A Bayesian Spatial-Temporal Model for $PM_{2.5}$ and Mortality

Kenny Flagg

Introduction

- Particulate matter — solid particle pollutants suspended in air
- Human-generated particles generally less than 2 μm in diameter; naturally-occuring particles are larger (Charlesworth, De Miguel, and Ordóñez 2011)
- Two sizes are studied: PM_{10} (< 10 μm) and $PM_{2.5}$ (fine particulate matter, < 2.5 μm)
- Exact chemical composistion varies
- Concern about human health effects of As, Cd, Cr, Hg, Mn, Ni, Pb, V, and others in $PM_{2.5}$

Bayesian Models

- Choi, Fuentes, and Reich (2009) use a Bayesian spatiotemporal model for daily mortality counts in North Carolina in 2001
    - $PM_{2.5}$ concentration as risk factor — several timescales
    - Control for age, gender, race/ethnicity, weather
    - Use all recorded natural and cardiovascular deaths
- Test "harvesting hypothesis" — short-term exposure affects frail individuals more than healthy individuals
- They ultimately use the model to estimate relative risk

Bayesian Models

- Use data to update prior distribution into posterior distribution
- Two-stage model
  - Stage 1: Spatiotemporal model for $PM_{2.5}$ with meteorological covariates
  - Stage 2: Generalized Poisson model for daily mortality counts by county, with demographic and meterological covariates
  - Posterior prediction from stage 1 used as prior for stage 2

Bayesian Models

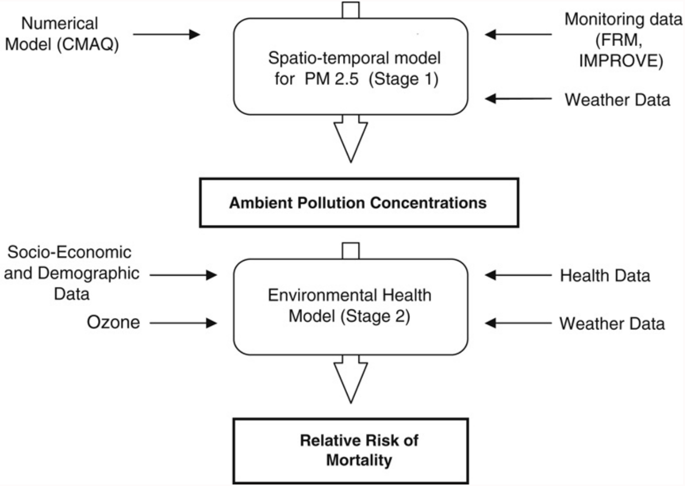

Figure 1 from Choi et al.

Stage 1: $PM_{2.5}$ Model

- Measurement error model
  - $PM_{2.5}$ measured in $\mu g/m^3$ at location $\mathbf{s}$ and day t
  - Observed: $\widehat{Z}(\mathbf{s},t) = Z(\mathbf{s},t) + e(\mathbf{s},t)$
    - $e(\mathbf{s},t) \overset{iid}{\sim} N(0,\sigma^2)$
  - Latent: $Z(\mathbf{s},t) = \mathbf{M}^T(\mathbf{s},t)\zeta + e_Z(\mathbf{s},t)$
    - $e_Z(\mathbf{s},t)$ normal with mean $\psi_Z e_Z(\mathbf{s},t-1)$ and covariance $Cov(e_Z(\mathbf{s},t), e_Z(\mathbf{s}',t)) = \sigma_Z^2 \exp(-||\mathbf{s}-\mathbf{s}'||/\phi_Z)$
  - $\mathbf{M}(\mathbf{s},t)$ is a vector of weather covariates, ζ is a vector of model coefficients
- Priors
  - $\sigma \sim Unif(0,5)$, $\sigma_Z \sim Unif(0,100)$
  - $\psi_Z \sim N(0,10)$, $\phi_Z \sim Unif(0,500)$
  - Prior for ζ not mentioned

Stage 1: $PM_{2.5}$ Model

- Averaged over area of county to compute $Z_j(t)$, the $PM_{2.5}$ concentration for county j on day t
- Fourier transform to decompose $Z_j(t)$ time series into 5 different timescales, $Z_{jl}(t)$, $l = 1, 2, 3, 4, 5$
- A similar "stage 0" model predicted $\mathbf{M}(\mathbf{s}, t)$ at the locations where $PM_{2.5}$ was measured

Stage 1: $PM_{2.5}$ Model

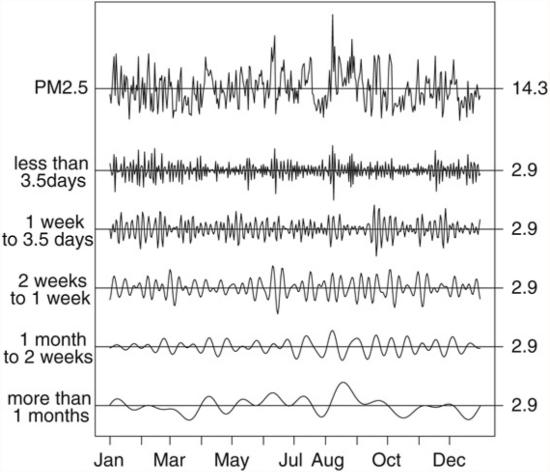

Figure 4a from Choi et al.

Stage 2: Daily Mortality Model

- Generalized Poisson model (Famoye 1993)
  - Observe mortality count $Y_j(t)$ in county j on day t
  - $Y_j(t)$ has mean $\mu_j(t)$ and variance $\mu_j(t)[1+\alpha\mu_j(t)]^2$
  - $\log(\mu_j(t)) = \gamma_j + \mathbf{x}_j^T(t)\beta + S(\mathbf{M}_j(t))$
    - γ_j normally distributed random effect with mean μ_{γ_j}
    - The γ_j have a conditional autoregressive structure (CAR) that depends on the adjecency matrix of the counties, with scale σ_γ and reaction parameter ρ (Banerjee, Carlin, and Gelfand 2004)
  - $\mathbf{x}_j(t)$ a vector including the $Z_{jl}(t)$, demographic variables, and interactions; β a vector of coefficients
  - $S(\mathbf{M}_j(t))$ a smooth spline function of the meteorological variables

Stage 2: Daily Mortality Model

- Priors
  - $\mu_{\gamma_j} \stackrel{iid}{\sim} N(0, 100)$
  - $\sigma^2_\gamma \sim InvGamma(0.5, 0.0005)$
  - ρ uniform with bounds that "guarantee that the variance matrix [is] positive definite"
  - β uses a spacial case of multivariate CAR, the multivariate intrinsic autoregressive (MIAR) prior (Gelfand and Vounatsou 2003)
  - α prior not mentioned

Model Fitting

- Markov chain Monte Carlo implemented using WinBUGS and R
- Two chains run for 2,000 iterations each (plus 3,000 iterations burn-in)
- Took "a couple of days to run"
- Convergence assessed with $\widehat{R}$, autocorrelation functions, trace plots

Results

- Reported posterior mean and SD of log relative risk by season and timescale
- $\log(RR) = 1{,}000\beta_{jlk}$ is the percent increase in mortality rate per 10 $\mu g/m^3$ increase in $PM_{2.5}$ concentration
- RR larger in winter and summer than in spring and fall
- RR larger for long timescales than short timescales
- The possible confounders did not have large interactions with $PM_{2.5}$ or strong main effects
  - No evidence of harvesting

Results

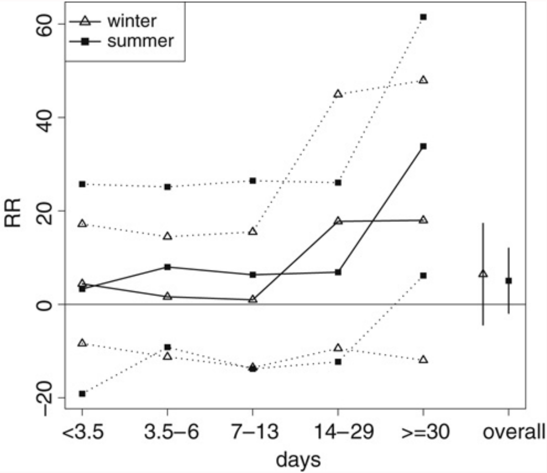

Figure 5d from Choi et al.

Discussion

- Tradeoffs between population-level inference and understanding effects on individuals
- Broad populations
  - Most other studies of particulate exposure also use complicated Bayesian models
  - $PM_{2.5}$ measured at monitoring stations may be a poor proxy for individual exposure (Özkaynak et al. 2013)
- Narrow populations
  - Riediker et al. (2004) studied young male NC highway patrol officers with monitoring devices in patrol cars
- Can't connect specific chemicals to mortalities
- Good starting point for more directed investigations

References

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2004). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall.

Charlesworth, S, E De Miguel, and A Ordóñez (2011). "A review of the distribution of particulate trace elements in urban terrestrial environments and its application to considerations of risk". In: *Environmental Geochemistry and Health* 33.2, pp. 103–123.

Choi, Jungsoon, Montserrat Fuentes, and Brian J Reich (2009). "Spatial–temporal association between fine particulate matter and daily mortality". In: *Computational Statistics and Data Analysis* 53.8, pp. 2989–3000.

Famoye, Felix (1993). "Restricted generalized Poisson regression model". In: *Communications in Statistics-Theory and Methods* 22.5, pp. 1335–1354.

Gelfand, Alan E and Penelope Vounatsou (2003). "Proper multivariate conditional autoregressive models for spatial data analysis". In: *Biostatistics* 4.1, pp. 11–15.

Özkaynak, Halûk et al. (2013). “Air pollution exposure prediction approaches used in air pollution epidemiology studies”. In: *Journal of Exposure Science and Environmental Epidemiology* 23.6, pp. 566–572.

Riediker, Michael et al. (2004). “Particulate matter exposure in cars is associated with cardiovascular effects in healthy young men”. In: *American Journal of Respiratory and Critical Care Medicine* 169.8, pp. 934–940.