(Ladies and) Gentleman: there is lots of room left in Hilbert Space.

Douglas Nychka

National Center for Atmospheric Research

Dan Cooley, Colorado State Univ.

National Science Foundation *UW Madison June 2010*

Outline

- *Reproducing kernels*
- *Climate Change*
- *Extremes in precipitation.*

Gentleman: There is lots of room left in Hilbert Space.

Saunders Maclane

Quoted in Reed and Simon, *Functional Analysis*

All Hilbert spaces are wrong, but some are useful.

The basic problem and our task

Scientist: I have some irregular observations of a curve.

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

Scientist: What can you tell me about g?

Nychka: I don't know. What can *you* tell me about g?

Scientist: Well, g is not very wiggly and values of g close to each other seem to be similar.

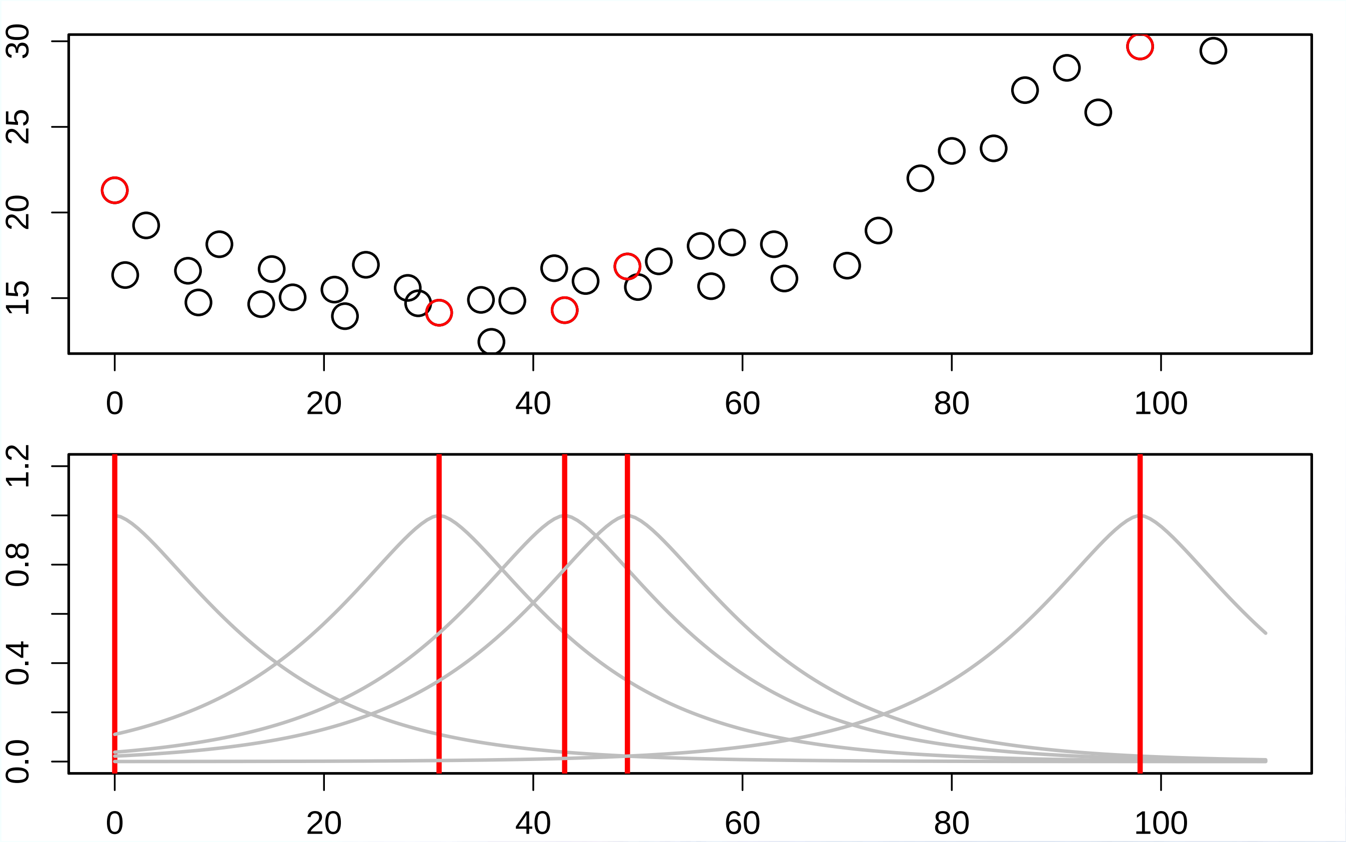
Some of my colleagues think g is almost linear. But it could be almost any shape – can you really fit a function like that?

Nychka: There's lots of room left in Hilbert space.

A great idea for an estimator

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

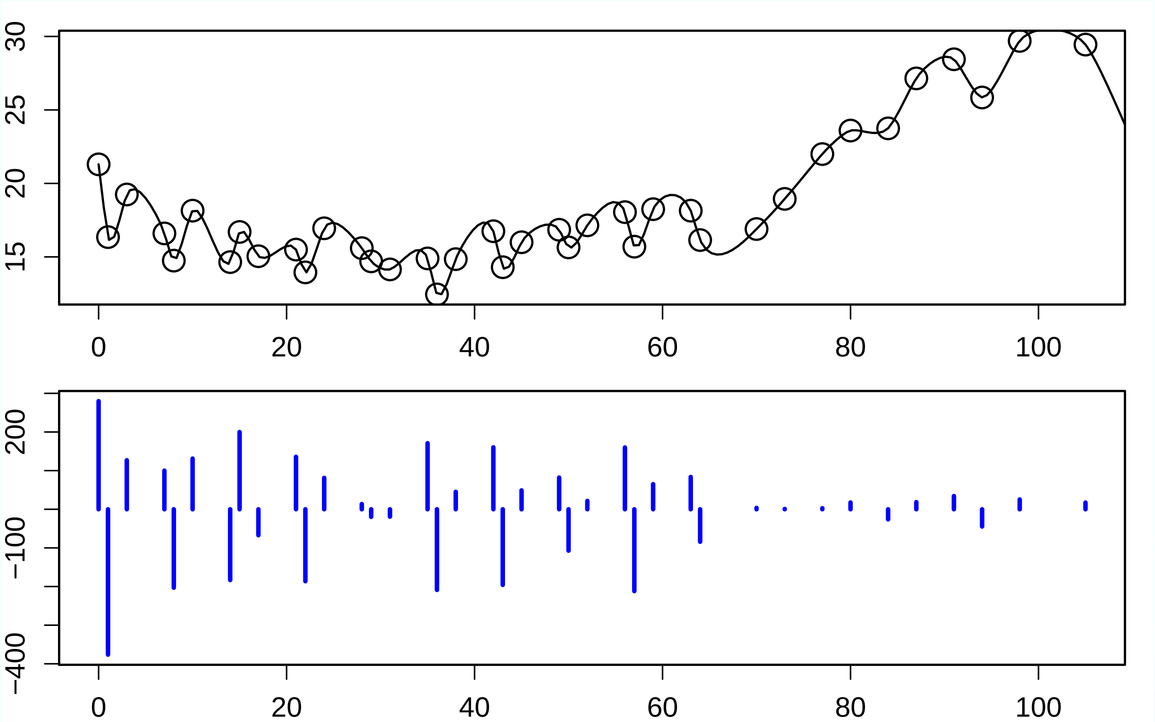
Put a bump at each location and blend them together to match the data.



The blending bit

$$\hat{g}(x) = c_1\phi(x - x_1) + c_2\phi(x - x_2) + ... + c_n\phi(x - x_n)$$

Find $\boldsymbol{c}$ by solving a linear set of equations



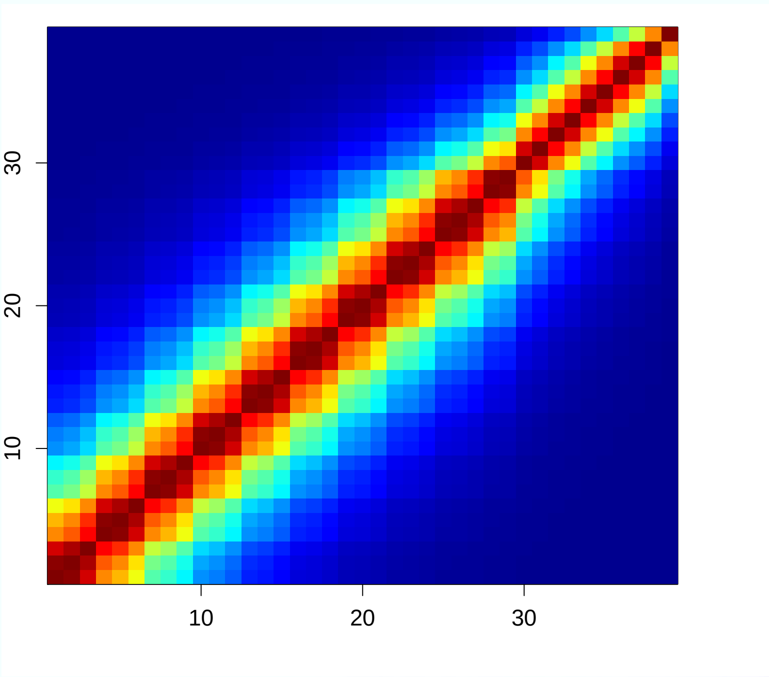
The c coefficients

With some measurement error or a smoothed estimate the c are proportional to the residuals!

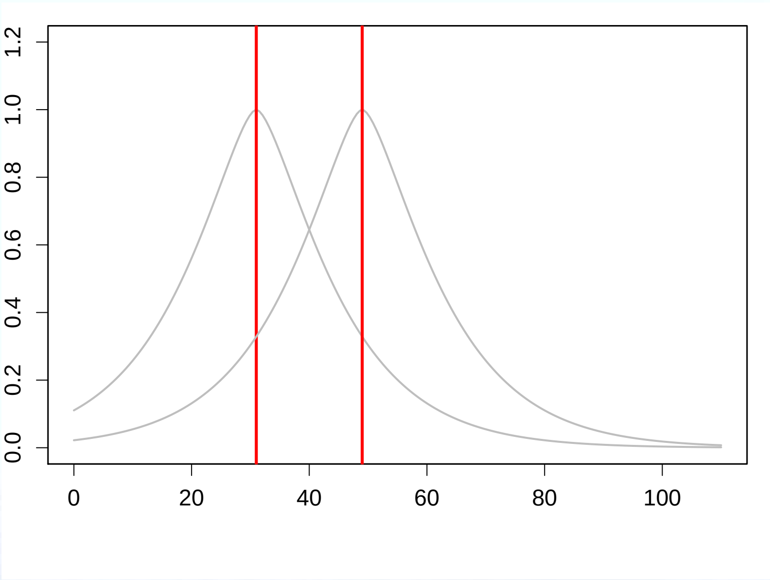
ϕ seems pretty important here

How about choosing ϕ so that the linear system is invertible e.g. $\phi(x_i - x_j)$ is positive definite.

The matrix $\phi(x_i - x_j)$



The "angle" between basis functions is the intersection with the nodes.



Bootstrapping our way to Hilbert space

If $k(x, x') = \phi(x - x')$ *is positive definite*

$\Rightarrow$ k *is a reproducing kernel (RK)*

All linear combinations of the reproducing kernel

$\Rightarrow$ *Hilbert space.*

e.g. The size (norm) of:

$$g = \alpha k(x, x_1) + \beta k(x, x_2)$$

is

$$||g|| = \sqrt{\alpha^2 k(x_1, x_1) + 2\alpha\beta k(x_1, x_2) + \beta^2 k(x_2, x_2)}$$

RK gives the basis functions and their size in one package.

Scientist: Looks like a great estimator and we don't have to assume anything!

Nychka You choose a reproducing kernel you choose a model.

Scientist: What kind of model?

The RK measures the size of g, usually in terms of wiggles
– numerical folks like this.

or
your g is a stochastic process with its covariance given by RK
– Bayesians like this. Frequentists call this Kriging so they can use it.

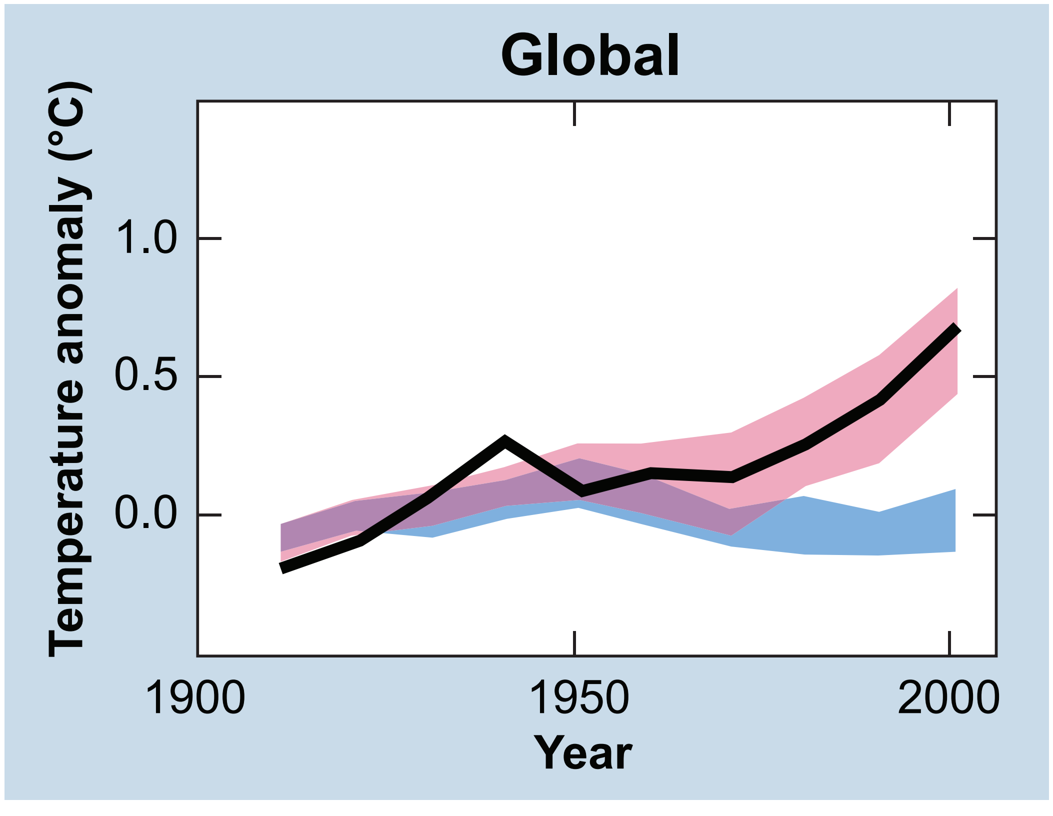
Scientist: Do I really have to choose?

Nychka Let's take a look at your data!

Humans and climate

IPCC Fourth Assessment Report:

Most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations.

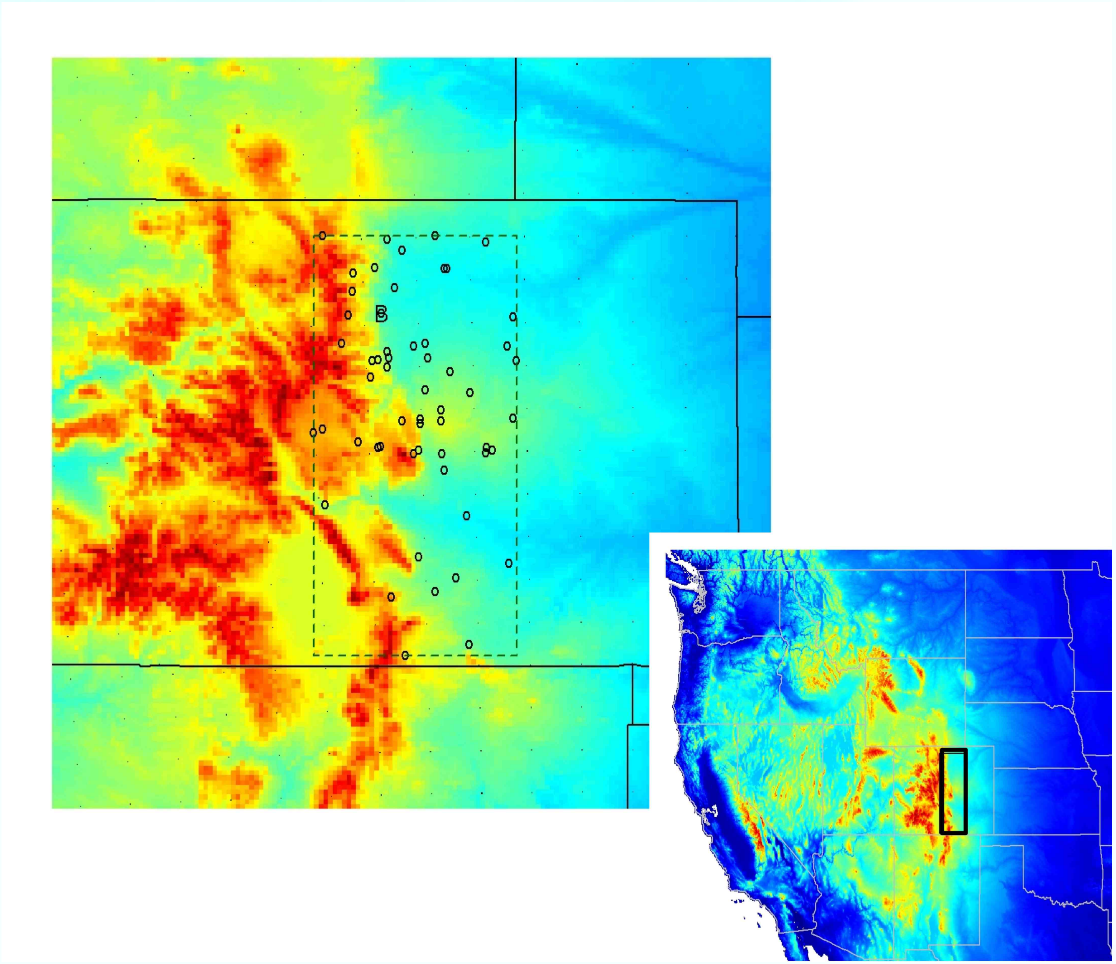


Observations

Climate models:
with human greenhouse gases
without

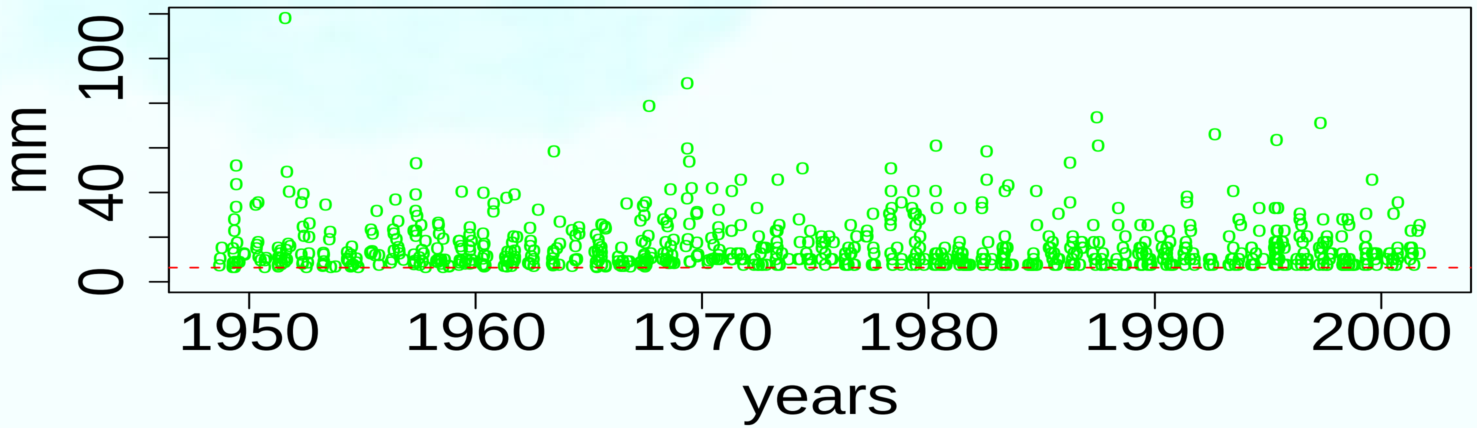
Sometimes extreme weather events are important.

Extreme rainfall in Colorado



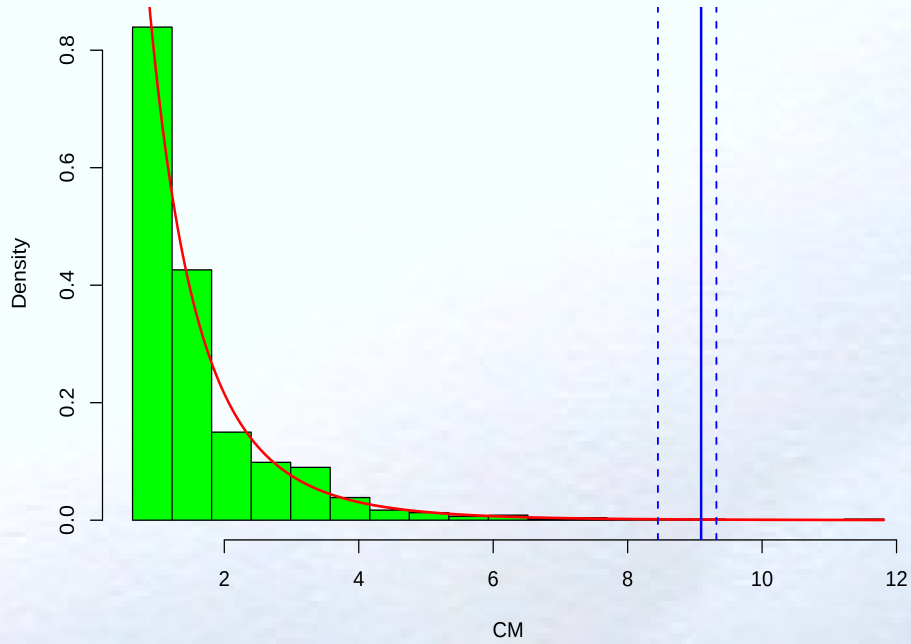
What is the size of a daily rainfall event that happens about once every 25 years? 100 years?

50 years of data from Boulder, CO



Fit using Generalized Pareto distribution and the estimated 25 year event, $\approx 9cm$.

(This fit depends on a scale and shape parameter.)



A spatial model

- Use extreme value statistical theory to approximate the distribution of large values $\rightarrow$ three parameters. probability of exceedence, q, shape ξ , log scale ϕ.

$$P\{\mathbf{Y} - u > y | \mathbf{Y} > u\} = \left(1 + \frac{\xi\, y}{\exp \psi}\right)^{-1/\xi}$$

- Assume that the parameters of the distribution vary over space according to a spatial process.

  e.g . $\xi(x)$ and $\psi(x)$

- Consider more general "spatial" coordinates besides just longitude and latitude.

  e.g Annual precipitation and elevation.

My equation slide

*-2 * log posterior density:*

Sample from the posterior:

$$-\sum_i \log L(\boldsymbol{Y}_i, \xi(x_i), \psi(x_i)) + \boldsymbol{\xi}^T \Sigma_1^{-1} \boldsymbol{\xi} + \boldsymbol{\psi}^T \Sigma_2^{-1} \boldsymbol{\psi}$$

i indexes the 50+ weather stations

Spline estimator:

minimize over ξ and ψ in two Hilbert spaces:

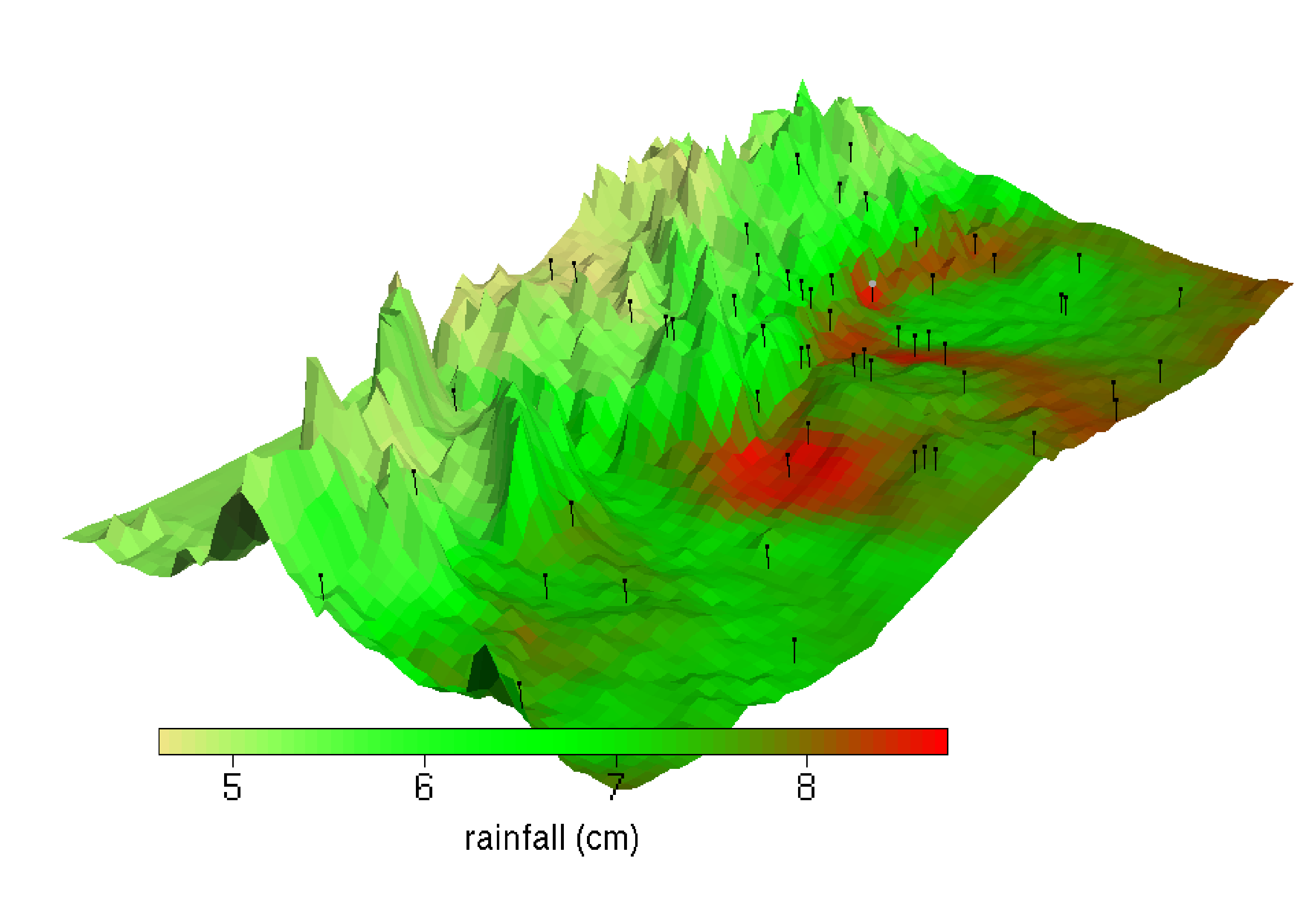
$$-\sum_i \log L(\boldsymbol{Y}_i, \xi(x_i), \psi(x_i)) + \lambda_1 [\xi, \xi]_1 + \lambda_2 [\psi, \psi]_2$$

- Spatial covariances are also the RKs for the inner products and Hilbert spaces.

- The Bayes approach yields a Monte Carlo sample of likely surfaces

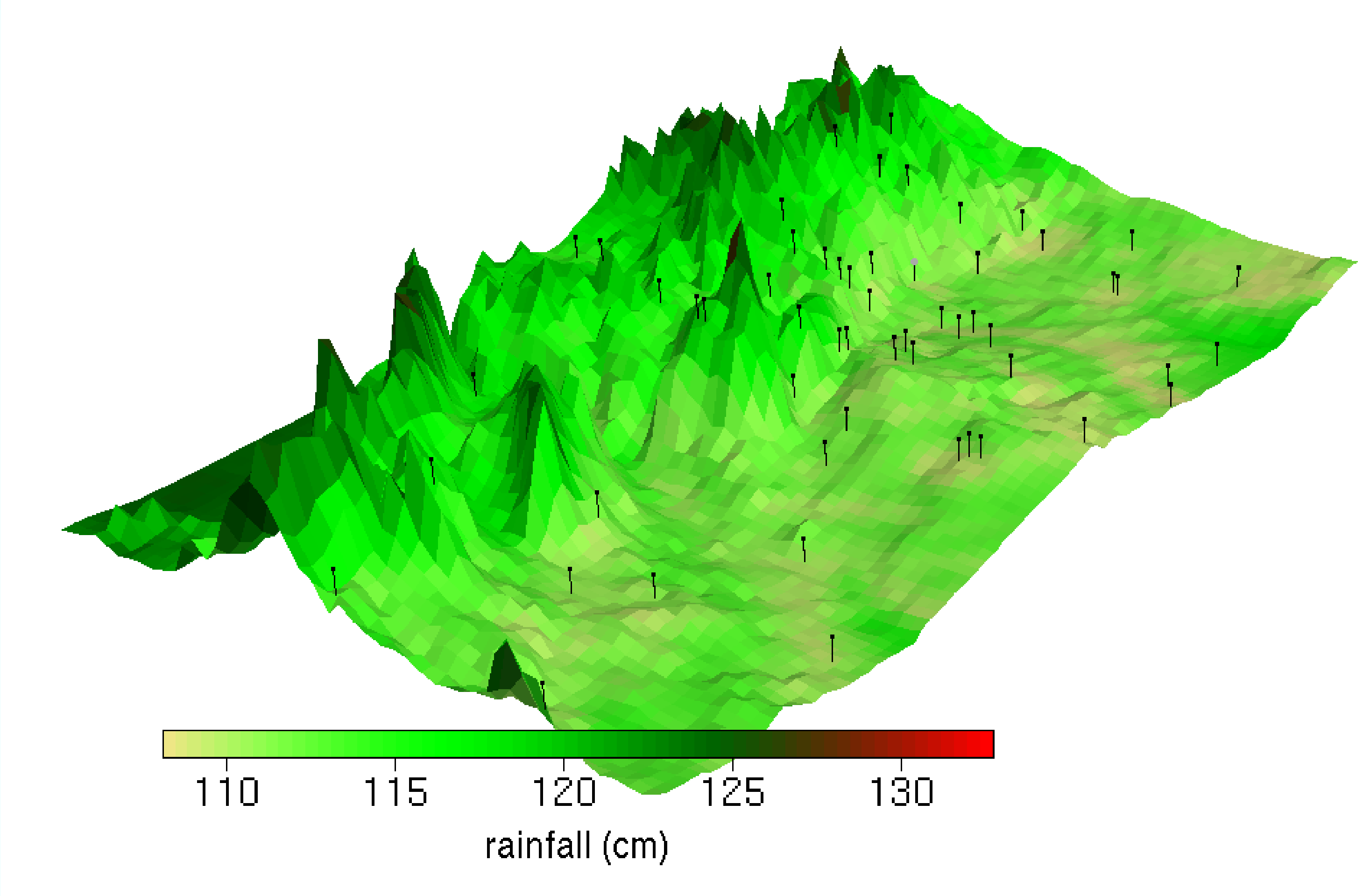
Rainfall extremes for the Colorado Front Range

Spline surface superimposed on elevations.



Estimated 25 year return values for daily precipitation.

95% upper bound on uncertainty



95 percent posterior quantile w/r to the mean.

- The initial problem of interpolation of 1-d data helped to develop the ideas of reproducing kernels and basis functions.

- Questions about climate depend on the scale: global or regional or local – means or extremes.

- Spatial models have at their core spline-like variational problems and reproducing kernels as basis functions.

There are still many curve and surface problems connected with understanding our Climate System.

There is lots of room left in Hilbert Space!

Thank you!

Questions?