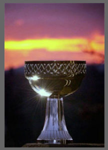
Prologue: A statisticians view of the carbon problem

Douglas Nychka,
`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 *NCAR/IMAGe summer school Jul 2007*

Sherlock Holmes solves an inverse problem

Sherlock Holmes. *The speckled band.*

... "Good-morning, madam," said Holmes cheerily. "My name is Sherlock Holmes. This is my intimate friend and associate, Dr. Watson, before whom you can speak as freely as before myself.

... 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. Sherlock Holmes ran her over with one of his quick, all-comprehensive glances.

"You must not fear," said he soothingly, bending forward and patting her forearm. "We shall soon set matters right, I have no doubt. You have come in by train this morning, I see."

"You know me, then?"

"No, but I observe the second half of a return ticket in the palm of your left glove. You must have started early, and yet you had a good drive in a dog-cart, along heavy roads, before you reached the station."

The lady gave a violent start and stared in bewilderment at my companion.

"There is no mystery, my dear madam," said he, smiling. "The left arm of your jacket is spattered with mud in no less than seven places. The marks are perfectl 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."

"Whatever your reasons may be, you are perfectly correct," said she.

What we see.

Observations are made on CO_2 concentrations either remotely or directly. These can be:

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

But call this the j_{th} 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 4×5 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_js

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

$\rightarrow$ *Concentrations at other times*

$\rightarrow$ *Noisy and irregular observations.*

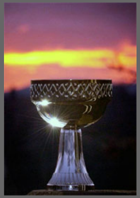
- **There is not a direct equation for the sources in terms of the observations!**
- **Limited observations make estimates of the concentrations incomplete.**
- **The presence of measurement error makes exact estimates of the sources ambiguous.**

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