Spatial statistics has two main branches.

Point processes: Analyses of *locations where specific events occur*.

Example: Locations of trees of a given species.

Spatially-referenced data: Analyses of *data measured at known locations*.

Example: Concentrations of a soil pollutant at specific points.

Example: Counts of stomach cancers in each county of Minnesota.

We'll talk about the second branch.

This is another class of models that can be represented as mixed linear models at the price of pruning away parts of an extensive literature.

Analyses of spatially-referenced data

Distinctive concern: Spatial correlation of the measured quantity.

Models usually represent the vague intuition that measurements are more similar at near locations than distant locations.

Geostatistical (point-referenced) data:

- Measurements are taken at specific map coordinates.
- Interpolation is meaningful and is often the main goal.
- Such interpolation amounts to two-dimensional smoothing.

Areal data:

- Each measurement is a total or average over a region.
- Spatial nearness is specified by defining pairs of neighboring regions.
- Interpolation is no longer meaningful.
- Goal: Smooth a map or "borrow strength" to improve estimates.
- These analyses are now being applied to non-spatial problems.

Geostatistical models as mixed linear models

A very common form for this kind of model:

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s}), \text{ where}$$

$\mathbf{s}$ is a vector of coordinates in space,
$y(\mathbf{s})$ is a scalar measured at $\mathbf{s}$,
$\mathbf{x}(\mathbf{s})$ is a row vector of predictors at $\mathbf{s}$ plus intercept,
$w(\mathbf{s})$ is the realization of a scalar-valued Gaussian process at $\mathbf{s}$,
$\epsilon(\mathbf{s})$ is an error process, most often "white noise", iid $N(0, \sigma_e^2)$.

With observations at n locations $\mathbf{s}_i$, in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ form,
the $\mathbf{x}(\mathbf{s})$ are the rows of $\mathbf{X}$,
$\mathbf{Z} = \mathbf{I}_n$, $\mathbf{u}_i = w(\mathbf{s}_i)$, and $\mathbf{G}$ is implied by the GP for $\mathbf{w}(\mathbf{s})$,
$\mathbf{R} = \sigma_e^2 \mathbf{I}_n$.

Geostatistical models: What is **G**?

SAS's MIXED procedure (v. 9.2) offers 12 spatial forms for **G** and **R**.

The most common **G** in the literature (?) is the Matérn family:

$$G_{ij} = \sigma_s^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\delta_{ij}}{\rho} \right)^{\nu} \mathrm{K}_{\nu} \left(\frac{\sqrt{2\nu}\delta_{ij}}{\rho} \right),$$

Γ is the gamma function,

δ_{ij} is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$,

K_{ν} is the modified Bessel function of the 2^{nd} kind of order $\nu > 0$,

ρ scales distance,

ν controls the smoothness of correlation as a function of distance.

For $\nu = 0.5$ and $G_{ij} = \sigma_s^2 \exp(-\delta_{ij}/\rho)$.

As $\nu \to \infty$, $G_{ij} \to \sigma_s^2 \exp(-\delta_{ij}^2/\rho)$.

Geostatistical models: Hot issues

The hottest issue right now is that $\mathbf{G}$ is a dense matrix so as n increases, computing rapidly becomes impractical.

Lots of different fixes have been proposed.

My favorite: Nearest Neighbor Gaussian Processes (Datta, Banerjee).

Data analytic tools for these models are crude at best.

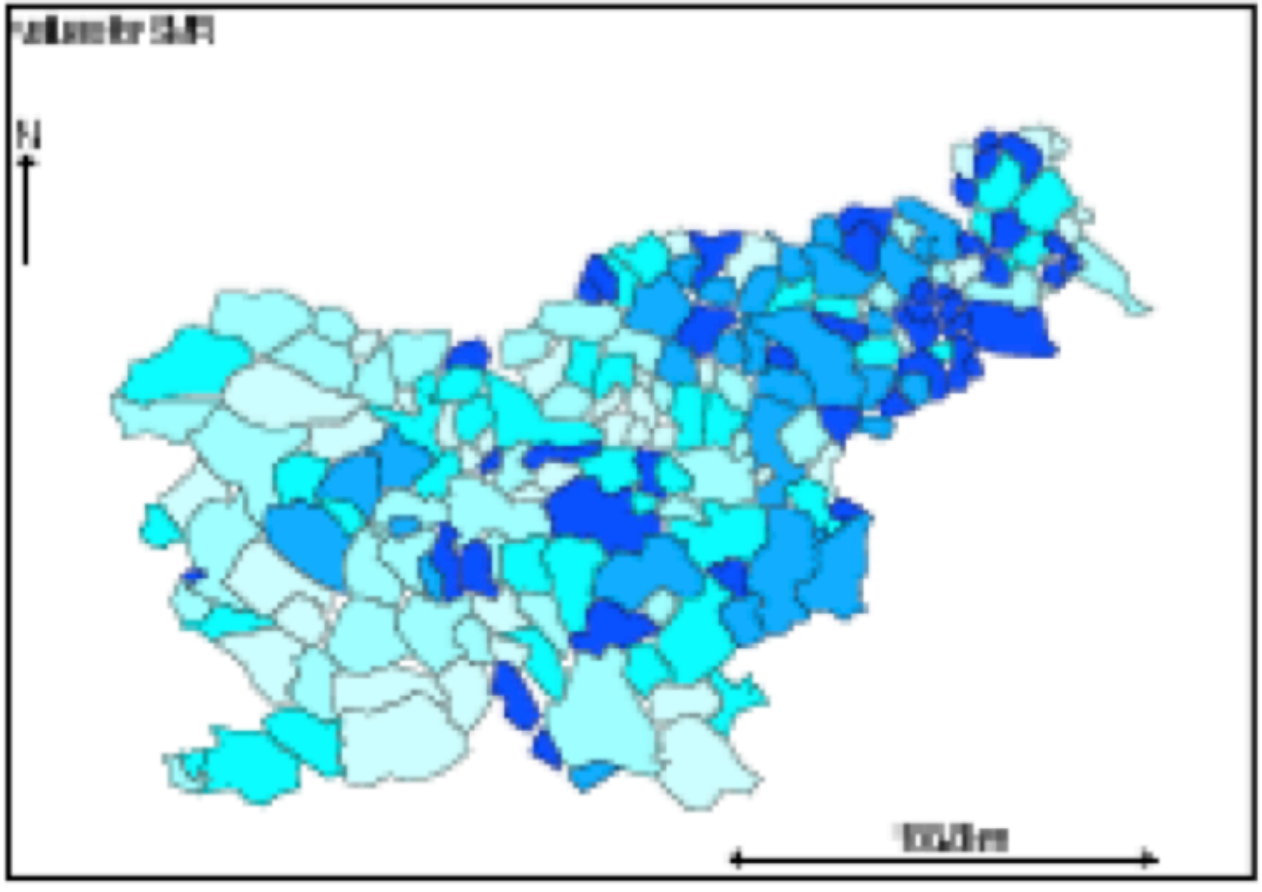
Is the GP random effect new- or old-style? It can be either.

- Smoothing/interpolation: New style.
- To describe variation between (say) days: Old style.

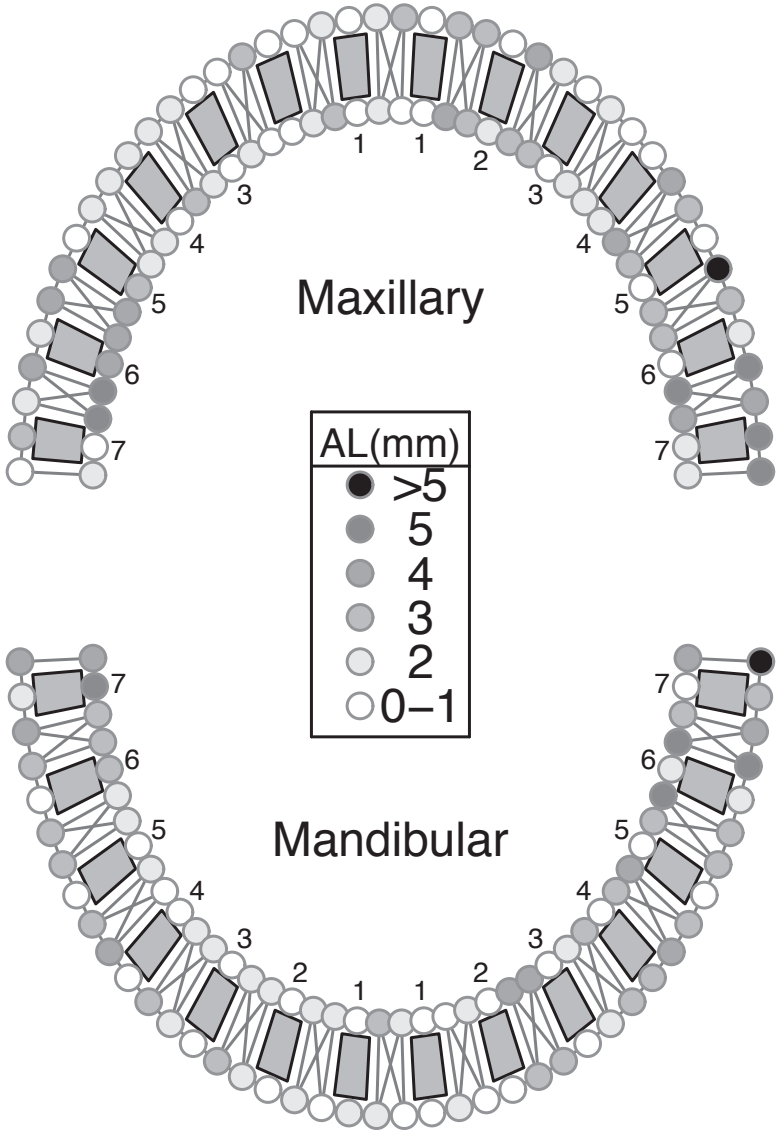
Areal models, examples: Disease mapping

This is the standard usage:

Two municipalities are a neighbor pair if they share a boundary.



Areal models, examples: Periodontal measurements



This is a less common usage:

Two tooth sites are a neighbor pair if we define them as adjacent.

The gray lines are neighbor pairs.

This has two "islands", one each for the upper and lower arches.

We'll mostly consider one model: Improper CAR (ICAR)

The model $\boldsymbol{\delta} \sim$ ICAR is often specified by its conditionals:

$$\delta_i | \boldsymbol{\delta}_{(-i)} \sim N(\textstyle\sum_{j \sim i} \delta_j / m_i, \sigma_s^2 / m_i),$$

where m_i is the number of region i's neighbors.

This is equivalent to the improper density:

$$f(\boldsymbol{\delta} | \sigma_s^2) \propto \exp\left(-\tfrac{1}{2\sigma_s^2} \boldsymbol{\delta}' \mathbf{Q} \boldsymbol{\delta}\right),$$

where $\mathbf{Q}$ has $Q_{ii} = m_i$ and $Q_{ij} = -1$ if $i \sim j$ and 0 otherwise.

Another equivalent form uses pairwise differences:

$$f(\boldsymbol{\delta} | \sigma_s^2) \propto \exp\left(-\tfrac{1}{2\sigma_s^2} \textstyle\sum_{i \sim j} (\delta_i - \delta_j)^2\right).$$

Intuition: Small σ_s^2 forces neighbors to be similar.
Large σ_s^2 allows neighbors to be different.

Including the ICAR in a mixed linear model

The ICAR model is almost always used as a random effect:

$$\mathbf{y} = \mathbf{X}_o\beta_o + \mathbf{I}_n\boldsymbol{\delta} + \boldsymbol{\epsilon}, \text{ for } \boldsymbol{\delta} \sim \text{ICAR}(\mathbf{Q}, \sigma_s^2).$$

$\mathbf{Q}$ is singular; here's how to make this model explicit.

Let $\mathbf{Q}$ have spectral decomposition $\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}'$, where

- $\mathbf{V}$'s columns are $\mathbf{Q}$'s eigenvectors (orthonormal).
- $\mathbf{D}$ is diagonal with non-negative diagonal entries

  $d_1 \geq \cdots \geq d_{n-I-1} > d_{n-I} = \cdots = d_n = 0,$

  where I is the number of islands in the spatial map.
- $\frac{1}{\sqrt{N}}\mathbf{1}_N$ is always an eigenvector with eigenvalue 0.

Re-parameterize the RE: $\boldsymbol{\theta} = \mathbf{V}'\boldsymbol{\delta}$; $\boldsymbol{\theta}$ has precision $\mathbf{D}/\sigma_s^2$.

We start with $\mathbf{y} = \mathbf{X}_o\boldsymbol{\beta}_o + \mathbf{I}_n\boldsymbol{\delta} + \boldsymbol{\epsilon}$, for $\boldsymbol{\delta} \sim \text{ICAR}(\mathbf{Q}, \sigma_s^2)$.

Re-parameterize the RE: $\boldsymbol{\theta} = \mathbf{V}'\boldsymbol{\delta}$; $\boldsymbol{\theta}$ has precision $\mathbf{D}/\sigma_s^2$,

$\Rightarrow \boldsymbol{\theta}_1$ $(n-l)\times 1$ has precision matrix $\mathbf{D}_1$, covariance $\mathbf{D}_1^{-1}$

$\boldsymbol{\theta}_2$ $l \times 1$ has precision matrix $\mathbf{0}$.

Partition $\mathbf{V}$ conformably to $\boldsymbol{\theta}$: $\mathbf{V} = [\mathbf{V}_1|\mathbf{V}_2]$; $\mathbf{V}_1$ is $n \times (n-l)$.

$$\begin{aligned} \text{Then} \quad \mathbf{y} &= \mathbf{X}_o\boldsymbol{\beta}_o + \mathbf{V}\mathbf{V}'\boldsymbol{\delta} + \boldsymbol{\epsilon} \\ &= \mathbf{X}_o\boldsymbol{\beta}_o + [\mathbf{V}_1|\mathbf{V}_2]\left[\frac{\boldsymbol{\theta}_1}{\boldsymbol{\theta}_2}\right] + \boldsymbol{\epsilon} \\ &= \mathbf{X}_o\boldsymbol{\beta}_o + \mathbf{V}_2\boldsymbol{\theta}_2 + \mathbf{V}_1\boldsymbol{\theta}_1 + \boldsymbol{\epsilon}. \end{aligned}$$

This is an MLM: $\mathbf{X} = [\mathbf{X}_o|\mathbf{V}_2]$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_o'|\boldsymbol{\theta}_2']'$, $\mathbf{Z} = \mathbf{V}_1$, $\mathbf{u} = \boldsymbol{\theta}_1$.

The ICAR implicitly specifies a FE for the l island means.

The re-expression highlights the ICAR's implicit REs

These random effects have design matrix $\mathbf{Z} = \mathbf{V}_1$ and $\mathbf{G} = \sigma_s^2 \mathbf{D}_1^{-1}$.

The columns of $\mathbf{V}_1$ and their d_j are determined by the spatial map.

Diagonals of $\mathbf{G} = \sigma_s^2 \mathbf{D}_1^{-1}$ differentially shrink the REs $\mathbf{u}$ toward 0:

Columns $\mathbf{V}_{1j}$ with large d_j: these u_j shrink a lot.
Columns $\mathbf{V}_{1j}$ with small d_j: these u_j shrink little.

Variation in $\mathbf{y} \propto \mathbf{V}_{1j}$ with large d_j is smoothed into error.
Variation in $\mathbf{y} \propto \mathbf{V}_{1j}$ with small d_j stays in the fit.

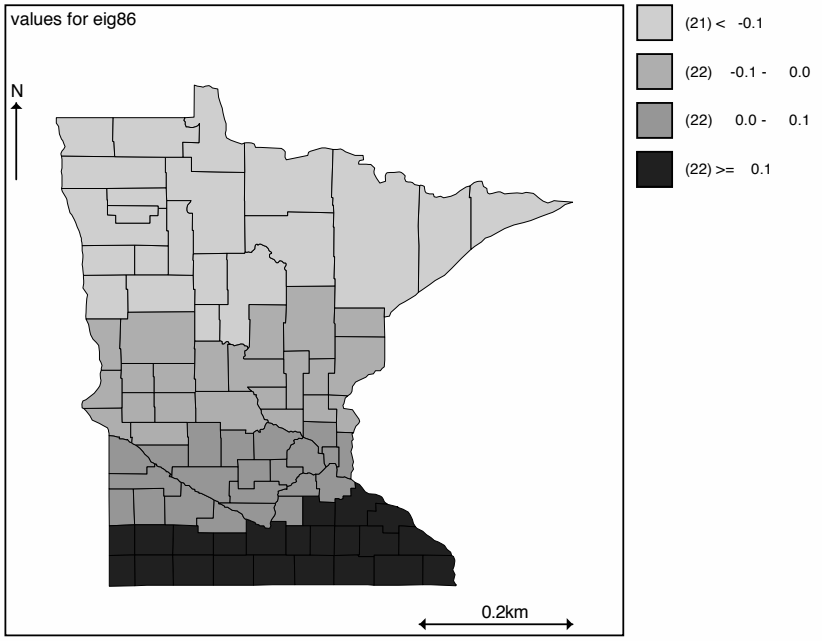
$\mathbf{1}_N$ is an eigenvector of $\mathbf{Q}$ with eigenvalue 0 $\Rightarrow$

$\mathbf{V}_1$'s columns are contrasts.

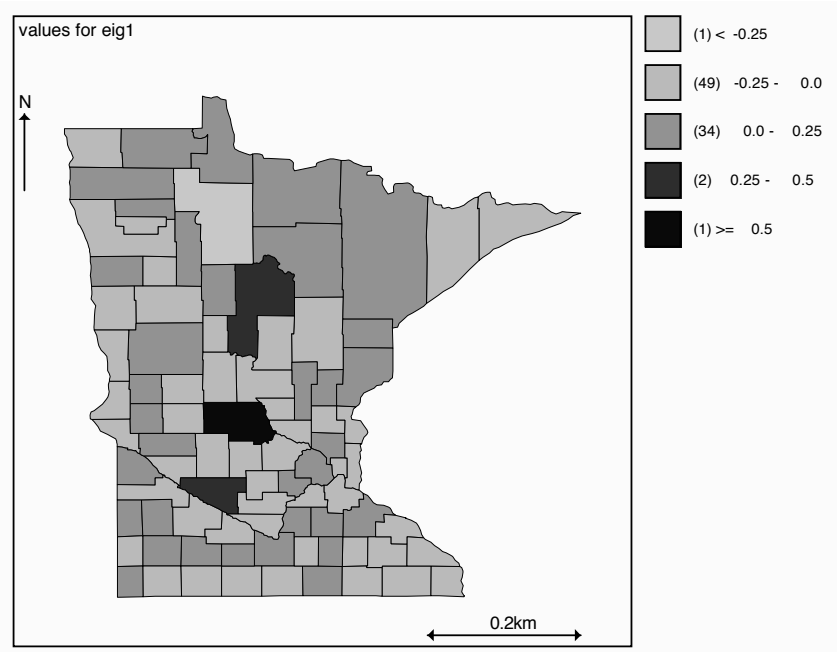
$\mathbf{V}_1$'s columns are interpretable (sort of)

The counties of Minnesota: neighbor pair = share a boundary.

$\mathbf{V}_{1,86}$



$\mathbf{V}_{1,1}$



u_{86} is least shrunk

u_1 is most shrunk

$\mathbf{V}_{1,86}$ is roughly linear north-to-south (low frequency);

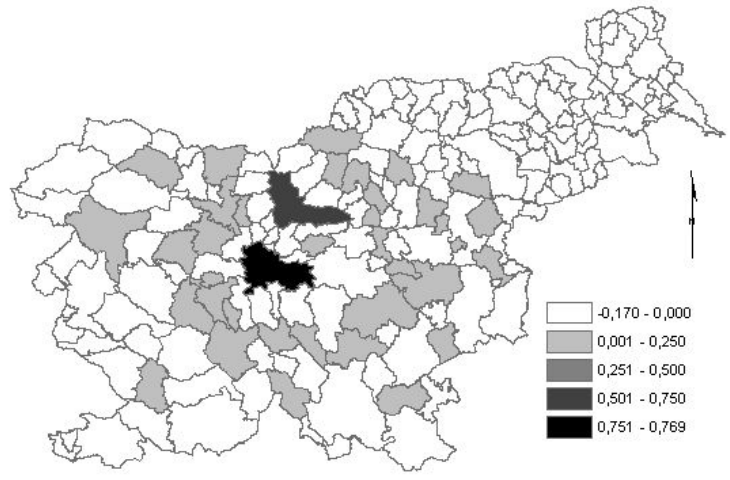
$\mathbf{V}_{1,1}$ is high-frequency.

The municipalities of Slovenia: neighbor pair = share a boundary.

$\mathbf{V}_{1,193}$

u_{193} is least shrunk

$\mathbf{V}_{1,1}$

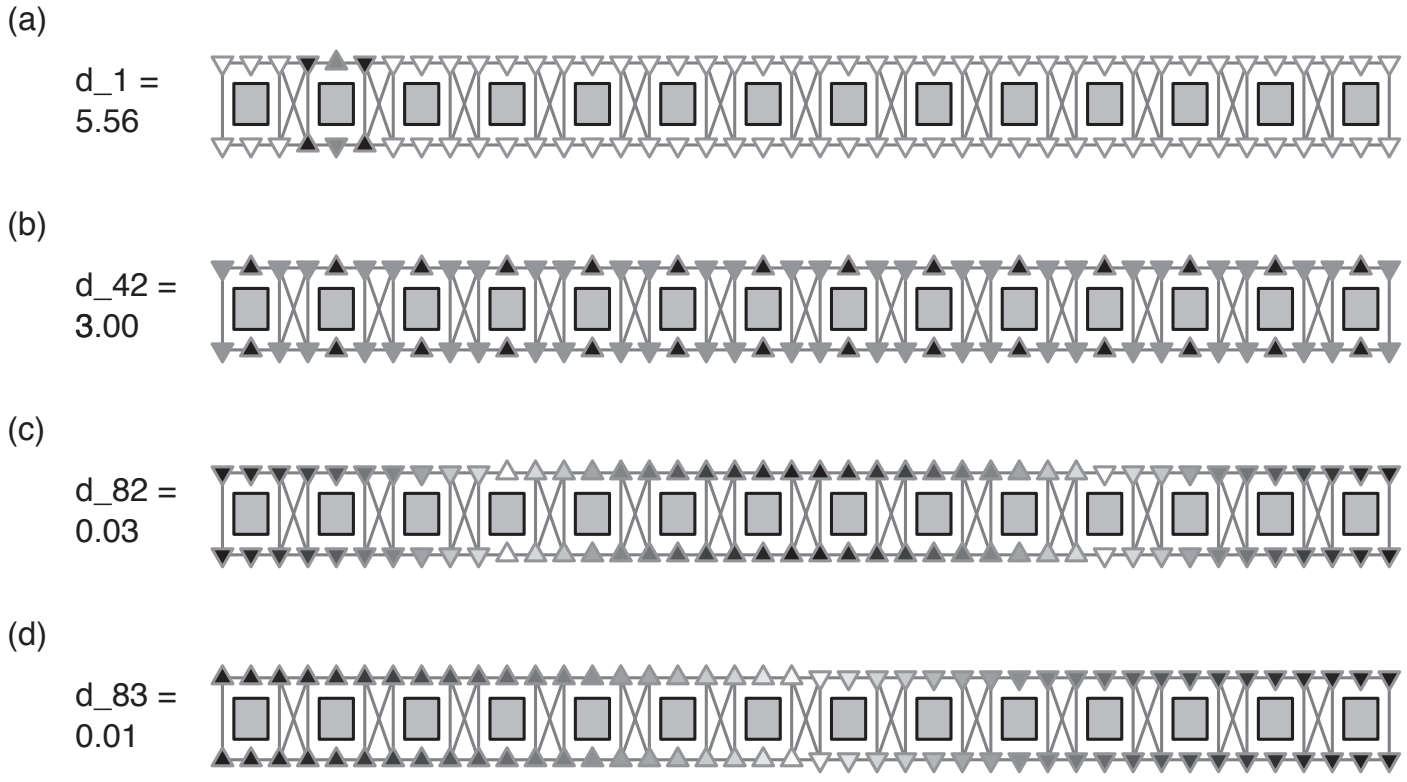


u_1 is most shrunk

$\mathbf{V}_{1,193}$ is roughly linear in Slovenia's long axis;

$\mathbf{V}_{1,1}$ centers on the municipalities with the most neighbors.

Periodontal measurement sites (one arch); neighbors as drawn.



u_{83}, u_{82} shrunk least and 2^{nd} least; roughly linear and quadratic.

u_1 shrunk most; it captures local variation.

u_{42} is shrunk a middling amount – we'll see it later.

Simple & useful variant on the ICAR model

Leroux et al (1999):

$\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_s^2 \Sigma)$ where

$\Sigma^{-1} = \lambda \mathbf{Q} + (1 - \lambda)\mathbf{I}_n$ is non-singular

and $\mathbf{Q}$ is the ICAR model's precision matrix.

This generalizes the ICAR model by having a mixture of spatially-structured variation $\mathbf{Q}$ and heterogeneity $\mathbf{I}$.

Conveniently, if $\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}'$ is $\mathbf{Q}$'s spectral decomposition,

$$\begin{aligned} \Sigma^{-1} &= \lambda \mathbf{V}\mathbf{D}\mathbf{V}' + (1 - \lambda)\mathbf{I}_n \\ &= \mathbf{V}\ (\lambda \mathbf{D} + (1 - \lambda)\mathbf{I}_n)\ \mathbf{V}' \end{aligned}$$

so Σ^{-1} and $\mathbf{Q}$ have the same eigenvectors but different eigenvalues.

Modularity of MLMs: Spatially smoothed ANOVA

Consider the Minnesota map with $N = 87$ counties.

Suppose each county has data y_{ij} for $m = 3$ cancers, $j = 1, 2, 3$.

This is an unreplicated 2-way ANOVA with factors
cancer ("*CA*", m levels) and county ("*CO*", N levels).

People who do disease mapping find it desirable to smooth

- the county main effect: how the overall level of these three cancers varies between counties, and
- the cancer-by-county interaction: how contrasts in the cancers vary between counties.

This generalizes to any other factor (e.g., sex, age) crossed with counties.

Zhang Y, Hodges JS, Banerjee S (2009 *Ann. Applied Stat.)*

Setting up this spatially-smoothed ANOVA

For county i, let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im})'$ describe the m cancers.

Define the Nm vector $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_N')'$.

Then decompose $\boldsymbol{\delta}$ as

$$\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \ldots, \boldsymbol{\delta}_N')' = [\mathbf{X}|\mathbf{Z}] \begin{bmatrix} \Theta_{GM} \\ \boldsymbol{\theta}_{CA} \\ \boldsymbol{\theta}_{CO} \\ \boldsymbol{\theta}_{CO \times CA} \end{bmatrix}$$

where

- GM = grand mean
- CA = cancer
- CO = county

The next slides show $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{G}$.

Columns of $\mathbf{X}$ for Θ_{GM} and $\boldsymbol{\theta}_{CA}$

For county i, let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im})'$ describe the m cancers.

Define the Nm vector $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_N')'$.

Let H_{CA} be $m \times (m-1) \ni H_{CA}' H_{CA} = \mathbf{I}_{m-1}$; its columns are contrasts.

Then $\mathbf{X}\boldsymbol{\beta}$, capturing the grand mean and cancer main effect, is:

$$\left[\begin{array}{c|c} \frac{1}{\sqrt{Nm}} \mathbf{1}_{Nm} & \frac{1}{\sqrt{N}} \mathbf{1}_N \otimes H_{CA} \end{array} \right] \left[\begin{array}{c} \Theta_{GM} \\ \boldsymbol{\theta}_{CA} \end{array} \right]$$

The two parts of $\mathbf{X}$ are $Nm \times 1$ and $Nm \times (m-1)$ respectively.

Recall that $\mathbf{A} \otimes \mathbf{B} = ((a_{ij}\mathbf{B}))$

The parts of $\mathbf{Z}$ and $\mathbf{G}$ for the county main effect

For county i, let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im})'$ describe the m cancers.

Define the Nm vector $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_N')'$.

Let $\mathbf{Q}$ $N \times N$ encode neighbor pairs among counties as in an ICAR model.

Let $\mathbf{Q}$ have spectral decomposition $\mathbf{VDV}'$ as before.

$\frac{1}{\sqrt{N}}\mathbf{1}_N$ is the only eigenvector with eigenvalue 0; let $\mathbf{V} = [\mathbf{V}_1 | \frac{1}{\sqrt{N}}\mathbf{1}_N]$.

The part of $\mathbf{Zu}$ capturing the county main effect is:

$$\left[\mathbf{V}_1 \otimes \tfrac{1}{\sqrt{m}}\mathbf{1}_m\right] \boldsymbol{\theta}_{CO}$$

The corresponding part of $\mathbf{G}^{-1}$ is $\tau_0 \mathbf{D}_1$, where

$\mathbf{D}_1$ is $\mathbf{D}$ without its N^{th} row and column, and $\tau_0 > 0$ is unknown.

The parts of $\mathbf{Z}$ and $\mathbf{G}$ for the cancer-by-county interaction

For county i, let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im})'$ describe the m cancers.

Define the Nm vector $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_N')'$.

$\mathbf{Q}$ $N \times N$ encodes neighbor pairs among counties;
$\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}'$ and $\mathbf{V} = [\mathbf{V}_1 | \frac{1}{\sqrt{N}}\mathbf{1}_N]$.

The part of $\mathbf{Zu}$ capturing the cancer-by-county interaction is:

$$\left[\mathbf{V}_1 \otimes H_{CA}^{(1)} \ \ldots \ \mathbf{V}_1 \otimes H_{CA}^{(m-1)}\right] \begin{bmatrix} \boldsymbol{\theta}_{CO\times CA,1} \\ \vdots \\ \boldsymbol{\theta}_{CO\times CA,m-1} \end{bmatrix}$$

The part of $\mathbf{G}^{-1}$ corresponding to $H_{CA}^{(j)}$, the j^{th} contrast, is $\tau_j \mathbf{D}_1$, where $\mathbf{D}_1$ is $\mathbf{D}$ without its N^{th} row and column, and $\tau_j > 0$ is unknown.