Disease mapping with messy data

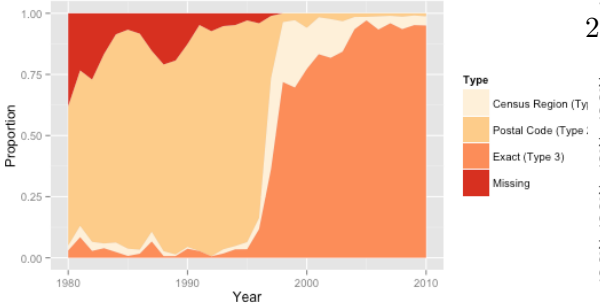
Patrick Brown

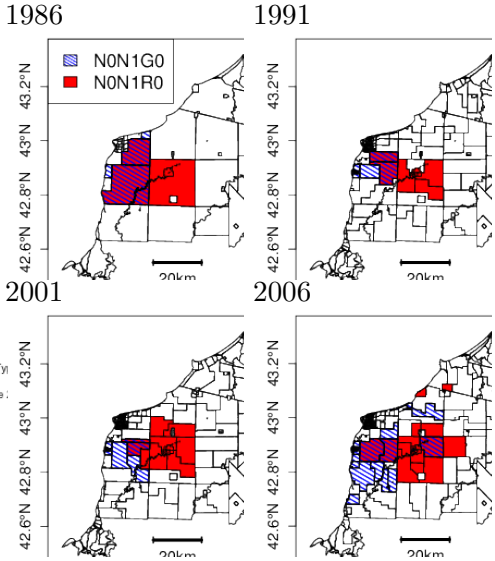
Cancer Care Ontario and University of Toronto

June 29, 2014

The Problem

- Estimate cancer risk at different locations
- mixture of exact locations, postal codes, and census regions
- completeness changes over time
- . . . as do boundaries

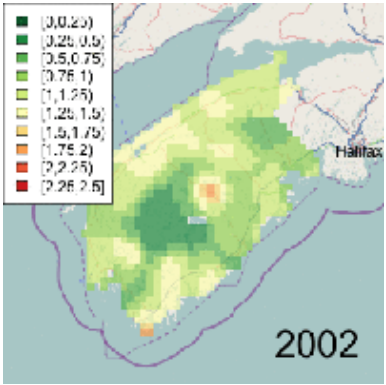




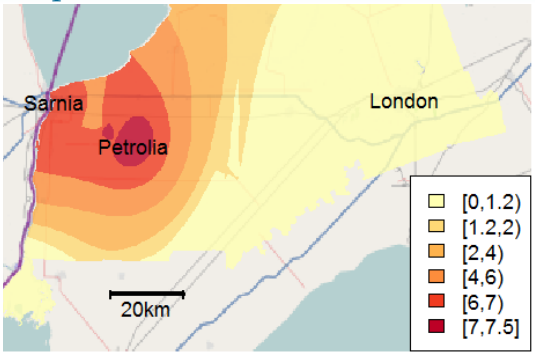
A ‘missing data’ problem

- Exact locations are ‘missing’ or ‘latent’ or ‘censored’
- **Hierarchical model:**

risk ⇒ case locations ⇒ data observed



a



- integrate out unknown locations
- Local-EM algorithm:
  pr(obsevations | risk)
- Bayes/MCMC:
  pr(risk | observaitons)

Computational issues

- Algorithms are demanding, must be run in parallel
- local-EM is naturally parallelizable
- MCMC less so. RMHMCMC is somewhat
- spatio-temporal: extremely high dimensions

Extensions

- Uncertainty in case ascertainment
- Uncertainty in covariates
- Model extensions:
  - multivariate data,
    i.e. two types of cancer
  - mixture models,
    normal v hazardous
  - directional effects