PICAR: An Efficient Extendable Approach for Fitting Hierarchical Spatial Models

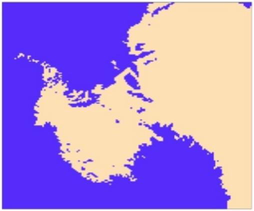
Ben Seiyon Lee
(joint work with Murali Haran)

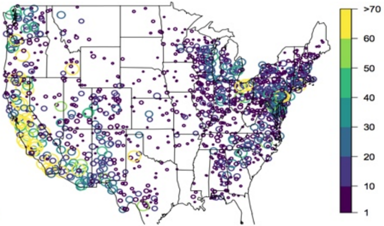
Department of Statistics, Penn State University

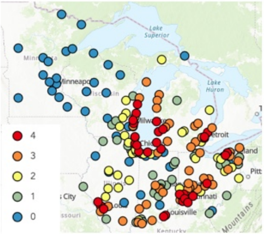
November 17, 2019

Hierarchical spatial models are popular models for spatial observations. Examples:

- **Glaciology:** Satellite imagery of ice presence/absence in Antarctica
- **Ecology:** Occurrence of animal species
- **Public Health:** Geo-referenced survey questionnaires







E.g. 1,000+ papers used these models in 2019 alone (Google Scholar)

Examples of Hierarchical Spatial Models

Hierarchical spatial models can combine information across different data sources while accounting for spatial dependence and other forms of variation. A few examples:

- Spatial generalized linear mixed models (SGLMMs)
  - Latent Gaussian process models (Diggle et al., 1998)
  - Latent Gaussian Markov random field models (Besag et al., 1991)
- Spatially-varying coefficient models (Gelfand et al., 2003)
- Cumulative-logit models for ordinal spatial data (Agresti, 2010)

This Talk

Motivated by challenges when fitting hierarchical spatial models

1. High-dimensional integration due to high-dimensional latent variables
2. Expensive matrix operations: $O(n^3)$ where n is data dimension
3. Highly correlated latent variables $\Rightarrow$ slow mixing Markov chains

Existing algorithms are either:

- Too slow for high-dimensional problems
- Not automated
- Black box approaches: cannot be easily extended by non-experts

Our new approach: **P**rojection-based **I**ntrinsic **C**onditional **A**uto**R**egression

- Fast: $\approx$ 4 hours for 20,000+ dimensional hierarchical model
- Extendable: can fit user-specified models using simple `Rstan` code
- Automated: basis is determined by algorithm

Outline

1. Hierarchical spatial models
2. Basis representation for spatial random effects
3. Our approach: PICAR
   1. Overview
   2. Tuning mechanisms
4. Simulation study and results
5. Applications
6. Discussion

Hierarchical Spatial Models

- Observation model for $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))$, $\{s_i\}$ are locations

$$Z(s_i) | W(s_i) \sim f(Z(s_i) | \beta, W(s_i), \epsilon(s_i)), \text{ for } i = 1, \ldots, n$$

- Process model for $\mathbf{W} = (W(s_1), \ldots, W(s_n))$:

$$\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma^2 R_\phi), \;\; \epsilon \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathcal{I})$$

  R_ϕ: Correlation matrix with parameter ϕ

- Priors for regression, covariance function parameters β, ϕ, σ^2, τ^2

Goal: Infer $\beta, \phi, \sigma^2, \tau^2$ based on posterior,

$$\pi(\beta, \phi, \sigma^2, \tau^2, \mathbf{W} \mid Z),$$

typically using Markov chain Monte Carlo.

In non-linear settings $\mathbf{W}$ poses major computational challenge

Example: Spatial Generalized Linear Mixed Models

Observation Model:

$$Z(s)|W(s) \sim f(\eta(s)|W(s))$$

where $f(\cdot)$ is Poisson, Negative Binomial, ...

$$\eta(s) = g(\mathbb{E}[Z(s)|\beta, W(s)]) = X(s)\beta + W(s) + \epsilon(s),$$

where $\eta(s)$ is link function, example log link

(Diggle et al., 1998)

Basis Representations

PICAR approach: Represent spatial random effects **W** with basis functions.

Reparameterized Spatial Random Effects: (cf. Cressie and Wikle, 2015)

$$\mathbf{W} \approx \mathbf{\Phi}\delta,$$

$$\delta \sim \mathcal{N}(0, \Sigma_\delta),$$

where:

- $\mathbf{\Phi}$ is an $n \times p$ basis function matrix
- $\delta \in \mathbb{R}^p$ are reparameterized spatial random effects (basis coefficients)
- Σ_δ

Benefits:

- **Dimension-Reduction:** $\delta \in \mathbb{R}^p$ with $p << n$
  (Later) dwarf mistletoe example $p = 520$ when $n = 25,431$
- **De-Correlation:** $\mathbf{\Phi}$ can de-correlate random effects for fast mixing

Basis Function Approaches for Spatial Models

Spatial linear models

- Eigenvector Spatial Filtering (Griffith, 2003)
- LatticeKrig and Radial Basis functions (Nychka et al., 2015)
- Random Projections (Banerjee et al., 2008)

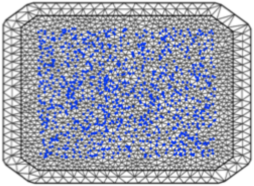
Non-linear hierarchical models

- Re-parameterizations of Spatial Random Effects (Christensen et al., 2006; Haran et al., 2003)
- Kernel Convolutions (Higdon, 1998)
- Predictive Processes (Banerjee et al., 2008)
- INLA (Rue et al., 2009; Lindgren et al., 2011)
- Moran's Basis for Areal Data (Hughes and Haran, 2013)
- Random Projections for hierarchical models (Guan and Haran, 2018, 2019; Park and Haran, 2019)

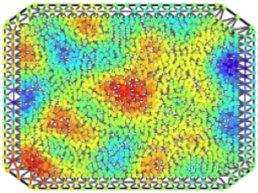
Sketch of Our Approach: PICAR

Projection-based **I**ntrinsic **C**onditional **A**uto**R**egression

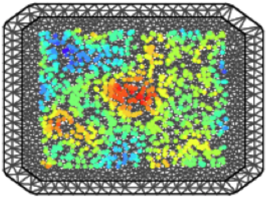
Part A:
Discretize spatial domain



Part B:
Use projected ICAR model on discrete space



Part C:
Use discrete model to interpolate on continuous space



Part B:

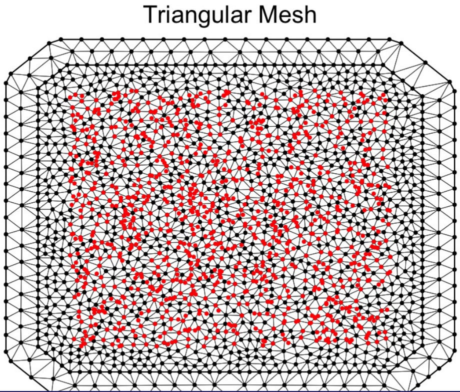
- Use adjacency implied by mesh to construct ICAR model
- Use projection + dimension reduction to obtain low-dimensional basis ⇒ low-dimensional, decorrelated spatial random effects

A. Mesh Construction

Use Delaunay triangulation to divide spatial domain into a set of non-intersecting irregular triangles via `R-INLA` (Lindgren et al., 2015)

Idea: Discretize latent spatial random field

1. Originally $n-$dimensional on observation locations: $\mathbf{W} \in \mathbb{R}^{\mathbf{n}}$
2. Now $m-$dimensional on mesh nodes $\tilde{\mathbf{W}} \in \mathbb{R}^{\mathbf{m}}$ with $m > n$



B. Basis #1: Moran's Basis Functions

Goal: Generate spatial field on mesh nodes $\tilde{\mathbf{W}}$ via empirical basis functions
Moran's Operator (Griffith, 2003; Hughes and Haran, 2013):

$$(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{m})\mathbf{A}(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{m}) \in \mathbb{R}^{m \times m},$$

where $\mathbf{A}$ is the adjacency matrix for mesh nodes, $\tilde{\mathbf{W}} \in \mathbb{R}^m$

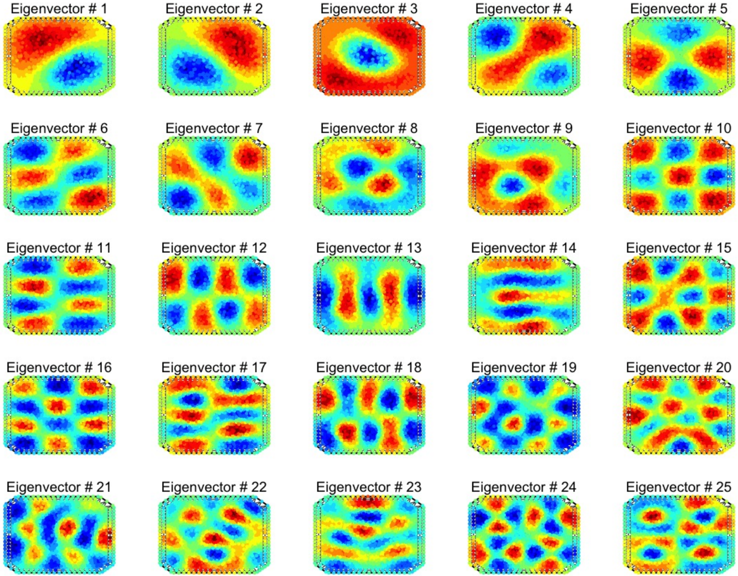
Moran's Basis M:

$\mathbf{M} \in \mathbb{R}^{m \times p}$ are the first $\mathbf{p}$ eigenvectors of the Moran's operator

1. **Dimension Reduction:**
   rank$(\mathbf{M}) = \mathbf{p}$ chosen via automatic heuristic with $p << m$
2. **De-correlated Weights:**
   Moran's basis is orthogonal $\Rightarrow$ spatial random effects $\delta \in \mathbb{R}^p$ are decorrelated

Outcome: $\tilde{\mathbf{W}} = \mathbf{M}\delta$

B. Eigenvectors of Moran's Operator



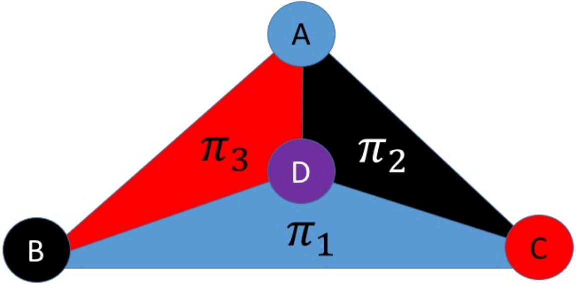
C. Basis #2: Piece-wise Linear Basis Functions

Basis Functions for Triangular Mesh (Lindgren et al., 2011):

- **Projector Matrix:** $\mathbf{A} \in \mathbb{R}^{n \times m}$

Objective: Use mesh nodes to interpolate points within triangular mesh.

- **Interpolation:** $\mathbf{W} \approx \mathbf{A}\tilde{\mathbf{W}}$, where $\mathbf{W} \in \mathbb{R}^n$ are observation locations and $\tilde{\mathbf{W}} \in \mathbb{R}^{\mathbf{m}}$ are mesh nodes.

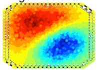


$$D \approx (\pi_1 \text{ x A}) + (\pi_2 \text{ x B}) + (\pi_3 \text{ x C})$$

Role of Basis Functions

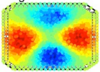
$\mathbf{M}$	δ
Moran's Basis	Coefficients/ Weights

Eigenvector # 1



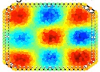
δ_1

Eigenvector # 5



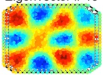
δ_5

Eigenvector # 10



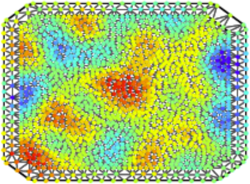
δ_{10}

Eigenvector # 15



δ_{15}

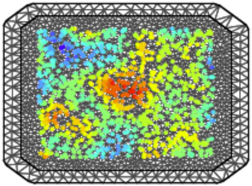
$\mathbf{M}\delta$



Idea:

- Imposes a latent spatial field on mesh nodes.
- Constructs latent field using **Moran's basis functions (M)** + Weights (**$\boldsymbol{\delta}$**).

$\mathbf{AM}\delta$



Idea:

- Projects latent field from mesh nodes onto observation locations.
- Uses Piecewise Linear Basis functions **(A)**.

PICAR Version of Spatial Hierarchical Model

Observation Model:

$$Z(s) \sim \prod_{i=1}^{n} f(\eta(s)|\beta, \delta)$$

$$\eta(s) = g(E[Z(s)|\beta, \delta]) = X(s)\beta + [\mathbf{AM}\delta](s),$$

where **A** is the projector matrix, **M** is the Moran's basis matrix, and δ are the dimension-reduced and de-correlated weights.

Process Model:

$$\delta \sim \mathcal{N}(0, \tau^{-1}(\mathbf{M}'\mathbf{QM})^{-1}),$$

where Q is a chosen $m \times m$ precision matrix.

Priors:

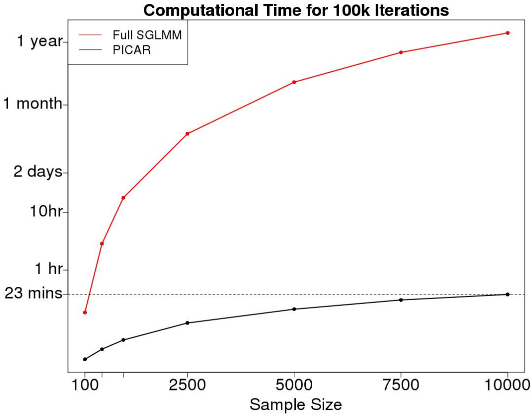
$$\tau \sim G(\alpha_\tau, \beta_\tau)$$
$$\beta \sim N(0, \Sigma_\beta),$$

where $\alpha_\tau, \beta_\tau,$ and Σ_β are the hyperparameters.

Computational Cost

Computational Complexity:

1. **Full SGLMM (gold standard):** $\mathcal{O}(\frac{1}{3}n^3)$
2. **PICAR:** $\mathcal{O}(np)$,
   where $n{=}\#$ of observations and $p = \text{rank}(\mathbf{M})$ with $p << n$



What is Being Approximated in PICAR?

Spatial Random Effects in:

1. Hierarchical spatial models:

$$\mathbf{W} \sim N(0, \sigma^2 R_\phi),$$

where R_ϕ is a correlation function, typically from the Matérn class.

2. PICAR (Basis Representation):

$$\mathbf{W} = \mathbf{AM}\delta, \qquad \delta \sim \mathbf{N}(\mathbf{0}, \tau^{-1}\mathbf{Q}^{-1})$$

$$\mathbf{W} \sim N(0, \tau^{-1}\mathbf{AMQ}^{-1}\mathbf{M}'\mathbf{A}')$$

Key Approximation:

$$\sigma^2 R_\phi \approx \tau^{-1}\mathbf{AMQ}^{-1}\mathbf{M}'\mathbf{A}'$$

Challenge: How do we improve this approximation?

Tuning Mechanisms

Idea: Select rank($\mathbf{M}$) and Q with the best predictive performance.

Tuning Mechanism #1:

Goal: Select p for the Moran's Basis ($p = \text{rank}(\mathbf{M})$).

Automatic Heuristic:

1. Construct Moran's basis $\mathbf{M}$ for varying p.
2. Fit GLM model using columns of $\mathbf{AM}$ as covariates.
3. Choose p that returns low cross-validated mean squared prediction error (CVMSPE).

Tuning Mechanism #2:

Goal: Choose prior precision matrices Q with low CVMSPE.

- Independent: $Q = \mathbf{I}$
- ICAR: $Q = (D - W)$
- CAR: $Q = (D - \phi W)$, where $\phi \in (0, 1)$

Binary Observations: Simulated Example

Overview:

- Generate spatial binary data with locations on the unit domain $[0,1]^2$
- $n_{mod} = 1000$ observations to fit model + $n_{cv} = 400$ for cross-validation.
- **Fixed Effects:** $\beta_1 = 1$ and $\beta_2 = 1$
- **Random effects** $W(s)$**:** Generated using the Matérn covariance function with parameters $\nu = 2.5$, $\sigma^2 = 1$, and $\phi = 0.2$.

Comparative Study:

1. Full SGLMM (gold standard)
2. PICAR with varying ranks for Moran's basis $(\mathbf{M})$.
3. PICAR with varying prior precision matrices Q.

Simulated Example: Moran's Basis Rank

Rank selection is important!

Rank	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	1.04 (0.77,1.31)	0.91 (0.64,1.16)	0.3	9.73
22	**1.09 (0.82,1.37)**	**0.93 (0.67,1.2)**	**0.27**	**10.73**
50	1.12 (0.83,1.41)	0.95 (0.67,1.23)	0.28	11.14
75	1.14 (0.85,1.44)	0.98 (0.69,1.26)	0.28	11.62
100	1.2 (0.9,1.5)	1 (0.71,1.29)	0.29	12.28
200	1.34 (1.01,1.66)	0.99 (0.69,1.31)	0.32	15.13
Gold Standard	1.03 (0.77,1.3)	0.89 (0.63,1.16)	0.29	3624.43

Table: Binary Observations simulated example results across Moran's basis ranks.

Count Data with Spatially Varying Coefficients

Goal: Demonstrate the ease of fitting user-defined models via PICAR

- We use **rStan**, a programming language for Bayesian inference

Simulated Example Overview:

- Generate spatial count data with locations on the unit domain $[0, 1]^2$
- $n_{mod} = 1000$ observations to fit model + $n_{cv} = 400$ for cross-validation.
- **Fixed Effects:** $\beta_1 = 1$ and $\beta_2 = 1$
- **Spatially-varying coefficients $\beta(s)$:** Generated using the Matérn covariance function with parameters $\nu = 2.5$, $\sigma^2 = 1$, and $\phi = 0.2$.
- **Random effects $W(s)$:** Generated similarly as $\beta(s)$.

Simulated Example for Varying Coefficient Model

Rank	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	0.92 (0.81,1.02)	0.94 (0.85,1.03)	3.59	0.52
50	0.77 (0.54,1.01)	0.99 (0.89,1.09)	2.31	7.68
63	**0.77 (0.48,1.06)**	**1.02 (0.9,1.12)**	**2.31**	**13.15**
75	0.86 (0.55,1.16)	1.03 (0.93,1.14)	2.68	22.07
100	0.95 (0.55,1.33)	1.06 (0.94,1.17)	2.82	46.55
200	1.07 (0.7,1.49)	1.07 (0.95,1.19)	3.8	226.49

Example Implementation in **rStan**

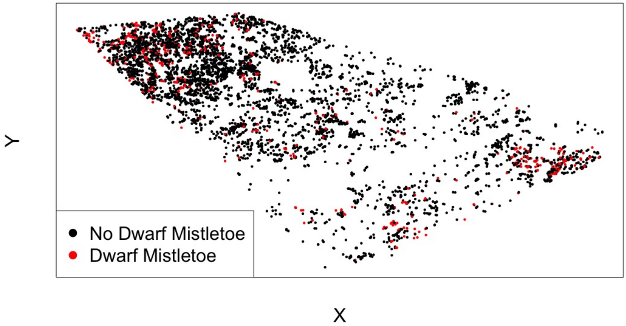
```
transformed parameters {
  vector[N] linpred;
  linpred = X*beta+M*delta+ X[,1].*(M*beta1S);
}

model {
  beta[1] ~ normal(0,100); // B1 prior
  beta[2] ~ normal(0,100); // B2 prior
  tau ~ gamma(0.5,2000); // Tau prior - random effects
  tau1S ~ gamma(0.5,2000); // Tau prior - coefficients
  delta ~ multi_normal_prec(zeros, tau * MQM); //Delta prior
  beta1S ~ multi_normal_prec(zeros, tau1S * MQM); //SVC prior
  y ~ poisson_log(linpred); // Likelihood
}
```

Minnesota Dwarf Mistletoe Infestation

Dwarf mistletoes are parasitic trees that affect the longevity of the black spruce. Black spruce: used for manufacturing paper

Observations

- $n = 25,431$ black spruce stands from the Department of Natural Resources (cf. Hanks et al., 2011)
- **Binary response:** Presence vs. absence of dwarf mistletoe
- **Covariates:** Age, basal area, height, and volume of spruce stand



Results

Computational Cost: $\sim$4 hours to model data via the PICAR approach.

- $\sim$ 2 hours to generate Moran's basis $\mathbf{M}$.
- $\sim$ 2 hours to draw 100k samples from the posterior distribution.

Parameter Inference:
Rank Chosen = 520

Type	β_1 (95% CI)	β_2 (95% CI)	β_3 (95% CI)	β_4 (95% CI)
Ind	0.56 (0.38,0.74)	-0.22 (-0.41,-0.02)	2.43 (2.03,2.79)	-0.19 (-0.32,-0.06)
ICAR	0.65 (0.47,0.83)	-0.16 (-0.35,0.02)	2.53 (2.16,2.91)	-0.2 (-0.33,-0.07)
CAR	0.56 (0.39,0.75)	-0.21 (-0.4,-0.03)	2.44 (2.08,2.8)	-0.19 (-0.33,-0.07)

Table: Fixed effects parameter estimates with 95% Credible Interval.

Discussion

1. We present PICAR, a new fast projection-based approach for high-dimensional hierarchical spatial models.
2. We demonstrate how it is easy to use PICAR to fit user-specified spatial hierarchical models via **rStan**
3. Our approach is fast ($\mathcal{O}(np)$) and provides comparable inference and prediction results to the gold standard approach.
4. Our approach applies to wide array of hierarchical spatial models such as: (1) SGLMMs, (2) spatially-varying coefficient models, and (3) cumulative-logit models for ordinal spatial data.
5. For high-dimensional data sets ($>$ 100,000) constructing the Moran's operator may be prohibitively expensive
6. Multivariate spatial models and space-time models still pose computational challenges: these are avenues for future work.

References I

Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20.

Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). Robust markov chain monte carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1):1–17.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Diggle, P. J., Tawn, J. A., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

References II

Griffith, D. A. (2003). Spatial filtering. In *Spatial Autocorrelation and Spatial Filtering*, pages 91–130. Springer.

Guan, Y. and Haran, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 27(4):701–714.

Guan, Y. and Haran, M. (2019). Fast expectation-maximization algorithms for spatial generalized linear mixed models. *arXiv preprint arXiv:1909.05440*.

Hanks, E. M., Hooten, M. B., and Baker, F. A. (2011). Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications*, 21(4):1173–1188.

Haran, M., Hodges, J. S., and Carlin, B. P. (2003). Accelerating computation in markov random field models for spatial data via structured mcmc. *Journal of Computational and Graphical Statistics*, 12(2):249–264.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159.

References III

Lindgren, F., Rue, H., et al. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19):1–25.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599.

Park, J. and Haran, M. (2019). Reduced-dimensional monte carlo maximum likelihood for latent gaussian random field models. *arXiv preprint arXiv:1910.09711*.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Thank you!

PICAR

Intrinsic Conditional Autoregressive (ICAR) Model

Objective: Model the latent intrinsic Gaussian Markov random field using mesh vertices $\tilde{W}(s)$.

Intrinsic Gaussian Markov Random Field:

$$\tilde{\mathbf{W}}|\tau \sim N(0, [\tau(\mathbf{D} - \mathbf{W})]^{-1}), \tag{1}$$

where:

- τ is the precision parameter
- $\mathbf{W} \in \mathbb{R}^{m \times m}$ is the neighborhood matrix where $\mathbf{W}_{ij} = 1$ when vertices i and j share an edge and $\mathbf{W}_{ij} = 0$ otherwise.
- $\mathbf{D} \in \mathbb{R}^{m \times m}$ where $\mathbf{D}_{i,i} =$ the number of neighbors for vertex i and 0 on the off-diagonals.

Poisson Simulated Example Results

Table: Results for one simulated Poisson dataset across varying dimensions of the Moran's Basis as well as INLA. For PICAR, time based on $\approx$ 250k iterations of the MCMC algorithm.

Dim	β_1	β_2	CVMPSE	Time (Minutes)
10	0.87 (0.76,0.98)	1.05 (0.94,1.16)	1.34	7.91
30	0.98 (0.86,1.09)	1.03 (0.93,1.14)	1.49	8.16
50	0.98 (0.86,1.1)	1.05 (0.95,1.17)	1.06	8.57
100	0.99 (0.86,1.11)	1.07 (0.96,1.19)	0.91	9.64
200	1 (0.87,1.13)	1.05 (0.93,1.17)	1.12	12.16
300	1 (0.87,1.15)	1.04 (0.92,1.17)	1.14	14.75
INLA	0.98 (0.86,1.11)	1.08 (0.97,1.19)	1.09	4.36

Simulation Study: Binary Data

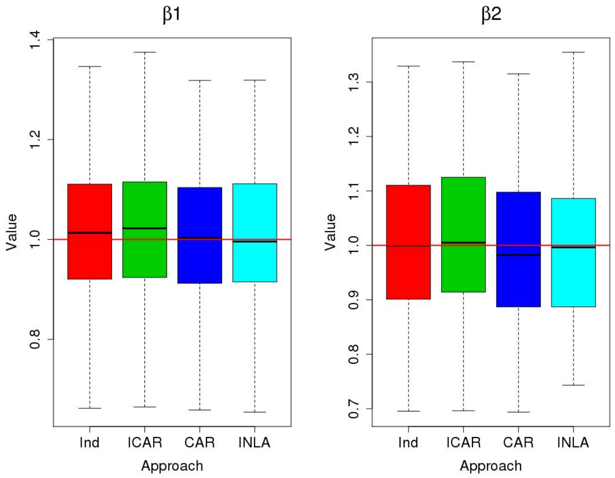


Figure: **Binary Data Simulation Study.** Posterior mean of coefficients β_1 and β_2 from 100 samples.

Simulation Study: Poisson Data

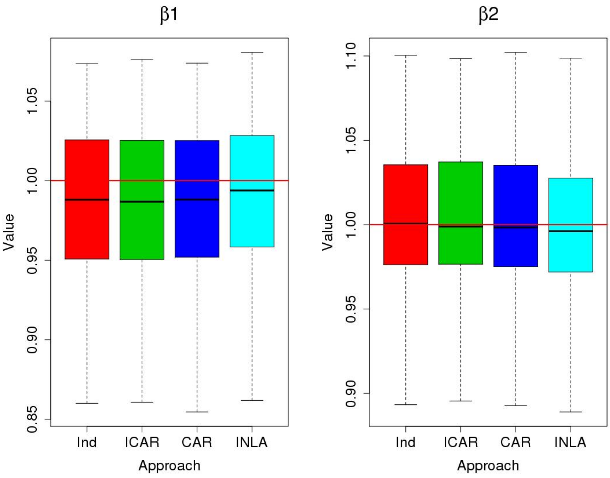


Figure: **Poisson Data Simulation Study.** Posterior mean of coefficients β_1 and β_2 from 100 samples.

Simulated Example: Precision Matrix Q

Choice of precision matrix Q has little effect on results.

Precision Matrix	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
Independent	1.07 (0.8,1.34)	0.92 (0.65,1.18)	0.28	9.53
ICAR	1.09 (0.82,1.37)	0.93 (0.67,1.2)	0.27	10.73
CAR	1.05 (0.79,1.33)	0.91 (0.65,1.18)	0.27	10.38
Gold Standard	1.03 (0.77,1.3)	0.89 (0.63,1.16)	0.29	3624.43

Table: Simulated example results across precision matrices.