

# Prologue: A statisticians view of the carbon problem

---

Douglas Nychka,

[www.image.ucar.edu/~nychka](http://www.image.ucar.edu/~nychka)

- What we see.
- What we want.
- The forward model connections
- Going backwards and the Bayesian solution.



*Supported by the National Science Foundation*

*MSRI summer school Jul 2006*

# What we see.

---

Observations are made on CO<sub>2</sub> concentrations either remotely or directly. These can be:

- Averages over a vertical column
- Averages over time.
- Something else.

But call this the  $j$ <sup>th</sup> observation,  $Z_j$

*We need to know exactly how  $Z_j$  is related to the actual concentrations.*

# What we want.

---

Divide the atmosphere and time up into many boxes and equally spaced intervals.

e.g. For Dave Baker's work a  $4 \times 5$  degree grid and hourly intervals.

$x_i$  are the vector of  $CO_2$  concentrations at time  $i$  for all the grid boxes.

What we really want are carbon surface sources  $u_k$ , the surface fluxes, that are actually driving the concentrations.

*Find the source fluxes given the data:  
Estimate the  $u_k$ 's given all the  $Z_j$ s*

# The models

---

*From concentrations to observations.*

$$Z_j = h_j(x_i) + \text{'measurement error'}$$

*From sources to concentrations*

$$x_{i+1} = \Phi(x_i) + G(u)$$

*These two equations connect the data with the sources.*

**Our (statistical) goal is to find values for the sources that are consistent with the observations.**

# Inverse Problems

---

*Our equations:*

*Sources + Initial concentration*

→ *Concentrations at other times*

→ *Noisy and irregular observations.*

- **There is not a direct equation for the sources in terms of the observations!**
- **Limited observations make estimates of the concentrations incomplete.**
- **The presence of measurement error makes exact estimates of the sources ambiguous.**

# Bayesian statistics and the inverse problem.

---

*Bayesian statistics is a formal way to introduce additional constraints to solve this problem.*

The ‘answer’ is a distribution for possible values for the sources in light of the data  
– not a single answer.

*The conditional distribution of  $u$  given  $\{Z_i\}$ .*



Also known as the posterior