Bayesian Models for Spatial Extremes

Application to inferring high values of ground-level ozone spatially

Eric Gilleland[1,3]

Dan Cooley[1,2], Douglas Nychka[1], and Uli Schneider[1]

National Center for Atmospheric Research (NCAR)
[1]NCAR Geophysical Statistics Project (GSP)
[2]Department of Applied Mathematics, University of Colorado at Boulder
[3]NCAR Research Applications Laboratory (RAL)

This work funded by the National Science Foundation (NSF)

Outline

- Background to Ozone Application
- Background to Extreme Value Analysis
  - Two approaches:
    1. Model all of the data, and look at extremes.
    2. Only model the extremes of the data.
- Previous Work:
  1. Space-time model
  2. Likelihood-based Extremes Model (w/ spatially coherent parameters)
- Bayesian Model for Spatial Extremes
- Summary and ongoing work

Background to Ozone Application: Air Quality Standards

As required by the Clean Air Act (CAA) of 1971, the EPA has established standards, known as the *National Ambient Air Quality Standards (NAAQS)*, to monitor and control ambient concentrations for six principal air pollutants (also referred to as criteria pollutants):

- carbon monoxide (CO),
- lead (Pb),
- nitrogen dioxide (NO_2),
- *ground-level Ozone (O_3)*,
- particulate matter (PM) and
- sulfur dioxide (SO_2)

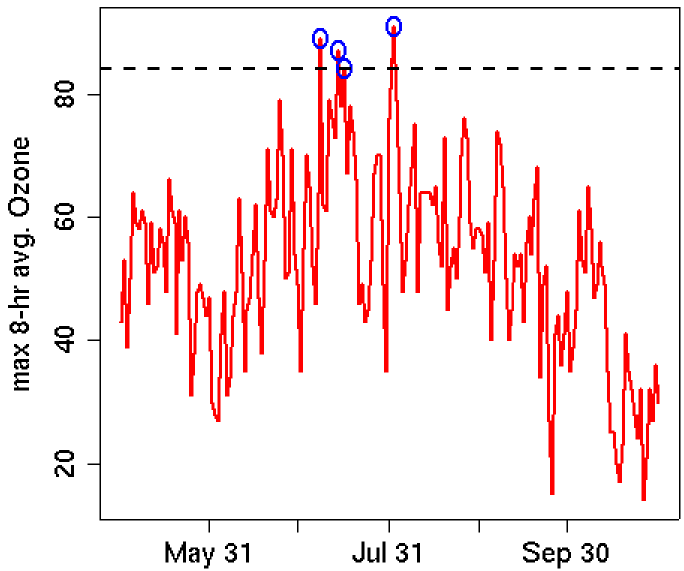
Background to Ozone Application: New NAAQS for ozone

If the three-year average of the

4^{th}-highest daily

maximum 8-hour average ozone (FHDA)

exceeds 84ppb, then attainment is not met.



Background to Ozone Application: New NAAQS for ozone

May make sense from a health/environment perspective.

Background to Ozone Application: New NAAQS for ozone

May make sense from a health/environment perspective.

Presents a challenging statistical problem.

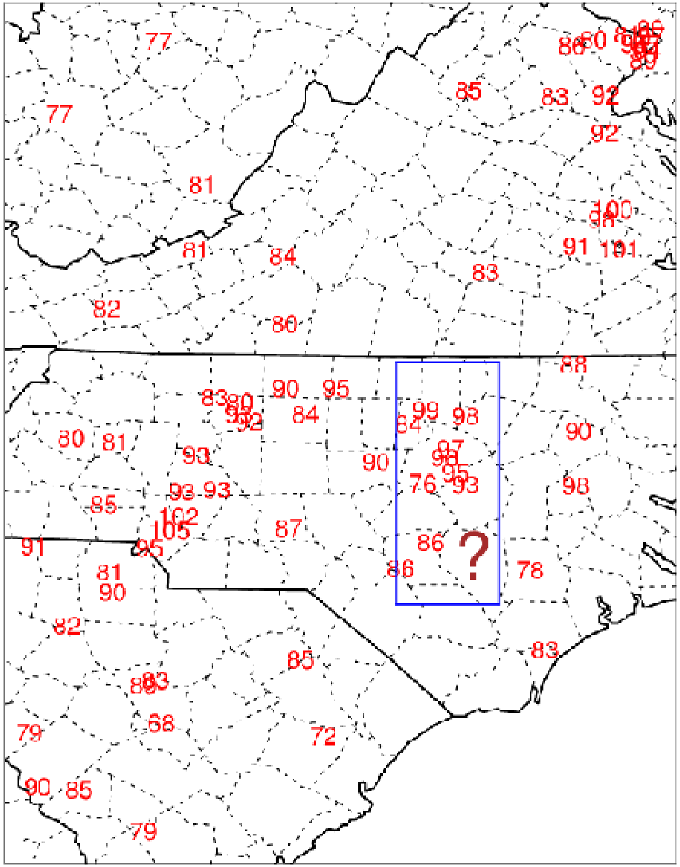
Background to Ozone Application: New NAAQS for ozone

Attainment/Non-attainment regions

Regions with at least one ozone monitor are out of attainment when the FHDA exceeds 84ppb (at any one monitoring station).

Regions without any monitoring can still be declared out of attainment, but there is no way for the EPA to determine high ozone exposure in such areas.

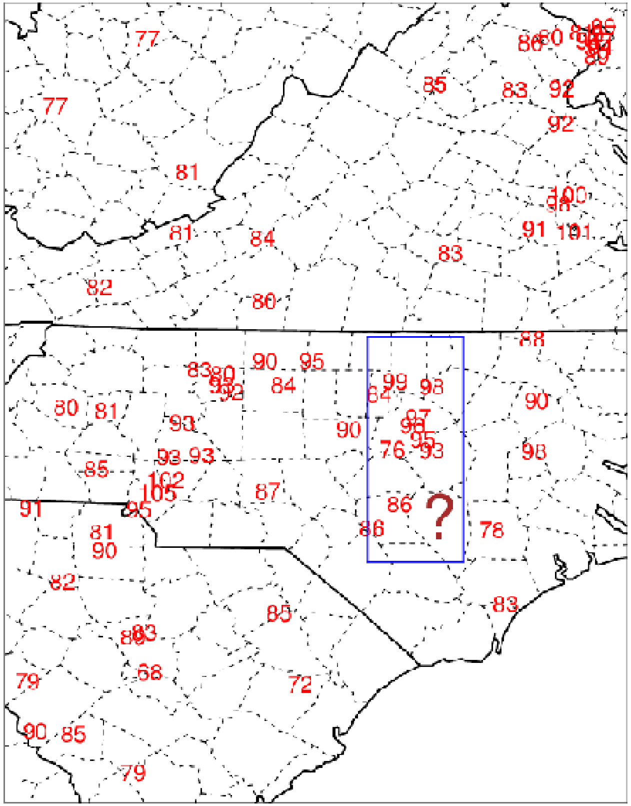
Background to Ozone Application: New NAAQS for ozone



Goal: To infer extreme ozone spatially.

Background to Ozone Application: Data

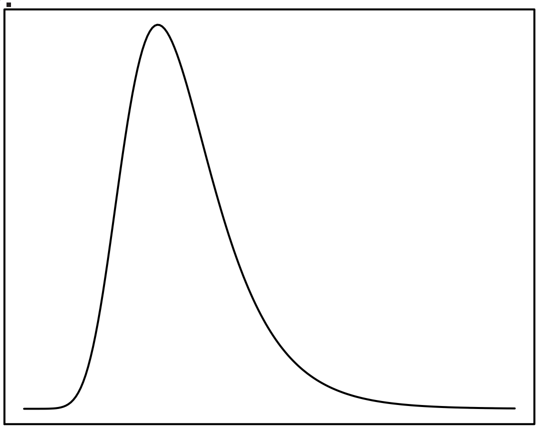
Here, we focus on a small homogeneous region over North Carolina (and surrounding areas) with 5 years (seasons) of data, 72 locations.



Background to Extreme Value Analysis: "Ordinary" vs Extreme Value Statistics

"Ordinary" Statistics: Tries to describe main part of distribution; may ignore outliers.

Extremes: Tries to characterize the tail of the distribution; keeps only the extreme observations.



Extreme Value Theory:

Asymptotic justification for univariate distributions.

- GEV – models block maxima
- GPD – models threshold exceedances

Generalized Pareto Distribution [Pickands, 1975]

$$P\{Z > y + u | Z > u\} \approx \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}$$

σ is scale parameter

ξ is shape parameter (controls tail behavior)

Generalized Pareto Distribution [Pickands, 1975]

$$P\{Z > y + u | Z > u\} \approx \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}$$

σ is scale parameter

ξ is shape parameter (controls tail behavior)

But, how do we incorporate spatial structure here?

Three Previously Studied Approaches

Space-Time Model

- Determine an AR model for **every** location, even the unobserved ones. [more]
- Using spatially-coherent shocks, simulate every day of an Ozone season.
- Build up the distribution of the FHDA.

Naive Model

- Straightforward application of kriging to FHDA field. [more]

Likelihood-based Spatial Extremes Model

- Uses a Generalized Pareto with spatially cohesive scale parameters. [more]

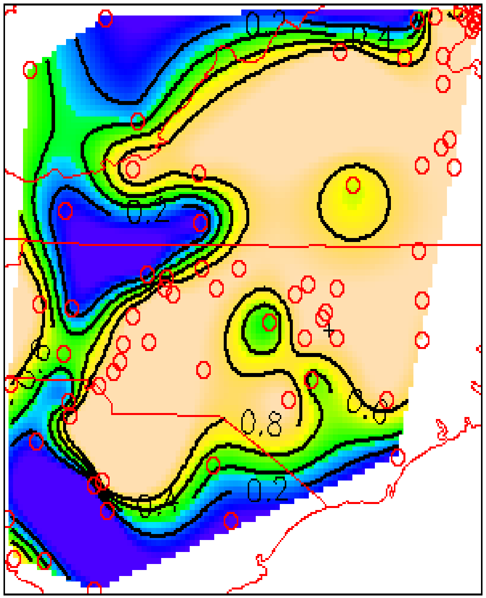
Comparing the Space-Time and naive models

		TPS	Space-Time	Naive	
				Variogram	Correlation
MPSE	1995	2.23	2.67	5.68	5.27
	1996	2.49	2.85	5.96	5.90
	1997	2.91	3.01	6.41	6.02
	1998	2.75	2.93	5.35	4.85
	1999	4.34	2.94	6.76	6.22
CV RMSE	1995	5.34	4.73	5.19	5.33
	1996	5.61	4.84	5.51	5.68
	1997	6.27	4.59	6.03	6.05
	1998	5.00	3.25	4.98	4.93
	1999	6.25	4.91	6.47	6.30

Probability of exceeding the standard

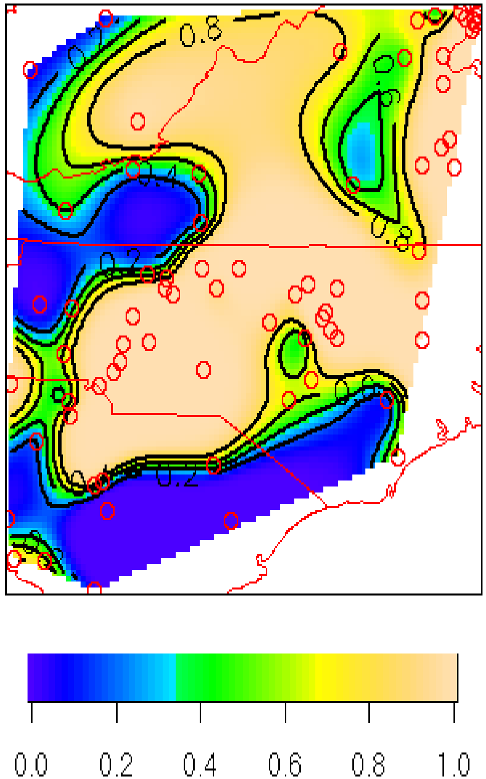
Space-Time Model

(a)



Extreme-Value Model

(b)



Summary of these models

- Simplicity of the naive model approach is desirable.
- Space-Time model yields consistently lower MSE from cross-validation.
- Space-Time model can account for “complicated” spatial features without resorting to non-standard techniques.
- Space-Time MPSE is consistently too optimistic.
- Extreme value models good alternative to modelling the tail of distributions.
- Two very different approaches yield similar results

Bayesian Hierarchical Model - Base Level

Let $Z_j(x_i)$ be the ozone concentration recorded at the station located at x_i on day j.

Model Foundation:

We assume that ozone concentrations $Z_j(x_i)$ that exceed a threshold u are GPD, whose parameters depend on the station's location.

$$P\{Z_j(x_i) - u > y | Z_j(x_i) > u\} = \left(1 + \frac{\xi_i y}{\tilde{\sigma}_i}\right)^{-1/\xi_i}$$

Bayesian Hierarchical Model - Level 2

$$Z_j(x_i) \sim GPD(\tilde{\sigma}_i, \xi_i)$$

Since $\tilde{\sigma}_i > 0$, we choose to model $\phi_i = \log \tilde{\sigma}_i$.

Bayesian Hierarchical Model - Level 2

$$Z_j(x_i) \sim GPD(\exp(\phi_i), \xi_i)$$

Structure of ϕ:

We model ϕ with standard geophysical methods, choosing a multivariate normal prior distribution.

$$\phi_i = f(\vec{\alpha}, \text{covariates}_i) + \epsilon,$$

where $\epsilon \sim MVN(0, \Sigma)$, and $\Sigma_{i,i'} = g(\vec{\beta}, ||x_i - x_i'||_2)$.

Bayesian Hierarchical Model - Level 2

$$Z_j(x_i) \sim GPD(\exp(\phi_i), \xi_i)$$

Structure of ϕ:

We model ϕ with standard geophysical methods, choosing a multivariate normal prior distribution.

$$\phi_i = f(\vec{\alpha}, \text{covariates}_i) + \epsilon,$$

where $\epsilon \sim MVN(0, \Sigma)$, and $\Sigma_{i,i'} = g(\vec{\beta}, ||x_i - x_i'||_2)$.

We let:

$$f(\vec{\alpha}, \text{covariates}) = \alpha_0 + \alpha_1(\text{covariate 1}) + \ldots \alpha_n(\text{covariate n})$$

$$g(\vec{\beta}, ||x_i - x_i'||_2) = \beta_0 * exp(-\beta_1 * ||x_i - x_i'||_2)$$

We can model ξ similarly.

Bayesian Hierarchical Model - Level 3

$$X_j(x_i) \sim GPD(\exp(\phi_i), \xi_i)$$

$$\phi_i = f(\vec{\alpha}, \text{covariates}) + \epsilon,$$

$$\epsilon \sim MVN(0, \Sigma), \text{ where } \Sigma_{i,i'} = \beta_0 * exp(-\beta_1 * ||x_i - x_i'||_2)$$

Priors for α: Non-informative $\alpha_i \propto Unif(-\infty, \infty)$

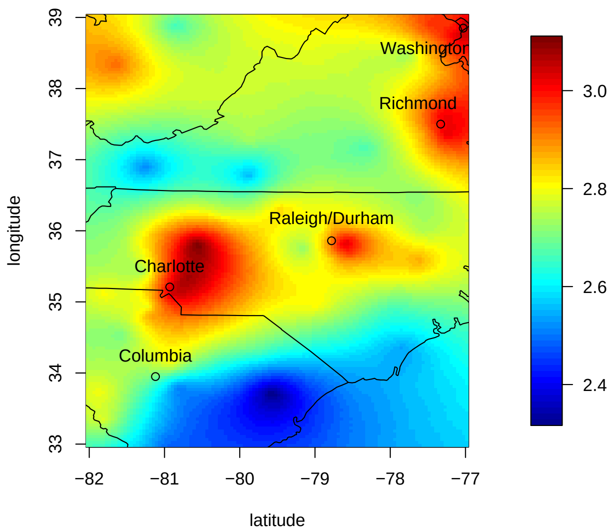
Priors for β: Informative $\beta_0 \propto Unif(0.01, 0.10)$ $\beta_1 \propto Unif(\frac{3}{250}, \frac{3}{50})$

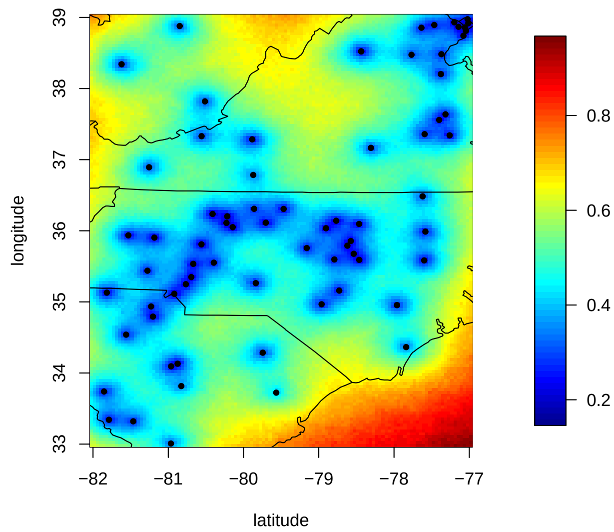
Threshold Selection

- Low enough to have enough data (i.e., low variance).
- High enough for the GPD to be appropriate (i.e., low bias).

A threshold of 65 ppb was found to be reasonable for these data

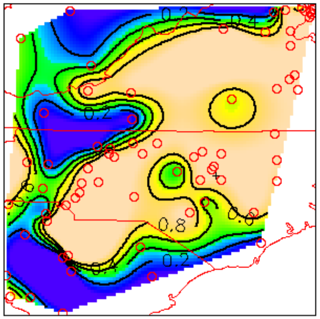
Initial Results



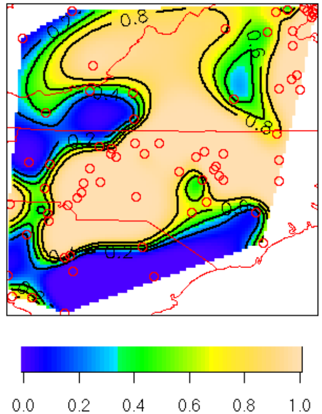


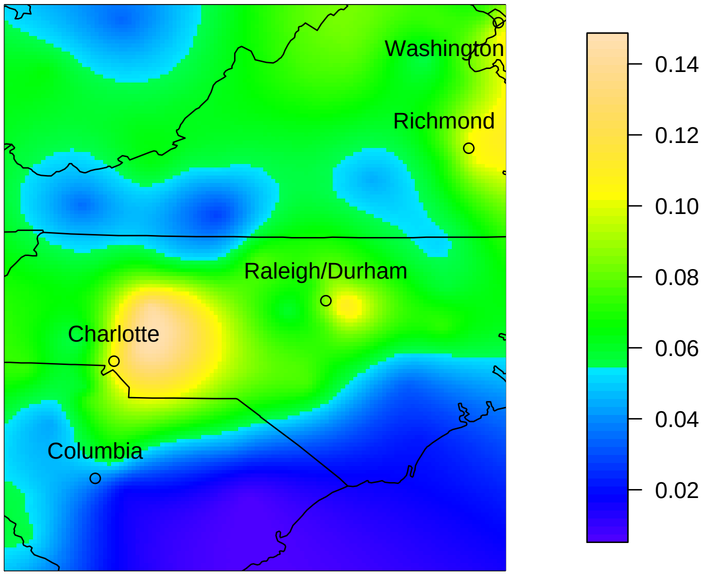
Initial Results

(a)



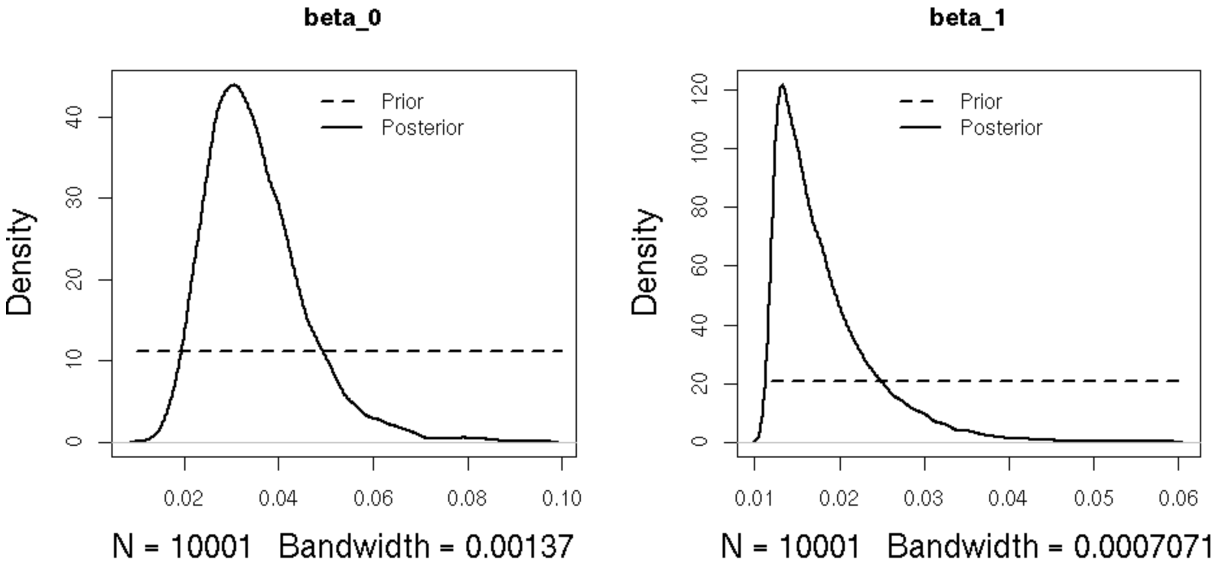
(b)





Initial Results

Posterior for β_0 and β_1



Ongoing and Future Work

- Have a flexible model that accounts for the tail behavior for ozone.

****

- Run more models/do model comparison
- Add covariates (e.g., population, elevation, ...).
- Test prior sensitivity
- Bayesian model checks
- Obtain probabilities of exceeding the standard in a more sophisticated way.

Stop!

Space-Time Approach: Space-Time Model

Let $Y(\mathbf{x}, t)$ denote the daily 8-hr max Ozone for m sites over n time points. Consider,

$$Y(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \sigma(\mathbf{x})u(\mathbf{x}, t),$$

where $u(\mathbf{x}, t)$ is a de-seasonalized zero mean, unit variance space-time process, $i.e.$

$$u(\mathbf{x}, t) = \rho(\mathbf{x})u(\mathbf{x}, t-1) + \varepsilon(\mathbf{x}, t),$$

where $|\rho(\mathbf{x})| < 1$, the spatial shocks, $\varepsilon(\mathbf{x}, t)$, are independent over time, but spatially correlated with covariance function

$$\text{Cov}(\varepsilon(\mathbf{x}, t), \varepsilon(\mathbf{x}', t)) = \sqrt{1-\rho^2(\mathbf{x})}\sqrt{1-\rho^2(\mathbf{x}')}\psi(d(\mathbf{x}, \mathbf{x}'))$$

Space-Time Approach: Space-Time Model

Algorithm to predict FHDA at unobserved location, $\mathbf{x}_0$.

1. Simulate data for an entire Ozone season

   (a) Interpolate spatially from $u(\mathbf{x}, 1)$ to get $\hat{u}(\mathbf{x}_0, 1)$.

   (b) Also interpolate spatially to get $\hat{\rho}(\mathbf{x}_0)$, $\hat{\mu}(\mathbf{x}_0, \cdot)$ and $\hat{\sigma}(\mathbf{x}_0)$.

   (c) Sample shocks at time t from $[\varepsilon(\mathbf{x}_0, t)|\varepsilon(\mathbf{x}, t)]$.

   (d) Propagate AR(1) model.

   (e) Back transform $\hat{Y}(\mathbf{x}_0, t) = \hat{u}(\mathbf{x}_0, t)\hat{\sigma}(\mathbf{x}_0) + \hat{\mu}(\mathbf{x}_0, t)$

2. Take fourth-highest value from Step 1.

3. Repeat Steps 1 and 2 many times to get a sample of FHDA at unobserved location.

[back]

Space-Time Approach: Space-Time Model

Distribution for the AR(1) shocks

$[\varepsilon(\mathbf{x}_0, t)|\varepsilon(\mathbf{x}, t)]$ (Step 1c) given by

$$\text{Gau}(\mathbf{M}, \boldsymbol{\Sigma})$$

with

$$\mathbf{M} = \mathbf{k}'(\mathbf{x}_0, \mathbf{x})\mathbf{k}^{-1}(\mathbf{x}, \mathbf{x})\varepsilon(\mathbf{x}, t)$$

and

$$\boldsymbol{\Sigma} = \mathbf{k}'(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{k}'(\mathbf{x}_0, \mathbf{x})\mathbf{k}^{-1}(\mathbf{x}, \mathbf{x})\mathbf{k}(\mathbf{x}, \mathbf{x}_0),$$

where $k(\mathbf{x}, \mathbf{y})$ represents the covariance between two spatial locations.

[back]

Geostatistical Approach: Naive Model

Covariance

Estimate a covariance function for the FHDA field, and use it to predict an unobserved location.

$$\hat{Y}(\mathbf{x}_0) = \mathbf{k}'(\mathbf{x}_0, \mathbf{x})\mathbf{k}^{-1}(\mathbf{x}, \mathbf{x})\mathbf{Y}$$

where $\mathbf{Y}$ is the observed FHDA, $\mathbf{k}(\mathbf{x}, \mathbf{y})$ is the covariance between two locations $\mathbf{x}$ and $\mathbf{y}$. This has variance,

$$k(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{k}'(\mathbf{x}_0, \mathbf{x})\mathbf{k}^{-1}(\mathbf{x}, \mathbf{x})\mathbf{k}(\mathbf{x}, \mathbf{x}_0)$$

Geostatistical Approach: Naive Model

Covariance

Two types of covariance: ψ_v and ψ_m. back

More on $\sigma(\mathbf{x})$

$$\sigma(\mathbf{x}) = P(\mathbf{x}) + e(\mathbf{x}) + \eta(\mathbf{x})$$

with P a linear function of space, e a smooth spatial process, and η white noise (nugget).

- As $\lambda \longrightarrow \infty$, the surface tends toward just the linear function.
- As $\lambda \longrightarrow 0$, the surface will fit the data more closely.

log of joint distribution

$$\sum_{i=1}^{n} \ell_{\mathsf{GPD}}(y(\mathbf{x}_i, t), \sigma(\mathbf{x}_i), \xi) -$$

log of joint distribution

$$\sum_{i=1}^{n} \ell_{\mathsf{GPD}}(y(\mathbf{x}_i, t), \sigma(\mathbf{x}_i), \xi) -$$

$$\lambda(\boldsymbol{\sigma} - \mathbf{X}\boldsymbol{\beta})^T K^{-1}(\boldsymbol{\sigma} - \mathbf{X}\boldsymbol{\beta})/2 - \log(|\lambda K|) + C$$

log of joint distribution

$$\sum_{i=1}^{n} \ell_{\text{GPD}}(y(\mathbf{x}_i, t), \sigma(\mathbf{x}_i), \xi) -$$

$$\lambda(\boldsymbol{\sigma} - \mathbf{X}\boldsymbol{\beta})^T K^{-1}(\boldsymbol{\sigma} - \mathbf{X}\boldsymbol{\beta})/2 - \log(|\lambda K|) + C$$

K is the covariance for the prior on σ at the observations.

This is a penalized likelihood:
The penalty on $\boldsymbol{\sigma}$ results from the covariance and smoothing parameter λ. back