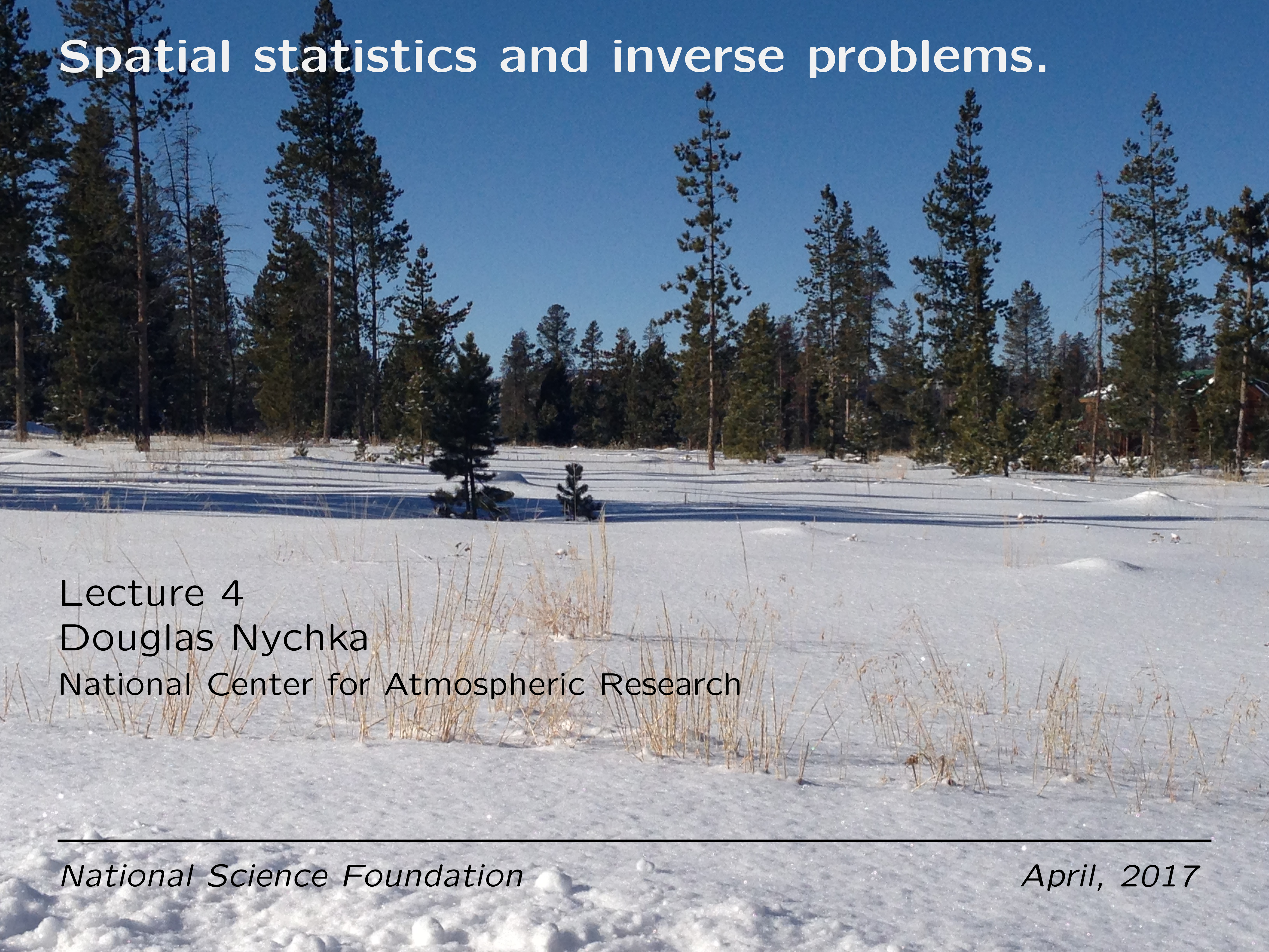


Spatial statistics and inverse problems.

Lecture 4

Douglas Nychka
National Center for Atmospheric Research

National Science Foundation *April, 2017*

Outline

- CO2 reconstruction problem from ice cores
- Hierarchical model (Data, process, priors)
- Blending spatial fields of snow cover
- Reconstructing the solar atmosphere

Main ideas:

- separating what is observed from what is to be estimated.
- model the unknown function using spatial statistics
- maximum likelihood/ Bayesian inference for model parameters
- use the LatticeKrig model to handle large data sets.

An Additive model

Connection with data:

$$\boldsymbol{y}_i = L_i(g) + e_i$$

- Observations made at irregular locations
  – or as a linear functional

- ... and some random error added.

Representing the surface:

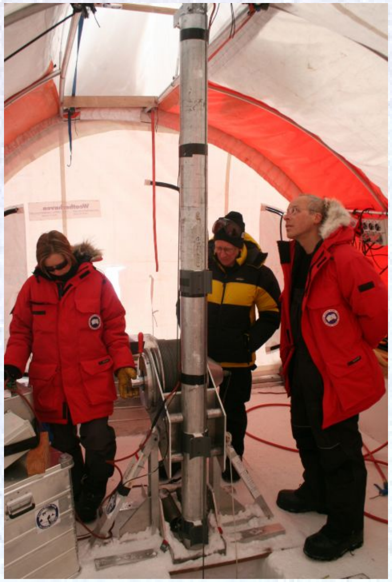
g is a Gaussian stochastic process or explicilty

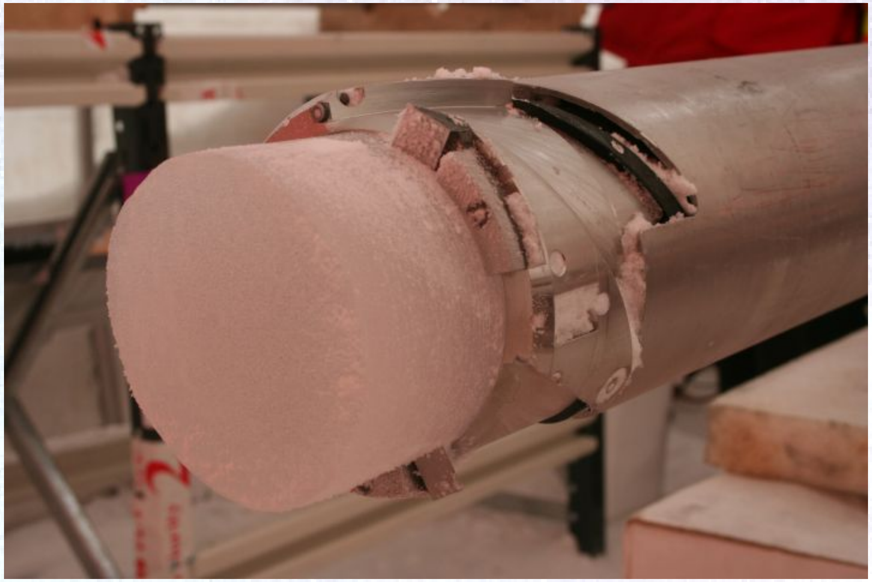
$g(x) = \sum_j \phi_j(x) c_j$ and $L_i(g) = \sum_j L_i(\phi_j) c_j$

Goal: Estimate g and quantify the uncertainty.

What were the annual CO_2 concentrations for the atmosphere over the past 2000 years?

Law Dome, East Antarctica

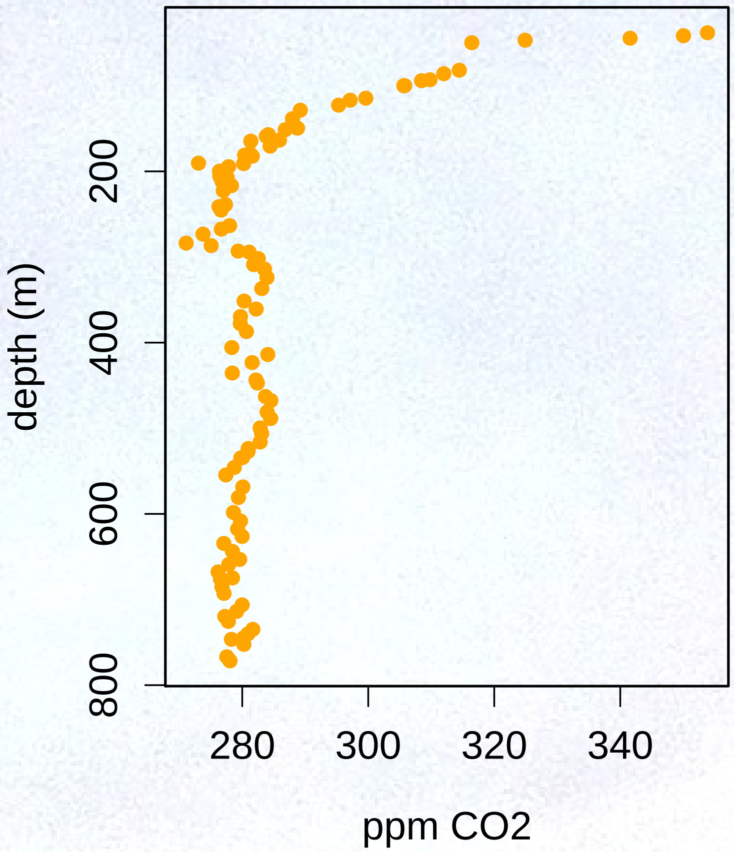




Ice core drill and glaciologists from the Australian Antarctic Division and Antarctic Climate and Ecosystems CRC, Law Dome, East Antarctica

Observations

CO_2 concentrations as a function depth.



Conceptually: Depth is related to the *time* air was trapped in the core.

$$y(\text{depth}) = \mathcal{F}(g(\text{time}))$$

$\mathcal{F}$ developed by Trudinger et al. (2013)

Inverse problem: Invert CO_2 concentrations by depth to concentrations by time.

$$g(\text{time}) = \mathcal{F}^{-1}(y(\text{depth}))$$

Firn Ice Inverse problem harder for upper ice layers that have not completely consolidated.

A Hierarchical (geophysical) model

The goal:
Estimate $g(t)$, CO_2 atmospheric concentration at time t.

Data level: A statistical model for the data given you know the process of interest

Process level: A statistical/physical model for the process – does not depend on the observations but possibly on other parameters.

Priors: Probability distribution that indicate likely ranges for parameters.

(I like to cheat on the Priors!)

Each level is based on conditional probabilities.

Gelfand Notation

X, Y, Z are random variables

- $[X]$ probability density function for X

  e.g. $[X] = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2} = N(\mu, \sigma^2)$

- $[X, Y]$ joint probability density function for X and Y

  e.g. $[X, Y] \sim \frac{1}{\sqrt{2\pi}} e^{-(x^2 - 2\gamma xy + y^2)} = N(\boldsymbol{\mu}, \boldsymbol{\Omega})$

- $[Y|X]$ conditional probability density function for Y *given* X

  e.g. $[X|Y]$ is $N(\gamma y, 1 - \gamma^2)$

Properties

- By definition: $[X, Y] = [Y|X][X]$
- Iterating: $[X, Y, Z] = [Y|X, Z][X|Z][Z]$

Bayes Theorem

$$[X|Y] = \frac{[Y|X][X]}{[Y]}$$

From "Y given X" to "X given Y" !

Back to Hierarchical Models

Y - observations at depth ,
$\boldsymbol{g} = g_1, g_2, ..., g_n$ process at closely spaced times (yearly,)
θ parameters

Data level: $[Y|\boldsymbol{g}, \theta]$

Process level: $[\boldsymbol{g}|\theta]$

Priors: $[\theta]$

Joint distribution:

$$[Y|\boldsymbol{g}, \theta][\boldsymbol{g}|\theta][\theta]$$

Now Bayes theorem

Inversion (also called posterior)

$$[\boldsymbol{g}, \theta | Y] \text{ proportional to } [Y | \boldsymbol{g}, \theta][\boldsymbol{g} | \theta][\theta]$$

take logs

$$log([\boldsymbol{g}, \theta | Y]) = log([Y | \boldsymbol{g}, \theta]) + log([\boldsymbol{g} | \theta]) + log([\theta]) + \text{constants}$$

fit to data weak constraint on $\boldsymbol{g}$ parameters

- This is a complete probability density that quantifies uncertainty in result.
- For fixed parameters we recognize $[g|Y, \theta]$ as “Kriging”

Data model

The measurements:

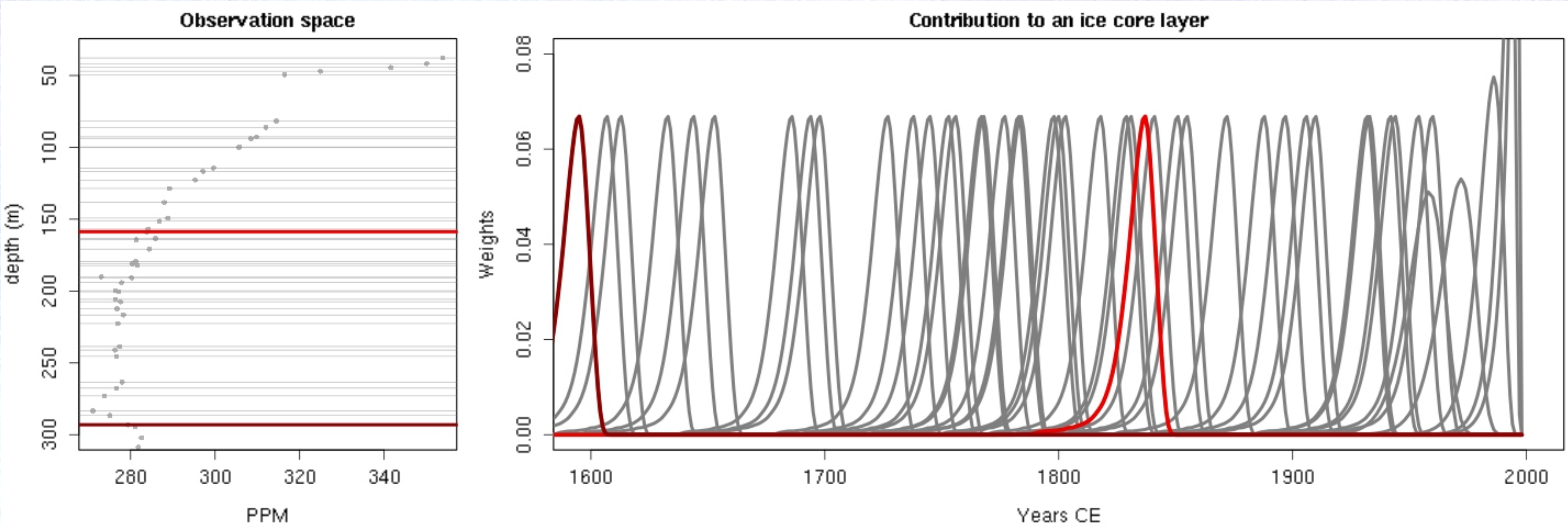
$$\boldsymbol{y} = W\boldsymbol{g} + \boldsymbol{e}$$

$\boldsymbol{e} = e_1, e_2, ..., e_m$: measurement errors – assumed to be $N(0, \sigma^2)$

$W\boldsymbol{g}$: Expected concentrations at observed depths
– the forward model $\mathcal{F}(g)$

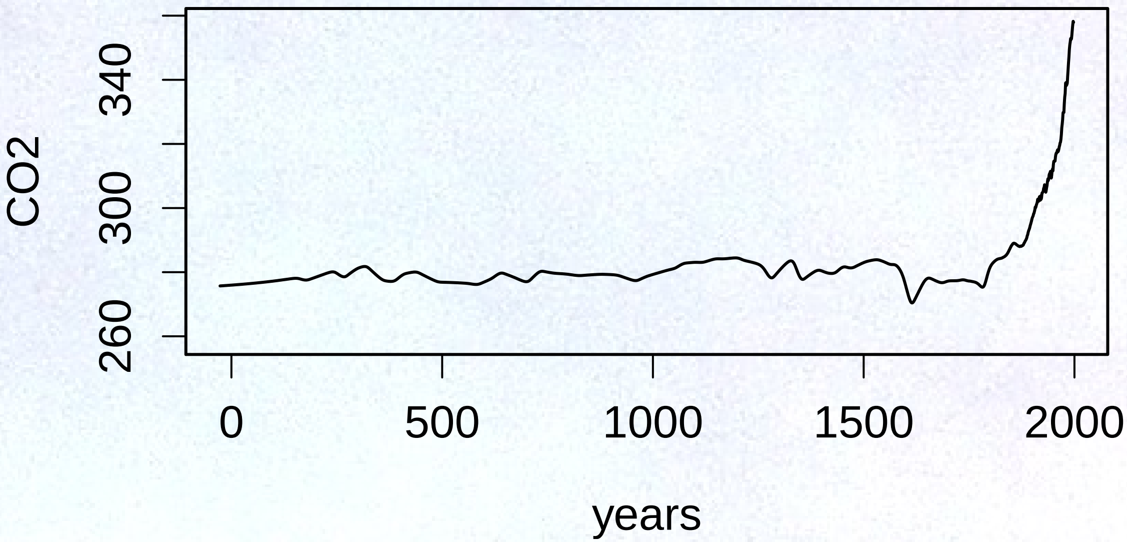
Effect of indirectly observing CO_2.

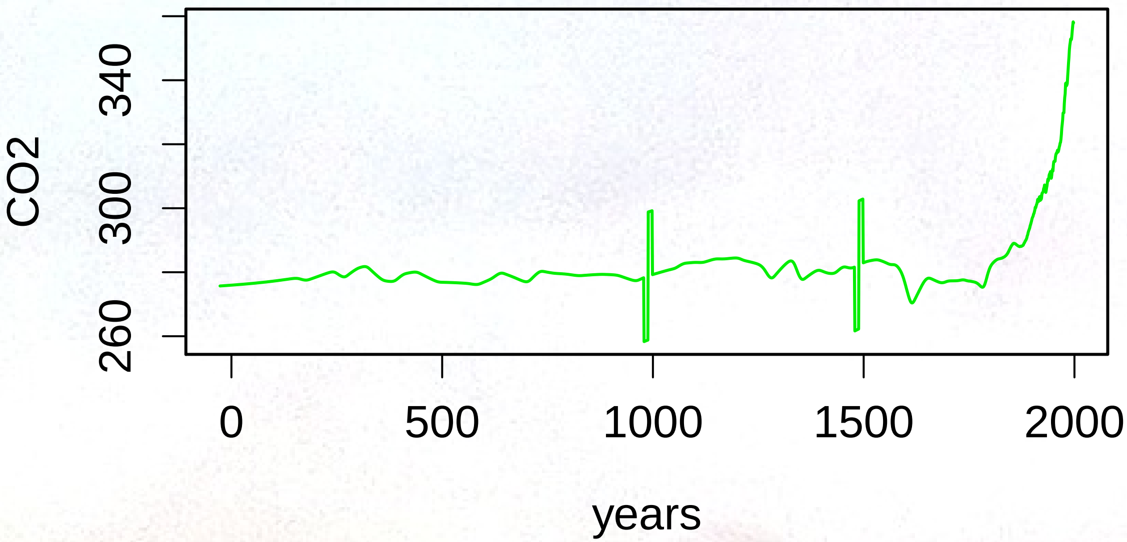
Rows of W for most recent ice layers.



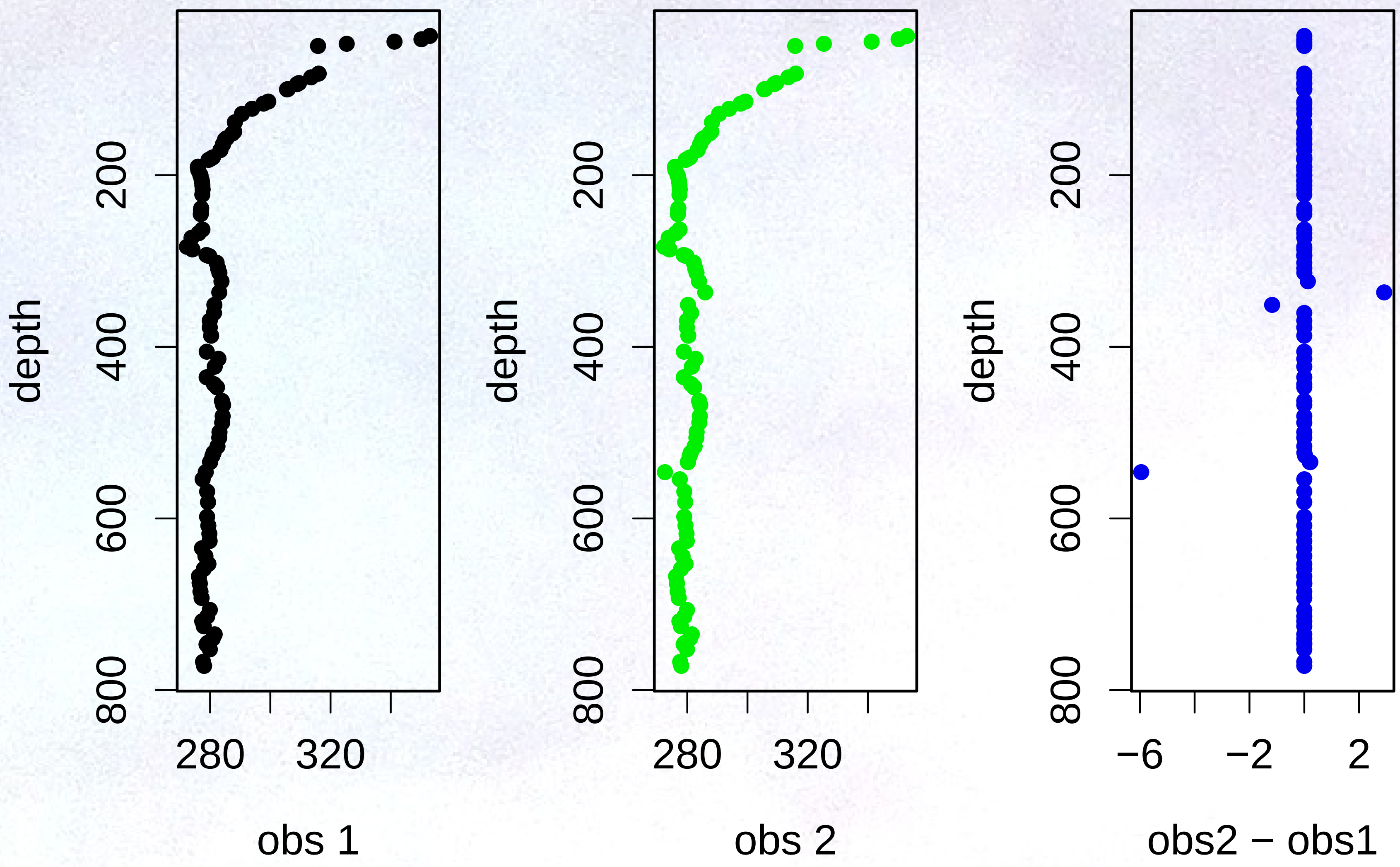
Problems caused by the weights

Two possible CO_2 annual series:





What we would see in ice core



- Large differences in g only give subtle differences in the observations.
- Easily masked by noise.

Process level

Annual concentrations based on a linear trend, human emissions covariate, and autoregressive "noise".

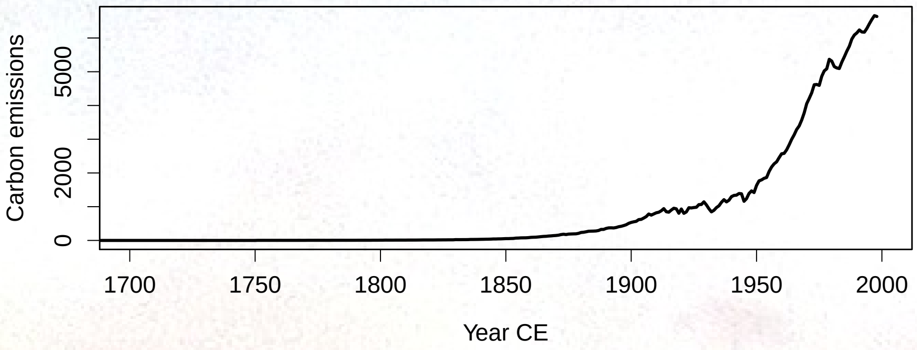
$$g_t = \alpha_1 + \alpha_2 t + \alpha_3 \text{Emissions}(t) + u_t$$

$$u_t = \beta u_{t-1} + v_t$$

v_t are uncorrelated normals and the variance of u_t is ρ

This can be written as $\boldsymbol{c}$ is $N(T\boldsymbol{\alpha}, \rho\boldsymbol{\Sigma}(\beta))$

Recent carbon emissions: Emissions(t)

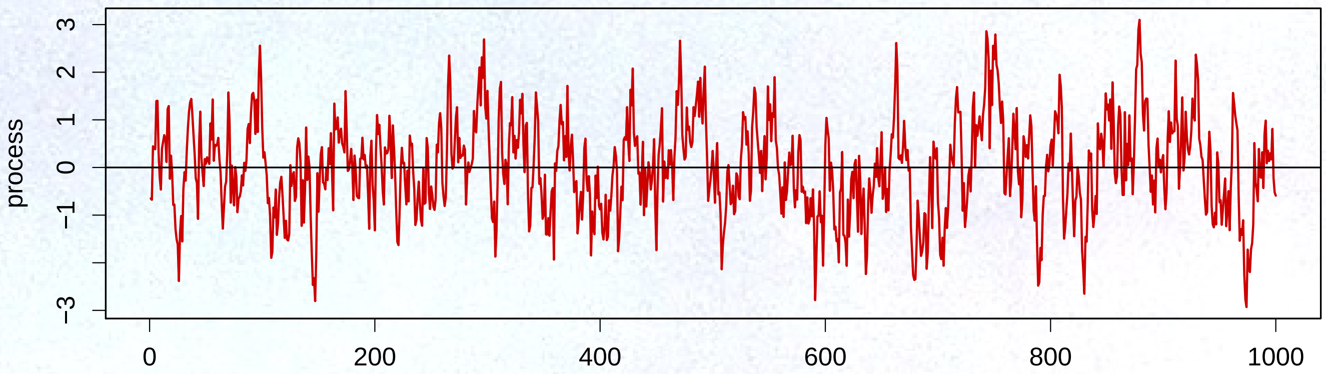


more on the autoregressive noise

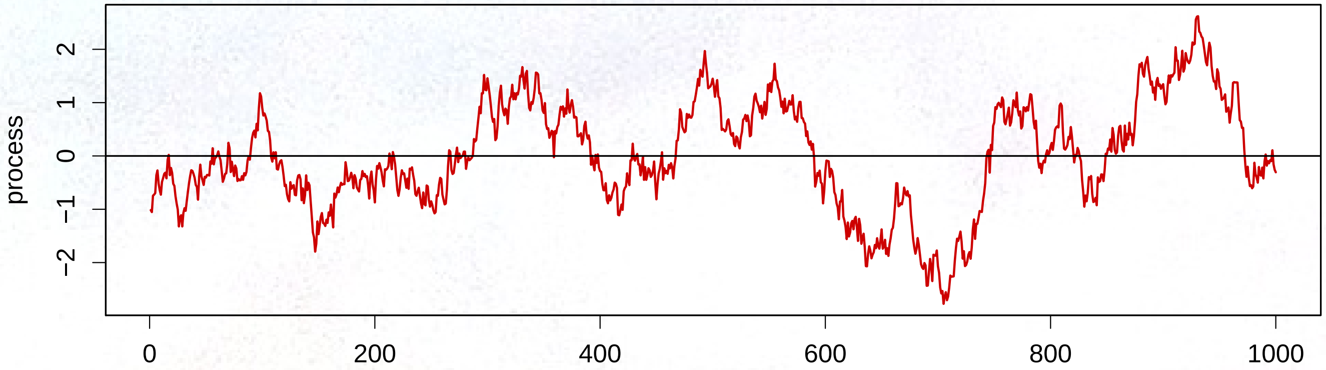
AR 1 process:

$u_t = \beta u_{t-1} +$ white noise – same as $corr(u_t, u_s) = e^{-|t-s|/\delta}$

$\beta = .8$



$\beta = .99$



- Just a 1-d spatial process with exponential covariance function.

Prior level – statistical parameters

- σ^2 measurement error variance
- $\alpha_1, \alpha_2, \alpha_3$ regression parameters
- β autoregressive parameter (correlation)
- ρ variance of autoregressive process.

Empirical Bayes

For this application we will estimate these parameters by maximum likelihood and just use these values.

Maximum likelihood $\equiv$ *uniform (or flat) prior distribution*

Finding parameters

Can integrate out g

$$\boldsymbol{y} \sim MN(\quad WT\boldsymbol{\alpha}, \quad \rho W \boldsymbol{\Sigma}_\beta W^T + \sigma^2 I)$$

Σ_β the covariance matrix for $\boldsymbol{g}$.

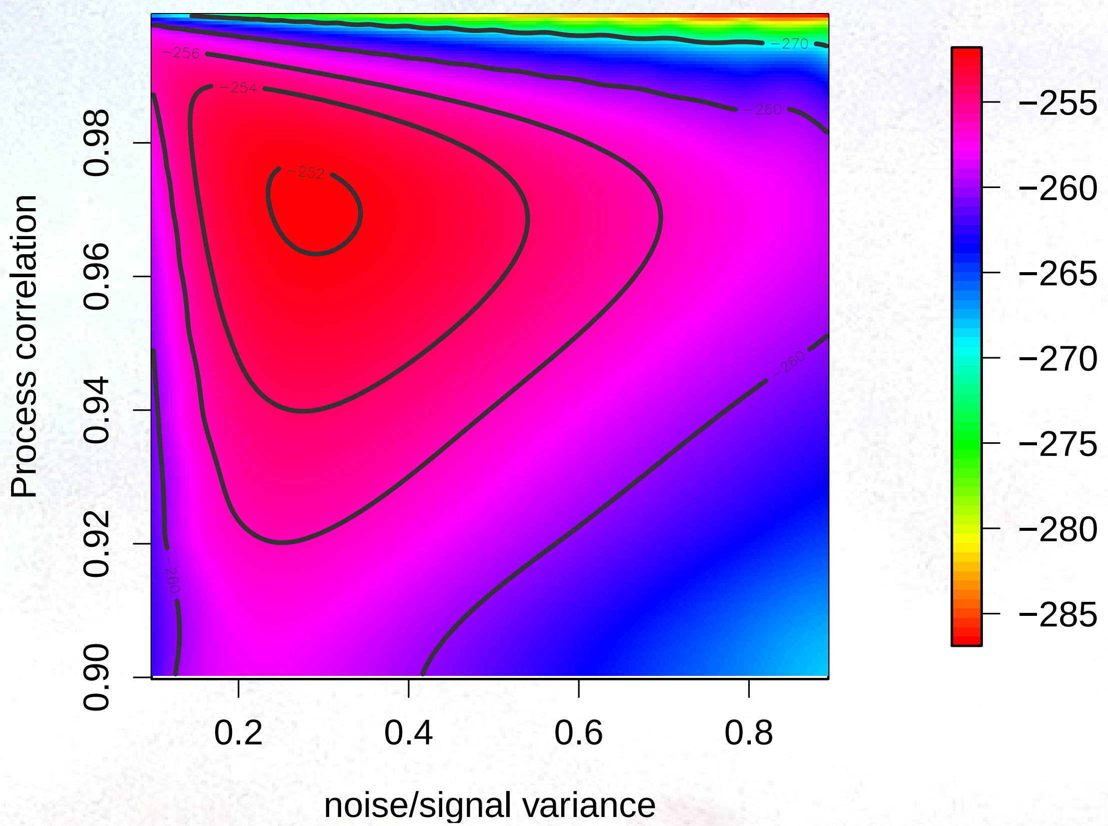
- Find parameters by maximum likelihood
- or add in a prior for a Bayesian inference.

The likelihood surface

Likelihood can be maximized analytically over all parameters except for AR1 parameter and " noise to signal " ratio (λ from Lecture 1).

Maximize these last two numerically.

log Likelihood surface



Finding the concentrations

Kriging step.

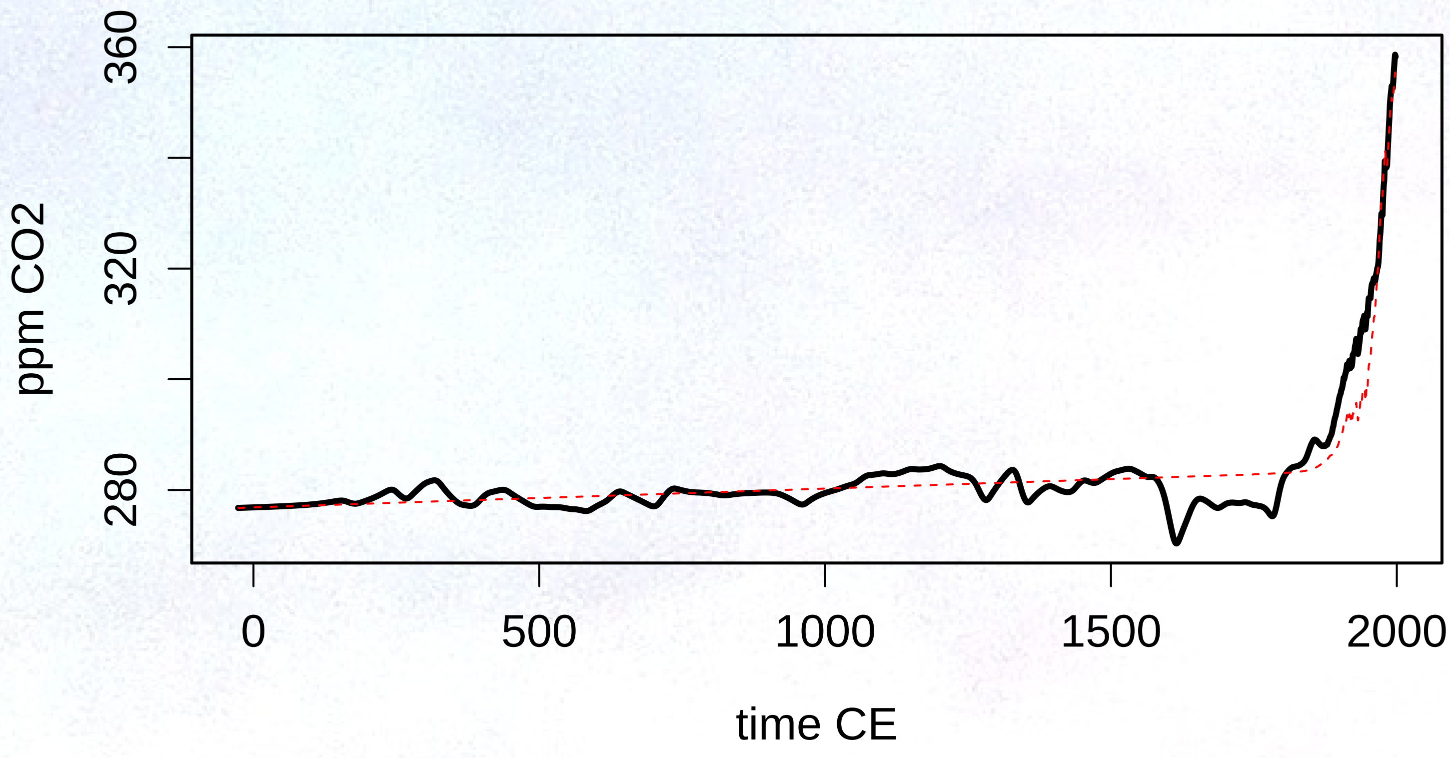
- $[g(.)|\theta, Y]$ is multivariate normal.

- The conditional expectation is also the penalized least squares solution.

$$\hat{g} = (W^T W + \lambda(\Sigma_\beta)^{-1})^{-1} W^T (\boldsymbol{y} - WT\alpha)$$

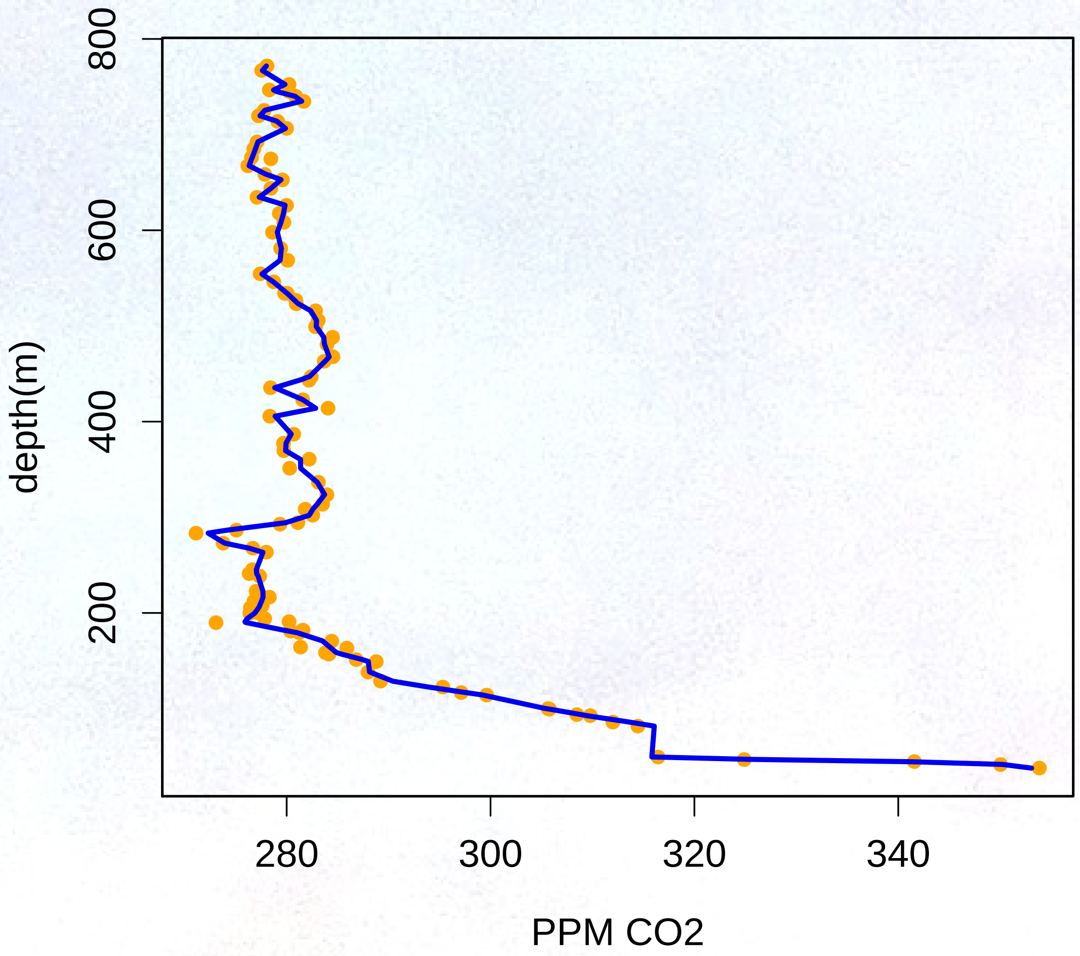
Estimate of $g(t)$

dashed linear trend and scaled emission



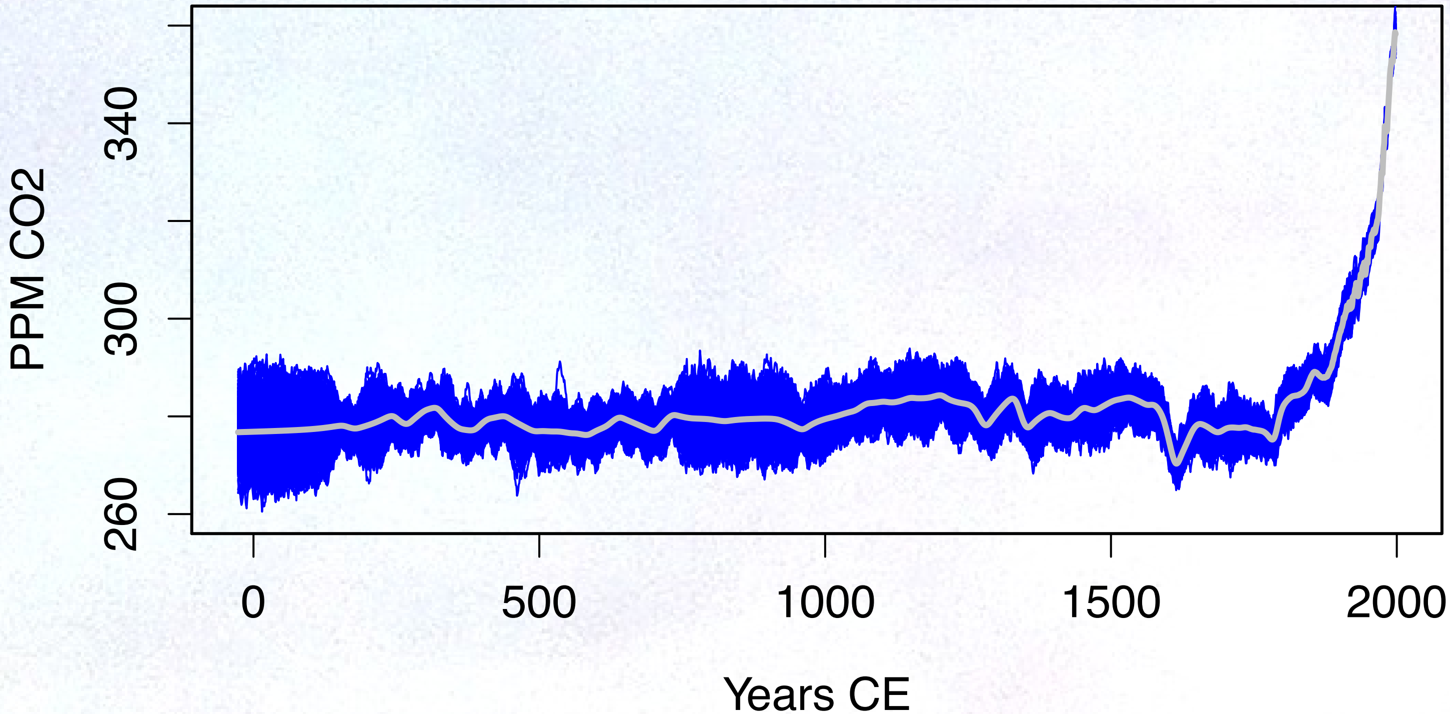
Fit to Observations

observations, predicted from concentrations



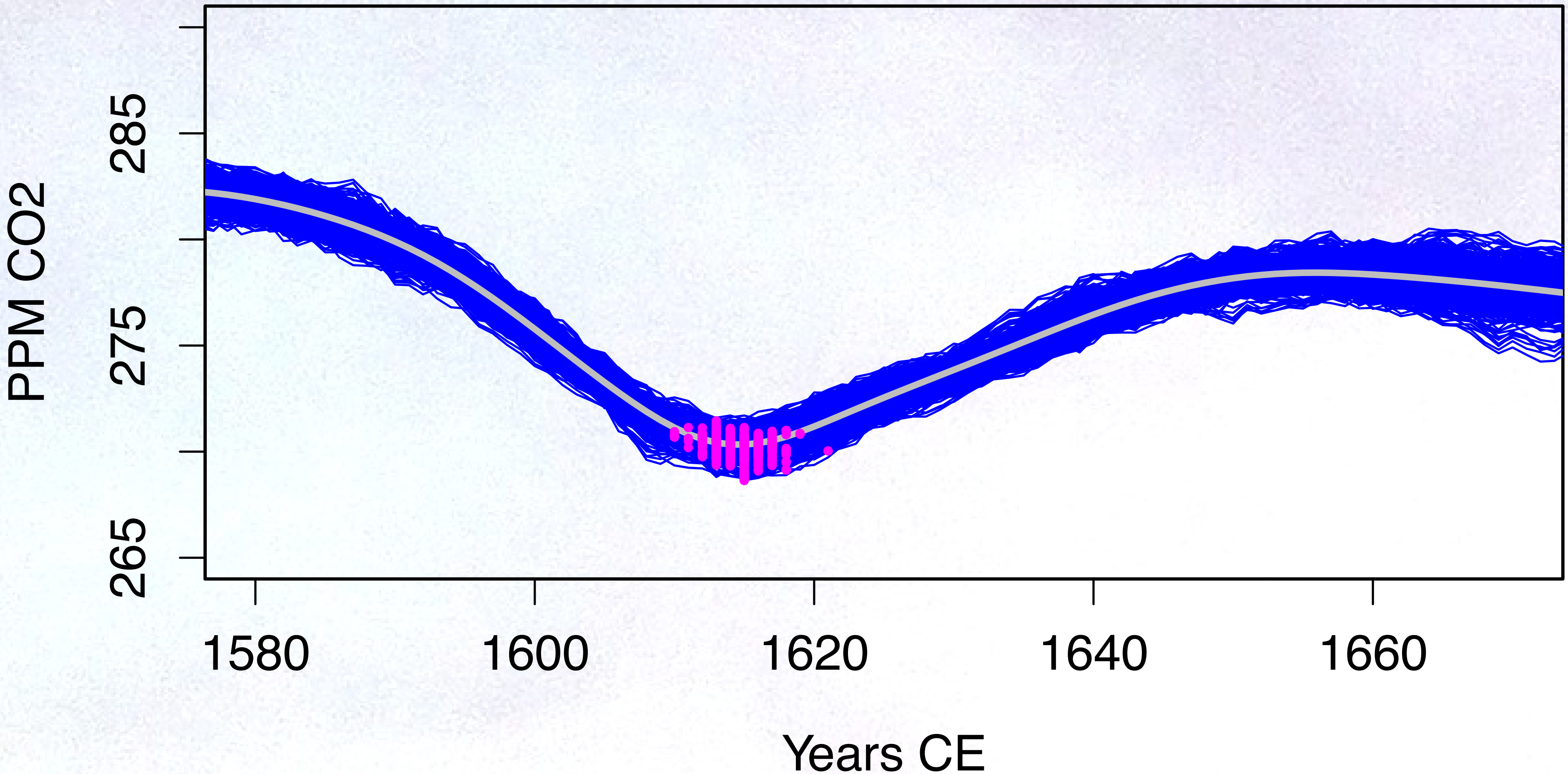
Ensembles

500 draws from the posterior density , mean



Dip around 1610

Based on running decadal mean of concentrations



Points indicate uncertainty about size and timing of minimum.

Bending gridded data products

For larger spatial data sets it is useful to switch to a basis/coefficient representation for the process such as LatticeKrig.

Process:

$$g(x) = \sum_j \phi_j(x) c_j, \quad \boldsymbol{c} \sim MN(0, \rho Q^{-1})$$

Data Model:

$$Y_i = L_i(g) + e_i$$

Snow water equivalent (SWE)

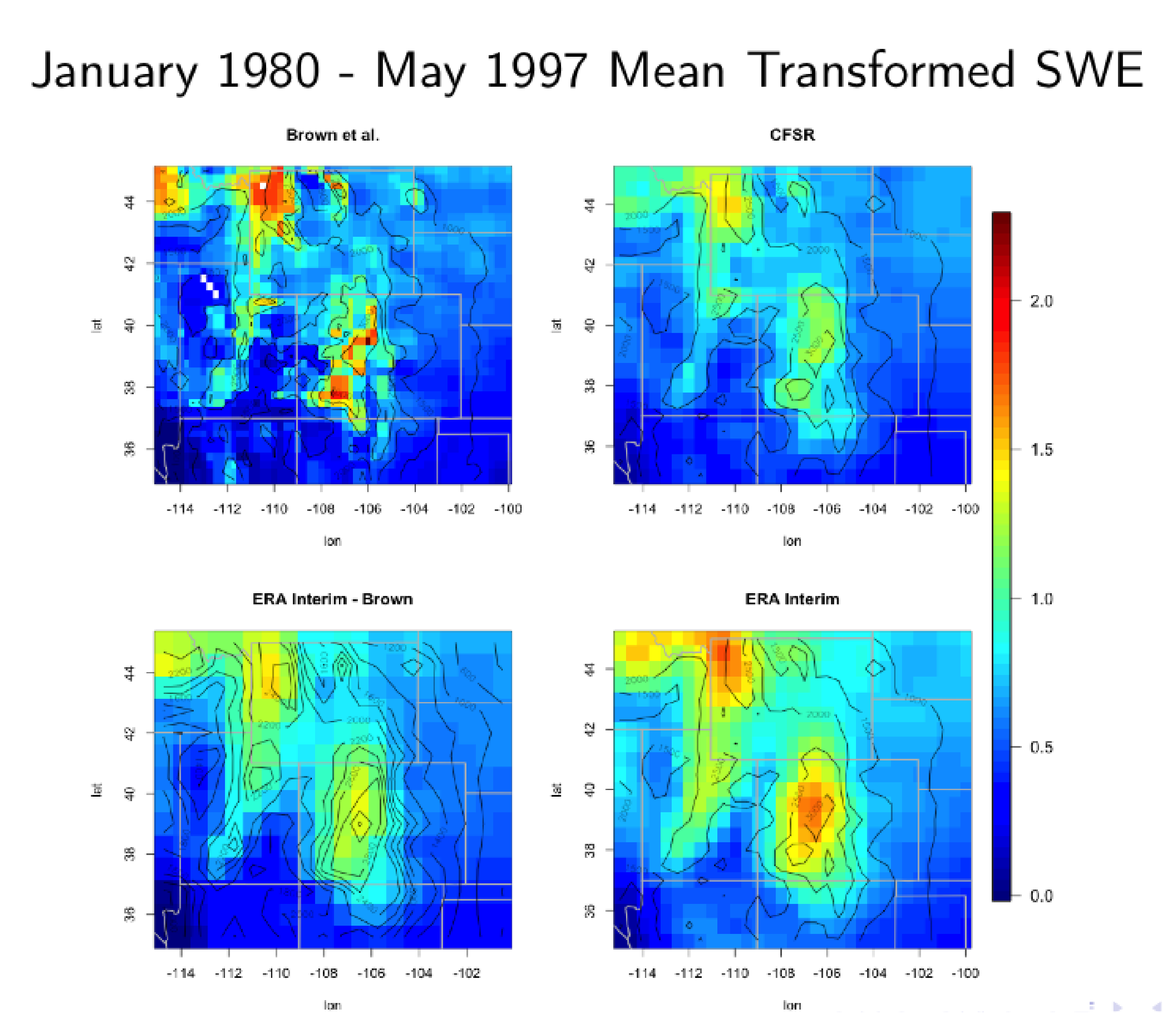
(Colette Smirniotis, Rachel McCrary)

- Observations are four different SWE data products
monthly snow accumulation for the Rocky Mountain Region at different resolutions and grids.

- $L_i(g) = \boldsymbol{X}_i^T \boldsymbol{c}$ an integral or pointwise evaluation that maps into the grid boxes for the different data products.

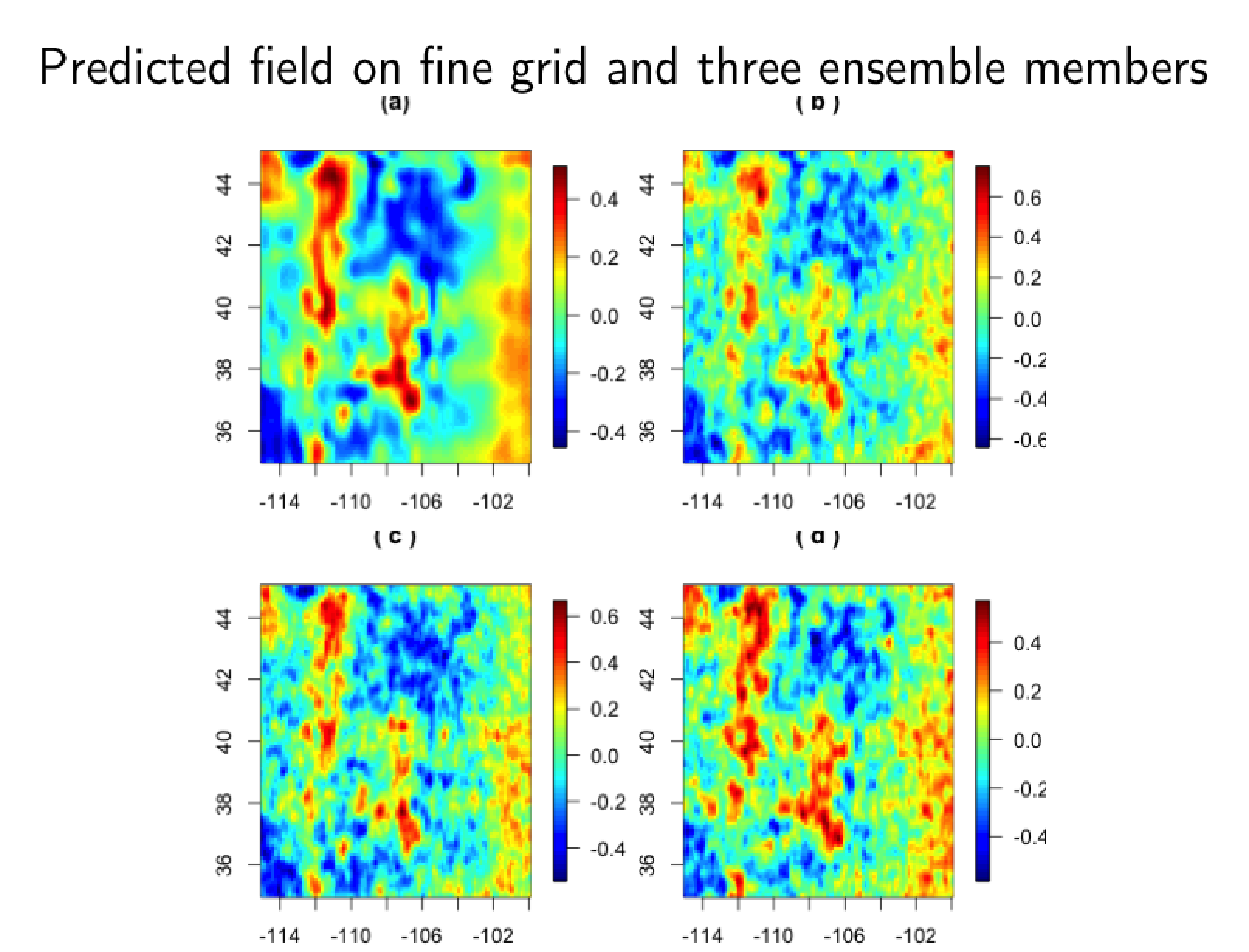
- Goal is a common SWE field (g) with uncertainty estimates.

The "observations"



Blended result

Estimate using LatticKrig and three draws from a conditional simulation.



Electron density in the solar corona

(Kevin Delmasse, Sarah Gibson and Luke Burnett)

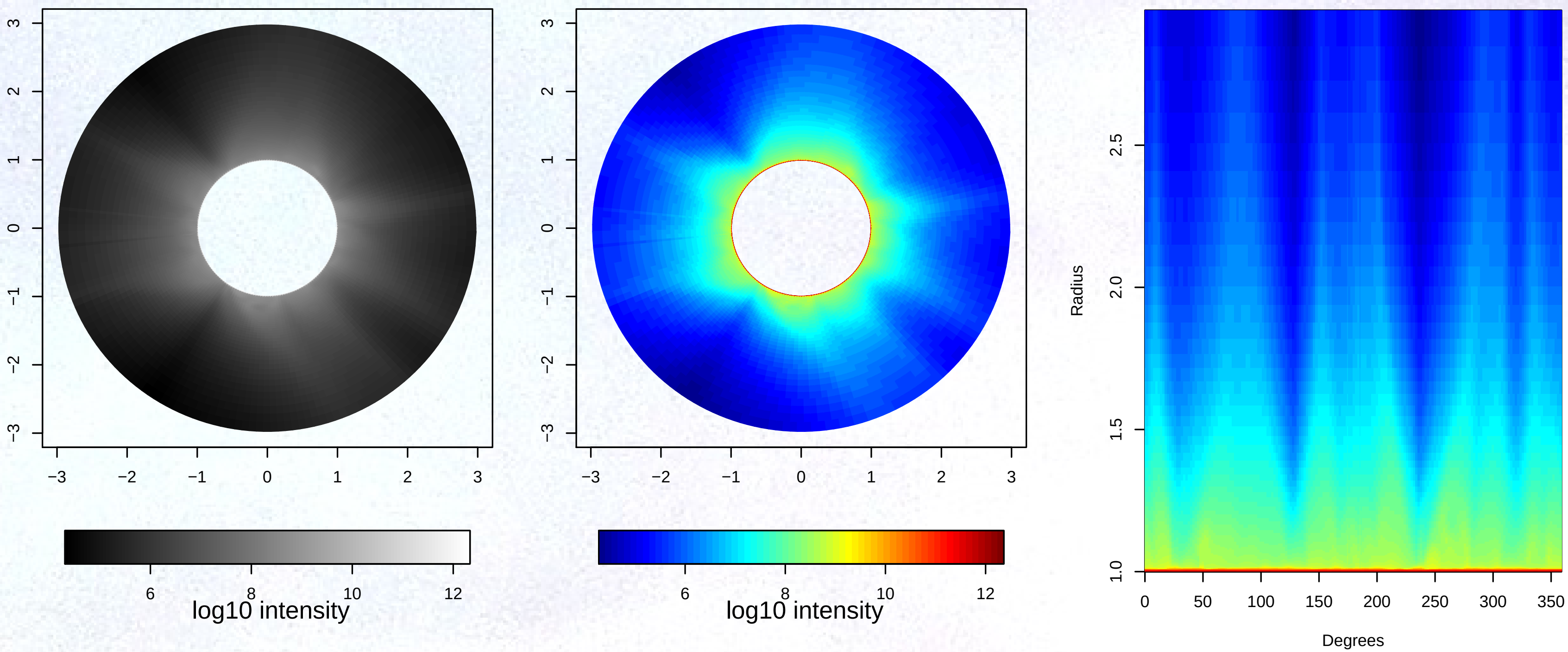
- Observations are integrals through corona.

- Goal is reconstruction of the density based on different viewing angles.

- This is a pilot project for the 3-d reconstruction and for magnetic fields.

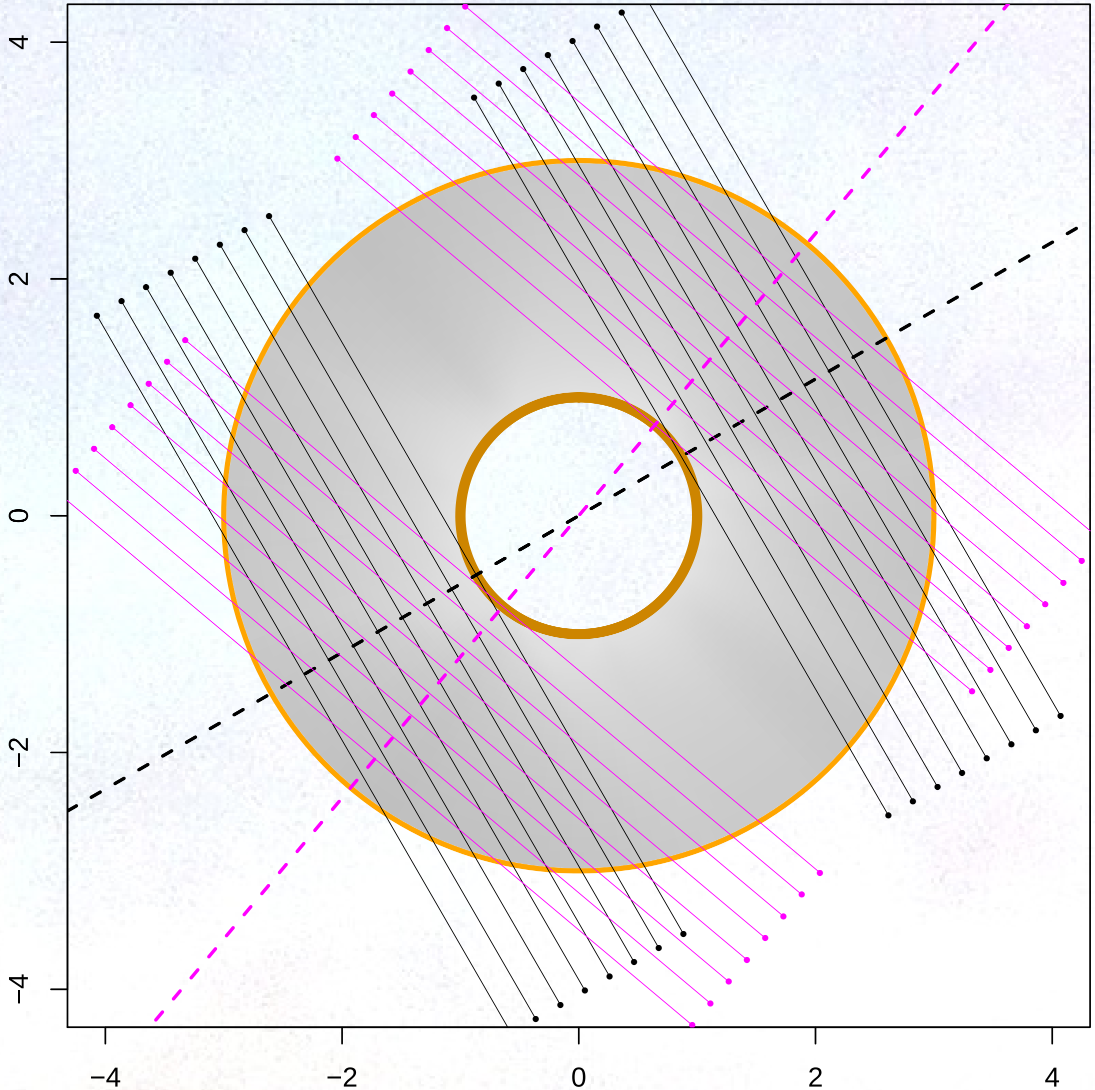
- Observational functionals are line integrals based on the viewers line of sight.

Equitorial slice for electron density
(*Predictive Science product* time = 2144^{th} Carrington rotation)



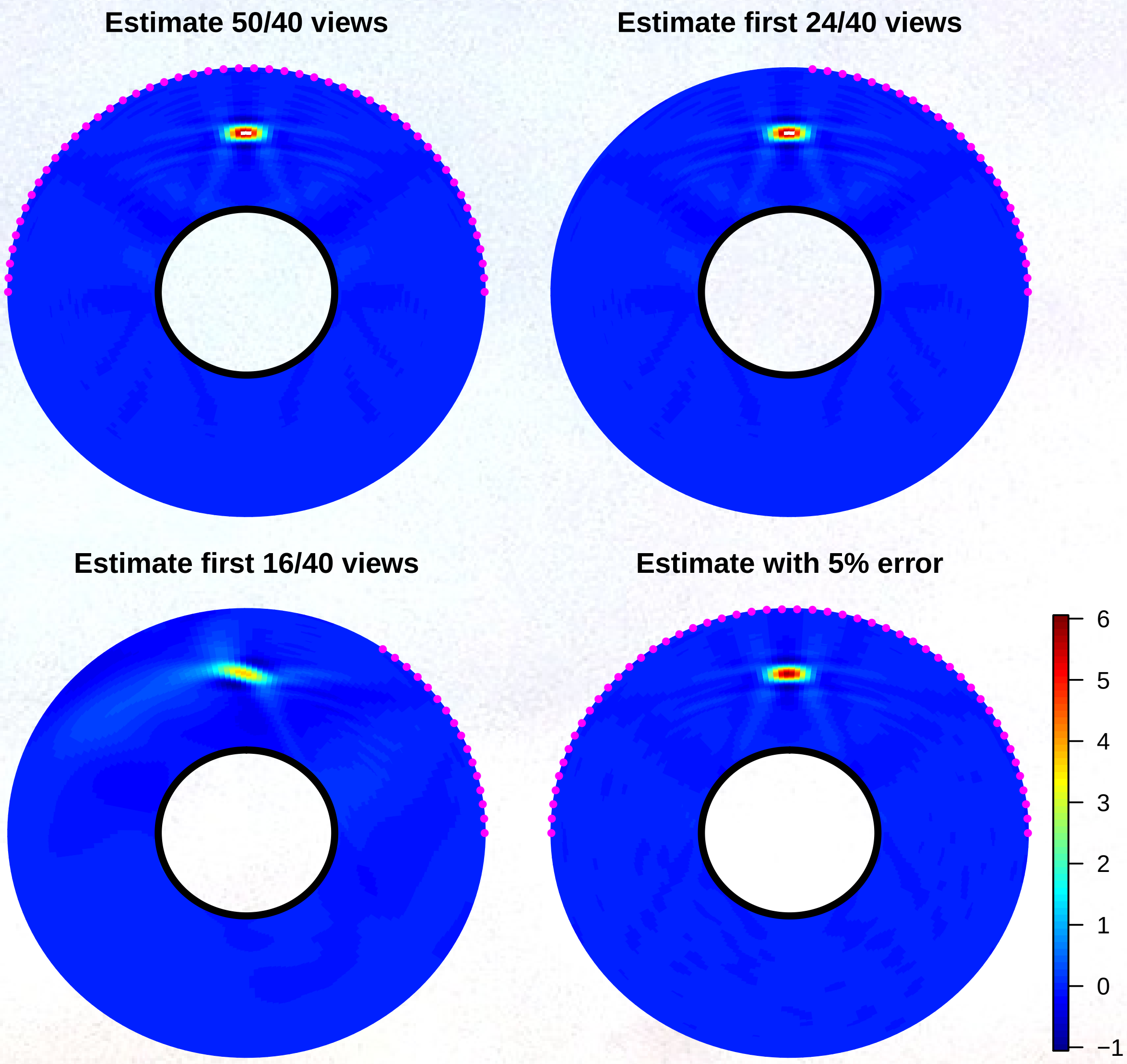
Observations of the Solar Corona

Two viewing angles each with 16 lines of sight: (2/16)



Reconstructions of simple phantom

LatticeKrig with $\sim$ 5000 basis functions,
50 angles with 40 lines of sight each.



Summary

- Break problem into observations and process models.

- Use spatial statistical models to find estimates of the function or field

- Basis function models help for large data sets.

Thank you!

Questions?