Inferring Dinosaurs from their Footprints

Inez Fung

August 2 2013

NCAR ASP Graduate Colloqium: Carbon Climate Connections in the Earth System

Dinosaur footprint near Enciso, La Rioja, Spain http://commons.wikimedia.org/wiki/File:Enciso-dinosaur-footprint-detail.jpg

Simple Example

- **I have two independent, but different, pieces of information about T at a point: T_1, T_2.**
- **If the info are (assumed) perfect, then the best estimate of T is**

$$T_a = \frac{1}{2}(T_1 + T_2)$$

- **But the info is not perfect… different approaches for estimating T_a**

(1) Least Squares Method

- **Two observations to estimate T_{truth} *(unknown)***

$$T_1+\varepsilon_1;\ \ T_2+\varepsilon_2;\quad E(\varepsilon_1)=E(\varepsilon_2)=0;\ \ E(\varepsilon_1{}^2)=\sigma_1^2;\ \ E(\varepsilon_2{}^2)=\sigma_2^2$$

- **Find T_a=$T_{analysis}$: the best approx to T_t=T_{truth}**

$$T_a=a_1T_1+a_2T_2;\quad \underbrace{E(T_a)=E(T_t)}_{unbiased};\ a_1+a_2=1$$

- **Choose a_1 and a_2 to minimize RMS error in T_a:**

$$\sigma_a^2=E[(T_a-T_t)^2]=E[(a_1(T_1-T_t)+(1-a_1)(T_2-T_t))^2]$$

$$\frac{d\sigma_a^2}{da_1}=0\rightarrow a_1=\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2};\ \ a_2=\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\ \ ;\ \frac{1}{\sigma_a^2}=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}$$

(1) Least Squares - continued

$$T_a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} T_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} T_2$$

- T_a is the weighted average of T_1 and T_2
- If uncertainty in T_1 is large, then T_2 is given greater weight.

(2) Variational (cost function) approach

- **Minimize cost function J**

$$J(T) = \frac{1}{2}\left[\frac{(T-T_1)^2}{\sigma_1^2} + \frac{(T-T_2)^2}{\sigma_2^2}\right]$$

$$\frac{\partial J}{\partial T} = 0 \quad \text{for } T = T_a$$

Note: Weighting of info according to uncertainty

- **Bayesian inversion:**
  - Given $P(X|\Phi)$: prior probability of CO2 (X) for an (unknown) flux (e.g. from model)
  - Find $P(\Phi|X)$: posterior probability of flux given obs of CO2

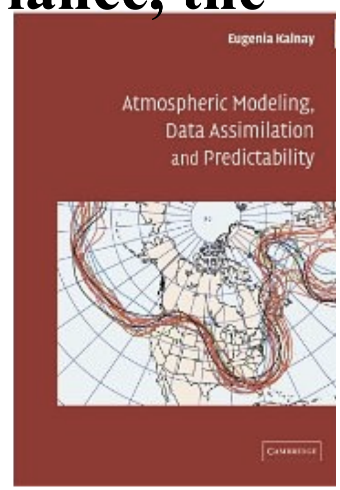
$$J(\Phi) = \frac{1}{2}\left[\frac{(X_{model}(\Phi) - X_{obs})^2}{\sigma_{obs}^{\ 2}} + \frac{(\Phi - \Phi_{prior})^2}{\sigma_{prior}^{\ 2}}\right]$$

(3) Simple sequential assimilation and Kalman Filter

- Let $T_b = T_1$ (background; prior knowledge); $T_o = T_2$ (obs)
- Analysis

$$T_a = T_b + W \underbrace{(T_o - T_b)}_{\substack{\text{obs innovation}\\ \text{obs increment}}}; \quad W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}; \quad \sigma_a^2 = (1 - W)\sigma_b^2$$

- The "analysis" is obtained by adding to the 1st guess (T_b, prior or background) the innovation, optimally weighted
- The optimal weight W is the background error variance relative to the total variance. The greater the background variance, the greater the info from the observations
- The precision (inverse of variance) of the analysis is the sum of the precision of the background and obs
- The error variance of the analysis is the background variance, reduced by a factor =1-optimal weight W



(3a) Carbon Data Assimilation: Kalman filter

Choose X=state vector = all the variables we are trying to estimate

At every assimilation time step n

1. **Gather observations Y_o^n**
2. **convert model variable to the form observed $H(X^n)$ (e.g.if X is CO2 conc;**
   i. **Atm model to forecast from last time step:** $$\underbrace{X_b^n}_{\text{background, forecast}} = \underbrace{\mathfrak{M}(X_a^{n-1})}_{\text{model to advance } \Delta t}$$
   ii. **Select station, average etc.**
3. **Optimize**

$$X_a^n = X_b^n + W\underbrace{(Y_o^n - H(X_b^n))}_{\text{obs innovation, obs increment}}; \quad W = \frac{(\sigma_b^n)^2}{(\sigma_b^n)^2 + (\sigma_o^n)^2};$$

$$(\sigma_a^n)^2 = (1-W)(\sigma_b^n)^2$$

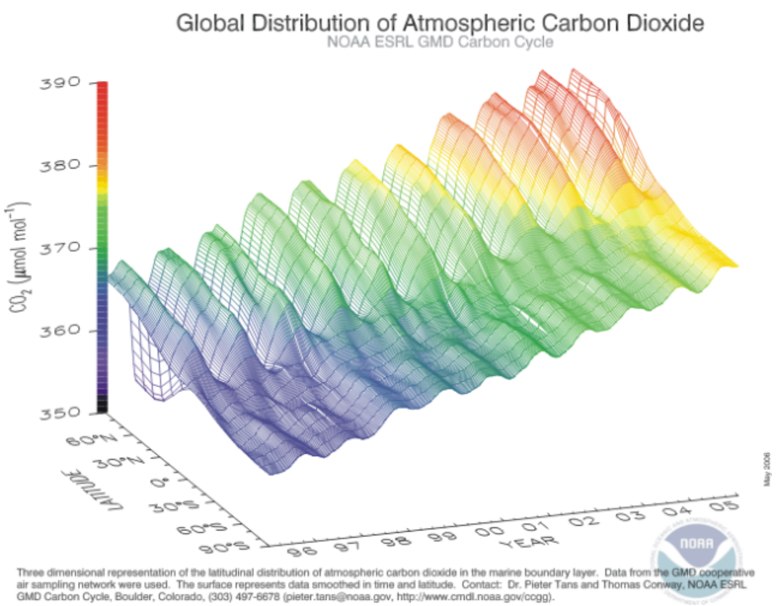
Estimating Carbon Fluxes

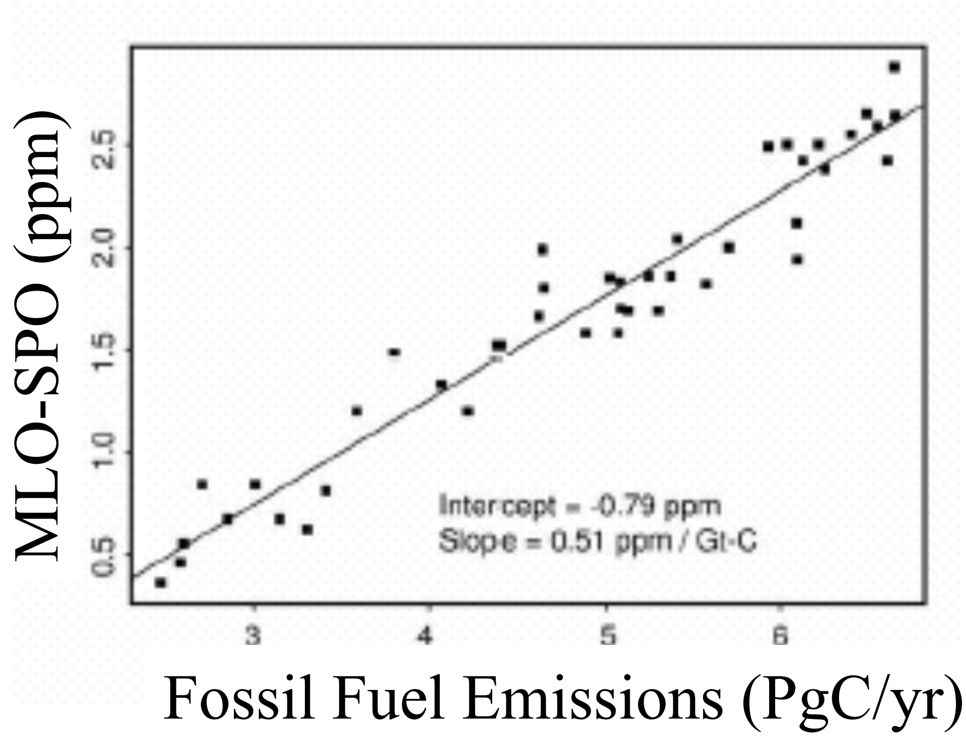
- **Most estimation of fluxes use some form of least squares, minimize some RMS metric** (e.g. Tans et al. 1990)
- **Inversions typically use variational approach. Typically done once for the entire observing period. Error is constant through time** (e.g. TransCom, Bousquet et al. 2000,Gurney et al. 2003, …)
- **Data assimilation is done every assimliation time step (e.g. 3 hours). May choose variational approach (3DVar, 4DVar), or Kalman Filter. Error evolves with time** (e.g. Peters et al. 2007; Engelen et al. 2009; Baker et al. 2010; Liu et al. 2012).

INFERRING CARBON FLUXES

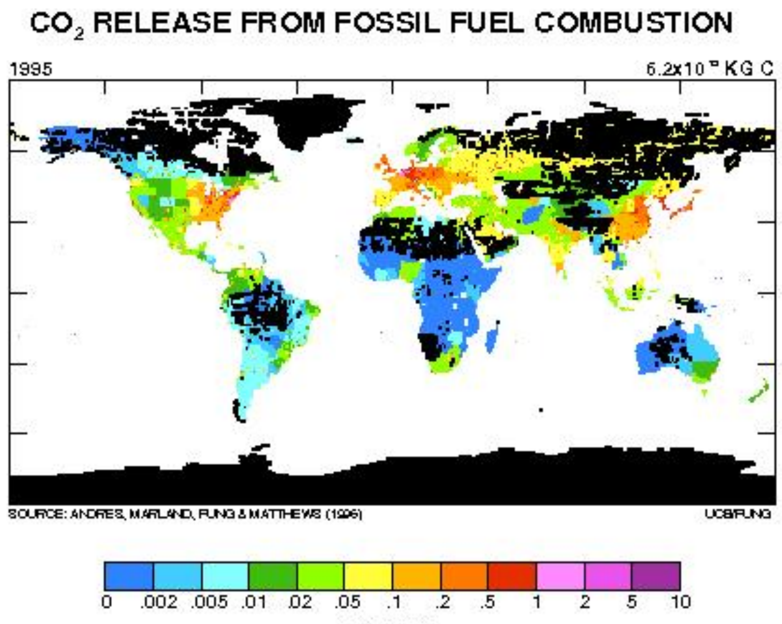
Atm Observations – Britt Stephens' talk

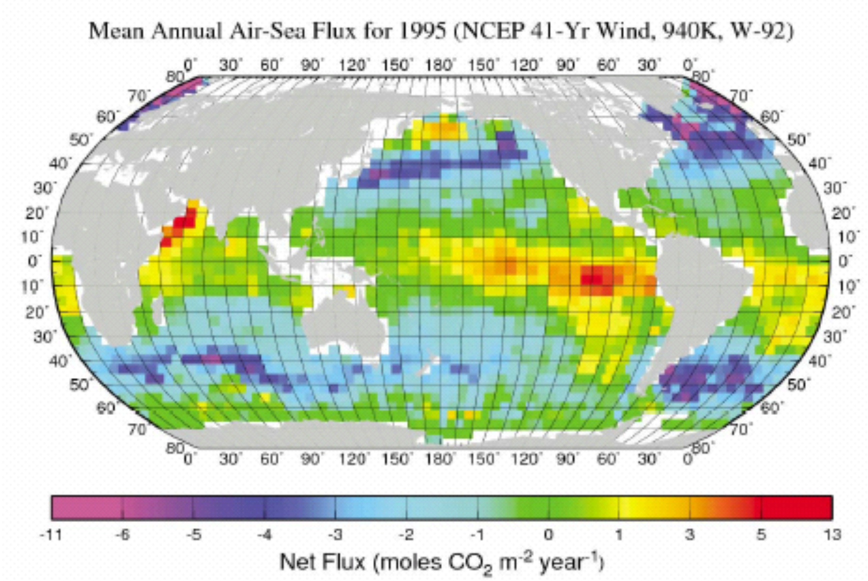
Today, we'll focus on the surface CO2 data from NOAA (GLOBALVIEW)





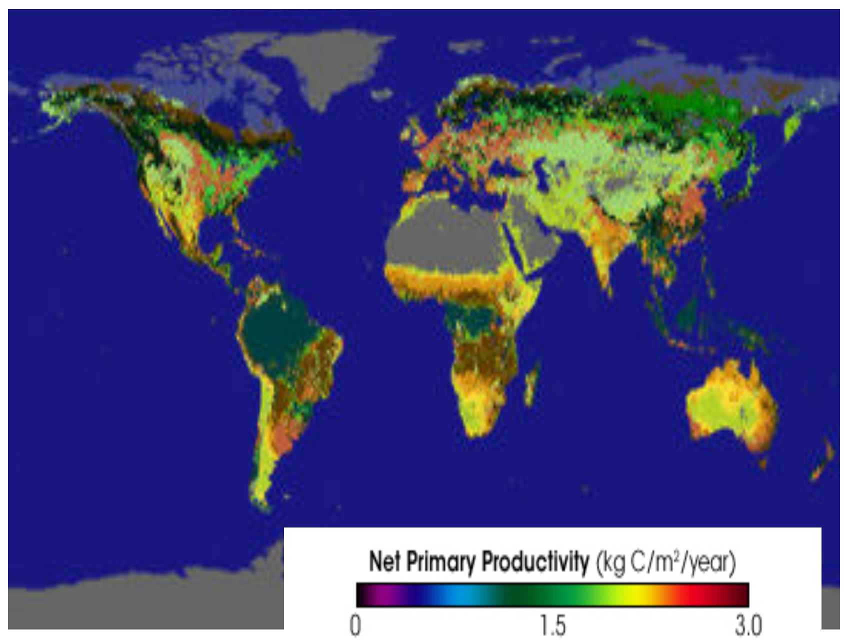
What We've Got: The Flux Priors + an Atm Transport Model





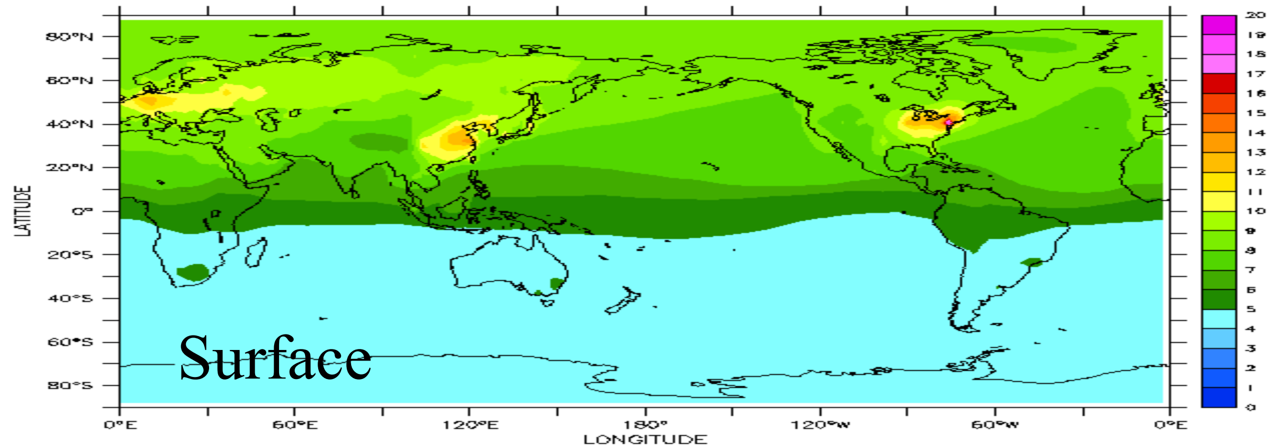
$$\frac{\partial C}{\partial t} + \underbrace{\Im(C)}_{Atm\_transport+mixing} = \underbrace{F\big|_{z=0}}_{SourcesSinks}$$

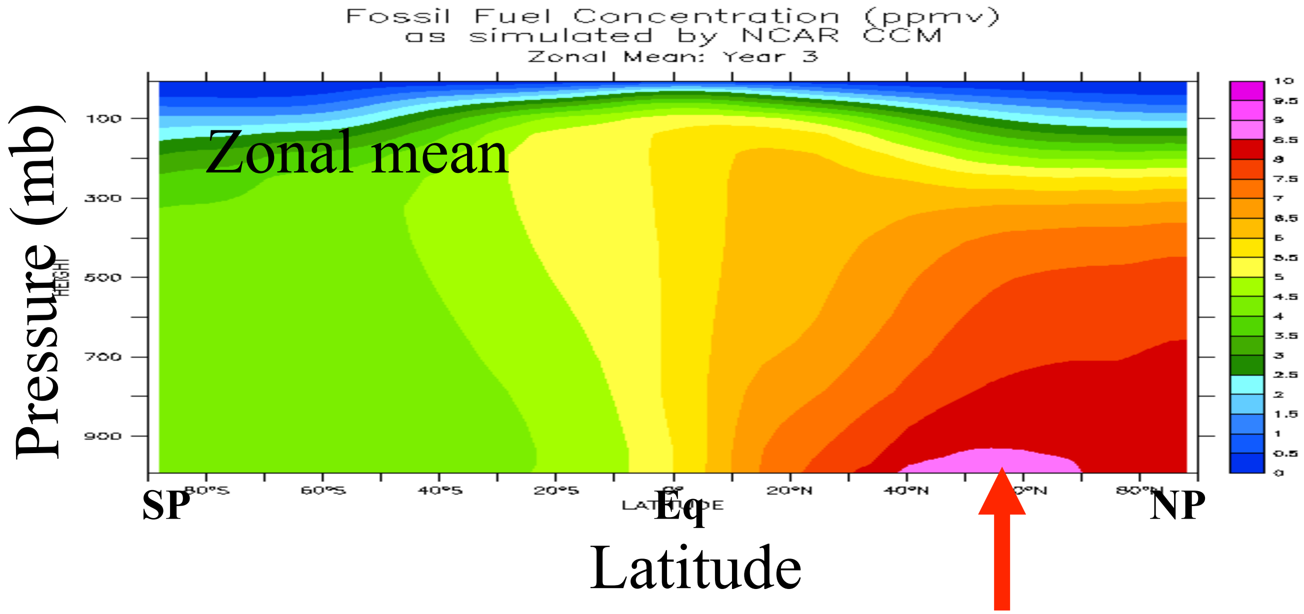
$$F = \underbrace{FF}_{\text{"well-known"}} + LandUse + \underbrace{(F_{oa} - F_{ao})}_{extrapolation\ of\ sparse\ obs} + (F_{ba} - F_{ab})$$



Example I: A Simpler Model - reduce 3D atm to 2 hemispheres

Atm CO2 distribution from FF emission, NCAR CSM





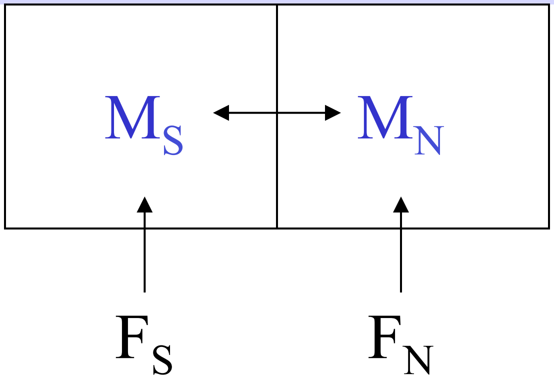
Note:
N-S gradient
Vertical gradient

INVERSE MODELING: (1) A SIMPLE MODEL OF THE ATM: PERFECT DATA

Example I: Interhemispheric Mixing: Two-Box Model, everything is perfect.

$$\frac{\partial M_N}{\partial t} = -\frac{M_N - M_S}{\tau} + F_N$$

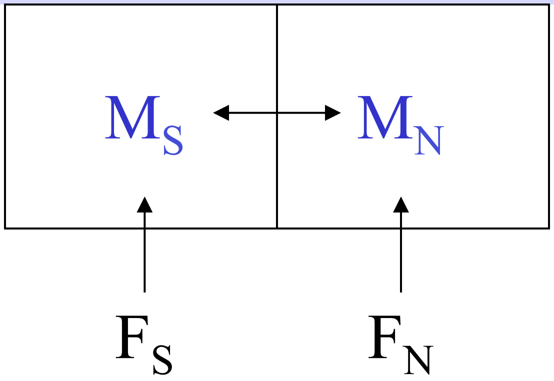
$$\frac{\partial M_S}{\partial t} = +\frac{M_N - M_S}{\tau} + F_S$$



Example 1: Interhemispheric Mixing: Two-Box Model, everything is perfect.

$$\frac{\partial M_N}{\partial t} = -\frac{M_N - M_S}{\tau} + S_N$$

$$\frac{\partial M_S}{\partial t} = +\frac{M_N - M_S}{\tau} + S_S$$

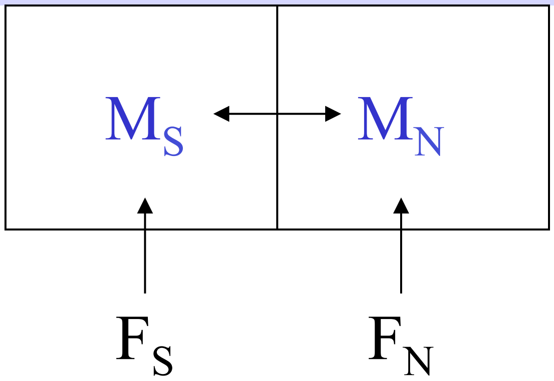


$$\frac{\partial (M_N - M_S)}{\partial t} = -2\frac{M_N - M_S}{\tau} + (F_N - F_S) = 0 \, @ \, SteadyState$$

$$\tau = 2\frac{M_N - M_S}{F_N - F_S}$$

Interhemispheric exchange time τ determined from inert tracers (e.g. CFC, with S_s=0): ~1-2 years

Ex I: 2-Box Model Applied to the Carbon Cycle



$$M_N - M_S = \frac{\tau}{2}(F_N - F_S)$$

Consider the case 100% FF is in the atm

$$F_N = 8\ PgC/yr; \quad F_s = 0; \quad \tau = 1\ yr$$

$$\rightarrow M_N - M_S = 4\ PgC$$

Recall $1\ PgC \rightarrow 0.5\ ppmv$ *if mixed in entire atm.*

$1\ PgC \rightarrow 1\ ppmv$ *if mixed in a hemisphere.*

$$\rightarrow X_N^{column} - X_S^{column} = 4\ ppmv$$

Guess (3D model) surface gradient γ=1.5 x column mean gradient

$$\rightarrow X_N^{sfc} - X_S^{sfc} = 6\ ppmv$$

But $(X_N^{sfc} - X_S^{sfc})_{obs} = 4\ ppmv$

Only 50% airborne. Sinks!

Inverse Problem: find the sinks

$$Obs: \quad (X_N^{sfc} - X_S^{sfc})_{obs} = 4 \ ppmv$$

$$\rightarrow (X_N^{column} - X_S^{column})_{obs} = \frac{4}{\gamma} = 2.7 \ ppmv$$

$$\rightarrow M_N - M_S\big|_{obs} = 2.7 \ PgC$$

$$Model: \ M_N - M_S = \frac{\tau}{2}(F_N - F_S)$$

$$Invert \ model \rightarrow F_N - F_S = 2\frac{M_N - M_S\big|_{obs}}{\tau} = 5.4 \ PgC/yr$$

$$(sources_N - sinks_N) - (sources_S - sinks_S) = 5.4 \ PgC/yr$$

$$(8 \ PgC/yr - sinks_N) - (0 - sinks_S) = 5.4 \ PgC/yr$$

$$\rightarrow sinks_N - sin\,ks_S = 8 - 5.4 = 2.6 \ PgC/yr$$

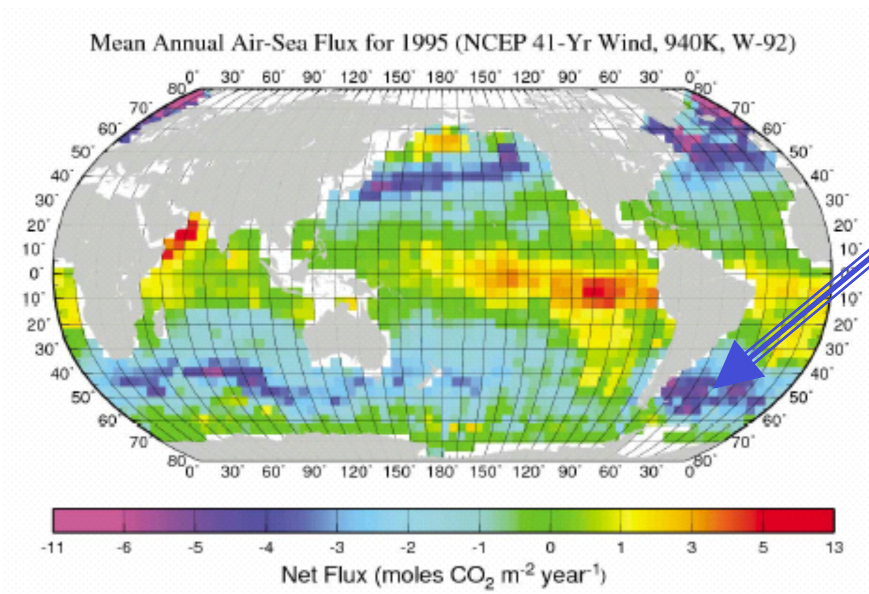
Where are the Carbon Sinks?

Budget $\quad sinks_N + sinks_S = 4.0\ PgC/yr$

Gradient $\quad sinks_N - sinks_S = 2.6\ PgC/yr$

$\rightarrow sinks_N = 3.3\ PgC/yr; \quad sinks_S = 0.7\ PgC/yr$

Northern sinks > Southern Sinks !!!!!!!



"Data/Obs": Huge C sink in the large expanse of southern ocean; but large uncertainty in obs

Northern ocn "better observed" → large Northern land sink!!!

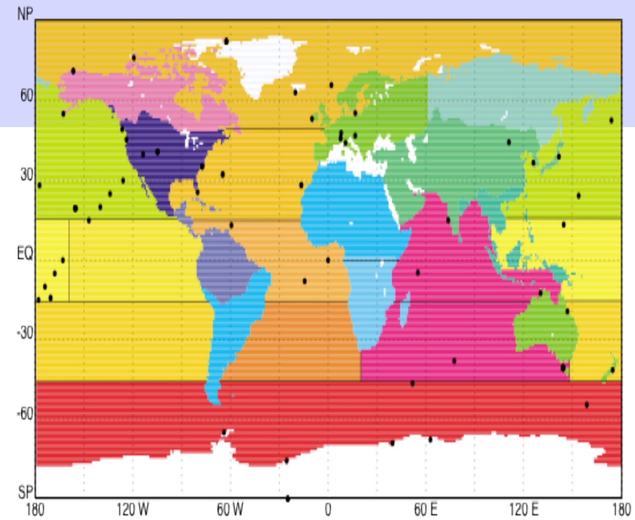
INVERSE MODELING: (2) PERFECT 3D ATM MODEL; DATA WITH UNCERTAINTY

Example II: Perfect 3D atm circulation model. Steady state

(1) Forward Step

- <u>Premise</u>: Atm CO_2 = linear combination of response to each source or sink

1.0 Divide surface into "basis regions"



1.1: Specify unitary source (e.g. 1 PgC/year) each year from each region

$$\widehat{s}_k(x, y)$$

1.2: Simulate atm CO_2 "basis" response with atm general circulation model

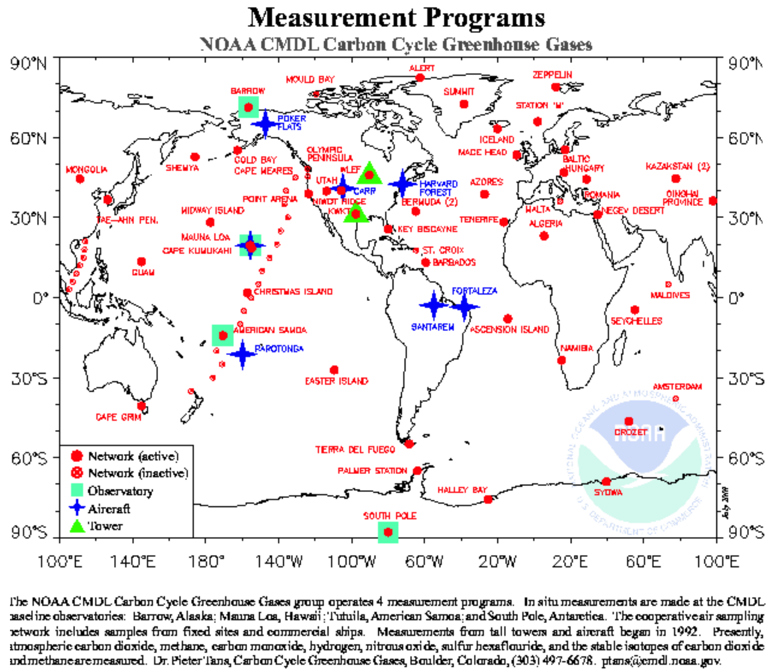
$$\widehat{s}_k(x,y) \xrightarrow{\text{transport model}} \widehat{c}_k(x,y,z,t)$$

1.3 Reconstruct fluxes and concentrations: unknown source strength Φ_k

$$S = \sum_{k-regions} \underbrace{\Phi_k}_{unknown} \times \widehat{s}_k(x,y)$$

$$c_{model}(x,y,z) = \sum_k \underbrace{\Phi_k} \times \widehat{c}_k(x,y,z)$$

Ex II: (Step 2) Bayesian Inversion: perfect circulation model



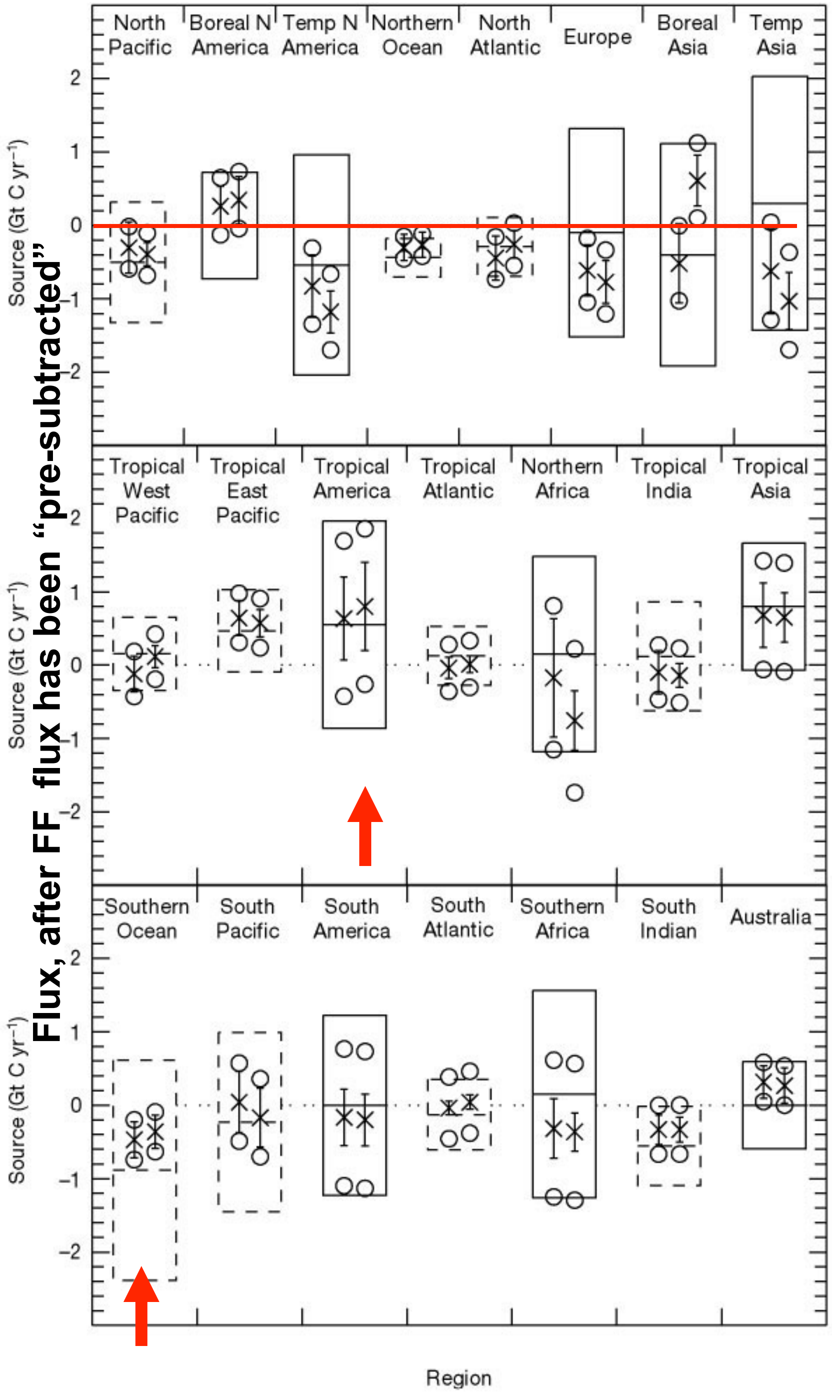
Inversion: Seek the optimal source/sink combination $\{\Phi_k\}$ to match atmospheric CO_2 data: *minimize*

$$J = \frac{1}{2}\left[\sum_{stn} \frac{\left[C_{obs}(stn) - \sum_{k-regions} \Phi_k \times \hat{c}_k(stn)\right]^2}{\sigma_{stn}^{\;2}} + \sum_{k-regions} \frac{\left[\Phi_k - \Phi_k^{prior}\right]^2}{\left[\sigma_k^{prior}\right]^2}\right]$$

- **Obs. Network –**
  - **mainly remote marine locations**

Trying to infer information over land

Undetermined; non-unique solutions; prior estimates of source/sinks as additional constraints



Ex IIa: Posterior from many "perfect" circulation models

$$\Phi_k^{prior} \pm \sigma_k^{prior}$$

"Analysis" from Model m:

$$\Phi_{m,k}^{posterior} \pm \sigma_{m,k}^{posterior}$$

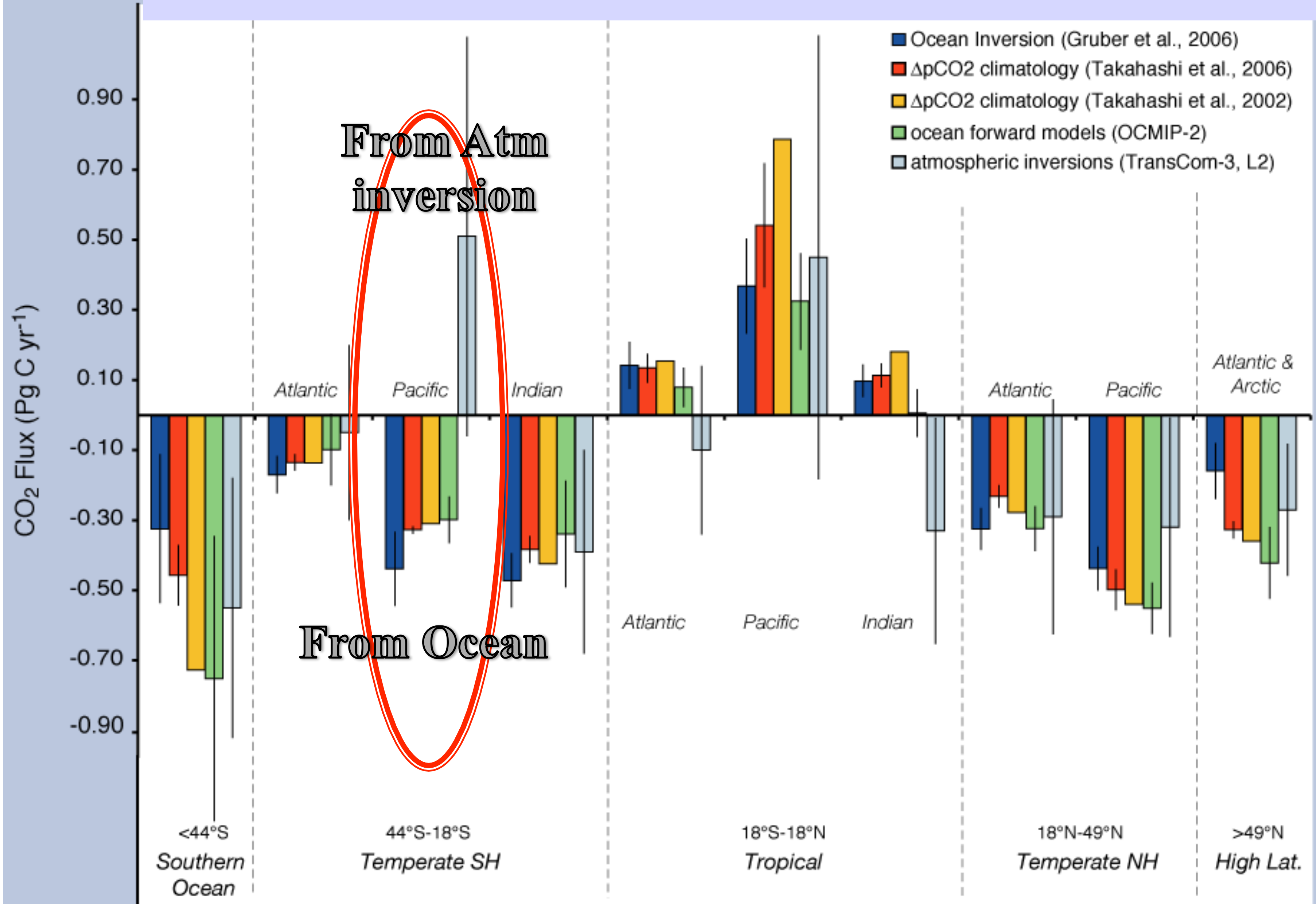
Multi-model Mean and std_dev of $\Phi_{m,k}^{posterior}$

Multi-model Mean of $\sigma_{m,k}^{posterior}$

Little innovation in tropics, Africa
Great innovation in S. Ocean

Gurney et al. Nature 2005

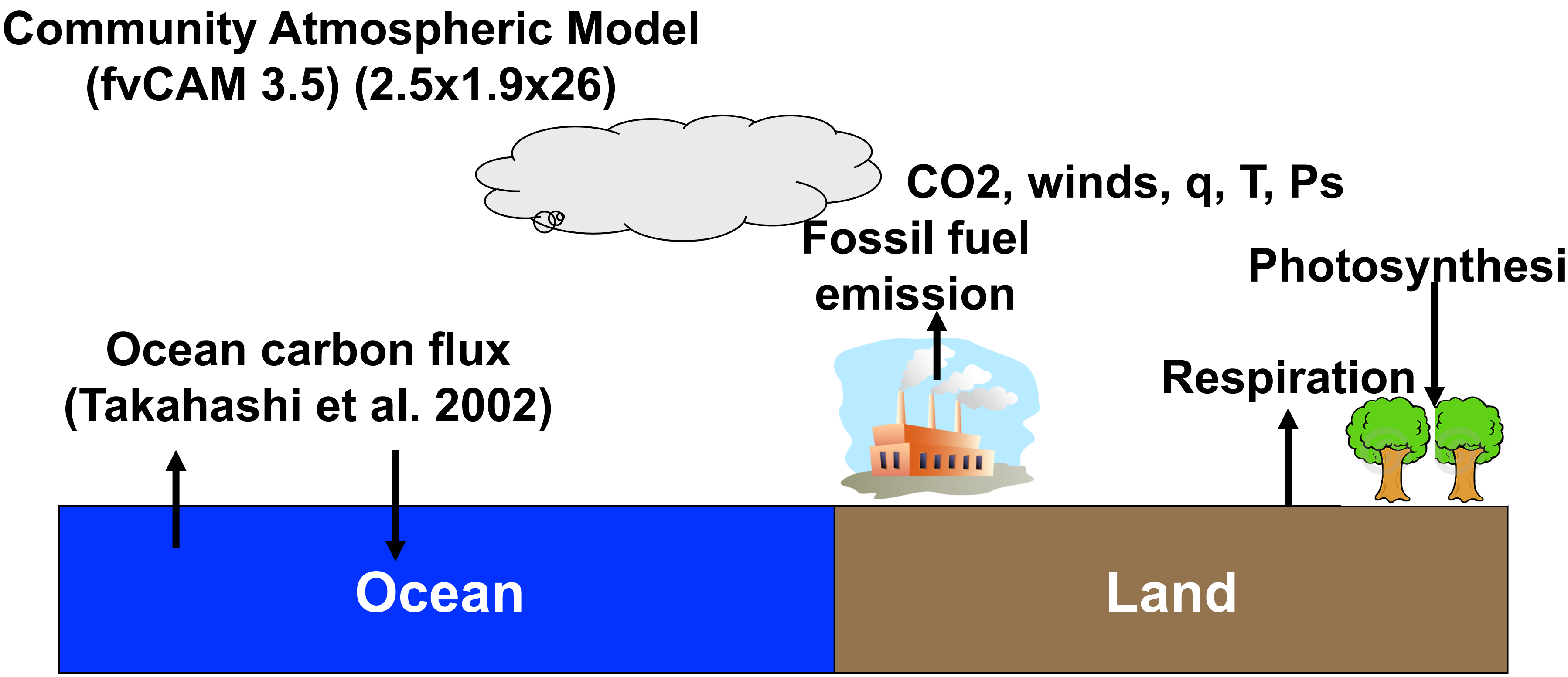
Do the different methods agree about regional fluxes?



Gruber et al. (2006)

CARBON DATA ASSIMILATION: LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

Step 1: Forecast. Integrate Carbon-Climate Model forward for 6 hours; ensemble of 64 members



- CO_2 is transported as a tracer in CAM 3.5.
- Land carbon flux: 6-hourly flux from biogeochemical model.
- Model produces CO_2 distribution that matches major features in surface CO_2 obs
- Time period: 2003.

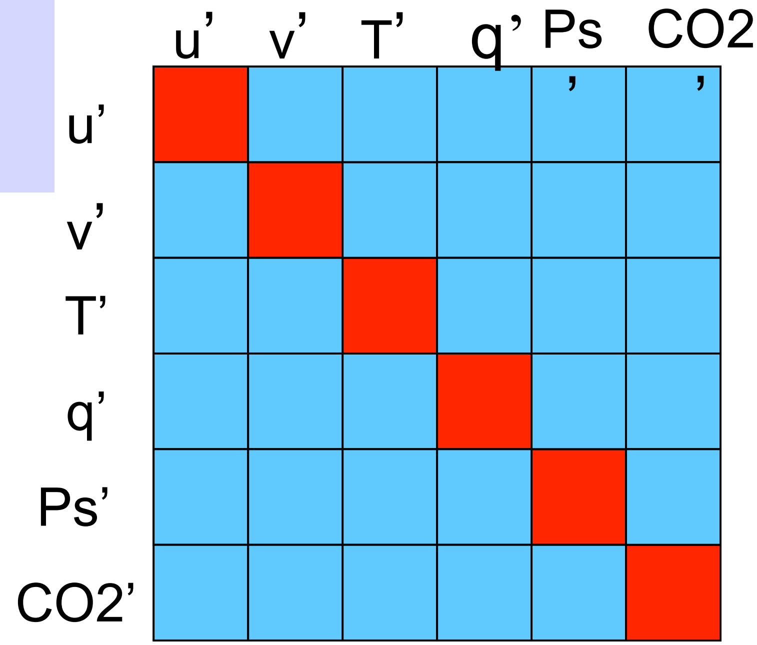
Step 2: Error Statistics of Forecast (background)

- $\mathbf{x}=\{u, v, T, q, Ps, CO2\}$
- K ensemble members of forecast →

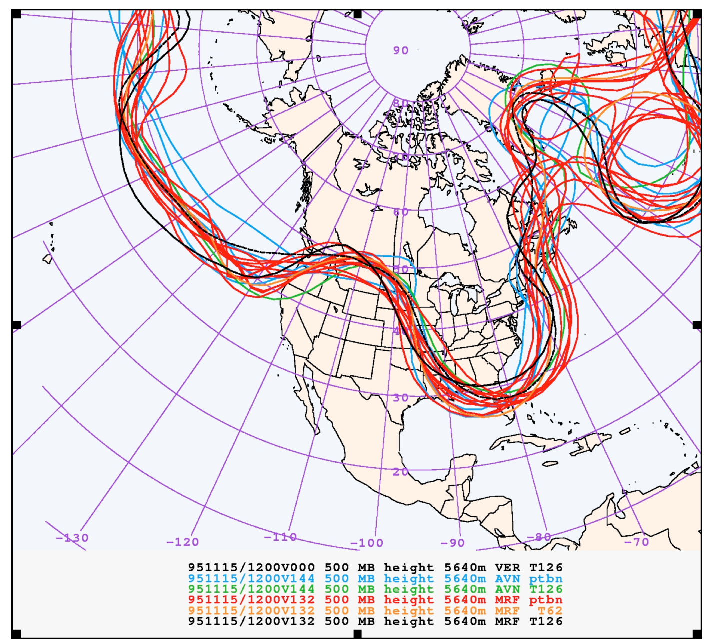
  x_i^b;   i=1…K=64
- Calculate ensemble mean $\overline{\mathbf{x}}^b$
- Calculate Covariance

$$\mathbf{P}^b = \frac{1}{K-1}\sum_{i=1}^{K}(\mathbf{x}_i^b - \overline{\mathbf{x}}^b)(\mathbf{x}_i^b - \overline{\mathbf{x}}^b)^T$$

$$= \begin{pmatrix} u'u' & \cdots & u'CO2' \\ \vdots & \ddots & \vdots \\ CO2'u' & \cdots & CO2'CO2' \end{pmatrix}$$



std dev in u’ etc → error of the day
Large std dev → atm is dynamically unstable

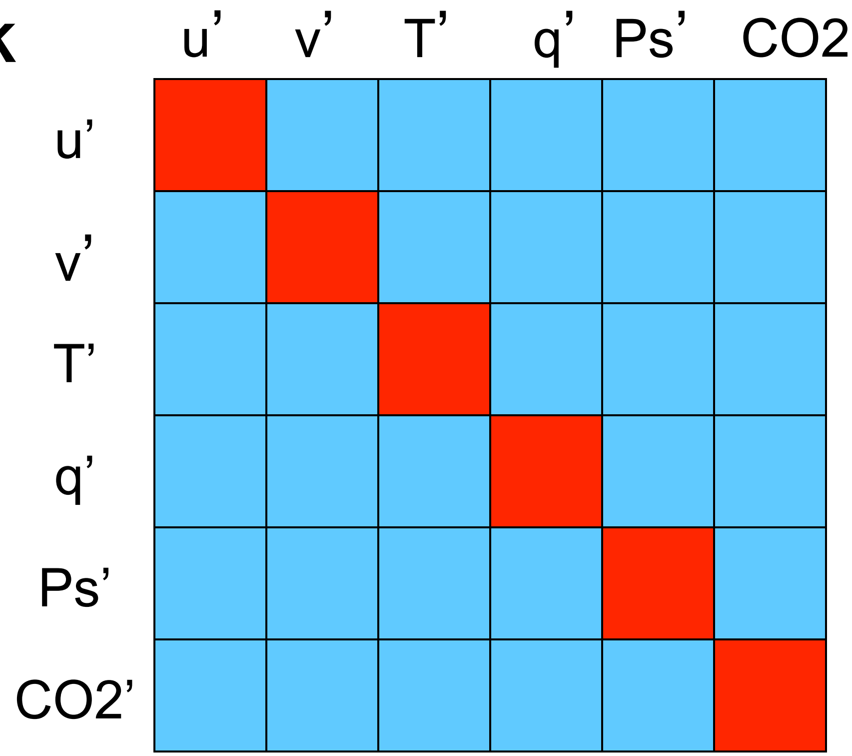


Forecast error statistics in EnKF

- **Run K ensemble members-> x_i^b; i=1..K**
- $\overline{\mathbf{x}}^b$ **: ensemble mean.**

$$\mathbf{P}^b = \frac{1}{K-1}\sum_{i=1}^{K}(\mathbf{x}_i^b - \overline{\mathbf{x}}^b)(\mathbf{x}_i^b - \overline{\mathbf{x}}^b)^T$$

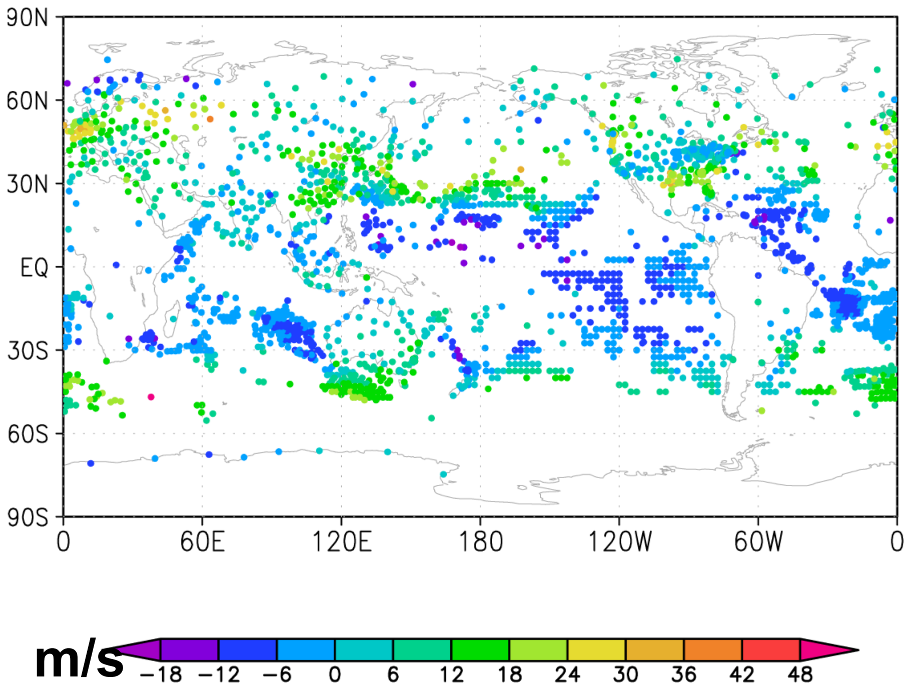
$$= \begin{pmatrix} u'u' & \cdots & u'CO2' \\ \vdots & \ddots & \vdots \\ CO2'u' & \cdots & CO2'CO2' \end{pmatrix}$$



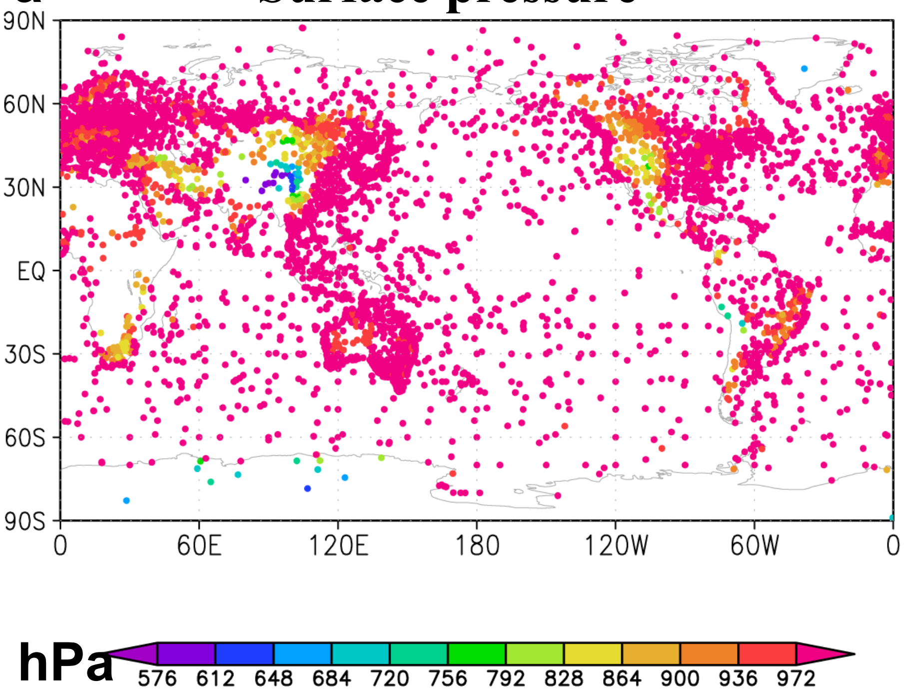
- **Propagate info from the dynamical variables with observations to the dynamical variables with no observation.**
- **From location with observation to locations with no obs.**

Meteorological observations
radiosonde, satellite, ships, …

Zonal wind within 500hPa and 600hPa

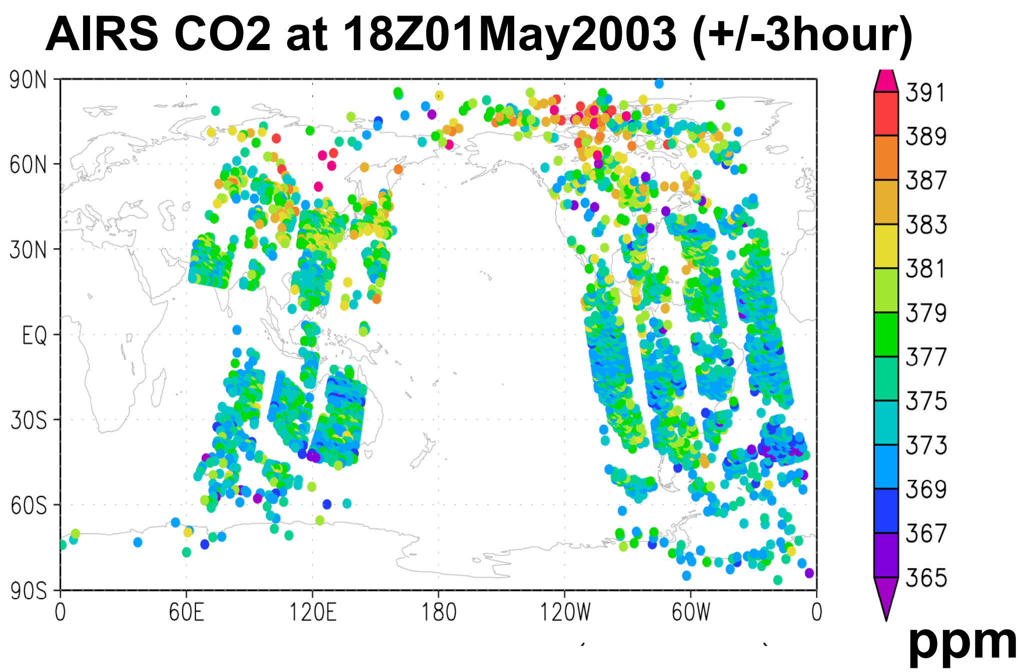


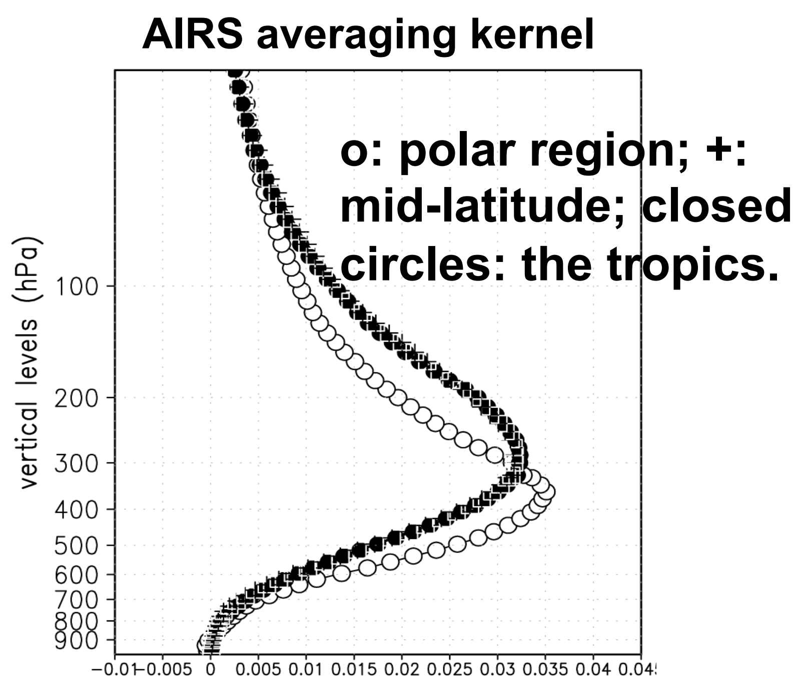
Surface pressure



10^6 observations within 6-hour.

CO2 obs from AIRS satellite





>2000 obs in 6 hours

Sensitive to CO2 in mid-troposphere

Step 3: Apply Observations Operator *H(x)*

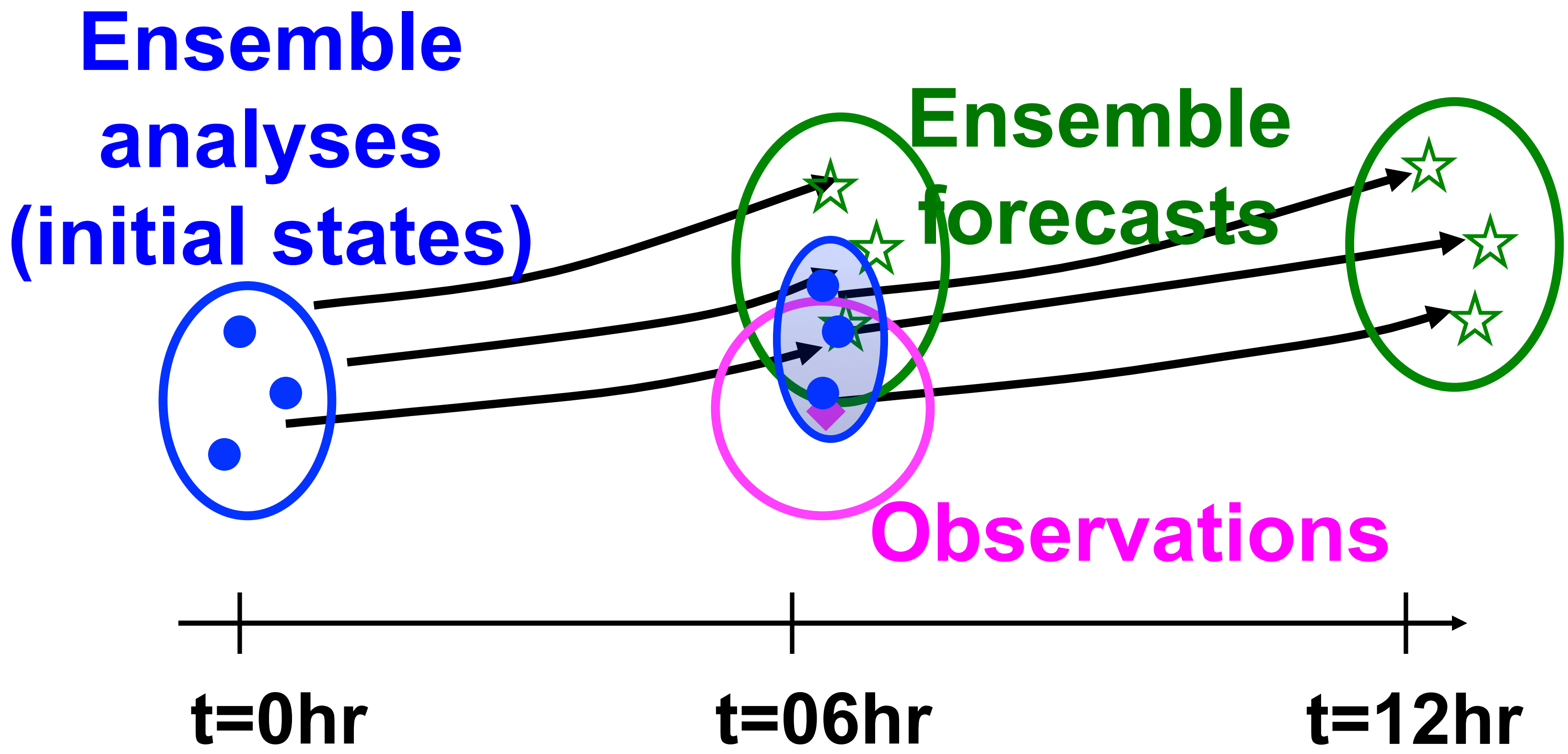
Example:

- **x={u,v,T,q,Ps,CO2} from model at every grid box**

- **Obs: e.g. column CO2 from satellite at certain (x,t)**
  - ***H(CO2)* does the column average (using satellite averaging kernel), interpolate/average to location of obs, select times of obs**

- **Obs: e.g. radiance measured by satellite**
  - ***H(CO2)* employs a radiative transfer model to calculate the associated radiance at the wavelength, time, location of measurement**

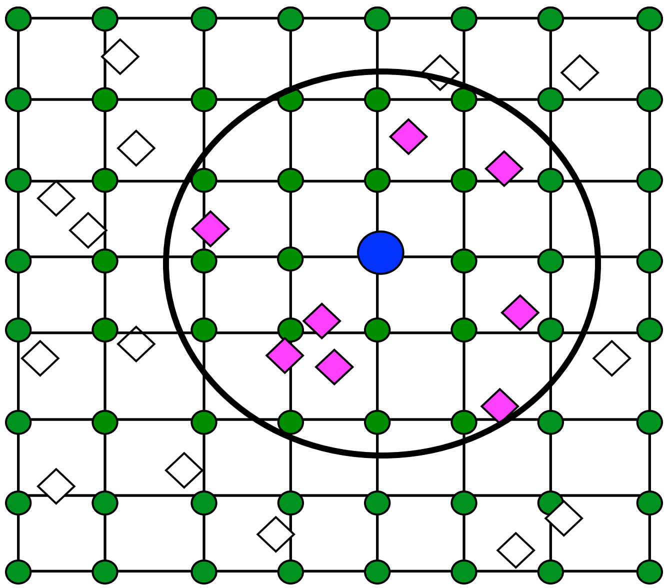
Step 4: Ensemble Kalman Filter (EnKF)



- Background error changes with time;
- Obtain ensemble analyses.

Local Ensemble Transform Kalman Filter (LETKF, Ott et al., 2004, Hunt et al., 2007)

Schematic 2-dimension local patch



- ✓ The LETKF solves analysis states at each grid point.
- ✓ The LETKF assimilates observations within a local volume (both horizontal and vertical); Choice of local volume is guided by effective correlation length from P^b.

Summary steps of LETKF

1) Global 6 hr ensemble forecast starting from the analysis ensemble

2) Choose the observations used for each grid point

3) Compute the matrices of forecast perturbations in ensemble space X^b

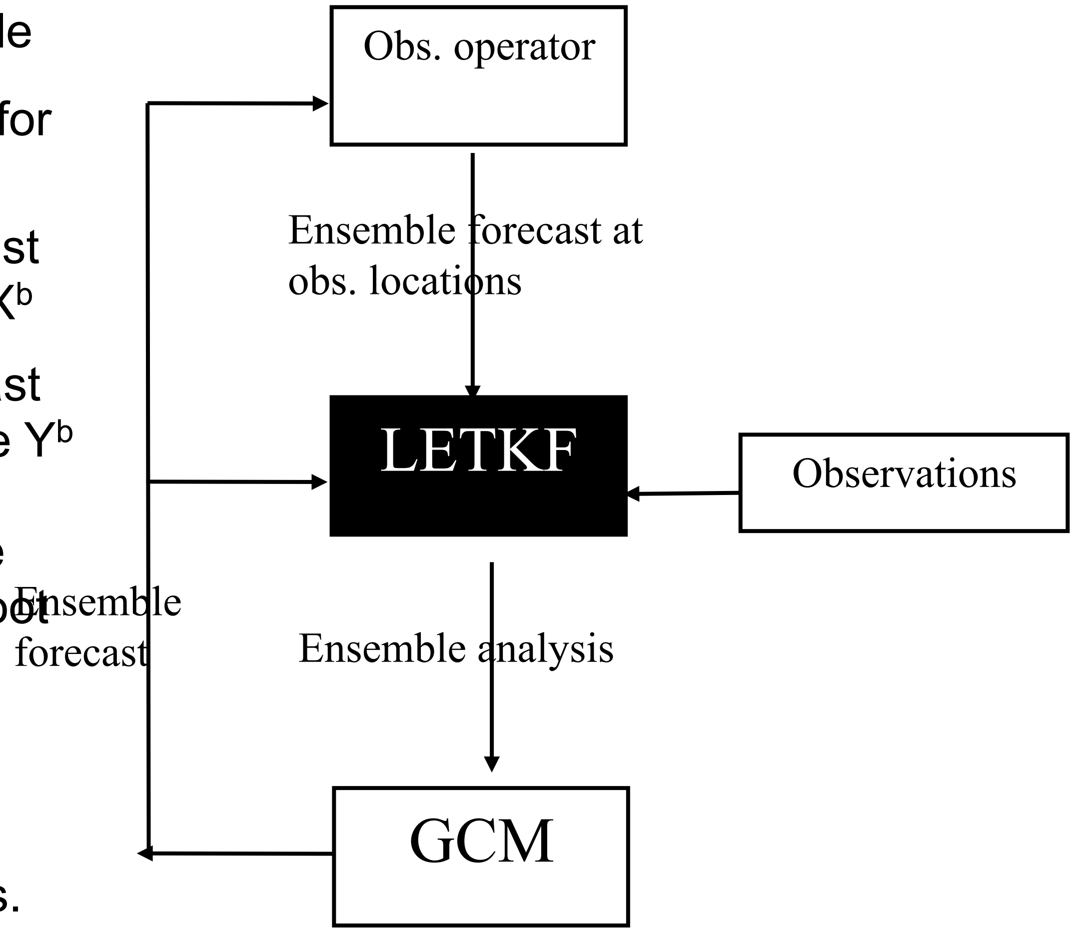
4) Compute the matrices of forecast perturbations in observation space Y^b

5) Compute P^b in ensemble space space and its symmetric square root

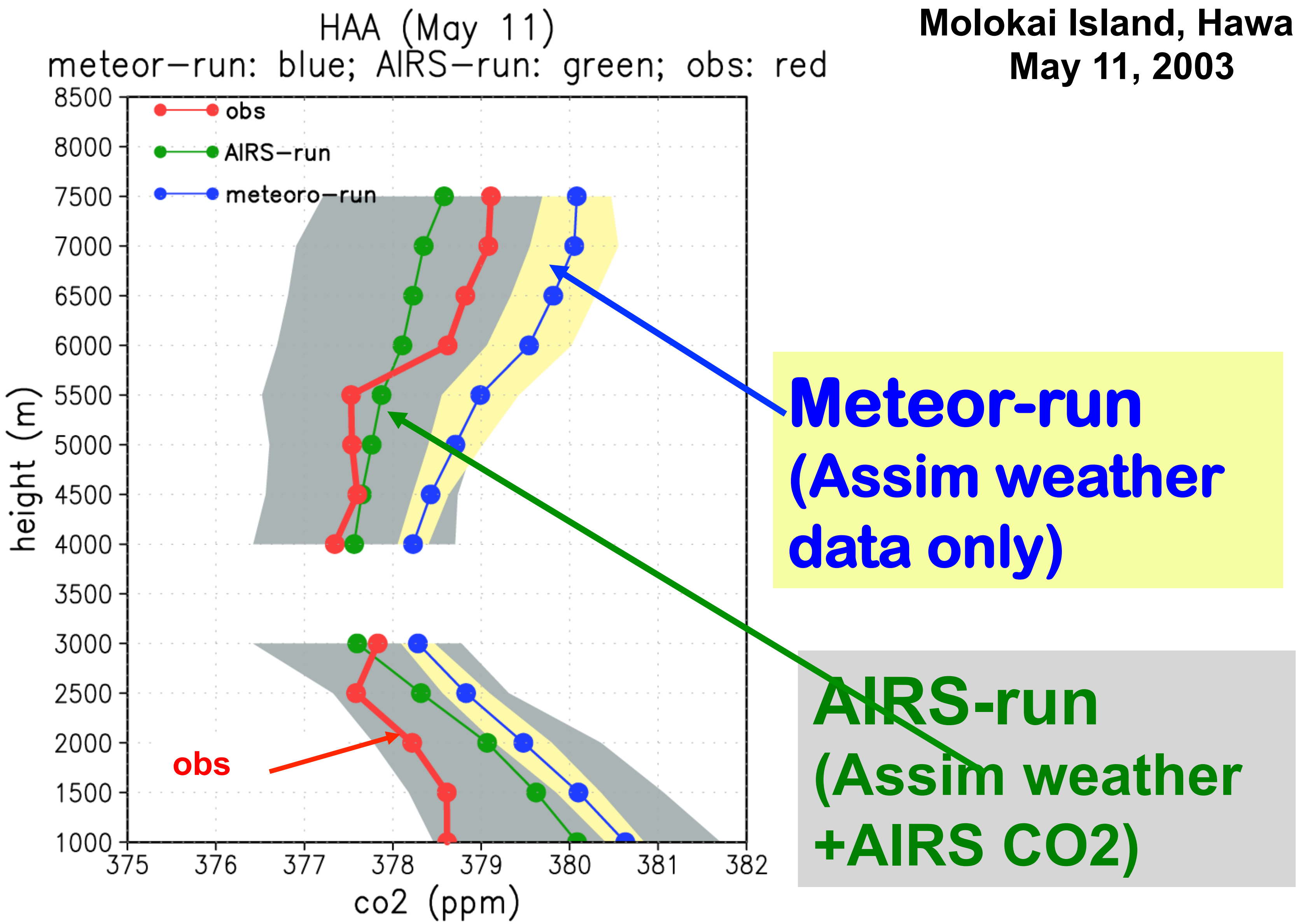
6) Compute w^a, the k vector of perturbation weights

7) Compute the local grid point analysis and analysis perturbations.

8) Gather the new global analysis ensemble.



Analysis ensemble: mean + spread



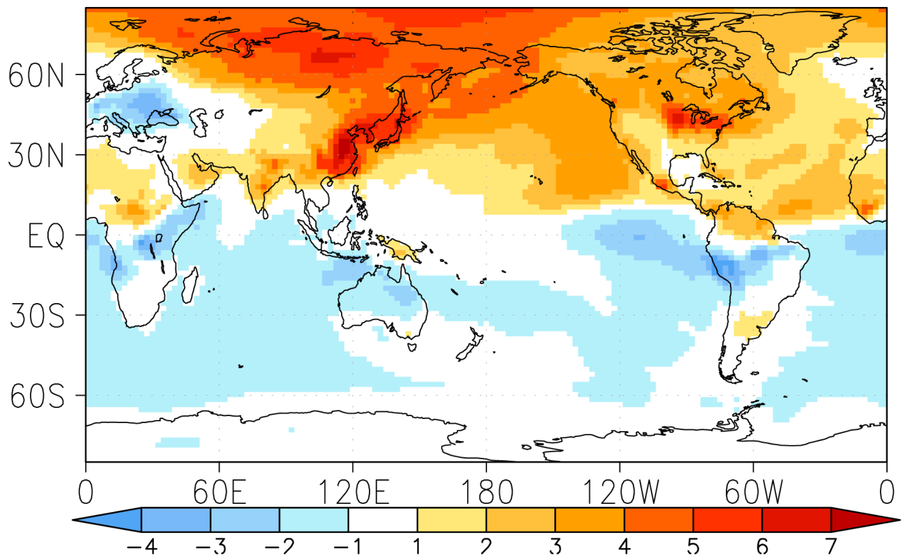
Molokai Island, Hawaii.
May 11, 2003

Vertical Gradient of CO2

CO2(925hPa)-CO2(500hPa)

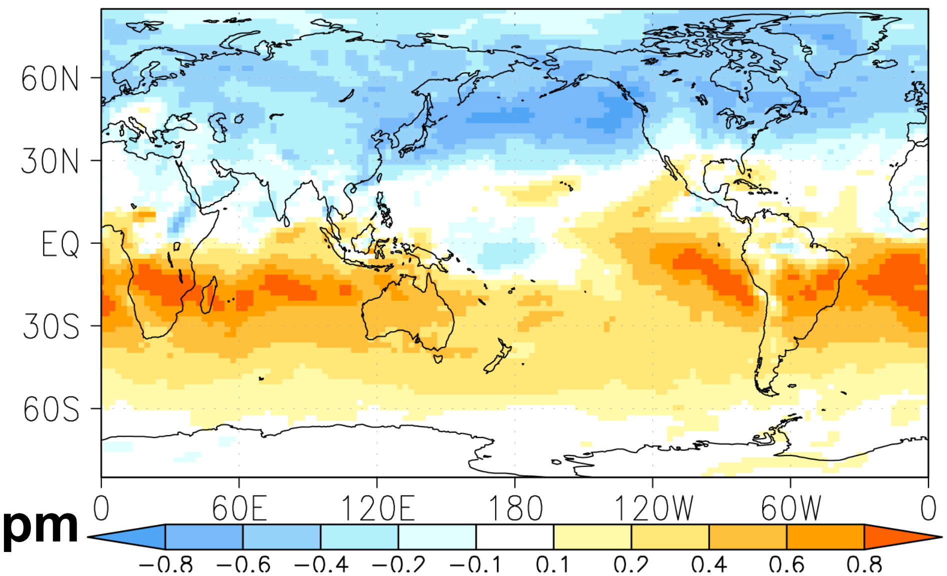
May 2003

Met-run



ppm

(AIRS-run)-(Met-run)

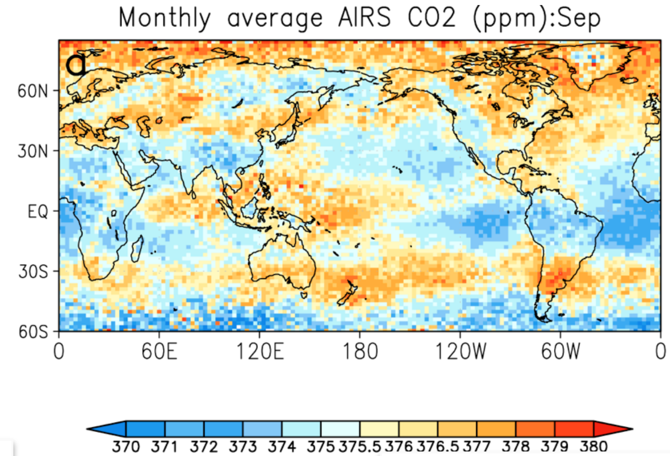


Model: not enough vertical mixing

Assimilation --> first global CO2(z) from obs

CO2 at ~500hPa

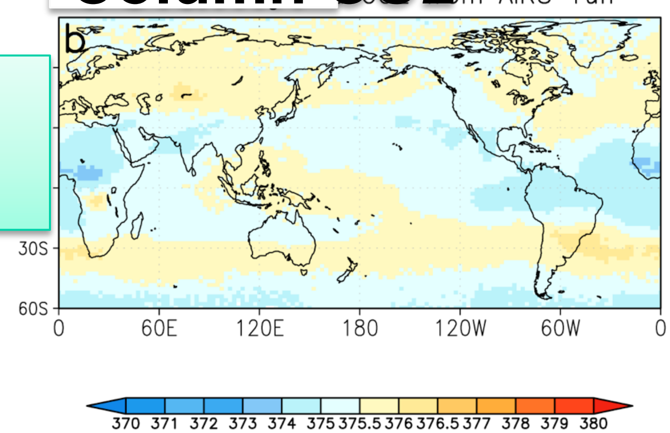
AIRS ob

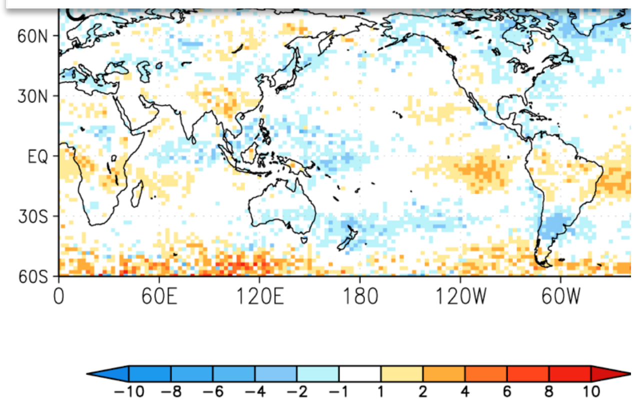


Column CO2

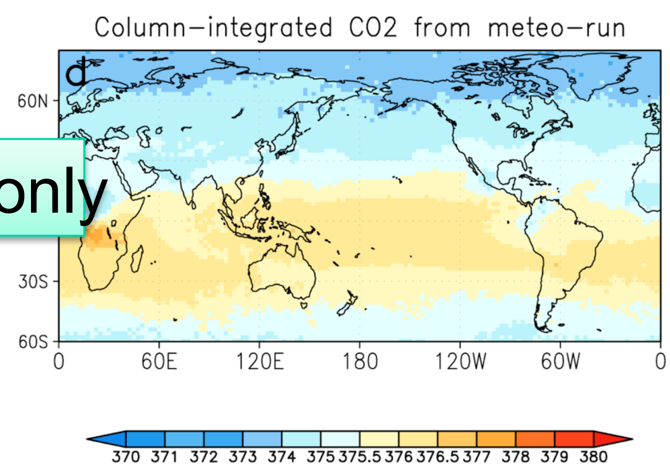
Column minus AIRS

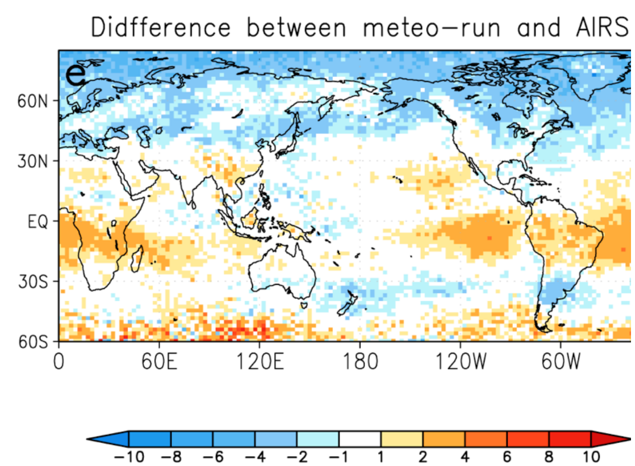
Assim Meteorology +AIRS





Assim Meteorology only





Summary: Inferring CO_2 fluxes is largely a problem of least squares fit

- Problem is under-determined
- Paucity of CO_2 observations, esp over land where fluxes are variable in space and time, and the Southern Ocean
- Paucity of obs re vertical gradient of CO_2
- Need to build up "background", "prior", "forecast"
- Need to improve estimates of uncertainty in obs, and uncertainty in "background" for proper weighting. N.B. Uncertainty in obs .ne. uncertainty in measurement (representativeness…)