A Unified Approach to Spatial Modeling Using Markov Random Fields?

Chris Paciorek
Department of Statistics; University of California, Berkeley
and
Department of Biostatistics; Harvard School of Public Health

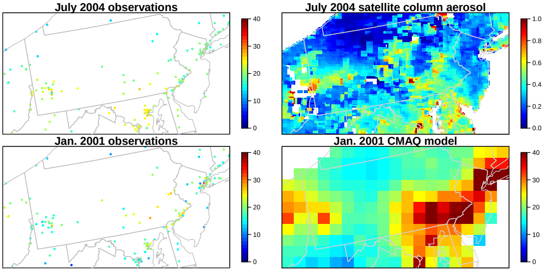
www.biostat.harvard.edu/˜paciorek

Research supported by U.S. NCI P01 CA134294-02

February 2011

Modeling point and areal data

- Both point and areal data are common.
- Particularly with the increase in computer code output and remote sensing information, many environmental analyses involve both point and areal data.



- How could we model both types of data using a single framework?

Latent process representations

1. Use a latent Gaussian process.
   1. Relate the process to point locations in the standard way.
   2. One can take integrals over areas to derive the areal data distribution (Kelsall and Wakefield 2002, Fuentes and Raftery 2005)
2. Use a Markov random field on a fine regular grid.
   1. Relate point observations to the relevant grid cells, possibly with covariate adjustments for within-cell heterogeneity.
   2. Relate areal observations to weighted averages of grid cells overlapped by a given area.

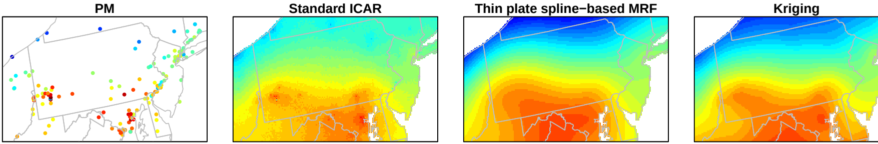
Main points

1. MRF models can be used to model both point and area data, and of course have nice computational properties.
2. Standard nearest neighbor MRF models ('ICAR') represent processes that are not smooth.
    - Only certain higher-order MRF models make sense. Simply expanding the neighborhood gives unexpected results.

Using a standard MRF model for point data

Options:

1. Create areas (e.g., by tesselation) centered around each observation.
   - Obvious problems with prediction and if new data arrive.
2. An alternative: use the MRF to represent a latent process on a fine grid and relate observations to relevant grid cells.



Proposed MRF-based GLMM

1. For Y_i either a point or area observation, let $\mu_i = E(Y_i|X_i, g)$, with

$$h(\mu_i) = X_i\beta + K_i g$$

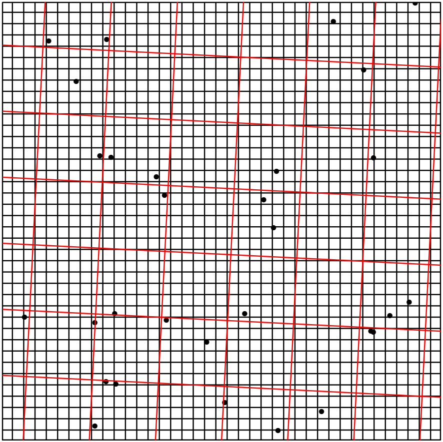
where K_i is the ith row of a mapping matrix, K.

2. Represent the unknown, latent spatial process, g, as a piecewise constant surface on a fine regular grid,

$$g \sim \mathcal{N}(0, (\kappa Q)^{-})$$

where Q is a (singular) MRF precision matrix.

Mapping data to the fine grid



MRF models

- **ICAR** - intrinsic (improper) conditional auto-regressive model: neighbors given a weight of one, all others a weight of zero.
- **TPS-MRF** - MRF approximation to a thin plate spline: weights derived from discrete approximation on grid to the usual thin plate spline penalized likelihood (Rue and Held 2005, Yue 2009 PhD).

$$J(g) = \int \int_{\Re^2} \left[\left(\frac{\partial^2 g(s_1, s_2)}{\partial s_1^2} \right)^2 + 2 \left(\frac{\partial^2 g(s_1, s_2)}{\partial s_1 \partial s_2} \right)^2 + \left(\frac{\partial^2 g(s_1, s_2)}{\partial s_2^2} \right)^2 \right] ds_1 ds_2.$$

- **HICAR** - ICAR with equal weights on higher-order neighbors.
- **DICAR** - ICAR with weights decaying with distance (Hrafnkelsson & Cressie 2003; Pettit et al. 2002)
- **EAR** - extended auto-regressive model (Linder, unpub.)
- **Matern-MRF** - MRF approximation to a Matern covariance GP (Lindgren et al. 2011)
  - TPS-MRF is a limiting case for Matern with $\nu = 1$.

MRF Model Precision Matrices

Matrix elements for a single location (single row)

ICAR

	−1	
−1	4	−1
	−1	

TPS−MRF

		1		
	2	−8	2	
1	−8	20	−8	1
	2	−8	2	
		1		

HICAR(3)

			−1			
	−1	−1	−1	−1	−1	
	−1	−1	−1	−1	−1	
−1	−1	−1	28	−1	−1	−1
	−1	−1	−1	−1	−1	
	−1	−1	−1	−1	−1	
			−1			

DICAR(3)

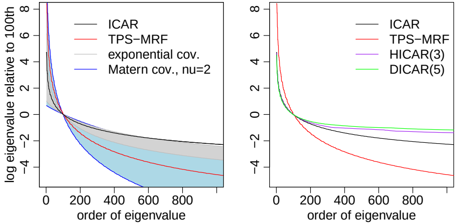
			−0.05			
	−0.06	−0.11	−0.15	−0.11	−0.06	
	−0.11	−0.39	−1	−0.39	−0.11	
−0.05	−0.15	−1	7.5	−1	−0.15	−0.05
	−0.11	−0.39	−1	−0.39	−0.11	
	−0.06	−0.11	−0.15	−0.11	−0.06	
			−0.05			

What do we know about ICAR model properties?

- Besag and Mondal (2005): ICAR model approaches two-dimensional Brownian motion (de Wijs process) as the grid resolution increases.
  - Continuous but non-differentiable sample paths.
- CAR models in one dimension are equivalent to AR(1) models in discrete time, which can be expressed as GPs for discrete time points with exponential covariance.
- Question: can higher-order neighbors (HICAR) and distance-based weights (DICAR) give us smoother representations?

Eigenvalues of MRF inverse precision matrices

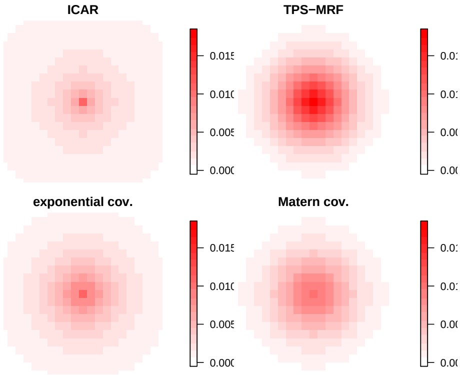
$Q^{-} = \Gamma \Lambda^{-} \Gamma^{\top}$; 100 x 100 grid



- ICAR model puts high weight on high-frequency eigenvectors and low weight on low-frequency eigenvectors; even more extreme than exponential-based GP.
- Higher-order (HICAR) and distance-based (DICAR) versions are even more extreme.
- [Visual comparison indicates eigenvectors are essentially the same. Projection of various model Q matrices on ICAR eigenvectors gives very similar results.]

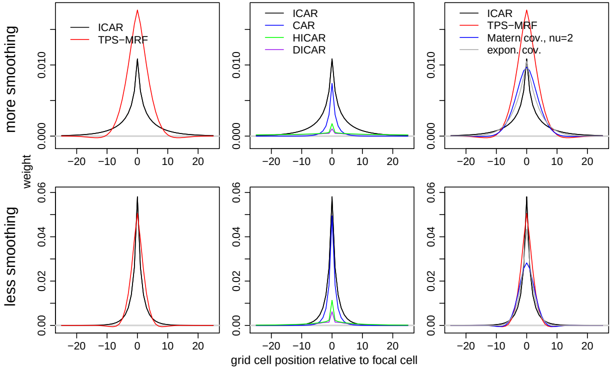
Equivalent kernels of MRF specifications, $\hat{g} = Sy$

Two-dimensional image



Equivalent kernels of MRF specifications

One-dimensional cross-section

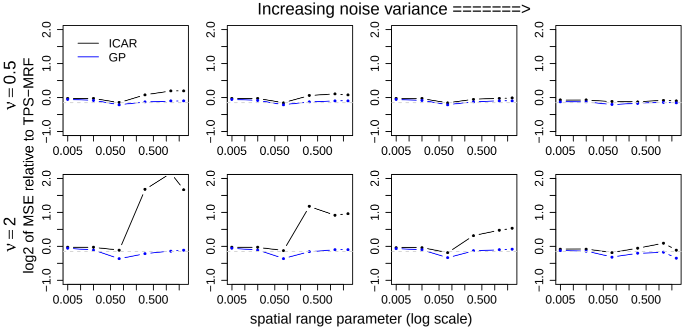


- Note heavy ICAR weight on focal cell, reduced weight nearby and large weight in tails.
    - Recall the weird behavior in the ICAR fit of point data.

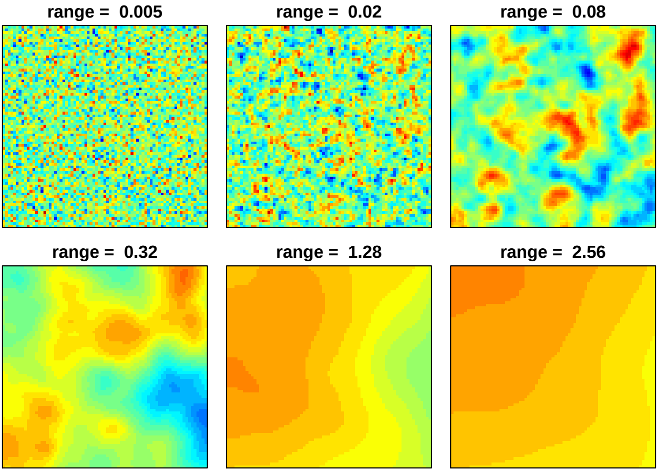
Analytic comparison of prediction error

n=100; averages over multiple sets of randomly-sampled spatial locations

- The MSE, $E_g E_Y(\sum_i(\hat{g}-g)^2)$, where $\hat{g}=(I+\lambda Q)^{-1}Y$, can be calculated analytically and only λ is unknown in the fitting.
- This suggests an 'oracle' analysis where we set the penalty parameter, $\lambda=\tau^2\kappa$ fixed at the value that optimizes MSE.

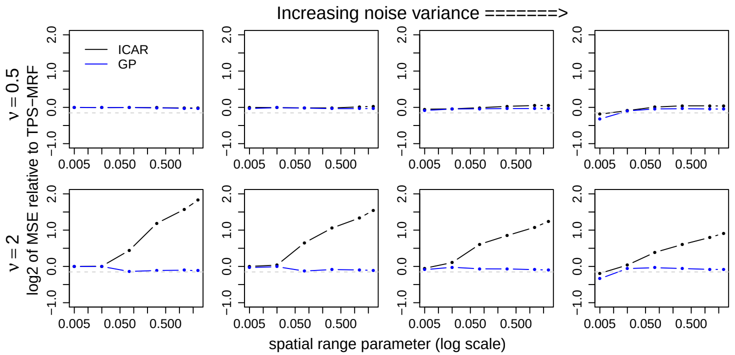


Surface types considered

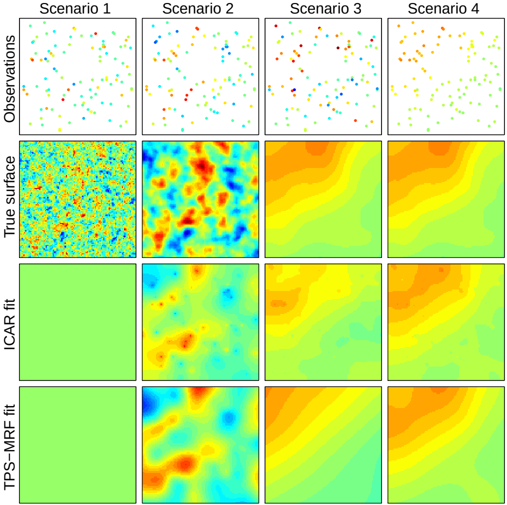


Analytic comparison of prediction error (2)

10000 observations; one per grid cell

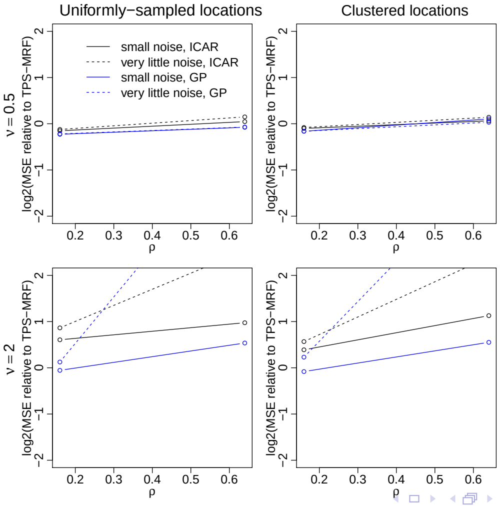


Example fits of ICAR and TPS-MRF models



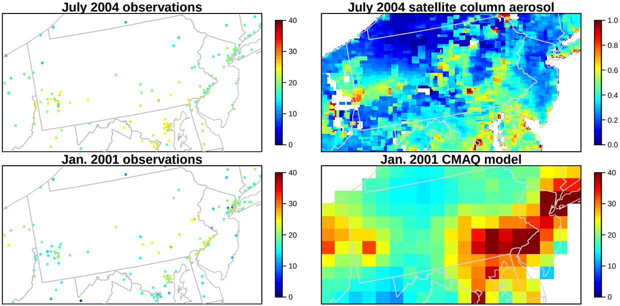
Simulation results

$n = 100$, little noise



Example: Modeling particulate matter (PM)

- Regulatory agencies collect PM monitoring data (point data).
- Proxy information available from atmospheric modeling and from remote sensing, but quality is an issue.
  - I'll call the systematic difference between the truth and the proxy 'discrepancy'.



Flexible Spatial Discrepancy Modeling

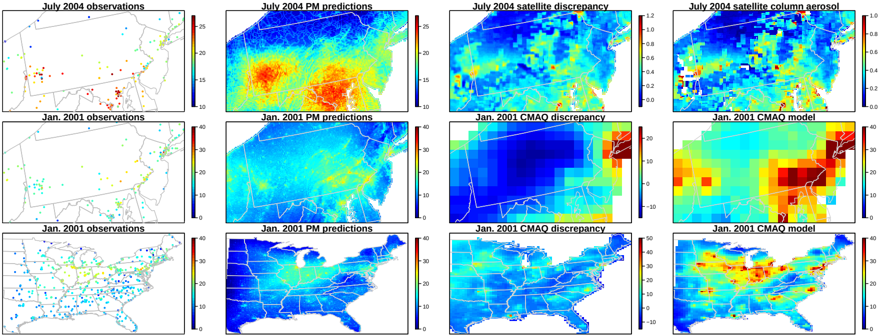
- Consider additive bias as a spatial discrepancy process, $D(\cdot)$:

$$\begin{aligned} Y &\sim \mathcal{N}(\mu_y(x) + K_y L, \sigma_y^2) \\ A &\sim \mathcal{N}(K_A D + \beta_1 K_A L, \sigma_a^2) \\ L &\sim \text{MRF}(\mu_L(x), Q_L) \\ D &\sim \text{MRF}(\mu_D(x), Q_D) \end{aligned}$$

- Latent processes, $L(\cdot)$ and $D(\cdot)$, are represented on a fine grid.
- We can explore the relationship of the proxy and gold standard through analysis of the spatial scales of $D(\cdot)$.
- $\mu_y(x)$ involves the effect of covariates that explain sub-grid scale variation in the point measurements, while $\mu_L(x)$ and $\mu_D(x)$ are covariate effects on the grid-scale process and the discrepancy term, respectively.

Predicted PM

Y | $\mathbf{PM = L}$ | **D** | **A**



Conclusions from the example

- The MRF approach helped in terms of computational challenges and combining information.
- In the example, the proxy information did not help with prediction - including the discrepancy component was critical for reasonable modeling.
- More statistical methodology is needed for combining data with highly-structured proxy information in the face of discrepancy.

Computation for the basic model

1. With normal data, one has a marginal likelihood (integrating over g) that involves large, but sparse, matrices and therefore efficient calculation.
    1. This allows either maximization or MCMC in the hyperparameter space.
2. For non-normal data:
    1. PQL: Consider the model to be a GLMM with Kg playing the role of the usual Zb notation.
        - The forthcoming glmmGS package for R implements the PQL approach (Breslow and Clayton 1993) based on sparse matrix calculations for Q and K.
    2. INLA (Rue et al. 2009): software at r-inla.org should be able to fit the model proposed here, including making use of the sparsity of Q and K.

Other computationally-efficient spatial representations

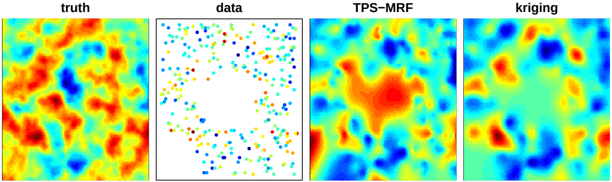
- Reduced-rank kriging (Kammann and Wand 2003; Banerjee et al. 2008)
  - Good for low-frequency surfaces but requires many knots for high-frequency variability.
- Tapering (Furrer et al. 2006; Kaufman et al. 2008)
  - Induces sparse covariance so effectiveness in modeling low-frequency surfaces is unclear.

Advantages of proposed MRF approach

- The approach handles both point and area data in an aggregation-consistent fashion.
- Computational efficiency
- MRF-based thin plate spline approximation provides for smooth latent spatial surfaces, unlike standard ICAR and extensions to higher-order neighbors.
- Basic extension to space-time based on specification of time 'neighbors' also provides for sparse matrix calculations and gives a Kronecker product structure for the precision.

Potential disadvantages of the proposed MRF approach (1)

- (1) Extrapolation issues:
  - GP specifications naturally handle extrapolation (on boundary or in 'large' gaps [relative to correlation range]) by mean reversion and increased uncertainty.
  - Basis functions local to gaps and near boundaries may be poorly estimated in spline models, causing 'ballooning'.



Potential disadvantages of the proposed MRF approach (2-3)

- (2) GP models have both marginal variance and a correlation range parameters, while MRF models have a single precision parameter.
  - Zhang (2004) shows non-identifiability in infill asymptotic regime with no nugget.
  - In applications/finite samples, does the additional GP parameter improve performance?
  - Hypothesis: spline models and GPs perform similarly when data are uniformly distributed in space, but with gaps [relative to correlation range], GPs may be more robust.
- (3) Nonstationarity is not addressed, but note Yue and Speckman (2010) TPS-MRF extension with spatially-varying penalty and Lindgren et al. (2011)