Statistics, Numerical Models and Ensembles

Douglas Nychka,
Reinhard Furrer, Dan Cooley, Claudia Tebaldi,
Linda Mearns, Jerry Meehl and Richard Smith (UNC).

- Spatial prediction and data assimilation
- Precipitation extremes
- Combining IPCC climate model exp.

Supported by the National Science Foundation
CAS2K5, Annecy, FR, Sep 2005
Another way of summarizing the talk

Part 1: Observations are in the wrong place!

Part 2 Observations do not measure what we want!

Part 3 Not sure what we have observed!
What do you want to know? e.g. $\theta$
What have you measured? e.g. $Y$
Relate them using a probability distribution. e.g. Data = parameter + error

$\theta = \theta + \epsilon$

Characterize reasonable values for $\theta$ given the data
The statistical method

For complicated problems

- Use Bayesian models and Monte Carlo methods to generate a statistical ensemble for $\theta$.
- The ensemble mean is a good estimate for $\theta$.
- The spread is a good measure of uncertainty for $\theta$. 
Part 1: Observations are in the wrong place!

Air quality
Spatial Prediction

Predict surface ozone where it is not monitored.

Ambient daily ozone in PPB June 16, 1987, US Midwestern Region.
A model for the spatial field

- The ozone surface has a mean and variance that can vary over space.
- The correlation of ozone at two different locations has a known form.
- Ozone follows a Gaussian distribution
An ensemble approach

- Start with an ensemble of fields that are distributed according to one’s best guess or forecast — without consulting the data.
- Update each member of the ensemble using the observed data.

The sample mean and covariances among the ensemble members could be used for the update calculations.

This is the same algorithm used in the ensemble Kalman filter for numerical weather prediction.
Some ensemble members for ozone
Uncertainty of ozone at center of region

Predictions across 100 members.
Spatial Prediction

The ensemble mean

A Kriging, Bayes, OI, BLUE solution
A real ensemble forecast.

Updates done as in ozone example

DART T85 CAM GPH at 500hPa

20 of 80 members for 00Z 01 Feb 2003

CONTOUR FROM 5320 TO 5800 BY 80
Part 2 Observations do not measure what we want!
Precipitation extremes

How will climate change effect extreme precipitation?
Extremes in precipitation are used to determine flood potential for urban areas, for dam and roadway specifications and also have extensive ecological importance.

• How does one estimate extremes where no observations are made?

• How does one determine a possible “25 year event” from 20 years worth of data?

Typically extremes are described by the return period: “A 25 year event” = probability of seeing this value (or higher) in a given year is 1/25 or 4%
The Western US
Colorado Front Range
Observed precipitation for Boulder, CO

Daily precipitation amounts
Observed precipitation for Boulder, CO

Daily precipitation amounts thresholded at 2.5 cm

Distribution above threshold:
A spatial model for precipitation extremes

- Use extreme value statistical theory to approximate the distribution of large values → three parameters.
- Assume that the parameters of the distribution vary over space. (see Part 1.)

If you know the parameters of the distribution this can easily be converted to a 25 year return level.
Distribution fit to the Boulder exceedances

... and the estimated 25 year event ($\approx 9\,cm$)
Six ensemble members for 25 year event

25 year return level based on all daily met stations in the Front Range
Ensemble mean of the 25 year return level
Ensemble mean of the 25 year return level

Elevation and return level (cm)
Part 3 Not sure what you have observed!
Data and the IPCC

What will the climate be like in 2100?

How much data do we have to answer this question. The most recent experiments to support the fourth report of the International Panel on Climate Change amount to an archive of approximately 100 Tb.

- More than 20 different climate models/modeling centers represented.
- Several different future scenarios.
- Multiply responses e.g. temperature, precipitation,
Some “Data”

Present winter temperatures, 9 AOGCMs

Future - Present

Is this really a massive data set?
Standard IPCC regions
A Statistical Model

Observations  =  truth\textsubscript{P} + error

Model Present  =  truth\textsubscript{P} + model/region bias\textsubscript{1} + error.

Model Future  =  truth\textsubscript{F} + model/region bias\textsubscript{2} + error.

Climate change  =  truth\textsubscript{F} - truth\textsubscript{P}
Some ensemble members for regional change

Future – Present DJF temperature (A2)
Western North America temp. change
Ensemble distributions: A summary for the AR4
Statistical ensembles are a useful way to estimate spatial fields and characterize uncertainty.

Statistical methods can be used to estimate complex indirect features.

The size of ”massive” data sets may not be massive. Statistics can be used to gauge the representativeness of a sample.
Thank you!