Spatial data: models and analysis

Lecture 2

Douglas Nychka,
National Center for Atmospheric Research

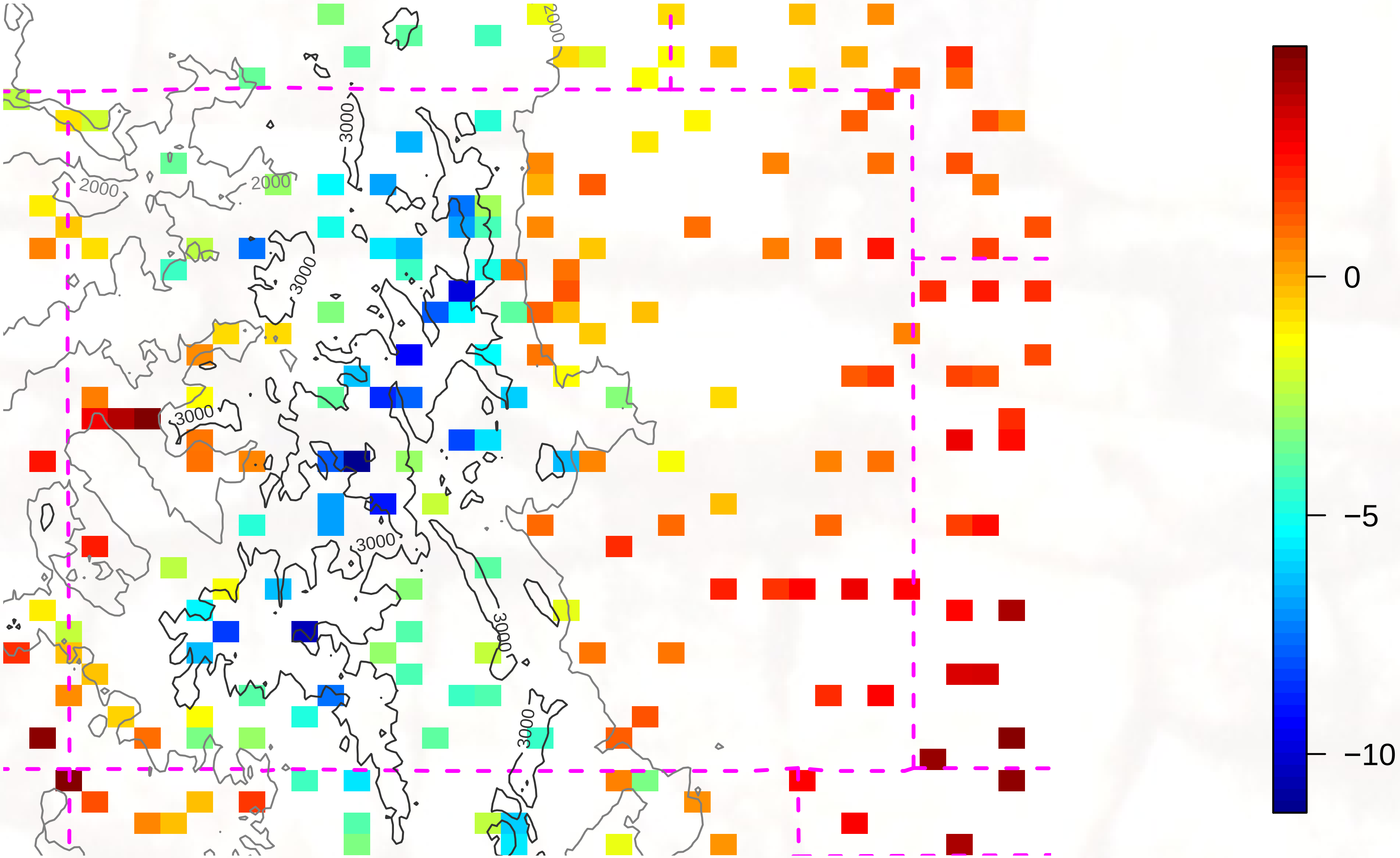
Supported by the National Science Foundation

April, 2017

Outline

- Additive statistical model
- Spatial Processes and covariance functions
- Kriging
- Colorado Climate
- `fields` R package

Colorado MAM average tmin



Goals:

1. Prediction: Determine the climate at locations where there are not stations.

2. Uncertainty: Quantify the error in the predictions.

3. Summarize the spatial/temporal structure: One tool for comparing observations to models and models to models for physical insight

An additive statistical model

Spatial data

$$y_i = \sum_k Z_{i,k} \boldsymbol{d}_k + g(\boldsymbol{x}_i) + \epsilon_i$$

or by vectors

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{d} + \boldsymbol{g} + \boldsymbol{\epsilon}$$

Observation = fixed component + spatial process + error

Main Ideas

MODEL STEP

Use observed data to tease out a statistical model for g

COMPUTING/ESTIMATION STEP

- Estimate the statistical parameters

- Find the distribution of g and $\boldsymbol{d}$ *given* the observations

e.g. This is Kriging, or Bayesian statistics or splines

Useful to view g as being a Gaussian process described by a mean function and a covariance function.

Spatial processes and covariance functions

Covariance functions $g(\boldsymbol{x})$ is a random surface with $E[g(\boldsymbol{x})] = \mu(\boldsymbol{x})$

Covariance function k has two spatial arguments

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = COV(g(\boldsymbol{x}_1), g(\boldsymbol{x}_2))$$

- Usually $\mu(\boldsymbol{x})$ is constant or even zero.
- Often k only depends on distance of separation between locations.

Exponential

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho e^{-\text{ distance}(\boldsymbol{x}_1, \boldsymbol{x}_2)/\theta}$$

Isotropic

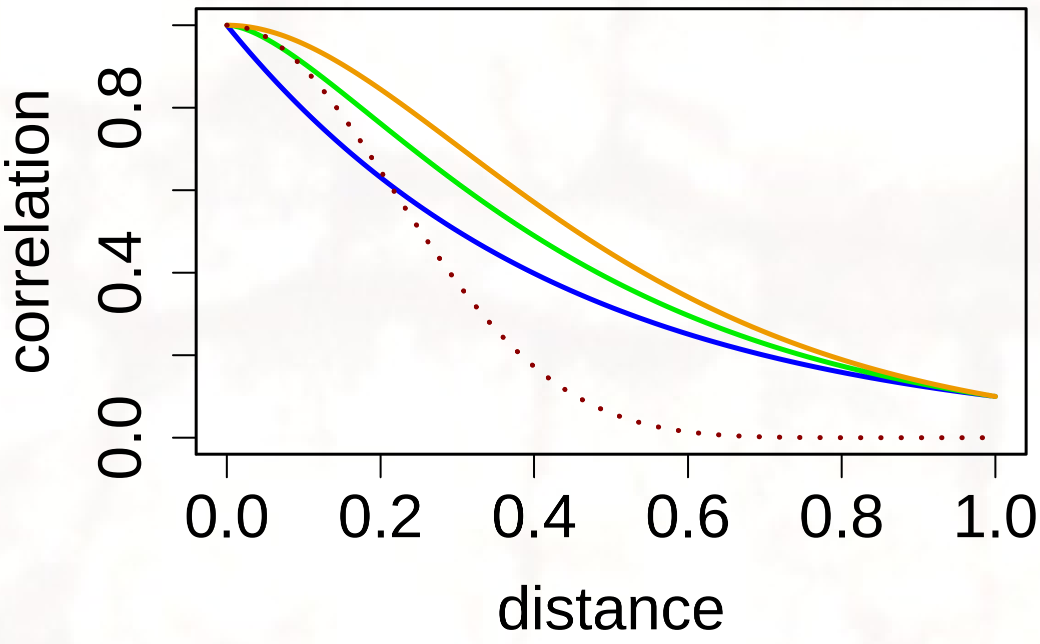
$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho \Phi(\text{distance}(\boldsymbol{x}_1, \boldsymbol{x}_2)/\theta)$$

Families of correlation functions

Matern:

$\phi(d) = \rho\psi_\nu(d/\theta))$ with ψ_ν a Bessel function.

$\nu =$ .5 Exponential, 1.0 , 2.0



- θ a range parameter
- ν smoothness at 0.
- ψ_ν is an exponential for $\nu = 1/2$ as $\nu \to \infty$ Gaussian.
- As ν increases the process is smoother.

Wendland:

Polynomial that is exactly zero outside given range.

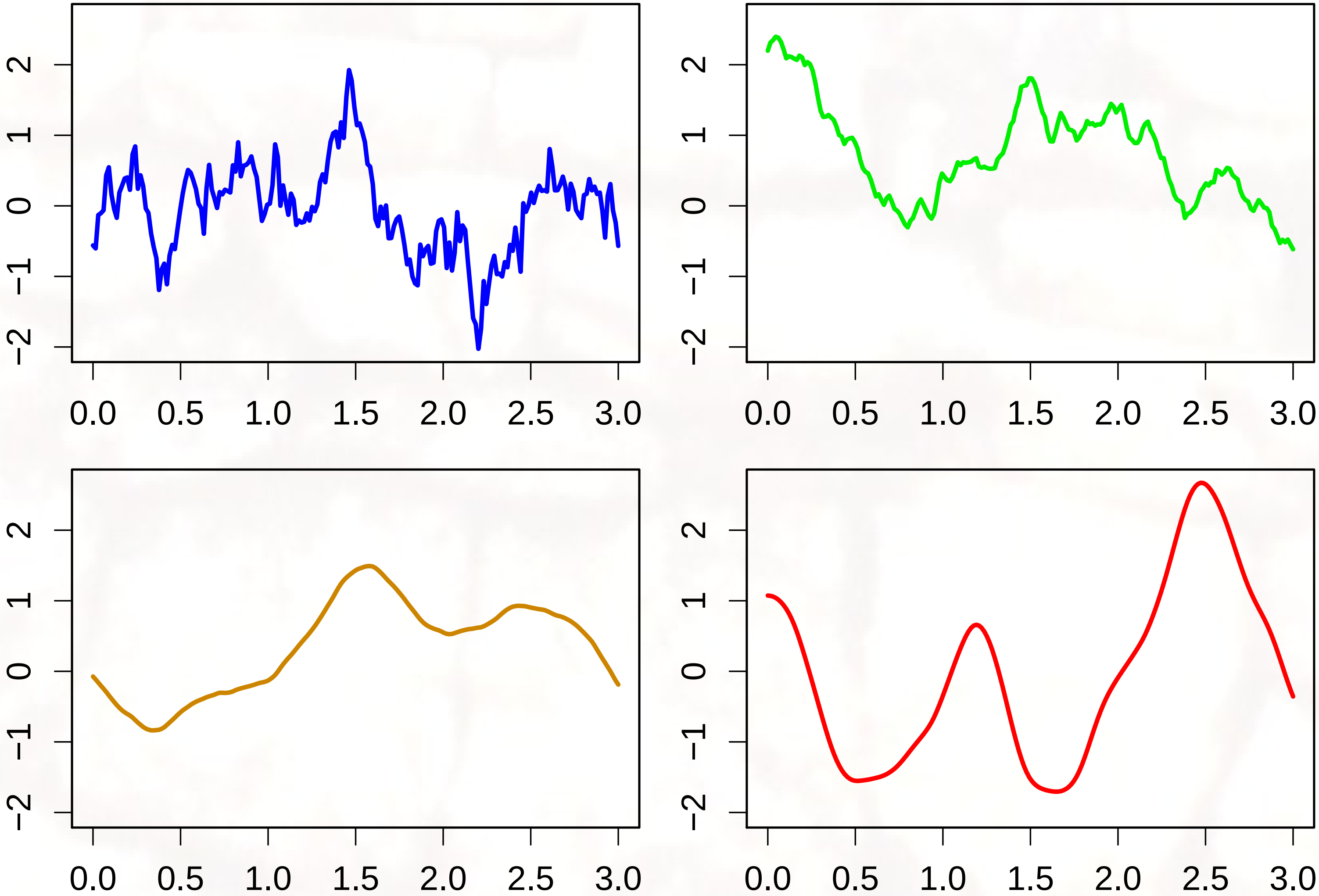
Compactly supported Wendland covariance (d=2, k=3)

What do these processes look like?

Varying the smoothness:

Matern (.5)

Matern(1.0)



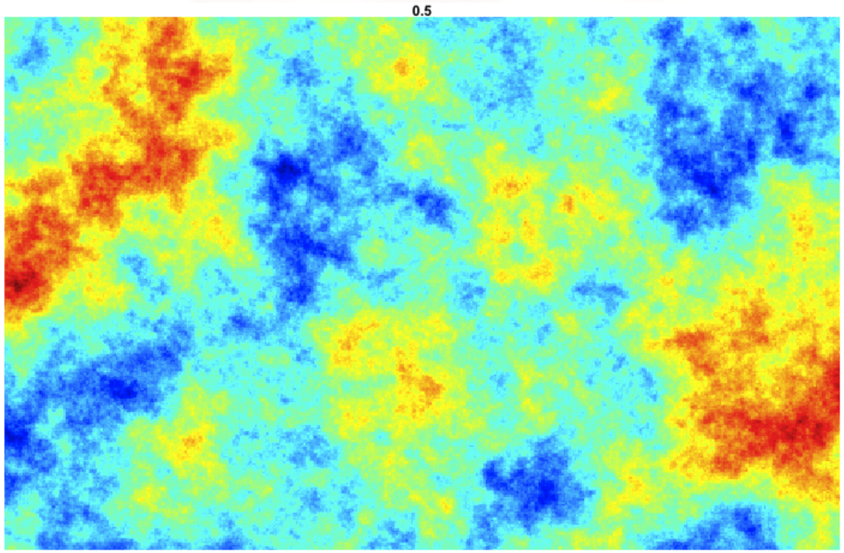
Matern (2.0)

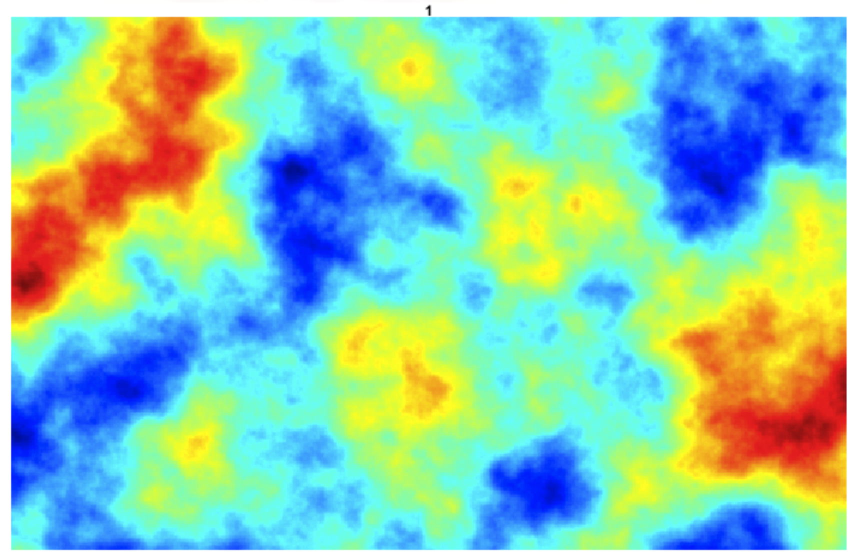
Wendland (2.0)

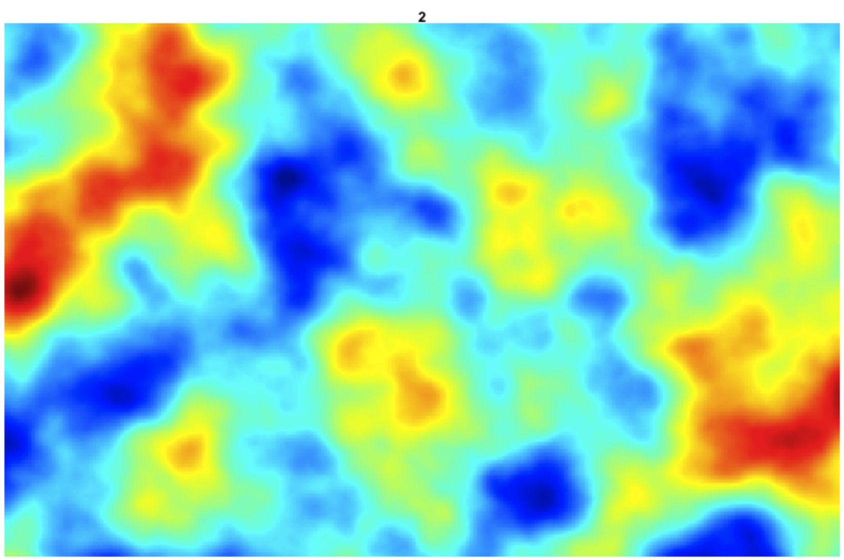
What do these processes look like?

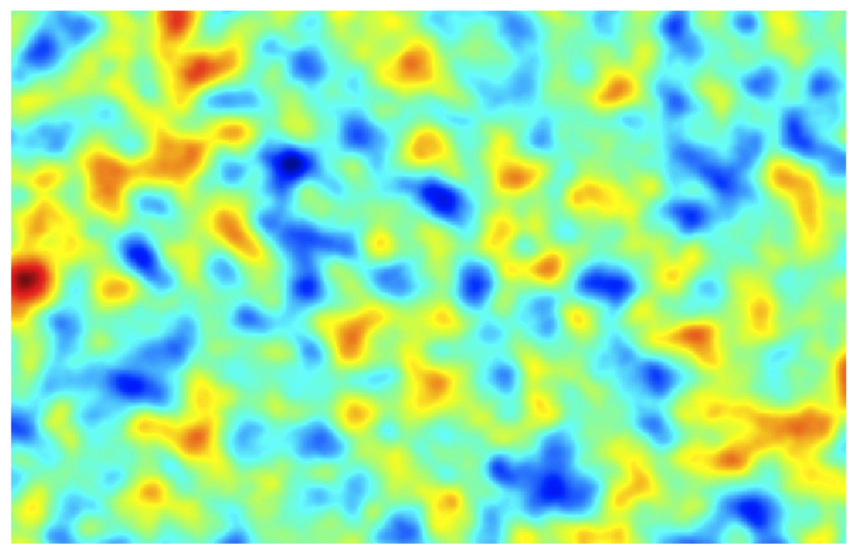
Varying the smoothness:

Matern (.5) Matern(1.0)









Matern (2.0) Wendland (2.0)

Revisiting the additive model

Observations:

$$y_i = \sum_k Z_{i,k} \boldsymbol{d}_k + g(\boldsymbol{x}_i) + \epsilon_i$$

or by vectors

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{d} + \boldsymbol{g} + \boldsymbol{\epsilon}$$

Observation = fixed component + spatial process + error

- ϵ_i's are uncorrelated $N(0, \sigma^2)$

Spatial process:

- $g(x)$ mean zero Gaussian process $VAR(g(\boldsymbol{x})) = \rho$
- Covariance function $\rho k_\theta(.,.)$ with θ some parameters.

A multivariate normal world

Covariance matrix of g at observations:

$K_\theta[i,j] = k_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$
so
$E(g(\boldsymbol{x}_i)g(\boldsymbol{x}_j)) = \rho K_\theta[i,j]$

The observation model can be collapsed as multivariate normal

$$\boldsymbol{y} \sim MN(\boldsymbol{Zd}, \rho K_\theta + \sigma^2 I)$$

$g(\boldsymbol{x})$ given $\boldsymbol{y}$ and all parameters is multivariate Gaussian

Follows the standard form for the conditional distribution of a multivariate Gaussian.

Some Background

Joint covariance matrix of (U,V) is Σ

$$\Sigma = \begin{bmatrix} \Sigma_{u,u} & \Sigma_{u,v} \\ \Sigma_{v,u} & \Sigma_{u,v} \end{bmatrix} \tag{1}$$

Recall if $\boldsymbol{U}$ and $\boldsymbol{V}$ are multivariate normal and $E(U) = E(V) = 0$.

$$[\boldsymbol{V}|\boldsymbol{U}] \sim MN(\quad \Sigma_{v,u}(\Sigma_{u,u}^{-1})\boldsymbol{U}, \quad \Sigma_{v,v} - \Sigma_{v,u}(\Sigma_{u,u}^{-1})\Sigma_{u,v} \quad)$$

Kriging

What is Kriging?

- Assume that the covariance parameters are known or estimated by variogram fitting or maximum likelihood.

- Find the parameters in the linear model by generalized least squares.

- With covariance and linear model parameters fixed, find the conditional distribution for g given $\boldsymbol{y}$

$\hat{g}$ is the conditional expectation of g *given* $\boldsymbol{y}$

e. g. $\hat{g}(x) = E[g(x)|\boldsymbol{y}]$

Prediction standard errors for g based on conditional variances.

Lecture 1 revisited

- $\widehat{g}(\boldsymbol{x})$ is type of spline using a specific roughness penalty

- $\widehat{g}(\boldsymbol{x})$ is a penalized least squares estimate with a specific Q and a specific set of basis functions.

Beyond Kriging

Likelihood for statistical parameters

$$\boldsymbol{y} \sim MN(\boldsymbol{Zd}, (\rho K + \sigma^2 I))$$

Likelihood

$$\frac{1}{\sqrt{2\pi}^n} \quad e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{Zd})^T(\rho K+\sigma^2 I)^{-1}(\boldsymbol{y}-\boldsymbol{Zd})} \quad |\rho K + \sigma^2|^{-1/2}$$

log Likelihood

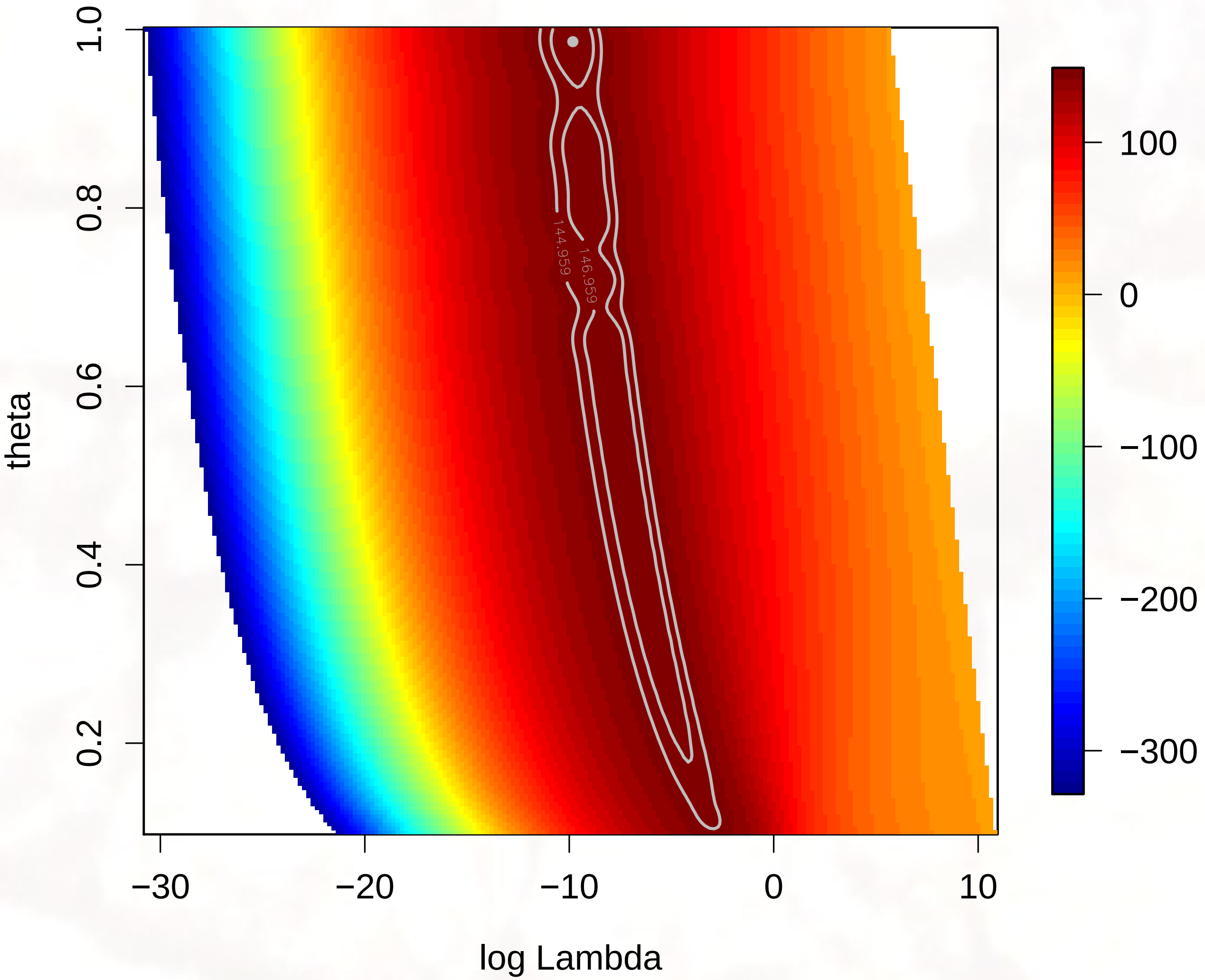
$$-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{d})^T(\rho K+\sigma^2 I)^{-1}(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{d})-\frac{1}{2}log(\mathrm{det}(\rho K+\sigma^2 I))+\mathrm{stuff}$$

With $\lambda = \sigma^2/\rho$

$$-\frac{1}{2\rho}(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{d})^T(K+\lambda I)^{-1}(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{d})-\frac{1}{2}log(\mathrm{det}(K+\lambda I))+-\frac{n}{2}log(\rho)+\mathrm{stuff}$$

- Can maximize this analytically for ρ and β
- K may depend on other parameters e.g. scale, shape

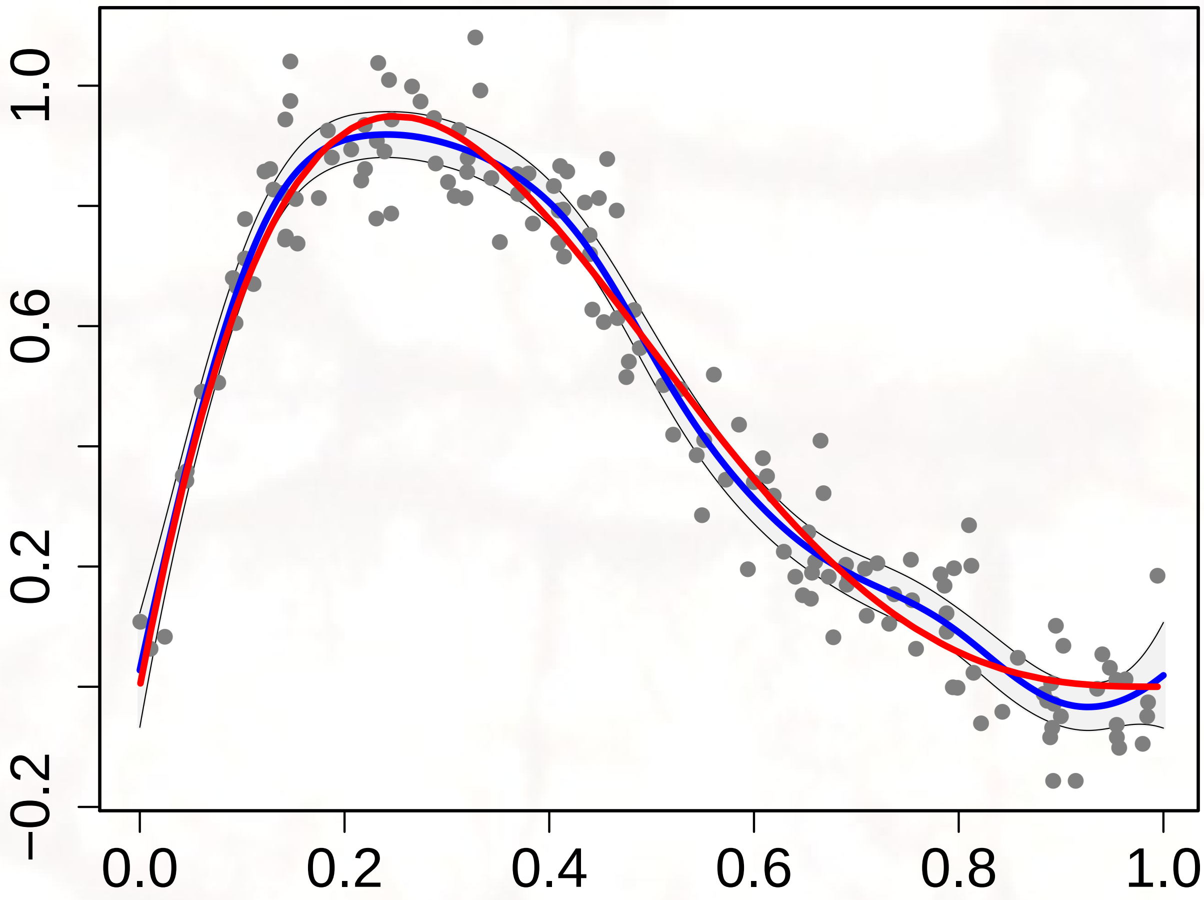
log Likelihood for example



Large uncertainty in the range/scale parameter is typical.

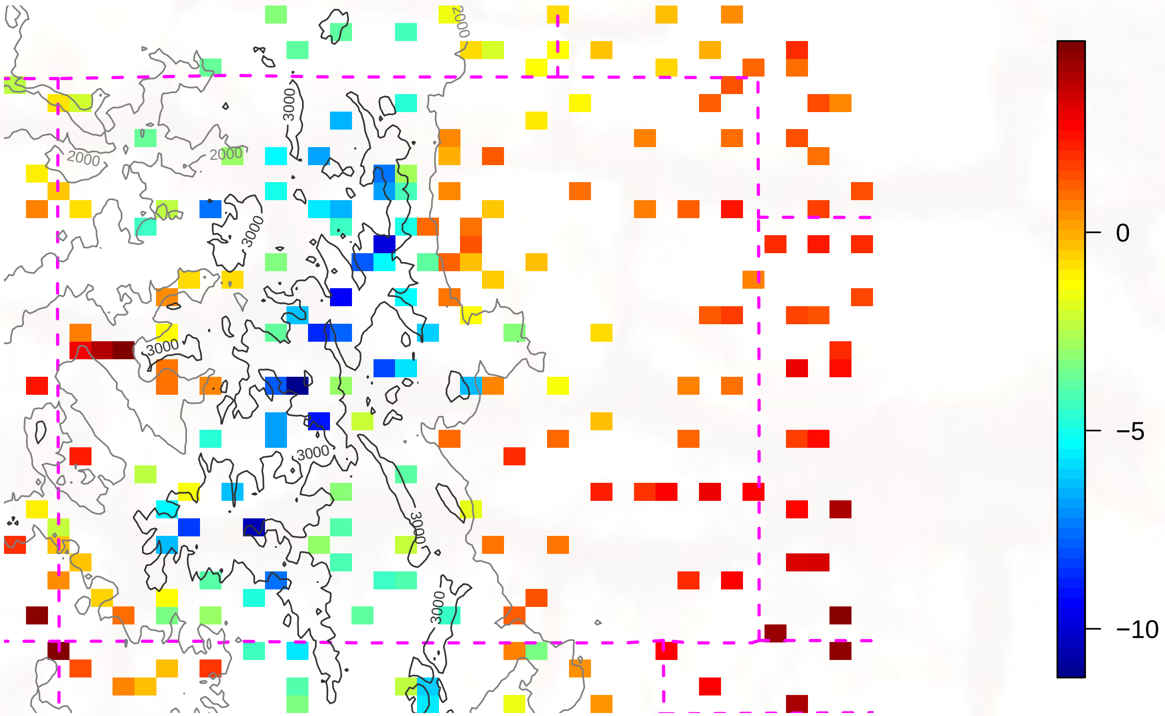
Curve estimate with uncertainty

True function , Estimate, $\hat{g}(x)$

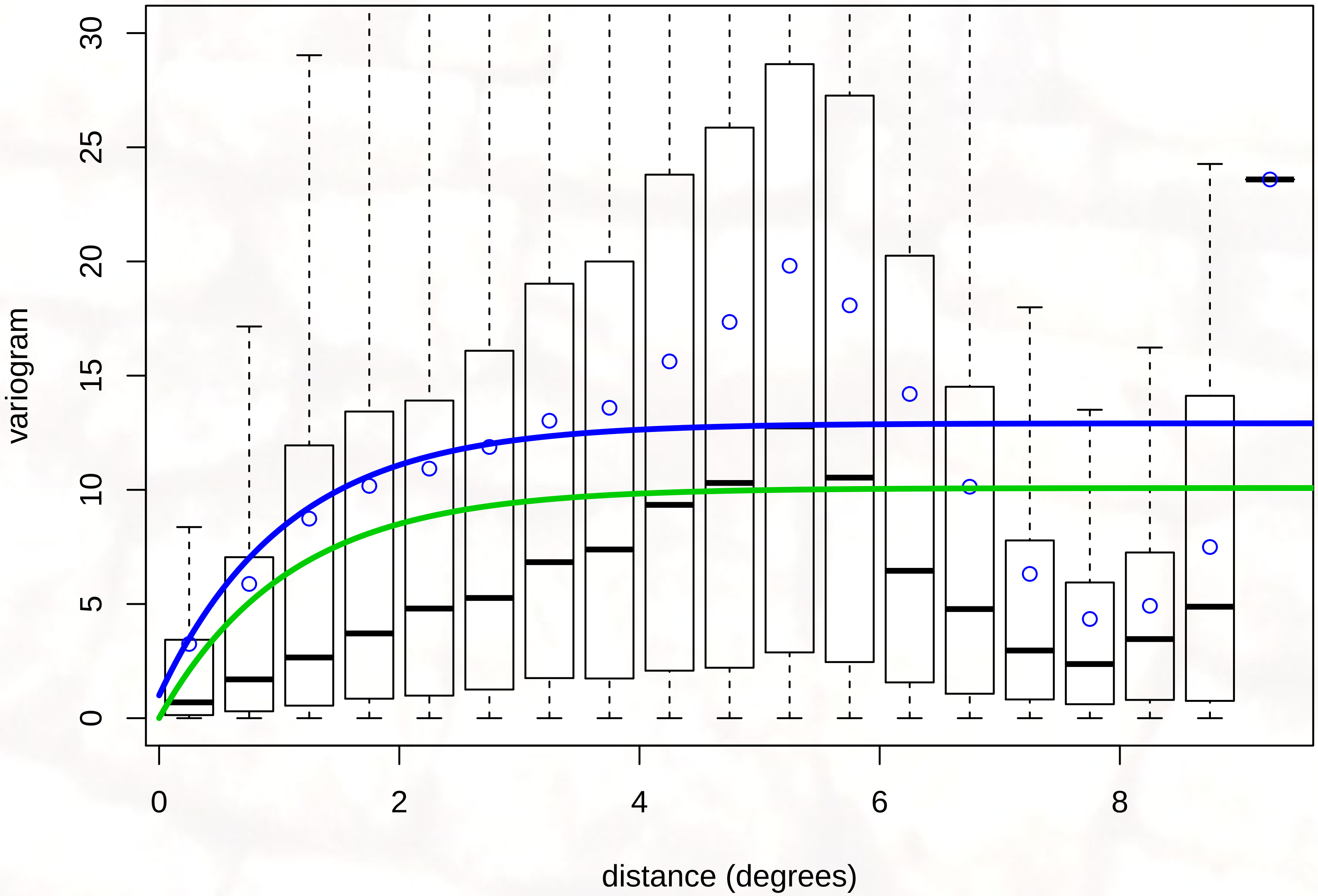


fixed part is linear, Matern covariance
smoothness = 2, $\hat{\theta} = .98$, $\hat{\sigma} = .08$

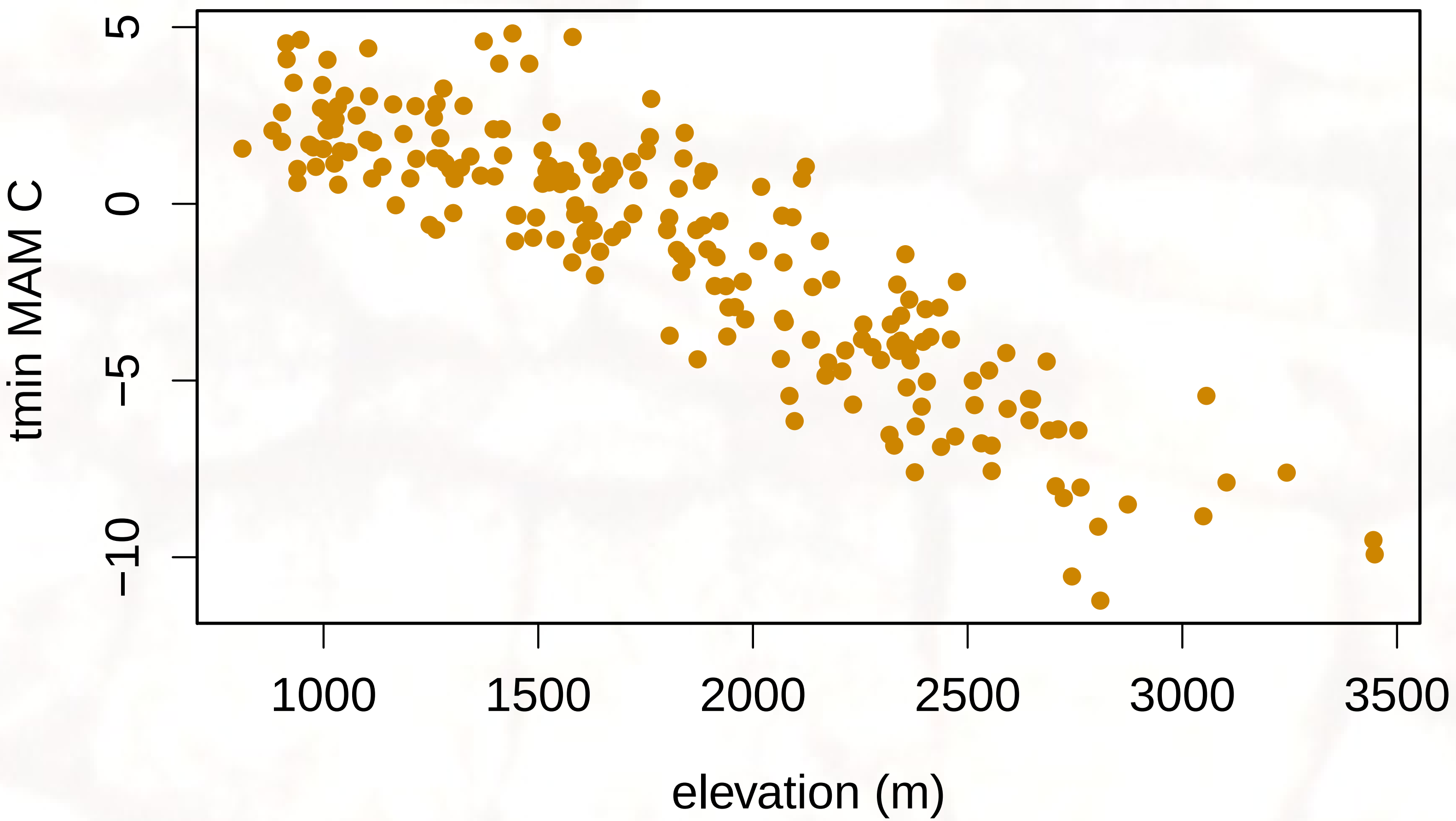
Colorado springtime temperatures



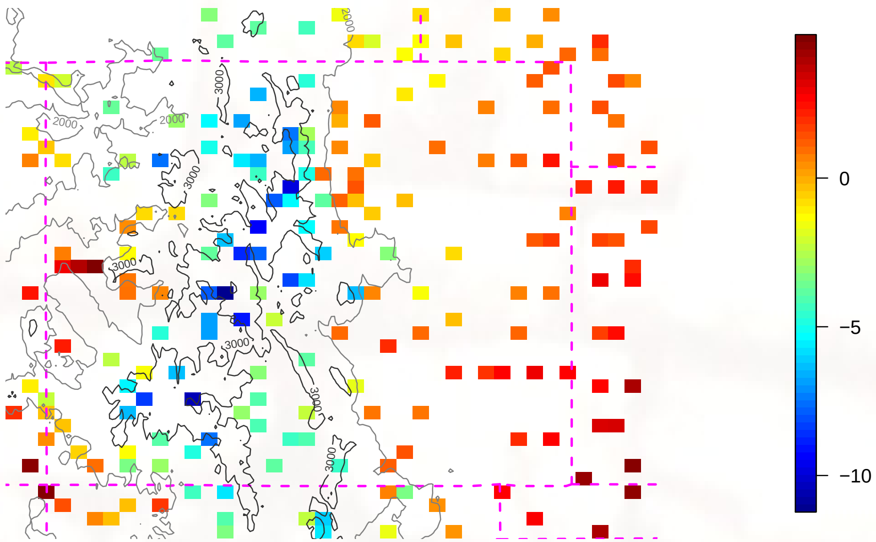
Variogram for the MAM station averages.



What about elevation?



Fitting the MAM Colorado temps



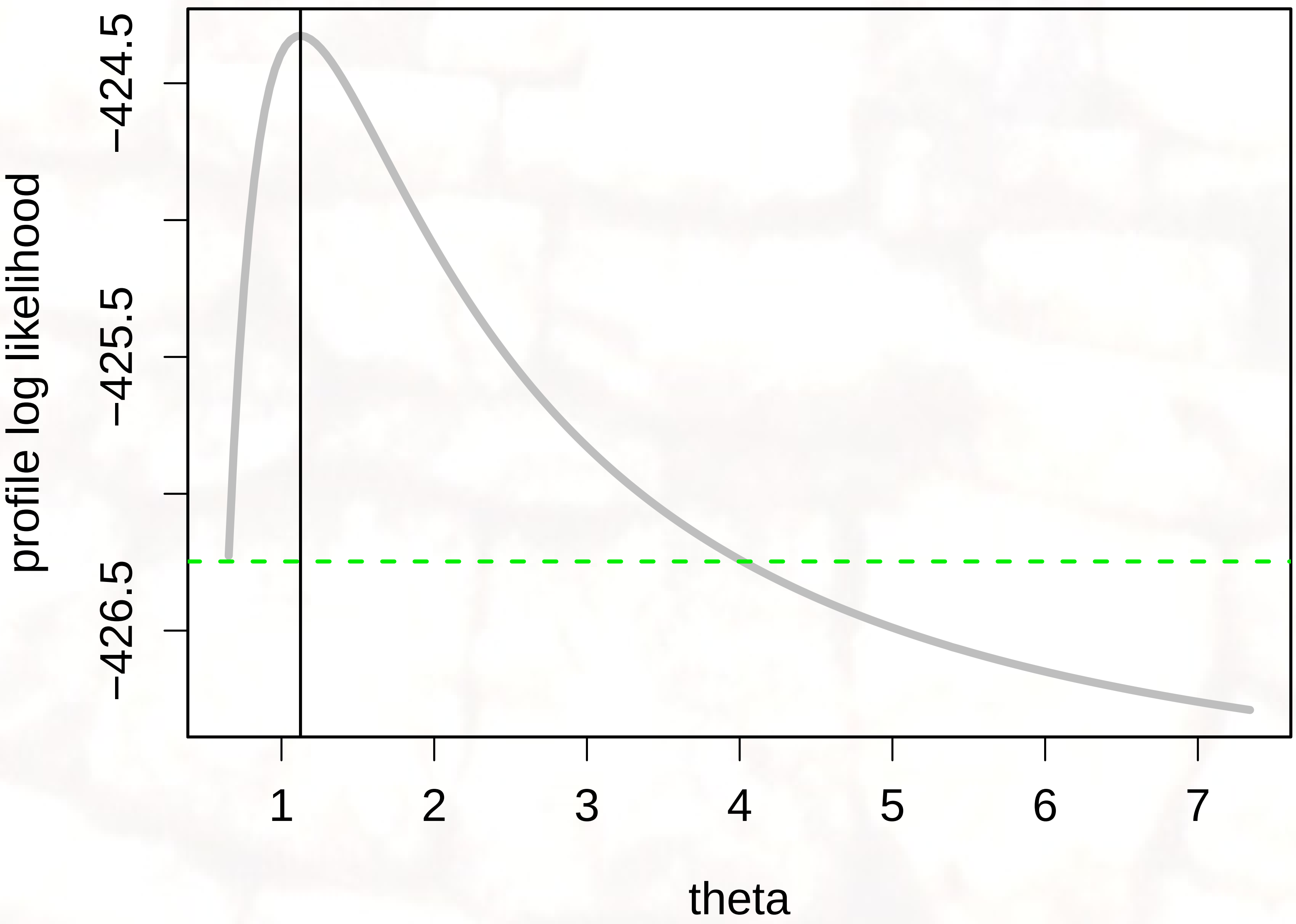
$$\text{Observed}_i = \boldsymbol{d}_1 + \boldsymbol{d}_2 \text{ lon} + \boldsymbol{d}_3 \text{ lat} + \boldsymbol{d}_4 \text{elevation}$$
$$+ \text{ g}(\boldsymbol{x}_i) + \text{error}$$

Use maximum likelihood to find:

- $\boldsymbol{d}$,
- VAR(g), VAR(error)
- range and smoothness (??) of Matern.

Profile likelihood for θ

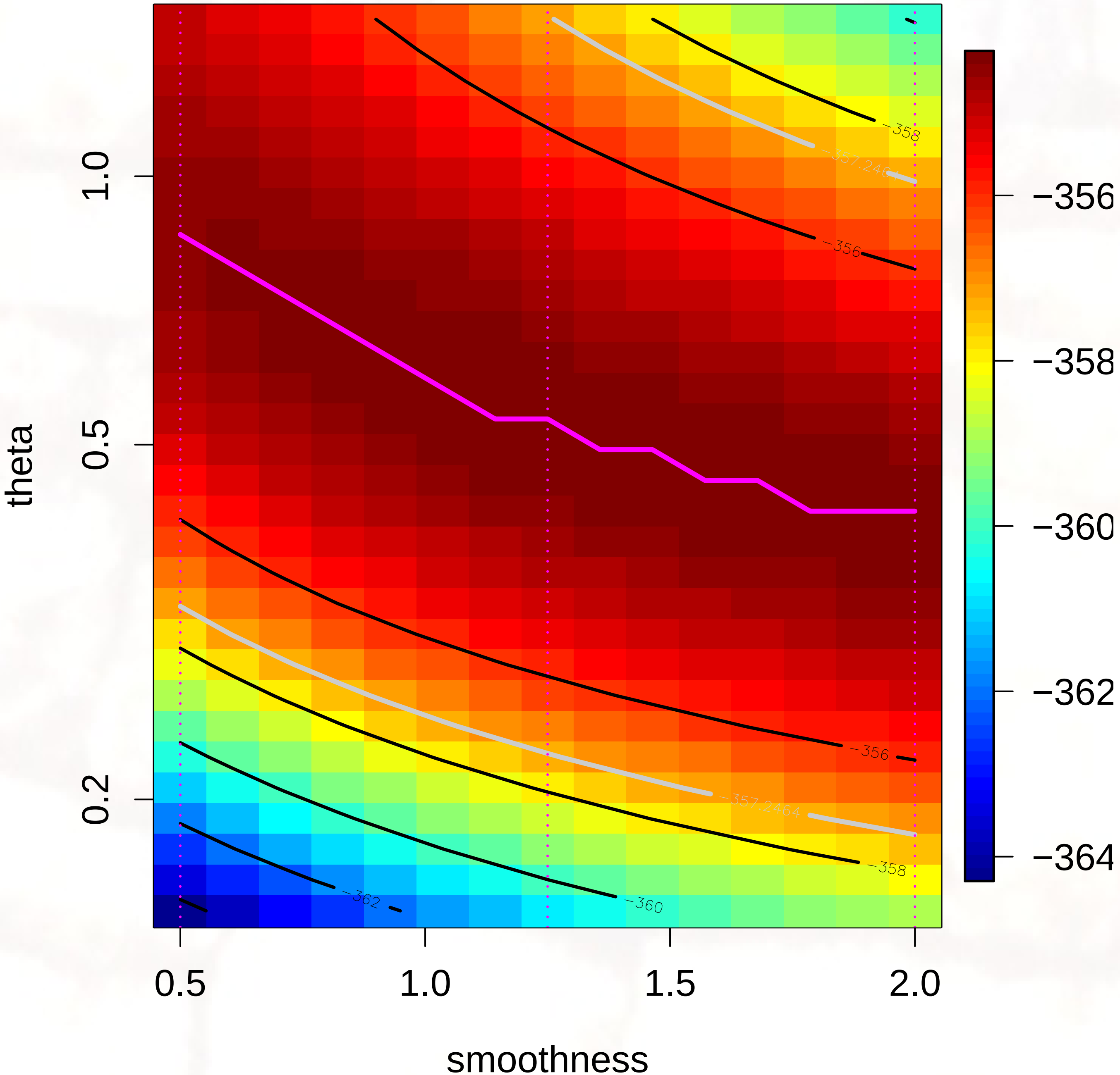
No elevation term and smoothness of 1.0



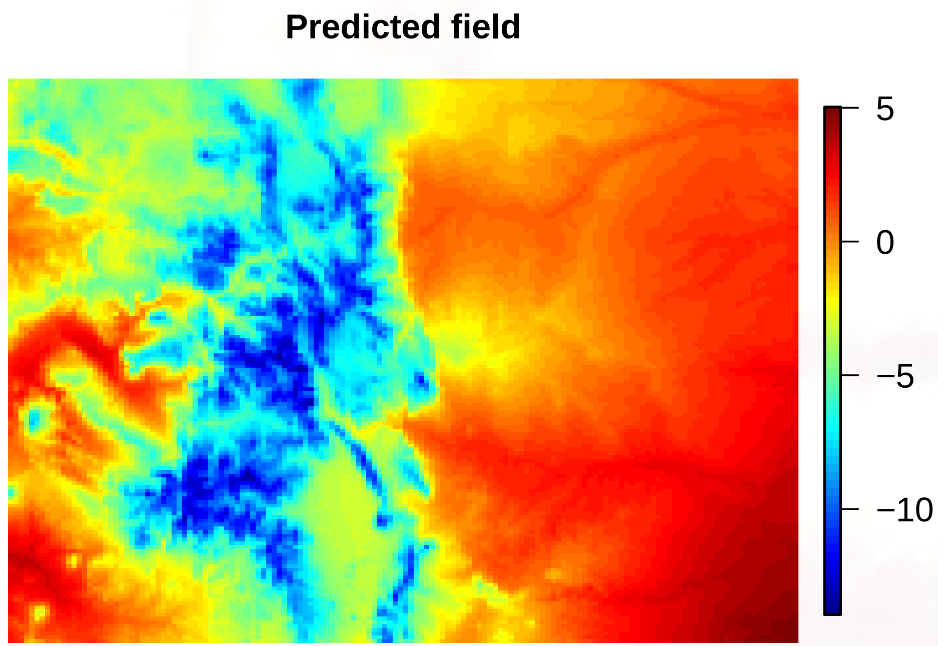
Line at approximate 95% confidence bound.

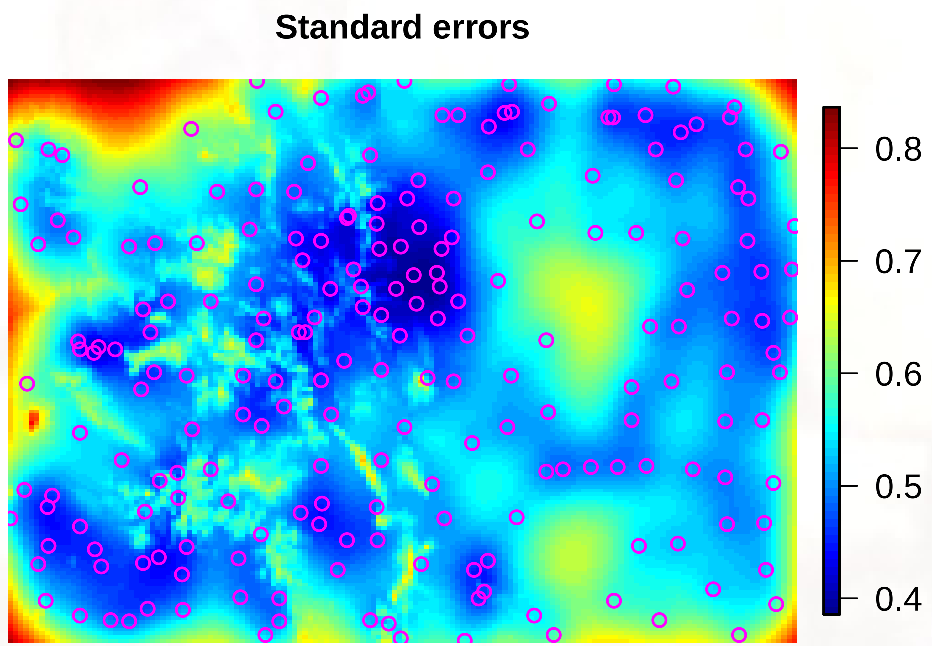
Likelihood θ and smoothness

Marginal maxima, 95% confidence bounds

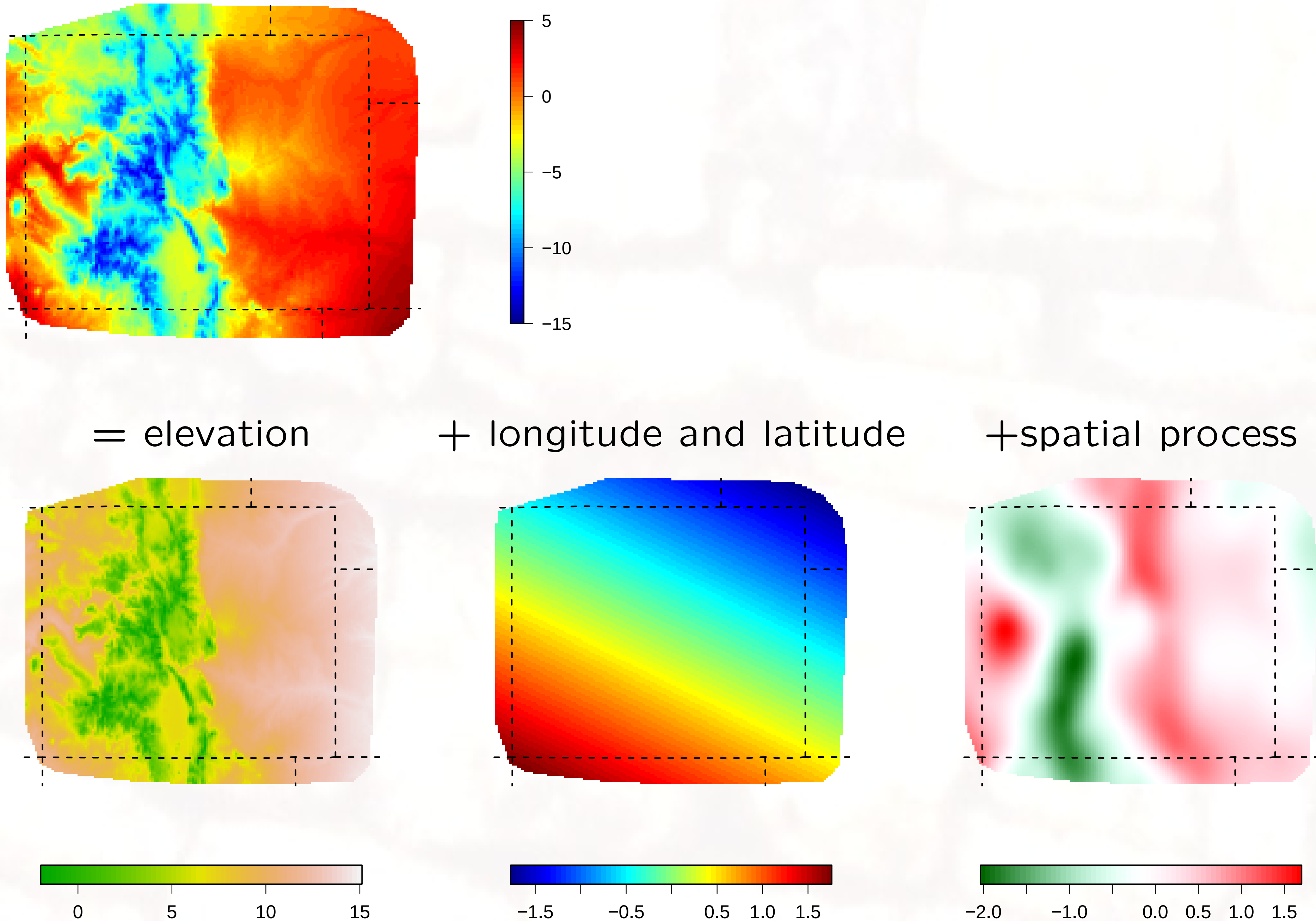


Predicted temperature field





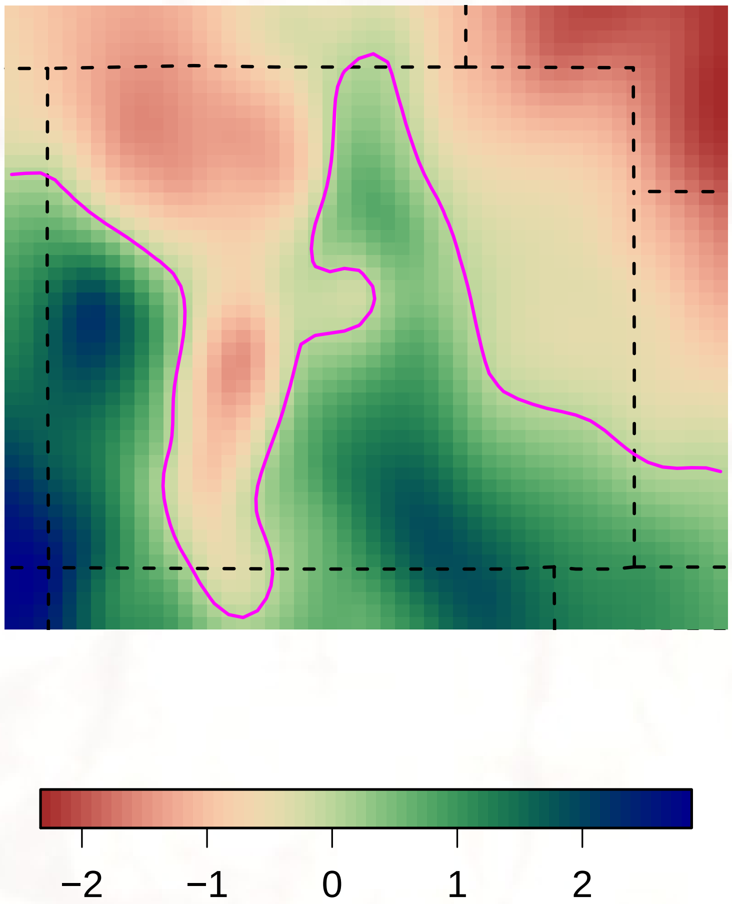
Decomposition of the fit



Inference beyond $\widehat{g}$

The mean surface w/o elevation

Components based on lon/lat and smooth surface.



What is uncertainty of the zero contour?

Conditional simulation

For fixed covariance parameters sample conditional distribution of the field given station data.

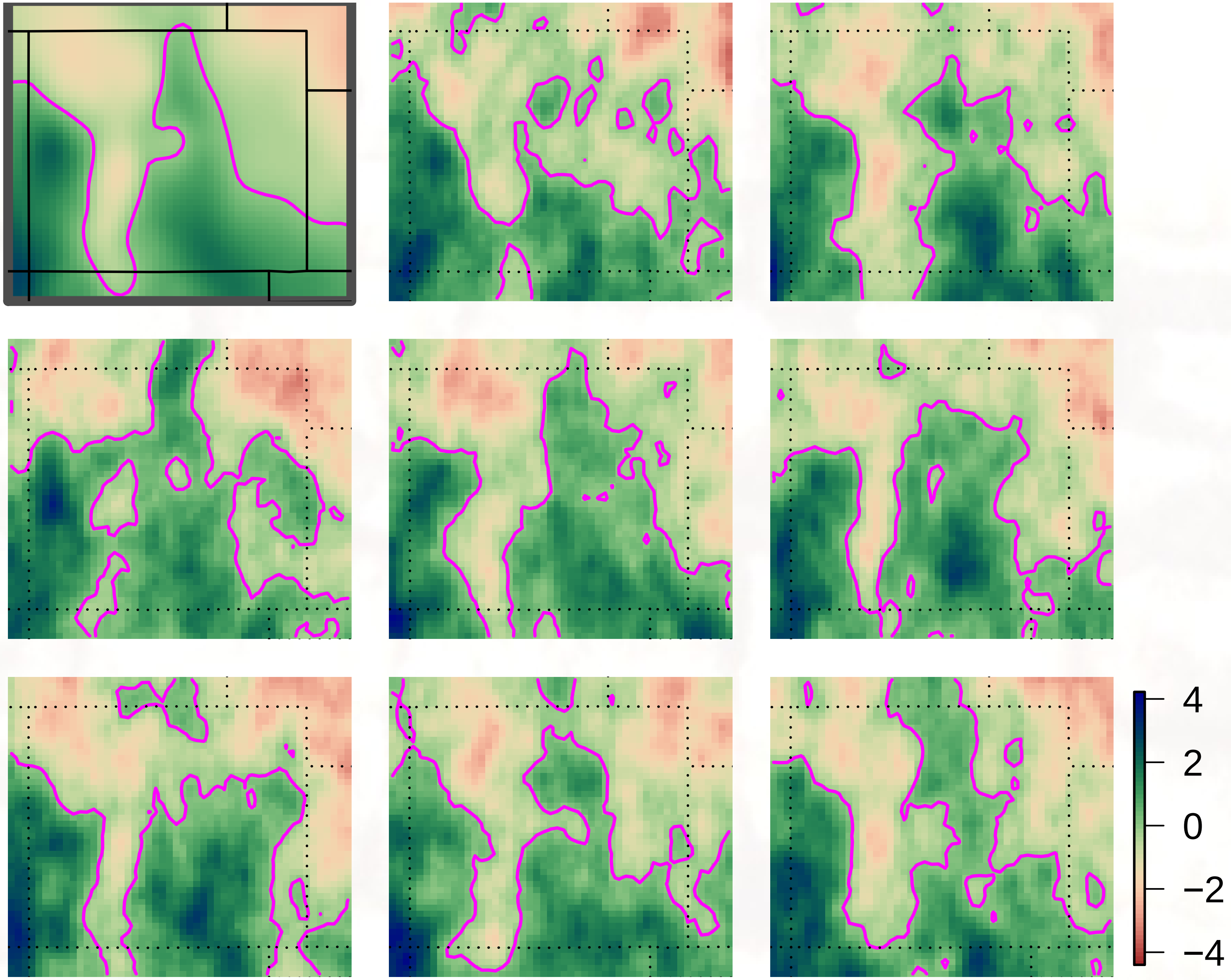
- Any subset of $g(\boldsymbol{x})$ *given* $\boldsymbol{y}$ is multivariate normal.

- The standard deviation of the conditional samples is a Monte Carlo estimate of the prediction standard error.

What is uncertainty of the zero contour?

- Conditional samples can be used for hard inference problems.

Ensemble for Colorado MAM temps



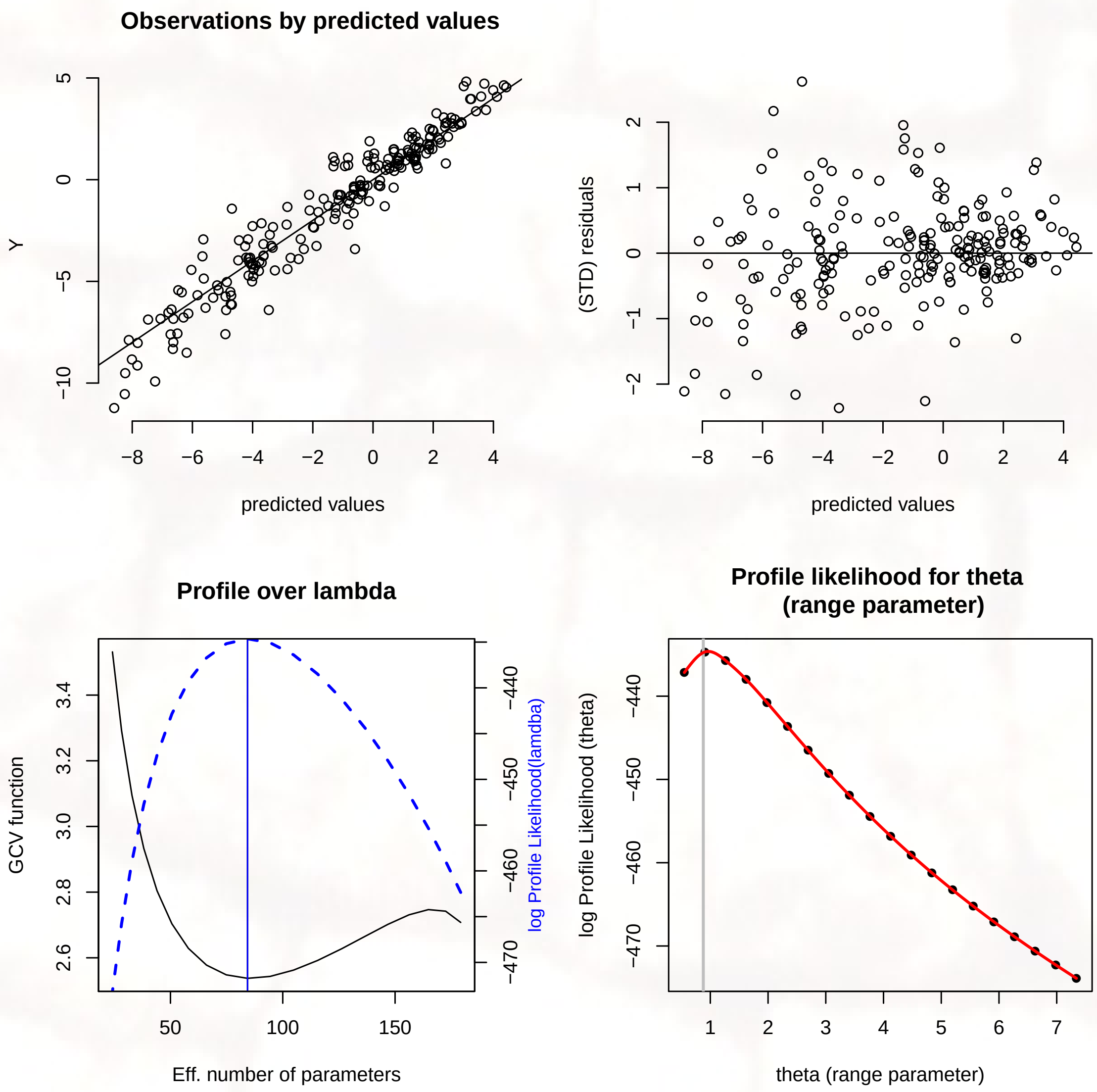
In R

`x` and `y` Colorado locations and MAM temps.

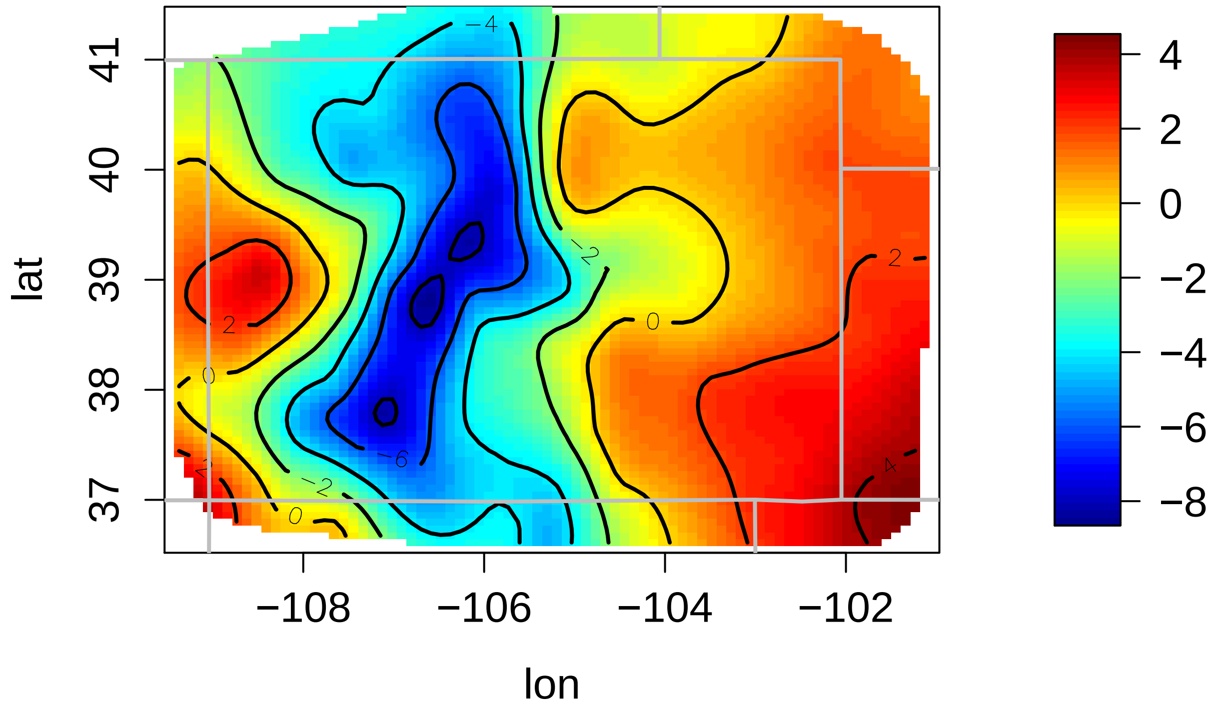
```
fit1<- spatialProcess( x,y)

fit1E<- spatialProcess( x,y, Z = elev)
```

```
set.panel(2,2)
plot( fit1)
```



```
surface( fit1)
US( add=TRUE)
```



Thank you

Variogram as an EDA tool

How to estimate the spatial from a single field?

- Plot for a spatial data set or spatial field plot $\frac{(\boldsymbol{y}_i - \boldsymbol{y}_j)^2}{2}$ against the distance of separation. "On the average" this should be follow the theoretical curve that is the variance of the data *minus* the covariance function .

covariance $\rightarrow$ variogram or variogram $\rightarrow$ covariance

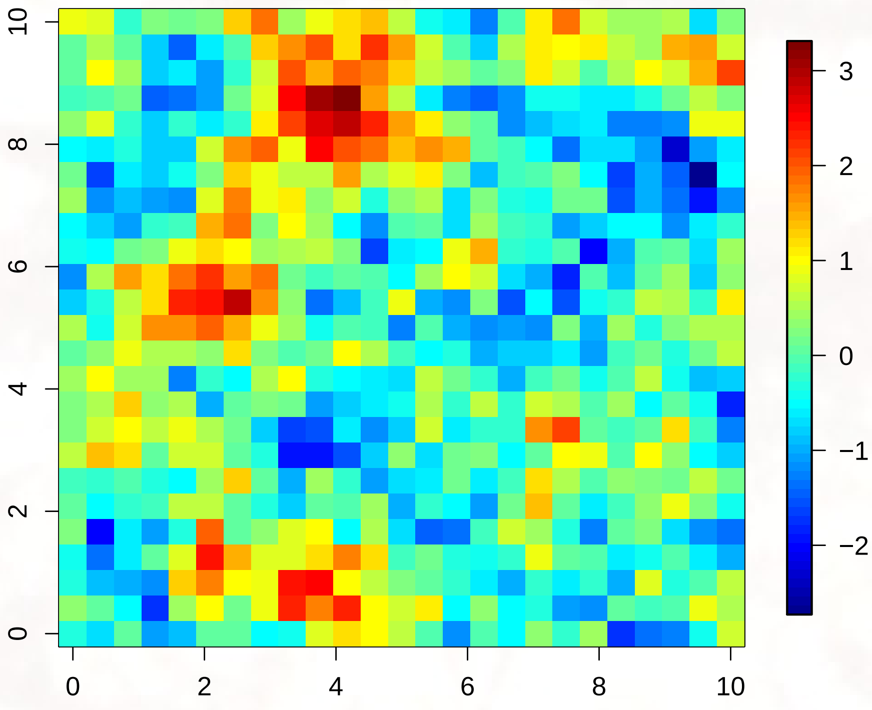
- If one can find find the variogram then one can transform back to the covariance function.

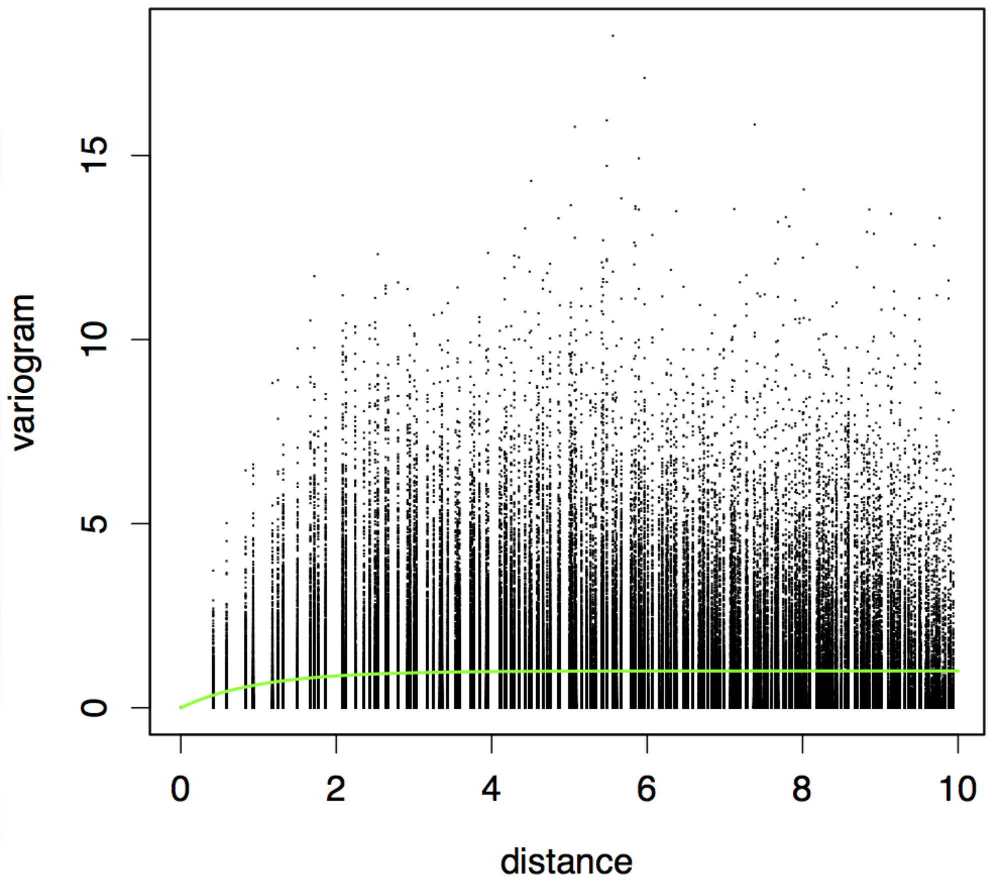
- Need to be careful about how the variogram behaves close to zero distance. The variogram estimates $\rho + \sigma^2$ right at zero – not just ρ

- Great EDA tool – terrible for actually estimating parameters!

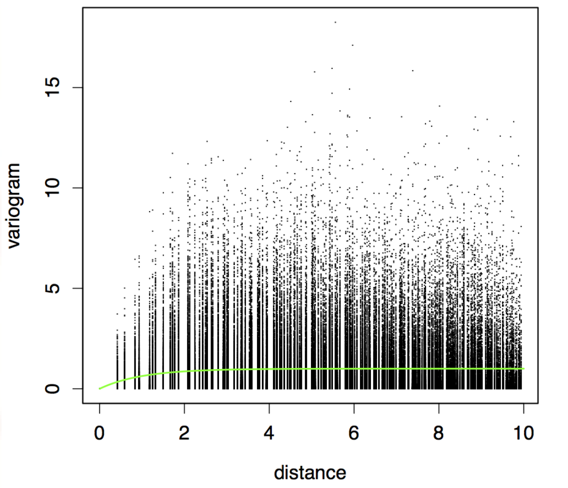
A variogram example

Sample field with true exponential variogram on a 25×25 grid.





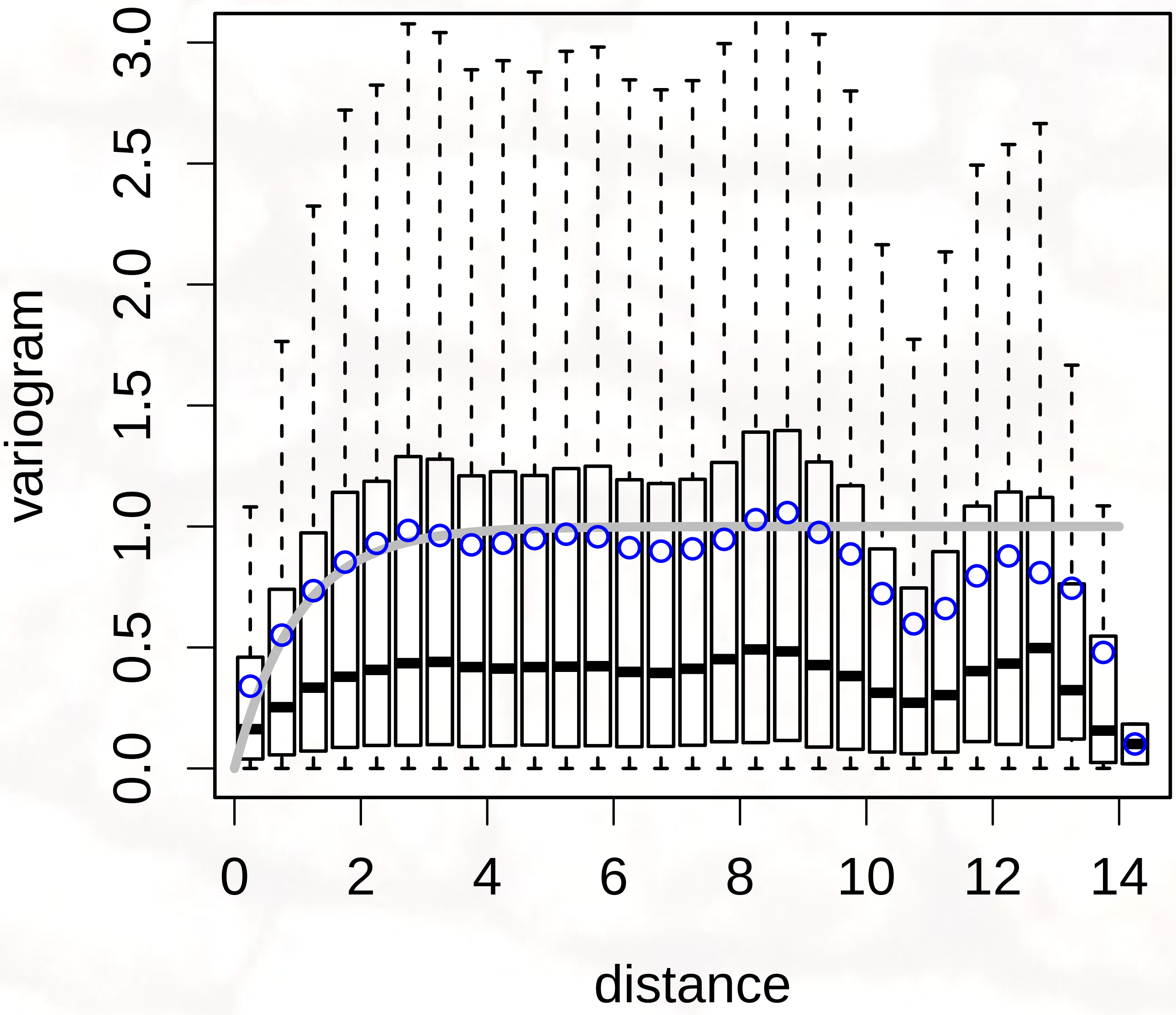
Based on (625 *choose* 2) total pairs
$\approx$ 200K differences: $(y_i - y_j)^2/2$



Houston, we have a problem.

Improving the variogram

Binning observations by distance ranges, finding box plots and adding the mean for each bin.



Identifying the nugget σ^2

Recall the additive model:

$$y_i = g(x_i) + \epsilon_i$$

Correlations among the observations due to the smooth field but the measurement error is uncorrelated. Adding measurement error to the example ($\sigma = .4, \sigma^2 = .16$)