \psi | \boldsymbol{Y}]$$

proportional to the product of the likelihood times the prior

$$Pr[\boldsymbol{Y} | \boldsymbol{\beta}, \psi] Pr[\psi] Pr[\boldsymbol{\beta}].$$

Defers spatial correlation to the prior rather than the likelihood.

# Spatial priors

- Could model *joint* distribution

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

- Could also model *conditional* distribution

$$\psi_i | \psi_{j \neq i} \sim N\left(\frac{\sum_{j \neq i} c_{ij}\psi_j}{\sum_{j \neq i} c_{ij}}, \frac{1}{v_{CAR} \sum_{j \neq i} c_{ij}}\right), i = 1, \ldots, N.$$

  where $c_{ij}$ are *weights* defining the neighbors of region $i$.
- Adjacency weights: $c_{ij} = 1$ if $j$ is a neighbor of $i$.

# CAR priors

- The conditional specification defines the *conditional autoregressive* (CAR) prior (Besag 1974, Besag et al. 1991).
- Under certain conditions on the $c_{ij}$, the CAR prior defines a valid multivariate joint Gaussian distribution.
- Variance covariance matrix a function of the *inverse* of the matrix of neighbor weights.

# Fitting Bayesian models

- Posterior often difficult to calculate mathematically.
- Markov chain Monte Carlo: Iterative simulation approach to model fitting.
- Given *full conditional* distributions, simulate a new value for each parameter, holding the other parameter values fixed.
- The set of simulated values converges to a sample from the posterior distribution.
- Alternative: *integrated nested Laplace analysis* using the `inla` package (example code).

# Complete model specification

$$Y_i|\boldsymbol{\beta}, \psi_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \psi_i)),$$

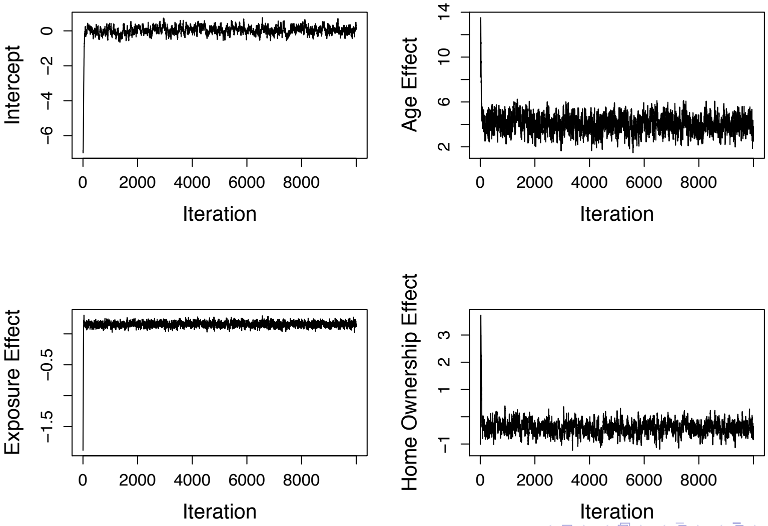
$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

$$\beta_k \sim \text{Uniform}.$$

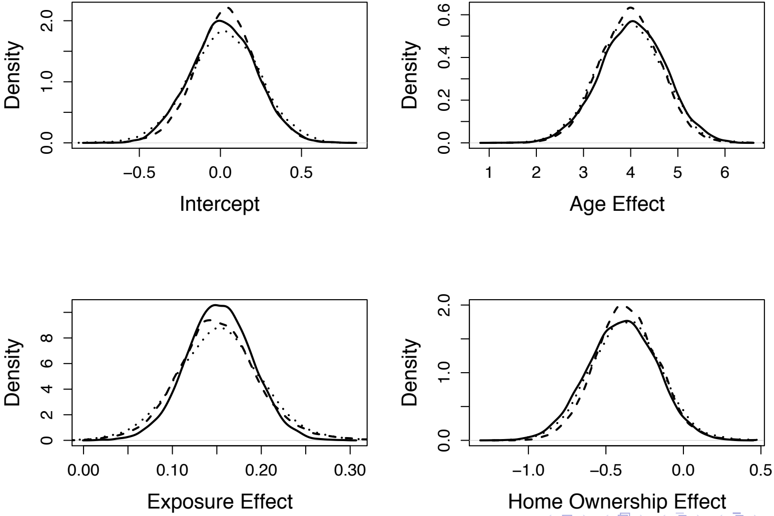
$$\psi_i|\psi_{j\neq i} \sim N\left(\frac{\sum_{j\neq i} c_{ij}\psi_j}{\sum_{j\neq i} c_{ij}}, \frac{1}{v_{CAR}\sum_{j\neq i} c_{ij}}\right), i = 1, \ldots, N.$$

$$\frac{1}{v_{CAR}} \sim \text{Gamma}(0.5, 0.0005).$$

# MCMC trace plots



# Posterior densities



# MCMC posterior estimates

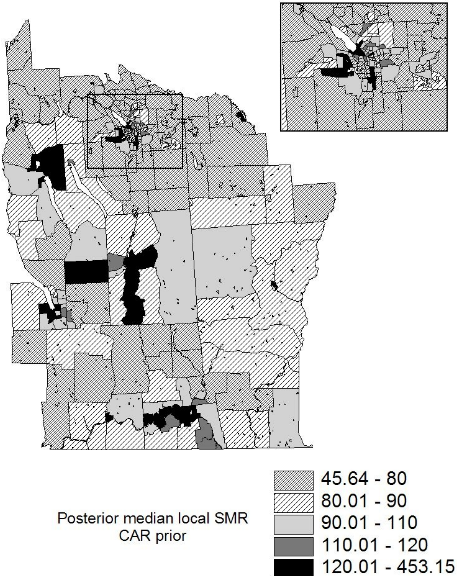
| Covariate | Posterior Median | 95% Credible Set |
|---|---|---|
| $\beta_0$ | 0.048 | (-0.355, 0.408) |
| $\beta_{65}$ | 3.984 | (2.736, 5.330) |
| $\beta_{TCE}$ | 0.152 | (0.066, 0.226) |
| $\beta_{home}$ | -0.367 | (-0.758, 0.049) |

# But there's more!

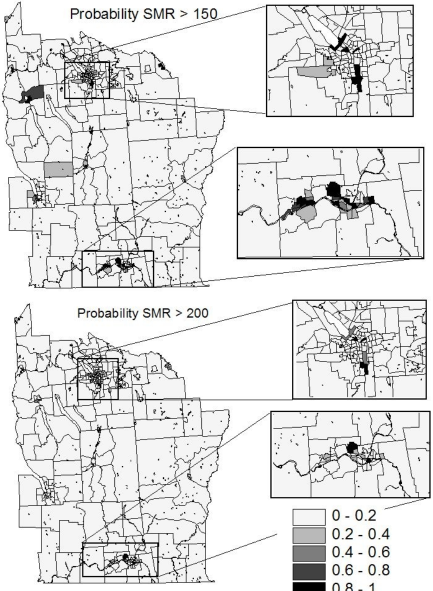
- A nifty thing about MCMC estimates:

  We get posterior samples from any function of model parameters by taking that function of the sampled posterior parameter values.
- Gives us posterior inference for $SMR_i = Y_{i,fit}/E_i$.
- Also can get $Pr[SMR_i > 200|\boldsymbol{Y}]$ and map these *exceedence probabilities*.

# Posterior median SMRs



# Posterior exceedence probabilities



# What do we have?

- Associations between local covariates and local outcomes (counts and rates).
- Spatial correlation between random intercepts (inside the link function).
- (Aside: This is a clever idea since we can use a multivariate Gaussian distribution for correlation...).
- Result: Local rates adjusted for covariates *and* smoothed by borrowing information.
- Many examples in the literature, and many extensions, we'll start with one tomorrow!

# Conclusions

- What method to use depends on what you want data you have and what question you want to answer.
- All methods try to balance trend (fixed effects) with correlation (here, with random effects).
- All models wrong, some models useful.
- Trying more than one approach often sensible.

# Slippery Slopes: What if *slopes* vary in space?

- Collaborators: Paul Gruenewald, Dennis Gorman, Li Zhu, Carol Gotway, and David Wheeler
- References:

  - Waller et al. (2008) Quantifying geographical associations between alcohol distribution and violence... *Stoch Environ Res Risk Assess* **21**: 573-588.
  - Wheeler and Caldor (2009) As assessment of coefficient accuracy...*J Geogr Systems* **9**: 573-588.
  - Wheeler and Waller (2009) Comparing spatially varying coefficient models... *J Geogr Systems* **11**: 1-22.
  - Finley (2011) Comparing spatially-varying coefficient models...*Methods in Ecology and Evolution* **2**: 143-154.

# What do we want to do?

- Quantify associations between outcomes and covariates as observed in data.
- Adjust for spatial correlation (spatial regression) using a random intercept with a CAR prior.
- What if strength of association varies across space?
- Usually, we assume $\beta$ is the same at every location, what if it varies (but is spatially correlated)?
- Can we have a *random slope*? Can we use CAR priors for that?

# What about spatially varying associations?

- Fix it: Geographically weighted regression (GWR)
  - Fotheringham et al. (2002)
- Model it: Spatially varying coefficient (SVC) models
  - Leyland et al. (2000), Assuncao et al. (2003), Gelfand et al. (2003), Gamerman et al. (2003), Congdon (2003, 2006)

# Data example: Alcohol, illegal drugs, violent crime

- Outcome: Rates (number of cases per person per year) of violent crimes (police/sheriff reports).
- Covariates: Alcohol distribution (licenses and sales), illegal drug arrests (police/sheriff reports).
- Potential confounders: Sociodemographics (census).
- Linked to common spatial framework (census tracts) via GIS.

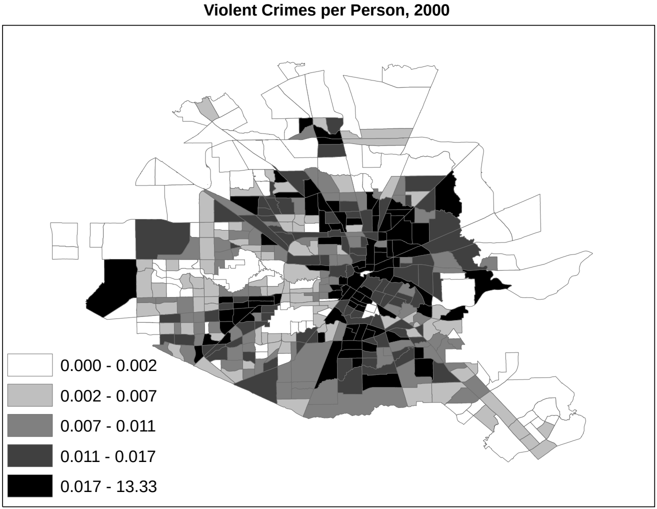
# Translation complications

- When are crime data like disease data?
  - Counts from small areas.
  - Per person "rate" of interest.
- When are crime data not like disease data?
  - Outcome not as "rare".
  - Police vs. medical records.
  - Residents not only ones at risk.

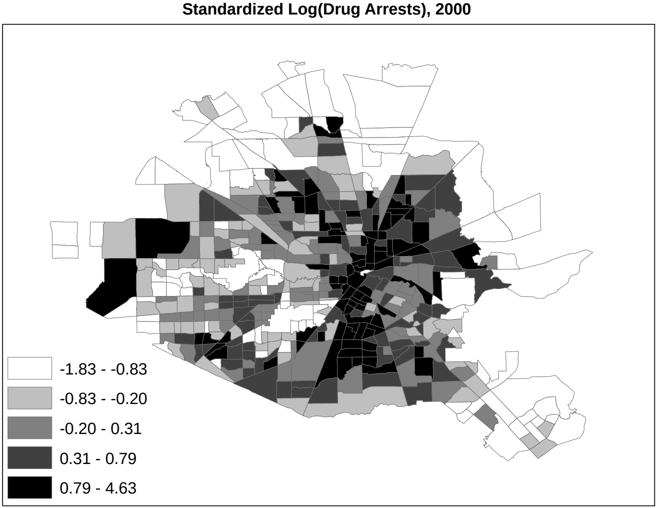
# Data description

- Spatial support: 439 census tracts (2000 Census).
- Violent crime (murder, robbery, rape, aggravated assault) "first reports" for year 2000 from City of Houston Police Department website.
- Gorman et al. (2005, *Drug Alcohol Rev*) report less than 5% discrepancy with 2000 Uniform Crime Reports.
- 98% of reports geocoded to the census tract level.
- Alcohol data (locations of active distribution sites in 2000) from Texas Alcoholic Beverage Commission (6,609 outlets), 99.5% geocoded to the tract level.
- Drug law violations (also from City of Houston police data). 98% geocoded to the tract level.

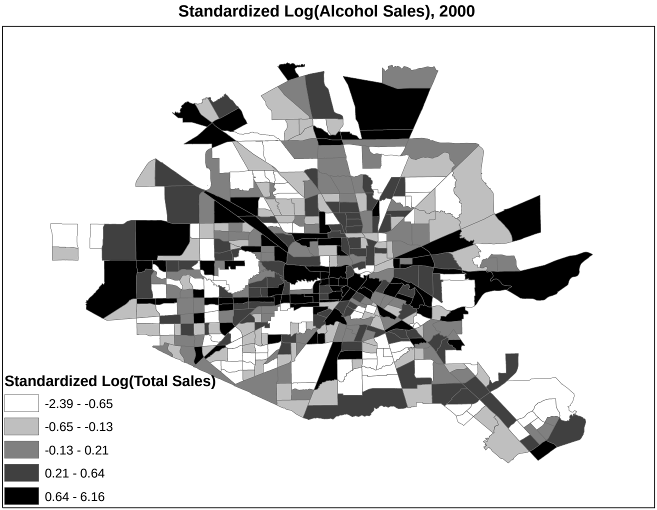
# Violent Crime reporting rates, Houston, 2000



# Standarized log(drug arrests), Houston, 2000



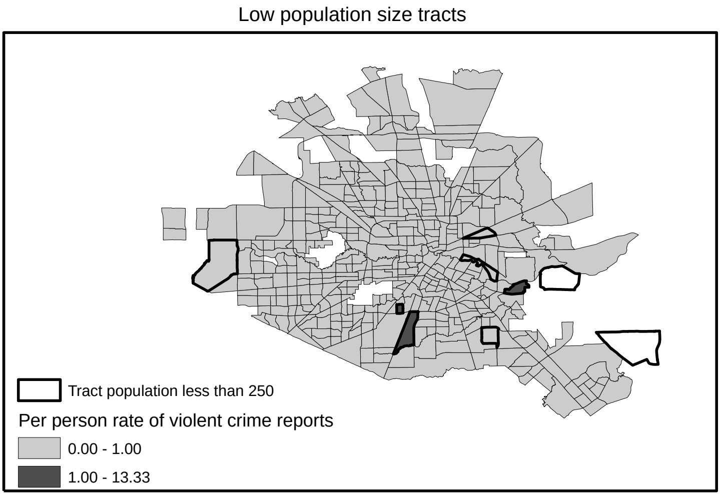
# Standarized log(alcohol sales), Houston, 2000



# Data "features"

- 7 of 439 tracts have extremely small population sizes: 1, 3, 4, 16, 34, 116, and 246.
- Tracts typically have 3,000-5,000 residents.
- Local rates for such tracts are extremely unstable (e.g., 40 reports, 3 residents).
- Actually a motivating a reason for including the spatially varying intercept: borrow information across regions.

# Low population tracts and high rates



# Basic Poisson regression

- Let $Y_i$ = number of reports in tract $i, i = 1, \ldots, 439$.
- Suppose $Y_i \sim \text{Poisson}(E_i \exp(\mu_i))$, where $E_i$ = the "expected" number of reports under some null model.
- Typically, $E_i = n_i R$ where all $n_i$ individuals in region $i$ are equally likely to report.
- $\exp(\mu_i)$ = "relative risk" of outcome in region $i$.
- We add covariates in linear format (within $\exp(\cdot)$): $\mu_i = \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}$.
- Same "skeleton" for both GWR and SVC.

# Why do we have $E_i$?

- $E_i = n_i R$ represents an "offset" in the model and lets us use Poisson regression to model *rates* as well as *counts*.

$$\begin{aligned} E[Y_i] &= E_i \exp(\beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\ &= \exp(\ln(E_i) + \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\ &= \exp(\ln(n_i) + \ln(R) + \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\ \log(E[Y_i]) &= \ln(n_i) + \ln(R) + \beta_0 + \beta 1 x_{alc,i} + \beta_2 x_{drug,i} \end{aligned}$$

- GWR offset: $\ln(n_i)$, SVC offset: $\ln(n_i) + \ln(R)$.

# GWPR (Nakaya et al., 2005)

- Geographically weighted Poisson regression.
- $\widehat{\beta}_{GWPR} = (\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{s})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})$.
- $\boldsymbol{A}(\boldsymbol{s})$ = diagonal matrix of Fisher scores.
- $\boldsymbol{Z}(\boldsymbol{s})$ = Taylor-series approximation to transformed outcomes.
- Update $\boldsymbol{A}(\boldsymbol{s})$, $\boldsymbol{Z}(\boldsymbol{s})$ and $\widehat{\beta}_{GWPR}$ until convergence.

# Fitting in *R*

- Waller et al. (2007) use `GWR 3.0` software.
- In `R`: `maptools` will read in `ArcGIS`-formatted shapefile (files) into *R*.
- `spgwr` fits linear GWR and GLM-type GWR.

# SVC

- $\mu_i = \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i} + b_{1,i} x_{alc,i} + b_{2,i} x_{drug,i} + \phi_i + \theta_i$.
- $\beta_0, \beta_1, \beta_2 \sim$ Uniform.
- Random intercept has 2 components (Besag et al. 1991):

$$\theta_i \overset{ind}{\sim} N(0, \tau^2)$$

$$\phi_i | \phi_j \sim N\left(\frac{\sum_j w_{ij}\phi_j}{\sum_j w_{ij}}, \frac{1}{\lambda \sum_j w_{ij}}\right).$$

  where $w_{ij}$ defines neighbors, and $\lambda$ controls spatial similarity.
- $\theta_i$ allows overdispersion (smoothing to global mean).
- $\phi_i$ follows conditionally autoregressive distribution (smoothing to local mean), generates MVN but more convenient for MCMC.

# Defining the SVCs

- $\boldsymbol{b}_1, \boldsymbol{b}_2$ also given spatial priors and allowed to be correlated with one another.
- We use a formulation by Leyland et al. (2000) which defines

$$(b_{1,i}, b_{2,i})' \sim MVN((0,0)', \boldsymbol{\Sigma})$$

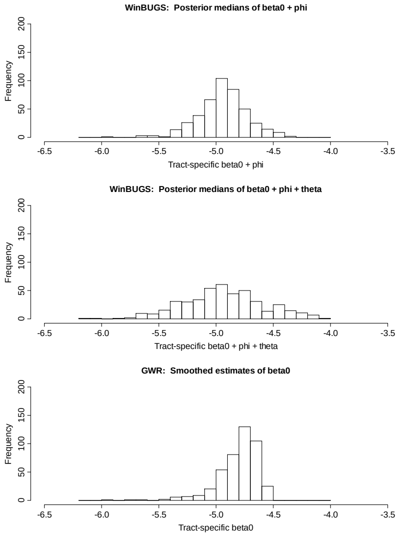
.

- Typically fit via MCMC (but often runs slowly).

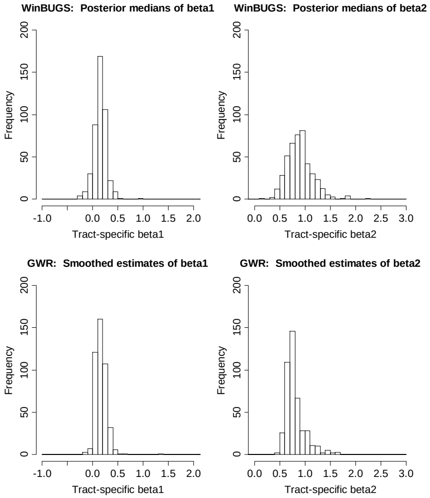
# Implementation

- Example code using `spgwr` library.)
  - Convergs in minutes.
- MCMC to fit SVC model.
  - Converged in hours.
- Fit several versions of SVC model and compared fit via deviance information criterion (Spiegelhalter et al., 2003).
- Best fit included spatial varying coefficients, random intercept, and correlation between alcohol and drug effects.

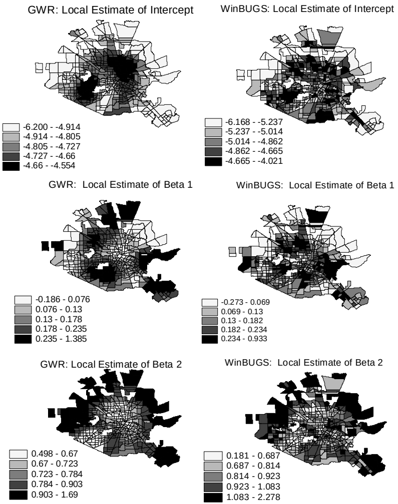
# Results: Intercept



# Results: Alcohol sales and drug arrests



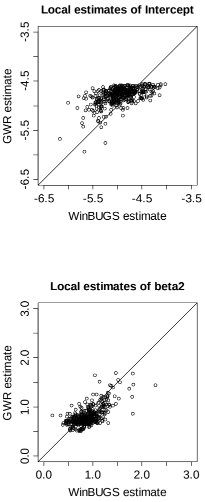
# Estimated effects



# Similarities

- Alcohol: Increased impact in western, south-central, and southeastern parts of city.
- Illegal drug: Increased impact on periphery, lower influence in central and southwestern parts of city.
- Intercept: Increased risk of violence in central area, above and beyond that predicted by alcohol sales and illegal drug arrests.
- But, associations not too close...

# Results: tract-by-tract



# Differences

- GWR much smoother based on global best fit for *bw*.
- SVC used adjacency-based smoothing and a different amount of smoothing for each covariate.
- GWR: collineary between surfaces (Wheeler and Tiefelsdorf, 2005).
- SVC: Model based approach removes (or at least reduces) collinearity.

# Let's try it out!

- Houston data on violent crime, alcohol sales, and illegal drug arrests.
- `ArcGIS` shapefile.
- Required `R` libraries: `maptools` (to read in shape file), `RColorBrewer` (to set colors), `classInt` (to set intervals of values for mapping), and `spgwr` (for GWR).

# Conclusions

- GWR and SVC very different approaches to the same problem.
- Qualitatively similar in results, but not directly transformable.
- GWR fixed problems within somewhat of a black box.
- SVC allows probability model-based inference with lots of flexibility but at a computational cost (both in set-up and implementation).
- Further research:
  - Wheeler and Waller (2009): Attempt to set up SVC model to more closely mirror amount of smoothing in GWR.
  - Collinearity "ribbons".
  - Griffith (2002) eigenvector spatial filtering to adjust collinearity. Interpretability?