On Spatial Disease Mapping Models Using Directed Acyclic Graphs

Sudipto Banerjee

Joint work with Abhirup Datta (Johns Hopkins Univ) and James S. Hodges (Univ of Minnesota Twin Cities)

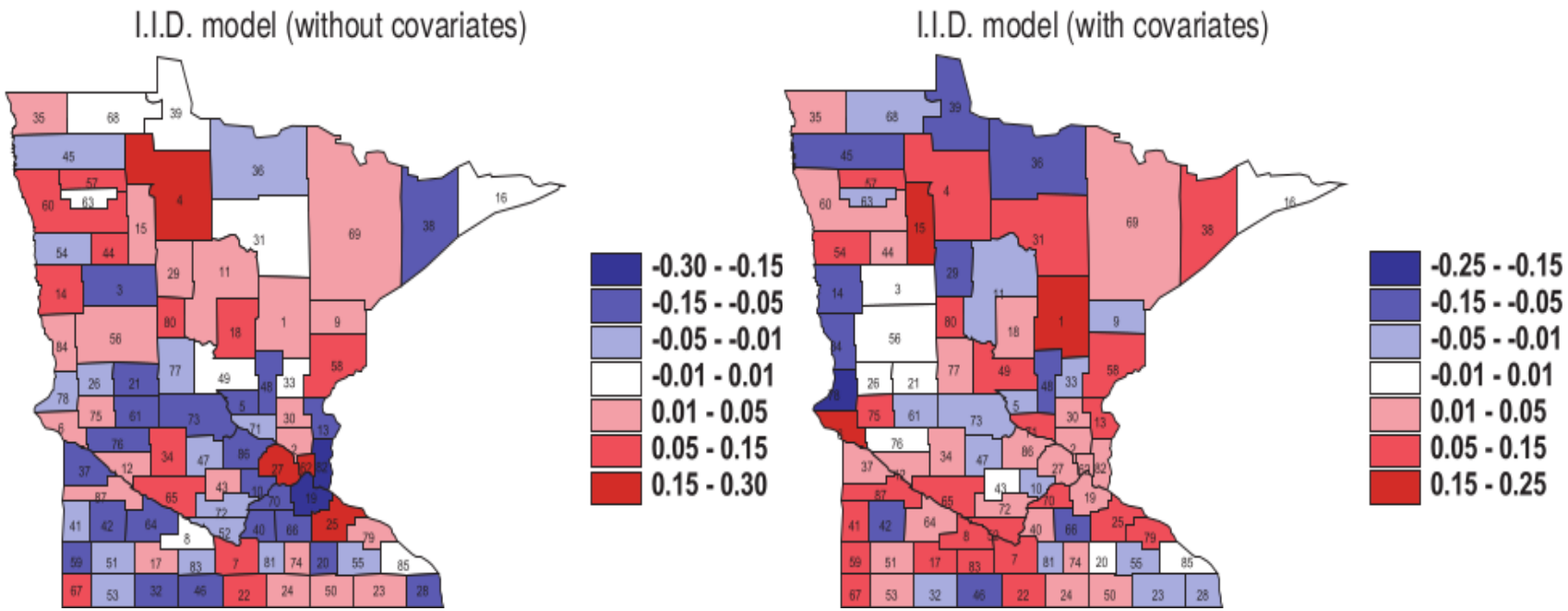
University of California, Los Angeles, USA

Department of Biostatistics UCLA

Disease Mapping: Mapping Random Effects

Model-based estimates of random effects across 87 counties in Minnesota

$$[\text{Infant Mortality Rates}] = [\text{Intercept}] + [\text{Fixed Effects}] + [\text{County-wise Random Effects}]$$



CAR models

- Conditional autoregressive (CAR) model (Besag, 1974; Clayton and Bernardinelli, 1992)
- Areal data modeled as a graph or network: V is the set of vertices (regions)
- $i \sim j$ if regions i and j share a common border
- Adjacency matrix $A = (a_{ij})$ such that $a_{ij} = I(i \sim j)$
- n_i is the number of neighbors of i
- CAR model:

$$w_i \,|\, w_{-i} \sim N\left(\frac{\rho}{n_i} \sum_{j \,|\, i \sim j} w_j, \tau_w n_i\right)$$

Hierarchical GLM for Spatial disease mapping

- At unit (region) i, we observe response y_i and covariate x_i
- $g(E(y_i)) = x_i^\top \beta + w_i$ where $g(\cdot)$ denotes a suitable link function

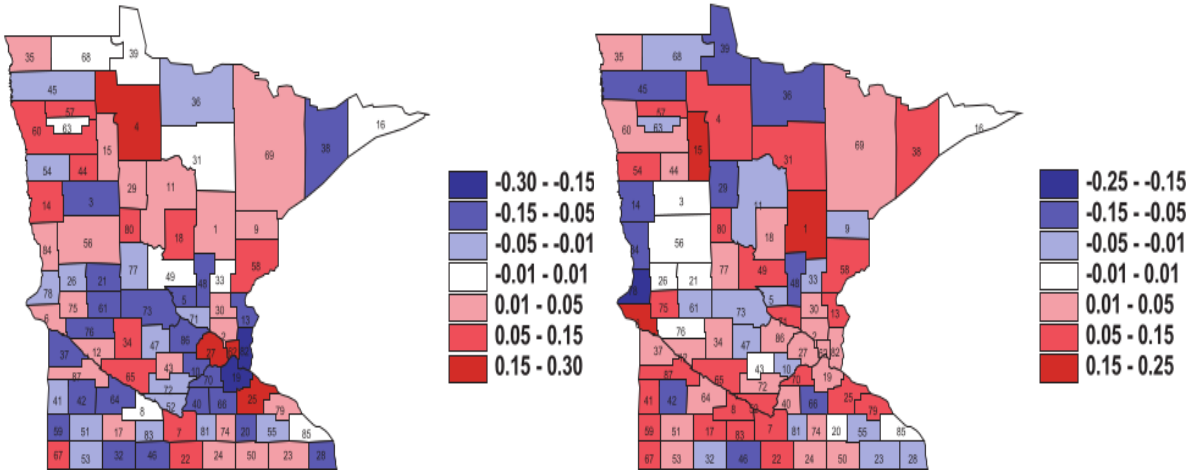
$$p_2(\beta, \tau_w, \rho) \times N(w \,|\, 0, \tau_w(D - \rho A)) \times \prod_{i=1}^{k} p_1(y_i \,|\, x_i^\top \beta + w_i)$$

- p_1 denotes the density corresponding to the link $g(\cdot)$

Disease Mapping: Mapping Random Effects

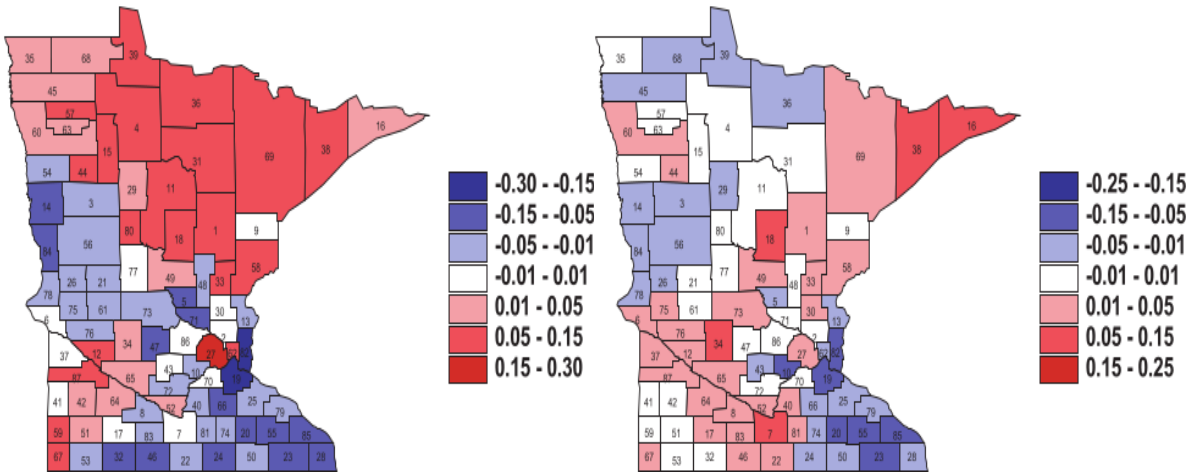
[Infant Mortality Rates] = [Intercept] + [Fixed Effects] + [County-wise Random Effects]

I.I.D. model (without covariates)



I.I.D. model (with covariates)

CAR model (without covariates)



CAR model (with covariates)

SAR models

- Simultaneous Autoregressive (SAR) model (Whittle, 1954)
- Instead of taking the conditional route, SAR model proceeds by simultaneously modeling the random effects

$$w_i = \rho \sum_{i \neq j} b_{ij} w_j + \epsilon_i \text{ for } i = 1, 2, \ldots, k$$

- $\epsilon_i \overset{ind}{\sim} N(0, \tau_i)$ are errors independent of w
- A common choice is to define $b_{ij} = I(i \sim j)/n_i$
- Joint distribution: $w \sim N(0, (I - \rho B)^\top F (I - \rho B))$, $B = (b_{ij})$ and $F = diag(\tau_1, \tau_2, \ldots, \tau_k)$

Interpretation of ρ in proper CAR and SAR models

- Calibration of ρ as a correlation, e.g., (as reported in Banerjee et al. 2014)

$$\rho = 0.80 \text{ yields } 0.1 \leq \text{Moran's } I \leq 0.15,$$
$$\rho = 0.90 \text{ yields } 0.2 \leq \text{Moran's } I \leq 0.25,$$
$$\rho = 0.99 \text{ yields Moran's } I \leq 0.5$$

- So, used with random effects, scope of spatial pattern may be limited

Interpretation of ρ in proper CAR and SAR models

- ρ cannot be interpreted as correlation between neighboring w_i's (Wall, 2004; Assuncao and Krainski, 2009)

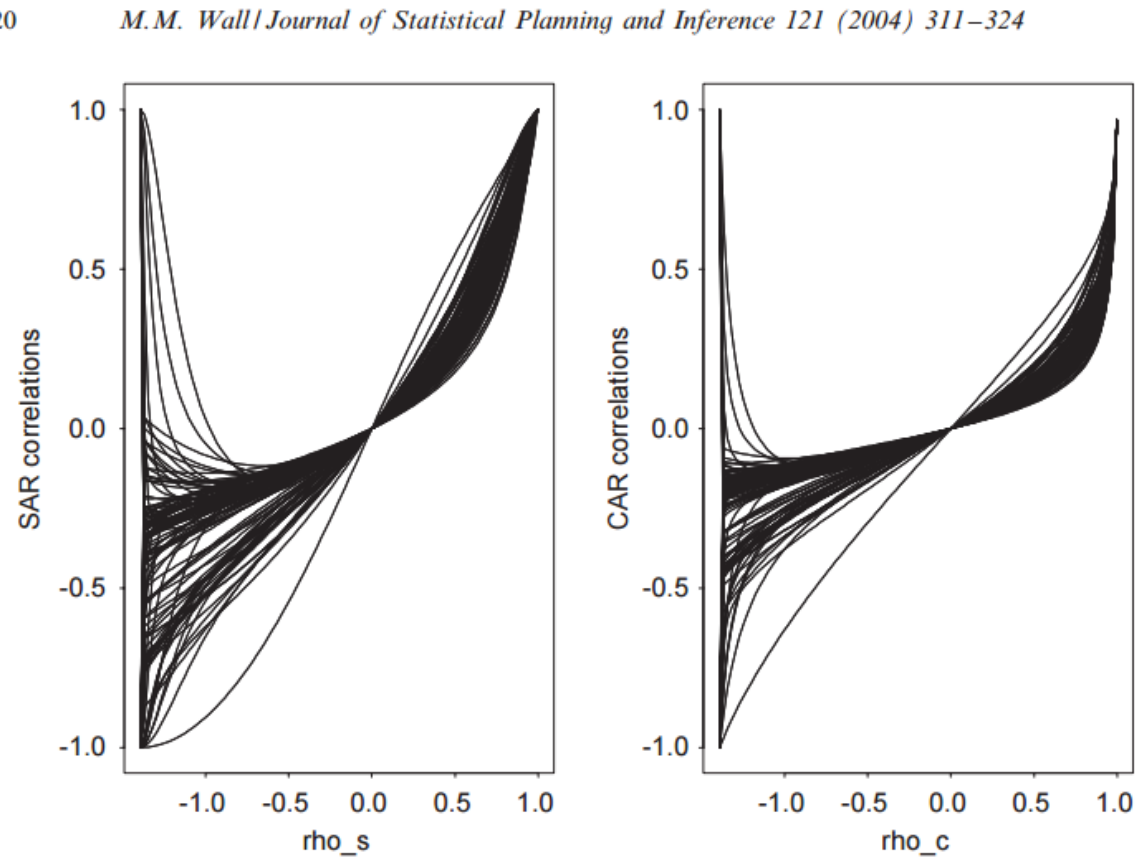


Figure: Neighbor pair correlations as a function of ρ for proper CAR and SAR models over the graph of US states

What if we construct a Directed Acyclic Graph

$$p(w) = p(w_1)p(w_2 \mid w_1)p(w_3 \mid w_1, w_2) \cdots p(w_n \mid w_1, w_2, \ldots, w_{k-1})$$

$$\begin{aligned}
w_1 &= \epsilon_1 \; ; \quad \epsilon_1 \sim N(0, \tau_1) \\
w_2 &= b_{21} w_1 + \epsilon_2 \; ; \quad N(0, \tau_2) \\
w_3 &= b_{31} w_1 + b_{32} w_2 + \epsilon_3 \; ; \quad N(0, \tau_3) \\
&\vdots \\
w_k &= b_{k1} w_1 + b_{k1} w_2 + \ldots + b_{k,k-1} w_{k-1} + \epsilon_k \; ; \quad N(0, \tau_k)
\end{aligned}$$

- $w = Bw + \epsilon$, where $\epsilon \sim N(0, F)$ and $F = \text{diag}(\tau_1, \ldots, \tau_k)$

$B = (b_{ij})$ is now a strictly lower triangular matrix.

What if we construct a Directed Acyclic Graph

- Advantages of lower triangular B:
  - $w \sim N(0, (I-B)^\top F(I-B))$ is a proper distribution for any choice of lower triangular B
  - $\det(L^\top F L) = \prod_{i=1}^{n} \tau_i$ where $F = \text{diag}(\tau_1, \ldots, \tau_k)$ and $L = I - B$
  - $w^\top L^\top F L w = \tau_1 w_1^2 + \sum_{i=2}^{k} \tau_i (w_i - \sum_{\{j<i\}} w_j b_{ij})^2$
  - Likelihood $N(w \,|\, 0, (I-B)^\top F(I-B))$ can be computed using $O(k+s)$ flops where s denotes the sparsity (number of non-zero entries) of B.
  - Even if k is large, evaluation of likelihood is fast if each region only shares border with a few others

Choice of B and F

- How to specify B and F?
- Sparsity of B is desirable
- $b_{ij} = 0$ for j outside neighbor sets $N(i)$
  - Pros: For graphs with a fixed topological order, neighbor sets can be chosen: $N(i) = \{j \mid j \sim i, j < i\}$
  - Cons: There is no covariance function on arbitrary graphs from which we can obtain non-zero b_{ij}'s and F

Autoregressive models on trees

- $D = (d_{ij})$ is the shortest distance matrix on the graph
- If the graph was a tree (no loops), then $\rho^D = \{\rho^{d_{ij}}\}$ is then a valid *autoregressive* correlation matrix (AR(1) model on a tree, Basseville et al., 2006).
- Areal graphs are loopy and are not usually trees

Local embedded spanning trees

- Embedded spanning trees (EST) of a graph G is a subgraph of G which is a tree and spans all the vertices of G
- Note that to specify $w_i = \sum_{j \in N(i)} b_{ij} w_j + \epsilon_i$ we only need a joint distribution on $\{i\} \cup N(i)$
- Let G_i denote the subgraph of G which includes vertices $\{i\} \cup N(i)$ and the edges among them
- The subgraph T_i of G_i which only contains the edges $\{i \sim j \,|\, j \in N(i)\}$ is an embedded spanning tree of G_i
- Use the local embedded spanning trees T_i to specify the b_{ij}'s and τ_i

Directed acyclic graph autoregressive (DAGAR) model

- AR_i denotes the $AR(1)$ distribution on T_i
- Solve for b_{ij} and τ_i such that $E_{AR_i}(w_i \mid w_{N(i)}) = \sum_{j \in N(i)} b_{ij} w_j$ and $\tau_i = 1/Var_{AR_i}(w_i \mid w_{N(i)})$
- No edge is left out !

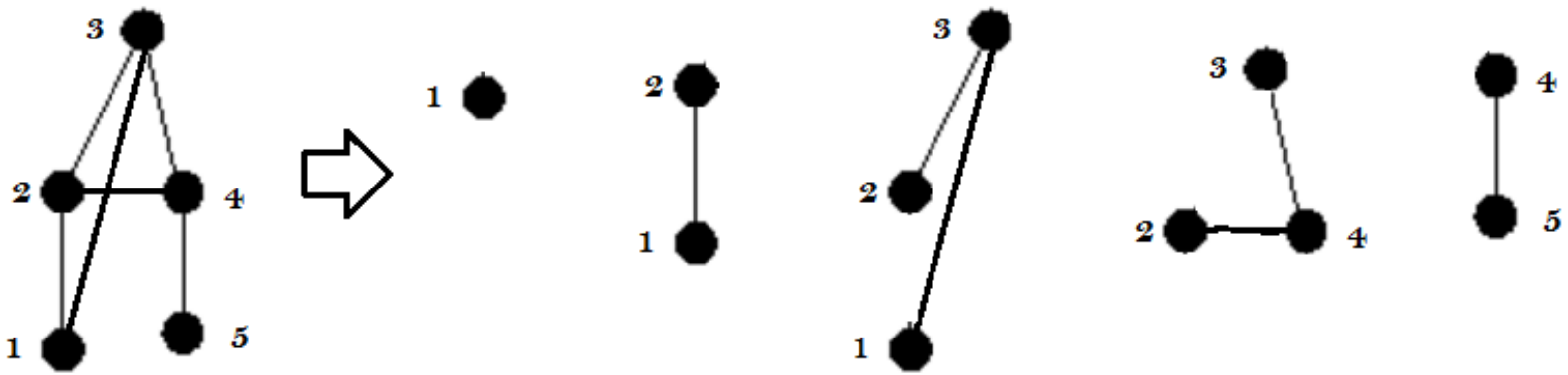


Figure: Decomposing a graph into a sequence of embedded spanning trees

Properties of DAGAR models

- $b_{ij} = b_i = \rho/(1 + (|N(i)| - 1)\rho^2)$
- $\tau_i = (1 + (|N(i)| - 1)\rho^2)/(1 - \rho^2)$
- $\det(Q_{DAGAR}) = \prod_{i=1}^{k} \tau_i$
- Positive definite for any $0 \leq \rho \leq 1$
- Interpretability of ρ:
  - If the graph is a tree, then DAGAR model is same as the AR(1) model on the tree i.e. correlation between d^{th} order neighbors is ρ^d for $d = 1, 2, \ldots$
  - If the graph is a closed two-dimensional grid, then each neighbor pair correlation is ρ
- $p_{DAGAR}(w)$ can be stored and evaluated using $O(e + k)$ flops where e is the total number of neighbor pairs

Averaging Cholesky decompositions over orderings

- Priors over the space of all $k!$ orderings/permutations
- Averaging over Cholesky decompositions of precision matrices:

$$Q = \frac{1}{k!} \sum_{\pi} P_\pi L_\pi D_\pi L_\pi^{\mathrm{T}} P_\pi^{\mathrm{T}}$$

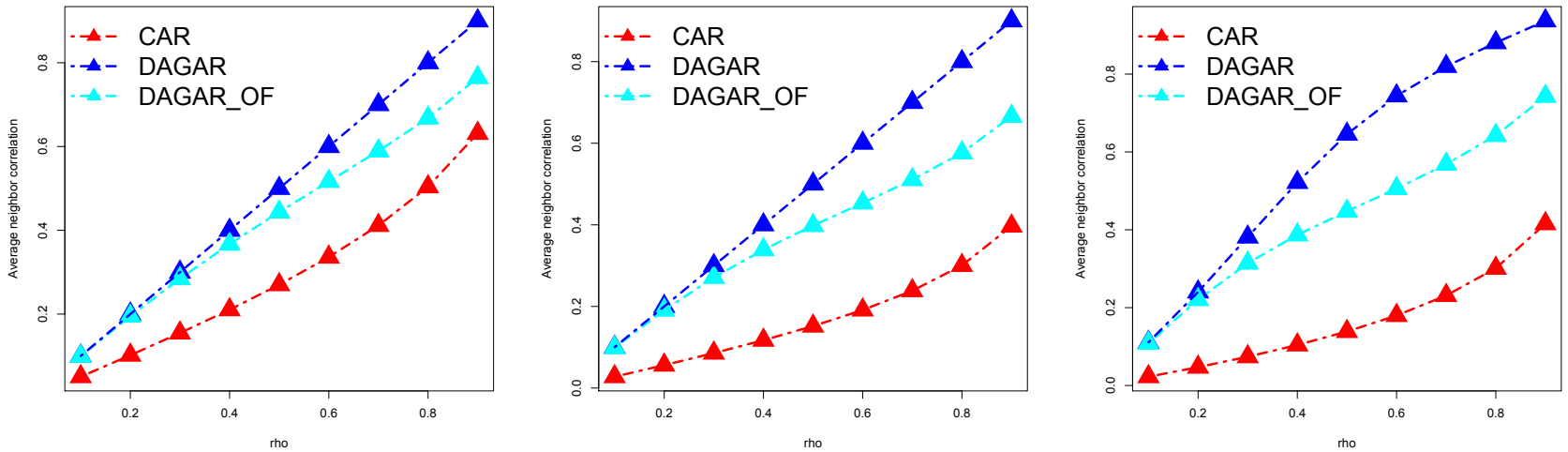
- The above produces closed-form expressions for Autoregressive models:

$$Q_{ii} = 1 + \frac{n_i \rho^2}{2(1-\rho^2)} + \frac{\rho^2}{(1-\rho^2)} \sum_{j \sim i} f(\rho, n_j)$$

$$Q_{ij} = \frac{\rho}{1-\rho^2} I(i \sim j) + \frac{\rho^2}{1-\rho^2} \sum_{k \in N(i) \cap N(j)} g(\rho, n_k) \; .$$

- New class of DAGAR models for areal/lattice data (Datta et al., 2017)

Interpretation of ρ



(a) path graph (b) grid graph (c) USA state map

Figure: Average neighbor pair correlations as a funcion of ρ for proper CAR and DAGAR models

Simulated data analysis

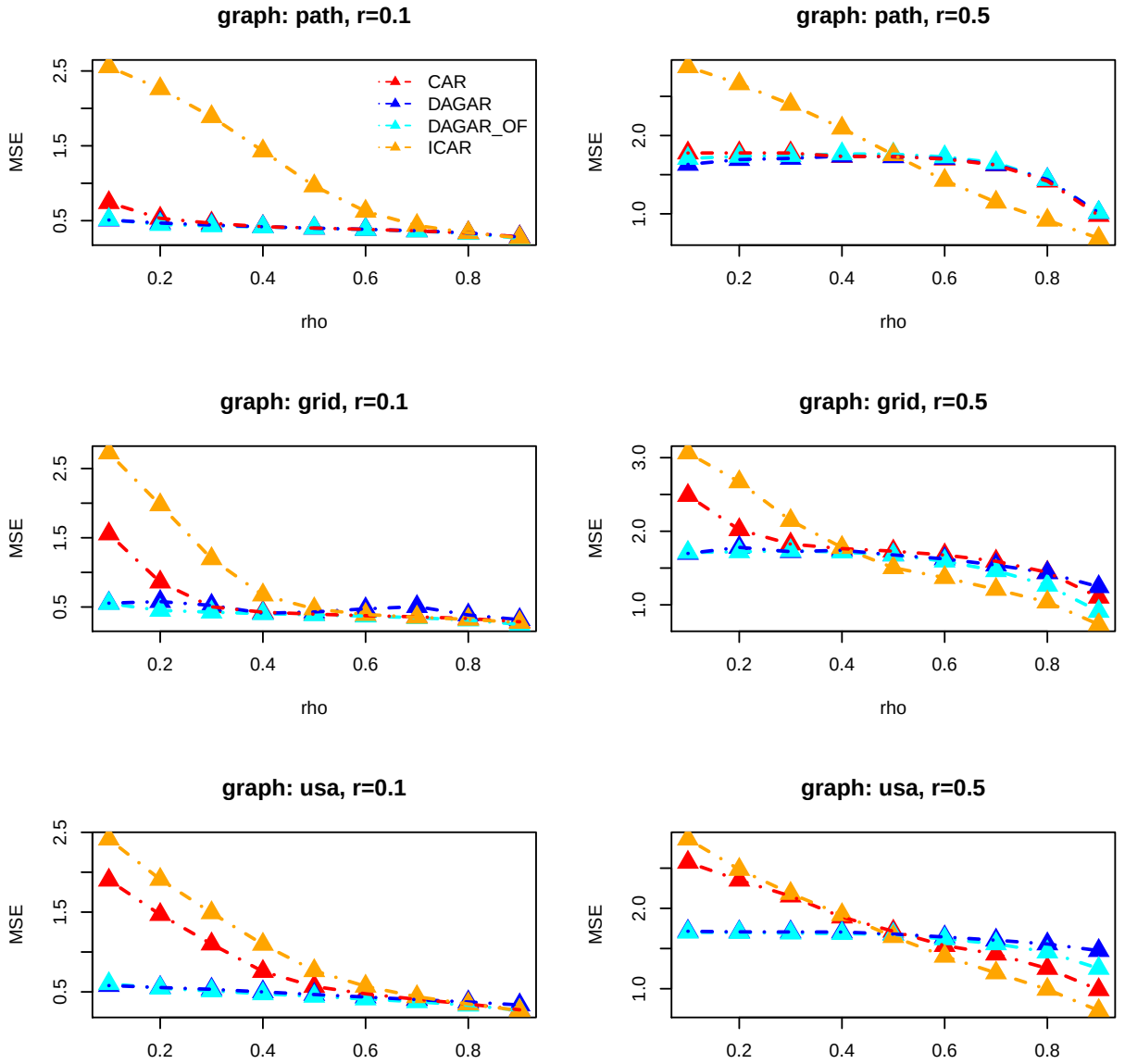
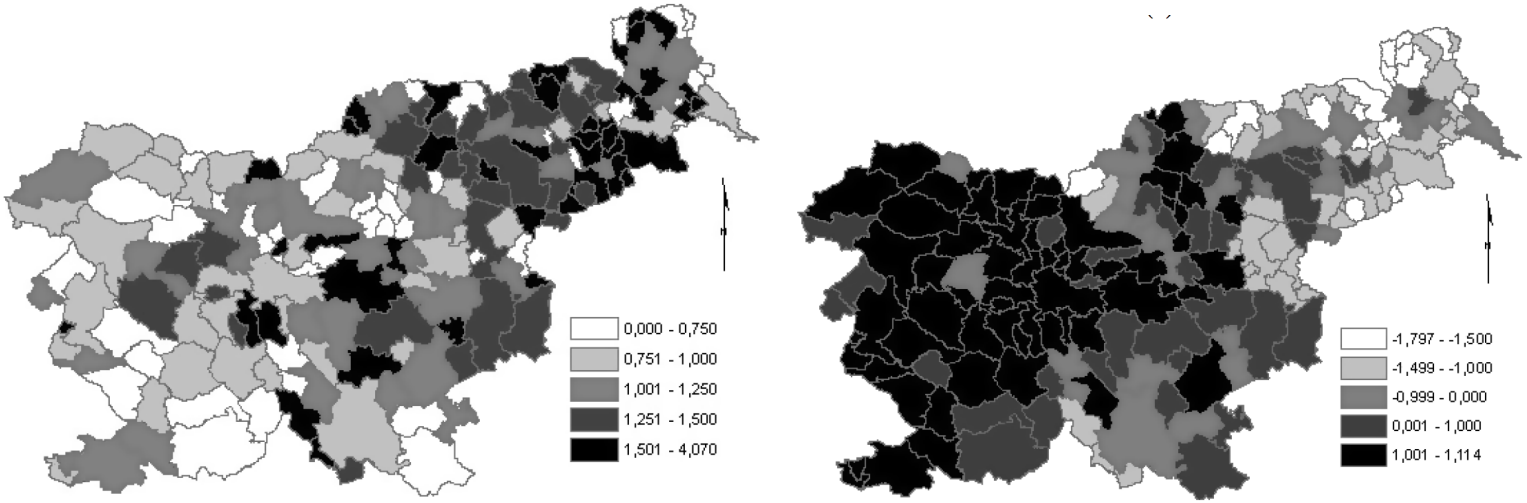


Figure: Mean square error as a function of ρ and $r = \tau^2/\sigma^2$ for DAGAR and CAR models

Slovenia stomach cancer data



(a) Standardized cancer incidence (b) Socio-economic score

Figure: Slovenia stomach cancer data

- Observed (O_i) and expected (E_i) number of cancer counts for each of the 194 municipalities of the country
- $O_i \sim Poisson(E_i \exp(\alpha + \beta SE_i + w_i))$ where $w \sim N(0, \tau_w Q(\rho))$

Slovenia stomach cancer data

Table: Parameter estimates with confidence intervals and model comparison metrics

	α	β	ρ	DIC	LPPD$_{LOOCV}$[1]
CAR	0.09 (0.02, 0.16)	-0.12 (-0.19, -0.04)	0.33 (0.02, 0.86)	1097	1170
DAGAR	0.11 (0.03, 0.18)	-0.12 (-0.19, -0.06)	0.08 (0.004, 0.24)	1091	1127
DAGAR$_{OF}$	0.11 (0.05, 0.17)	-0.12 (-0.18, -0.06)	0.06 (0.003, 0.2)	1090	1133

- Zadnik and Reich (2006) observed spatial confounding with ICAR model ($\hat{\beta}_{ICAR} = -0.02(-0.10, 0.06)$)
- Here for all three models the CIs for β lie outside zero
- Estimates of ρ are much smaller than 1
- Estimates of β here are closer to those obtained in the non-spatial (NS) analysis ($\hat{\beta}_{NS} = -1.4(-0.17, -0.10)$)

[1] Log-predictive posterior density using Leave one out cross validation

Summary

- DAGAR models for areal data constructed from sparse Cholesky factors
- Scalability for large areal data
- Ordered vs order-free DAGAR
  - For all analysis, ordered model performed very similar to the order-free model
  - Ordered model is faster with theoretical results about interpretability of ρ
- DAGAR models are positive definite and can be directly used to model or simulate any multivariate data on graphs (like imaging or social network data)
- Better performance than CAR modes for many scenarios
- DAGAR available at `https://arxiv.org/pdf/1704.07848.pdf`

Thank You!