Environmental Exposure Assessment in Spatial Health Modeling: why is it important?

Andrew B Lawson
Division of Biostatistics and Bioinformatics
Department of Public Health Sciences
MUSC

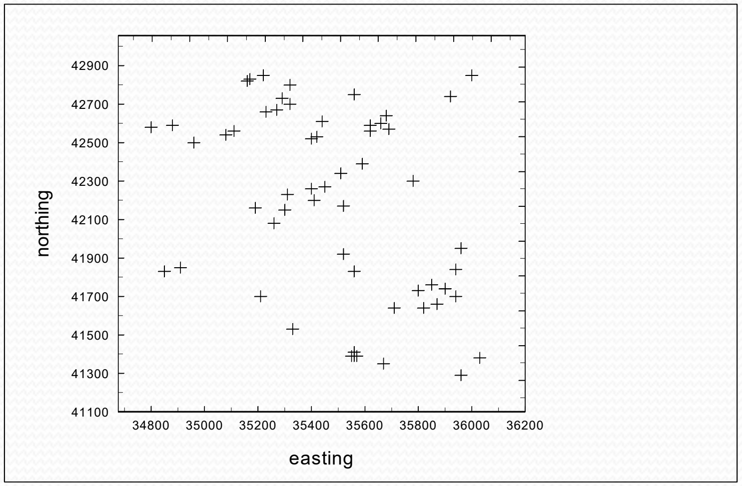
TIES International Conference, Edinburgh, UK
2016

What I will talk about:

- Background
  - Small area health data
  - Environmental stressors and etiology
  - Modeling approaches
- Case Studies/scenarios
- Mixtures (in models and in predictors)
- Challenges

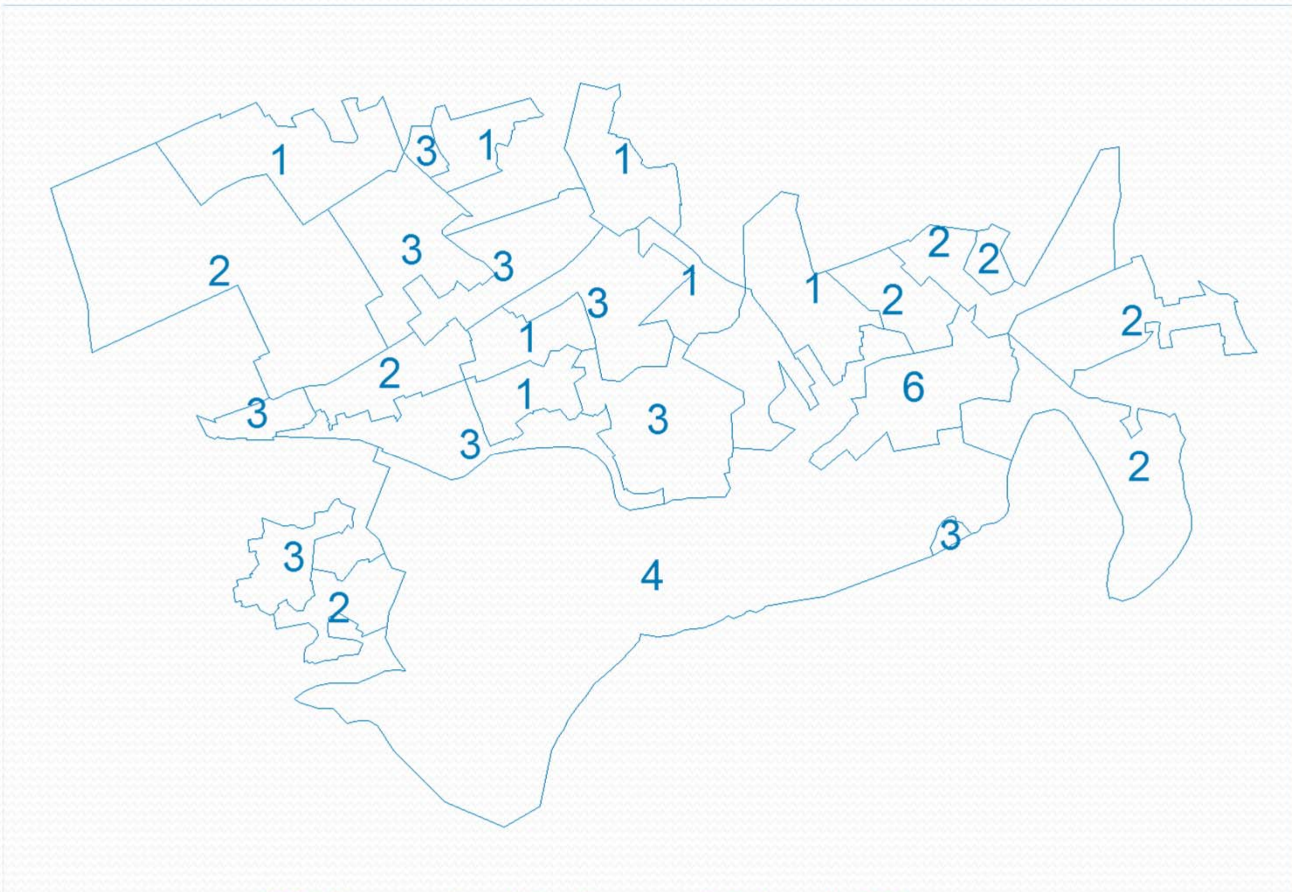
Background

- Small area health data:
  - Context: geo-referenced health outcomes

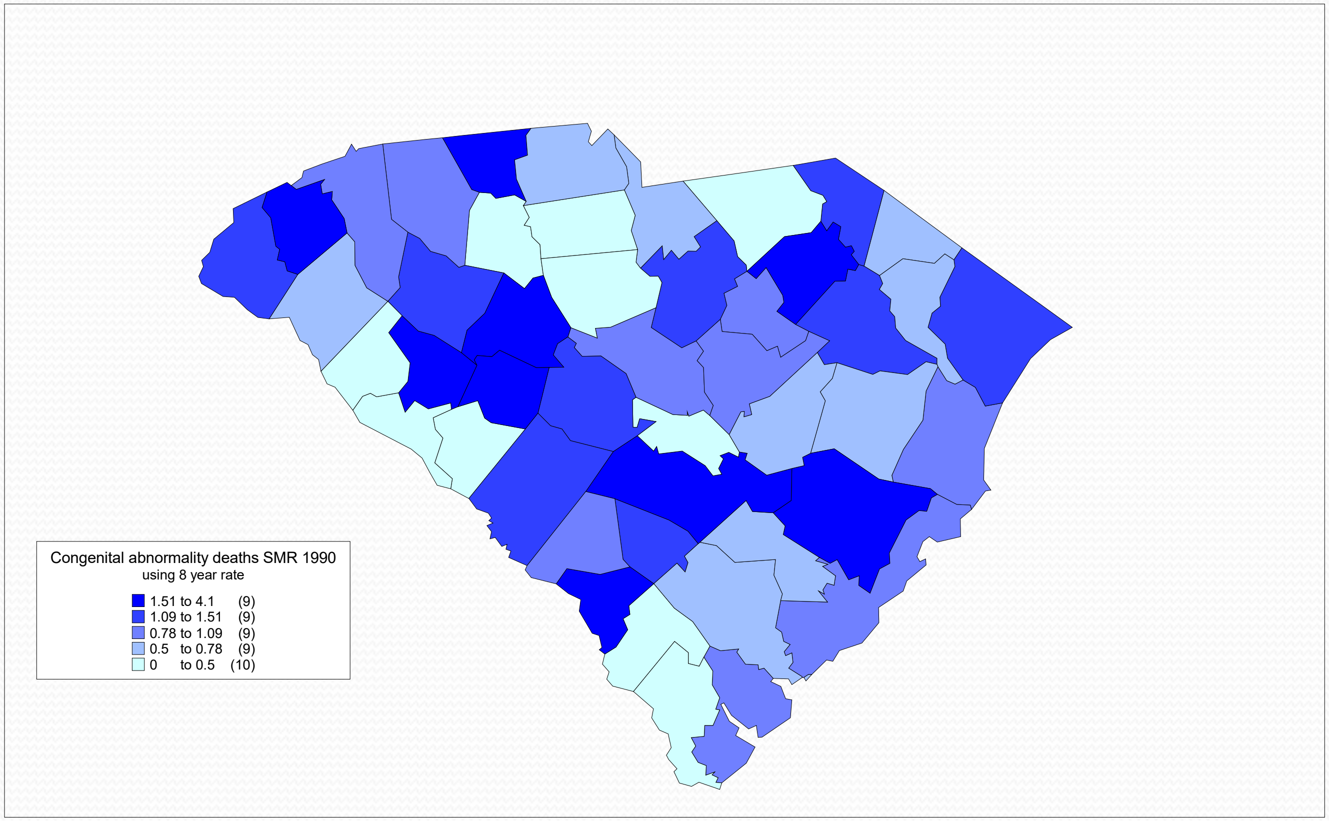


Larynx cancer incidence in Lancashire NW England for the period

26 Census tracts in Falkirk, Central Scotland: counts of respiratory cancer deaths 1978 - 1983



South Carolina congenital deaths 1990



Environmental stressors and aetiology

- Air pollution
  - Particulate matter PM 10, PM2.5, speciated versions
  - Mold, pollen
  - Toxics: PCBs, PFOS, heavy metals (arsenic, lead, mercury……)
- Soil pollution
  - Pesticide residues; arsenic, lead, mercury,……..
  - Long term concentrations/accumulations
- Water contamination
  - toxics, BOD, etc

Exposure pathways I

- Air pollution may relate to respiratory health outcomes and indeed often acute outcomes are strongly related to air events:
  - E.g. Anto et al (1993) Barcelona harbor soya bean dust events
- Proximity to sources of air pollution is important
  - Roadways
  - Incinerators
  - Docks
  - Power plants
  - Waste dump sites

Exposure Pathways II

- Some pathways are less obvious:
- Soil pollution:
  - Could be a *surrogate* for air pollution
  - Accumulation over time
  - Direct ingestion
  - Indirect ingestion: fruit and vegetables
  - Could be double exposures (air and soil)
- Water pollution
  - Direct ingestion
  - Mediator effects (eg infected fish)

Health outcomes

- Acute respiratory outcomes:
  - Asthma, COPD, URI
- Chronic exacerbation
  - COPD, CHD, MI, some cancers (larynx, lung, leukaemia, NHL), mental health
- Latency issues
  - Lag time before display of symptoms
  - affects cancers
  - Residential history becomes important when lag time is considered

Confounding effects

- Care must be taken to make sure that environmental stressors are not confounded with other effects…such as migration.
- Residential history is VERY important especially in long latency diseases
  - Sellarfield example
  - Rosyth example

Small environmental effects

"....many current methods of clustering often have poor power for detecting the small increases in risk often associated with environmental exposures." (after Waller, 1996)

- There is a need for reasonably precise tools for the detection of environmental insults.

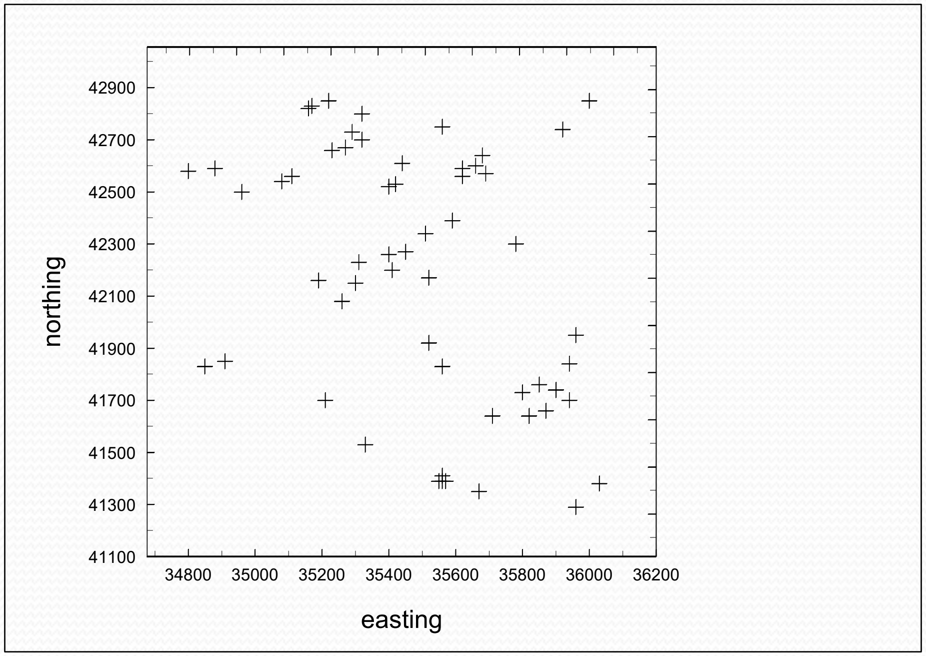
Examples

1. Putative source analysis with case event data
2. Ecological analysis: melanoma and sunlight
3. Mixtures: intellectual delay and soil metals

Case event data

- Count data is the commonest format found in spatial epidemiology
- However this is just an aggregation of case event data where the (residential) location of a case of disease is the primary data focus
- Often case event data is important when small spatial scales are of interest (1-10kms for example)

Example: larynx cancer in NW England



Case event notation

- Define the study area as T

s_i : x,y coordinate pair of the i th location

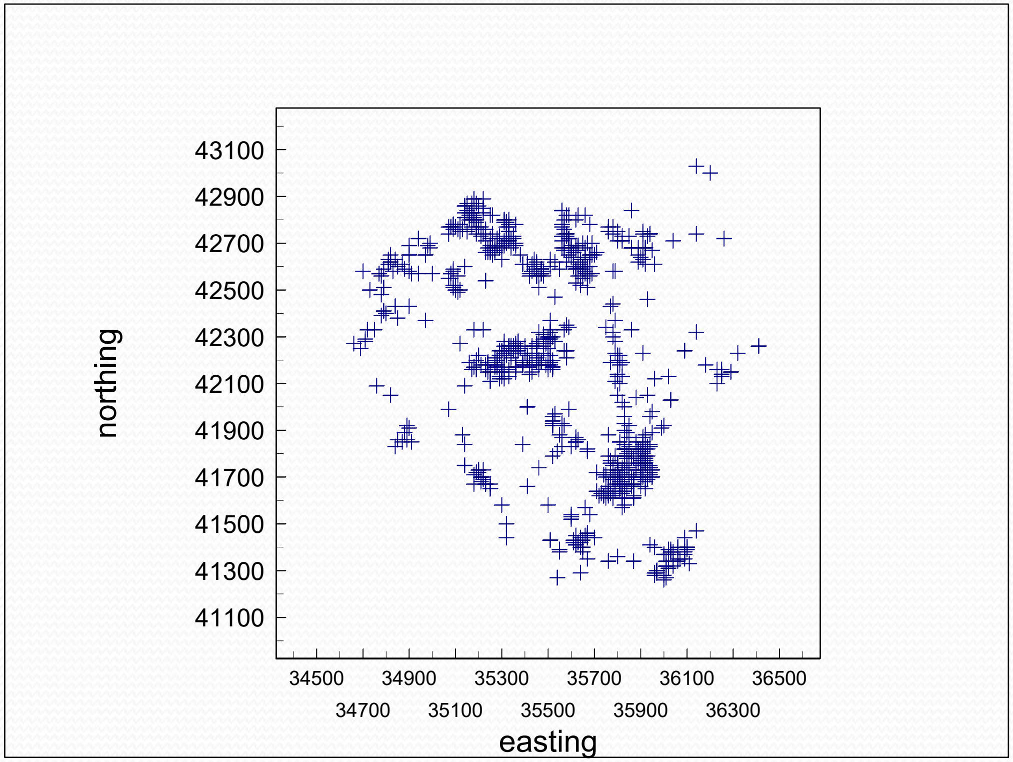
m events in T

$\{s_i\}$, $i = 1,....,m$

Control disease

- Usually the cases have associated with them a control disease realization
- This is used as a geographical control for the case distribution (acting like a expected count in the count data examples)

Control: lung cancer



Control notation

$$n \text{ control locations in T}$$

$$\{s_j\},\ j = m+1,\ m+n$$

- Hence we treat the controls and cases as one vector of length $m+n$

Conditional Logistic models

- Instead we use CONDITIONING to give us a simpler labeling approach
- Intensity of the case and control events is defined to be

$$
\begin{aligned}
&\mathit{control} : \lambda_0(s) \\
&\mathit{case} : \lambda_0(s)\lambda_1(s) \\
&\text{modeled part:} \\
&\lambda_1(s) = \exp(\eta(s))
\end{aligned}
$$

Conditional Logistic models

Assume that the complete vector is used for a binary label so that

$$y_i = \begin{cases} 1 & if \ \ s_i \in \{s_i\}, \ i = 1,\ldots,m \\ 0 & \quad\quad otherwise \end{cases}$$

- Hence, y_i is 1 for case and 0 for a control

Logistic spatial models

- Then:

$$y_i \sim Bern(p_i)$$

$$p_i = \frac{\lambda_1(s_i)}{1 + \lambda_1(s_i)}$$

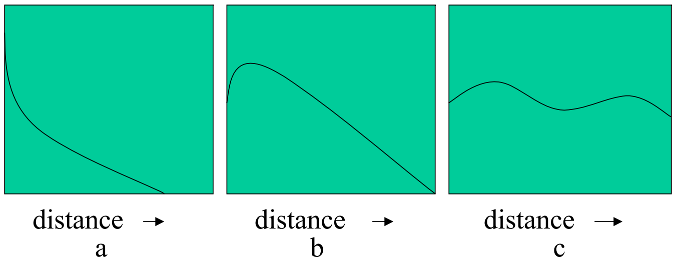
$$If \;\; \lambda_1(s_i) = \exp(x_i'\beta)$$

$$where \;\; x_i'\beta \text{ is a linear predictor}$$

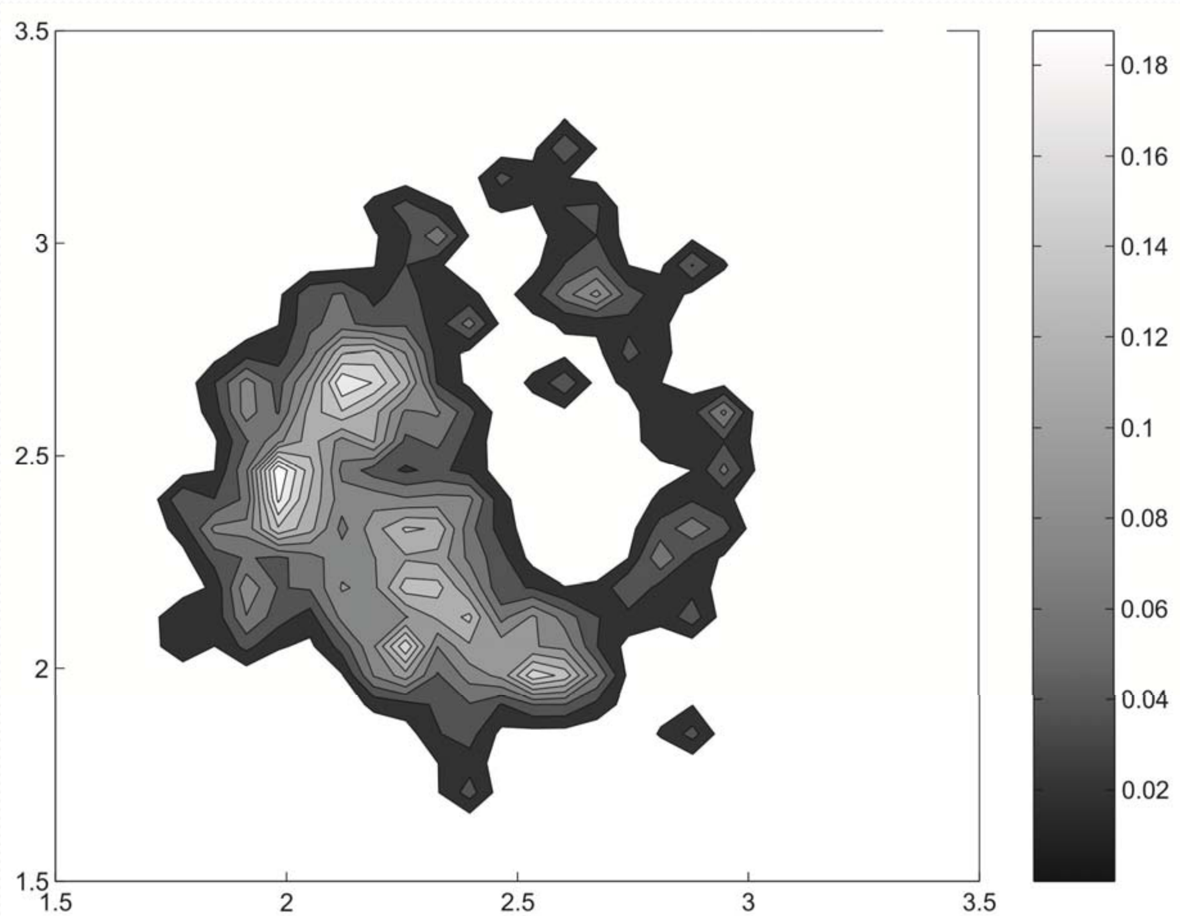
- This is just a logistic regression formulation
- Hence as long as you have covariate information at the locations of controls and cases you can assume a conditional logistic spatial model

Exposure Modeling

- How does incidence relate to source
  - Distance effect
  - Directional effect
  - More complex interaction



Plume simulation : Weibull-Von Mises spatial deposition map



Typical example

- Location (s), distance from a pollution source (d), age (x) as variables must be available for all cases and controls

$$\lambda_1(s_i) = \psi_0 \exp\{\gamma_1 d_i + \gamma_2 x_i\} = \exp\{\gamma_0 + \gamma_1 d_i + \gamma_2 x_i\}$$

$$d_i = \| s_i - s_0 \| \quad \text{distance from source}$$

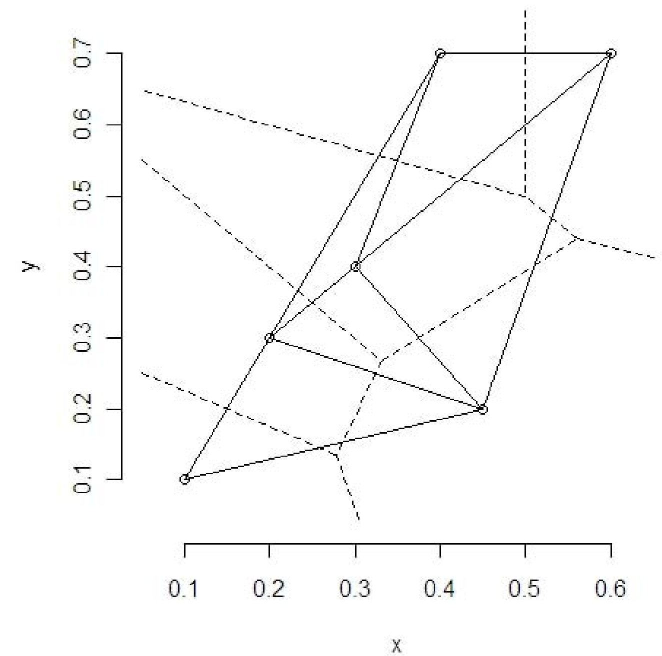
s_0 is the source location

Addition of Random effects

- It is easy to add various types of REs
- UH can be added via an individual level zero mean Gaussian effect:V~N(0,tau)
- CH is slightly different: A CAR model cannot be simply applied here
- Can use a CAR if you can defined neighborhoods?
- Otherwise must use a full MVN geostatistical model

Delauney Neighbors

X 0.1 0.2 0.4 0.45 0.6 0.3
Y 0.1 0.3 0.7 0.2 0.7 0.4
Num 2 4 3 4 3 4



Example

- Larynx and lung cancer (NW England)
- Variables: x, y, case indicator, distance, age
- Using Delauney neighbors to define adjacencies
- Distance is the exposure surrogate

Models

Model	DIC	pD
I D only	447.45	0.44
II D+V	439.74	41.12
III D+V+A	366.67	89.01
IV D+A	444.69	1.82
V D+V+U	447.4	5.67
VI D+V+U+A	352.94	118.10

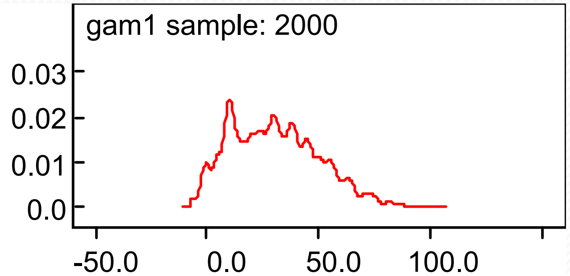
Lowest DIC is model VI

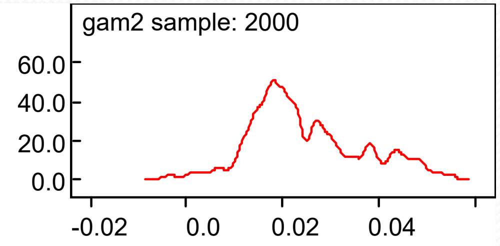
- D: distance; V: UH; U: CH; A: age

Model VI results

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
gam0	-7.623	1.475	0.2146	-10.64	-7.43	-5.53	10001	2000
gam1	30.56	19.92	1.523	-1.001	29.31	73.13	10001	2000
gam2	0.02479	0.01175	0.001538	0.003891	0.02232	0.04966	10001	2000



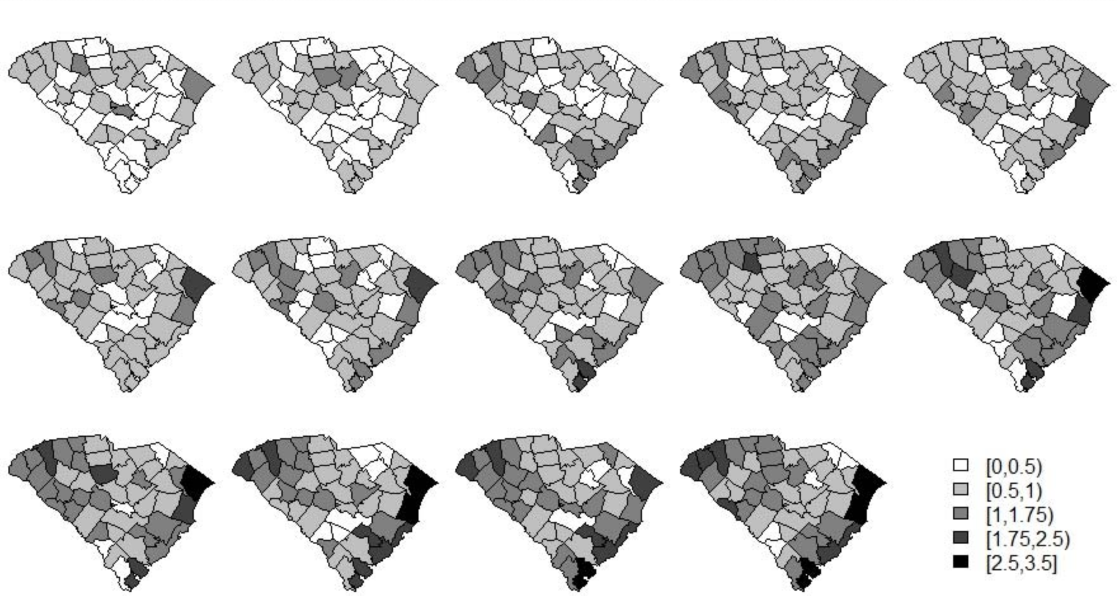


Reference

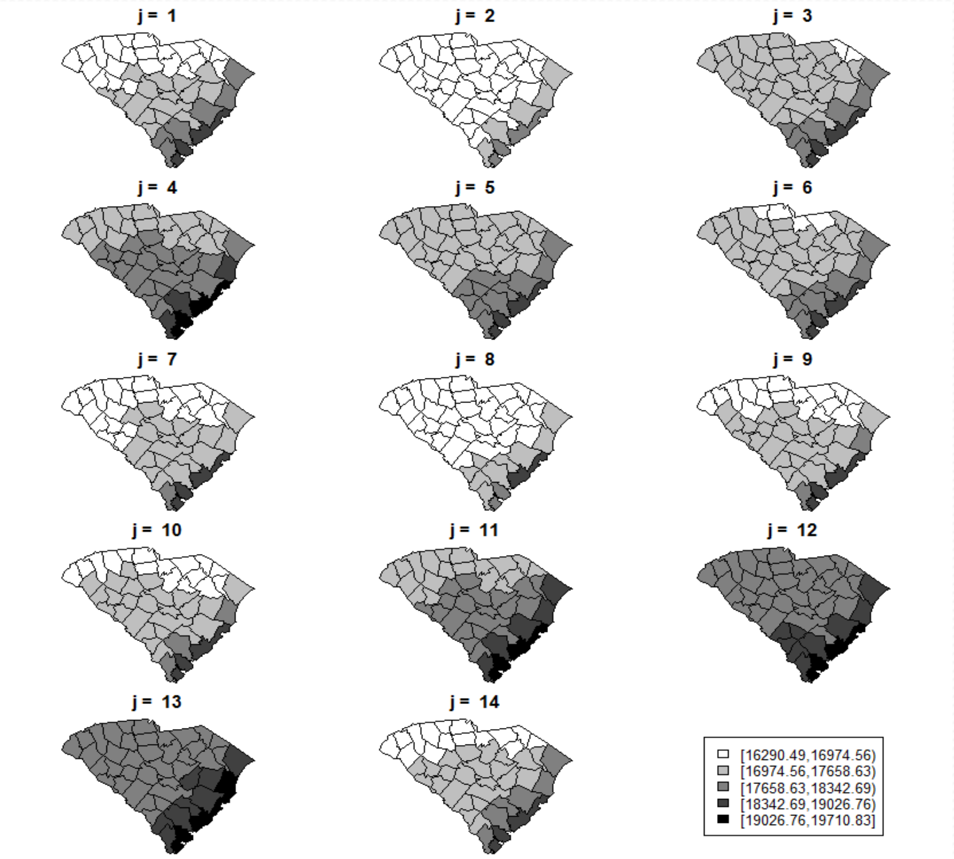
- Lawson, A. B. (2012) Bayesian Point Event Modeling in Spatial and Environmental Epidemiology. *Statistical Methods in Medical Research* 21, 5, 509-530

Model selection with environmental linkage

- Melanoma and sunlight in SC (14 years: 1996-2009)
  - Ecological relation between sunlight exposure and melanoma incidence hypothesized



Average daily sunlight (KJ/m^2)



Models

- Additional confounders and unobserved effects
  - unemployment percentage (ST)
  - Percent African American (S)
  - Median income (S)
  - Random effects (S, T and ST)

Selected models for melanoma

Model		Model Probabilities			Favored Linear Predictor
F2	Spatial $p_{[illegible]}$	0.41 (0.04,0.88)	0.29 (0.01,0.64)	0.29 (0.01,0.64)	$\alpha_1 x_{1i} + \alpha_2 x_{2i}$
	Spatio-Temporal	0.29 (0.01,0.66)	0.30 (0.01,0.65)	0.40 (0.03,0.86)	$\alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j$
F3		0.33 (0.02,0.69)	0.34 (0.02,0.68)	0.34 (0.01,0.68)	NA
F4		0.17 (0.16,0.18)			$\alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j$

F4 general model

- Log risk

$$\log\left(\theta_{ij}\right) = \alpha_0 + p\omega_i^S + (1-p)\omega_{ij}^{ST} + \psi_{ij}$$

- With

$$\omega_i^S = \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i$$
$$\omega_{ij}^{ST} = \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j$$

Parameter estimates: F4 model

α_{31}	**-0.067**
α_{32}	-0.106
α_{33}	-0.143
α_{34}	-0.167*
α_{35}	-0.154
α_{36}	-0.213*
α_{37}	-0.141
α_{38}	-0.141
α_{39}	-0.197*

- This model consist of two space-time covariates (sunlight and unemployment rate) and a temporal random effect.
- Posterior mean estimates of the sunlight parameter for 9 years (1999-2007)
- Mostly poorly estimated and skewed.

Reference

- Carroll, R., Lawson, A.B. et al (2016) Spatio-temporal Bayesian model selection for disease mapping

Environmetrics (to appear)

Mixtures

- Many environmental health problems involve the assessment of the effect of *multiple* predictors and hence mixture of exposures.
  - Particulate matter can be speciated into multiple different types: size based or chemical composition

  (PM_{10},$PM_{2.5}$,……carbon, etc)
  - Soils can have pesticide metal residues (>20 chemicals)

The National Institute of Environmental Health Sciences (NIEHS) has a major focus in mixtures:

http://www.niehs.nih.gov/research/supported/exposure/mixtures/index.cfm

Mixture modeling

- First issue is that of multicolinearity:
  - Some predictors can mask effects of other predictors
  - Unique contribution of each predictor can be difficult to assess when multiple predictors are present.
  - Possible approaches could be orthogonal projection/decomposition, grouping, or Bayesian adaptive regression trees (BART)
- Second issue is that of sequencing:
  - If you have measured multiple exposures *in sequence* how do you assess the unique effect of these on a single health outcome (i.e. the outcome is *not* longitudinal)
  - This typically is the problem with residential history studies or pregnancy exposures?

Classic example

- Outcome is normal or abnormal birth (binary)
  - Sequence of exposures during pregnancy is available but no longitudinal outcome data
  - A number of studies are examining this type of situation
  - One such attempt to assess monthly impact of metal exposure from soils was:

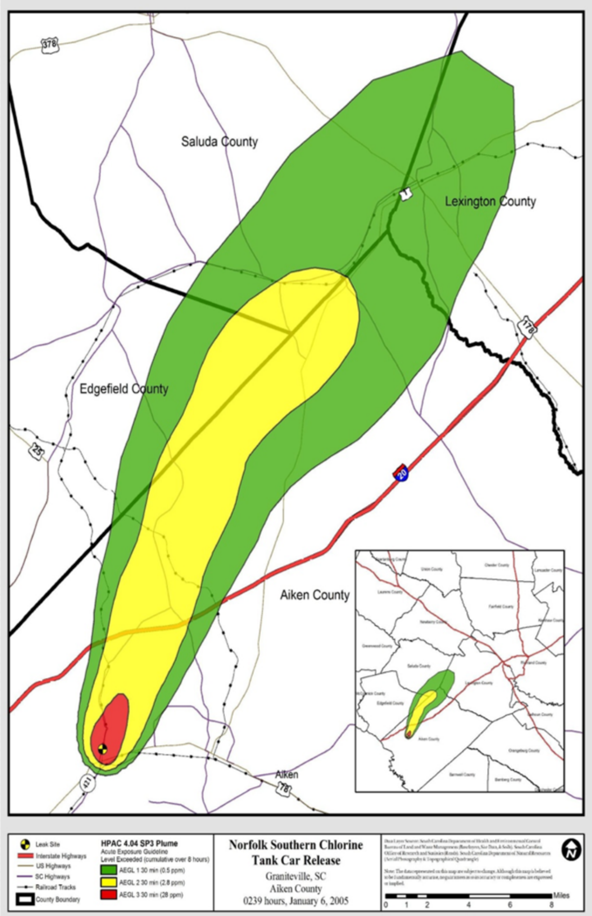
Onicescu, G., Lawson, A. B., et al (2014) Bayesian Importance parameter modeling of misaligned Predictors: soil metal measures related to residential history and intellectual disability in children. *Environmental Science and Pollution Research*., DOI: 10.1007/s11356-014-3072-8

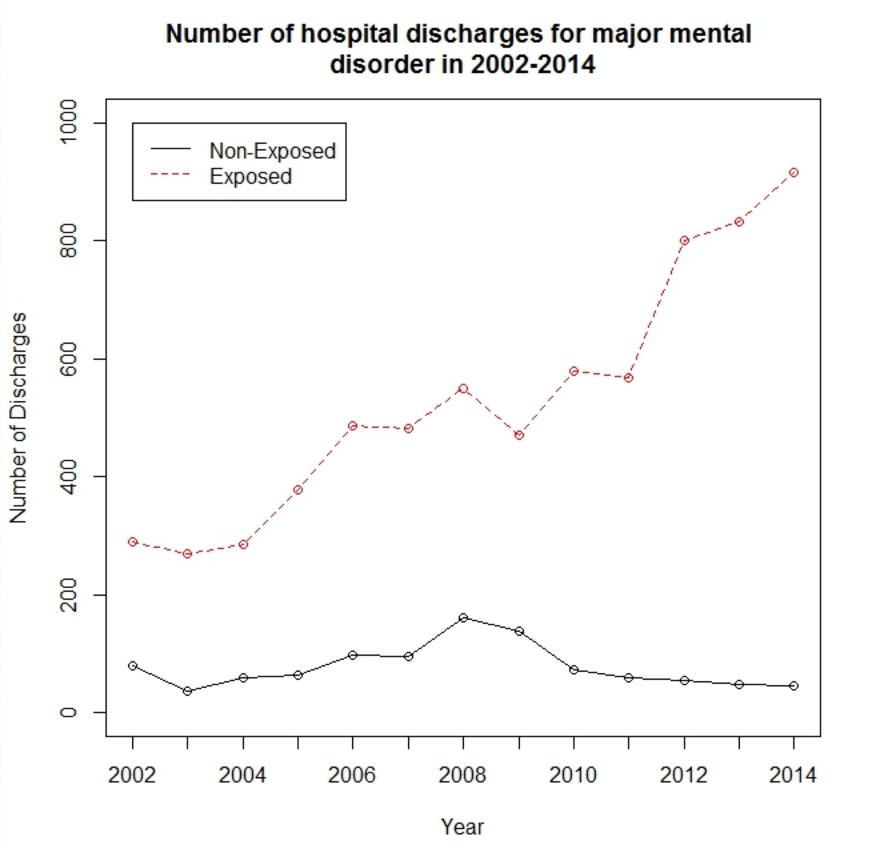
But Finally

Large environmental effects

- Although small effects are pervasive there are examples of large exposures where health outcomes are dramatic.
- Example: Graniteville chlorine gas disaster (2005)
  - On January 6th 2005, a 16 car train derailment led to an estimated release of 54,480 kg of liquid chlorine in Graniteville, South Carolina, USA. Over 5,000 residents were evacuated within one mile of the accident.
  - Nine deaths were initially reported, 71 individuals hospitalized, and at least 529 people were treated and released from local emergency departments.
  - In total, 1,384 casualties have been identified

Graniteville chlorine disaster (2005)





Summary

- Environmental exposure is often linked to a range of health outcomes
- These outcomes could be acute (asthma) or chronic (cancer or mental health ?)
- Spatial methods can be important in assessing (particularly chronic) exposures.
- Spatio-temporal methods should also be considered
- Mixtures of exposures is a major and challenging methodologic area
- Occasionally large environmental effects also lead to large scale health effects.