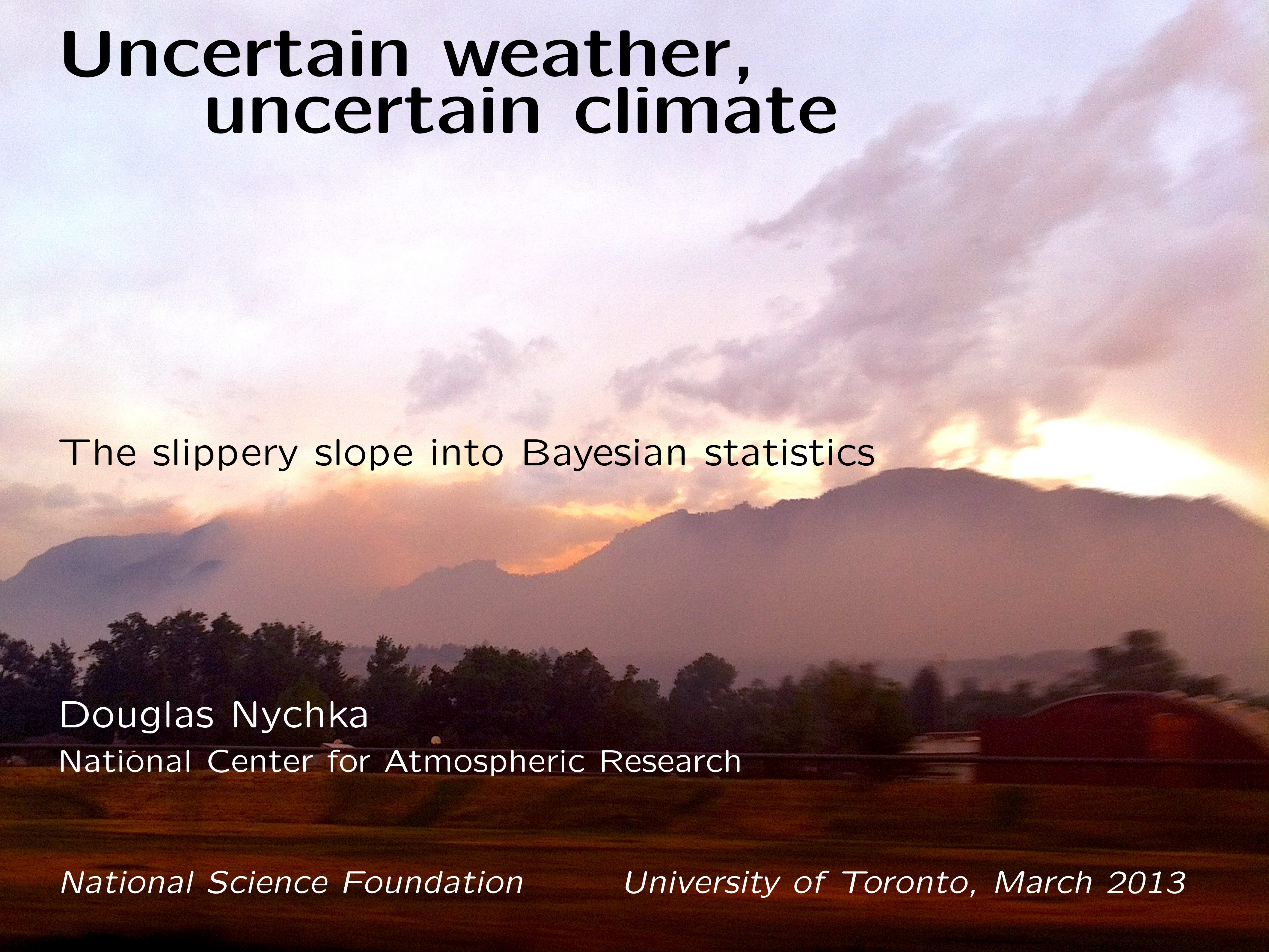


Uncertain weather, uncertain climate

The slippery slope into Bayesian statistics

Douglas Nychka
National Center for Atmospheric Research

National Science Foundation

University of Toronto, March 2013

Outline

- Inverse problems in 19th century
- Weather observations
- How to make a forecast
- Past climate: what we measure
- Reconstructing past climate

Big ideas

Divide up the problem into two parts:

- What do know about the state of the atmosphere?

- How is an observation related to what you know?

- **prior** information, **likelihood** of observating data,

$\rightarrow$ **posterior** probabilities

222b Baker Street

From "The Adventure of the Speckled Band"

"Good-morning, madam," said Holmes cheerily. "My name is Sherlock Holmes.

222b Baker Street

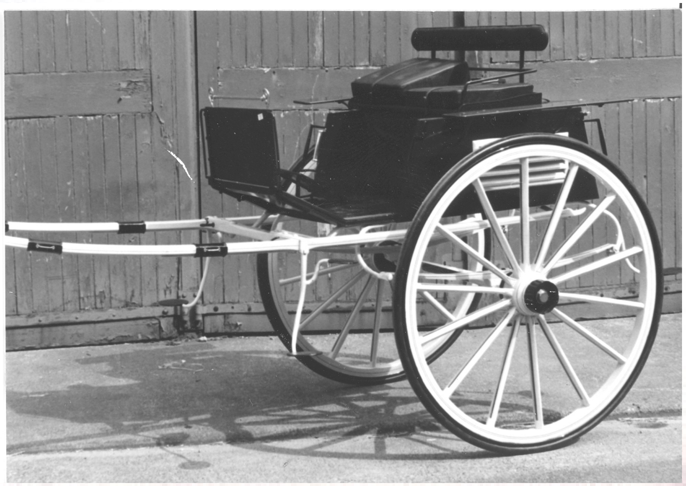
Her features and figure were those of a woman of thirty, but her hair was shot with premature grey, and her expression was weary and haggard.

222b Baker Street

"We shall soon set matters right, I have no doubt. You have come in by train this morning, I see. ... and yet you had a good drive in a dog-cart, along heavy roads, before you reached the station. "

"The left arm of your jacket is spattered with mud in no less than seven places. The marks are perfectly fresh.

There is no vehicle save a dog-cart which throws up mud in that way, and then only when you sit on the left-hand side of the driver."



Holmes' calculation

Before meeting Ms. Helen Stoner:

- A PRIOR probability of type of **vehicle**

Knowledge of vehicles effects:

- LIKELIHOOD of **observation** *given* type of **vehicle**

Combine prior with observation:

- POSTERIOR is a product:

Likelihood of **mud stains** *given* type of **vehicle**

$\times$ Probability of type of **vehicle**

Maximize over **vehicle**

Holmes' conclusion – the highest probability

- **vehicle** = **dog cart**

Some differences

Holmes' genius:

Uses observations without assuming much prior information!

Weather and climate applications:

- Our observations are not as decisive – we must rely more on the prior information.

- Have to use a computer to do the computation!

NCAR/U Wyoming, Yellowstone Supercomputer, 70,000+ processors.

Making a weather forecast

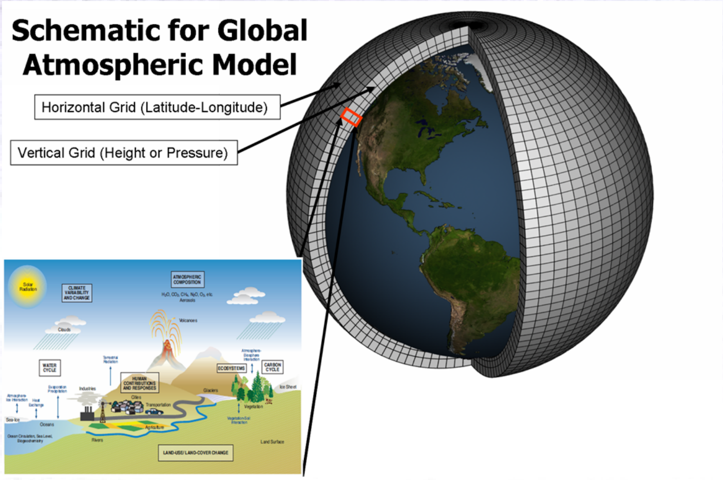
Credits: Data Assimilation Research Section, NCAR
Jeff Anderson, Tim Hoar, Kevin Raeder

The ensemble Kalman filter

Describing the atmosphere

- divide up the atmosphere into a large 3-d grid $\approx 144 \times 96 \times 27 = 1/3$ millon points

- the temperature, pressure, water vapor and the wind for each grid box for each time – need at least 6 variables to describe state.



the state vector at one time:

atmosphere = about 2 million numbers

Even this is only about 200km resolution.

Prior information for the atmosphere

The atmospheric state is uncertain.

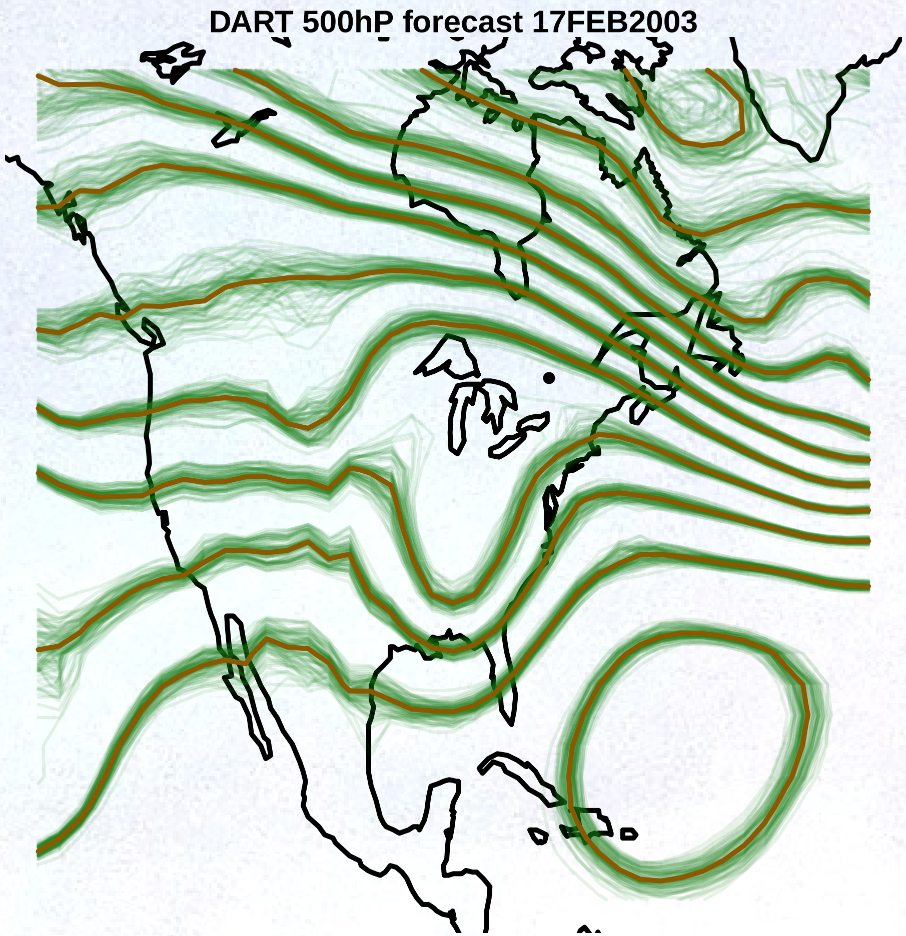
Represent the uncertainty with a sample (ensemble) of states
– all equally plausible and each physically consistent.

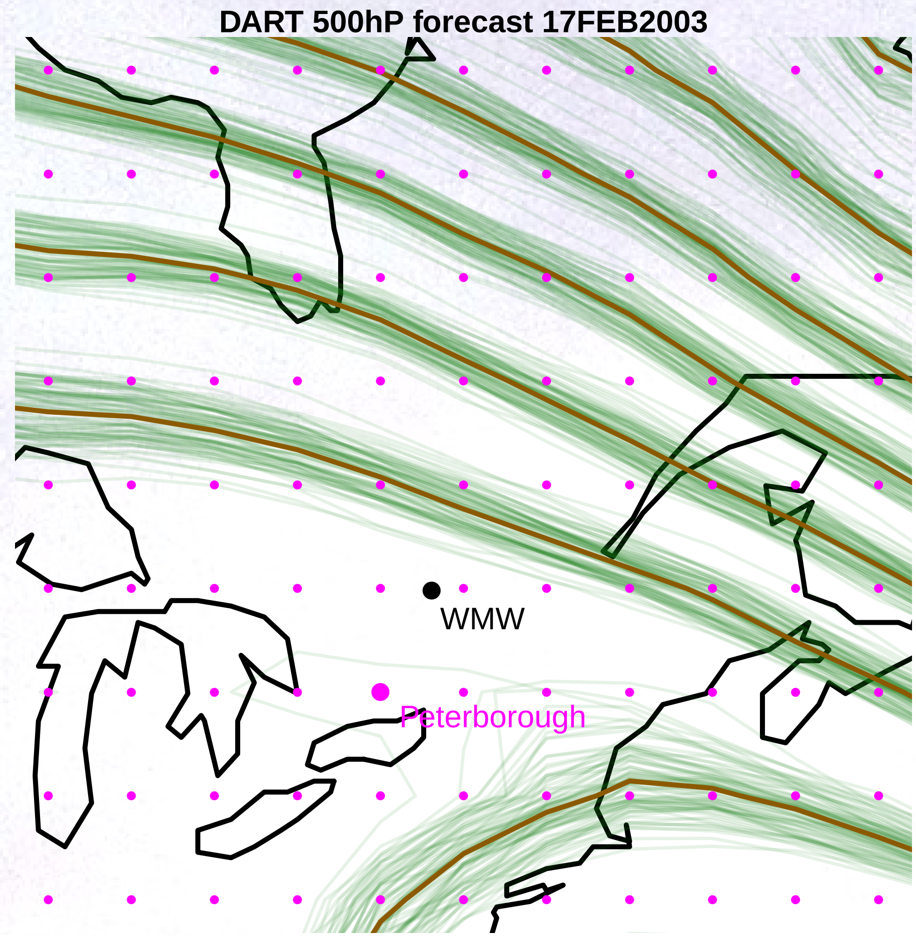
$$\mathbf{atmosphere}^1,\ \mathbf{atmosphere}^2,\ \mathbf{atmosphere}^3, \ldots$$

For this example there will be 80 members in the ensemble

An example

Height where the pressure is 500hP, FEB 17, 2003 (12Z)





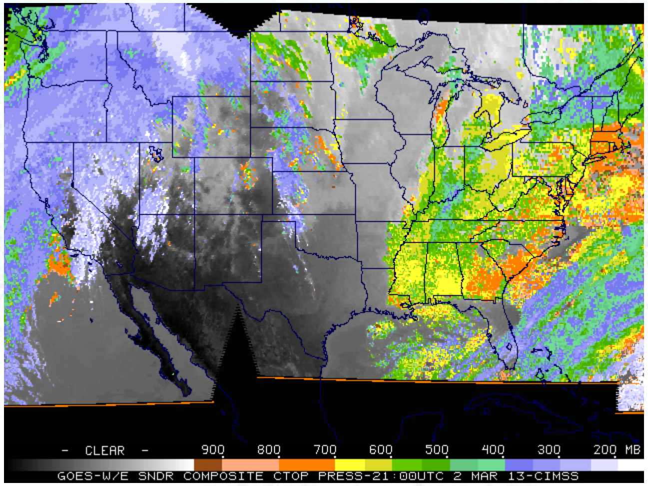
- 80 member ensemble
- Just a small, 2-d glimpse – the PRIOR is global and 3-d.

Observations for the atmosphere

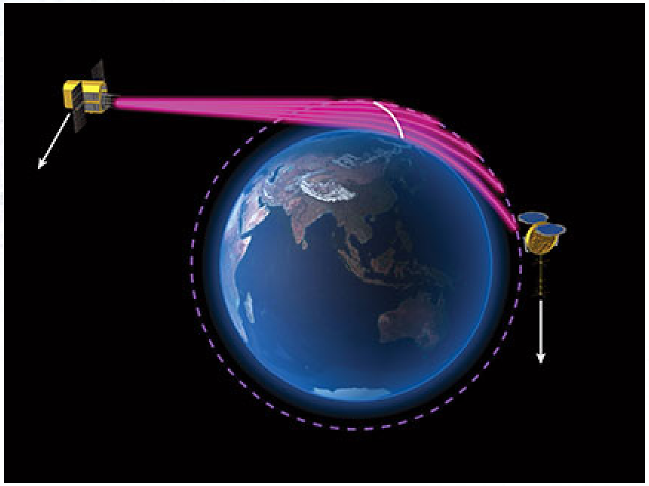
Surface observations

rawinsondes (balloons)

satellite images



remotely sensed

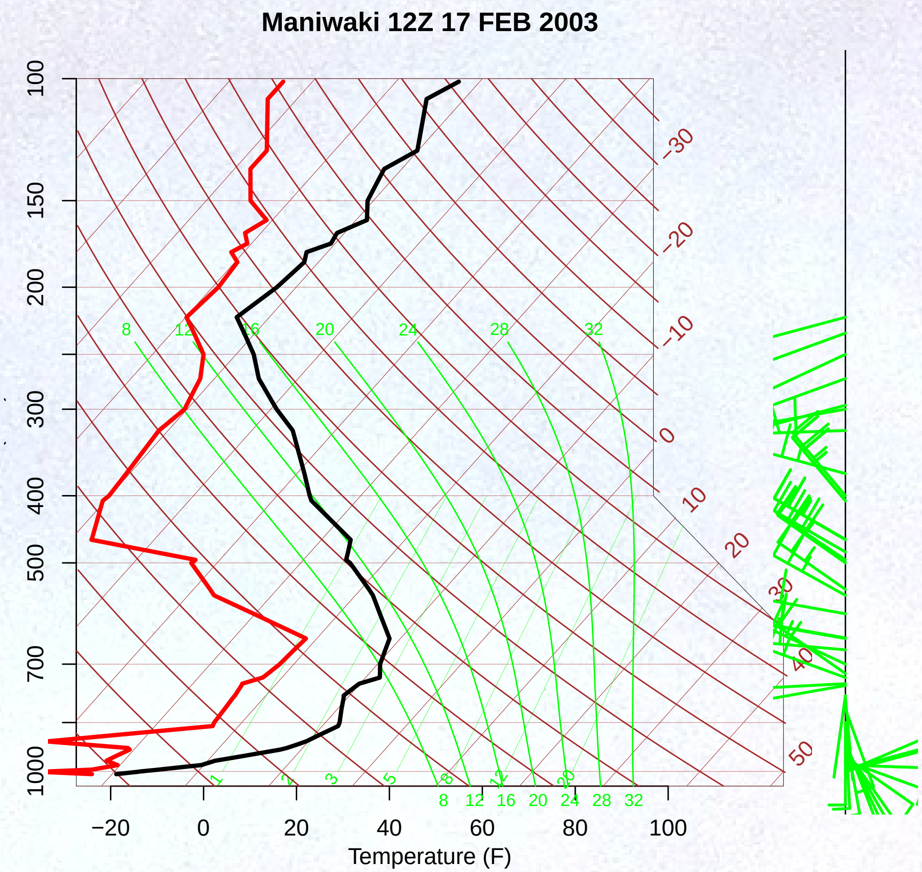


... ???

LIKELIHOOD: the probability of an **observation** *given a state of the* **atmosphere**

- All observations have some degree of measurement error.
- All can be related back to the state of the atmosphere.

Radiosonde Maniwaki , Quebec

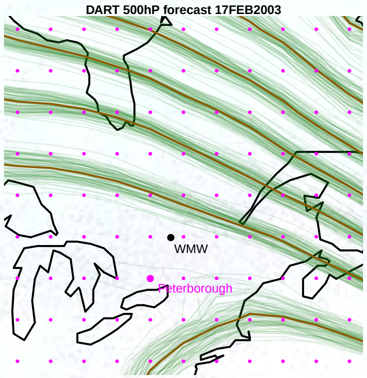


12Z FEB 17, 2003, height at 500hP measured at 5530 (m)

POSTERIOR is a product:

Likelihood of **observation** *given* **atmosphere**
$\times$ Probability of **atmosphere**

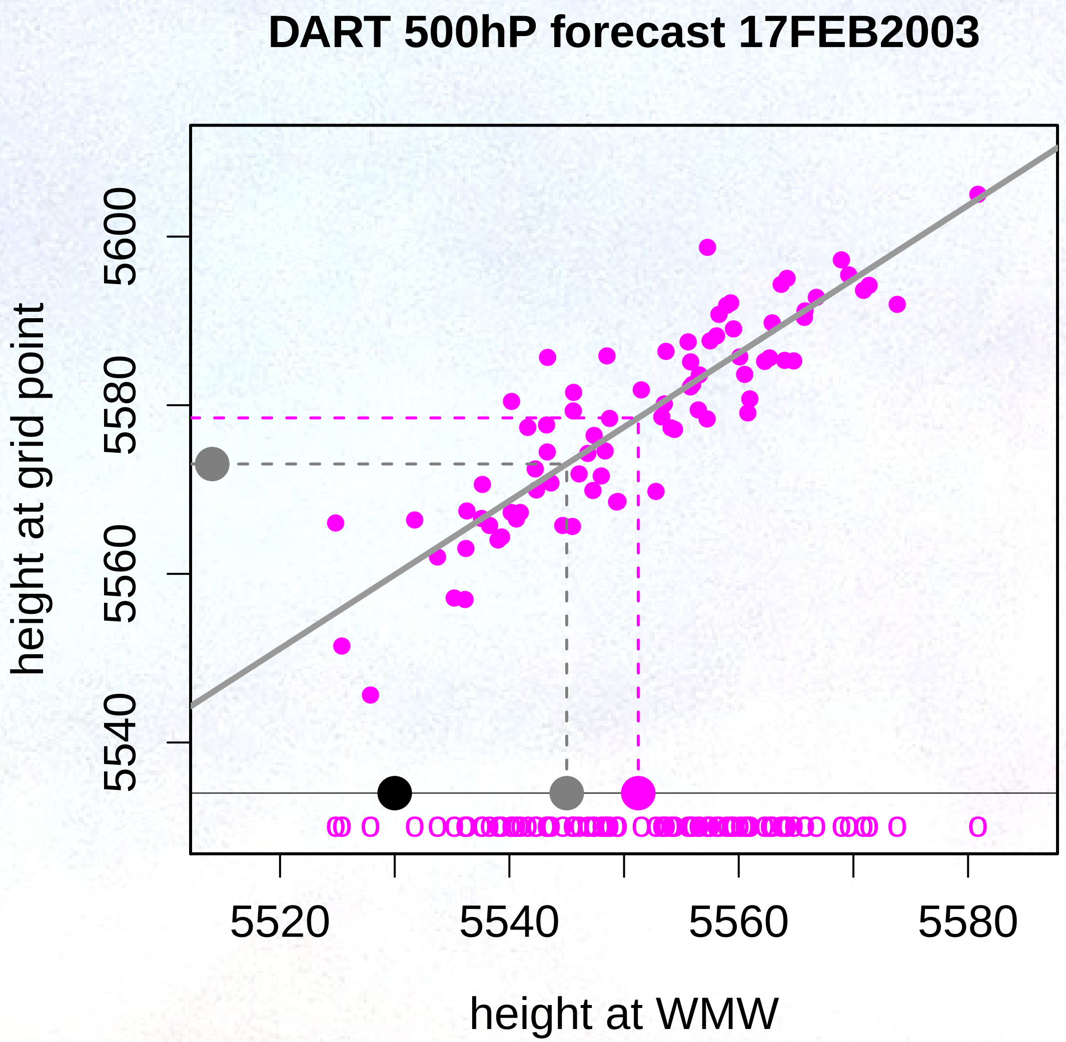
Difficult to compute exactly . . .



Use the ensemble

The algorithm

Updating the estimate at the Peterborough grid point.
Actual observation at 5530 (m). Prior prediction is 5551 (m)



- Combining **observation** and PRIOR prediction gives **POSTERIOR** estimate of 5545 (m).

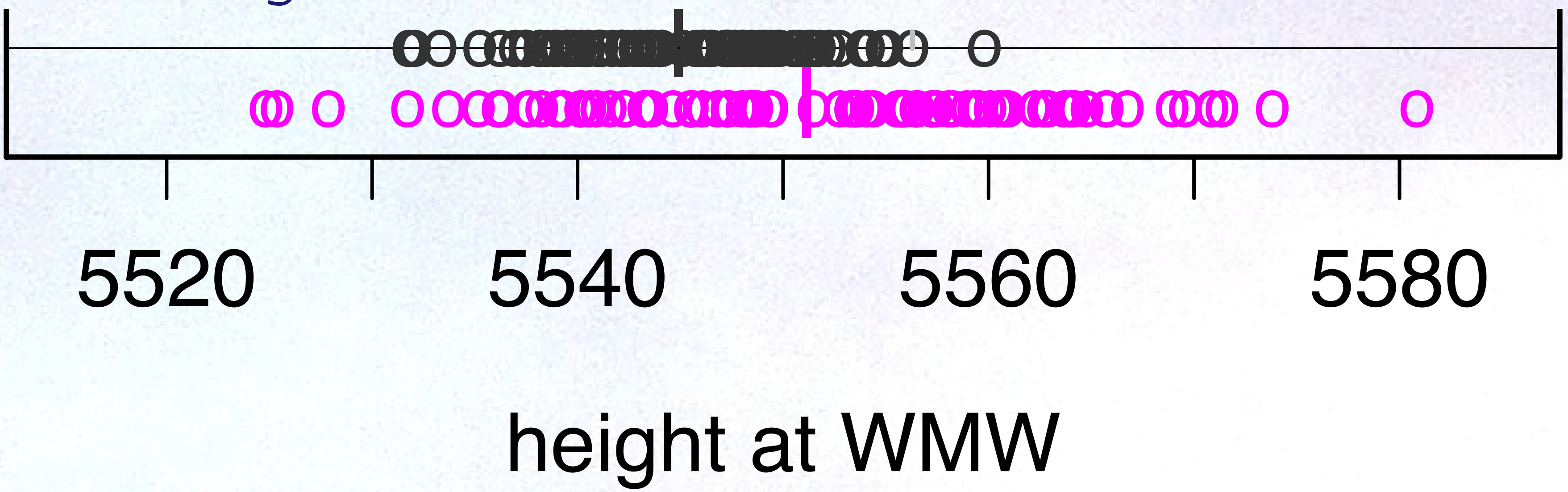
- Have the **PRIOR** distribution for the observation and at the grid point. *Fit a line by least squares.*

- Use scatterplot relationship to get **POSTERIOR** estimate for Peterborough grid box of 5573 (m).

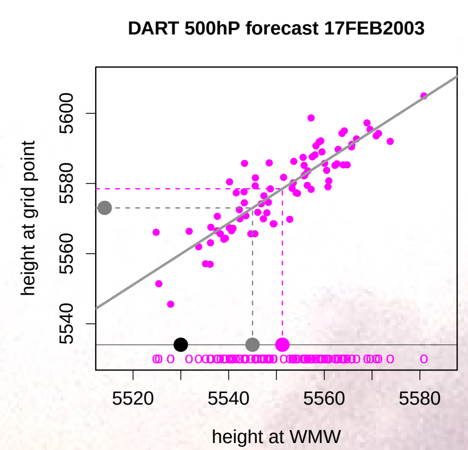
Updating the ensemble

- Shrink the ensemble members to the POSTERIOR estimate

500hP height at observation location



- Use scatterplot relationship to get **POSTERIOR** values of ensemble members at Peterborough grid box.



The full POSTERIOR ensemble

Repeat this algorithm for all available observations at this time and for every value in the state.

This is why we need a supercomputer!

Making a forecast

Use a weather model, M, to advance the current state into the future – this is the forecast.

atmosphere 6 hours later $= M($ **atmosphere**$)$

A state-of-the-art M:

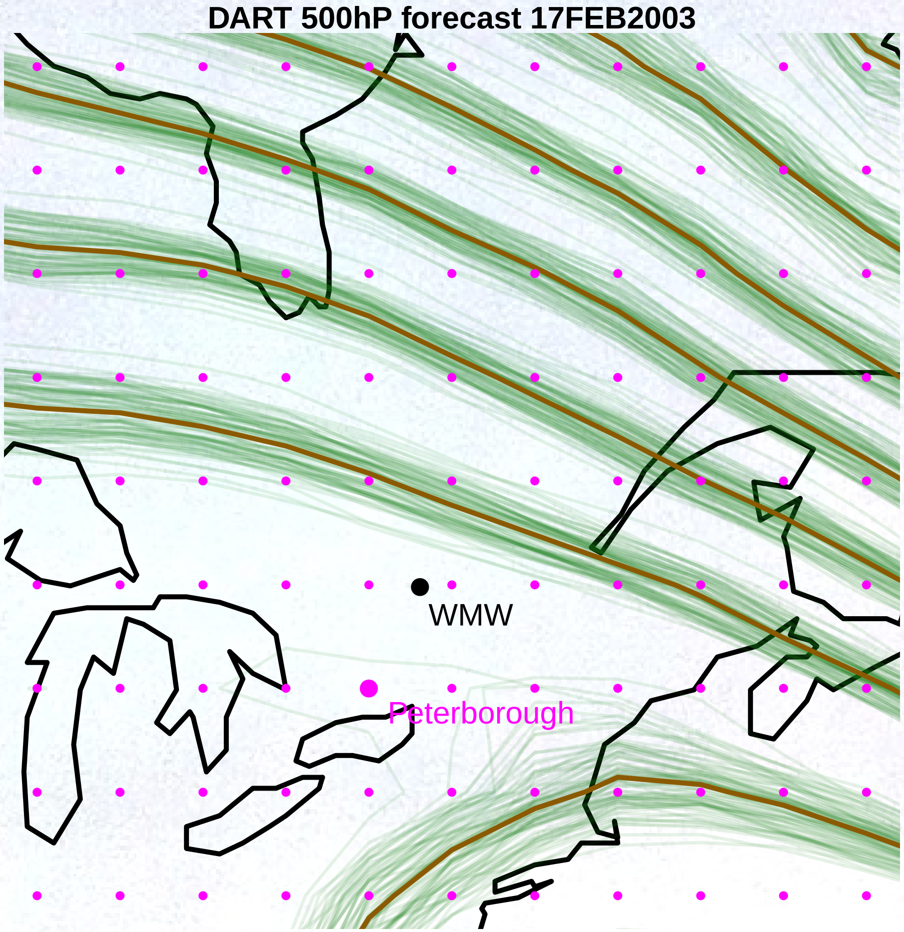
- based on detailed physics of the atmosphere
- millions of lines of code, requires a supercomputer

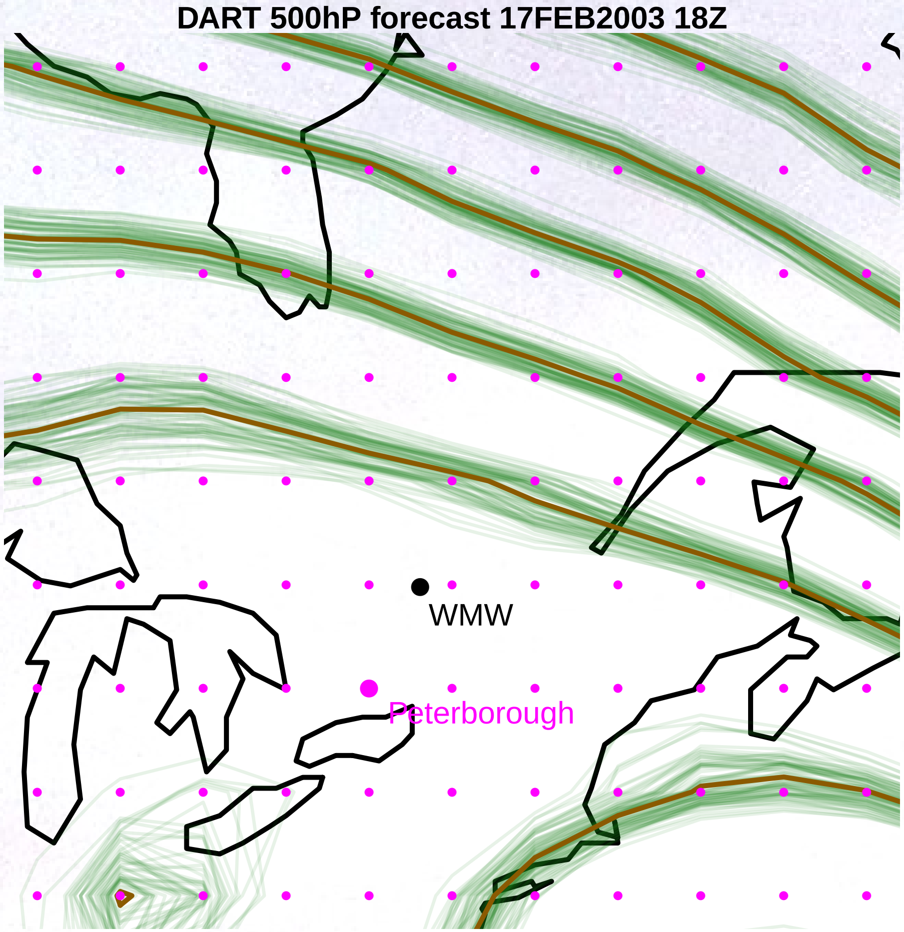
Apply M to the POSTERIOR ensemble members

atmosphere 6 hours later$_1 = M($ **atmosphere**$_1)$
atmosphere 6 hours later$_2 = M($ **atmosphere**$_2)$
...
atmosphere 6 hours later$_{80} = M($ **atmosphere**$_{80})$

The 6 hour forecast

PRIOR ensemble at 12Z POSTERIOR ensemble advanced 6 hours





The forecast becomes the new PRIOR ...

Estimating past climate

Credits: Martin Tingley and Peter Huybers, Harvard U

Bayesian Hierarchical Models

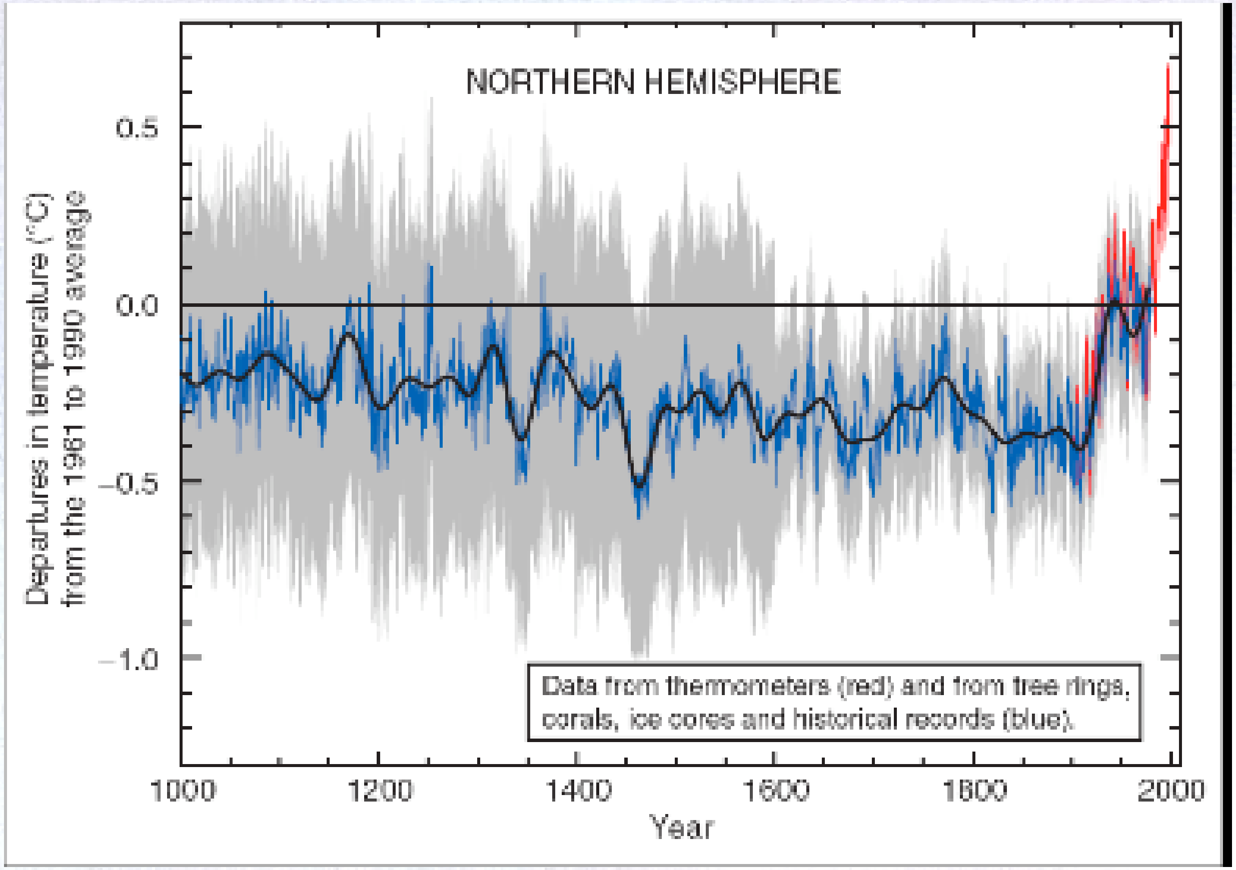
How do we know temperatures before thermometers?

Surface temperatures are related to other obervations:

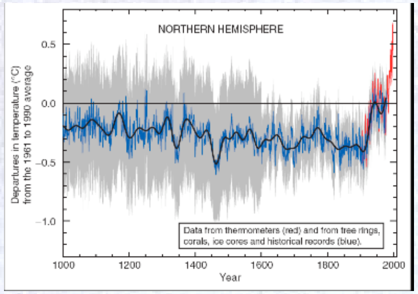
e.g. tree ring width and density, pollen, ice cores and lake sediments.

The Mann et al reconstruction

Northern Hemisphere temperatures 1000AD - 2000AD



Mann, Bradley and Hughes 1998 *Nature*.

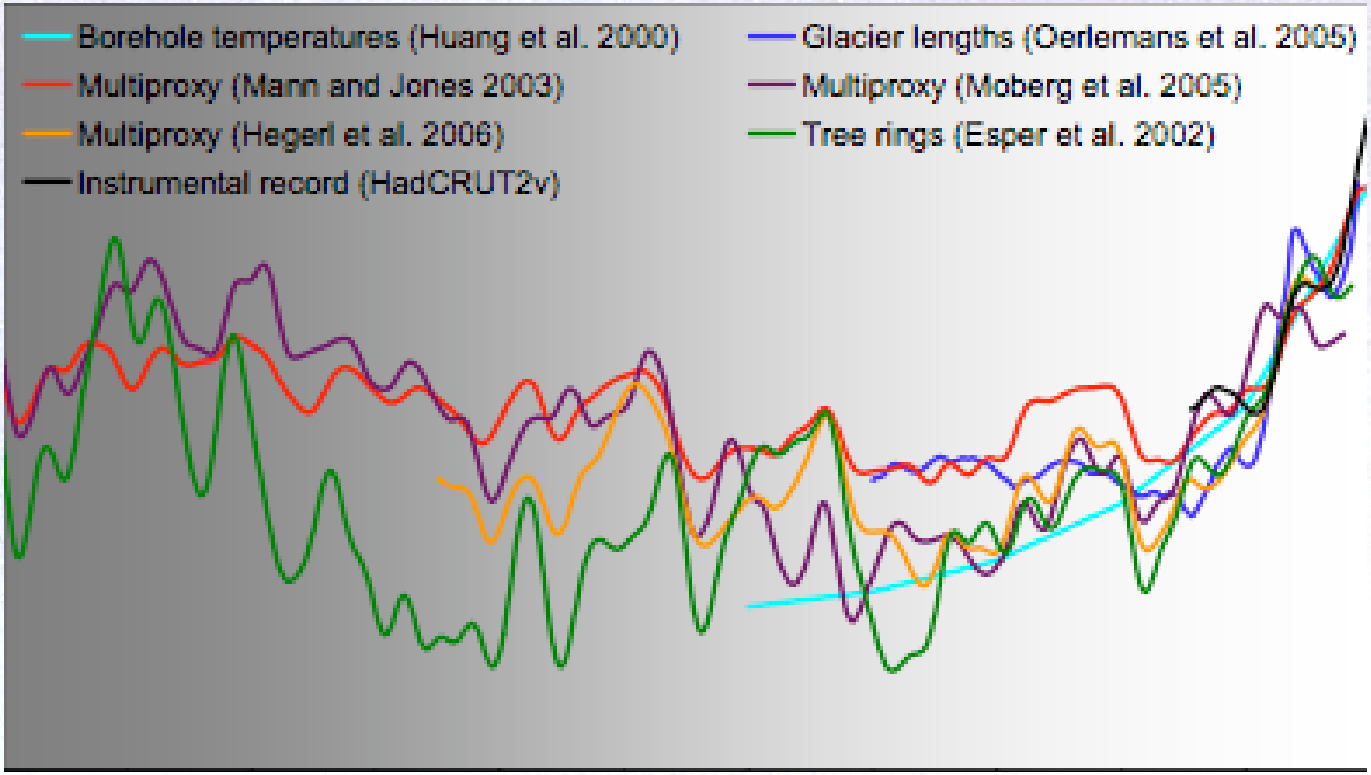


Mann et al. (1999)

"... the 1990s are likely the warmest decade, and 1998 the warmest year, in at least a millennium"

- Used informally (by others) to argue human influence on climate
- The scientific process: the field has moved on from this initial work.

An ensemble of NH reconstructions



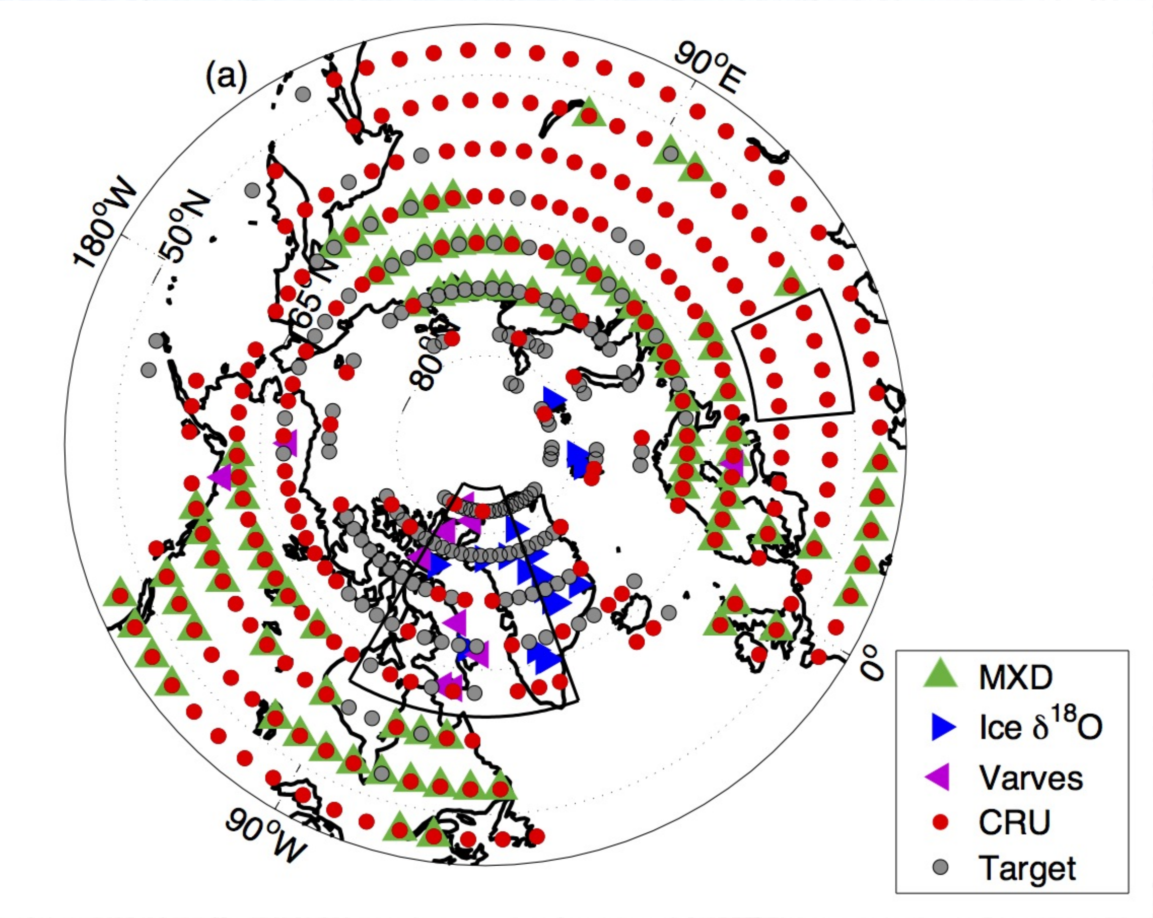
Circa June 2006:

National Academy of Sciences Report Surface Temperature Reconstructions for the past 2000 years.

Recent statistical work

Martin P. Tingley & Peter Huybers (2013)

Goal: Create a spatial estimate of annual temperatures for high Northern latitudes and for the past 600 years.



Prior information

State is annual **temperature** on 96, 5$\times$ 5 degree grid boxes above 45N.

Use a statistical model to capture smoothness of the annual temperatures

- over space
- from year to year.

$$\textbf{temperature}_{t+1} = M(\ \textbf{temperature}_t) + \text{ random shocks}$$

- M is simple
- random component mimics the effects of weather

LIKELIHOOD: the probability of an **proxy observation** *given the local state of the* **temperature**

- Each observation has a linear relationship with local temperature plus random error.

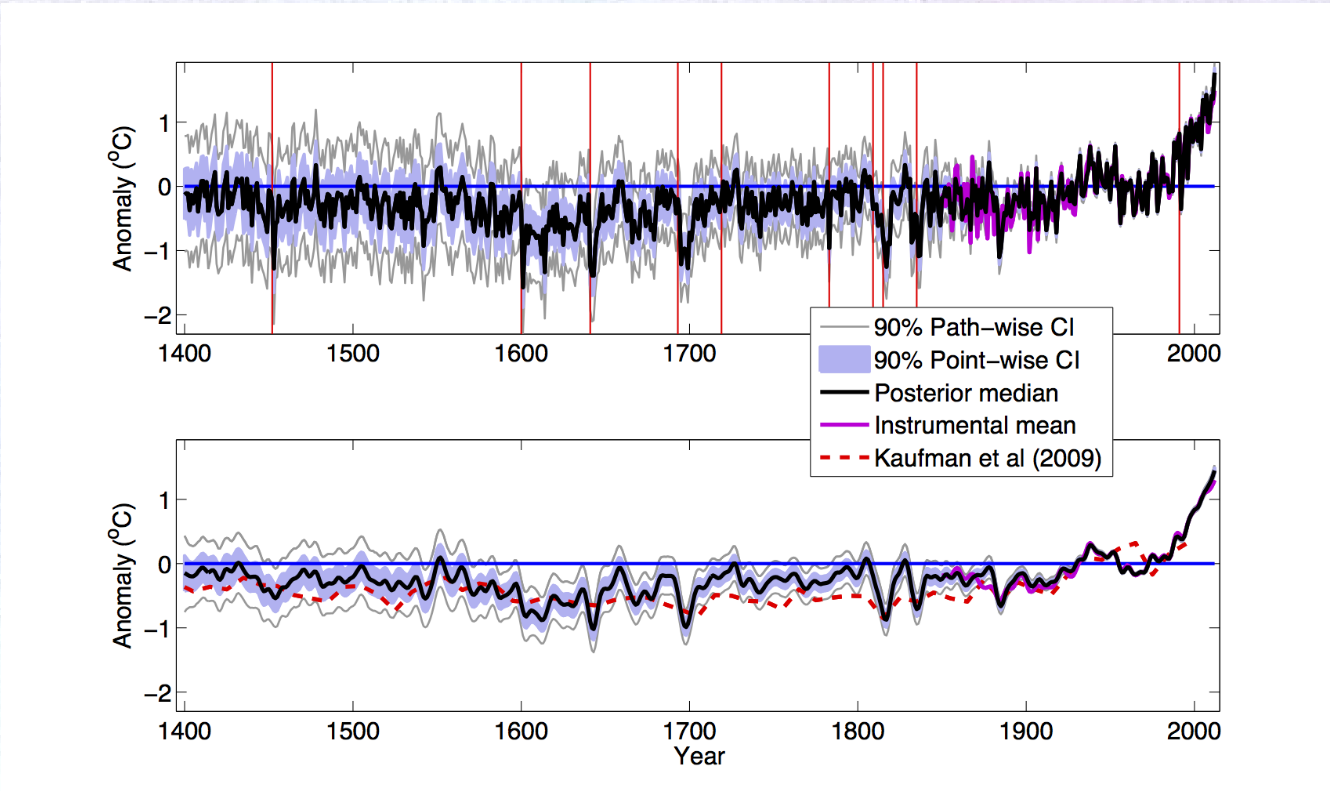
POSTERIOR is a product:

Likelihood of **proxy observation** *given* **temperature**
$\times$ Probability of **temperature**

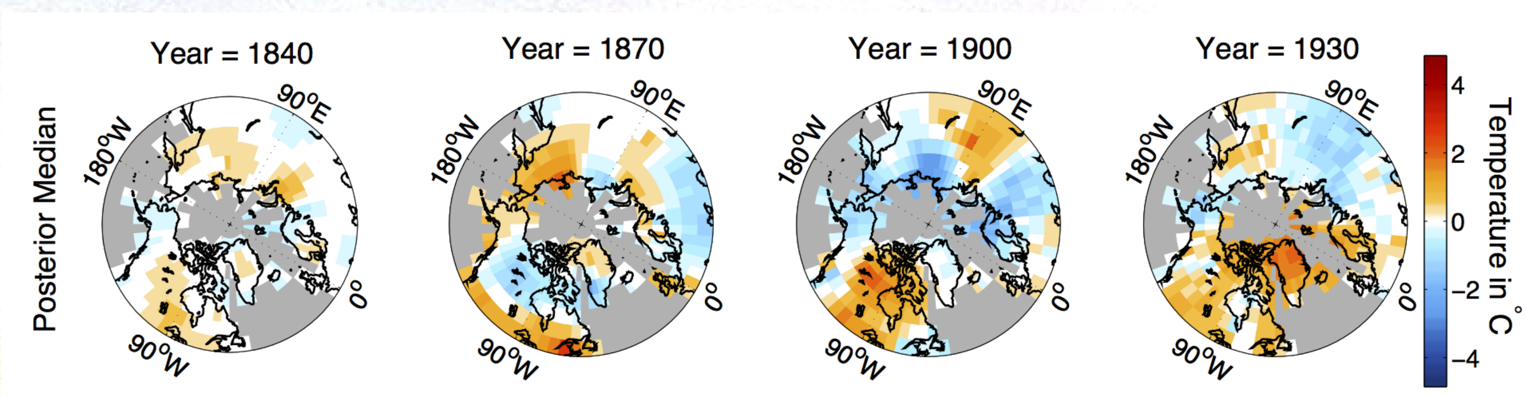
Find temperature histories that represent the POSTERIOR distribution

- No simple formulas – need to use Monte Carlo sampling
- Analysis based on 4000 sample histories
- Statistical parameters are estimated along with histories.

Average temperature for the high latitude region:



Snapshots of four years



Thank you!