PICAR: An Efficient Extendable Approach for Fitting Hierarchical Spatial Models

Murali Haran
(joint work with Ben Seiyon Lee)

Department of Statistics, Penn State University

December 15, 2019

This Talk

Motivated by challenges when fitting hierarchical spatial models

1. 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

This Talk

Motivated by challenges when fitting hierarchical spatial models

1. 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

This Talk

Motivated by challenges when fitting hierarchical spatial models

1. 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

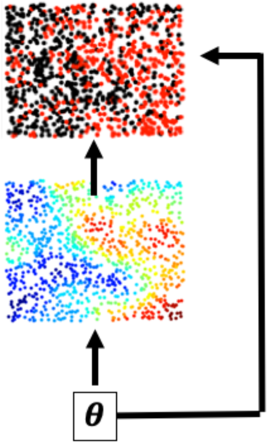
- Fast. E.g. $\approx$ 4 hours for $> 20,000$ dimensional problems. Other algorithms would take days/weeks or worse
- Extendable: can fit user-specified models using simple `stan` code
- Automated: basis functions determined by algorithm

What Are Hierarchical Spatial Models?

Data Model: $Z | W, \theta$

Process Model: (Latent) $W | \theta$

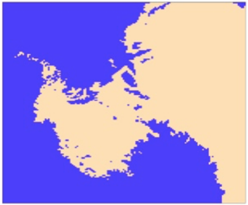
Parameter Model: θ



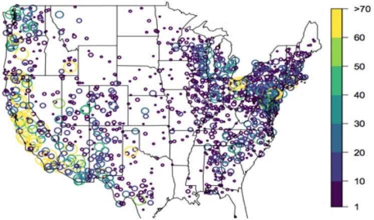
Uses of Hierarchical Spatial Models

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

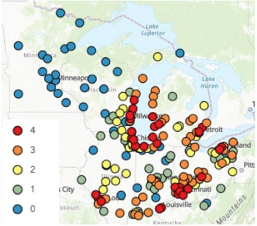
- **Glaciology:** Satellite imagery of ice presence/absence in Antarctica
- **Ecology:** Occurrence of animal species
- **Public Health:** Geo-referenced survey questionnaires



Binary



Count



Ordered Categorical

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

Examples

Hierarchical spatial models combine information across different data sources + accounts for spatial dependence and other forms of variation.

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)

Model Specification Example

Data Model:

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

Process Model:

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

Parameter Model:

Priors for β, ϕ, σ^2, and τ^2.

Notation:

- $f(\cdot)$: Probability distribution (e.g. Poisson, Negative Binomial)
- $g(\cdot)$: Link function
- $X(s)$: Matrix of covariates
- $W(s)$: Spatial random effects
- R_ϕ: Correlation function
- ϕ: Range parameter
- σ^2: Partial sill parameter
- τ^2: Nugget/measurement error

Goal: Infer β, ϕ, σ^2, τ^2 and $W(s)$.

Inference

Observations: $\mathbf{Z}$ Spatial Random Effects: $\mathbf{W}$ Parameters: θ

1. Maximum Likelihood:

$$\hat{\theta}_{ML} = \arg\max_{\theta} \int_{\mathbf{W}} L(\theta, \mathbf{W}|\mathbf{Z}) d\mathbf{W}$$

Integration may not be practical for high-dimensional $\mathbf{W}$

2. Bayesian Estimation:

$$\pi(\theta, \mathbf{W}|\mathbf{Z}) \propto p(\mathbf{Z}|\mathbf{W}, \theta) p(\mathbf{W}|\theta) p(\theta)$$

Sample from $\pi(\theta, \mathbf{W}|\mathbf{Z})$ via Markov chain Monte Carlo (MCMC) algorithm

How Does PICAR Speed Up Computation?

Observations: $\mathbf{Z}$ Spatial Random Effects: $\mathbf{W}$ Parameters: θ

Full Hierarchical Spatial Model

$$\pi(\theta, \mathbf{W}|\mathbf{Z})$$

$\mathbf{W}$ is high dimensional and highly correlated

- Large matrix operations $\rightarrow$ Each iteration of MCMC is expensive
- Slow mixing Markov chains $\rightarrow$ Longer MCMC runs

How Does PICAR Speed Up Computation?

Observations: $\mathbf{Z}$ Spatial Random Effects: $\mathbf{W}$ Parameters: θ

Full Hierarchical Spatial Model

$$\pi(\theta, \mathbf{W}|\mathbf{Z})$$

$\mathbf{W}$ is high dimensional and highly correlated

- Large matrix operations $\rightarrow$ Each iteration of MCMC is expensive
- Slow mixing Markov chains $\rightarrow$ Longer MCMC runs

Our Approach (PICAR)

$$\pi(\theta, \boldsymbol{\delta}|\mathbf{Z})$$

$\boldsymbol{\delta}$ is low dimensional and decorrelated

- Small matrix operations $\rightarrow$ Each iteration is cheap
- Fast mixing Markov chains $\rightarrow$ Shorter MCMC runs
- Expensive calculations (projection matrices) done beforehand

Basis Representations

Objective: Represent spatial random effects $\mathbf{W}$ with basis functions $\mathbf{\Phi}$

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

$$\mathbf{W} \approx \mathbf{\Phi}\delta, \qquad \delta \sim \mathcal{N}(0, \Sigma_\delta),$$

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

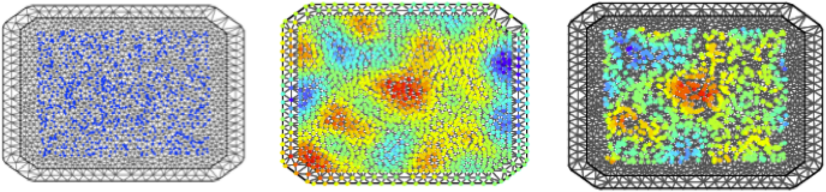
Sketch of Our Approach: PICAR

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

Part A:
Discretize continuous spatial domain using data

Part B:
Fit projected ICAR model on discrete space

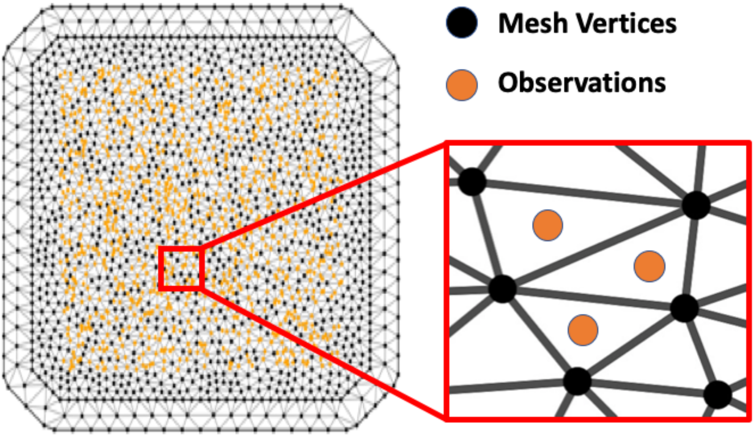
Part C:
Use discrete model to interpolate onto continuous space



Part A: Mesh Construction

Use data to discretize continuous spatial domain into a triangular mesh (see Hjelle and Dæhlen (2006))

- Generate mesh vertices $\tilde{\mathbf{W}} \in \mathbb{R}^m$ with $m > n$



Part B: Moran's Basis Functions

$\boxed{\text{Step 1:}}$ Create spatial basis functions on the mesh (m #nodes)

- Moran's basis $\mathbf{M} \in \mathbb{R}^{m \times p}$: first p eigenvectors of Moran's operator
- Moran's Operator (Griffith, 2003; Hughes and Haran, 2013)

$$(\mathbf{I} - \mathbf{11}'/\mathbf{m})\mathbf{A}(\mathbf{I} - \mathbf{11}'/\mathbf{m}) \in \mathbb{R}^{m \times m},$$

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

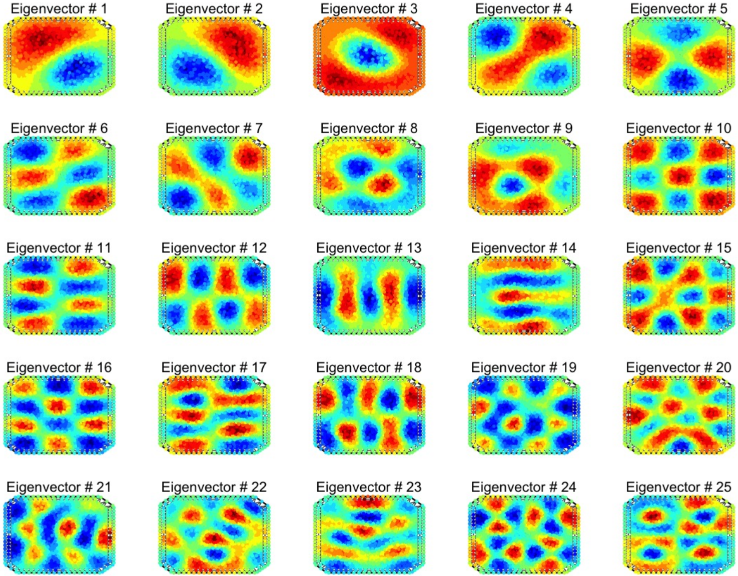
$\boxed{\text{Step 2:}}$ Construct spatial random field on mesh with basis functions

- $\tilde{\mathbf{W}} = \mathbf{M}\delta$, where $\tilde{\mathbf{W}}$ are mesh nodes and $\delta \in \mathbb{R}^p$ are weights.

Result:

1. Dimension Reduction: $\delta \in \mathbb{R}^p$ where $p << m$.
2. De-correlated Weights:
   Orthogonal basis functions $\Rightarrow$ weights δ are de-correlated

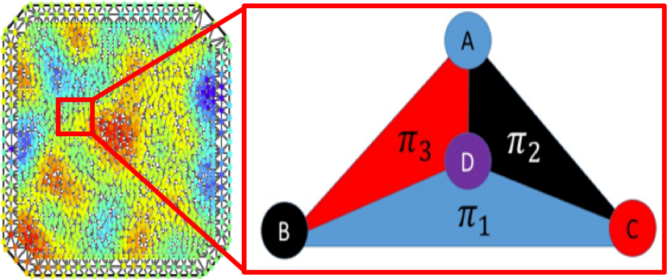
Eigenvectors of Moran's Operator



Part C: Piece-wise Linear Basis Functions

Piece-wise linear basis functions contained in an $n \times m$ projector matrix **A**.

- Projector matrix **A** interpolates points within triangular mesh
- Mesh nodes A, B, and C + Observation location D.



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

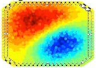
Result: Interpolate $\mathbf{W} \approx \mathbf{A}\tilde{\mathbf{W}}$, where $\mathbf{W} \in \mathbb{R}^n$ are observation locations and $\tilde{\mathbf{W}} \in \mathbb{R}^{\mathbf{m}}$ are mesh nodes.

Role of Basis Functions

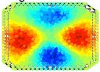
M

Moran's Basis

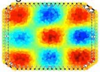
Eigenvector # 1



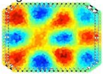
Eigenvector # 5



Eigenvector # 10



Eigenvector # 15



δ

Coefficients/ Weights

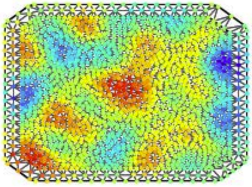
δ_1

δ_5

δ_{10}

δ_{15}

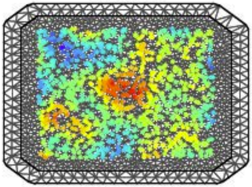
Mδ



Idea:

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

AMδ



Idea:

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

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 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 hierarchical spatial model (gold standard)
2. PICAR with varying ranks for Moran's basis ($\mathbf{M}$).

Simulated Example

Binary Spatial Data: 1,000 for fitting + 400 for validation
Used CVMSPE: cross-validation mean squared prediction error (MSPE) + time (minutes) for same run length

Rank	β_1 (95% CI)	β_2 (95% CI)	MSPE	Time
10	1.04 (0.77,1.31)	0.91 (0.64,1.16)	0.3	9.73
22 (Selected)	**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
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

Gains from fast mixing in MCMC + reducted cost per iteration:
For the random effects $\mathbf{W}$, the average effective samples per second is 5.8 (PICAR) vs. 0.016 (gold standard). $\sim$ 345-fold increase

Summary

1. PICAR: new projection-based approach for hierarchical spatial models
2. Automated approach, with heuristic for rank selection
3. Fast per iteration ($\mathcal{O}(np)$) + fast mixing MCMC
4. Provides comparable inference and prediction to full model
5. Studied many simulated and real data examples
   - large data sets (20k+)
   - many models: spatial generalized linear mixed models, spatially-varying coefficient models, cumulative-logit models for ordinal data
   - works well in each case

6. Can easily fit user-specified models via `stan`

Summary

1. PICAR: new projection-based approach for hierarchical spatial models
2. Automated approach, with heuristic for rank selection
3. Fast per iteration ($\mathcal{O}(np)$) + fast mixing MCMC
4. Provides comparable inference and prediction to full model
5. Studied many simulated and real data examples
    - large data sets (20k+)
    - many models: spatial generalized linear mixed models, spatially-varying coefficient models, cumulative-logit models for ordinal data
    - works well in each case

6. Can easily fit user-specified models via `stan`

Lee, B. and Haran, M. (2019) PICAR: An Efficient Extendable Approach for Fitting Hierarchical Spatial Models
`https://arxiv.org/abs/1912.02382`

PICAR Version of a Spatial Hierarchical Model

Observation Model:

$$\mathbf{Z} \sim f(\boldsymbol{\eta}|\beta, \delta)$$

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

where $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the projector matrix, $\mathbf{M} \in \mathbb{R}^{m \times p}$ is the Moran's basis functions matrix and $\delta \in \mathbb{R}^{p}$ are the weights.

Process Model: (Hughes and Haran, 2013)

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

where $\mathbf{Q}$ is the $m \times m$ precision matrix for an ICAR spatial model.

Priors:

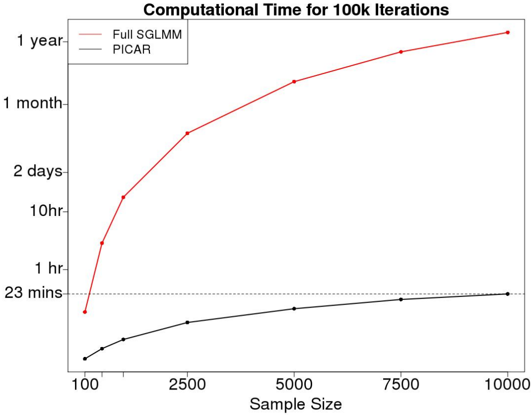
$$\tau \sim G(\alpha_\tau, \beta_\tau), \qquad \beta \sim N(0, \Sigma_\beta),$$

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

Computational Cost

PICAR scales linearly in n.

- **Full hierarchical spatial model (gold standard):** $\mathcal{O}(n^3)$
- **PICAR:** $\mathcal{O}(np)$, where $p = \text{rank}(\mathbf{M})$ with $p << n$

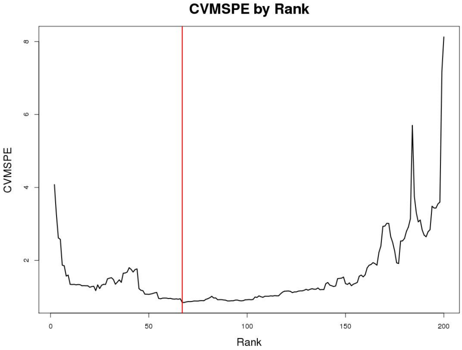


Automated Heuristic for Rank Selection

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

Automated Heuristic:

1. Construct Moran's basis $\mathbf{M}$ with varying p.
2. Fit a generalized linear model using basis functions as covariates (i.e. columns of $\mathbf{AM}$).
3. Choose p with lowest out-of-sample prediction error (CVMSPE).



Count Data with Spatially Varying Coefficients

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

- We use stan, 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

Table: Results for the simulated example of a spatially-varying coefficients model.

Example Implementation in **stan**

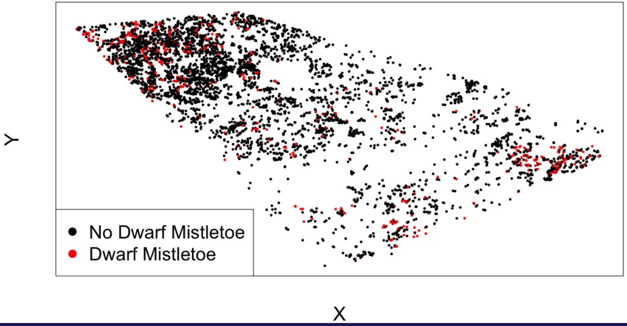
```
transformed parameters {
  vector[N] linpred;
  linpred = X*beta+M*delta+ X[,1].*(M*beta1S);
}

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 is used to manufacture paper.

Observations

- $n = 22,888$ 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 $\mathbf{M}$ for rank $p = 520$
- $\sim$ 2 hours to draw 100k samples from the posterior distribution.

Covariate	Mean (95% CI)
Age	0.65 (0.47,0.83)
Basal area	-0.16 (-0.35,0.02)
Height	2.53 (2.16,2.91)
Volume of spruce stand	-0.2 (-0.33,-0.07)

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

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.

Hjelle, Ø. and Dæhlen, M. (2006). *Triangulations and applications*. Springer Science & Business Media.

References III

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.

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

Hierarchical Spatial Models (Gaussian Observations)

Data Model:

$$Z(s)|W(s) \sim \mathcal{N}(X(s)\beta + W(s), \tau^2 I)$$

Process Model:

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

Parameter Model:

Priors for β, ϕ, σ^2, and τ^2.

Notation:

- $Z(s)$: Observations
- $X(s)$: Matrix of covariates
- $W(s)$: Spatial random effects
- R_ϕ: Correlation function
- ϕ: Range parameter
- σ^2: Partial sill parameter
- τ^2: Nugget/measurement error

Goal: Marginalize out $W(s)$ and infer β, ϕ, σ^2, and τ^2.

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.

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:

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

where $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{M} \in \mathbb{R}^{m\times p}$ are the basis function matrices.

Key Approximation:

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

Challenge: How do we improve this approximation?

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.

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.

Mixing in MCMC algorithms:
For the random effects $\mathbf{W}$, the average effective samples per second is 5.8 (PICAR) vs. 0.016 (gold standard), a $\sim$ 345x improvement!

Poisson Simulated Example Results

Table: Simulated example with count spatial observations. Parameter estimation, prediction, and model fitting time results across precision matrices.

Precision Matrix	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	1.09 (0.99,1.19)	1.01 (0.92,1.11)	1.96	8.84
50	1.05 (0.95,1.15)	1.02 (0.92,1.12)	1.74	9.87
62	**1.04 (0.94,1.14)**	**0.99 (0.89,1.09)**	**1.57**	**10.65**
75	1.03 (0.93,1.14)	0.99 (0.89,1.09)	1.66	10.39
100	1.05 (0.95,1.16)	0.98 (0.88,1.09)	1.71	11.07
200	1.08 (0.97,1.19)	0.98 (0.87,1.1)	1.81	13.49
Gold Standard	1.07 (0.97,1.17)	1.01 (0.91,1.12)	1.66	3803.84

Simulation Study: Binary Data

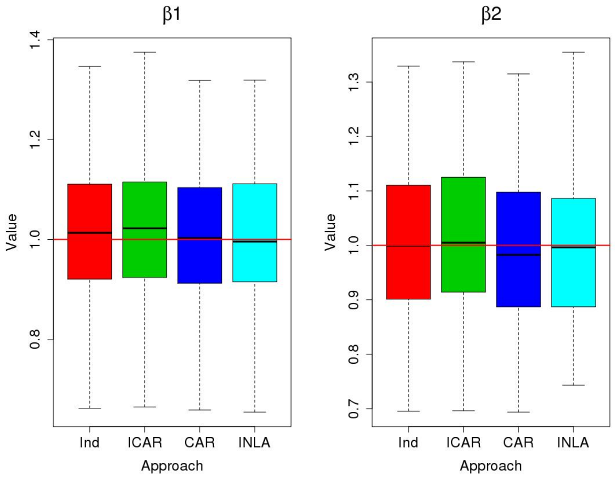


Figure: **Binary Data Simulation Study.** Posterior mean of coefficients β_1 and β_2 from 100 samples.

Simulation Study: Poisson Data

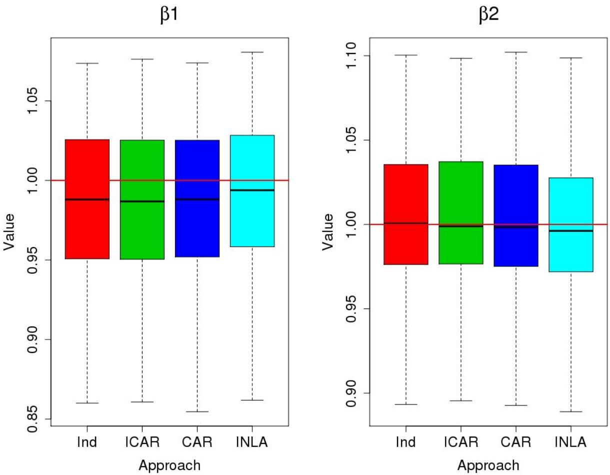


Figure: **Poisson Data Simulation Study.** Posterior mean of coefficients β_1 and β_2 from 100 samples.

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.

Tuning Mechanisms

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

Tuning Mechanism #1: Rank Selection

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

Automated Heuristic:

1. Construct Moran's basis $\mathbf{M}$ for varying p.
2. Fit GLM model using columns of $\mathbf{AM}$ as covariates.
3. Select p with lowest out-of-sample prediction error (CVMSPE).

Tuning Mechanism #2: Precision Matrix

Goal: Specify prior precision matrices Q.

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

- 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 W poses major computational challenges

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)

Simulated Example: Map of Obsrvations