Markov random fields

Help from David Bolin, Johan Lindström and Finn Lindgren

Julian Besag
1945-2010

Håvard Rue
1965-

Finn Lindgren
1973-

The big N problem

Log likelihood:

$$\ell(\sigma,\theta) = -\frac{n}{2}\log(2\pi\sigma) - \frac{1}{2}\log\det\Sigma(\theta) + \frac{1}{2\sigma}(Z-\mu(\theta)^{T}\Sigma(\theta)^{-1}(Z-\mu(\theta))$$

Prediction:

$$E(Z(s_0)\,|\,Z_1,\ldots,Z_N,\hat{\theta}) = \mu_0 + \Sigma_{0,Z}\Sigma_{Z,Z}^{-1}(Z-\mu_Z)$$

Covariance has $O(N^2)$ unique elements

Inverse and determinant take $O(N^3)$ operations

The Markov property

Discrete time:

$$(X_k \mid X_{k-1}, X_{k-2}, \ldots) = (X_k \mid X_{k-1})$$

A time symmetric version:

$$(X_k \mid \underset{\sim}{X}^{-k}) = (X_k \mid X_{k-1}, X_{k+1})$$

A more general version:

Let A be a set of indices >k, B a set of indices <k. Then

$$X_A \perp X_B \mid X_k$$

These are all equivalent.

On a spatial grid

Let δ_i be the neighbors of the location i. The Markov assumption is

$$P(Z_i = z_i \mid \underset{\sim}{Z}^{-i} = \underset{\sim}{z}^{-i}) = P(Z_i = z_i \mid Z_{\delta_i} = z_{\delta_i})$$

$$= p_i(z_i \mid z_{\delta_i})$$

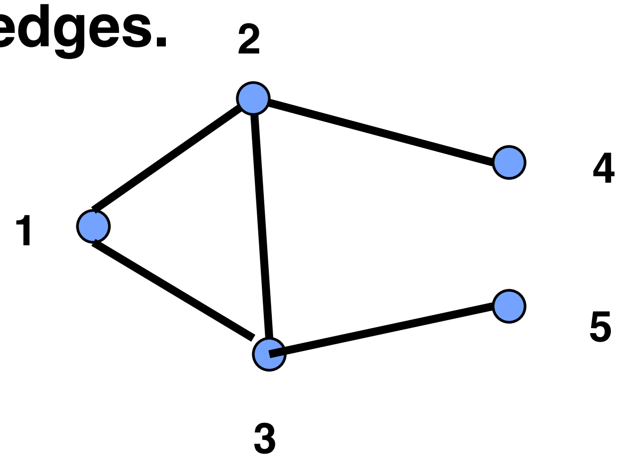
Equivalently for $i \notin \delta_j \quad Z_i \perp Z_j \mid \underset{\sim}{Z}^{-i,j}$

The p_i are called *local characteristics*. They are *stationary* if $p_i = p$.

A *potential* assigns a number $V_A(z)$ to every subconfiguration z_A of a configuration z. (There are lots of them!)

Graphical models

Neighbors are nodes connected with edges.



Given 2, 1 and 4 are independent.

Gibbs measure

The *energy* U corresponding to a potential V is $U(z) = \sum_A V_A(z)$.

The corresponding *Gibbs measure* is

$$P(z) = \frac{\exp(-U(z))}{C}$$

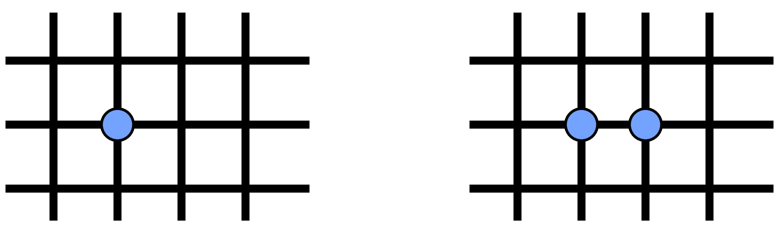
where $C = \sum_z \exp(-U(z))$

is called the *partition function*.

Nearest neighbor potentials

A set of points is a *clique* if all its members are neighbours.

A potential is a *nearest neighbor potential* if $V_A(z)=0$ whenever A is not a clique.



Markov random field

Any nearest neighbor potential induces a Markov random field:

$$p_i(z_i \mid \underset{\sim}{z}^{-i}) = \frac{P(\underset{\sim}{z})}{\sum_{z'} P(\underset{\sim}{z}')} = \frac{\exp(-\sum_{C\ \text{clique}} V_C(\underset{\sim}{z}))}{\sum_{z'} \exp(-\sum_{C\ \text{clique}} V_C(\underset{\sim}{z}'))}$$

where z' agrees with z except possibly at i, so $V_C(z)=V_C(z')$ for any C not including i.

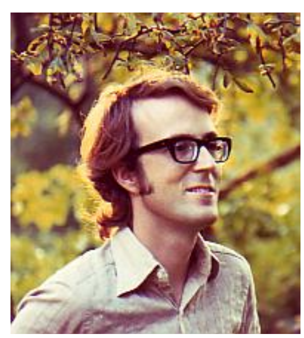
The Hammersley-Clifford theorem

Assume P(z)>0 for all z. Then P is a MRF on a (finite) graph with respect to a neighbourhood system Δ iff P is a Gibbs measure corresponding to a nearest neighbour potential.

Does a given nn potential correspond to a unique P?

John Hammersley
1920-2004

Peter Clifford
1944-



The Ising model

Model for ferromagnetic spin (values +1 or -1). Stationary nn pair potential $V(i,j)=V(j,i)$; $V(i,i)=V(0,0)=v_0$; $V(0,e_N)=V(0,e_E)=v_1$.

$$\operatorname{logit} P(Z_i = 1 \mid \underset{\sim}{Z}^{-i} = \underset{\sim}{z}^{-i})$$

$$= -(v_0 + v_1(z_{i+e_N} + z_{i-e_N} + z_{i+e_E} + z_{i-e_E}))$$

so $L(v) = \dfrac{\exp(t_0 v_0 + t_1 v_1)}{C(v)}$ **where**

$$t_0 = \sum z_i; \quad t_1 = \sum_i \sum_{j \sim i} z_i z_j$$

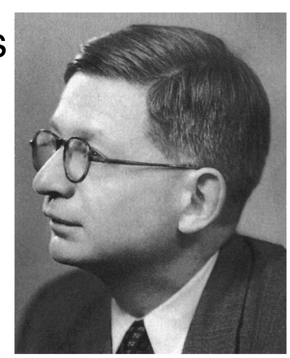
Interpretation

v_0 is related to the external magnetic field (if it is strong the field will tend to have the same sign as the external field)

v_1 corresponds to inverse temperature (in Kelvins), so will be large for low temperatures.

Ernst Ising
1900-1998

Rudolf Peierls
1907-1995



Phase transition

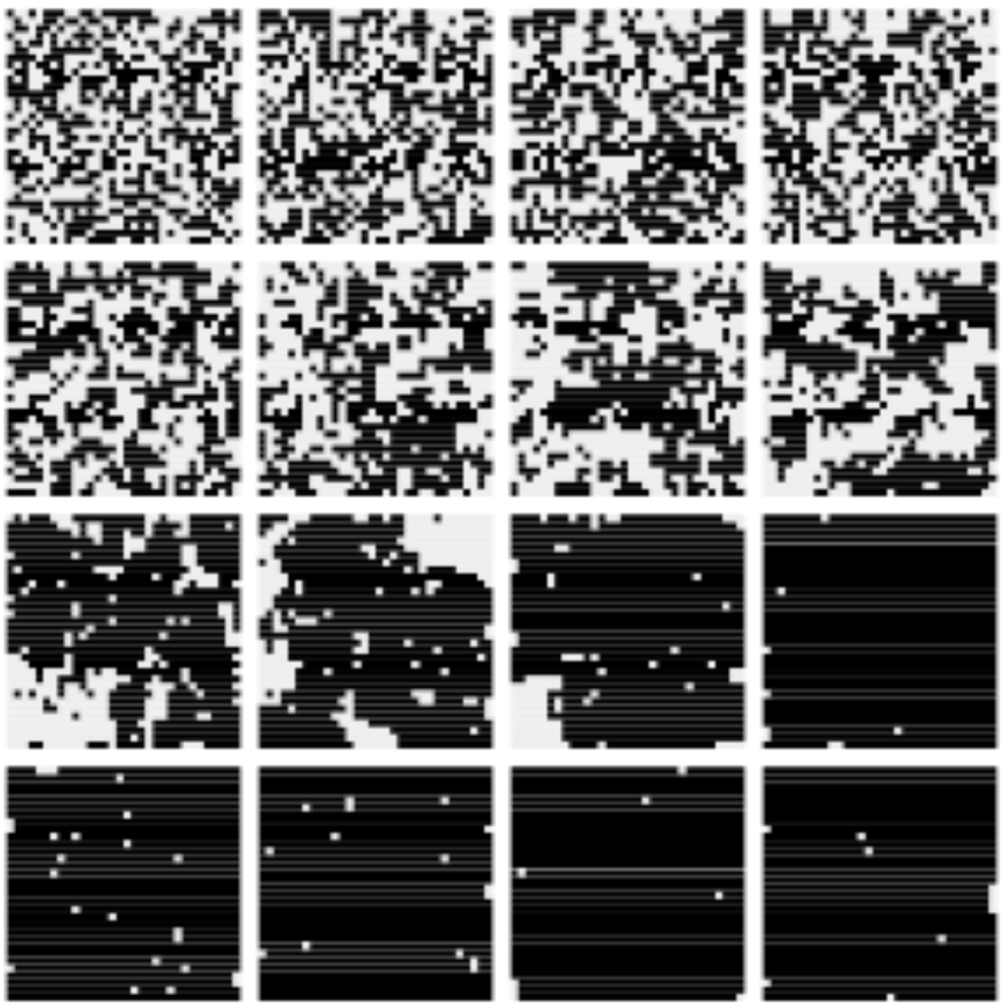
At very low temperature there is a tendency for spontaneous magnetization.

For the Ising model, the boundary conditions can affect the distribution of x_0.

In fact, there is a critical temperature (or value of v_1) such that for temperatures below this value, the boundary conditions are felt.

Thus there can be different probabilities at the origin depending on the values on an arbitrary distant boundary!

Simulated Ising fields



The auto-models

Let $Q(x)=\log(P(x)/P(0))$. Besag's auto-models are defined by

$$Q(\underset{\sim}{z}) = \sum_{i=1}^{n} z_i G_i(z_i) + \sum_{i=1}^{n} \sum_{j \sim i} \beta_{ij} z_i z_j$$

When $z_i \in \{0,1\}$ and $G_i(z_i)=\alpha_i$ we get the *autologistic* model

When $G_i(z_i) = \alpha_i z_i - \log(z_i!)$ and $\beta_{ij} \leq 0$ we get the *auto-Poisson* model

Pseudolikelihood

Another approximate approach is to write down a function of the data which is the product of the $p_i(x_{\delta_i})$, I.e., acting as if the neighborhoods of each point were independent.

This as an estimating equation, but not an optimal one. In fact, in cases of high dependence it tends to be biased.

Recall the Gaussian formula

If $$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$$

then $$(Y \mid X) \sim N(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X),$$
$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY})$$

Let $Q = \Sigma^{-1}$ be the *precision matrix*. Then the conditional precision matrix is

$$\left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} = Q_{YY}$$

Gaussian MRFs

We want a setup in which $Z_i \perp Z_j \,\middle|\, \underset{\sim}{Z}^{-i,j}$ whenever i and j are not neighbors.

Using the Gaussian formula we see that the condition is met iff $Q_{ij} = 0$.

Typically the precision matrix of a GMRF is sparse where the covariance is not. This allows fast computation of likelihoods, simulation etc.

An AR(1) process

Let $X_t \mid X_{t-1} = \phi X_{t-1} + \varepsilon_t$. The lag k autocorrelation is $\phi^{|k|}$. The precision matrix has $Q_{ij} = \phi$ if $|i-j|=1$, $Q_{11}=Q_{nn}=1$ and $Q_{ii}=1+\phi^2$ elsewhere, all other 0.

Thus Σ has N^2 non-zero elements, while Q has N+2(N-1)=3N-2 non-zero elements.

Using the Gaussian formula we see that

$$\mu_{i|-i} = \frac{\phi}{1+\phi^2}(x_{i-1} + x_{i+1}) \quad Q_{i|-i} = 1+\phi^2$$

Conditional autoregression

Suppose that

$$Z_i \,\big|\, \underset{\sim}{Z}^{-i} \sim N(\mu_i + \sum_{i \neq j} \beta_{ij}(x_j - \mu_j), \kappa_i^{-1})$$

This is called a Gaussian conditional autoregressive (CAR) model. WLOG μ_i=0. If also $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}$ these conditional distributions correspond to a multivariate joint Gaussian distribution, mean 0 and precision Q with $Q_{ii}=\kappa_i$ and $Q_{ij}= -\kappa_i\beta_{ij}$, provided Q is positive definite. If the β_{ij} are nonzero only when i~j we have a GMRF.

Likelihood calculation

The Cholesky decomposition of a pd square matrix A is a lower triangular matrix L such that $A=LL^T$.

To solve $Ay = b$ first solve $Lv = b$ (forward substitution), then $L^T y = v$ (backward substitution).

If a precision matrix $Q = LL^T$, log det(Q) = $2\sum \log\left(L_{i,i}\right)$. The quadratic form in the likelihood is $w^T u$ where $u=Qw$ and $w=(z-\mu)$. Note that

$$u_i = Q_{i,i} w_i + \sum_{j:j\sim i} Q_{i,j} w_j$$

Simulation

Let $x \sim N(0,I)$, solve $L^T v = x$ and set $z = \mu + v$.

Then $E(z) = \mu$ and $Var(z) = (L^T)^{-1} I L^{-1} = (LL^T)^{-1} = Q^{-1}$.

Spatial covariance

Whittle (1963) noted that the solution to the stochastic differential equation

$$\left(\Delta - \frac{1}{\phi^2}\right)^{(\kappa+1)/2} Z(s) = W(s)$$

has covariance function

$$\mathrm{Cov}(Z(s), Z(s+h)) \propto \left(\frac{\|h\|}{\phi}\right)^{\kappa} \mathcal{K}_{\kappa}\left(\frac{\|h\|}{\phi}\right)$$

Rue and Tjelmeland (2003) show that one can approximate a Gaussian random field on a grid by a GMRF.

Solution

Write $Z(s) = \sum_k \psi_k(s) w_k$

where $\psi_k(s)$ are piecewise linear on a (possibly nonregular) grid and w_k are appropriately chosen normal random variables.

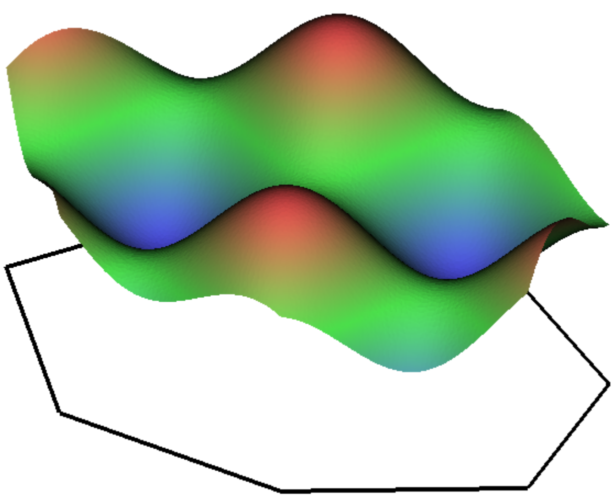
Let $A_{i\cdot} = (\psi_1(s_i), \ldots, \psi_N(s_i))$ and $A=(A_{i\cdot})$.

If Y is Z observed with spatial error η

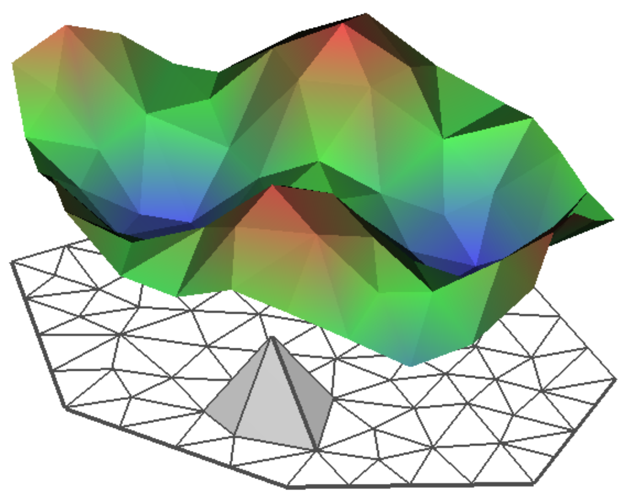
$$Y \mid w \sim N(Aw, Q_\eta^{-1}), \quad w \sim N(\mu, Q^{-1})$$

Basis functions

$x(\mathbf{u}) = \cos(u_1) + \sin(u_2)$

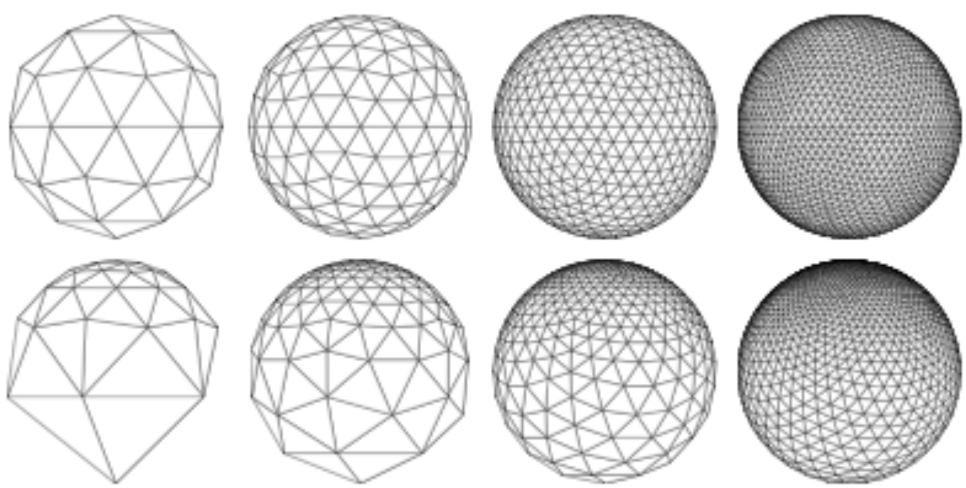


$x(\mathbf{u}) = \sum_k \psi_k(\mathbf{u})\, x_k$

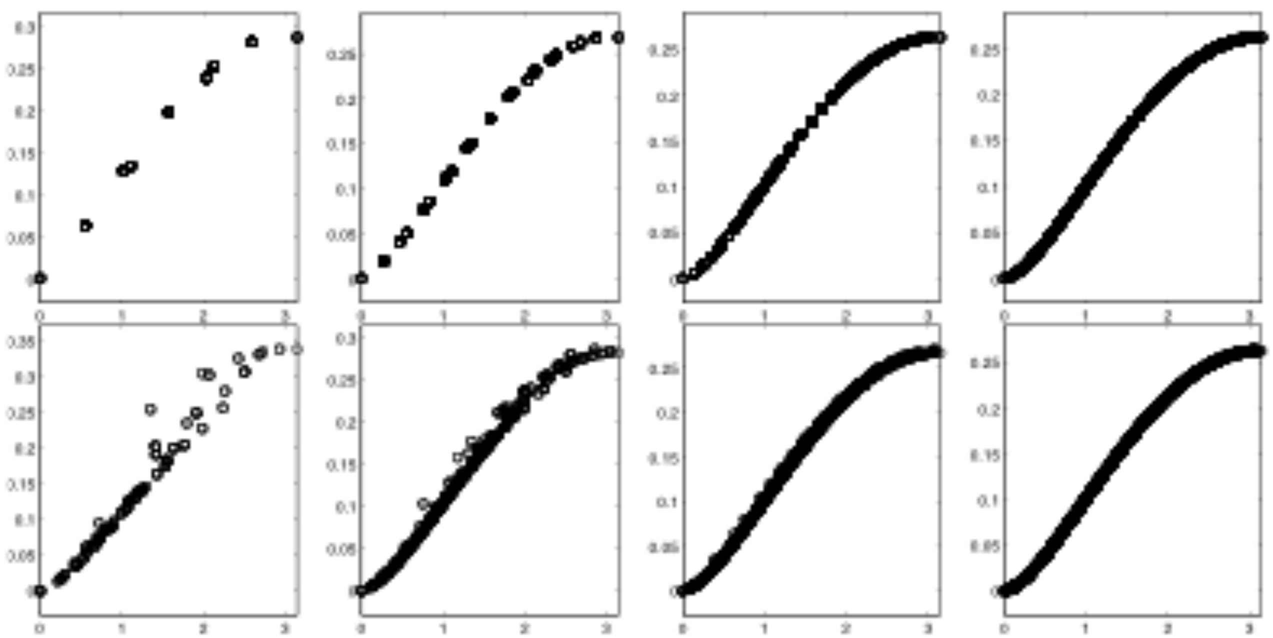


Unequal spacing

Lindgren and Rue show how one can use finite element methods to approximate the solution to the sde (even on a manifold like a sphere) on a triangulization on a set of possibly unequally spaced points.



Covariance approximation



Hierarchic model

Data model: $p(y|z;\theta)$

Latent model: $p(z|\theta)$

$$Z = Aw + \beta B \quad w \sim N(0, Q^{-1})$$

If Bayesian, parameter model: $p(\theta)$

For INLA, need

$$p(y|z,\theta) = \prod_i p(y_i|z_i,\theta)$$

INLA

Laplace's approximation:

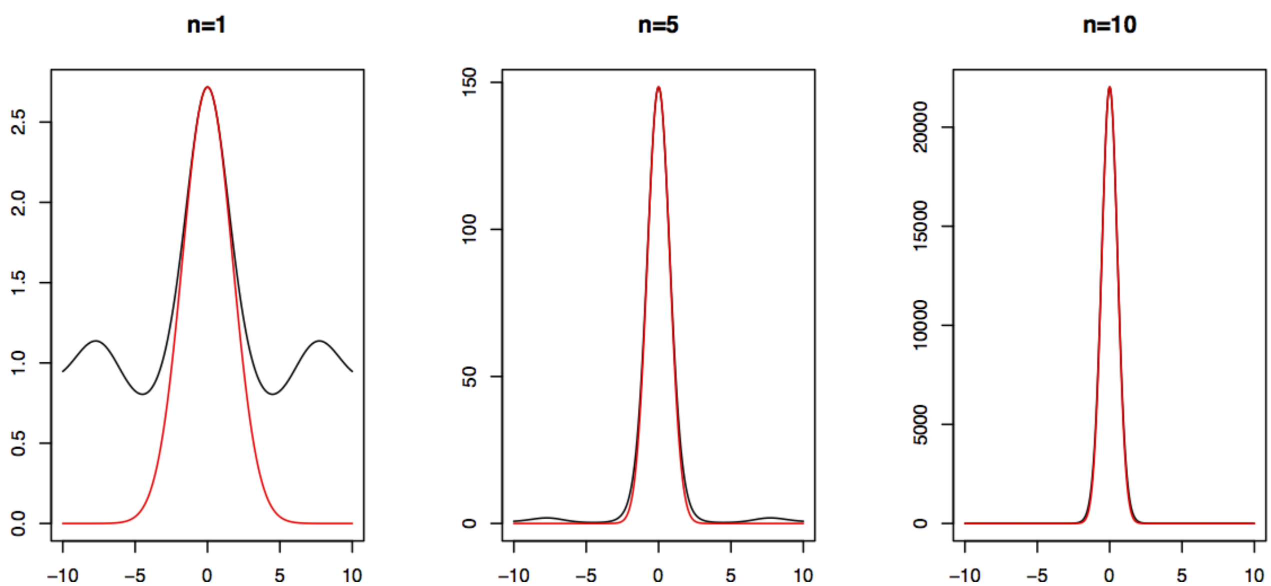
$x^*=\text{argmax}(f(x))$. Taylor expansion around x^*: $f(x) \approx f(x^*)+(x-x^*)^2 f''(x^*)/2$

$$e^{Nf(x)} \approx e^{Nf(x^*)} e^{-N\left|f''(x^*)\right|(x-x^*)^2/2}$$

$$= \sqrt{\frac{2\pi}{N\left|f''(x^*)\right|}} e^{Nf(x^*)} \phi\left(\sqrt{N\left|f''(x^*)\right|}(x-x^*)\right)$$

Interested in predictive distribution p(z|y) and posterior density p(θ|y)

f(x)=sin(x)/x



Posterior manipulation

$$p(y|x,\theta)p(x|\theta) = p(y,x|\theta) \equiv p(x|y,\theta)\,p(y|\theta)$$

Thus

$$p(\theta|y) \propto p(y|\theta)p(\theta) = \frac{p(y|x,\theta)p(x|\theta)}{p(x|y,\theta)}p(\theta)$$

Using the Laplace approximation on $f(x)=\log(\, p(x|y,\theta)/N)$ we get a Gaussian approximation

$$x \mid y,\theta \approx N(\mu^*,Q^*)$$

where μ^*,Q^* depend on Q, β and D^2f.

What INLA computes

Joint posterior of parameters (Laplace approximation)

Marginal posterior of latent variable (integral Laplace approximation or numerical integration)

Not computing the joint predictive distribution

Computational comparison

n=20×2500 obs, m=20×15000 kriging locs

Estimation O(n^3), storage O(n^2)≈20GB

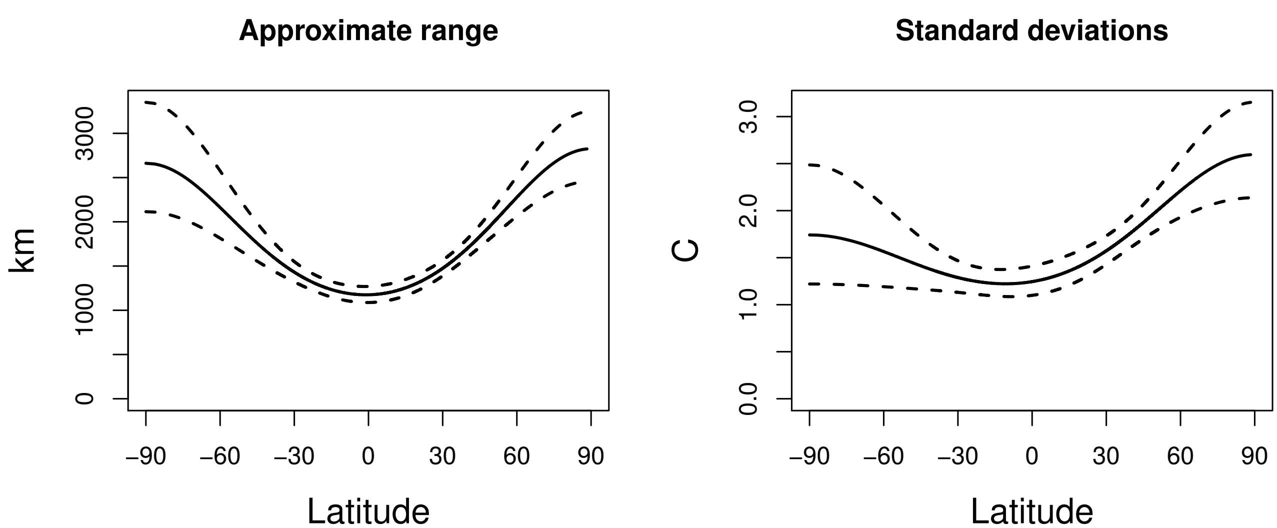
Kriging O($mn+n^3$), storage O($mn+n^2$)≈130GB

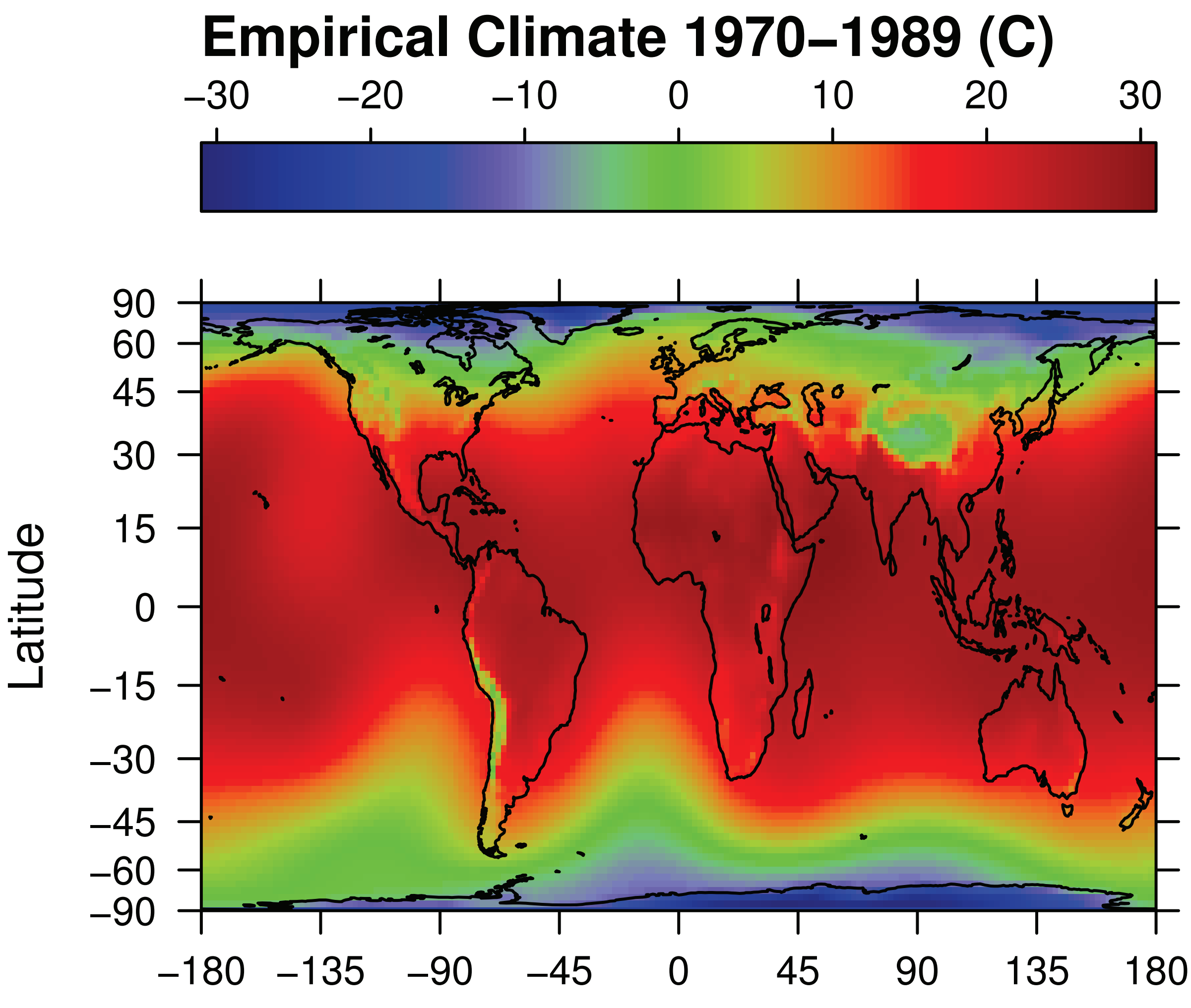
INLA Estimation + kriging O($m^{3/2}$), storage O(m+n)≈50MB

Global temperature analysis

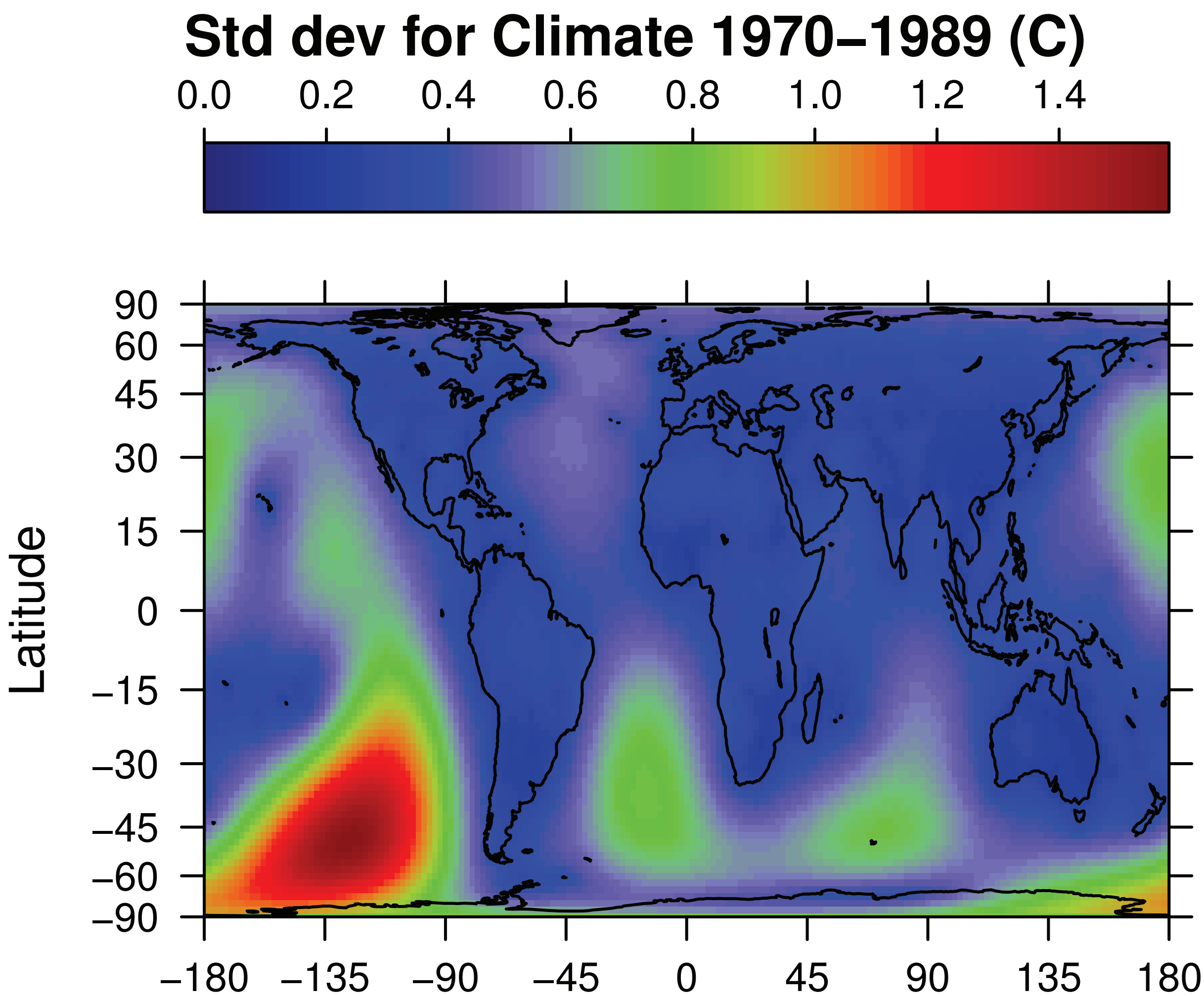
obs = climate + anomaly + elevation + error

Climate covariance parameters:

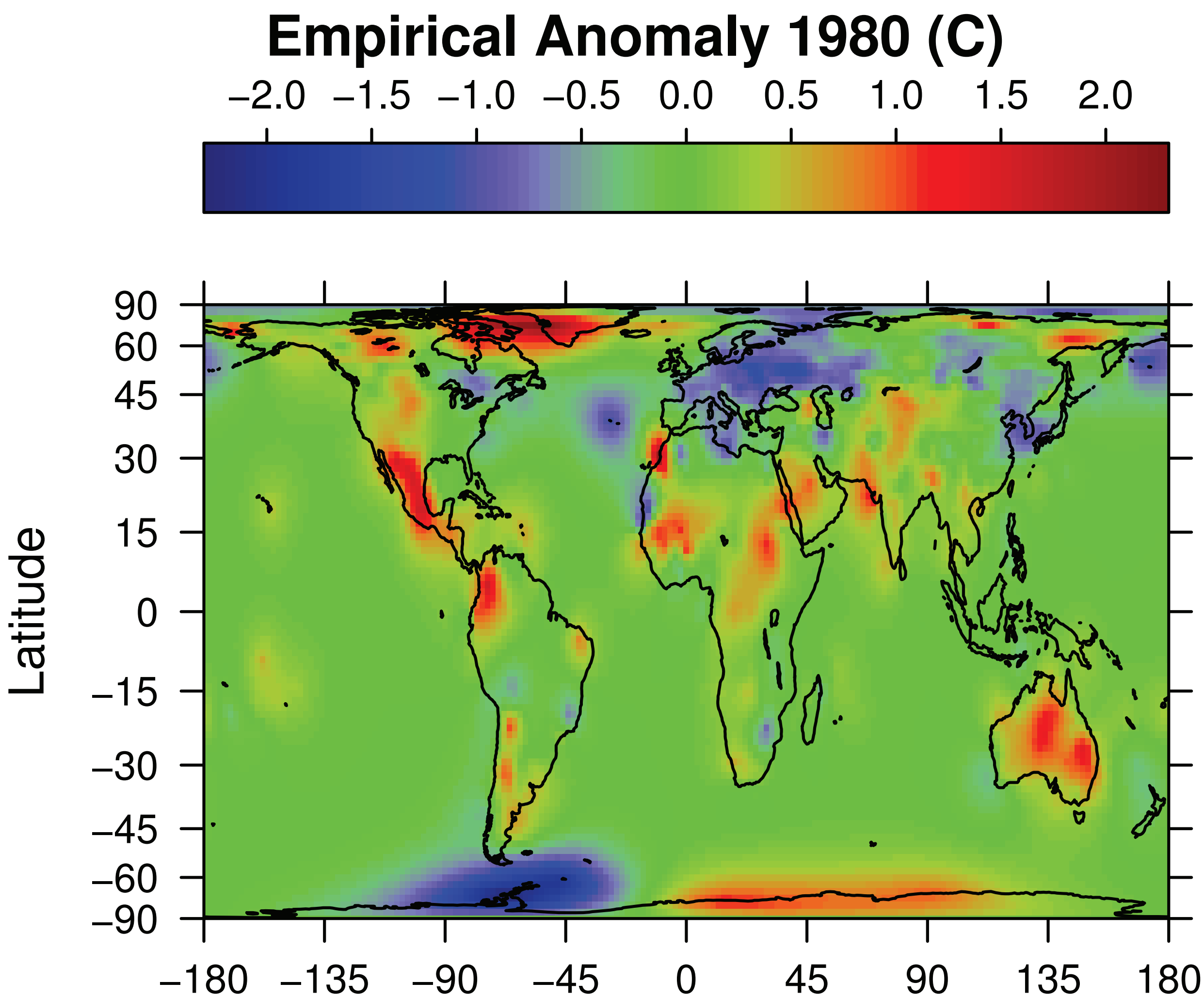


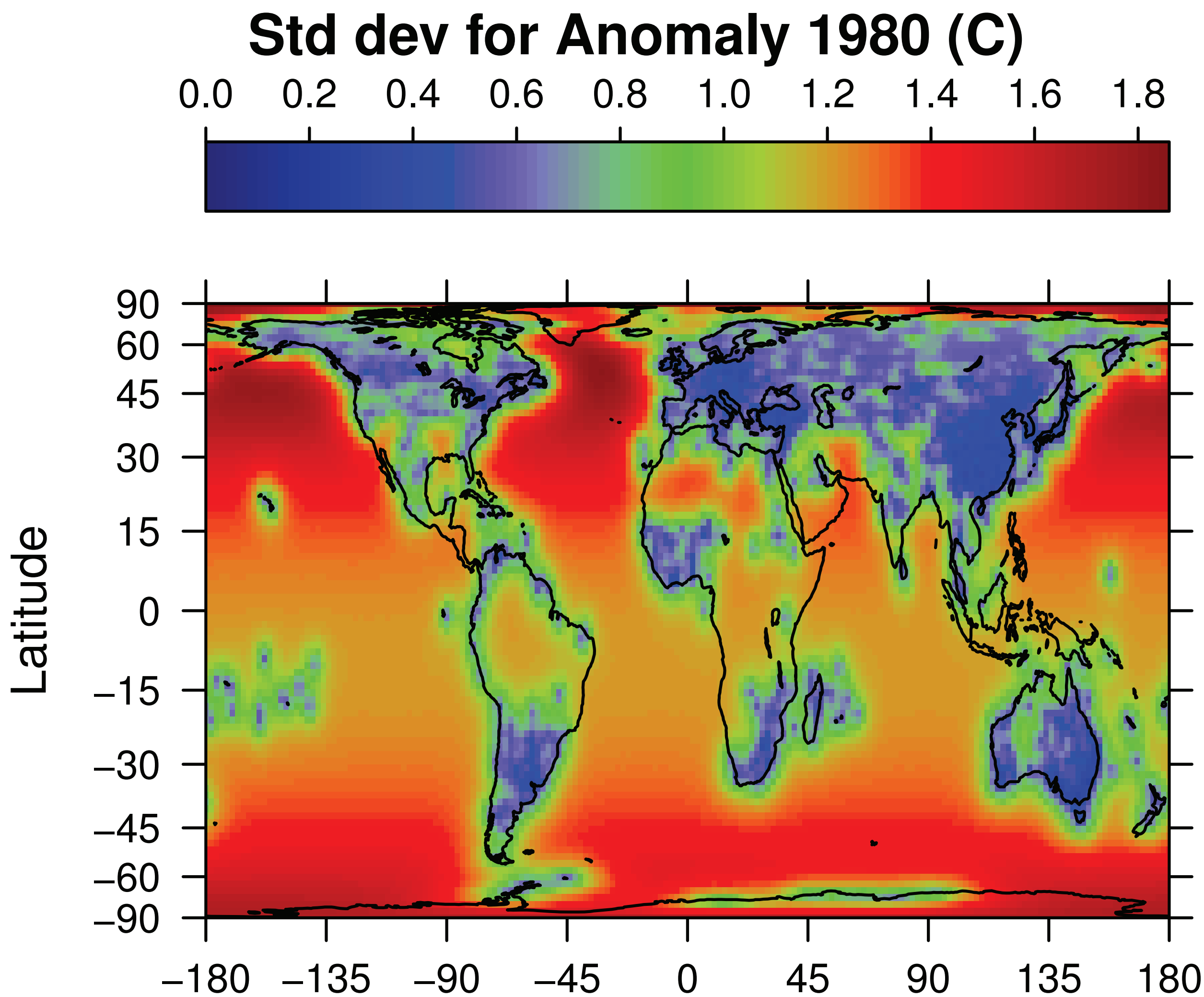


Empirical Climate 1970–1989 (C)
–30
–20
–10
0
10
20
30
Latitude
90
60
45
30
15
0
–15
–30
–45
–60
–90
–180
–135
–90
–45
0
45
90
135
180



Std dev for Climate 1970−1989 (C)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Latitude
90
60
45
30
15
0
−15
−30
−45
−60
−90
−180
−135
−90
−45
0
45
90
135
180


Empirical Anomaly 1980 (C)
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0
Latitude
90 60 45 30 15 0 −15 −30 −45 −60 −90
−180 −135 −90 −45 0 45 90 135 180


Std dev for Anomaly 1980 (C)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Latitude
90
60
45
30
15
0
–15
–30
–45
–60
–90
–180
–135
–90
–45
0
45
90
135
180