Some applications of climate Data Assimilation

- Really two half lectures;
- An example of data assimilation for climate and biogeochemistry;
- An unfinished approach to model ensembles and a lesson on why it's sometimes good to go back to first principles.

Uncertainties in the relationship between concentrations and emissions

P. J. Rayner (Univ. Melbourne)
E. Koffi (LSCE)
M. Scholze (U. Bristol)
P. Friedlingstein (LSCE)
M. Raupach (CSIRO)
T. Kaminski (FASTOPT)
J.-L. Dufresne (IPSL)

"It's only a model"

(Monty Python and the Holy Grail)

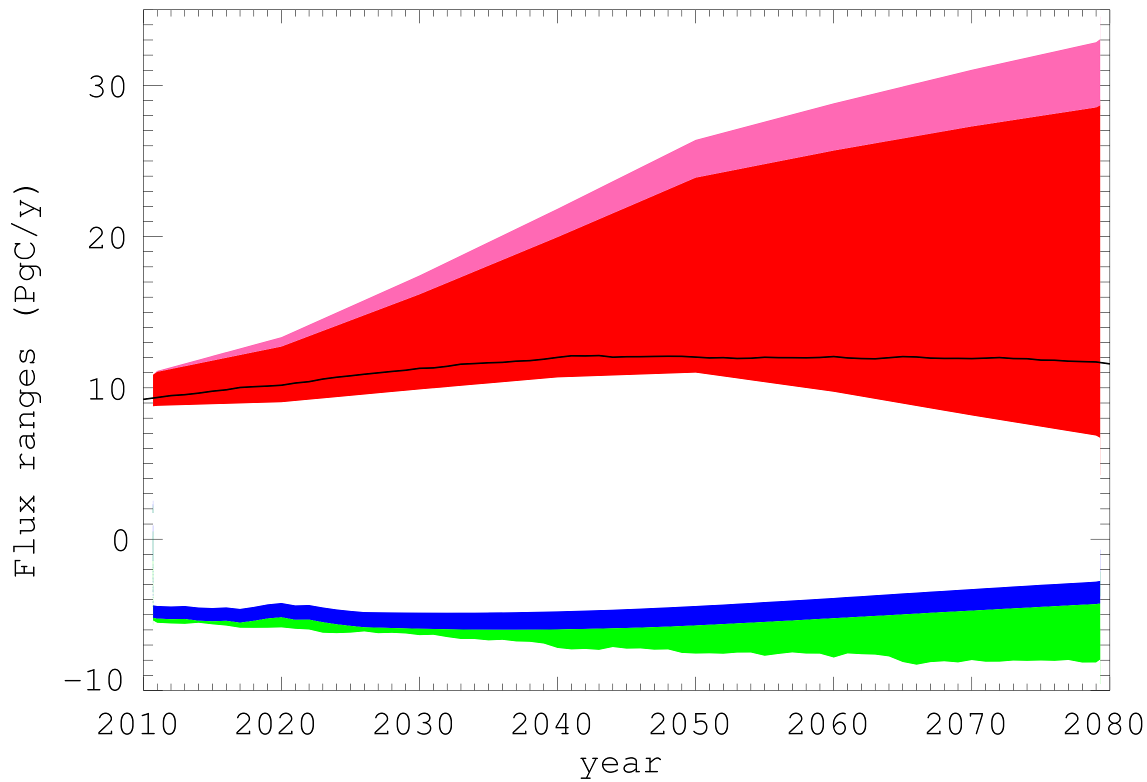
Papers

- This submitted to Philosophical Transactions of the Royal Society;
- Optimization from Koffi et al. (2010) almost submitted to Global Biogeochemical cycles.

Outline

- Uncertainties in the carbon cycle;
- A simple predictive model and its uncertainty;
- A little on sensitivity;
- Confronting the model with data;
- Conclusions.

Motivation



Ranges of global CO_2 fluxes. Red = anthropogenic, blue = ocean, green = land, pink = other vulnerabilities from Raupach et al., Tellus, 2010. Uptakes from IPCC-2007 Fig. 10.21. Black line shows emission scenario.

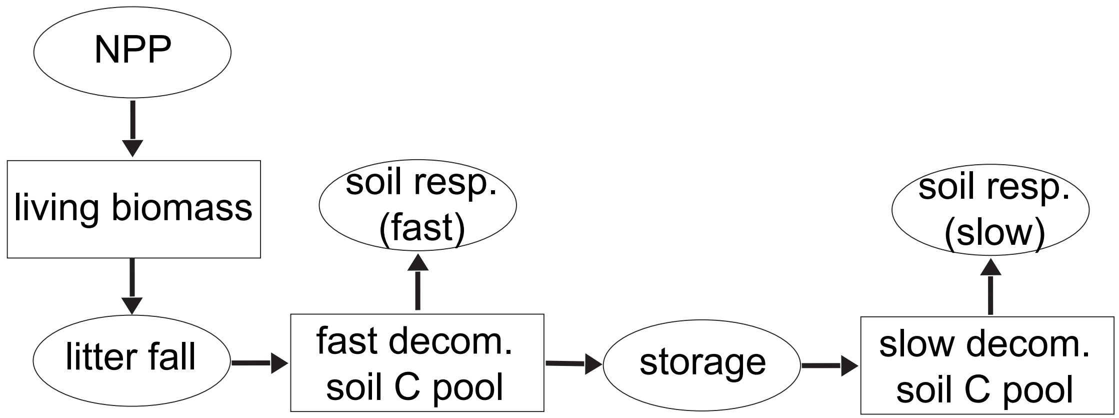
Sources of Terrestrial Model Uncertainty

- Different models include different processes;
- Equivalent processes are described with different equations;
- There are many uncertain parameters in these models.

Exploring Parameter Uncertainty

- Write simple box model of terrestrial carbon cycle
- Climate model $\rightarrow$ global model $\rightarrow$ simple model;
- Calculate sensitivities of future uptake to inputs;
- Calculate uncertainty of future uptake as function of uncertainty in input parameters;
- Assimilate current data and study reduced uncertainty on future uptakes.

Simple Model



$$\begin{aligned}\mathrm{NetUptake} = \quad & \mathbf{Production} - (1 - \mathbf{K}) \times \mathbf{LitterDecomposition} \\ & - \mathrm{SoilOutgassing}\end{aligned}$$

$$\mathrm{SoilOutgassing} \propto \mathrm{SoilPool} \times \omega^{\kappa} \mathbf{Q}_{10}^{T_a/10}$$

where ω = soil moisture and T_a = air temperature.

$$\frac{\partial \mathrm{SoilPool}}{\partial t} = \mathbf{K} \times \mathbf{LitterDecomposition} - \mathrm{SoilOutgassing}$$

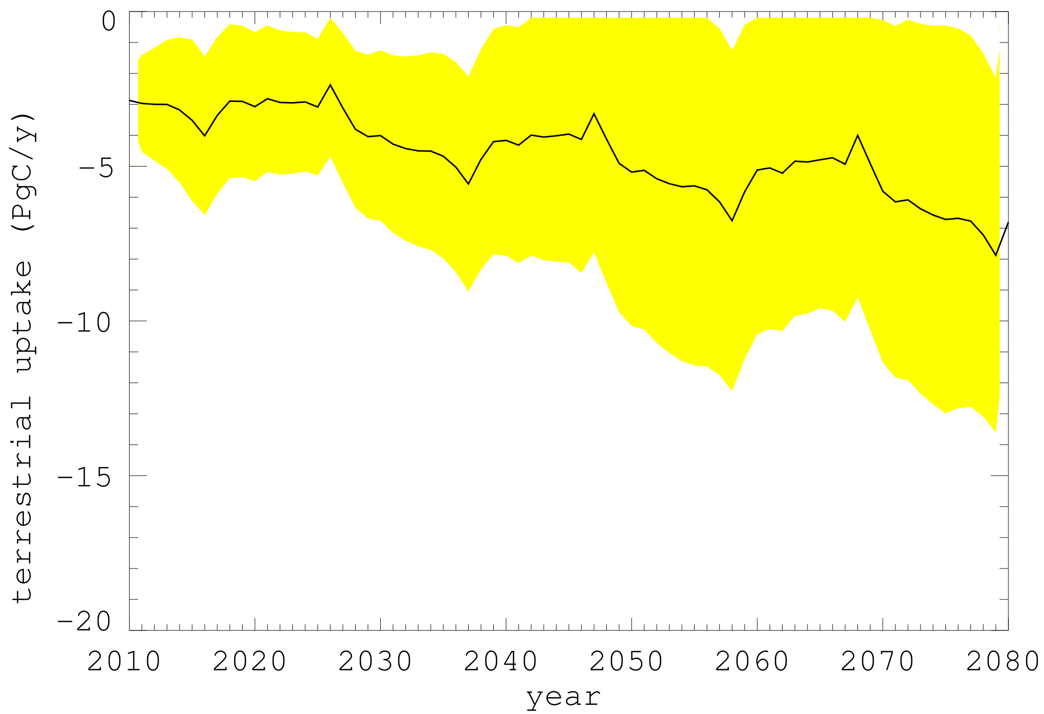
Technical Details

- Need derivatives of outputs of simple model and global model with respect to their inputs;
- Simple model can be differentiated by hand;
- Global model differentiated by the software "Transformation of Algorithms in FORTRAN" `http://www.fastopt.com`.

$$\text{Uncertainty(uptake)} = J \times \text{Uncertainty(parameters)} \times J^T$$

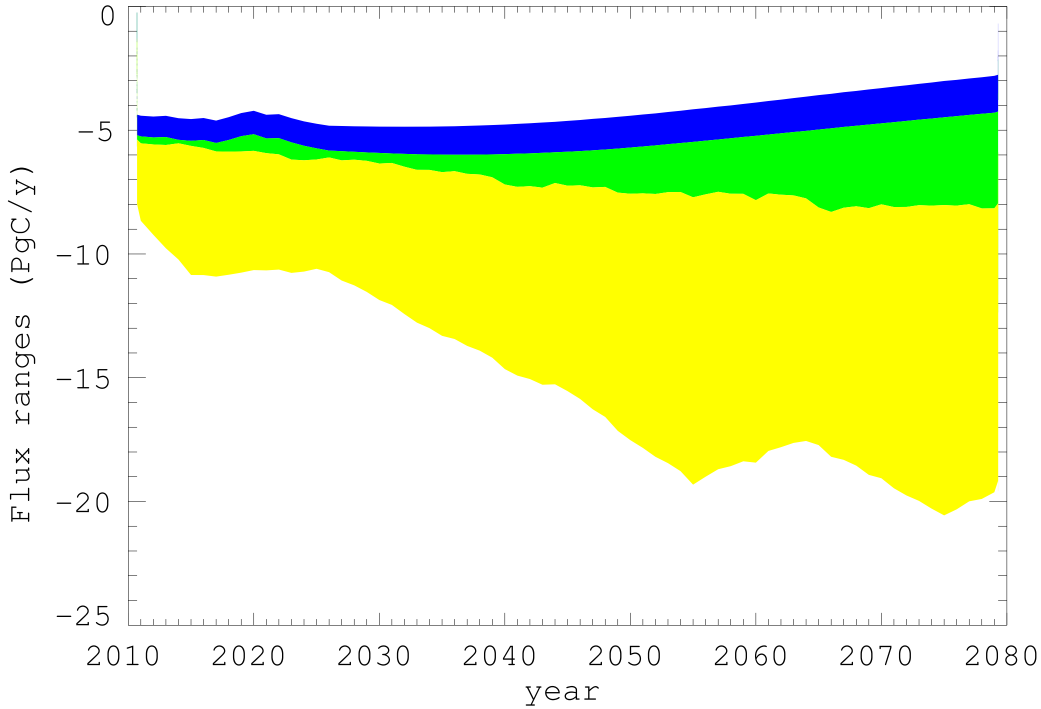
where J is derivative and T is transpose.

Uptake from Prior Model



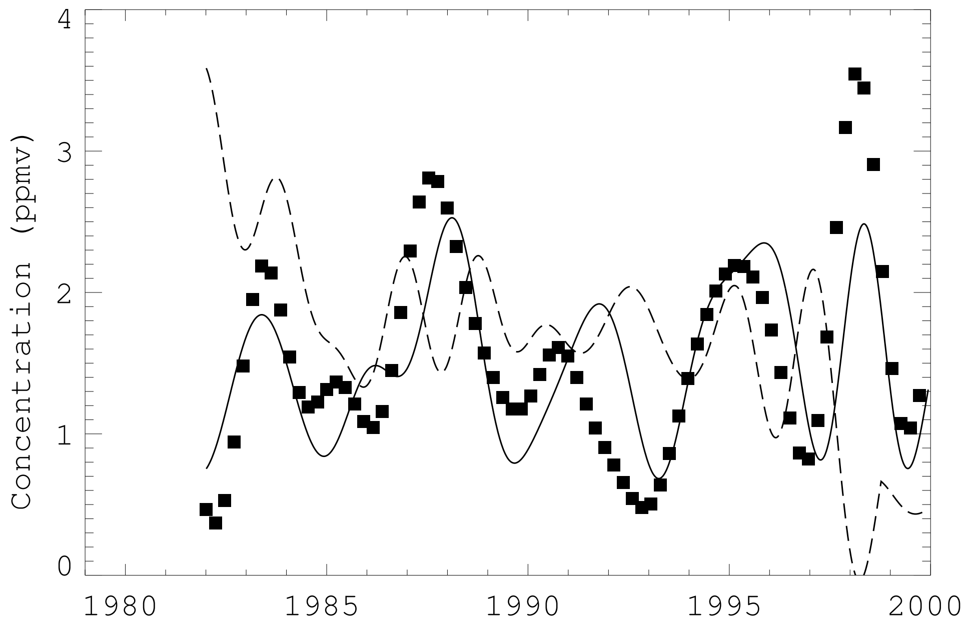
Terrestrial uptake (no climate change) from prior model and its 90% confidence interval. Uptake is anchored at its 2000–2010 value.

Comparison with Other Uncertainties



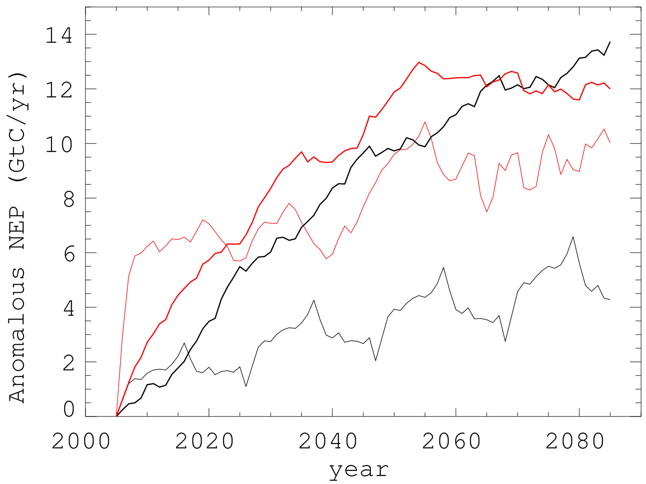
Range of uptakes, blue = ocean, green = land from IPCC models. The yellow band represents the 90% confidence interval of the uncertainty in the simple model.

Fitting Atmospheric Growth Rate



Smoothed global growth rate, squares = obs, dashed = prior, solid = optimised. Note the great improvement in phasing with the optimisation.

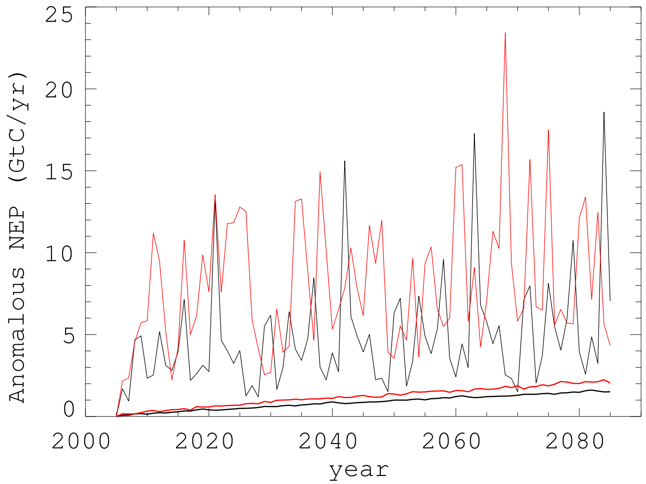
Comparing Uptakes



Decadal mean δNEP, 2000–2090. Black lines = current climate, red = climate change. Thin lines = prior parameters, thick = optimized.

- $\delta\text{NEP} = \text{NEP} - \overline{\text{NEP}}(t = 2000 - -2010)$
- Unrealistically rapid increase;
- High $\frac{\partial \text{GPP}}{\partial \text{CO}_2}$: 0.3PgC/yr/ppm cf FULLBETHY = 0.23 and LPJ = 0.19. ORCHIDEE anyone?

Comparing Uncertainties



Uncertainty in decadal mean δNEP, 2000–2090. Black lines = current climate, red = climate change. Thin lines = prior parameters, thick = optimized.

- Uncertainty in δNEP calculated as $\mathbf{C}(x) = J\mathbf{C}(p)J^T$ where $J = \frac{\partial x}{\partial p}$ and $\mathbf{C}$ is covariance;
- Uncertainties completely dominated by climate change;
- Partially reflects small uncertainty on photosynthesis parameters.

For those who prefer numbers

Case	sum (PgC)	Uncertainty (1σ PgC)
prior no-clim	278	126
prior clim-change	656	1141
optimized no-clim	717	78
optimized clim-change	799	107

Value and uncertainty for integrated $\delta\mathrm{NEP}$ from 2000–2090

Climate Feedback Parameter

- $G = \frac{\sum \delta \mathrm{NEP(climate)}}{\sum \delta \mathrm{NEP(noclimate)}}$;
- Can calculate $\frac{\partial G}{\partial p}$ (unpleasant) and hence uncertainty of G;

$$\sigma(G) = \sqrt{\nabla_p G \mathbf{C}(p) \nabla_p G^T}$$

- For $\mathbf{C}(p)$ diagonal (prior) this is simple sum.

Equations again

$$\text{SoilOutgassing} \propto \text{SoilPool} \times \omega^{\kappa} \mathbf{Q}_{10}^{T_a/10}$$

where $\omega =$ soil moisture and $T_a =$ air temperature.

$$\frac{\partial \text{SoilPool}}{\partial t} = \mathbf{K} \times \mathbf{LitterDecomposition} - \text{SoilOutgassing}$$

- $\sigma(G) = 3.54$ for prior;
- 80% from κ with rest from $\mathbf{Q}_{10}$ and $\mathbf{K}$;
- $\sigma(G) = 0.04$ posterior.

Conclusions

- Tangent linear models are fun;
- The carbon-cycle/climate feedback uncertainty is very large even within one model;
- For BETHY the sensitivity of respiration to soil moisture is the biggest contribution to uncertainty;
- The atmospheric record is sufficient to constrain this aspect of model dynamics.

Using Data Assimilation with Model Ensembles

- Work very much in progress;
- Similar efforts in physical climate.

Transcom

- How much uncertainty in inversions due to transport?
- Three phases: compare forward models, run known tracers, compare inversions;
- Law et al., Tellus, 1996, Denning et al., Glob. Biogeochem. Cyc. 1999, Gurney et al., Nature 2002, Baker et al., Glob. Biogeochem. Cyc., 2006.

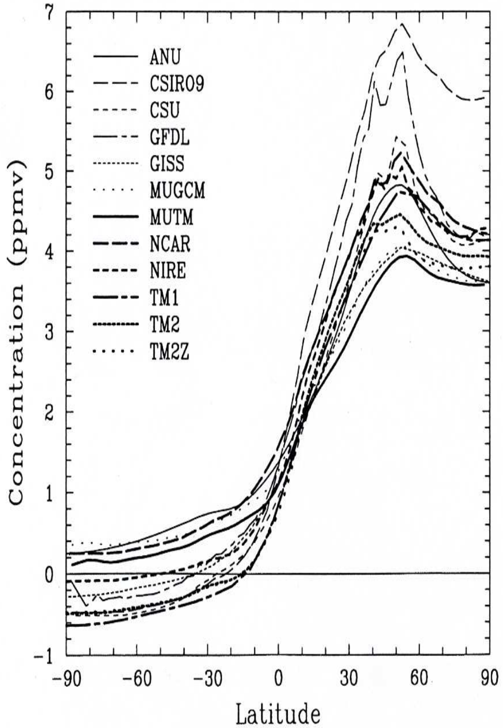


Fig. 3.1: Zonal annual surface mean concentration in ppmv due to fossil emissions.

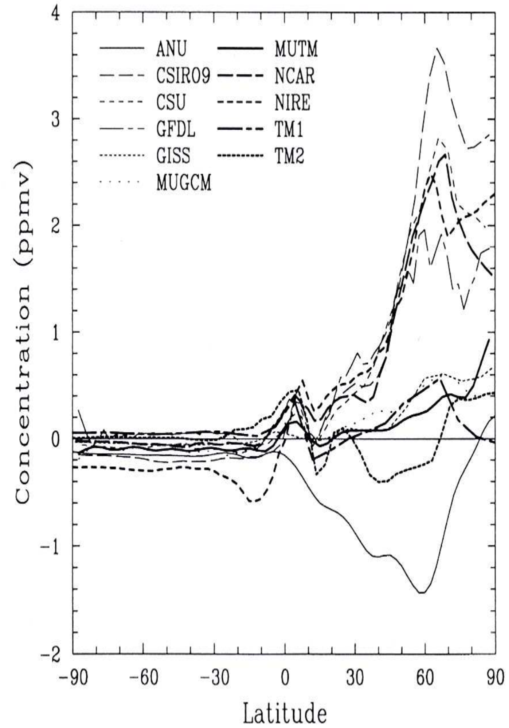


Fig 4.6: Zonal annual mean concentration in ppmv for the biosphere experiment.

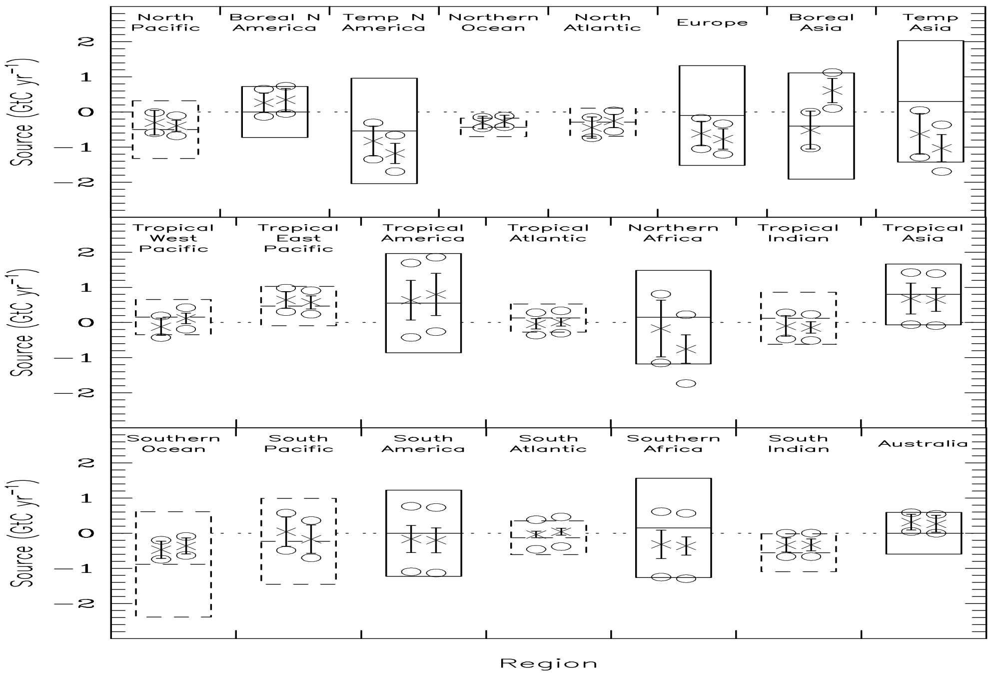
Zonal mean concentration from fossil fuel source

Zonal mean response to annually balanced biosphere source

Transcom III

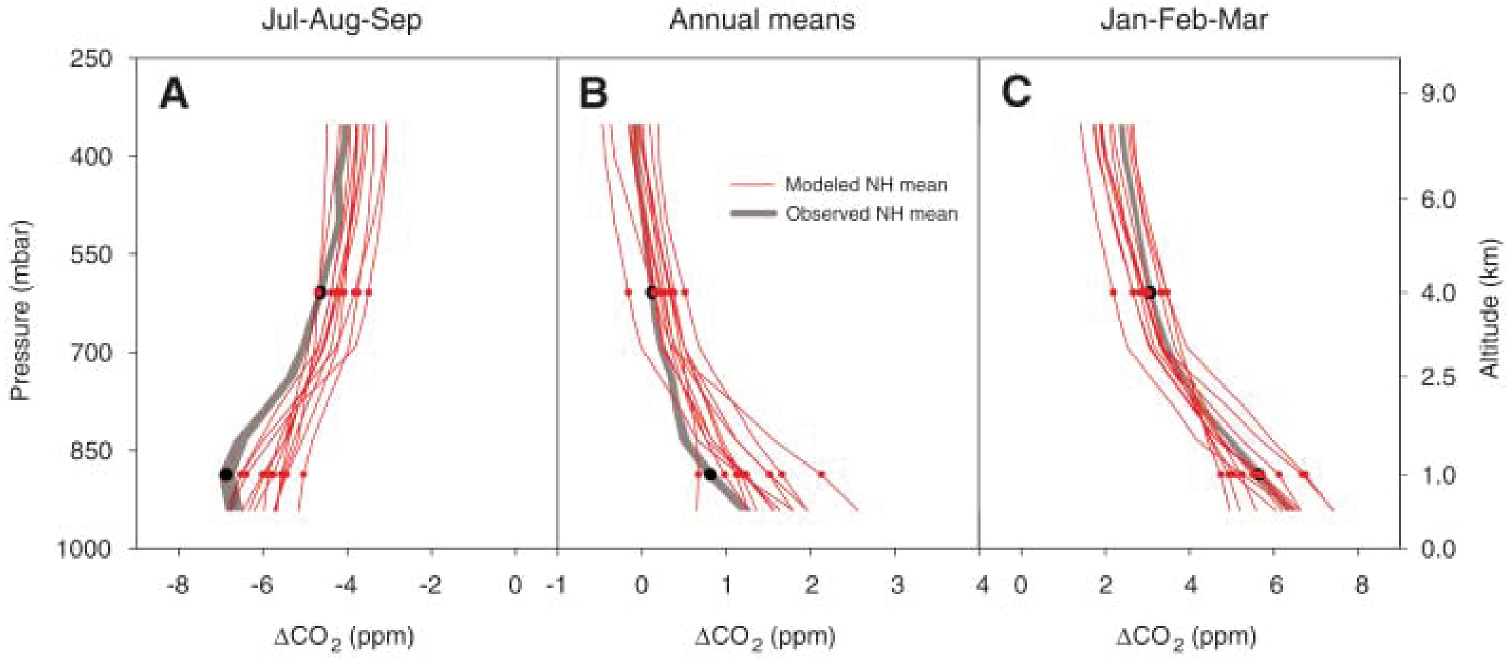
- Run inversions changing only response functions from different models;
- Data and uncertainties, prior and uncertainties and algorithm fixed.
- Annual mean case: 26 response functions, 17 models;
- Seasonal and interannual cases: 268 response functions, 12 models.

Gurney et al., Nature, 2002



Inversion results for the control (left bar) and no-biosphere (right bar). Mean fluxes are the 'X'. Positive = source. Prior flux and uncertainty: horizontal bar and boxes (land in green, ocean in blue). Within model uncertainty = circles, between model uncertainty = length of vertical bars. Regions are shown in their approximate north-south and east-west relationship.

Impact of Vertical Transport on Inversions



Notes

- Stephens et al., Science, 2007;
- Compared different models against independent climatology of profiles;
- Did not consider posterior uncertainty;
- Main point that not all models are equal.

The Approach

- Use statistical techniques to choose among an ensemble;
- Include the choice of model as an extra unknown;
- Produce a PDF among the models;
- Weight means etc by this PDF

Set-up

- Prior PDF for fluxes and data as for Gurney et al., Nature, 2002, i.e Gaussian;
- Uniform prior distribution for model choice (every model equally likely);
- Relative probability for each model depends on overlap between simulation and data (size of black triangle).

Making a long story short

- Collection of linear models $\mathbf{H}_1 \ldots \mathbf{H}_N$s
- Let $G(\vec{\mu}, \mathbf{C})$ be Gausian distribution with mean $\vec{\mu}$ and covariance $\mathbf{C}$

$$P(\vec{x}, \mathbf{H}) \propto G(\vec{x} - \vec{x}_0, \mathbf{C}(\vec{x}_0) * G(\vec{y} - \mathbf{H}\vec{x}, \mathbf{C}(\vec{y}))$$

$$P(\mathbf{H}) = \int d\vec{x} P(\vec{x}, \mathbf{H})$$

Skipping the painful algebra

$$P\mathbf{H} \propto \left[\det \mathbf{H}\mathbf{C}(\vec{x}_0)\mathbf{H}^T + \mathbf{C}(\vec{y})\right]^{-0.5} \exp -\frac{1}{2}(\vec{y} - \mathbf{H}\vec{x}_0)^T \left[\mathbf{H}\mathbf{C}(\vec{x}_0)\mathbf{H}^T + \right.$$

$$\sum P(\mathbf{H}_i) = 1$$

- Can be calculated without ever performing an inversion;
- Similar to maximum value of $P(\vec{x}, \mathbf{H})$.

Sample of Tabulated values

Model	$P(\mathbf{H})$
JMA-CDTM.maki	0.66
MATCH.law	0.29
MATCH.bruhwiler	0.02
SKYHI.fan	1e-7

- Unrealistically strong discriminant (7 orders of magnitude)
- Problem over-determined so many obs left to discriminate among models
- Model error $(\mathbf{C}(\vec{y})$ should not be independent.

Applications and Problems

- Using data assimilation to improve model structure as well as parameters;
- Choosing among an ensemble of models;
- Unusually dependent on proper formulation of uncertainties.

Summary for Today

- Climate DA is possible and can help us improve climate prediction;
- We can learn a lot by propagating uncertainty into prediction;
- We can extend DA beyond improving a particular model into the domain of model choice but it’s not easy in real cases.

Overall Summary

- Data assimilation best thought of as a statistical problem;
- Watch the statistics of inputs and results carefully;
- There's a lot to gain by considering more than just the best guess for unknowns and simulations;
- The basic theory is flexible enough for interesting extensions, like model choice but sometimes you have to go back to first principles.