Right Map, Right Model

Two examples using the BYM2 model

Mitzi Morris

2025-05-30

Stan Development Team

Talk Overview

- The Besag York Mollié model: workhorse of spatial epidemiology.
- Computational Concerns:
  - Coding the BYM2 model.
  - Getting the map you want from the geo-coordinates you have.
- Example: accounting for time plus roads, rails, and airplanes.
- Example: accounting for disconnected regions.

Spatial Smoothing For Areal Data

- Counts of (rare) events in small-population regions are noisy
  - Borrow information from *neighboring* areas
- Besag, 1973, 1974: Conditional Auto-Regressive Model (CAR), and **Intrinsic Conditional Auto-Regressive** (ICAR) model
  - Gaussian Markov Random Field (GMRF)
- CAR model has parameter α for the amount of spatial dependence;
  - CAR model requires computing matrix determinants
  - Cubic operation ($\mathrm{O}N^3$) (calculated at every *step* of the sampler)
- ICAR simplification: let $\alpha == 1$ (complete spatial dependence)
  - ICAR model computes pairwise distance between neighbors
  - Linear operation ($\mathrm{O}N$)

Neighbor Relationship

- The binary neighbor relationship (written $i \sim j$ where $i \neq j$) is encoded as
  - 1 if regions n_i and n_j are neighbors
  - 0 otherwise
- For ICAR models, neighbor relationship is
  - symmetric - if $i \sim j$ then $j \sim i$
  - not reflexive - a region is not its own neighbor ($i \nsim i$)
- Many possible definitions of "neighbor"
  - 'rook' - areas share a bounding line
  - 'queen' - areas share a boundary point
  - *mobility networks* - amount of travel between areas

Intrinsic Conditional Auto-Regressive (ICAR) Model

- Conditional specification: multivariate normal random vector ϕ,
  each ϕ_i is conditional on the values of its neighbors
- Joint specification rewrites to *Pairwise Difference*:

$$p(\phi) \propto \exp\left\{-\frac{1}{2}\sum_{i\sim j}(\phi_i - \phi_j)^2\right\}$$

  - Centered at 0, assuming common variance for all elements of ϕ.
  - Each $(\phi_i - \phi_j)^2$ penalizes the distance between the *values* of neighboring regions.
- ϕ is non-identifiable - adding a constant to ϕ washes out of $\phi_i - \phi_j$
- Sum-to-zero constraint centers ϕ

Besag York Mollié (1991) BYM Model

- Basic regression Lognormal Poisson regression plus 2 components for spatial smoothing
  - spatial effects: ICAR component
  - random effect: per-region standard Normal
- Formula: $\mu + x\beta + \phi + \theta$
  - μ is the fixed intercept.
  - x is the design matrix, β is vector of regression coefficients.
  - ϕ is an ICAR spatial component
  - θ is an vector of ordinary random-effects components.

BYM2 model: Riebler et al, 2016

- *Penalized Complexity* (Simpson et al, 2014)
  - **spatial and random effects must have** $\text{Var}(\phi_i) \approx 1$
  - combined spatial and random effects have mixing parameter, scale parameter, (following Leroux, 2000)
- BYM2 model:

$$\mu + x\beta + \left((\sqrt{\rho/s}\ )\,\phi^* + (\sqrt{1-\rho})\,\theta^* \right) \sigma$$

where:

  - $\sigma \geq 0$ is the overall standard deviation.
  - $\rho \in [0, 1]$ - proportion of spatial variance.
  - ϕ^* is the ICAR component.
  - $\theta^* \sim N(0, 1)$ is the vector of ordinary random effects
  - s is a scaling factor s.t. $\text{Var}(\phi_i) \approx 1$, computed from neighbor graph, i.e. s is *data*, not a parameter.

Coding Challenge: from Math to Model

$$\mu + x\beta + \left(\left(\sqrt{\rho / s} \;\right) \phi^* + \left(\sqrt{1 - \rho} \right) \theta^* \right) \sigma$$

```
parameters {
  real beta0; // intercept
  vector[K] betas; // covariates
  real<lower=0, upper=1> rho; // proportion of spatial variance
  sum_to_zero_vector[N] phi; // spatial effects
  vector[N] theta; // heterogeneous random effects
  real<lower = 0> sigma;  // scale of combined effects
}
model {
  vector[N] gamma = sqrt(rho / tau) * phi + sqrt(1 - rho) * theta;
  y ~ poisson_log(log_E + beta0 + xs_centered * betas + gamma * sigma);
  target += -0.5 * dot_self(phi[neighbors[1]] - phi[neighbors[2]]); // ICAR prior
  beta0 ~ std_normal();   betas ~ std_normal();
  rho ~ beta(0.5, 0.5);  theta ~ std_normal();  sigma ~ std_normal();
}
```

Stan ICAR Model

Encode neighbor information as graph edgeset, i.e. pairs of indices for neighbors i, j:

```
data {
  int<lower = 0> N;  // number of areal regions
  int<lower = 0> N_edges;  // number of neighbor pairs
  array[2, N_edges] int<lower = 1, upper = N> neighbors; // columnwise adjacent
  ...
```

Use Stan's `sum_to_zero_vector` to identify `phi`

```
parameters {
  sum_to_zero_vector[N] phi; // spatial effects
  ...
```

Use Stan's vectorized operations and multi-indexing to compute ICAR prior

```
model {
  target += -0.5 * dot_self(phi[neighbors[1]] - phi[neighbors[2]]);
  ...
```

Stan Constrained Parameter `sum_to_zero_vector`

```
sum_to_zero_vector[K] beta;
```

- On the unconstrained scale, beta is length K - 1 (because the last is determined by the first K - 1).
  - Stan computes on the *unconstrained* scale.
  - Constraining transform keeps variance of all elements equal.
- Outperforms "hard" and "soft" sum-to-zero constraints
  - "hard" sum-to-zero: $N^{th} element == - sum(elements 1 : N-1)
  - "soft" sum-to-zero: constrain `sum(beta) ~ Normal(0, epsilon)` epsilon = 0.001
- The larger, more complex the model, the bigger the difference.
- See Stan case study The Sum-to-Zero Constraint in Stan

From Geospatial Maps to Neighbor Graphs

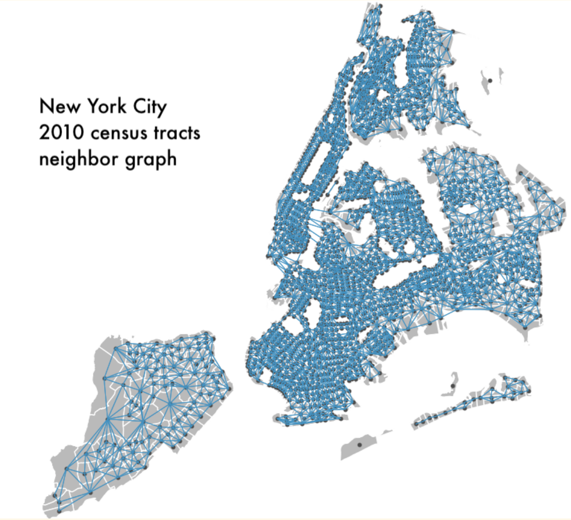
- GIS data: points, lines, and bounding polygons
  - Common format: shapefile
  - Python: `GeoPandas` - add support for geographic data to pandas objects.
  - R: `sf` Simple Features for R - "spatial analysis simplified"
- Neighbor graph
  - Graph nodes denote regions, edges denote neighbors
  - GDAL-based utilities compute neighbor graph from geo-dataframe `geometry` column.
- Shapefiles may not line up with your data.
  - Shapefile region IDs may differ from dataset IDs.
  - Map boundaries may differ from idealized boundaries.
- Neighbor graphs are difficult to edit.
  - Easy to get node indices wrong.
  - Difficult to maintain symmetry

Example Dataset: NYC traffic accidents

Dataset taken from Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan.

- Aggregated counts of car vs. child pedestrian traffic accidents, localized to US Census tract. Per-tract data includes:
  - raw counts of accidents, population
  - measures of foot traffic, car traffic
  - socio-economic indicators: median income, neighborhood transiency
- Starting point: NYC Planning Census Blocks
  - " These boundary files are derived from the US Census Bureau's TIGER data products and have been geographically modified to fit the New York City base map."

NYC Census Blocks



- Study focus is on *pedestrians*
- Map doesn't respect geography
  - East River separates Manhattan from Brooklyn, Queens
- Geography doesn't respect ICAR model
  - Graph must be fully connected - else elements of ϕ are undefined.

Ward et al Mobility Network Graph

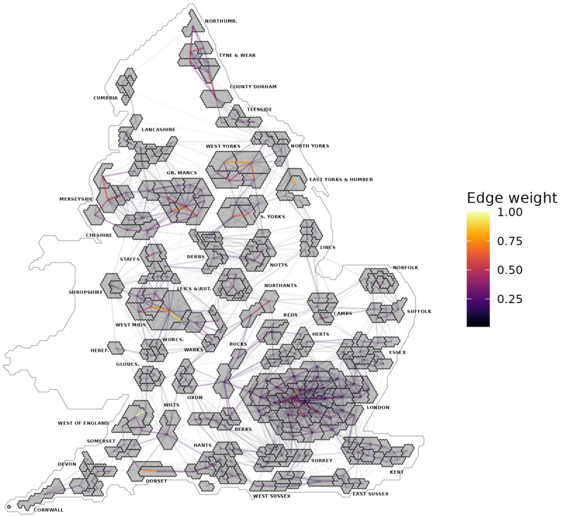


Fig 1. The mobility network overlaid on a hex map of England, where the edge weight describes the rate of journeys between different LTLA's, normalised to take values in [0,1]. Created from geographical files using the House of Commons Library, which is under the Open Parliament License v3.0.

- Online: Fig 1
- Northern england shows divide between east and west coasts

BYM2 Model for Spatio-Temporal Smoothing

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Bayesian spatial modelling of localised SARS-CoV-2 transmission through mobility networks across England

Thomas Ward, Mitzi Morris, Andrew Gelman, Bob Carpenter, William Ferguson, Christopher Overton, Martyn Fyles

Version 2 Published: November 13, 2023 • https://doi.org/10.1371/journal.pcbi.1011580

- Neighbor graph reflects *mobility networks*
- Extend the BYM2 model to include a daily-level random effect, i.e. three components:
  - Spatial effect θ_i
  - Ordinary per-region random effect ϕ_i
  - Ordinary per-region, per_day random effect $\phi_{i,t}$
  - Proportion of variance ρ is a simplex

Ward et al Simulation Study

Bayesian spatial modelling of localised SARS-CoV-2 transmission through mobility networks across England

Table 3

The average error for the naïve estimate (%), BYM2 (%), and total reduction in error (%) for the simulation scenarios.

Scenario	Average Error for Naïve Estimate (%)	Average Error for BYM2 Estimate (%)	Reduction in Error (%)
1	7.41	2.42	67.3
1.1	15.11	3.48	77.0
1.2	10.04	5.81	42.1
1.3	16.76	6.30	62.4
2	5.79	3.63	37.3
2.1	11.03	5.20	52.9
2.2	8.03	5.53	31.1
2.3	12.73	6.46	49.2
3	2.66	2.00	24.6
3.1	5.50	3.21	41.6
3.2	3.82	2.90	24.2
3.3	5.45	3.41	37.4

2017 BYM2 Case Study - Fully Connected Map

- Map editing process
  - Identify pairs of tracts to connect
  - Write utility functions to munge data structures
- Extremely tedious and error prone
- Must be scripted for reproducibility

BYM2 Multicomp Model

- Make the model respect the geography
- Extends the BYM2 model to account for disconnected graphs and islands, following the recommendations from A note on intrinsic Conditional Autoregressive models for disconnected graphs, Freni-Sterrantino et.al. 2018.
  - Component nodes are given the BYM2 prior
  - Singleton nodes (islands) are given a standard Normal prior
  - Compute per-connected component scaling factor
  - Impose a sum-to-zero constraint on each connected component
- See R and Python notebooks from GeoMED 2024 Workshop
  - GitHub Repository: mitzimorris/GeoMED_2024
  - Notebook "h6_bym2_multicomp"

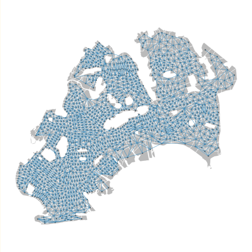
Scaling factor, following Riebler

- Geometric mean of the variances of the spatial covariance matrix
- Neighborhood structure matrix is the *precision* matrix
- Expensive to compute, but as data, is only done once

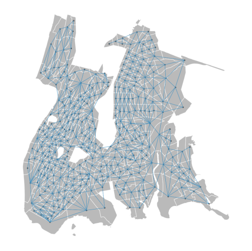
```
get_scaling_factor = function(nbs) {
    # Create ICAR precision matrix
    N = length(nbs)
    adj_matrix = nb2mat(nbs,style="B")
    Q =  Diagonal(N, rowSums(adj_matrix)) - adj_matrix

    # Get covariance matrix, compute the geometric mean of diagonal
    Q_pert = Q + Diagonal(N) * max(diag(Q)) * sqrt(.Machine$double.eps)
    Q_inv = q_inv_dense(Q_pert, adj_matrix)
    return(exp(mean(log(diag(Q_inv)))))
}
```

Different Maps Have Difference Scaling Factors

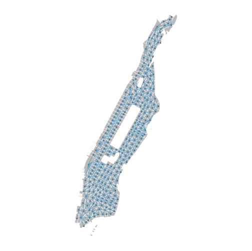
Brklyn-Queens: 0.71



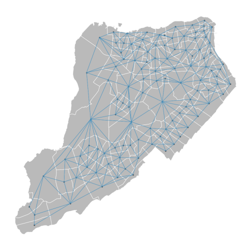
Bronx: 0.57



Manhattan: 0.8

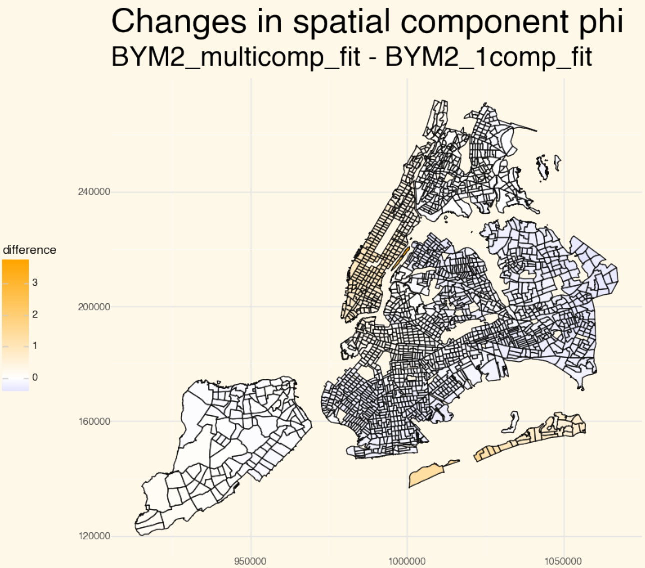


Staten Island: 0.36



Fully connected graph has scaling factor of 1.3

Comparison of ICAR component estimates



- Systematic differences due to per-component scaling factors
- Minor differences from removing spurious connections

References, Thanks!, and Questions?

- Simpson et al., 2014: Penalising model component complexity: A principled, practical approach to constructing priors
- Riebler et al., 2016: An intuitive Bayesian spatial model for disease mapping that accounts for scaling
- Freni-Sterrantino et al.,2018: A note on intrinsic conditional autoregressive models for disconnected graphs
- Morris et al., 2019: Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan
- Ward et. al., 2023 Bayesian spatial modelling of localised SARS-CoV-2 transmission through mobility networks across England