

# MODULE 9: Spatial Statistics in Epidemiology and Public Health

## Lecture 9: Disease Ecology

**Lance Waller** and Howard Chang

Disease Ecology: What do we want to do?

Point Processes: Chagas Disease in Peru

Spatial Regression: Raccoon Rabies

## Disease Ecology: What do we want to do?

Pattern and Process

## Point Processes: Chagas Disease in Peru

Cluster detection

Spatial relative risk

## Spatial Regression: Raccoon Rabies

Statistical estimation of landscape barriers

Wombling

Spatially varying coefficients

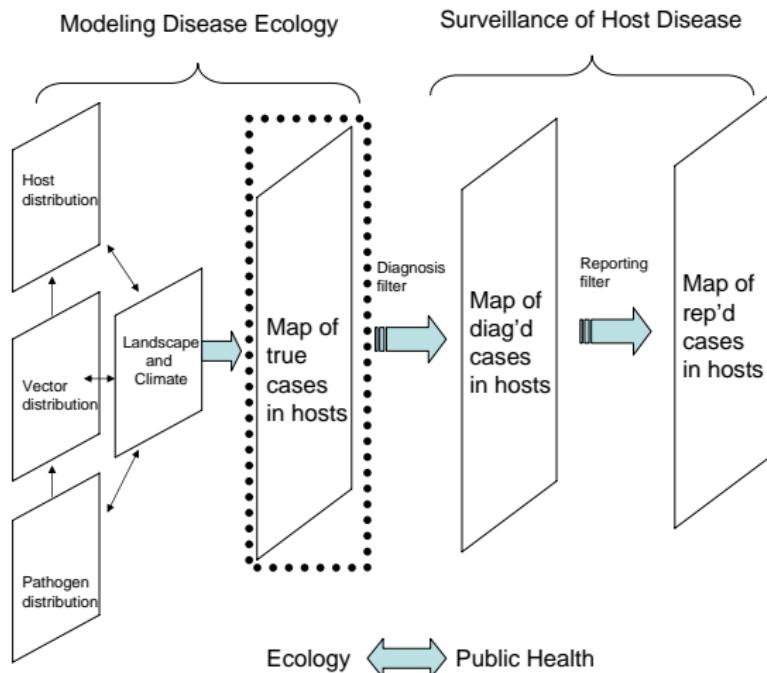
# Disease Ecology

- ▶ Interactions between virus, host, landscape.
- ▶ Landscape epidemiology (Pavlovsky, 1967), landscape ecology (Manel et al. 2003, *TrEE*), spatial epidemiology (Osfeld et al. 2005, *TrEE*), landscape genetics (host and virus) (Biek et al. 2006, *Science*), conservation medicine (Aguirre et al. 2002).
- ▶ People, animals, diseases, ecology, environment!
- ▶ Spatio-temporal data, mathematical models, genetic sequences, missing data, GIS!

# Epizoology and Epidemiology

- ▶ Most emergent infectious diseases have animal reservoir (WNV, Ebola, Avian influenza, Monkeypox, SARS, HIV/SIV).
- ▶ History of animal/human disease (Torrey and Yolken, 2005, *Beasts of the Earth*).
- ▶ Interesting intersection of modelers, ecologists, statisticians, medical geographers, ecological geneticists, public health researchers, epidemiologists.
- ▶ One Health.

# The “big picture”



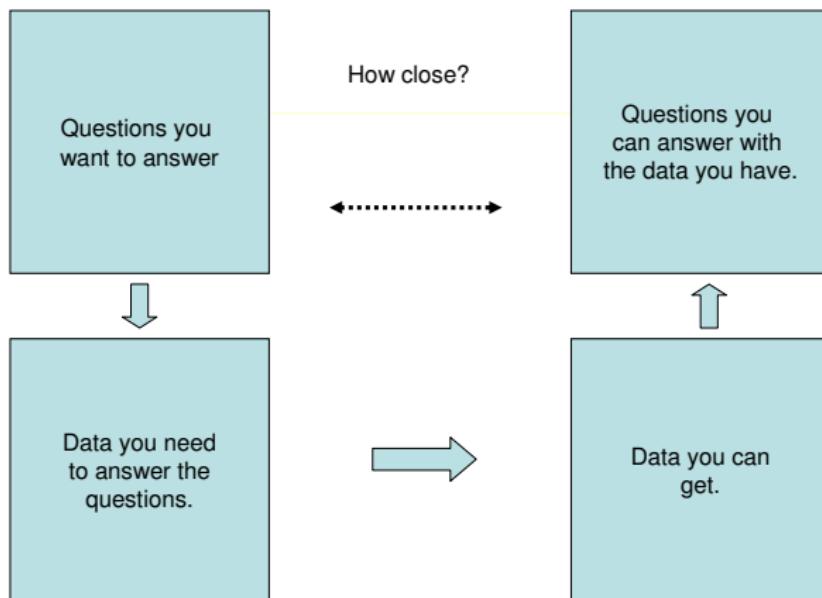
# Pattern and Process

- ▶ Our ultimate goal is understanding the ecological processes driving the patterns we see in our observations.
- ▶ When linking process (model or reality) to pattern (data), typically:
  - ▶ Ecology focus: Process to pattern
    - ▶ Emphasis on mathematical model, link to available data
  - ▶ Statistics: Pattern to process
    - ▶ Collected data, hypothesis test or analytic (e.g., regression) model.

# Ultimately futile exercise?

- ▶ Process may not yield unique pattern (e.g., chaos, stochasticity).
- ▶ Pattern may not reveal unique process without additional information (e.g., spatial point patterns, Bartlett (1964)).
- ▶ But the real question is, “Can we learn more than we already know?”
- ▶ If not, what additional data do we need?

# The whirling vortex



# Point Processes in Disease Ecology: Chagas disease in Peru

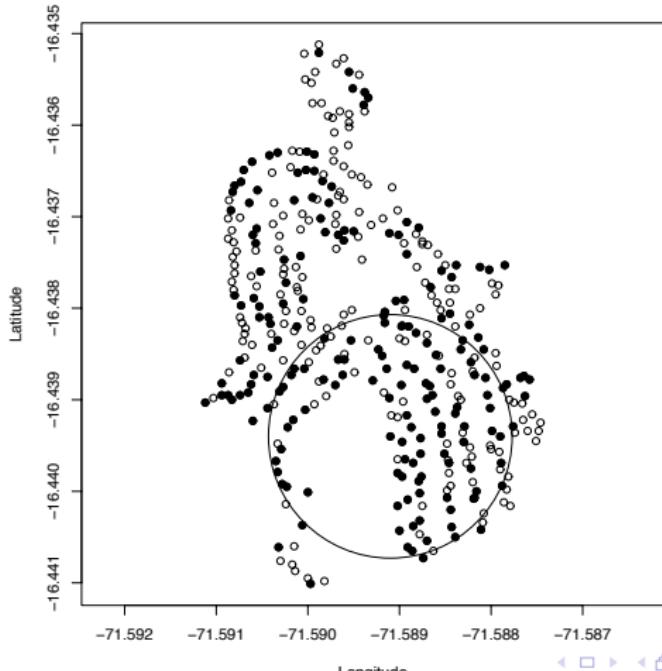
- ▶ Joint work with Michael Levy (Fogarty International Center, NIH)
- ▶ Chagas disease: Vector borne disease (infection with *T. cruzi*).
- ▶ Vector (in southern Peru): *Triatoma infestans*.
- ▶ Study area: Guadalupe, Peru (peri-urban).
- ▶ Fields surrounding rocky hilltops with houses.
- ▶ GPS all household locations.
- ▶ Spraying campaign, identify house locations, houses with vectors ("infested"), and houses with infected vectors ("infected").

# How to find a cluster?

- ▶ Consider two approaches: scan statistic and intensity estimators.
- ▶ Spatial scan statistic:
  - ▶ Define set of potential clusters (elements of scanning window).
  - ▶ Assign “score” to each potential cluster.
  - ▶ Find “most likely cluster” (MLC) as potential cluster with extreme score.
  - ▶ Evaluate significance of most likely cluster via Monte Carlo simulation.
  - ▶ Compare observed “score” of MLC to distribution of scores MLCs (regardless of location) under random assignment.
  - ▶ SaTScan software ([www.satscan.org](http://www.satscan.org)).

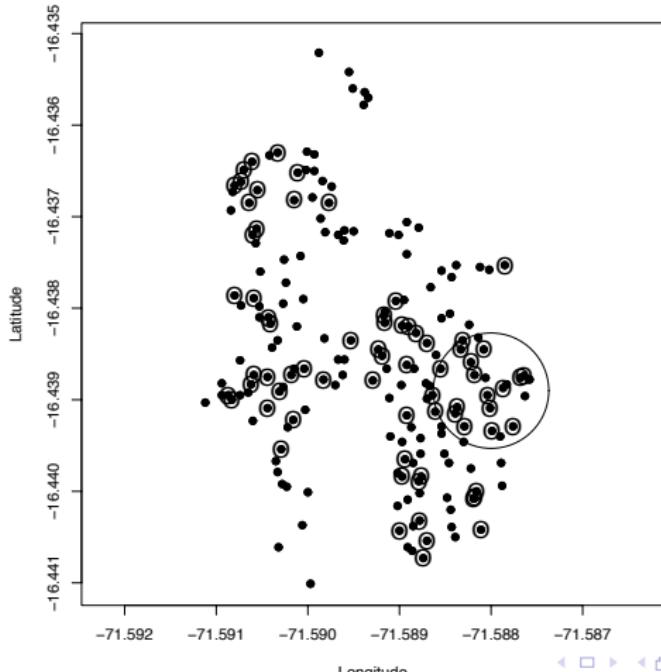
# SaTScan, Infested among households, Most likely cluster ( $p=0.002$ )

SaTScan, Most Likely Cluster, Infested,  $p$ -value = 0.002



# SaTScan, Infected among infested, Most likely cluster ( $p=0.181$ )

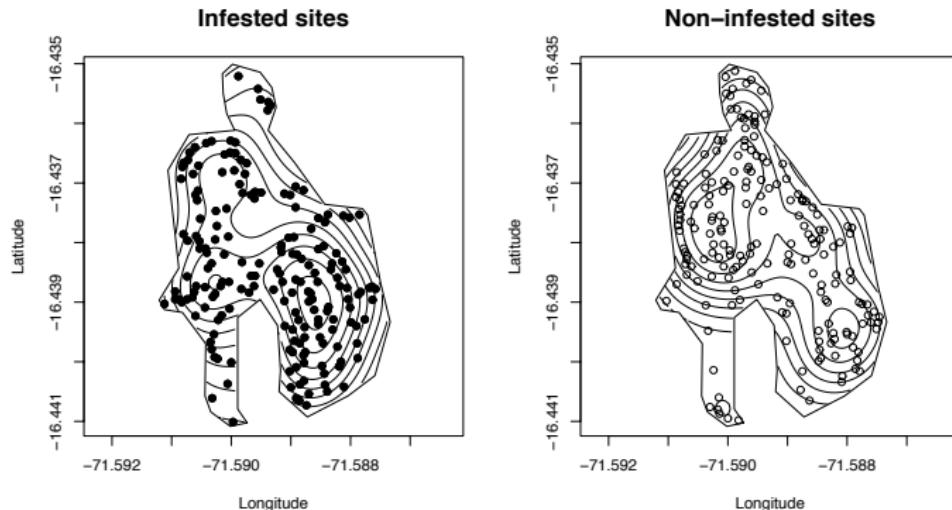
SaTScan, Most Likely Cluster, Infected,  $p$ -value = 0.181



# Chagas SaTScan conclusions

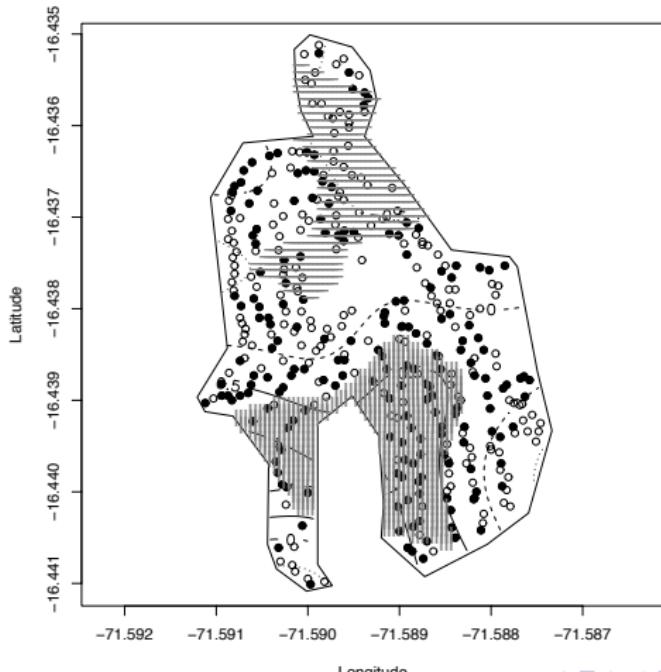
- ▶ Statistically significant cluster of infested households among all households.
- ▶ No statistically significant cluster of infected households among infested households.
- ▶ Note circular most likely cluster may include gaps (top of hill).
- ▶ What about non-circular clusters?

# Kernel intensity estimates, infested vs. all households

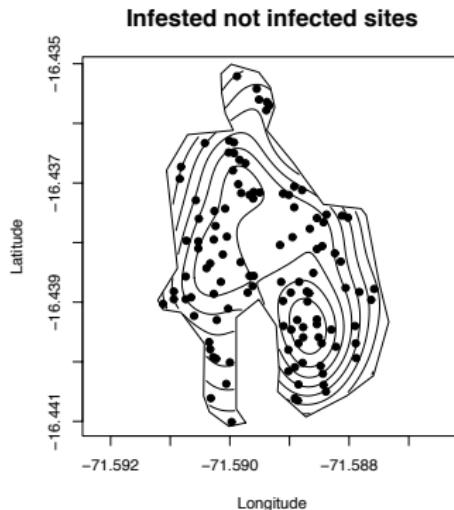
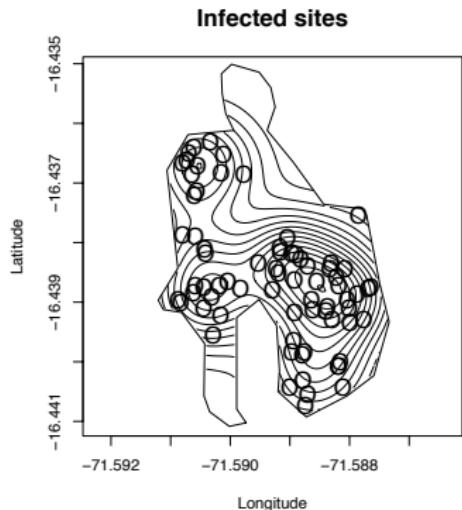


# Ratio of kernel intensity estimates, infested vs. all households

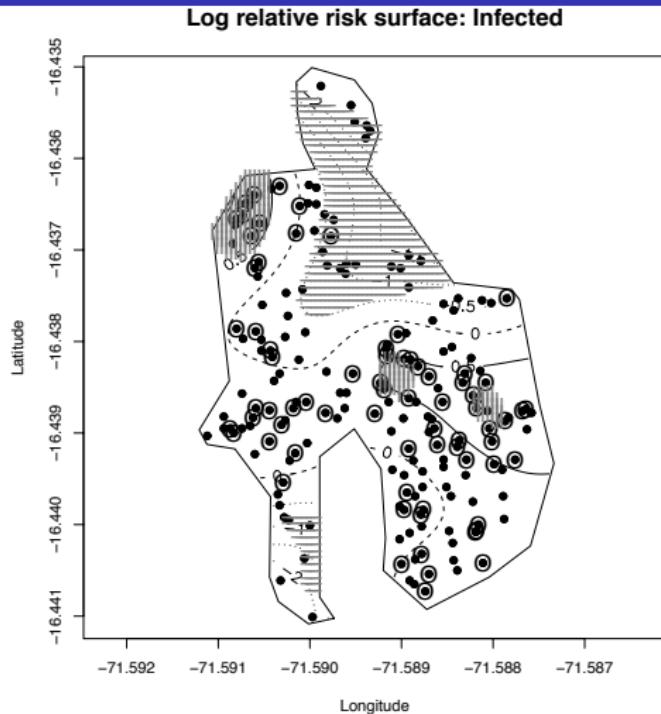
Log relative risk surface: Infested



# Kernel intensity estimates, infected vs. infested households



# Ratio of kernel intensity estimates, infected vs. infested



## Cluster conclusions

- ▶ Relative risk surface adds more geographical precision to patterns initially revealed by SaTScan.
- ▶ Large risk of infestation in the south.
- ▶ Within this some pockets of increased risk of infection.
- ▶ Area of lower risk missed by circular scan statistic, due to its irregular shape.
- ▶ Identifies areas for future studies.

## Chagas conclusions

- ▶ Significant cluster of infested households, but no clusters of infected households (circular clusters).
- ▶ Relative risk surface also suggests area of low risk (both infestation *and* infection) in northeast.
- ▶  $K$  functions suggest significant *clustering* of *infected* but not *infested* households.
- ▶ Taken together, results reveal different aspects of the underlying process.
- ▶ A single cluster does not define clustering, nor does clustering imply a single cluster.

# Chagas conclusions

- ▶ Infestation: pockets of higher and lower relative risk, but level of clustering not different between cases and controls.
- ▶ Infection: More clustered at small distances than infestation, but resulting clusters are smaller and more diffuse.
- ▶ Scale of clustering different between infestation and infection, and larger than typical range of individual vectors.
- ▶ Scale of clustering useful in targeted surveillance for human cases (Levy et al., 2007, *PLoS NTD*).

# Questions?

# Raccoon rabies



# What is rabies?

- ▶ Virus in family of Lyssa ("frenzy") virus.
- ▶ Reportable disease.
- ▶ Various strains associated with primary host (bat, dog, coyote, fox, skunk, and raccoon).
- ▶ Host cross-over, typically transmitted via bite/scratch.
- ▶ Most human infection from bat strains.

# Raccoon rabies

- ▶ Endemic in Florida and South Georgia.
- ▶ Translocation of rabid animal(s) to VA/WV border circa 1977.
- ▶ Wave-like spread since.
- ▶ Connecticut first appearance 1991-1996.
- ▶ Ohio 2005.
- ▶ Joint work with Leslie Real's lab in Population Biology, Evolution, and Ecology (David Smith, Colin Russell, Roman Biek, Scott Duke-Sylvester).

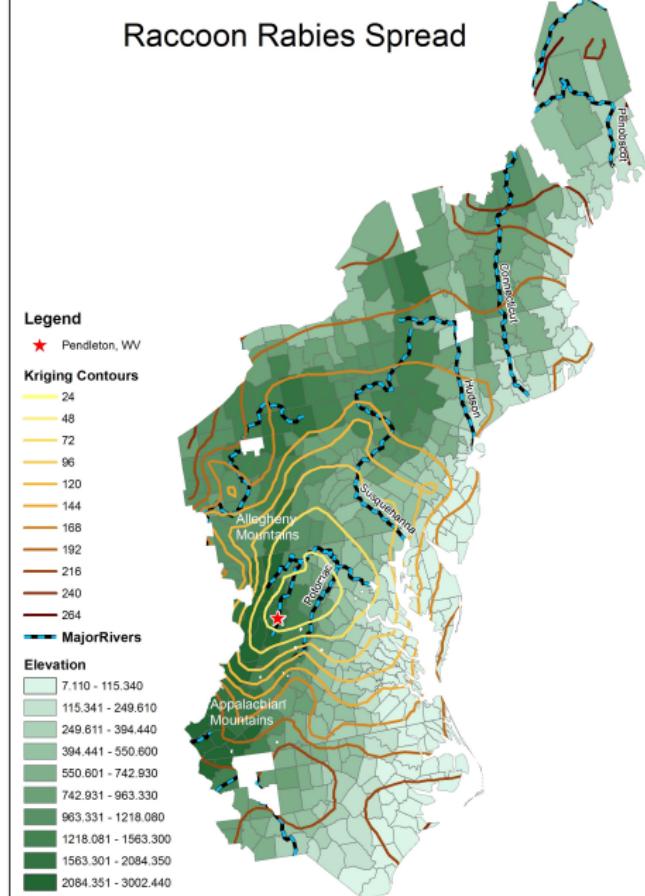
# Barrier estimation: What do we want?

- ▶ Goal: Measure effect of landscape features, (e.g., mountains and rivers) on the speed of raccoon rabies diffusion.
- ▶ Elevation, river or road presence significantly related to raccoon rabies counts (Recuenco et al. 2007) and transmission time (Russell et al. 2004).
- ▶ Landscape features may serve as either barriers or gateways to the spread of infectious disease.
- ▶ Find and visualize barriers: Do they align with certain landscape features?

# Data: What do we have?

- ▶ Time in months to first reported raccoon rabies case in 428 contiguous counties in the Eastern US (CDC).
  - ▶ 0 for origin county: Pendleton, WV.
- ▶ Mean elevation by county (USGS - Geographic Names Information System).
- ▶ Indicator for major river presence in county (ESRI data and a geographic information system (GIS)).
- ▶ Population density by county (US Census and ESRI).
- ▶ Distance between origin county and all counties.

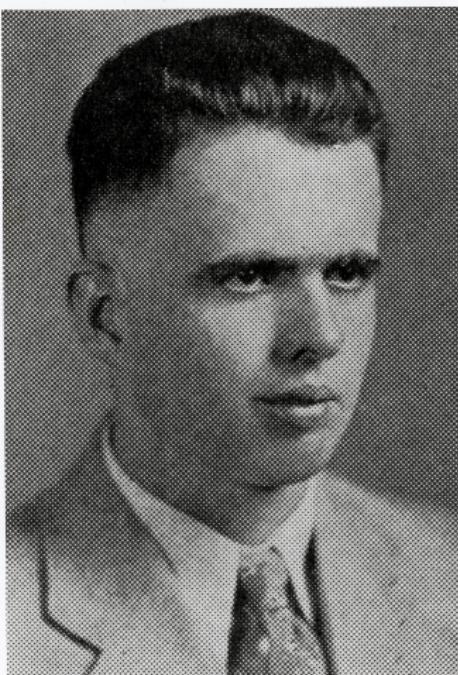
# Data



# Wombling

- ▶ Joint work with David Wheeler (Wheeler and Waller, *JABES*, 2008).
- ▶ Wombling: determine boundaries on a map by finding where local spread (change) is slower than elsewhere (Womble, 1951 *Science*).
- ▶ William H. Womble a bit of an elusive figure...

# William H. Womble (?)



# Google search: W.H. Womble Professor Robert Stencel



Which leads to...



# Are you ready to womble?

- ▶ Consider a set of potential boundaries and decide if each is a “real” boundary or not.
- ▶ Many algorithmic approaches both deterministic and “fuzzy”.
- ▶ Adopt a Bayesian hierarchical model for wombling (Lu and Carlin 2005).
- ▶ Bayesian approach provides a direct estimate of the probability that a line segment between two adjacent areas is a barrier (fuzzy boundary) in contrast to algorithmic versions based on thresholds, etc.

## Bayesian areal wombling

- ▶ Model time to first reported raccoon rabies case  $Y_i$ :

$$Y_i | \mu_i, \tau \sim N(\mu_i, 1/\tau)$$

where

$$\mu_i = \alpha + \phi_i$$

is the expected value of time to first case per county.

- ▶ Spatial random effects follow a conditionally autoregressive (CAR) prior  $\phi \sim CAR(\eta)$  with a mean random effect determined by its neighboring values.

# Bayesian areal wombling

- ▶ *Boundary likelihood value* (BLV) assigned to each potential boundary (here, edge between two counties), based on difference in expected (modeled) time to first appearance.

$$\Delta_{ij} = |\mu_i - \mu_j|$$

- ▶ Use MCMC to draw sample from posterior  $[\Delta_{ij} | \mathbf{y}]$  based on draws from posteriors  $[\mu_i | \mathbf{y}]$  and  $[\mu_j | \mathbf{y}]$ .
- ▶ This assigns a posterior probability for each edge, then display edges with  $p(\Delta_{ij} > c | \mathbf{y})$  for some threshold probability  $c$ .

# Wombling boundaries: $p(\Delta_{ij} > c | \mathbf{y})$

24 Months



72 Months



## Linking to local covariates

- ▶ Bayesian areal wombling provides estimates of barriers but does not allow direct inference regarding the impact of particular landscape barriers on the evidence for barriers.
- ▶ We could expand our fixed effect  $\alpha$  to  $\mathbf{X}'\boldsymbol{\beta}$  to include local covariates (e.g., elevation, boundary based on a river).
- ▶ However, what if the effect of elevation or presence of river varies from place to place?
- ▶ Russell et al. (2003, *PNAS*) suggest that river effect depends on direction of movement of the wave (perpendicular? Slower. Parallel? Faster.)

# Spatially varying coefficients

- ▶ We consider a *spatially varying coefficient* model with CAR priors on the covariate effects  $\beta$ , i.e.,

$$Y_i | \mu_i, \tau \sim N(\mu_i, 1/\tau)$$

where

$$\mu_i = \mathbf{X}'_i \boldsymbol{\beta}_i + \phi_i$$

- ▶ Spatial priors on elements of  $\boldsymbol{\beta}_i$ .
- ▶ More specifically, assign a multivariate CAR prior on the set of  $\boldsymbol{\beta}$  (Banerjee et al. 2004).

## MultiCAR details

- ▶  $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ip})'$
- ▶  $\beta_i | (\beta_{(-i),1}, \beta_{(-i),2}, \dots, \beta_{(-i),p}) \sim N(\bar{\beta}_i, \Omega/m_i)$   
where

$$\bar{\beta}_i = (\bar{\beta}_{i1}, \bar{\beta}_{i2}, \dots, \bar{\beta}_{ip})'$$

and

$$\bar{\beta}_{i1} = \sum_{k \in \kappa_i} \beta_{k1} / m_i$$

where  $\kappa_i$  = neighbor set for region  $i$ , and  $|\kappa_i| = m_i$ .

- ▶  $\Omega \sim \text{Inverse-Wishart}(\nu, 0.02 \cdot I_{p \times p})$ .

# Including covariates

- ▶ Include effects of (mean) elevation, presence of a major river, and the natural log of the (human) population density.
- ▶ Best fitting (via DIC) model includes spatial variation in all three (and intercept).

$$E[Y_i] = \beta_{i1} + \beta_{i2}(\text{mean elev}) + \beta_{i3}(\text{river}) + \beta_{4i}(\log(\text{pop dens}))$$

## Outline

Disease Ecology: What do we want to do?

Point Processes: Chagas Disease in Peru

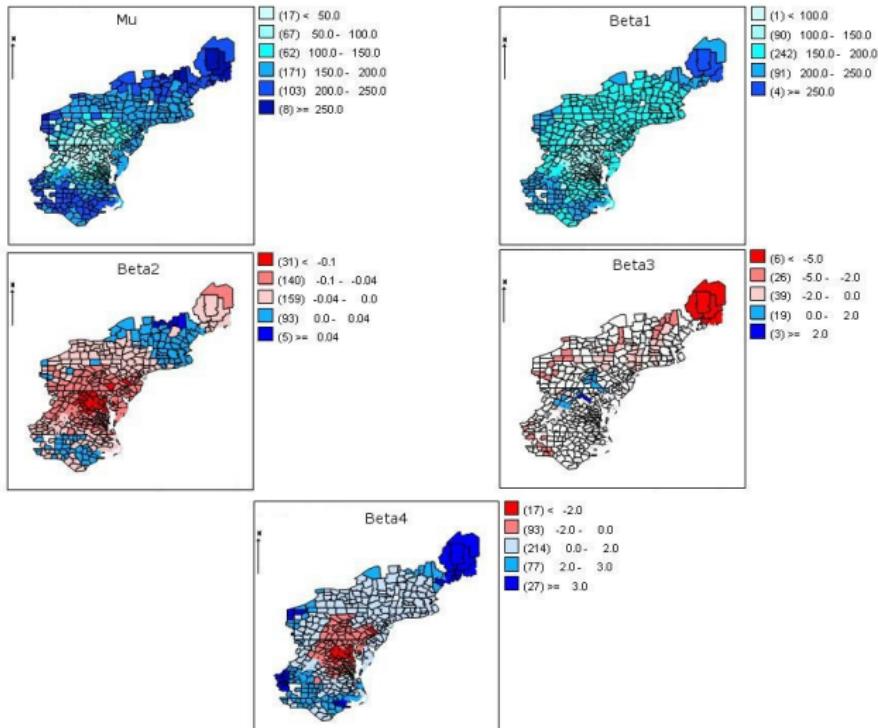
Spatial Regression: Raccoon Rabies

Statistical estimation of landscape barriers

Wombling

Spatially varying coefficients

$\beta_1$ : int,  $\beta_2$ : elev,  $\beta_3$ : river,  $\beta_4$ : log pop



# Findings/interpretations

- ▶ Map of posterior mean (MU): shows the overall wave or spread.
- ▶ Random intercept reveals local adjustments.
- ▶ River effect indicates increases in time until first appearance across Potomac and Susquehanna Rivers, decreases time for Hudson River and others.
- ▶ Elevation is not difference in elevation so not directly informing on elevation gradients as barriers, simply elevation impact on time until appearance.

SVC wombled boundaries:  $p(\Delta_{ij} > c | \mathbf{y})$

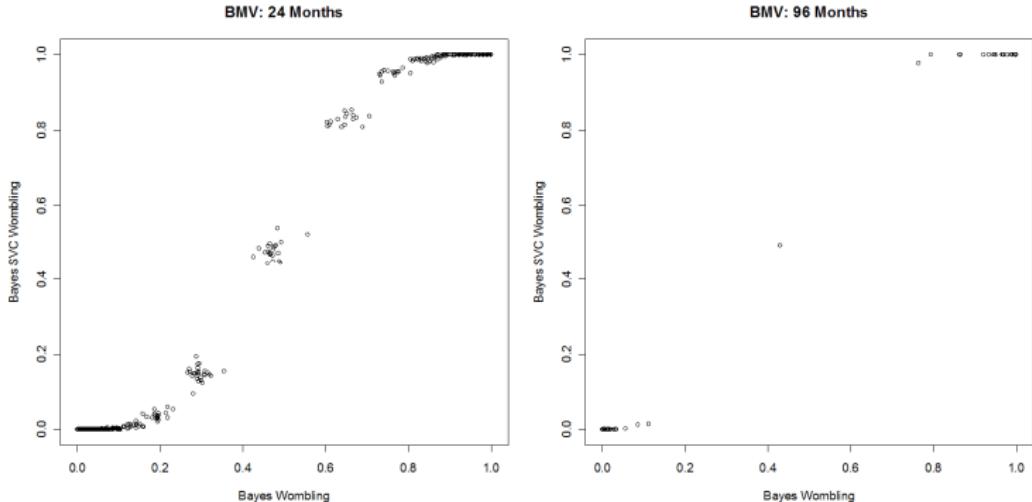
24 Months



72 Months



# Including covariates → better wombling?



# Overall Conclusions

- ▶ Much to be done to link mathematical models to statistical ideas.
- ▶ Disease ecology offers great setting for exploration.
- ▶ Models of transmission, interaction, observation.
- ▶ Mathematical models can inform statistics, statistics can inform models.
- ▶ Room to move past “ad-hockery”.
- ▶ Linking landscape features in a more meaningful (inferential) and spatial way.

# Closing

“... Nature’s dice are always loaded, ... in her heaps and rubbish are concealed sure and useful results.”

Ralph Waldo Emerson, *Nature*.

# References

- ▶ Smith et al. (2002) Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *PNAS* **99**, 3668-3672.
- ▶ Waller et al. (2003) Monte Carlo assessments of fit for ecological simulation models. *Eco Mod* **164**, 49-63.
- ▶ Waller (2010) Bridging gaps between statistical and mathematical modeling in ecology. *Ecology* **91**, 3500-3502.
- ▶ Wheeler and Waller (2008) Mountains, valleys, and rivers: The transmission of raccoon rabies over a heterogeneous landscape. *JABES* **13**, 388-406.