



Markov Random Fields and Regional Climate Models

Stephan R. Sain

Geophysical Statistics Project

Institute for Mathematics Applied to Geosciences

National Center for Atmospheric Research

Boulder, CO



Supported by NSF ATM/DMS.

Outline

- Overview of Markov random fields.
- Case study I: A multivariate analysis of a regional climate model ensemble.
- Case study II: Spatial extremes.
- Case study III: Functional ANOVA and NARCCAP.

The Movie

(jan2002ncep.mov)

The Sequel

(jan2002models.mov)

Spatial Data

- Let $s \in \mathfrak{R}^d$ indicate a generic data location in a d -dimensional space.
 - Typically, $d = 1, 2$, or 3 .
- Let s vary over an index set $D \subset \mathfrak{R}^d$ to generate a random field $\{Y(s) : s \in D \subset \mathfrak{R}^d\}$.
 - $Y(s)$ is an feature observed at location s .
- Three types of spatial data:
 - Geostatistical data
 - Lattice or areal/regional data
 - Point patterns

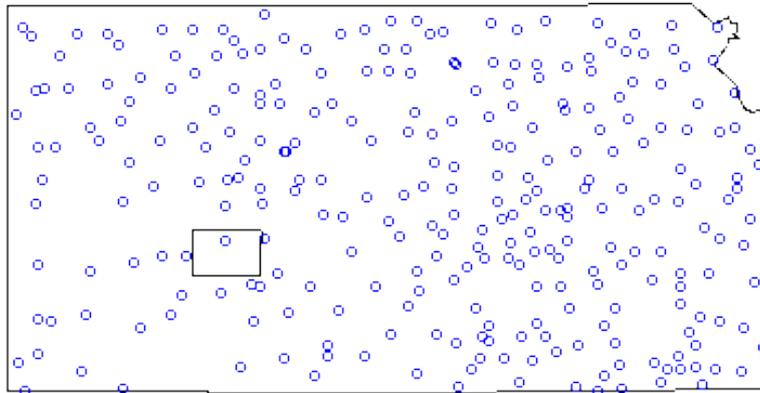
Geostatistical Data

- Let s vary over an index set $D \subset \mathbb{R}^d$ to generate a random field $\{Y(s) : s \in D \subset \mathbb{R}^d\}$.
 - D is a continuous, fixed set:
 - * $Y(s)$ can be observed everywhere within D
 - * the points in D are non-stochastic.
 - Y can either be continuous or discrete.
- Generally, some assumption of stationarity is made, a covariance function is adopted, and the goal is to reconstruct the underlying process that generated Y (Kriging).
 - Covariance is a function of the distance and/or direction between locations.

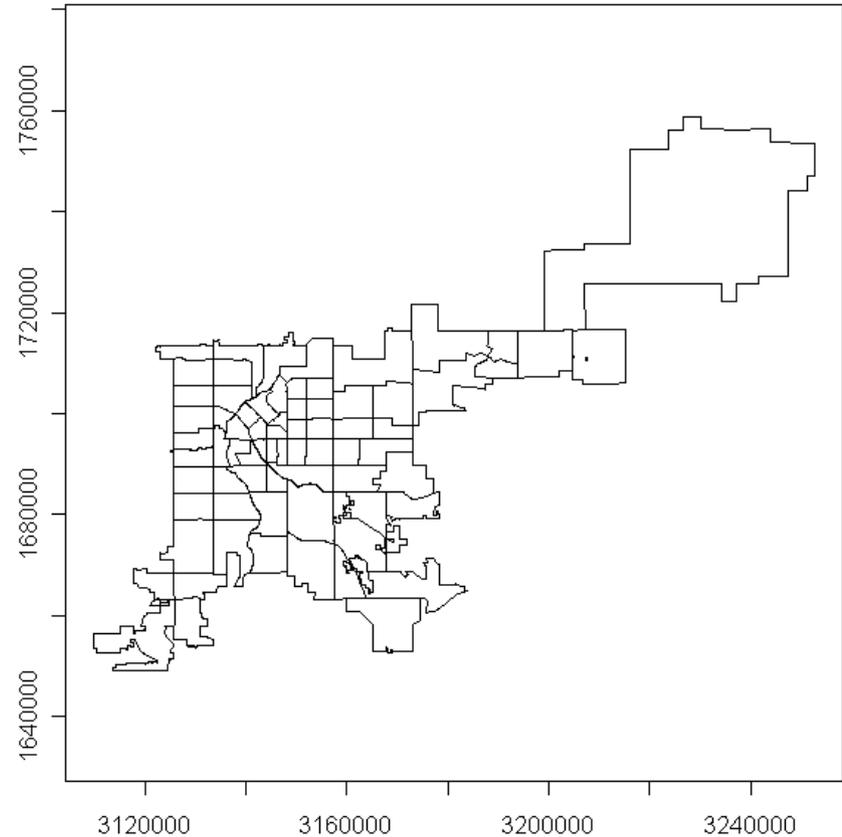
Lattice Data

- Let s vary over an index set $D \subset \mathfrak{R}^d$ to generate a random field $\{Y(s) : s \in D \subset \mathfrak{R}^d\}$.
- D is fixed and discrete, i.e. non-random and countable
- Y can either be continuous or discrete.
- Lattices can be regular, as on a grid, or irregular where there is no predictable pattern.
 - Examples: remote sensing, police precincts, zip-codes, census divisions, etc.
- Also referred to as regional or areal data.

Geostatistical vs Lattice Data



Kansas weather stations (2004)



Denver police precincts

Kriging Lattice Data?

- The goals of a spatial analysis of lattice data can be similar to geostatistical data, i.e. prediction, modeling, etc., but...
- The notation of $Y(s)$ might be somewhat misleading or confusing: is s a point location?
- Often, analysts define a representative point for a lattice site and use traditional geostatistical methods for analysis.
 - Many issues arise, in particular for irregular lattices: arbitrariness of representative points and distances, aggregation (e.g. unequal variances, observations not continuous, etc.), lack of well-defined locations for prediction, etc.
- Need for more formal approaches...

Spatial Autoregressive Models

- Geostatistical methods model spatial dependence through specification of a covariance function based on the distances between points.
- Spatial autoregressive models represent the data at a spatial location as a linear combination of neighboring locations.
 - Spatial dependence is induced through this autoregression and the neighborhood structure in the data.
- Two formulations: simultaneous autoregressive (SAR) models and conditional autoregressive (CAR) models.
 - Unlike temporal autoregressive models, these formulations do not necessarily yield the same model.

A Look Back...

- Consider a simple AR(1) time series model through a simultaneous specification:

$$Z_t = \mu + \rho(Z_{t-1} - \mu) + \epsilon_t, \quad i = 1, \dots, n,$$

where ϵ_t is a white-noise process.

- In matrix form:

$$\mathbf{Z} = \mu + \rho\mathbf{H}(\mathbf{Z} - \mu) + \epsilon$$

where

$$\mathbf{H} = \begin{pmatrix} \mathbf{0}' & \mathbf{0} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}$$

A Look Back...

- We can also define the conditional distributions:

$$f(Z_t|Z_{t-1}) \sim \mathcal{N}(\rho Z_{t-1}, \sigma^2), \quad i = 1, \dots, n.$$

- Both specifications give rise to a joint distribution:

$$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

SARs

- The SAR model is specified via

$$y_i = \mu_i + \sum_{j=1}^n g_{ij}(y_j - \mu_j) + \epsilon_i, \quad i = 1, \dots, n,$$

where

- μ_i is the mean of y_i
- ϵ_i are zero-mean, uncorrelated random shocks
- g_{ij} are spatial weights
 - * $g_{ii} = 0$
 - * In general, g_{ij} does not have to equal g_{ji} .
 - * Typically, $g_{ij} = 0$ if $j \notin N_i$.

SARs

- In matrix form,

$$\begin{aligned} \mathbf{Y} - \boldsymbol{\mu} &= \mathbf{G}(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\epsilon} \\ (\mathbf{I} - \mathbf{G})(\mathbf{Y} - \boldsymbol{\mu}) &= \boldsymbol{\epsilon}. \end{aligned}$$

- Some properties:

- $E[\mathbf{Y}] = \boldsymbol{\mu}$

- $\text{var}[\mathbf{Y}] = (\mathbf{I} - \mathbf{G})^{-1} \mathbf{S} (\mathbf{I} - \mathbf{G}')^{-1}$ with $\mathbf{S} = \text{var}(\boldsymbol{\epsilon})$ (diagonal)

- $\text{cov}[\boldsymbol{\epsilon}, \mathbf{Y}] = \text{cov}[\boldsymbol{\epsilon}, (\mathbf{I} - \mathbf{G})^{-1} \boldsymbol{\epsilon}] = (\mathbf{I} - \mathbf{G})^{-1} \mathbf{S}$

- Note that $\boldsymbol{\epsilon}$ and \mathbf{Y} are not independent, i.e. the shock at the i th site is not independent of the autoregressive variable at the j th site.

MRFs and CARs

- Besag (1974) showed that the collection of conditional distributions $f(y_i|y_{-i})$, $i = 1, \dots, n$ can be combined to form a joint distribution $f(y_1, \dots, y_n)$.

- The collection of Gaussian conditionals with

$$E[y_i|y_{-i}] = \mu_i + \sum_{j=1}^n b_{ij}(y_j - \mu_j) \quad \text{and} \quad \text{Var}[y_i|y_{-i}] = \tau_i^2,$$

gives rise to a joint Gaussian distribution,

$$\mathcal{N}(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{B})^{-1}\mathbf{M}),$$

so long as some symmetry conditions are met and the b_{ij} are chosen to ensure a positive-definite covariance matrix.

Factorization Theorem

- Assume we have a family of one-dimensional conditional distributions of the form $p(x_i|x_j, j \neq i)$. Then, for some fixed reference point (x_1^*, \dots, x_n^*) ,

$$\begin{aligned} & \frac{p(x_1, \dots, x_n)}{p(x_1^*, \dots, x_n^*)} \\ &= \prod_{i=0}^{n-1} \frac{p(x_1^*, \dots, x_i^*, x_{i+1}, x_{i+2}, \dots, x_n)}{p(x_1^*, \dots, x_i^*, x_{i+1}^*, x_{i+2}, \dots, x_n)} \\ &= \prod_{i=0}^{n-1} \frac{p(x_{i+1}|x_1^*, \dots, x_i^*, x_{i+2}, \dots, x_n)}{p(x_{i+1}^*|x_1^*, \dots, x_i^*, x_{i+2}, \dots, x_n)} \end{aligned}$$

MRFs and CARs

- Some conditions on $\{b_{ij}\}$:
 - $b_{ii} = 0$ and $b_{ij} = 0$ if $j \notin N_i$
 - $b_{ij}\tau_j^2 = b_{ji}\tau_i^2$ (symmetry)
 - Generally, the non-zero b_{ij} are assumed to be proportional to some fixed constants.
- The off-diagonal elements of the inverse covariance matrix are either:
 - zero, implying conditional independence between observations that are not neighbors, or
 - $-b_{ij}/\tau_i^2$, implying conditional dependence.

CAR vs SAR

- A matrix representation for the CAR model gives

$$\mathbf{Y} - \boldsymbol{\mu} = \mathbf{B}(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\delta}$$

where $\boldsymbol{\delta} = (\mathbf{I} - \mathbf{B})(\mathbf{Y} - \boldsymbol{\mu})$ are “pseudo-errors”.

– Note that

$$\text{cov}[\boldsymbol{\delta}, \mathbf{Y}] = \text{cov}[(\mathbf{I} - \mathbf{B})(\mathbf{Y} - \boldsymbol{\mu}), \mathbf{Y}] = (\mathbf{I} - \mathbf{B})\text{var}[\mathbf{Y}] = \mathbf{M}$$

- The matrix \mathbf{M} is diagonal, suggesting that the shock at location i is independent of the autoregressive variable at the j th site.

More on CAR vs SAR

- Assuming the means are the same, then the SAR and the CAR specification are the same if and only if

$$(\mathbf{I} - \mathbf{B})^{-1}\mathbf{M} = (\mathbf{I} - \mathbf{G})^{-1}\mathbf{S}(\mathbf{I} - \mathbf{G}')^{-1}$$

- Since \mathbf{M} is diagonal, any SAR can be represented as a CAR, but not vice versa – hence the CAR is more general.
- The CAR model immediately gives rise to the best (mean squared prediction error) predictor.
- There are issues with identifiability and consistency when estimating spatial dependence parameters g_{ij} for SAR models.

More on CAR vs SAR

- Likelihood computations with both SAR and CAR models are expensive, but the spatial dependence matrices (\mathbf{G} and \mathbf{B}) are typically sparse, making storage and computation more efficient.
- The conditional nature of the CAR specification and its interpretation has advantages when extending to multivariate spatial models and in conjunction with a hierarchical Bayesian model.

MRFs and CARs

- A very simple CAR covariance can be written as

$$(\mathbf{I} - \mathbf{B})^{-1}\mathbf{M} = \sigma^2(\mathbf{I} - \phi\mathbf{C})^{-1}$$

where

- $\mathbf{M} = \sigma^2\mathbf{I}$ (homogeneity) and \mathbf{C} is an adjacency matrix
 - ϕ is a spatial dependence parameter ($b_{ij} = \phi I_{j \in N_i}$)
- The conditional mean simplifies to

$$E[y_i | y_{-i}] = \mu_i + \phi \sum_{j \in N_i} (y_j - \mu_j)$$

- ϕ can be interpreted as partial or conditional correlation between two neighboring locations.

Multivariate CAR Models

- Let \mathbf{Y}_i be a p -dimensional random vector with a Gaussian conditional distribution with

$$E[\mathbf{Y}_i | \mathbf{Y}_{-i}] = \boldsymbol{\mu}_i + \sum_{j \in N_i} \boldsymbol{\Lambda}_{ij} (\mathbf{Y}_j - \boldsymbol{\mu}_j) \quad \text{var}[\mathbf{Y}_i | \mathbf{Y}_{-i}] = \boldsymbol{\Gamma}_i.$$

- Assuming

- $\boldsymbol{\Lambda}_{ij} \boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_i \boldsymbol{\Lambda}'_{ji}$ for $i, j = 1, \dots, n$ (symmetry)

- * $\boldsymbol{\Lambda}_{ii} = -\mathbf{I}$ and $\boldsymbol{\Lambda}_{ij} = \mathbf{0}$ for $j \notin N_i$

- $\text{Block}(-\boldsymbol{\Gamma}_i^{-1} \boldsymbol{\Lambda}_{ij})$ or $\text{Block}(-\boldsymbol{\Lambda}_{ij})$ is positive-definite

then $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$ is $N_{np}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_n)' \quad \text{and} \quad \boldsymbol{\Sigma} = \left(\text{Block}(-\boldsymbol{\Gamma}_i^{-1} \boldsymbol{\Lambda}_{ij}) \right)^{-1}.$$

Multivariate CAR Models

- The joint distribution: $\mathbf{Y} \sim N_{np} \left(\boldsymbol{\mu}, \left(\text{Block}(-\boldsymbol{\Gamma}_i^{-1} \boldsymbol{\Lambda}_{ij}) \right)^{-1} \right)$.
- Let:
 - $\boldsymbol{\mu}'_i = \mathbf{X}'_i \boldsymbol{\beta}$ where \mathbf{X}_i is a known q -vector for location i and $\boldsymbol{\beta}$ is a $q \times p$ matrix of parameters.
 - $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}$
 - $\boldsymbol{\Lambda}_{ij} = \boldsymbol{\Lambda}$ for $i < j$, $j \in N_i$.
 - * $\boldsymbol{\Lambda}_{ji} = \boldsymbol{\Lambda}'$.

Multivariate CAR Models

- The joint covariance can be written as

$$\Sigma = \Gamma^* \mathbf{H}^{-1} \Gamma^{*'}$$

where

$$\Gamma = \mathbf{I}_n \otimes \Gamma^{1/2}$$

and

$$\mathbf{H} = \begin{bmatrix} \mathbf{I} & -\mathbf{B}I(2 \in N_1) & \dots & -\mathbf{B}I(n \in N_1) \\ -\mathbf{B}'I(1 \in N_2) & \mathbf{I} & \dots & -\mathbf{B}I(n \in N_2) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{B}'I(1 \in N_n) & -\mathbf{B}'I(2 \in N_n) & \dots & \mathbf{I} \end{bmatrix}$$

- Note the reparameterization: $\mathbf{B} = \Gamma^{-1/2} \mathbf{\Lambda} \Gamma^{1/2}$

Multivariate CAR Models

- The form and value of \mathbf{B} ensures \mathbf{H} (and Σ) is positive-definite and controls the nature of the spatial dependence and interactions.
- The conditional mean can still be thought of as a weighted average of the observations at neighboring locations.
 - Weights are complicated functions of the within location correlation and spatial dependence parameters.
- The conditional or partial correlation is a function of both Γ and \mathbf{B} .

Thanks!



ssain@ucar.edu
www.image.ucar.edu/~ssain

- Sain, S.R., Furrer, R., and Cressie, N. (2008), “Combining ensembles of regional climate model output via a multivariate Markov random field model.”
- Cooley, D. and Sain, S.R. (2008), “Spatial hierarchical modeling of precipitation extremes from a regional climate model.”
- Kaufman, C. and Sain, S. (2008), “Bayesian functional ANOVA modeling using Gaussian process prior distributions.”