TM5 AT NOAA CMDL

Wouter Peters, Lori Bruhwiler, John Miller, Adam Hirsch, Kevin Schaefer, Arlyn Andrews, Pieter Tans

Activities

- TM5 projects at CMDL
- Transcom Continuous
- Ensemble Data assimilation
  - TM5 + CO_2 tracer version
  - TM5 + CH_4 tracer version (*)
  - TM5 + CO full chemistry version (*)

Our resources

- 2 modelers
- a big computer: ???
  - ??x2.?Ghz
  - ?? Tb fast access storage
  - intel compiler
- lots of data

TM5 project:

- TransCom continuous project
  - submissions coming in now
  - TM5 zoomed over US

TM5 project:

- Build a near real-time data assimilation system for North America to estimate regional fluxes of CO_2
  - North American Carbon Program
  - NOAA CMDL commitment to yearly updated maps of flux + uncertainty

TM5 project:

- Make a TM5 preprocessing system for NCEP meteorology
    - freely available, in-house
    - possibly nest to 20kmx20km

TM5 project:

- Model geologic sources of methane forward using GIS-derived emission distribution and estimated source strengths
  - stochastic approach
  - analyze gradients, signals at sites
  - compare to other sources

TM5 project:

- replace vertical diffusion scheme by simple 'mixed layer model'
  - assumes well-mixed below PBL height (h), concentration controlled by surface and entrainment flux and dh/dt
  - Diffusion does not work for non-convective PBL's anyway...?

TransCom Continuous

- Designed to compare models-to-models and models-to-observations at fine time scales

- Continuous recording of several tracers, and meteorological parameters for the year 2002 (2003 to be added)

- For 273 surface sites, and 89 surface-500 hPa sites

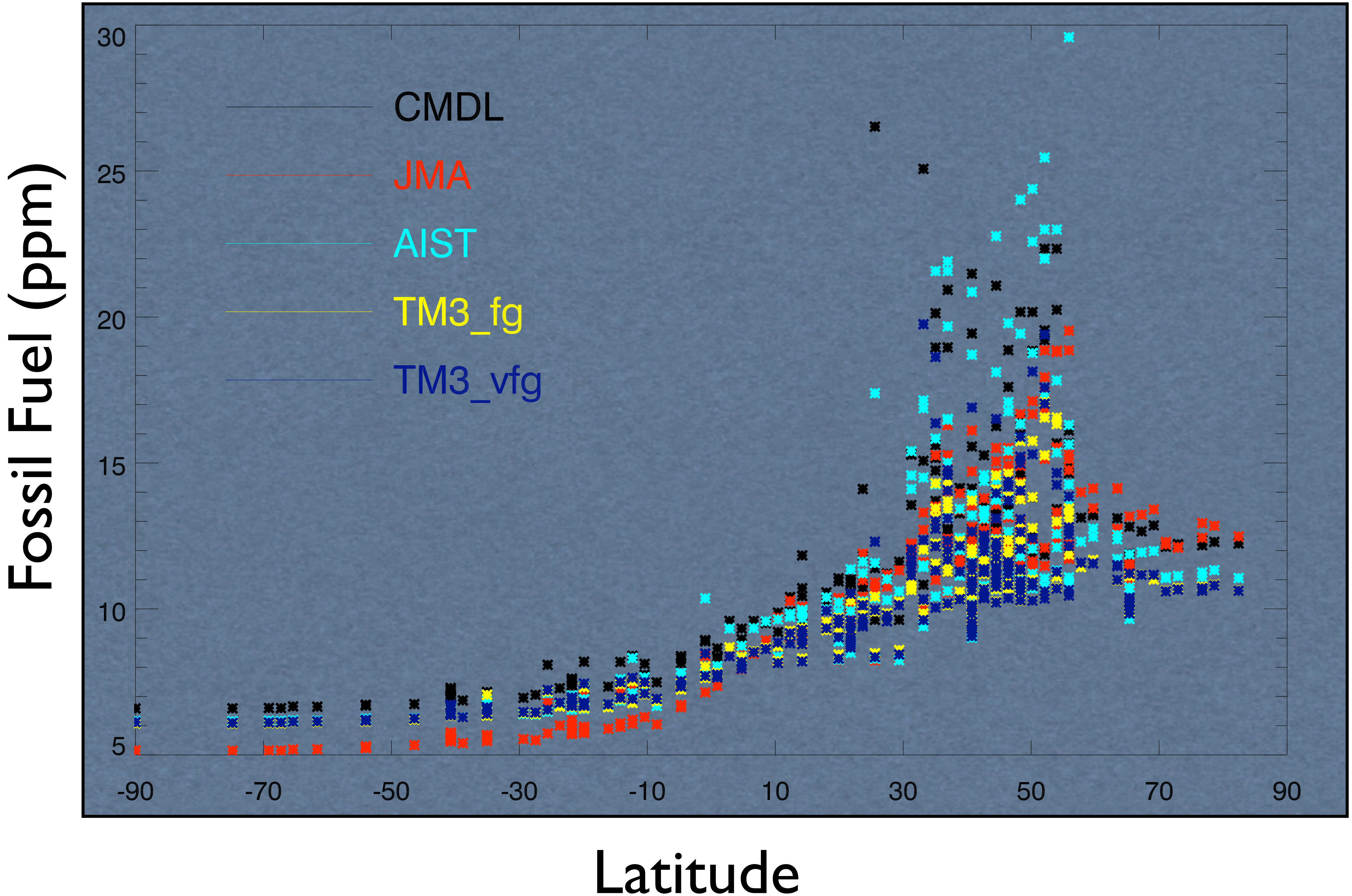
Tracers

Tracer name	Description	Flux time resolution
SiB	SiB biosphere model fluxes for 2002	hourly
SiB_day	SiB model daily average fluxes	daily
SiB_mon	SiB model monthly average fluxes	monthly
CASA	CASA biosphere fluxes with diurnal cycle	3 hourly
CASA_mon	CASA monthly fluxes	monthly
SF6	SF6 emissions	constant
radon	Radon emissions	constant
fossil98	Fossil emissions for 1998	constant
Taka02	Takahashi ocean fluxes, 2002 compilation	monthly

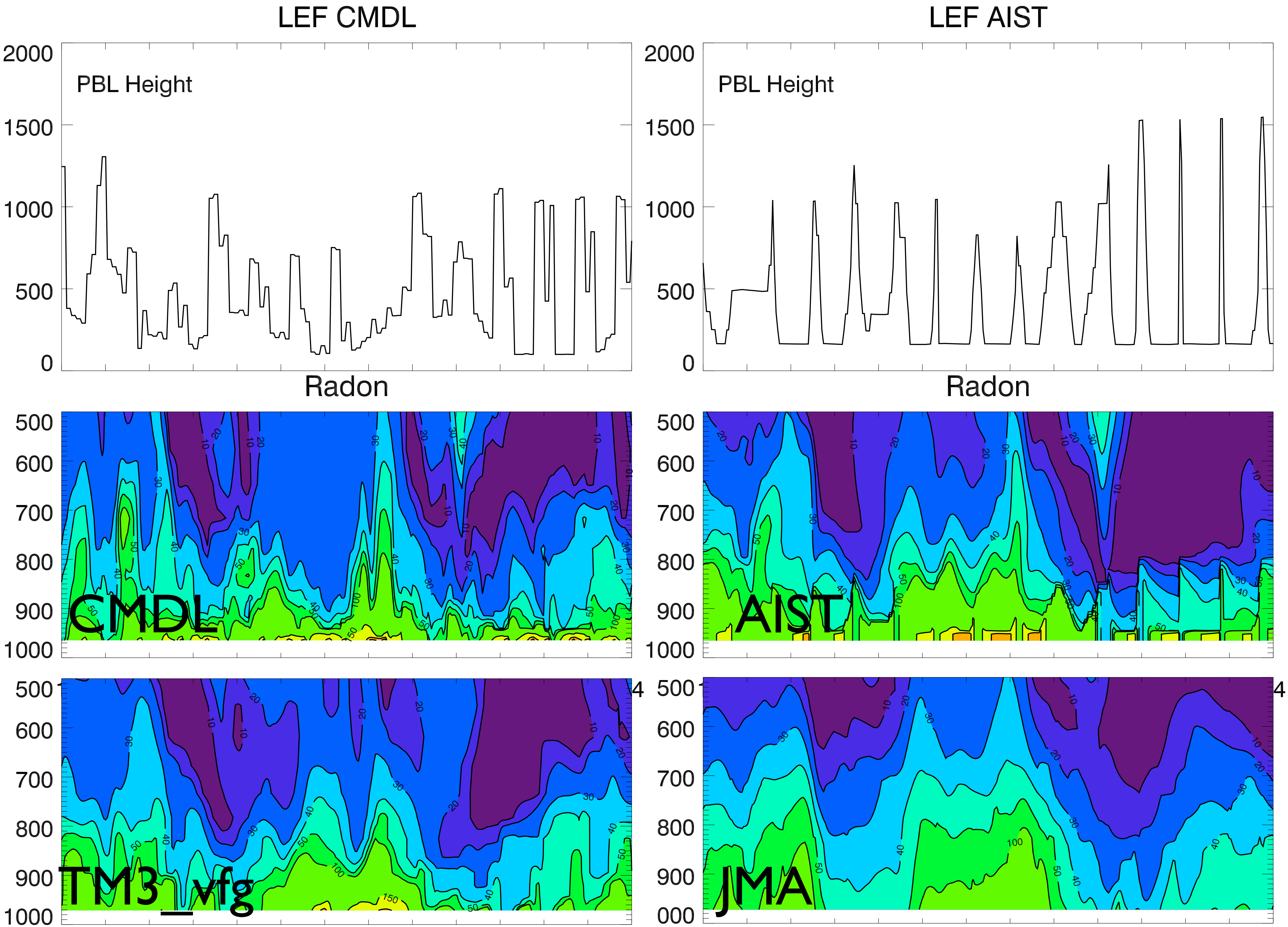
What's there

- CMDL, TM5 @ (6x4),(3x2),(1x1)x25, ECMWF
- MPI-Jena, TM3@1.25x1.25x??, NCEP
- AIST, STAG @ 1.25x1.25x60, ECMWF
- JMA, CDTM @ 2.5x2.5x32, JMA analysis
- LSCE, LMDZ @ ??
- To come: Denning, NASA, Fan, CSIRO, ...

A first look at the results...



A first look at the results...



Inverse modeling

$$J = (\mathbf{y}^{\circ} - \mathcal{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}^{\circ} - \mathcal{H}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}^{\mathbf{p}})^{\mathrm{T}}\mathbf{P}^{-1}(\mathbf{x} - \mathbf{x}^{\mathbf{p}})$$

Observations

first guess

- Matrix solutions to linear problem
- Adjoint + 4d-var to minimize cost function
- Kalman filter sliding window for efficiency
- Ensemble Kalman Filter

Data assimilation: 2 steps

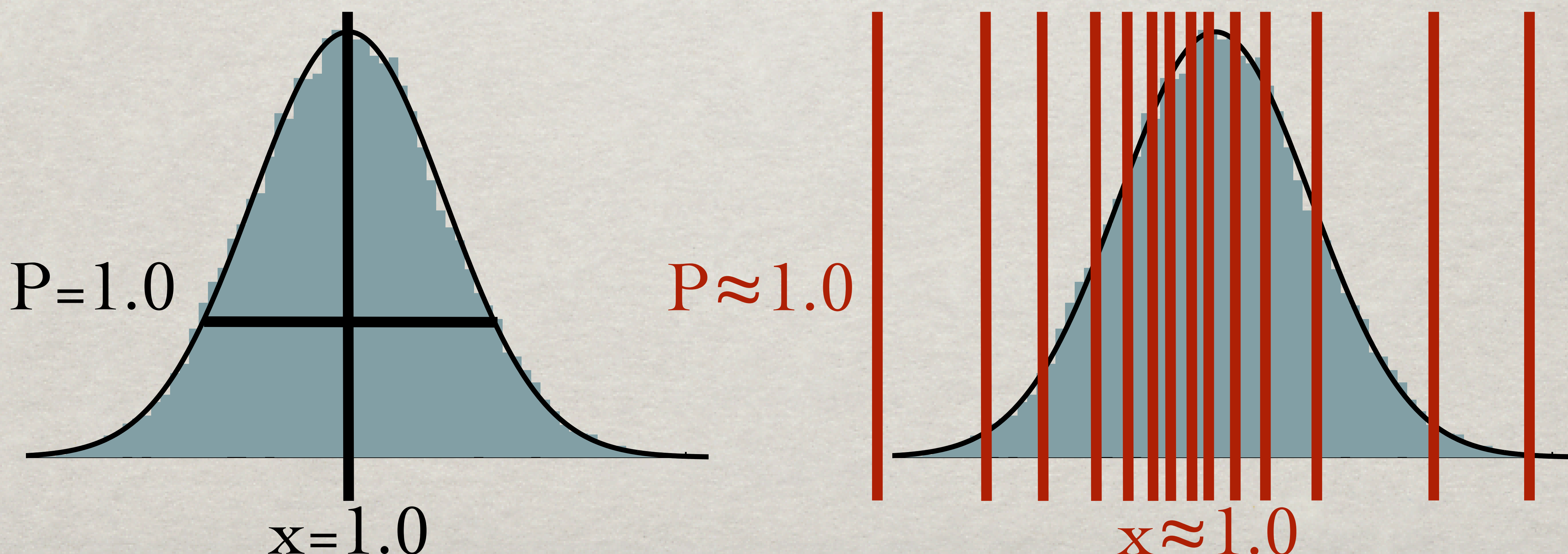
- 1) **Forecast**
  - take old state (fluxes,weather,CO_2 columns,...)
  - combine with a model of state propagation
  - forecast new state

New!

- 2) **Optimization**
  - sample forecast state
  - combine with observations
  - calculate optimal state

Ensemble DA

- Do not describe the PDF of the fluxes by its mean **x** and its covariance **P**, but create an ensemble of fluxes to reflect that same information:



Ensemble DA

- Predict concentrations for all ensemble members that span the prior PDF
- Optimize across the ensemble to find optimum fluxes
- Optimized ensemble represents posterior PDF

Ensemble DA (+)

- no base functions needed
- no adjoint needed
- only forward model runs
- weakly non-linear problems possible
- fully parallel

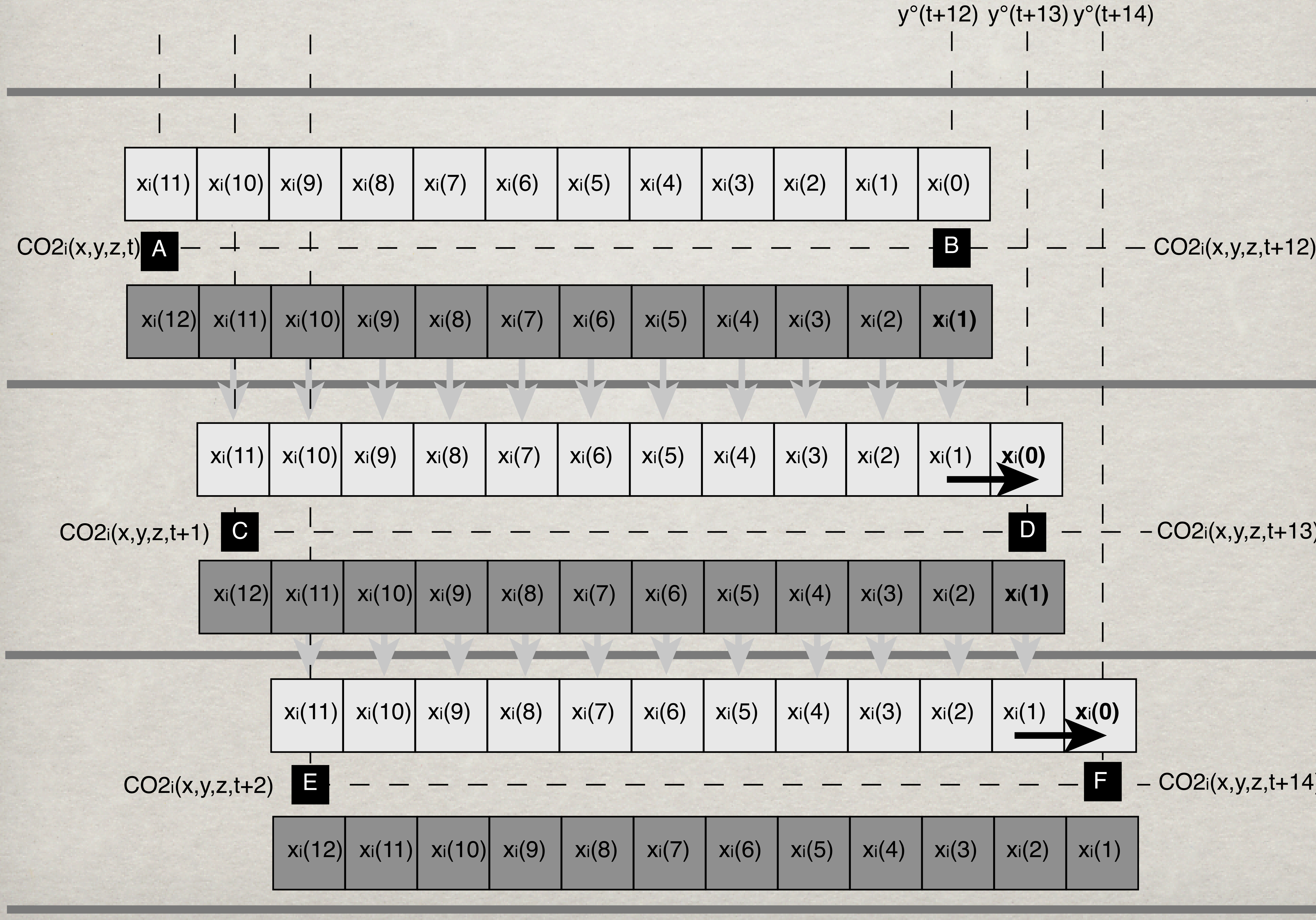
Ensemble DA (-)

- Success depends on statistics of ensemble
- Statistics depend on structure of **P**
- **P** depends on model and assumptions
- Base functions replaced by N parallel runs
- Posterior **P** has only N nonzero eigenvalues

TM5 version

- Massive parallel with 500 tracers on 50-120 processors, transport only
- Two new modules to coordinate assimilation code/ TM5 (subroutine)
- Model loops over one week timesteps, with several weeks of runtime in each loop
- All processors read/write, savefiles as binaries

Flow chart

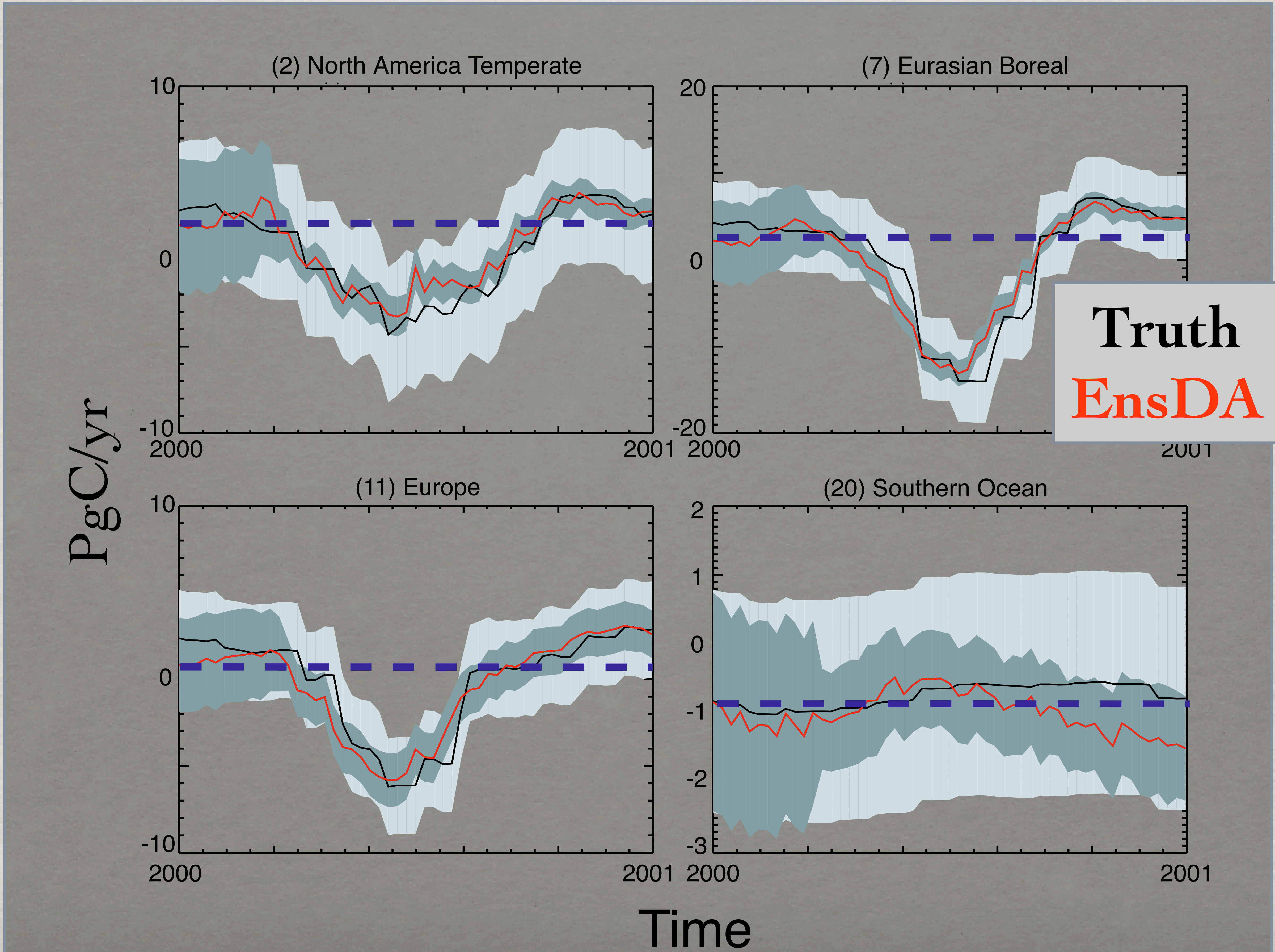


y°(t+12)
y°(t+13)
y°(t+14)
xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1) xi(0)
CO2i(x,y,z,t)
A
B
CO2i(x,y,z,t+12)
xi(12) xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1)
xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1) xi(0)
CO2i(x,y,z,t+1)
C
D
CO2i(x,y,z,t+13)
xi(12) xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1)
xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1) xi(0)
CO2i(x,y,z,t+2)
E
F
CO2i(x,y,z,t+14)
xi(12) xi(11) xi(10) xi(9) xi(8) xi(7) xi(6) xi(5) xi(4) xi(3) xi(2) xi(1)

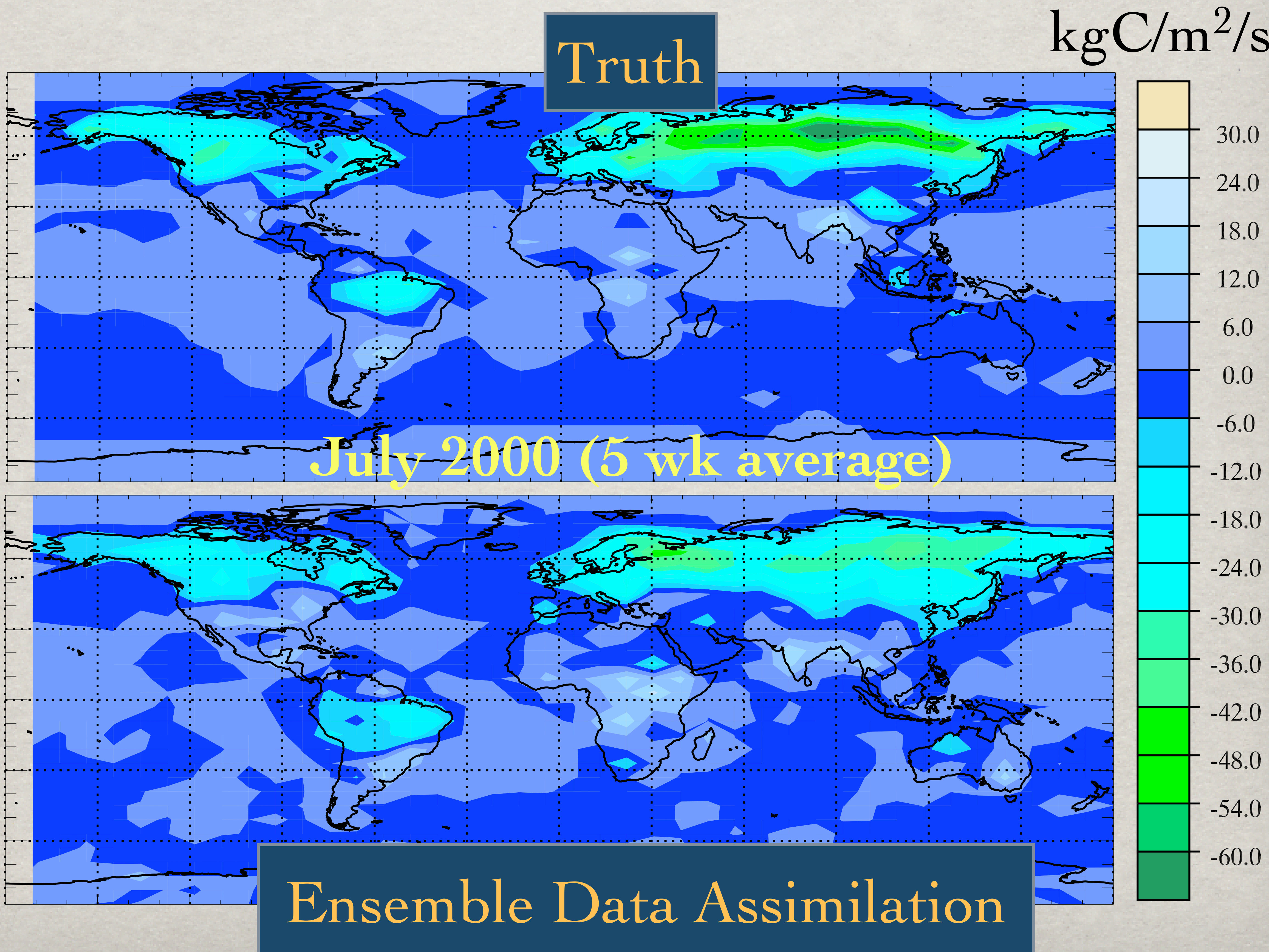
Pseudo data tests

- Global fluxes @ $9° \times 6°$ estimated each week
  - 1200x52 =62.400 unknowns
- from year 2000 NOAA CMDL flask network (no towers)
  - ~4800 observations
- Correlation structure imposed on solution
- No ‘bottom-up’ prior fluxes, M=I
- Covariances prescribed, not propagated

Pseudo data tests



Pseudo data tests



Summary

- Ensemble data assimilation offers an alternative way to optimize fluxes

- Large flexibility in # parameters, # and type of observations

- Efficient algorithm, no base functions needed, partial covariances returned

- Lots of CPU needed, more 'engineering' choices to be explored

Peters, W., J.B. Miller, J. Whitaker, A.S. Denning, A. Hirsch, M.C. Krol, D. Zupanski, L. Bruhwiler, and P.P. Tans, "An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations" *J. Geophys. Res.*, *submitted*, 2005.

- Terrestrial CO_2 fluxes vary at temporal and spatial scales that are beyond our capacity to observe from the atmosphere
- The mean CO_2 flux can only be observed after substantial integration of the signal
- Observations on the continent can only be interpreted when the smaller space and time variations are accounted for

- A large part of the spatial variability of CO_2 fluxes is due to the underlying biome type, and its characteristics
- A large part of the temporal variability of CO_2 fluxes is due to the effect of sunlight, temperature, soil moisture,...

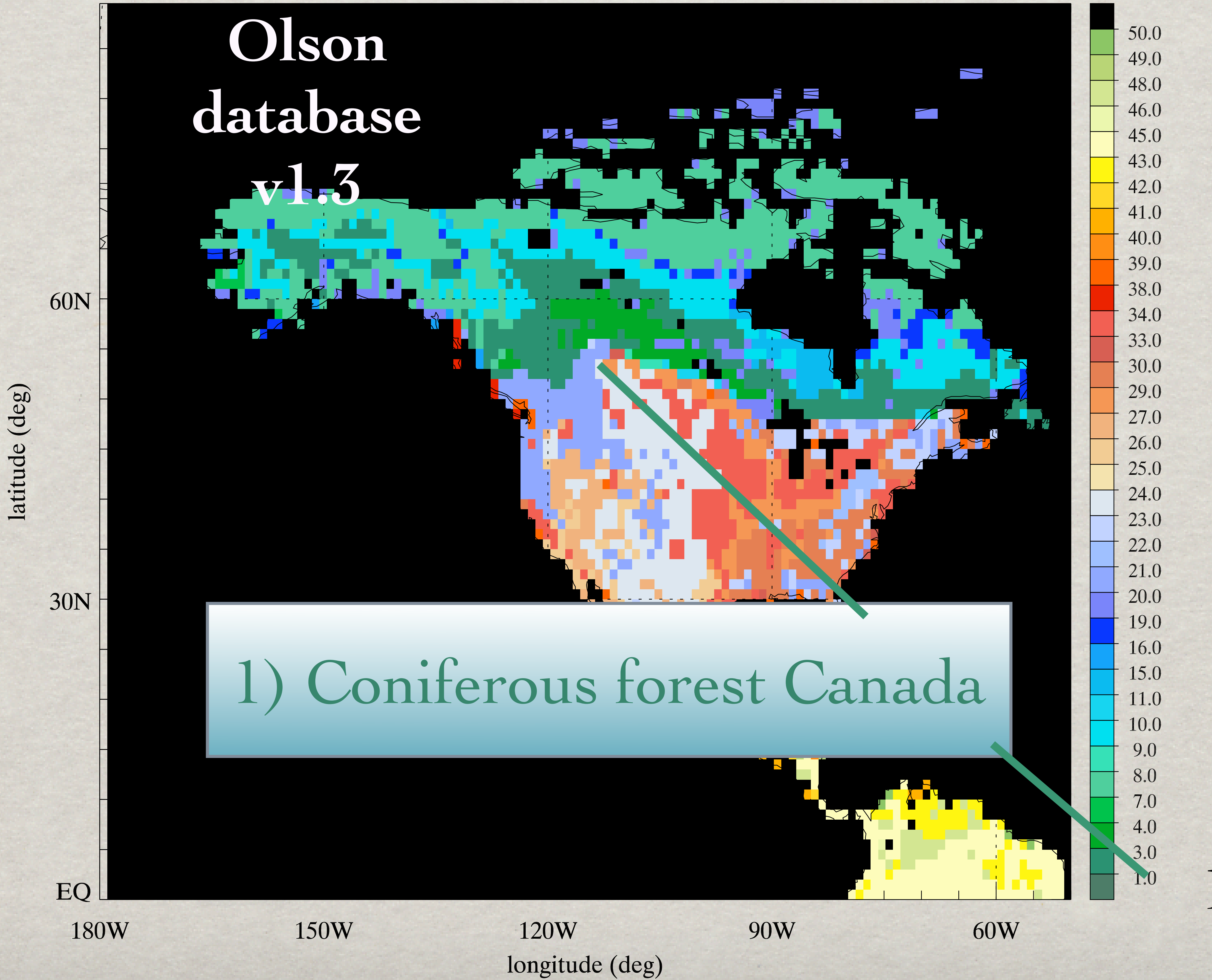
Gerbig et al, 2003 / Harvard / CSU approach:

- Solve for the (stable, predictable) response of an (known, observable) ecosystem to (variable, known) forcing, to create fluxes that are:
  - Resolved at appropriate scales
  - Optimized with observations

From biosphere to fluxes

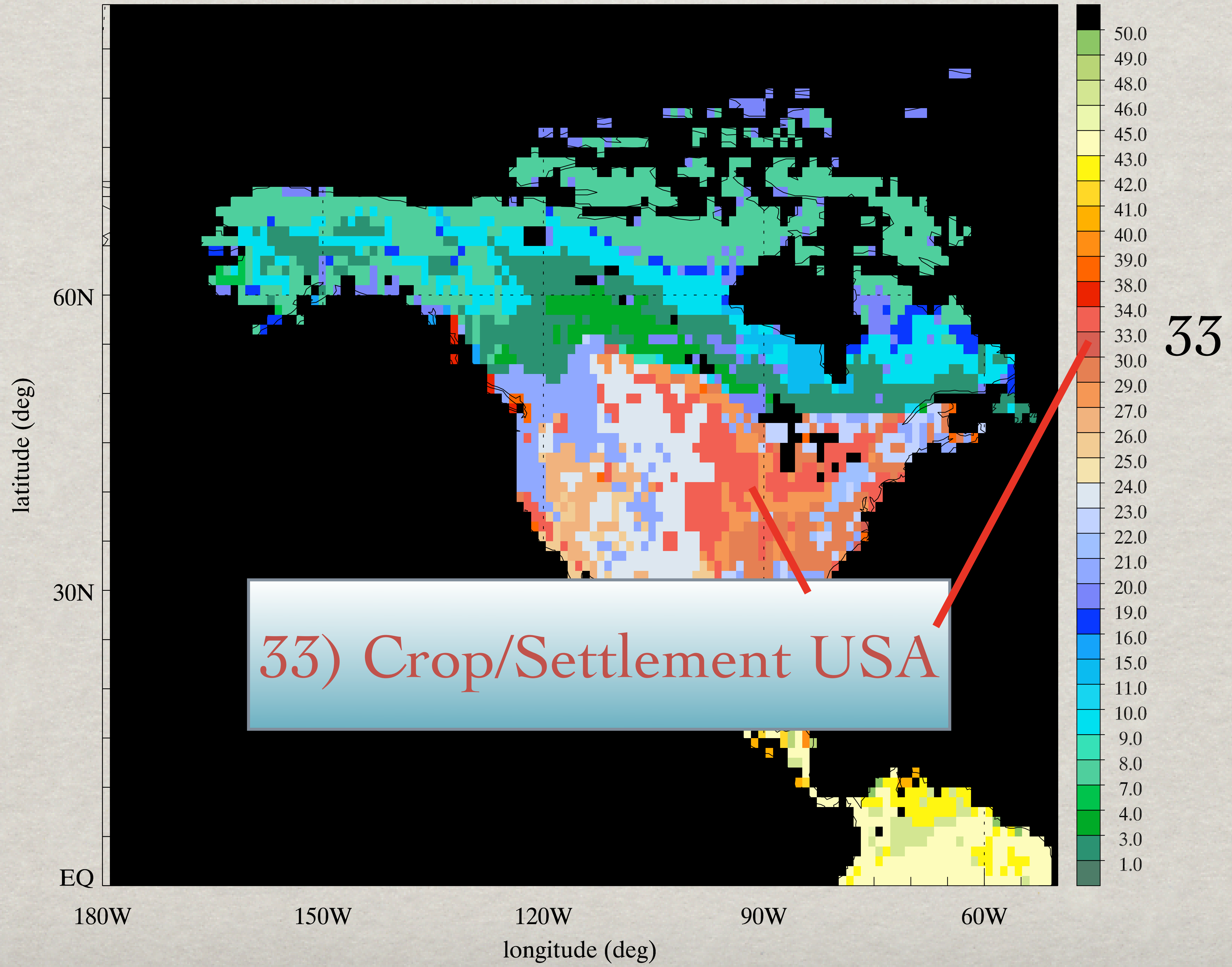
- Flux parameterization (Gerbig)
  - T, q, short wave radiation
- simple biosphere model (VPRM)
  - + LAI, NDVI, PAR
- full biosphere model (SiB, CASA)
  - + GPP, R formulations, reservoirs

The ecosystem regions



1

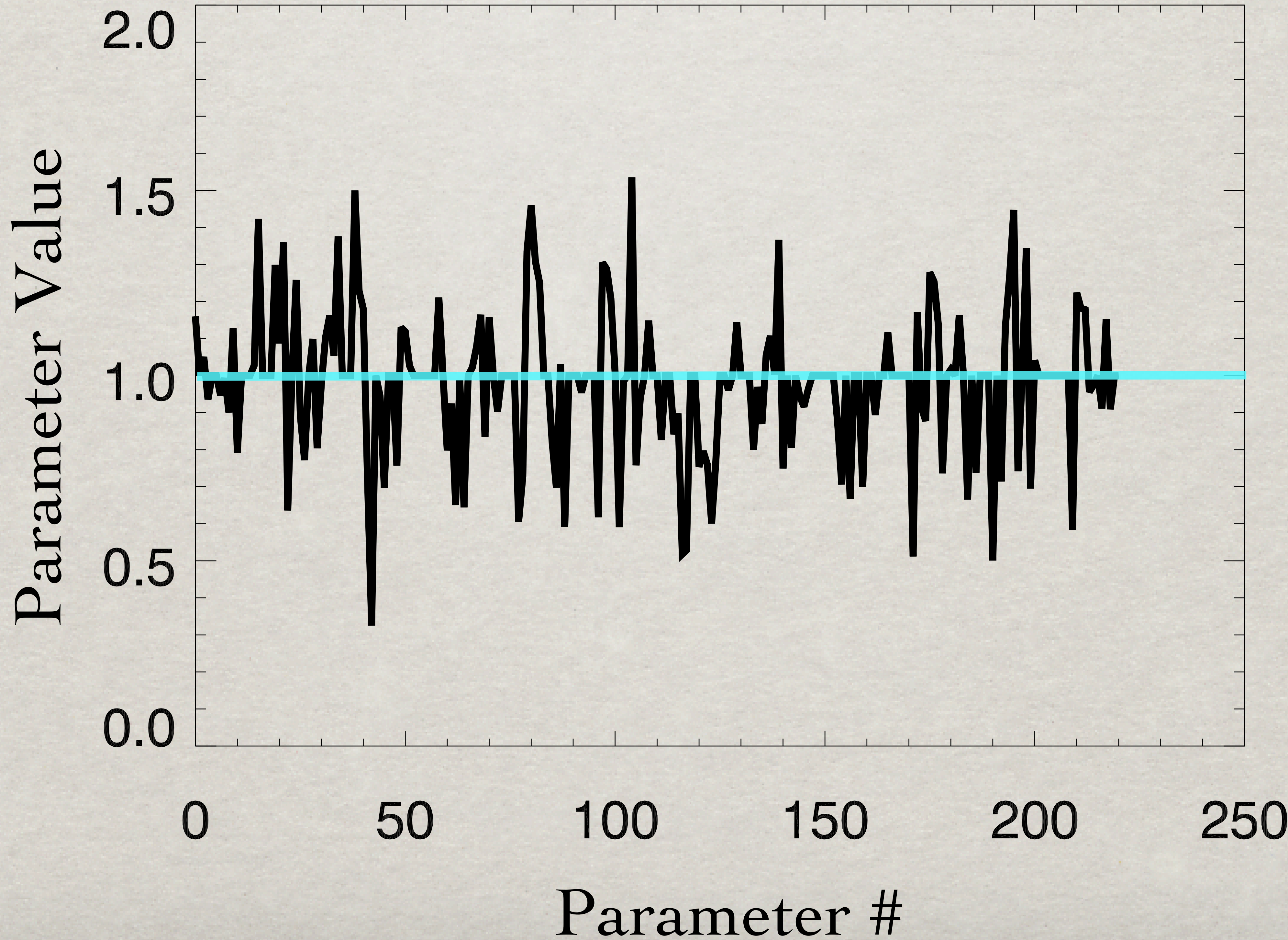
The ecosystem regions



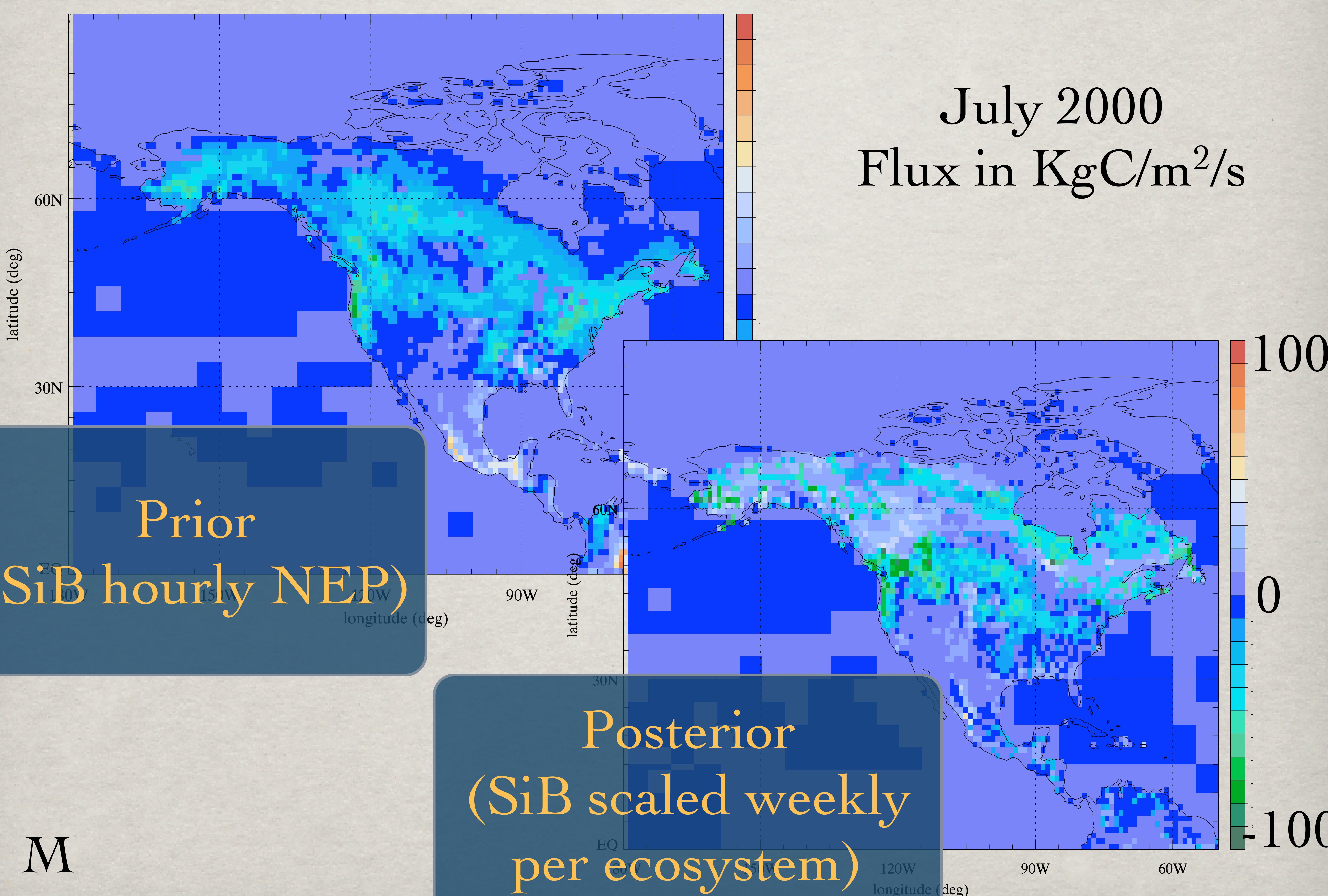
Data assimilation

- Prior flux = VPRM model driven by offline weather in TM5 as a function of 2 parameters for each ecosystem
- Posterior flux = VPRM model + TM5 weather +2 optimized parameter values for each ecosystem
- Ocean fluxes are now 22 regions and prescribed, but will become spatially explicit and a function of pCO2, SeaWifs, ...,?

Example: scaling of SiB NPP



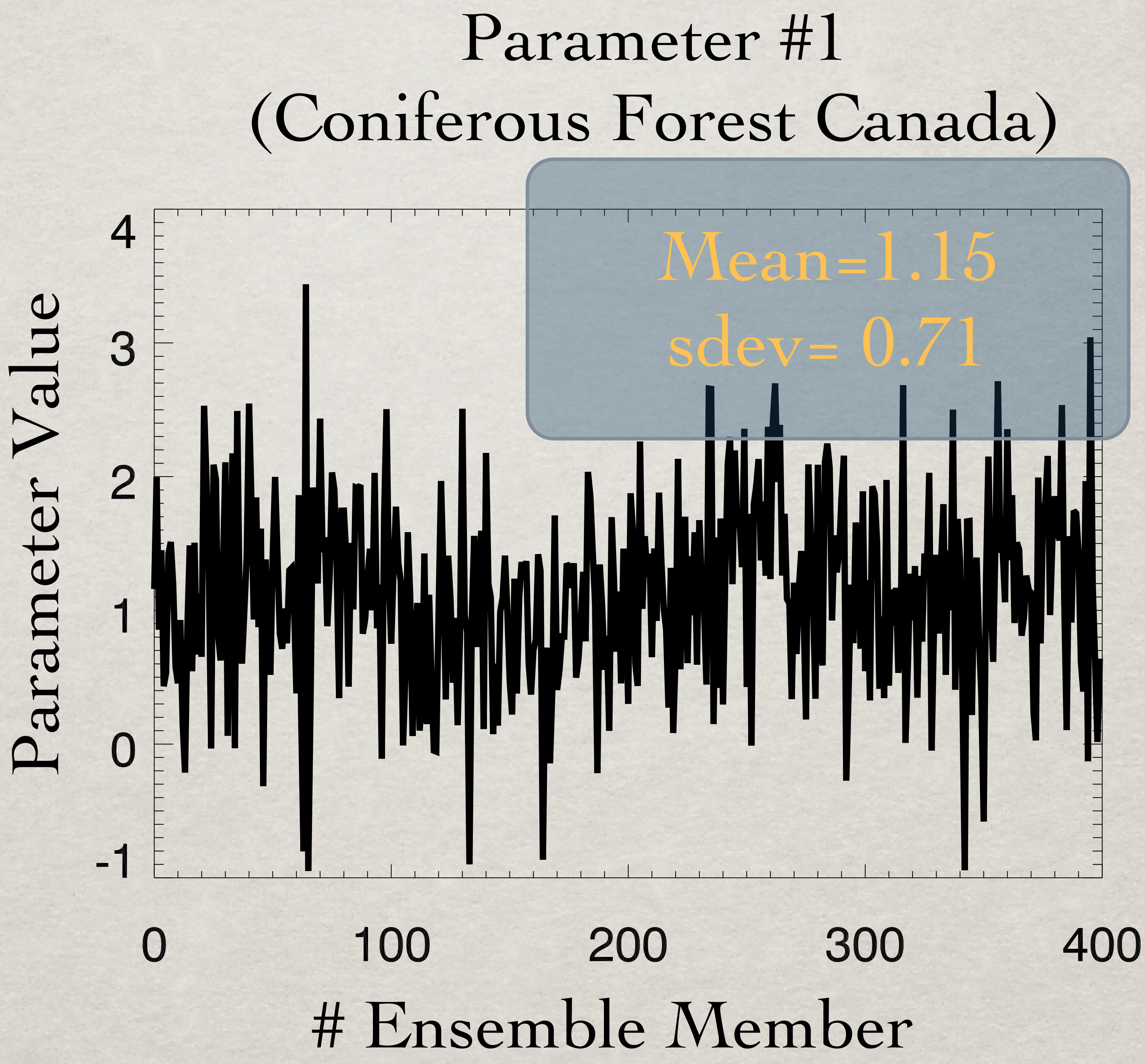
July 2000
Flux in KgC/m^2/s



Prior
(SiB hourly NEP)

Posterior
(SiB scaled weekly
per ecosystem)

M


Parameter #1
(Coniferous Forest Canada)
Mean=1.15
sdev= 0.71
Parameter Value
4
3
2
1
0
-1
0
100
200
300
400
Ensemble Member

June 24-27 hourly fluxes