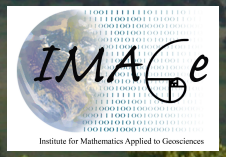
# Regional climate models, spatial data and extremes

Douglas Nychka

National Center for Atmospheric Research

*Stephan Sain , Cari Kaufman (UC-Berkeley)*




*National Science Foundation*

*EURANDOM, May 2009*

# Outline

- North American Regional Climate Change Program
- Logspline density estimates
- Functional and spatial data.
- Regional climate simulations under current climate.

**Challenges:**

Spatial and functional data, design and analysis of computer experiments, computational statistics for large problems.

# Probability of extreme events.

## Focus on the log density

Given observations $\{y_i\}$ from $f(y)$, a probability density function.

$$f(y) = e^{g(y)} \quad \text{or} \quad g(y) = log(f(y))$$

we are interested in the (simple) behavior of $g$ when $p$ is large. For example, $g$ could be linear for large $y$

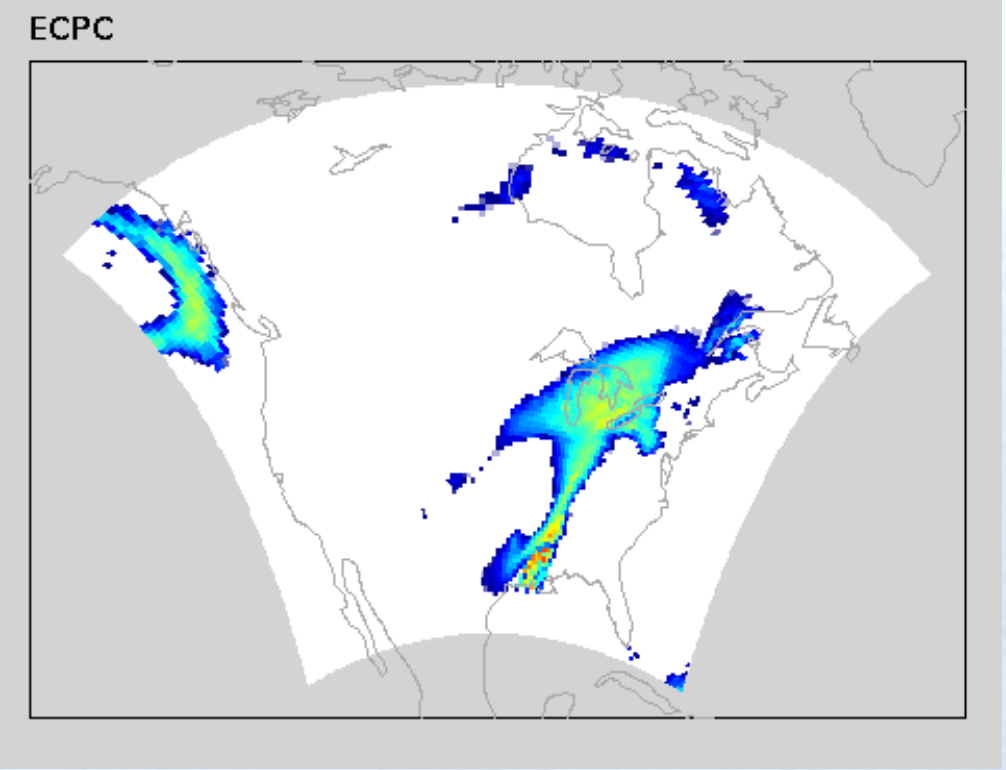
## Quantify how $g$ changes over space and for different cases

i.e. $g = g(p, \boldsymbol{x}, M)$ with $\boldsymbol{x}$ being a location and $M$ being a climate model.

# Regional Climate Simulations

**Geophysical models simulate weather under different climate states.**

Snapshot of 3 hour precipitation for Exp. Climate Prediction Center Regional forced by observations.



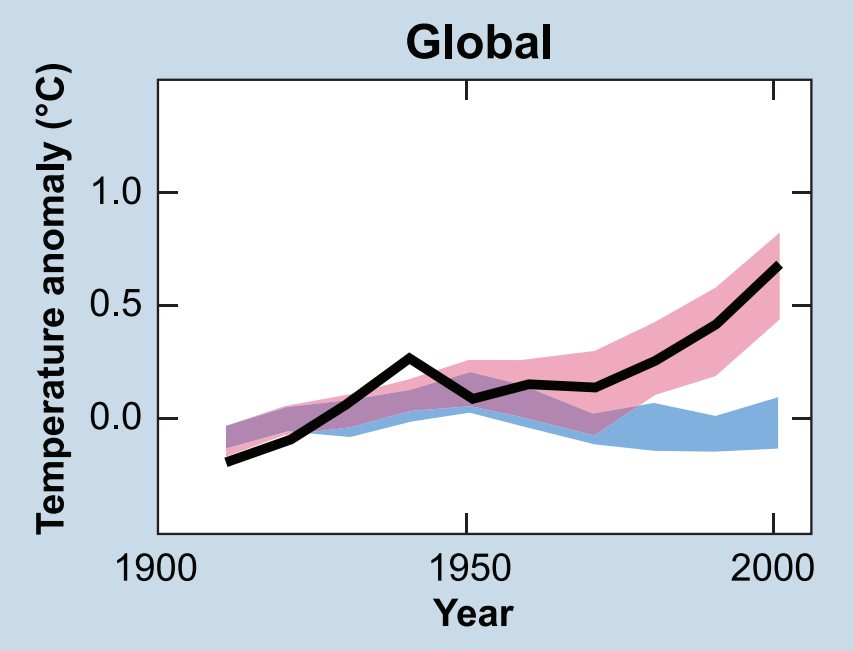
# Inferring changes in extreme weather

**Basic strategy** Use numerical simulations, such as *regional* climate models to determine the distribution of weather under different *global* conditions.

e.g. Change the amount of greenhouse gases in the atmosphere over time and simulate possible changes in climate.

**Numerical Experiments** Need to quantify the sources of uncertainty for different geophysical models and for different regions of the Earth.

# Global climate change experiments



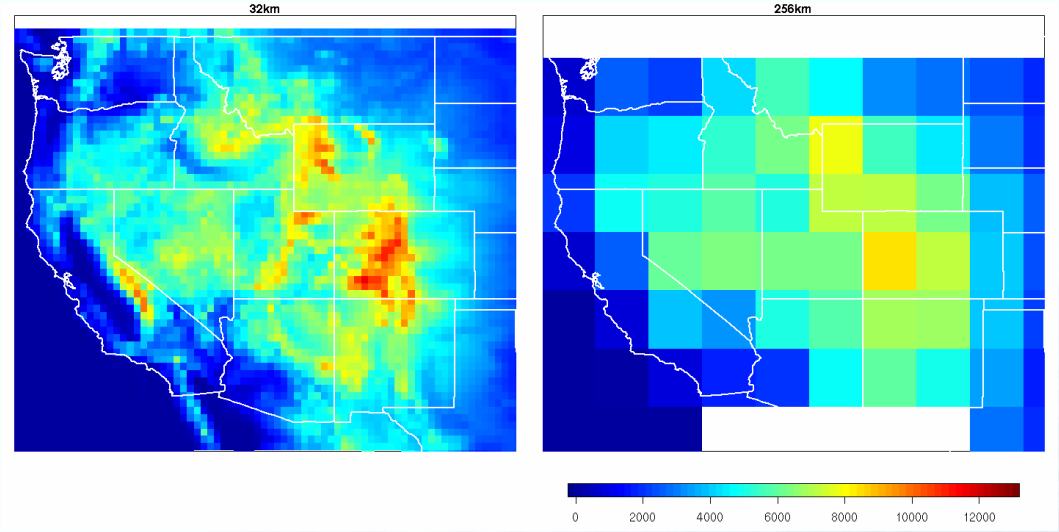

– observations
– with human greenhouse gases
– without

Summary figure from *Intergovernmental Panel on Climate Change, Fourth Assessement Report.* Used as evidence for attributing global warming to human activities.

# The problem of regional climate.

*The global models on their own do not give enough detailed information at regional and local scales.*

32km and 256km grid boxes for elevation



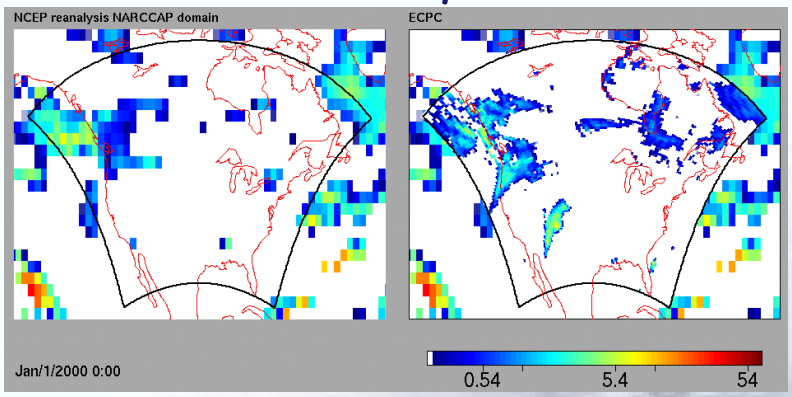

Many physical processes and features are not modeled explicitly!
E.g. thunderstorms or extreme weather events.

# NARCCAP

*4GCMS × 6RCMs: 12 runs as a balanced design*

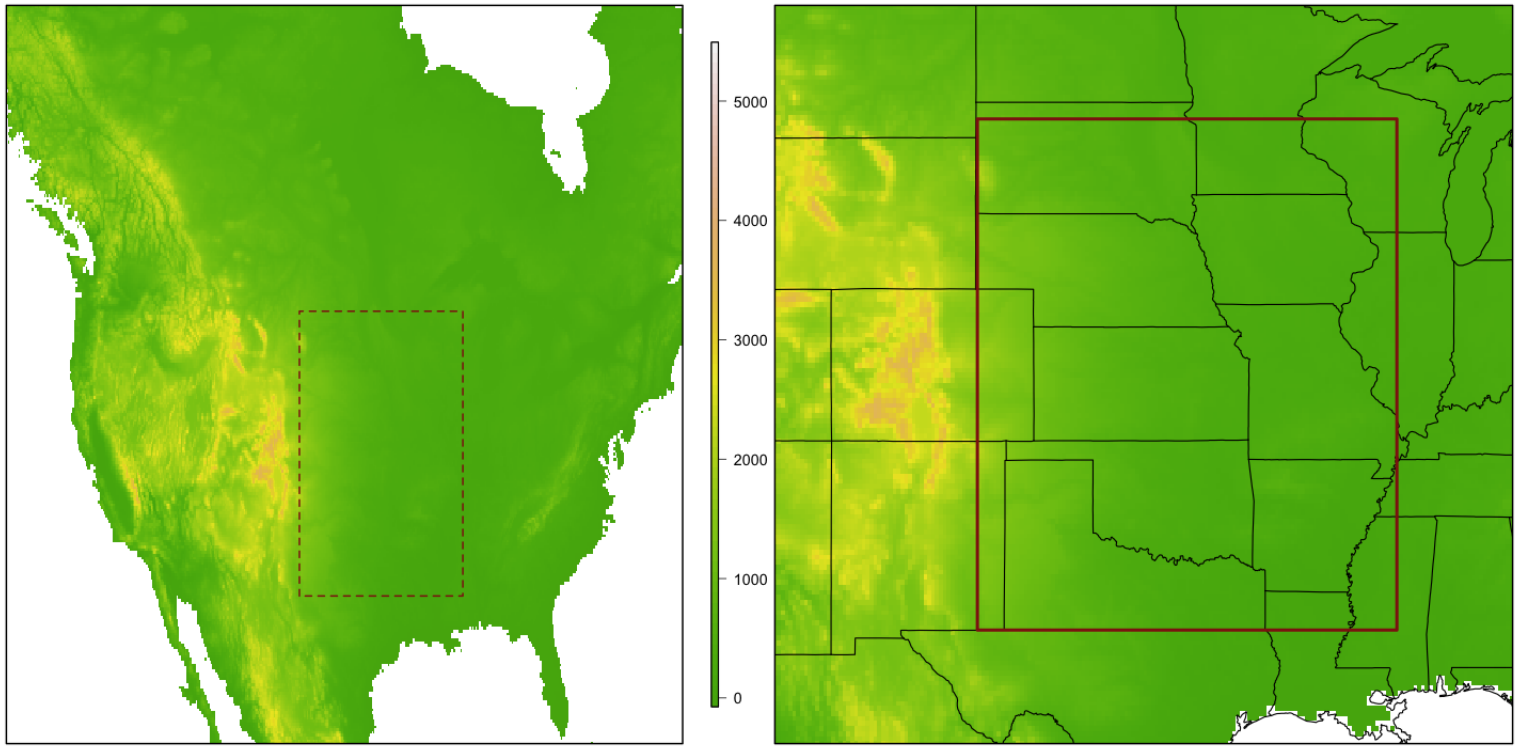
| | MM5I | RegCM | CRCM | HADRM | RSM3 | WRFP |
|---|---|---|---|---|---|---|
| GFDL | | ● | | ● | ● | |
| CGCM3 | | ● | ● | | | ● |
| HADCM3 | ● | | ● | ● | | |
| CCSM | ● | | | | ● | ● |
| Obs. (NCEP) | ● | ● | ● | ● | ● | ● |

*Surface Precipitation.*



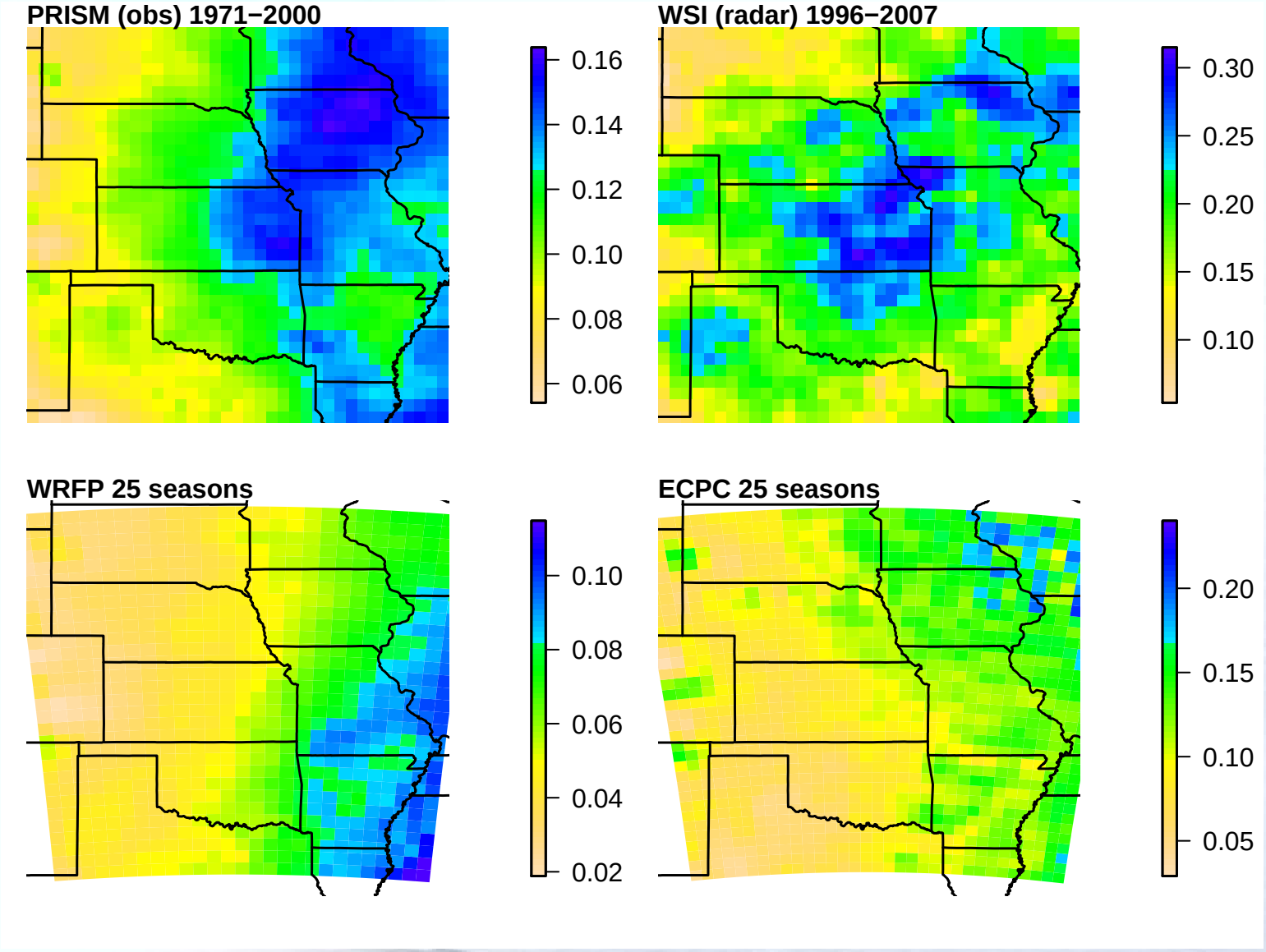

ECPC model forced by NCEP reanalysis

# Study region (for this talk)



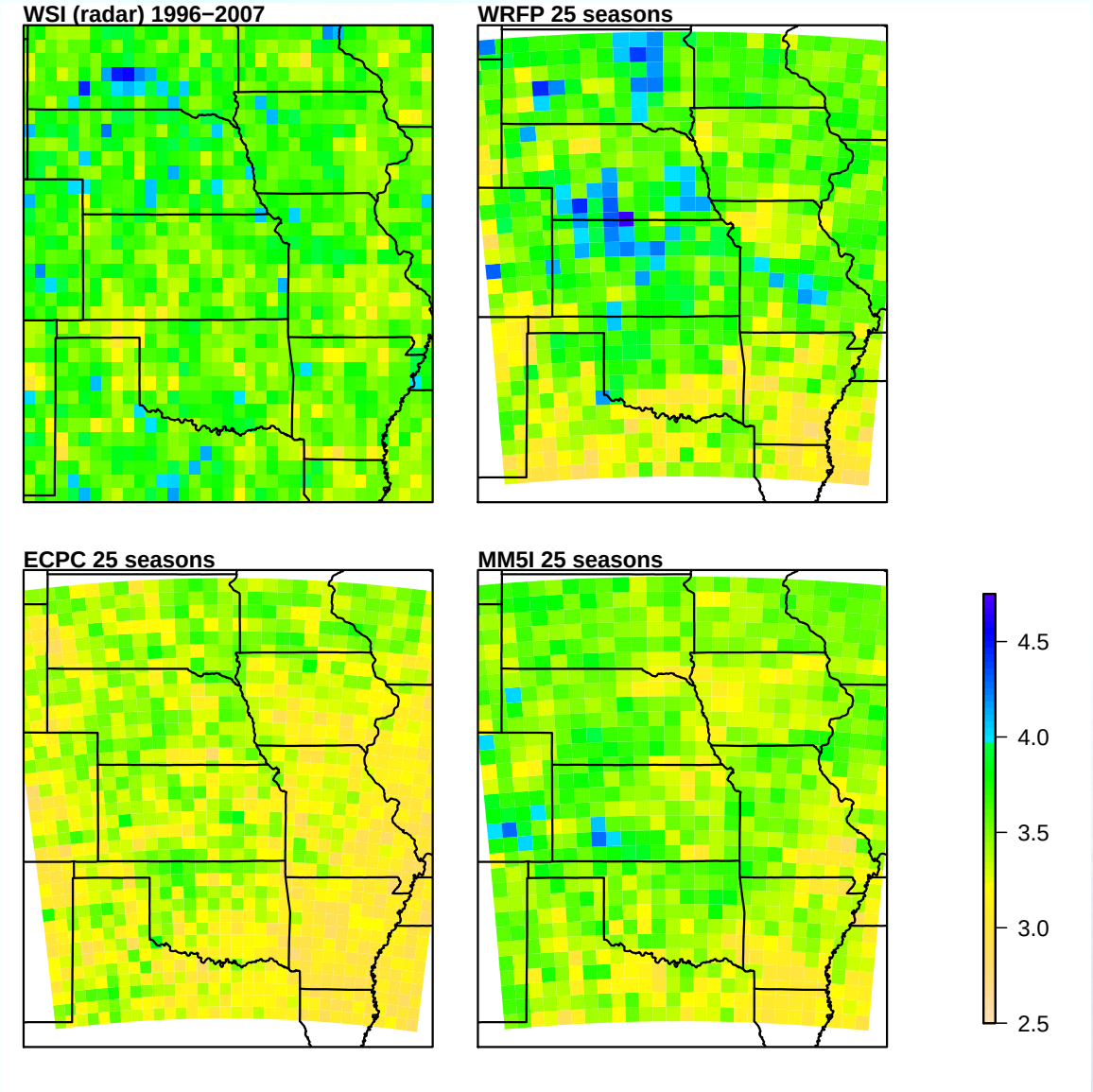

Includes $\approx$ 800 grid points from RCM simulations

# Mean summer precip (mm/hour)

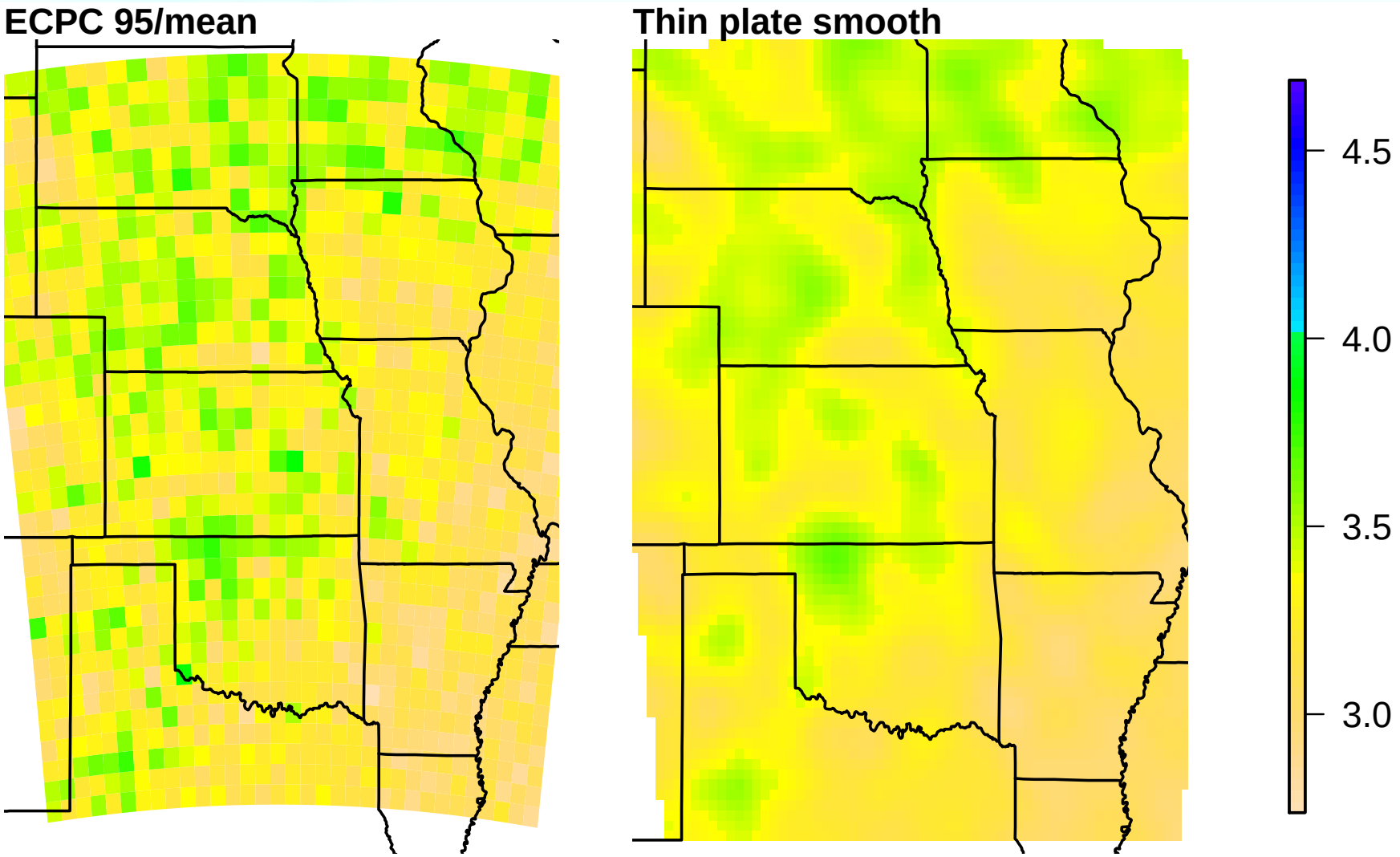



Maybe we should just stop here!

# 95% quantiles/ mean



# Spatial smoothing



# Summarizing the distribution

## log spline estimates

Stone, Hansen, Kooperberg, Truong (1997)

$$f(\widehat{y}) = \exp\{B_1(y)\beta_1 + B_2(y)\beta_2 + \ldots + B_2(y)\beta_2 + C(\boldsymbol{\beta})\}$$

Basis will be a cubic polynomial in between the $J$ "knots".

$$Y_1 < Y_2 < \ldots < Y_J$$

and linear outside of $Y_1$ and $Y_J$

*$\widehat{f}$ will have exponential tails.*

# Estimating parameters and knots

**When knots are known:** Estimate the parameters $\{\beta\}$ by maximum likelihood.

**Forward knot selection:** Starting with a minimum number of knots placed according to order statistics.

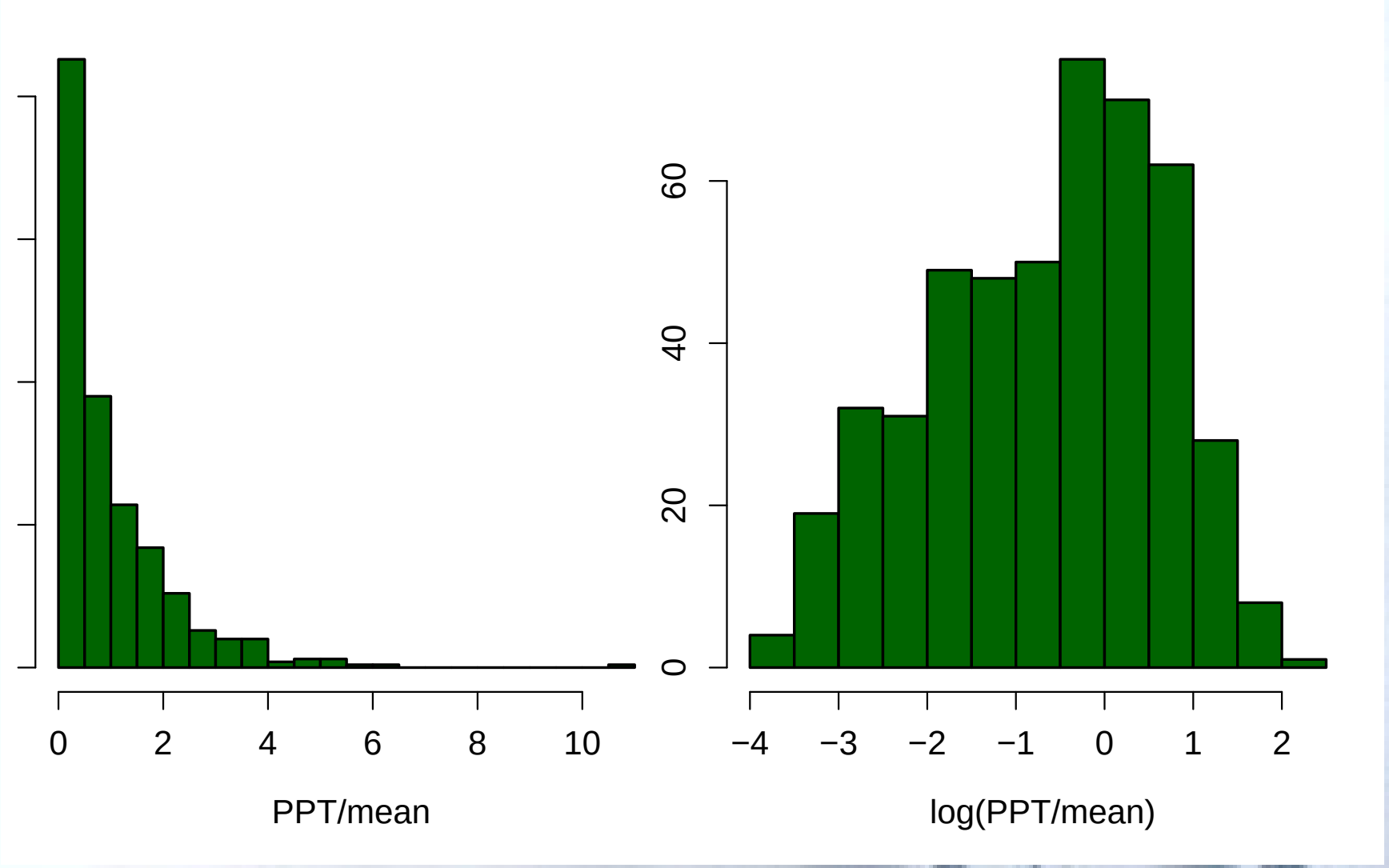
Add knots sequentially by maximizing a likelihood test statistic (Rao)

**Backward knot deletion:** Remove knots sequentially using a likelihood test statistic (Wald)

**Information criterion:** Select among all the sequences of models generated from forward and backward selection.

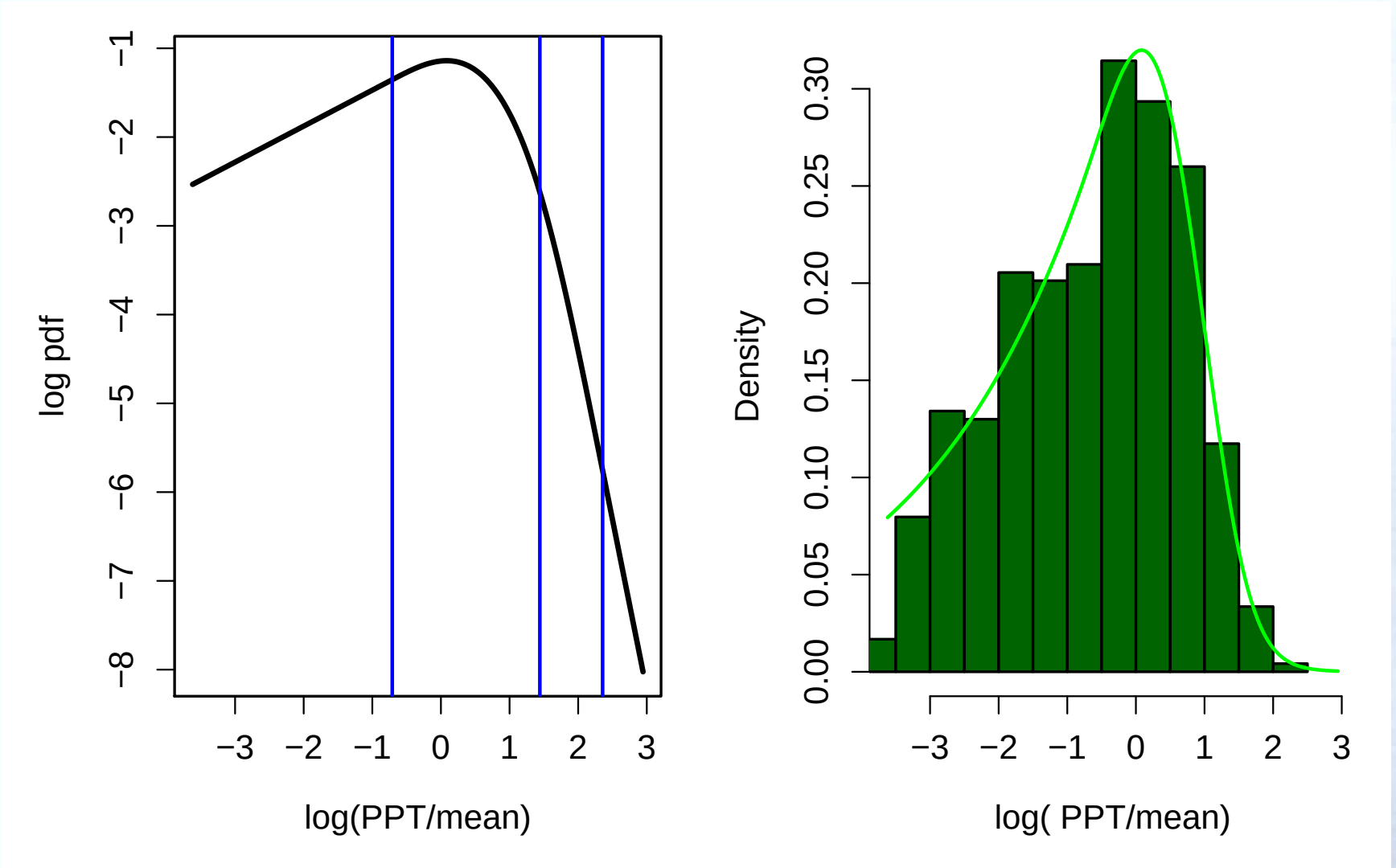
# An example

Daily precipitation from a grid box of ECPC.
Values relative to the mean and the logged transformation.



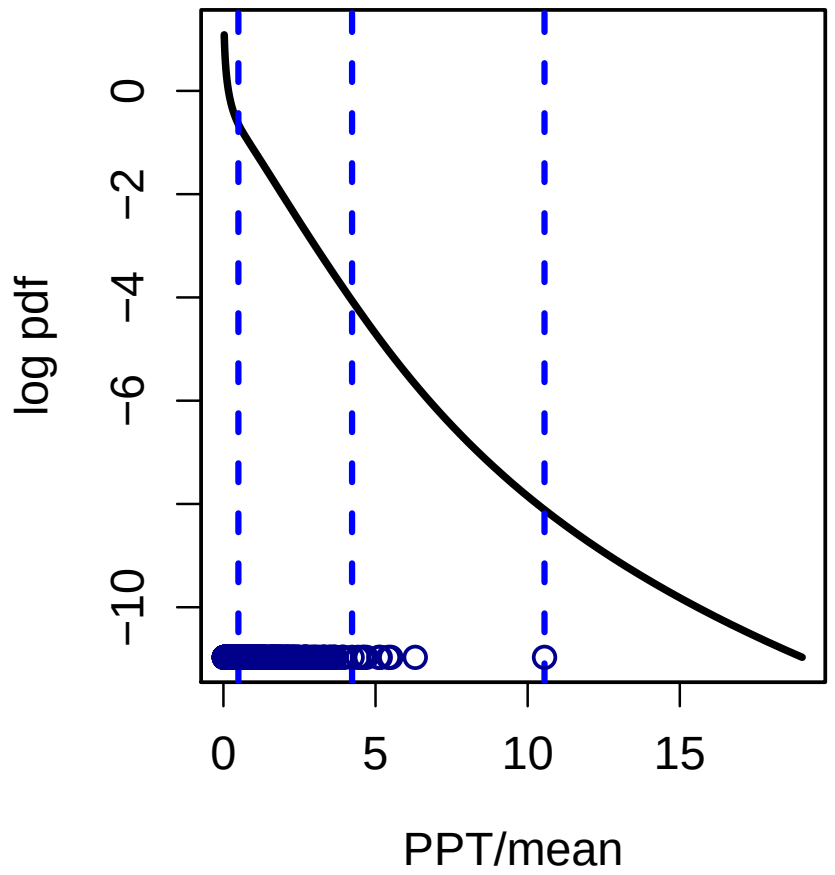
# Logspline fit

Fit having logged the precipitation values

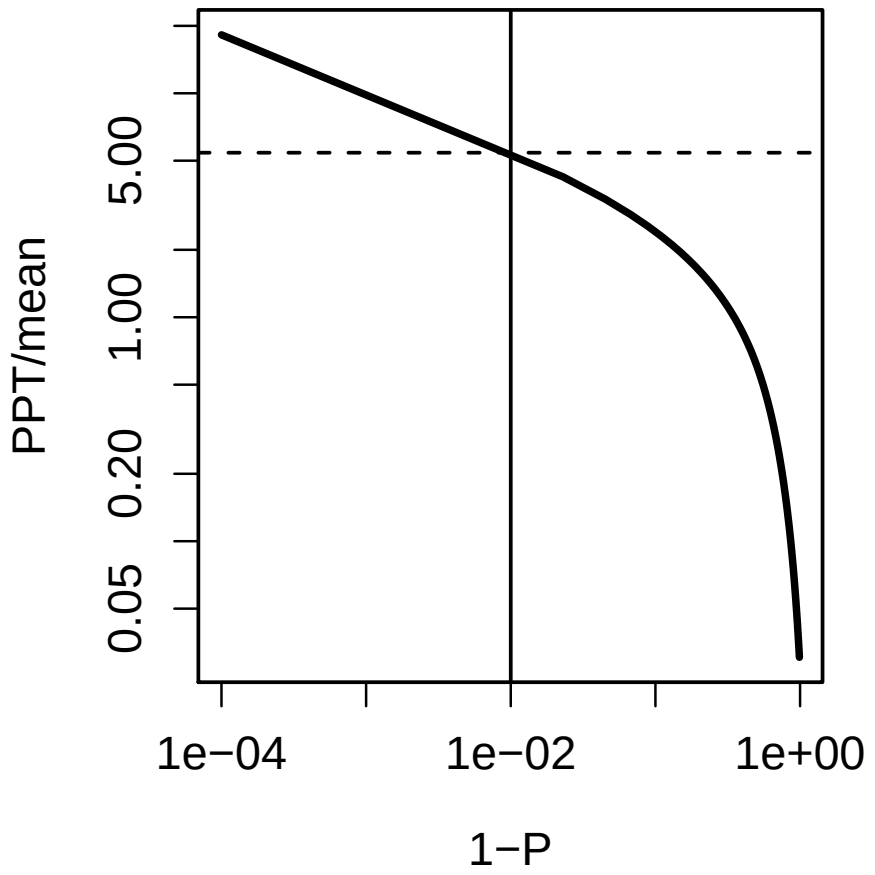


# Fit in unlogged scale

Log density




estimated tail probabilities.



# Building a distribution from EOFs

Approximate the log quantile function as a linear combination of strategic basis functions.

$$log[q(p)] = \sum_{i=1}^{M} \phi_j(p) u_j$$

The basis functions and coefficients are found by EOF/PC analysis.

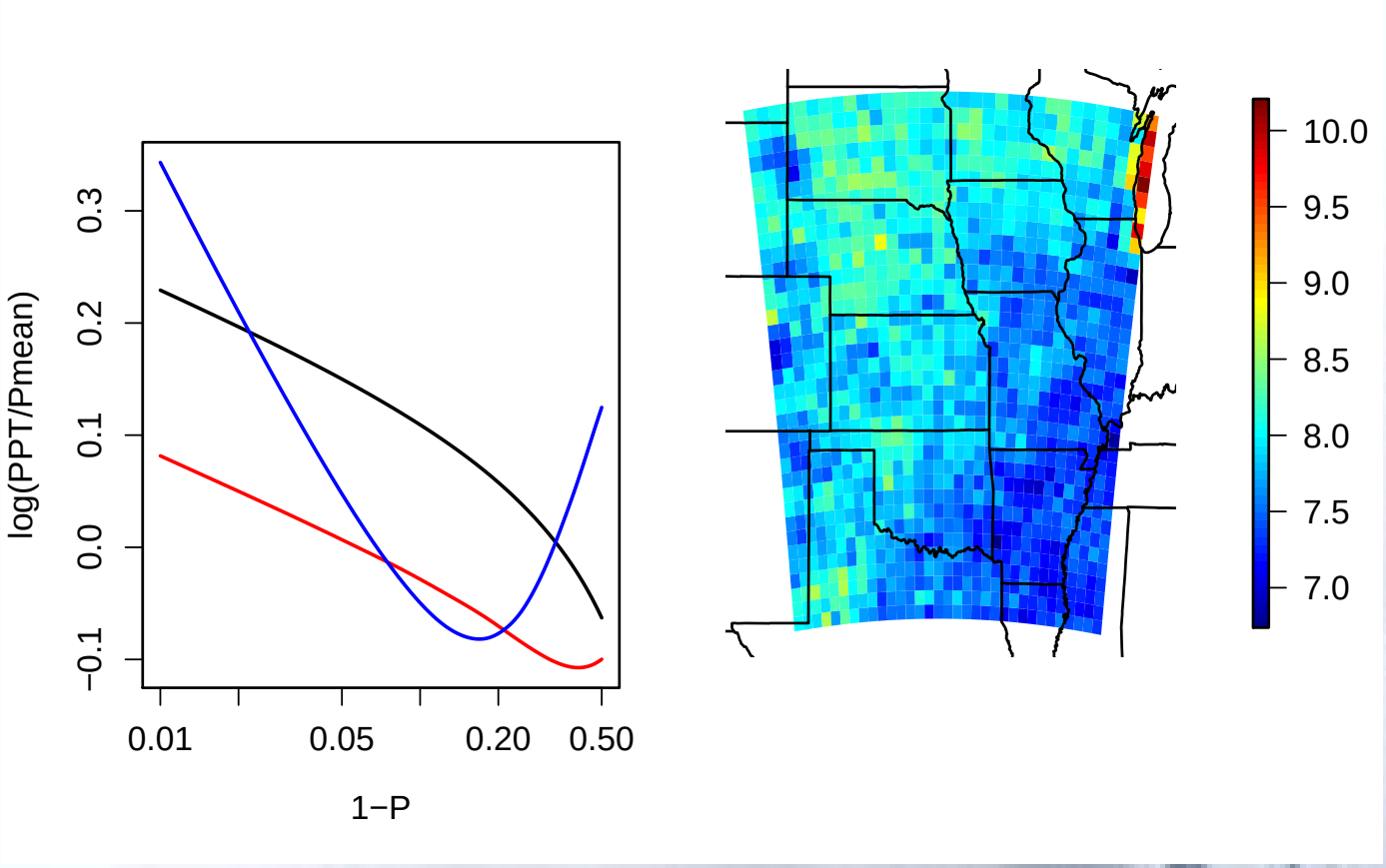
- There is a set of coefficients for every grid box of every model.
- Log form means basis functions have multiplicative effects.
- Working with quantiles simplifies finding a common range.

# EOFs for the ECPC model

**EOF analysis of the log quantile function for ECPC.**

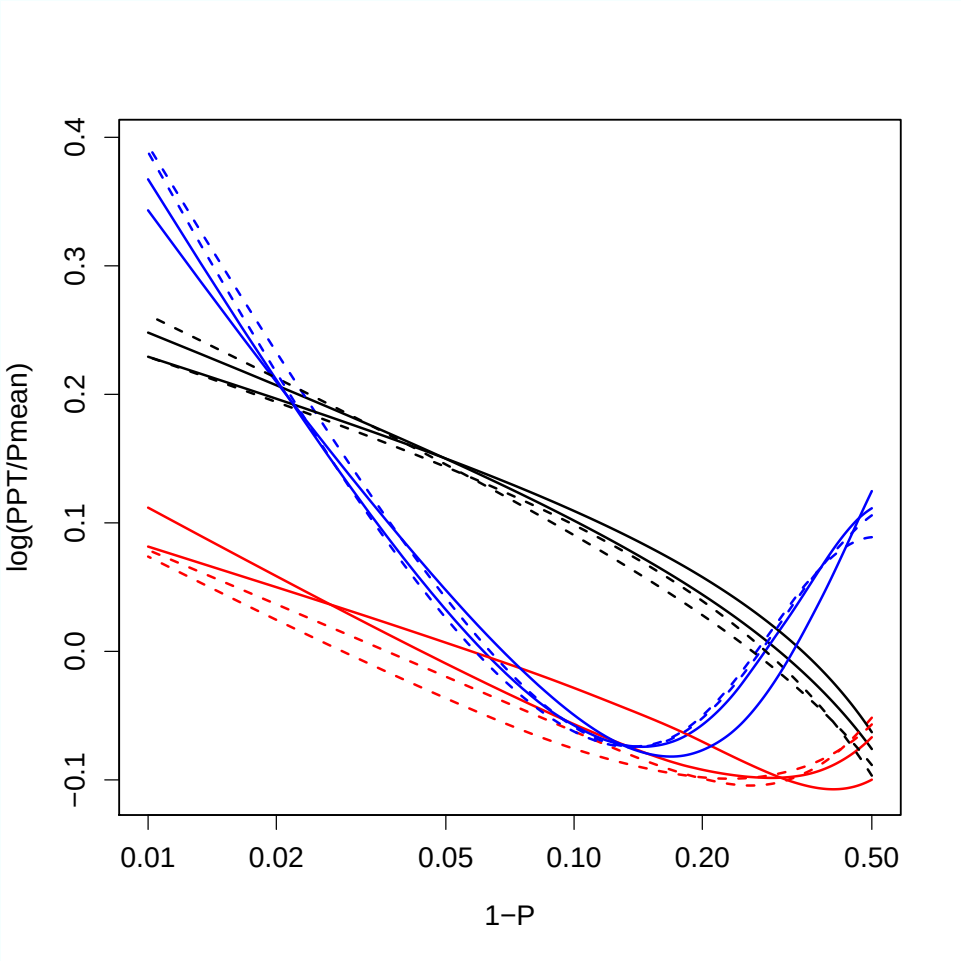
3 leading EOFs ( 1, 2, 3) coefficients for leading EOF.

Singular values: $265.0, 21.4, 9.6$

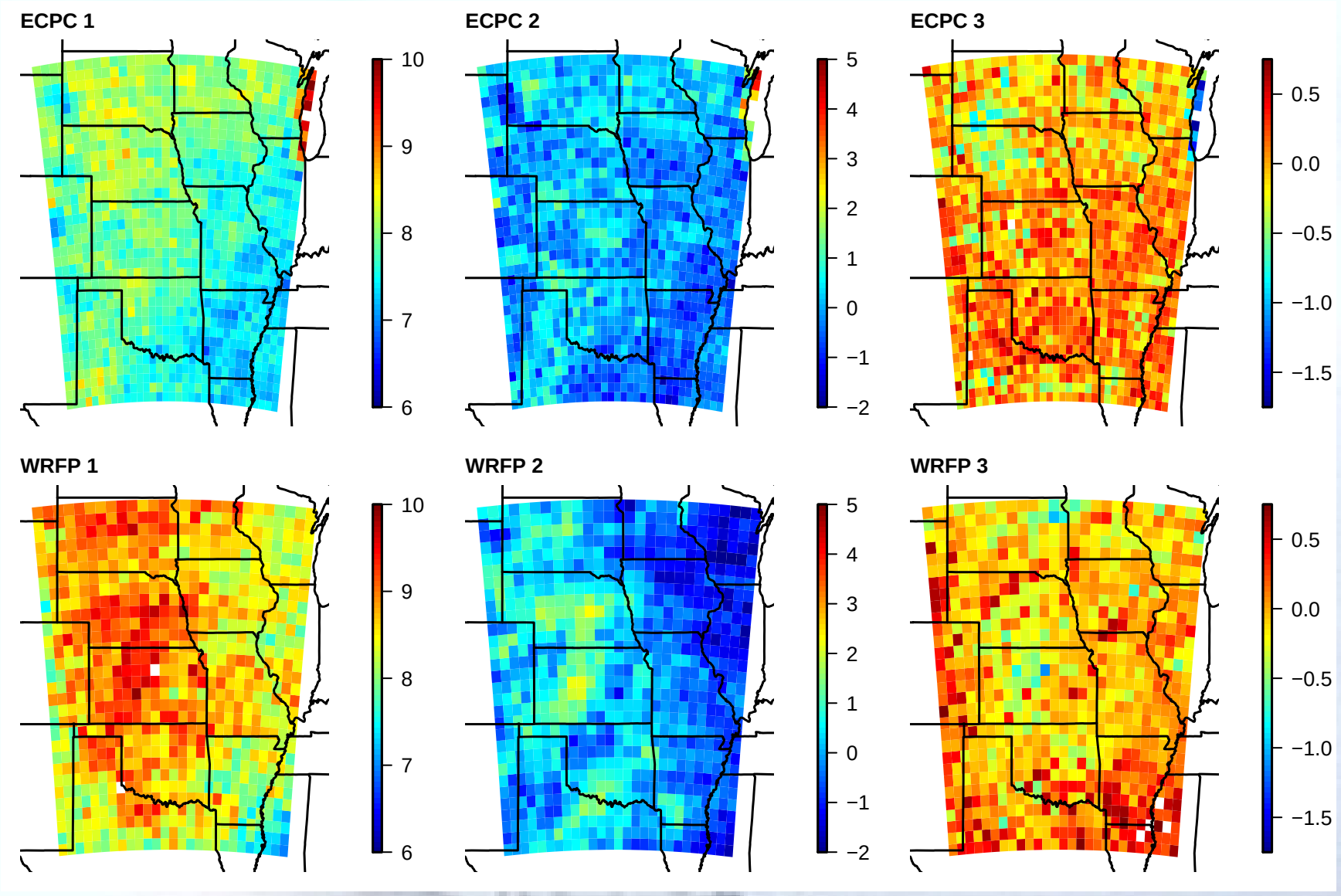


# Comparing basis functions

Comparing leading EOFs of 4 RCMs: ECPC, MM5I, WRFP, RCM3



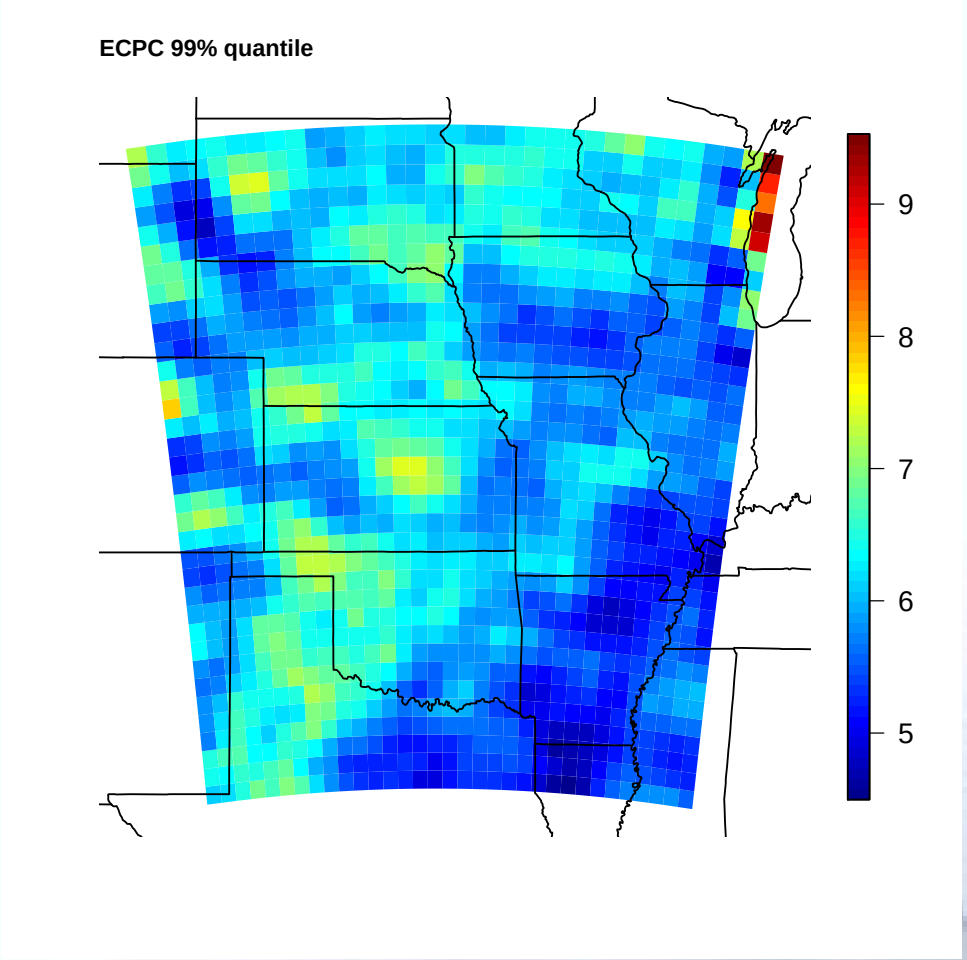
# Coefficients for ECPC and WRFP



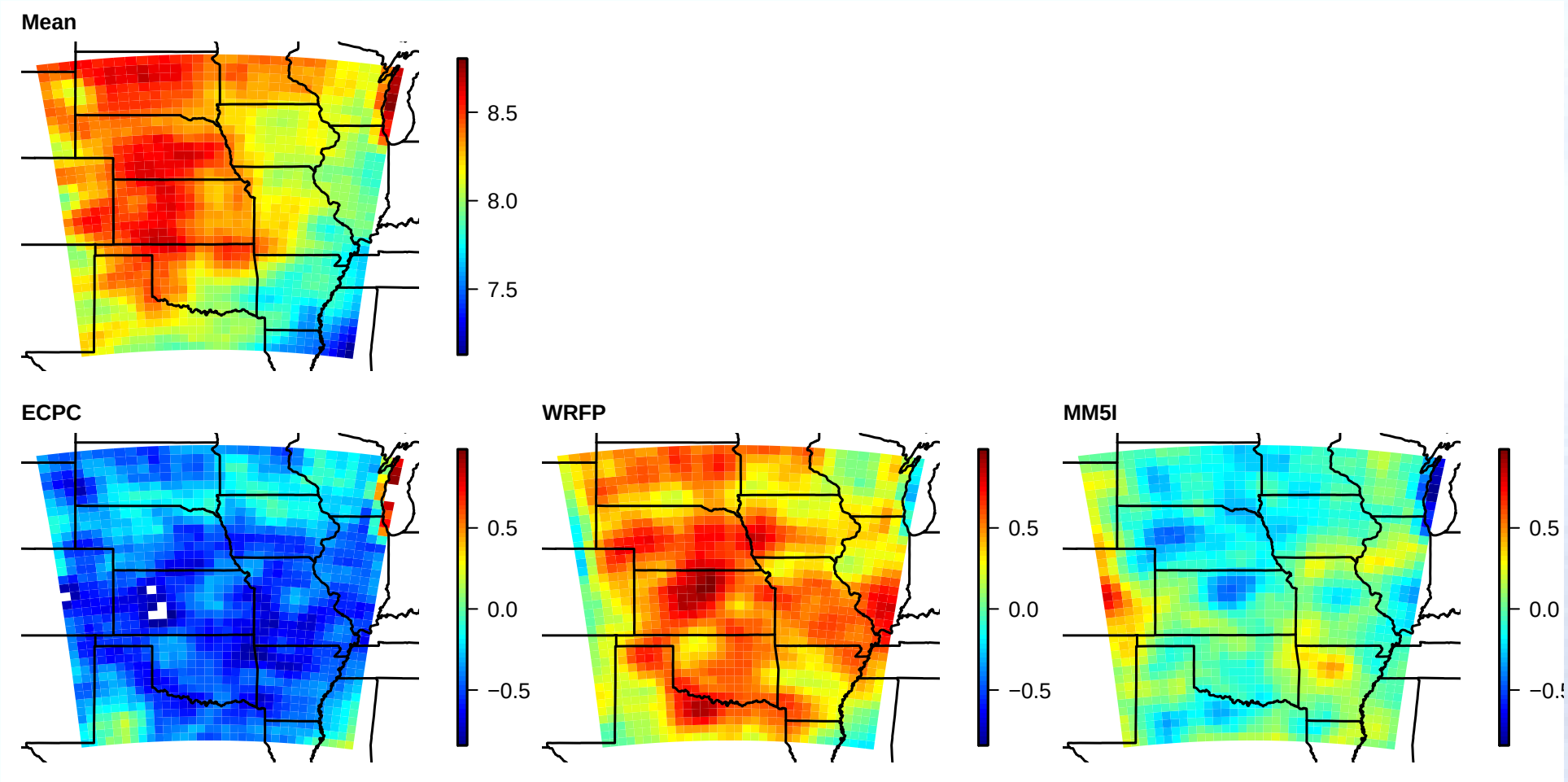
# Towards a hierarchical model ...

## 99% /mean estimate for ECPC

Use Tps to smooth three fields of coefficients
then reconstruct quantiles with 3 EOFs.



# An ANOVA summary of first EOF



# Summary

*Regional climate experiments have the potential to provide more detailed information about changes in extremes for future climate*

*logspline density estimates are a flexible method for fitting large data sets*

*EOFs are useful in reducing the dimension for comparison across space and across models*

*Some challenges are to add uncertainties bounds on these estimates – Bayesian methods are the easiest way to do this.*

# Thank you!