Introduction to data assimilation and the applications of EnKF to carbon data assimilation

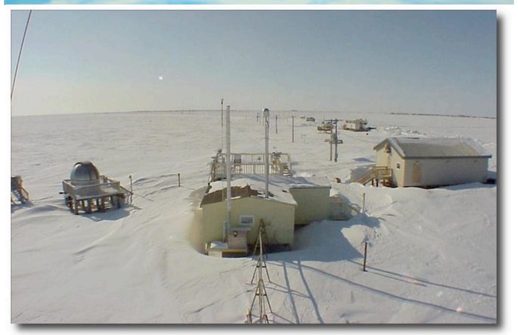
Junjie Liu

University of California, Berkeley

What is data assimilation?

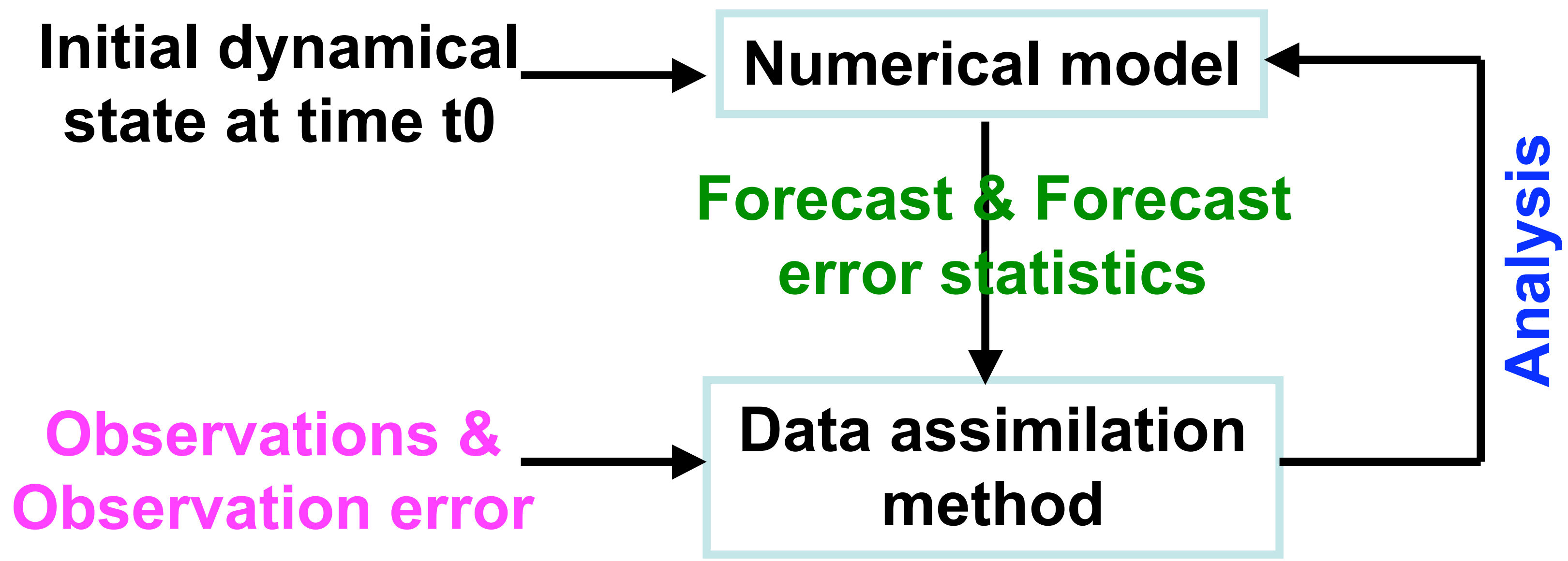
Data assimilation is the technique whereby incomplete observational data are combined with forecasts from an imperfect model to produce an optimal estimate of the evolving state of the system.

Why we need data assimilation in carbon problem?

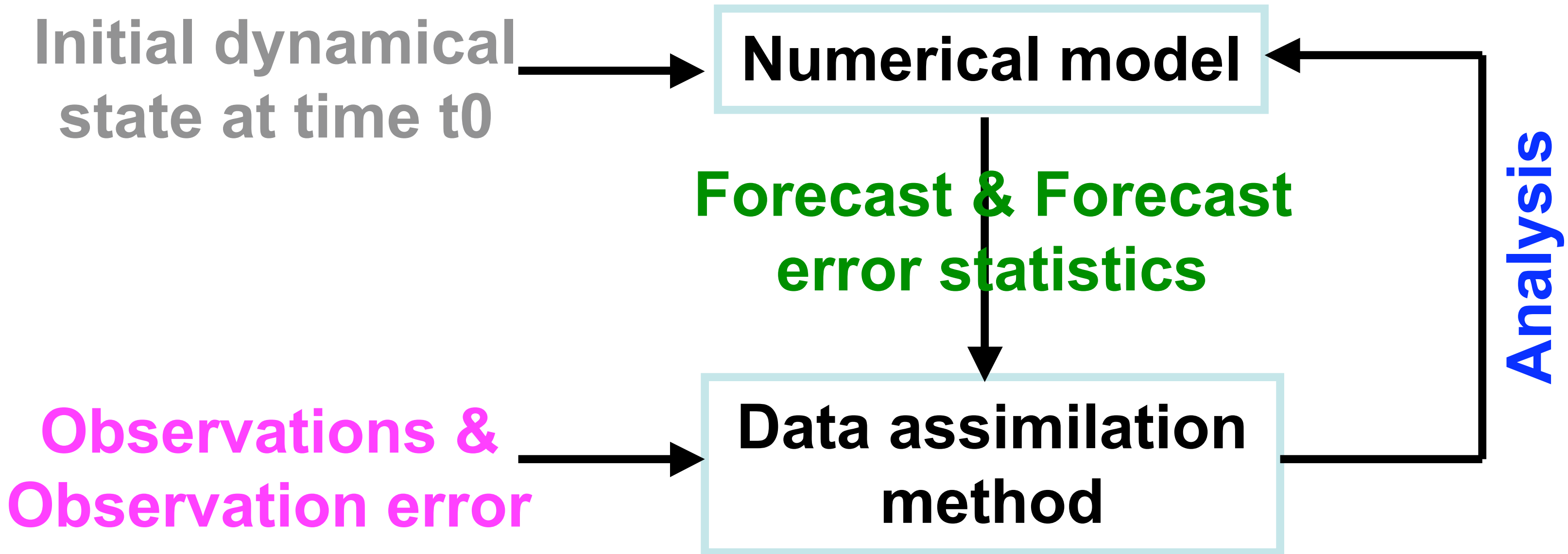


- Observations from different sources
- Different errors
- Data gaps
- Quantities not measured (i.e. model parameters)
- Quantities linked (i.e., surface flux)

Data assimilation cycle

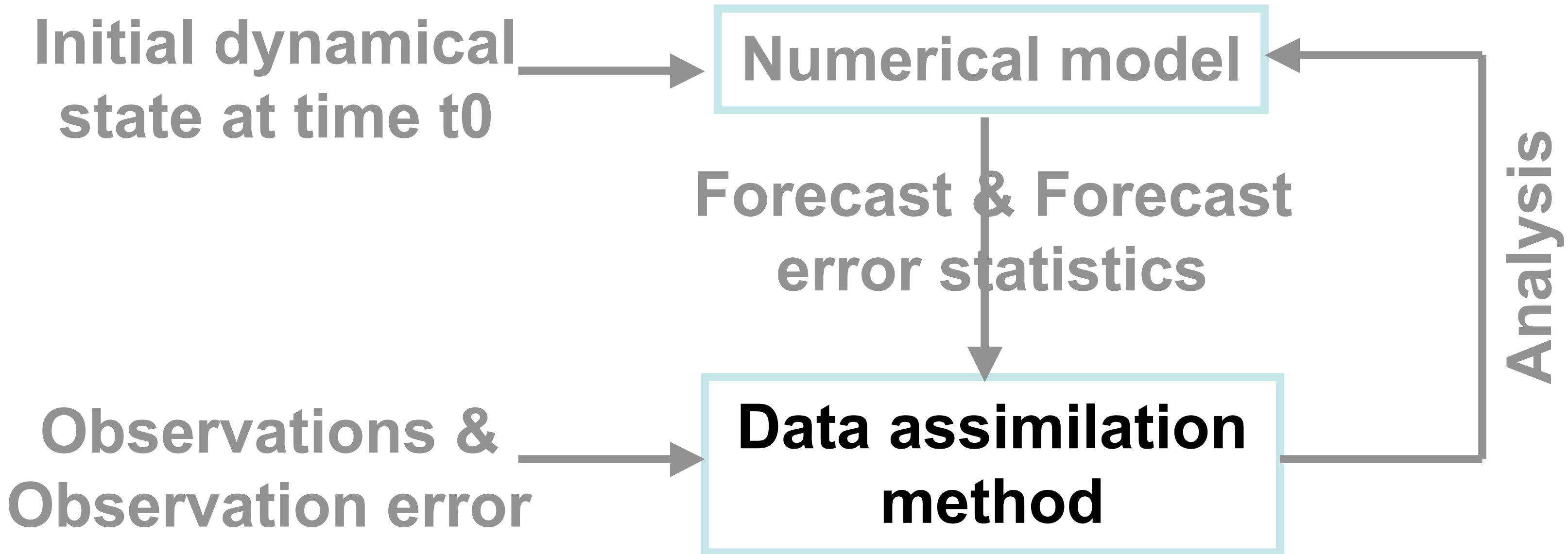


Data assimilation cycle



- Numerical model: Community Atmospheric Model 3.5 (CAM3.5)
- Observations: u, v, T, q, Ps and AIRS CO2
- CO2 is a tracer in CAM3.5 with fossil fuel, land and ocean fluxes from TransCom3 as forcing (nature run matches seasonal cycle and north-south gradient of observations etc.)

Data assimilation cycle



Ensemble Kalman Filter (EnKF) process

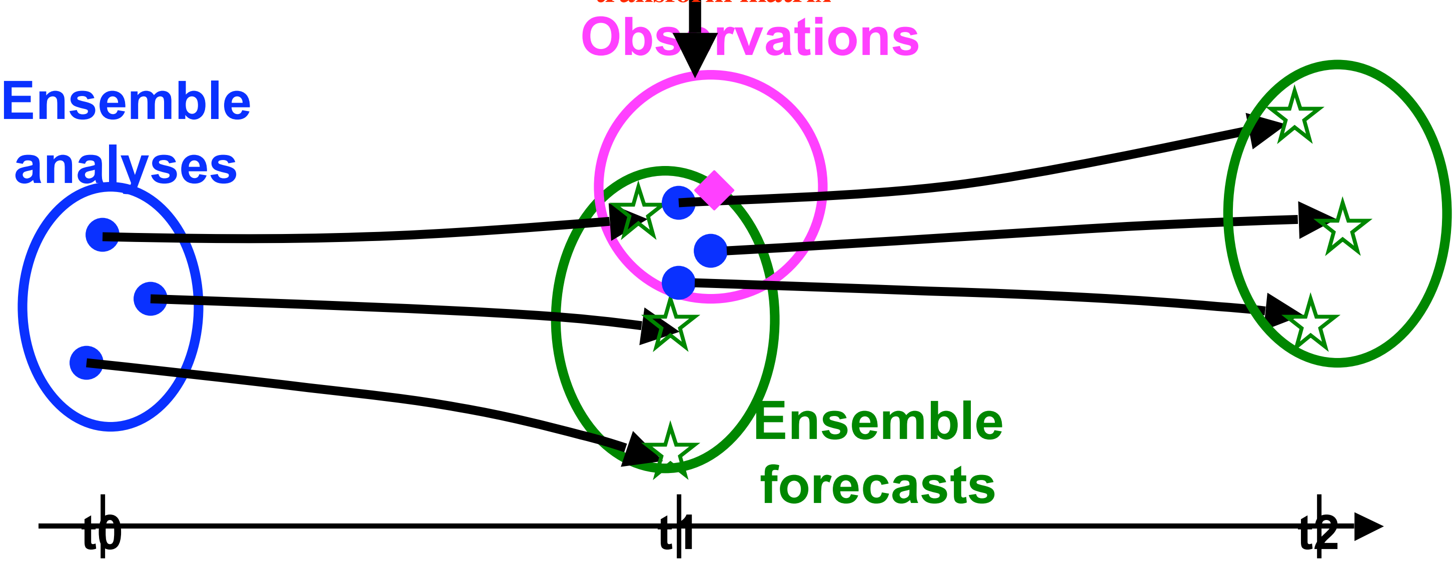
$$\underbrace{\bar{\mathbf{x}}^a}_{\text{analysis}} = \underbrace{\bar{\mathbf{x}}^b}_{\text{forecast}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} (\underbrace{\mathbf{y}^o}_{\text{observation}} - \underbrace{h(\bar{\mathbf{x}}^b)}_{\text{forecast at observation space}}); \mathbf{K} = F(\underbrace{\mathbf{R}}_{\text{observation error}} , \underbrace{\mathbf{P}^b}_{\text{foreca error}})$$

$$\underbrace{\mathbf{X}^a}_{\text{analysis perturbations}} = \underbrace{\mathbf{X}^b}_{\text{forecast perturbations}} \underbrace{(\tilde{\mathbf{P}}^a)^{\frac{1}{2}}}_{\text{transform matrix}} ; \tilde{\mathbf{P}}^a = F(\underbrace{\mathbf{R}}_{\text{observation error}} , \underbrace{\mathbf{P}^b}_{\text{foreca error}})$$

Ensemble Kalman Filter (EnKF) process

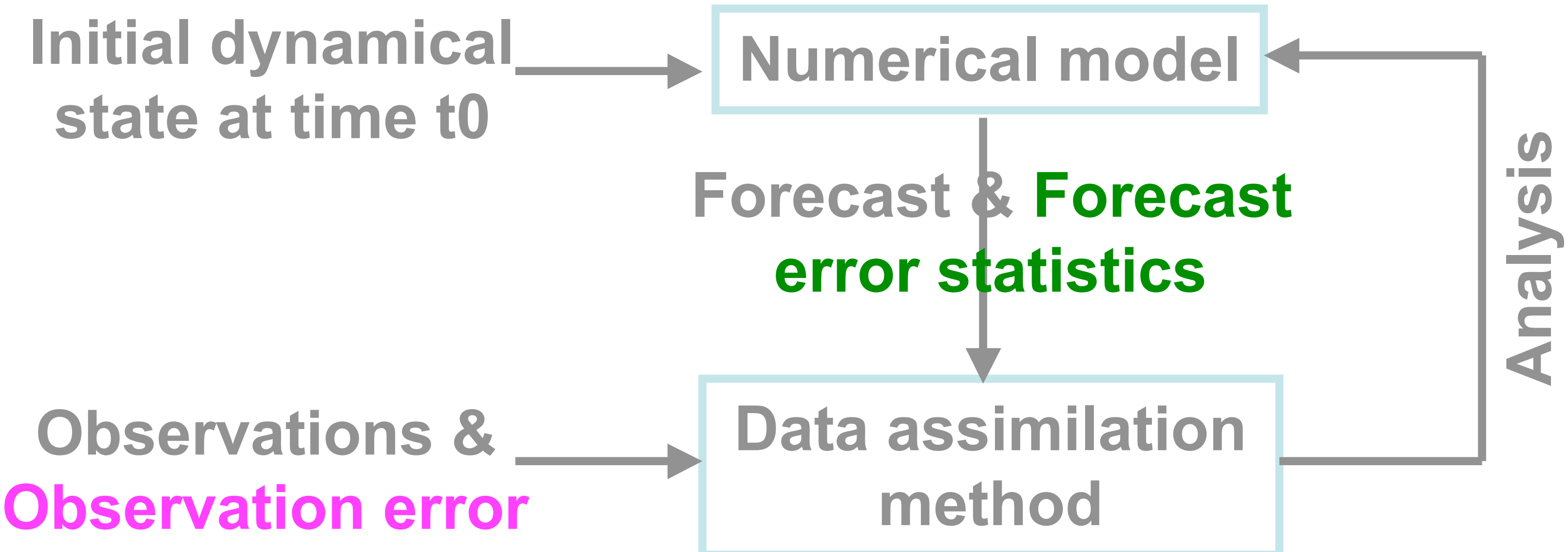
$$\underbrace{\bar{\mathbf{x}}^a}_{\text{analysis}} = \underbrace{\bar{\mathbf{x}}^b}_{\text{forecast}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} \Big(\underbrace{\mathbf{y}^o}_{\text{observation}} - \underbrace{h(\bar{\mathbf{x}}^b)}_{\text{forecast at observation space}} \Big); \mathbf{K} = F\Big(\underbrace{\mathbf{R}}_{\text{observation error}}, \underbrace{\mathbf{P}^b}_{\text{foreca error}} \Big)$$

$$\underbrace{\mathbf{X}^a}_{\text{analysis perturbations}} = \underbrace{\mathbf{X}^b}_{\text{forecast perturbations}} \underbrace{(\tilde{\mathbf{P}}^a)^{\frac{1}{2}}}_{\text{transform matrix}} ; \tilde{\mathbf{P}}^a = F\Big(\underbrace{\mathbf{R}}_{\text{observation error}}, \underbrace{\mathbf{P}^b}_{\text{foreca error}} \Big)$$



- Magenta: observation and its uncertainty; Green: Ensemble forecasts and its uncertainty; Blue: ensemble analyses and its uncertainty;
- Obtain ensemble analyses and forecasts;

Data assimilation cycle



Observation error statistics

1. Instrumental/retrieval errors given along the observations
2. Estimate observation error R during data assimilation

$$(\underbrace{\mathbf{y}^o}_{\text{obs}} - \underbrace{\mathbf{y}^a}_{\text{analysis at obs space}})(\mathbf{y}^o - \underbrace{\mathbf{y}^b}_{\text{forecast at obs space}})^T = \mathbf{R}$$

- **R: diagonal term + off diagonal term**
- **R captures the relationship among observation, forecast and analysis, and representation error.**
- **This could be used in estimating observation errors for observations without error statistics**

Forecast error statistics in EnKF

- **Using ensemble method to estimate:**

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

$$= \frac{1}{K-1}\mathbf{X}^b\mathbf{X}^{bT}$$

- **Estimate error variance (diagonal) along with covariance (off-diagonal term)**

Forecast error statistics in EnKF

- Using ensemble method to estimate:

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

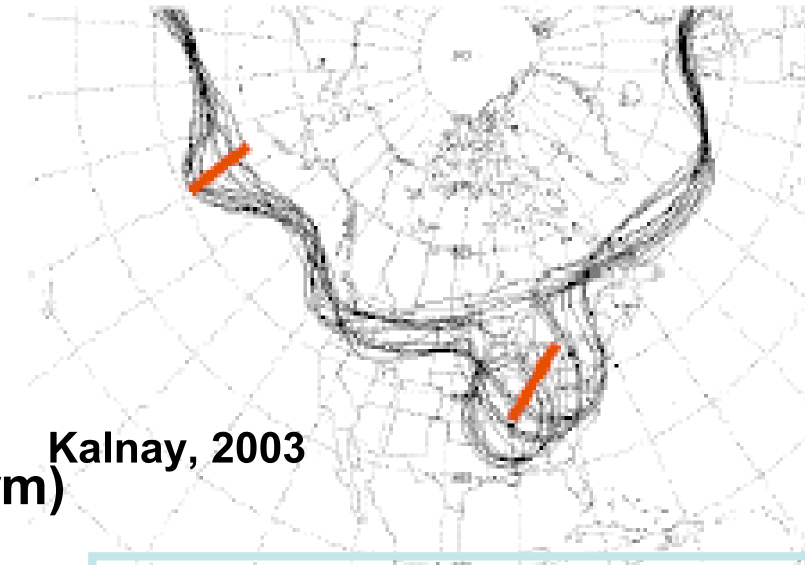
$$= \frac{1}{K-1}\mathbf{X}^b\mathbf{X}^{bT}$$

- Estimate error variance (diagonal) along with covariance (off-diagonal term)

- Ensemble member K(~100)<<n (10^7) grid points

  — Correlation exists between grid points which are only within a distance

  — The strong instabilities only have a few dominant shapes



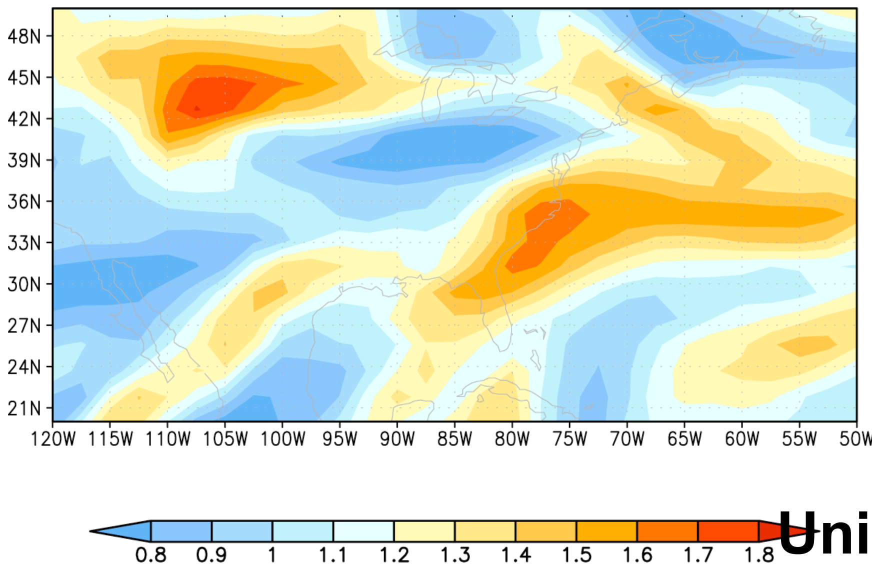
Kalnay, 2003

K ensemble forecasts of 500hPa height at a specific time

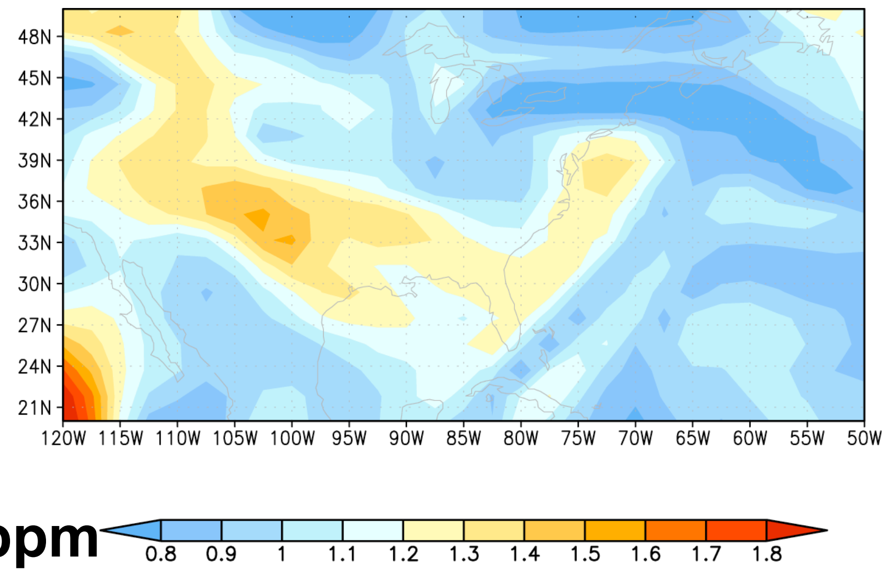
Forecast error statistics change with time

500hPa CO_2 ensemble spread from AIRS CO_2 assimilation

06Z18Feb2003



12Z19Feb2003



Unit: ppm

- **EnKF estimates time-changing forecast error**
- **CO_2 is a tracer, so affected by the time-changing dynamical flow**

The data assimilation process

Observations &
Observation error statistics

Forecast (background)&
Forecast error statistics

Data assimilation method

Compare and reject;
Interpolate and adjust

Analysis states

The data assimilation process

Observations &
Observation error statistics

Forecast (background)&
Forecast error statistics

Data assimilation method

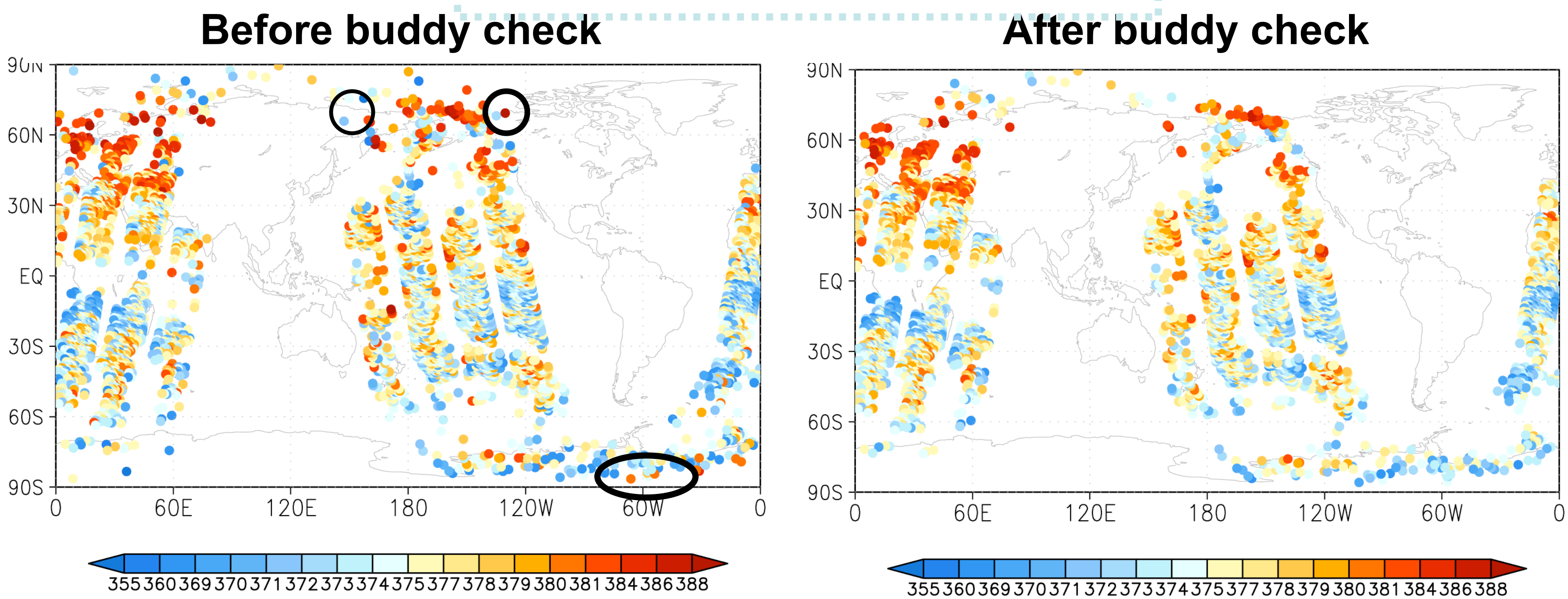
Compare and reject;
Interpolate and adjust

Analysis states

Compare and reject 1: buddy check

Buddy check: compare each obs to the mean of the adjacent obs

AIRS CO2 within 6 hours



The quality of the rejected obs is not necessarily bad by itself!

- **8% of AIRS CO2 observations were deleted in this way**

Compare and reject 2: compared to error statistics

Reject the obs when $abs(\mathbf{y}^o - \mathbf{y}^b)$ is larger than 5 (could be another number) times of observation error or forecast error

- **This method is only used in assimilating meteorological observations, not AIRS CO2 observations.**

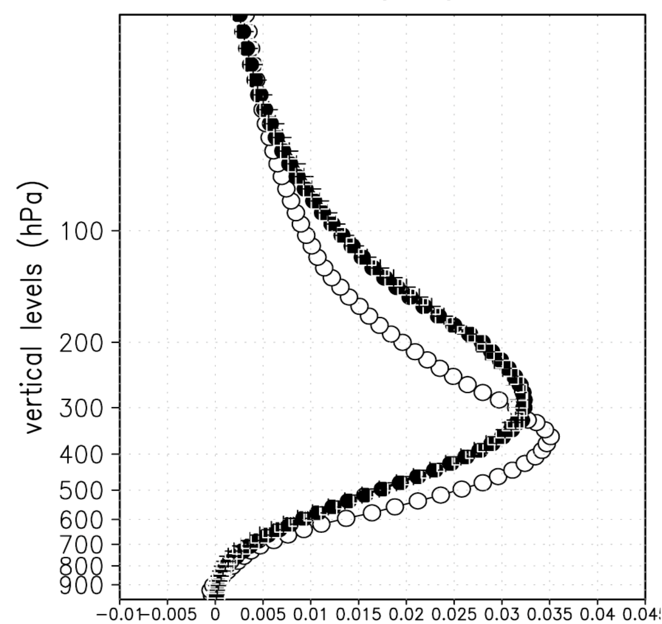
Interpolate and adjust (1)

Observation operator h(·): convert the model forecast (e.g. x^b=CO2(x,y,z,t)) to observation space e.g. CO2(x,y, t)

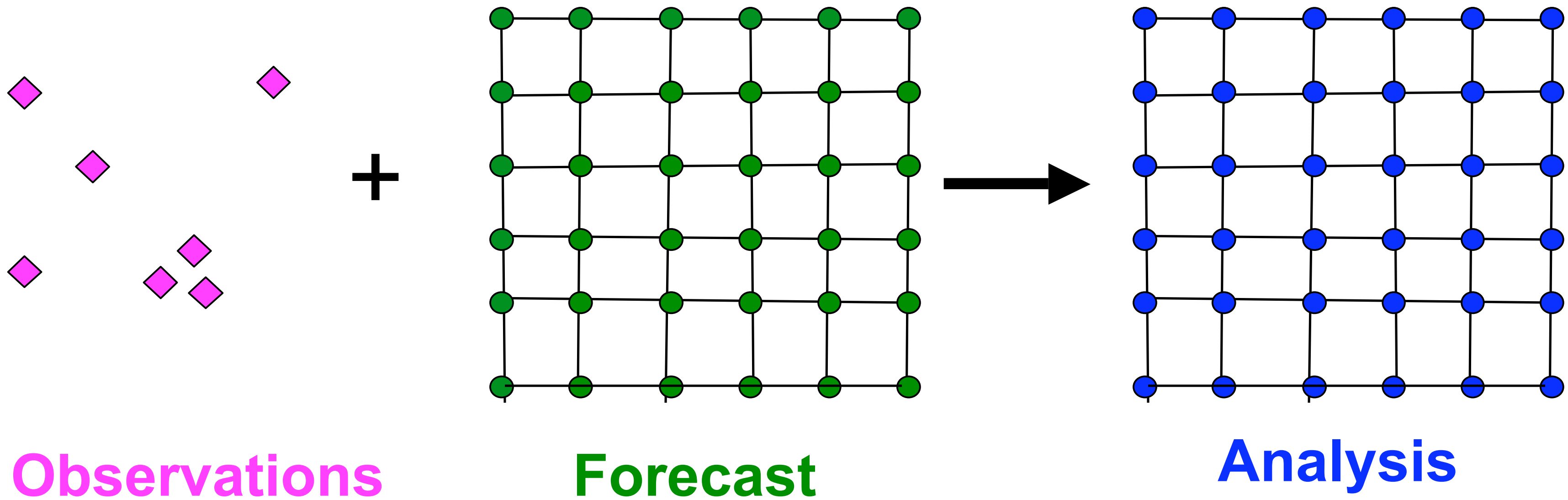
- **Assimilate AIRS CO2 retrievals:**

$$\underbrace{\mathbf{y}^b}_{\text{model forecast "obs"}} = \underbrace{\underbrace{\mathbf{A}}_{\text{avg kernal}}{}^T (\underbrace{S}_{\text{spatial interpolator}}}_{\text{obs operator}} (\underbrace{\mathbf{x}^b}_{\text{model forecast}}))$$

AIRS averaging kernel



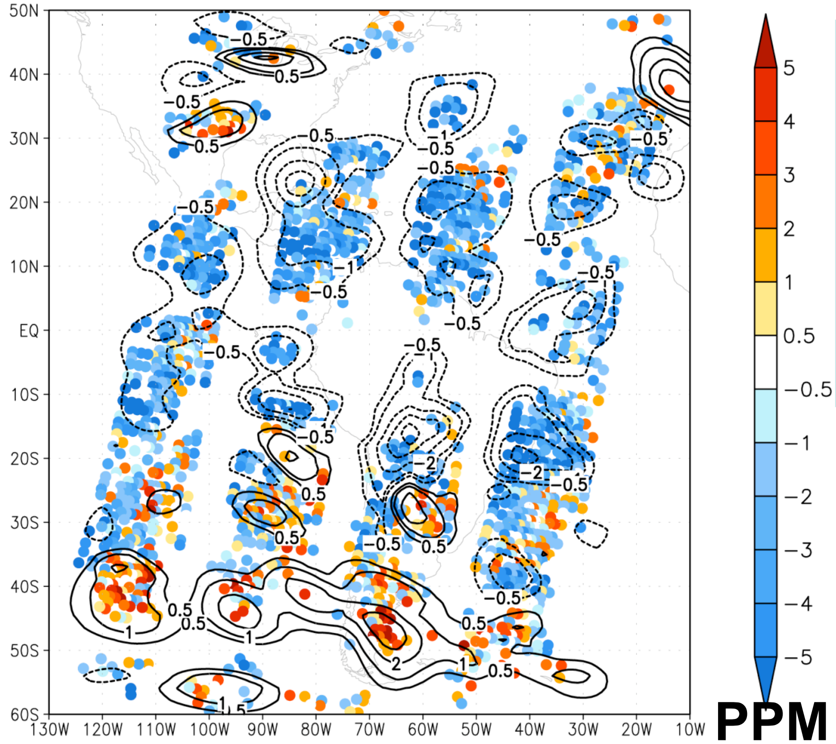
Interpolate and adjust (2)



- Fill in data void areas
- The interpolation and adjustment are based on the error statistics or numerical model

Example 1: horizontal interpolation

$$(\underbrace{\mathbf{x}^a}_{\text{analysis}} - \underbrace{\mathbf{x}^b}_{\text{forecast}})\ \text{(contour)};(\underbrace{\mathbf{y}^o}_{\text{observations}} - \underbrace{\mathbf{y}^b}_{\text{model forecast "obs"}})\ \text{(shaded)}$$

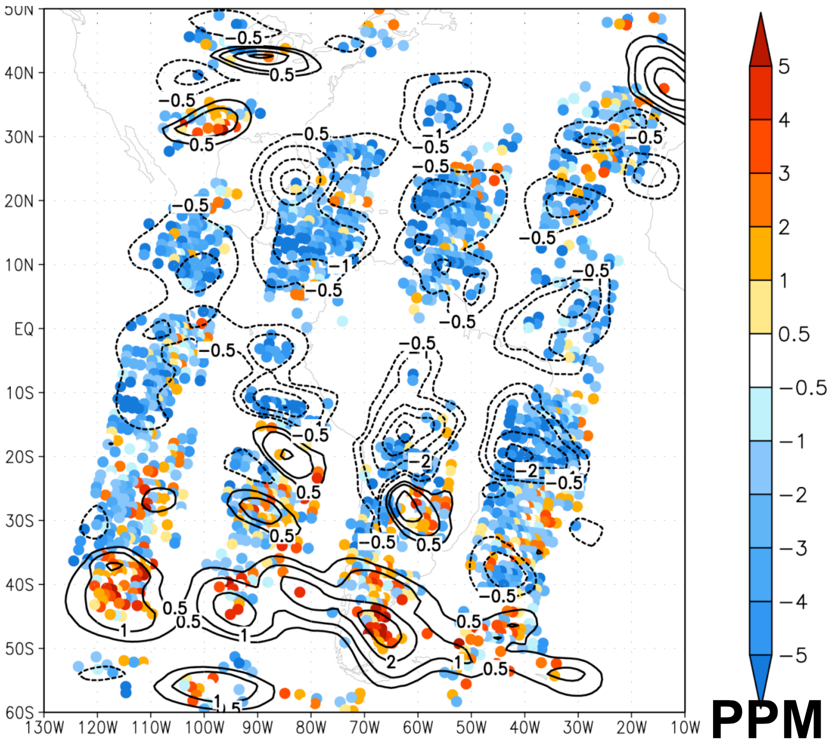


CO2 column integrated analysis increment and observation increment (AIRS CO2- model forecast "AIRS") agree with each other.

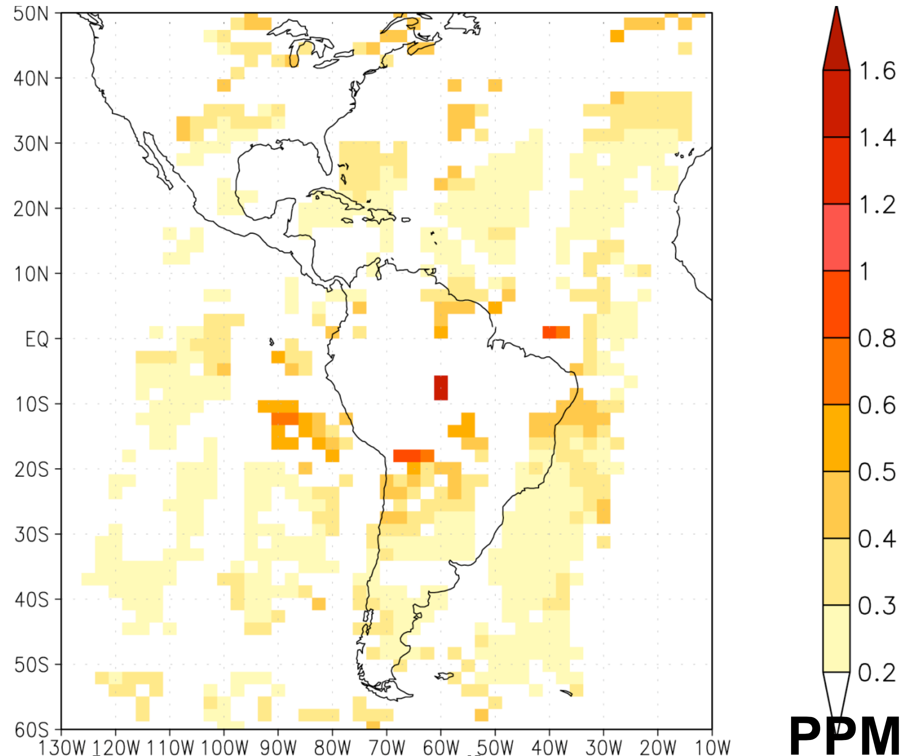
$$\underbrace{\overline{\mathbf{x}}^a}_{\text{analysis}} = \underbrace{\overline{\mathbf{x}}^b}_{\text{forecast}} + \underbrace{\mathbf{K}}_{\text{Kalman Gain}}(\underbrace{\mathbf{y}^o}_{\text{observations}} - \underbrace{\mathbf{y}^b}_{\text{model forecast "obs"}})$$

Error statistics are the bases for interpolation

Observation-forecast (shaded)
Analysis-forecast (contour)



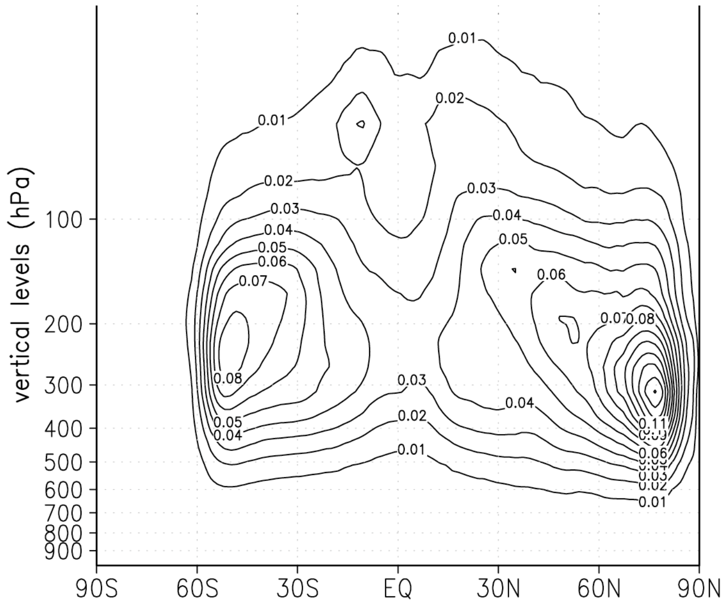
CO2 forecast ensemble spread
at AIRS CO2 space



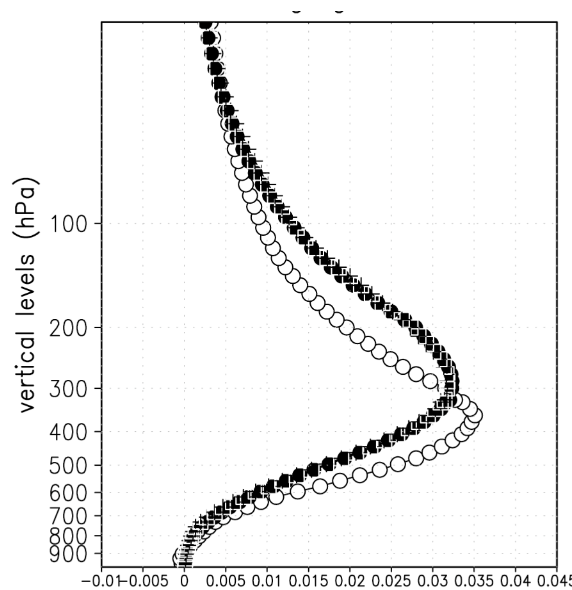
$$\underbrace{\mathbf{x}^a}_{\text{analysis}} = \underbrace{\mathbf{x}^b}_{\text{forecast}} + \underbrace{\mathbf{K}}_{\text{Kalman Gain}} \left(\underbrace{\mathbf{y}^o}_{\text{observations}} - \underbrace{\mathbf{y}^b}_{\text{model forecast "obs"}} \right)$$

Example 2: vertical interpolation

Time-averaged CO2 analysis increments



Averaging Kernel

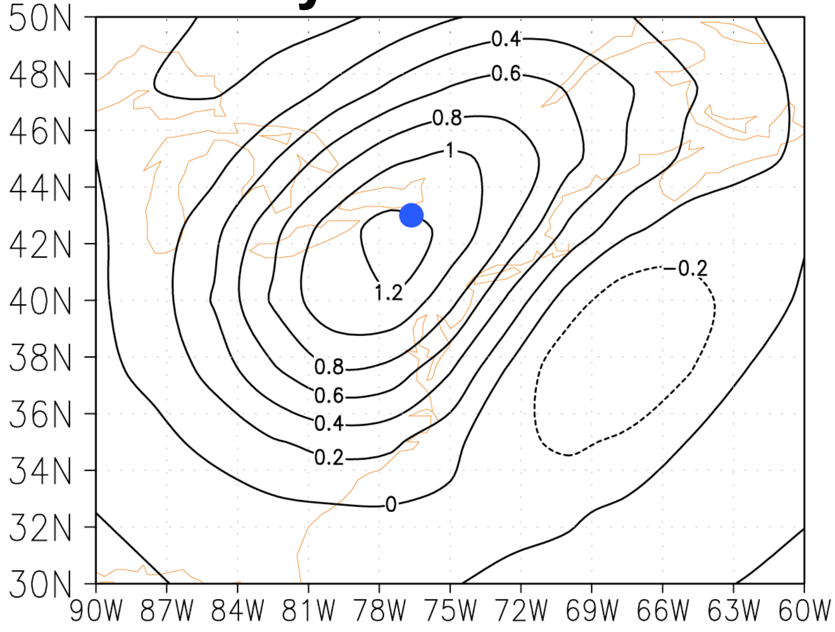


- **Obtain CO2 vertical profiles from column weighted CO2.**
- **Analysis increments peak at the similar levels of the peak of the averaging kernels.**

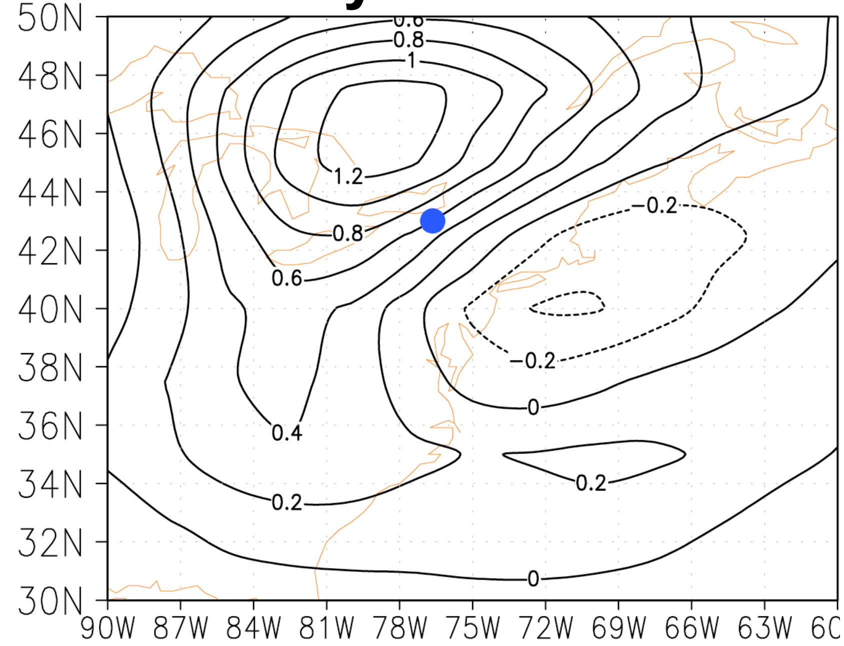
Example 3: inter-variable adjustment

Blue dot: Temperature (T) observation location

T analysis increment

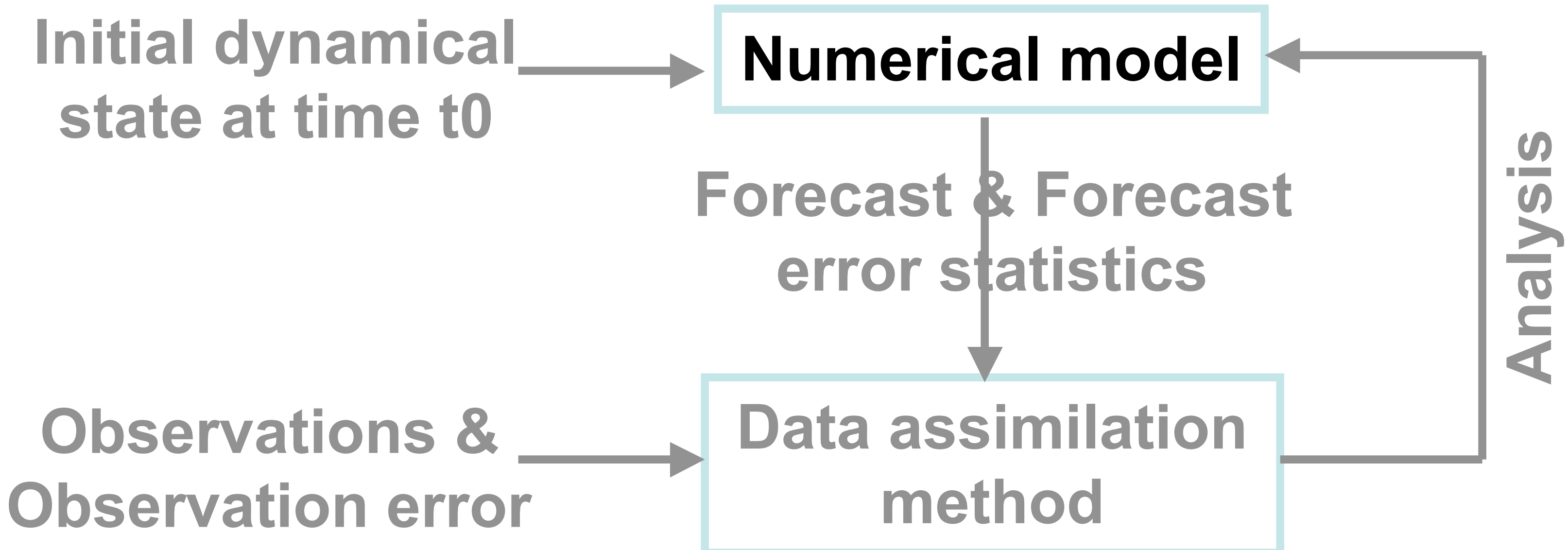


u analysis increment



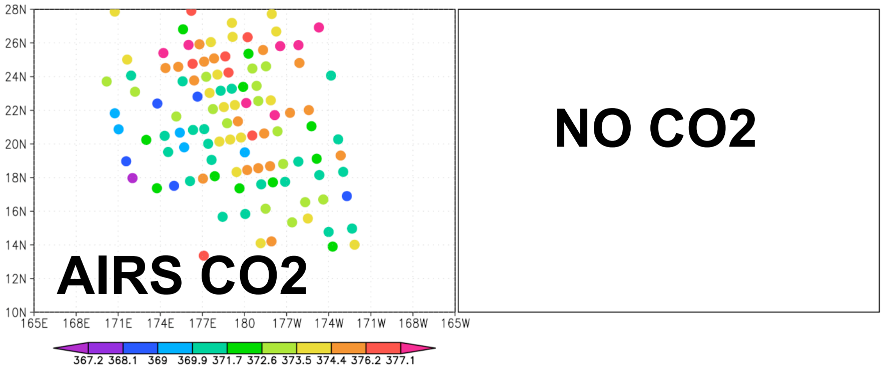
- **Zonal wind (u) analysis increments from assimilating T observation satisfy geostrophic balance**

Data assimilation cycle



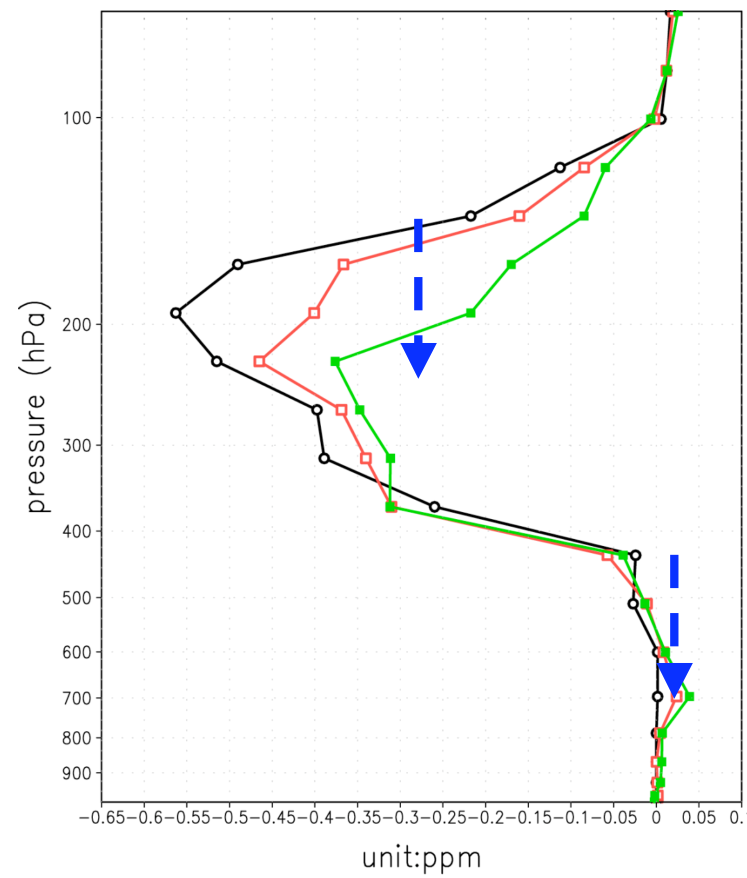
Propagate information through forecast

t0=00Z01Jan2003 t1=06Z01Jan2003



	exp1		exp2	
analysis at t0:	meteo+CO2	-	meteo	= black
	⇓		⇓	
forecast at t1:	$(CO2)_1$	-	$(CO2)_2$	= red
analysis at t1:	meteo+CO2	-	meteo	
	⇓		⇓	
forecast at t2:	$(CO2)_1$	-	$(CO2)_2$	= green

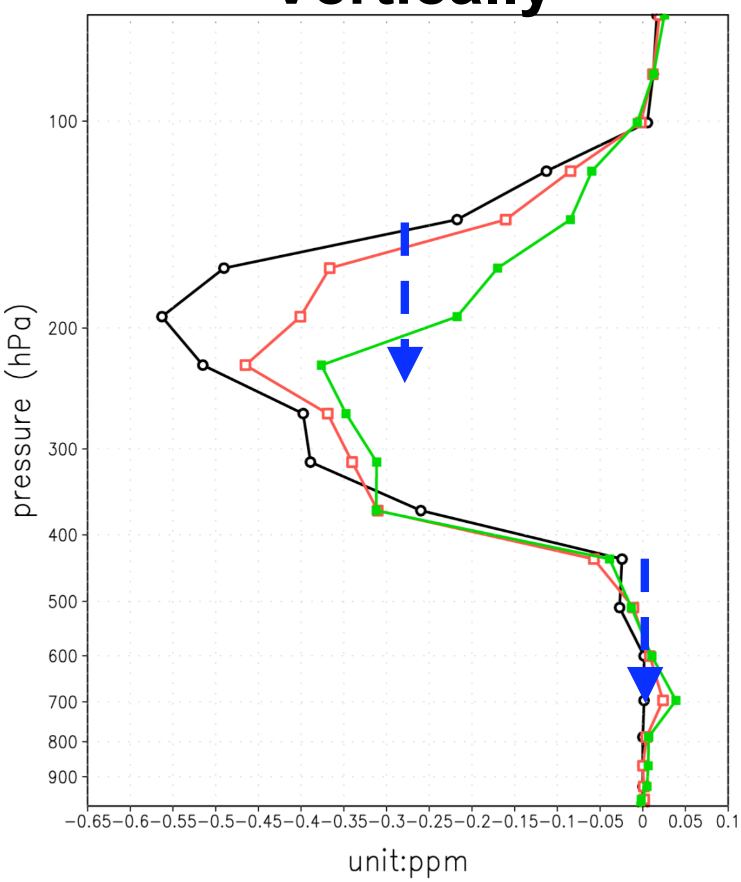
Vertically



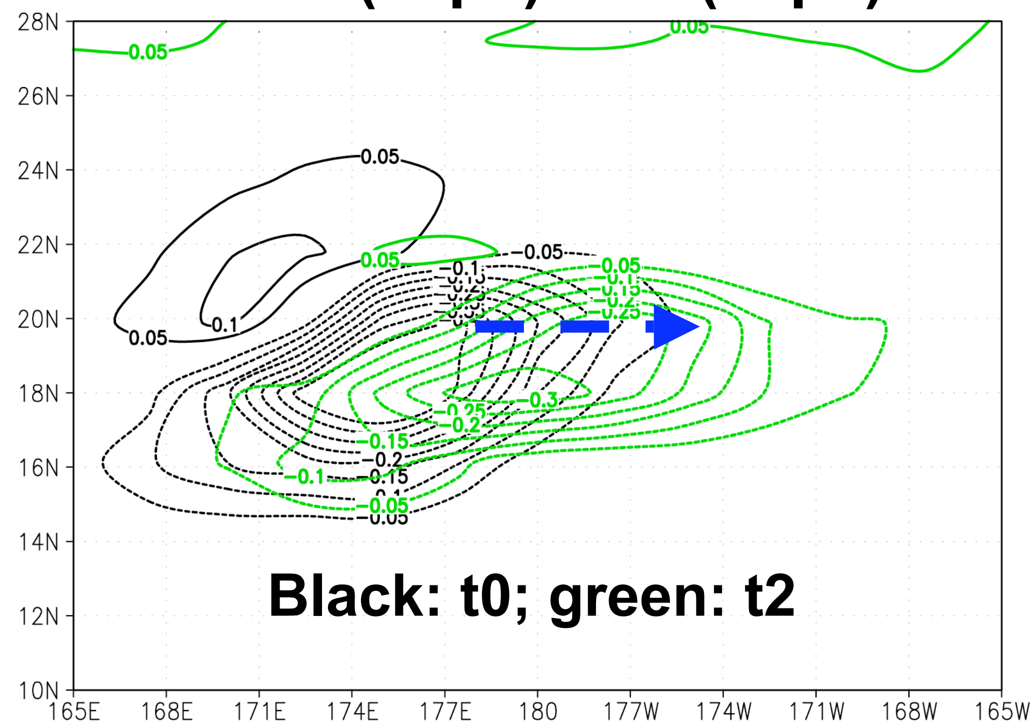
Lon=175ºE,lat=18°N

Propagate information through forecast

Vertically

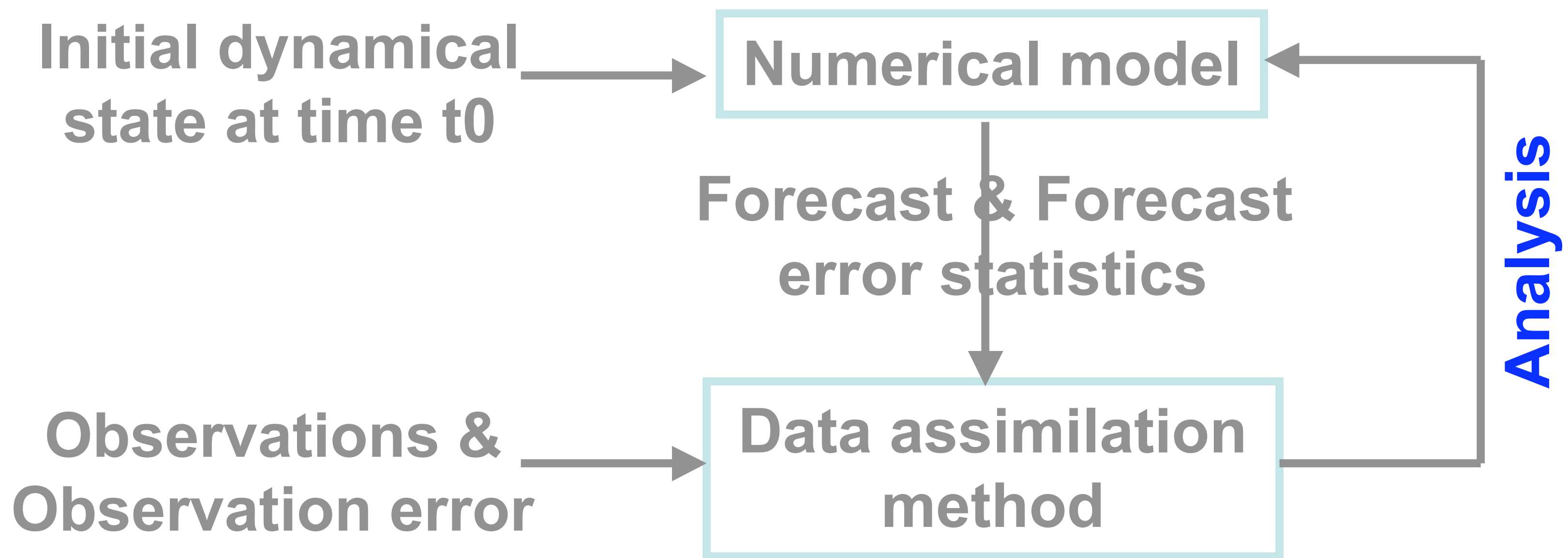


Horizontally 200hPa CO2(exp1)-CO2(exp2)



- **Propagate observation information both vertically and horizontally**

Data assimilation cycle



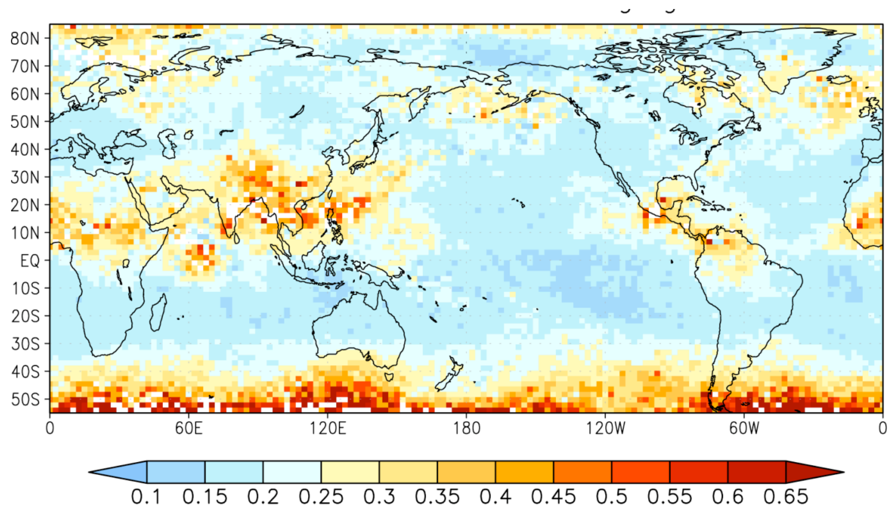
Analysis error statistics in EnKF

$$\underbrace{\mathbf{P}^a}_{\text{analysis error covariance}} \approx \frac{1}{K-1}\sum_{i=1}^{K}(\underbrace{\mathbf{x}_i^a}_{\text{analysis ensemble member}} - \underbrace{\overline{\mathbf{x}}^a}_{\text{analysis mean}})(\mathbf{x}_i^a - \overline{\mathbf{x}}^a)^T$$

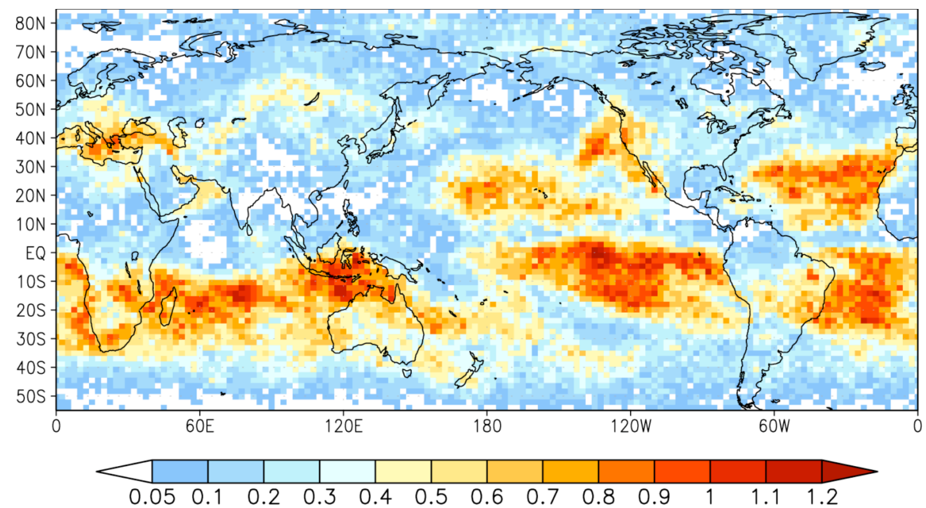
$$= \frac{1}{K-1}\mathbf{X}^a\mathbf{X}^{aT}$$

Relationship between analysis ensemble spread and observation coverage

CO2 analysis ensemble spread at observation space

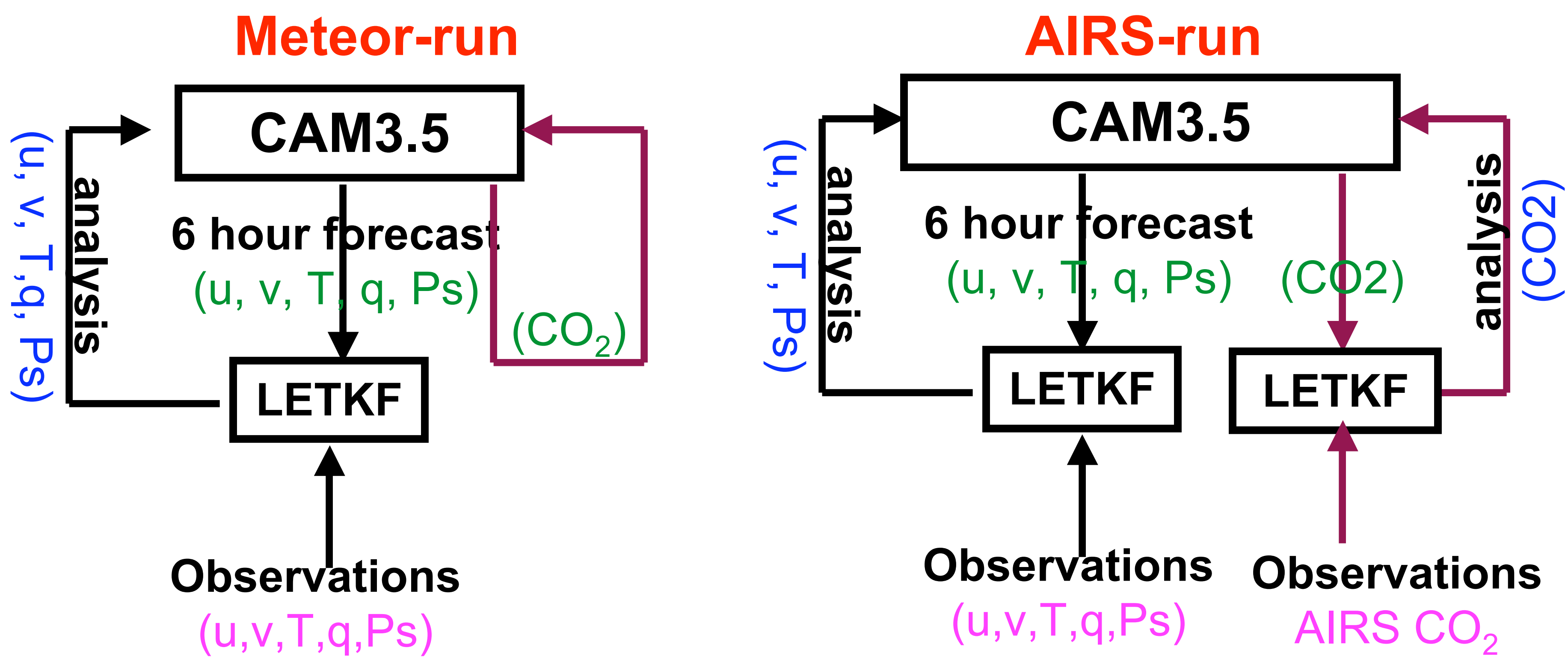


Average number of AIRS CO2 observations within 6-hour



- **Analysis ensemble spread is anti-correlated with the the CO2 observation coverage**

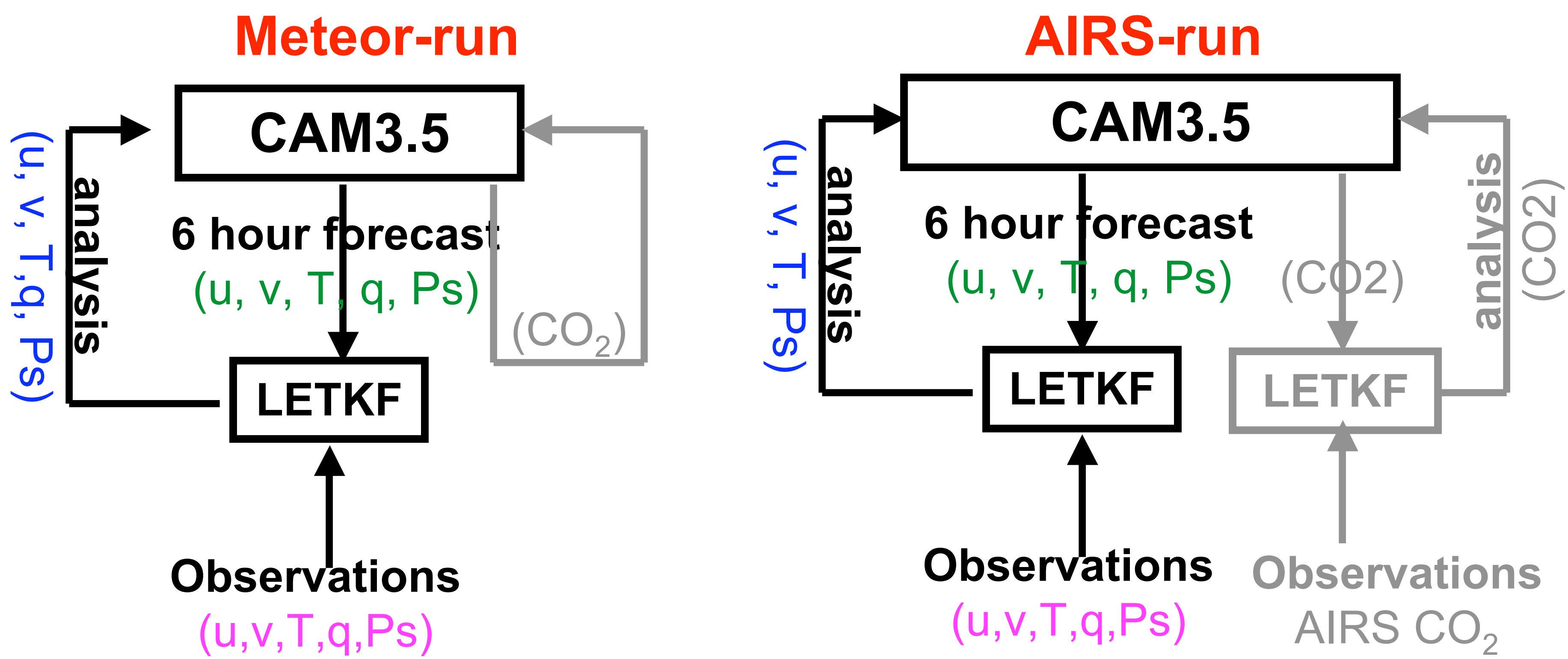
The impact of AIRS CO2 assimilation



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

- Assimilate meteorological observations along with AIRS CO2

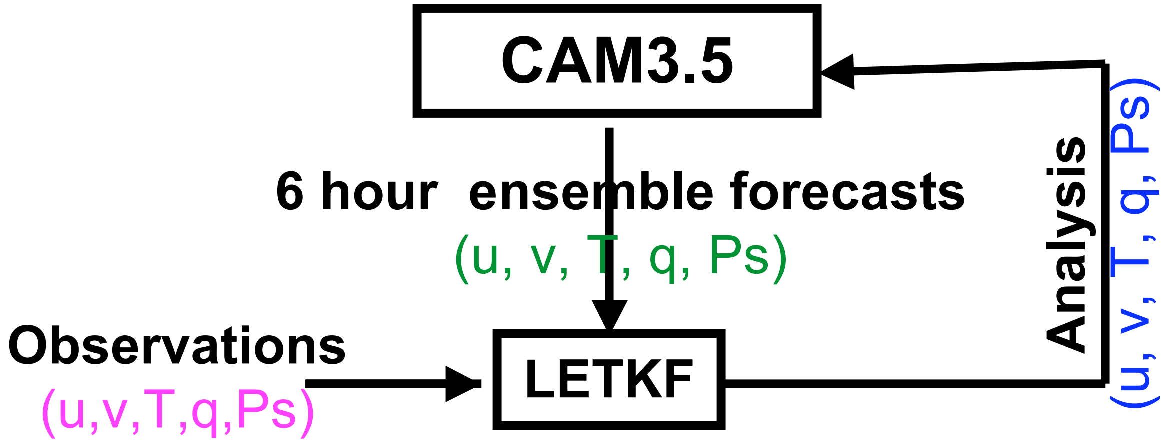
Why generating ensemble analyses of meteorological states?



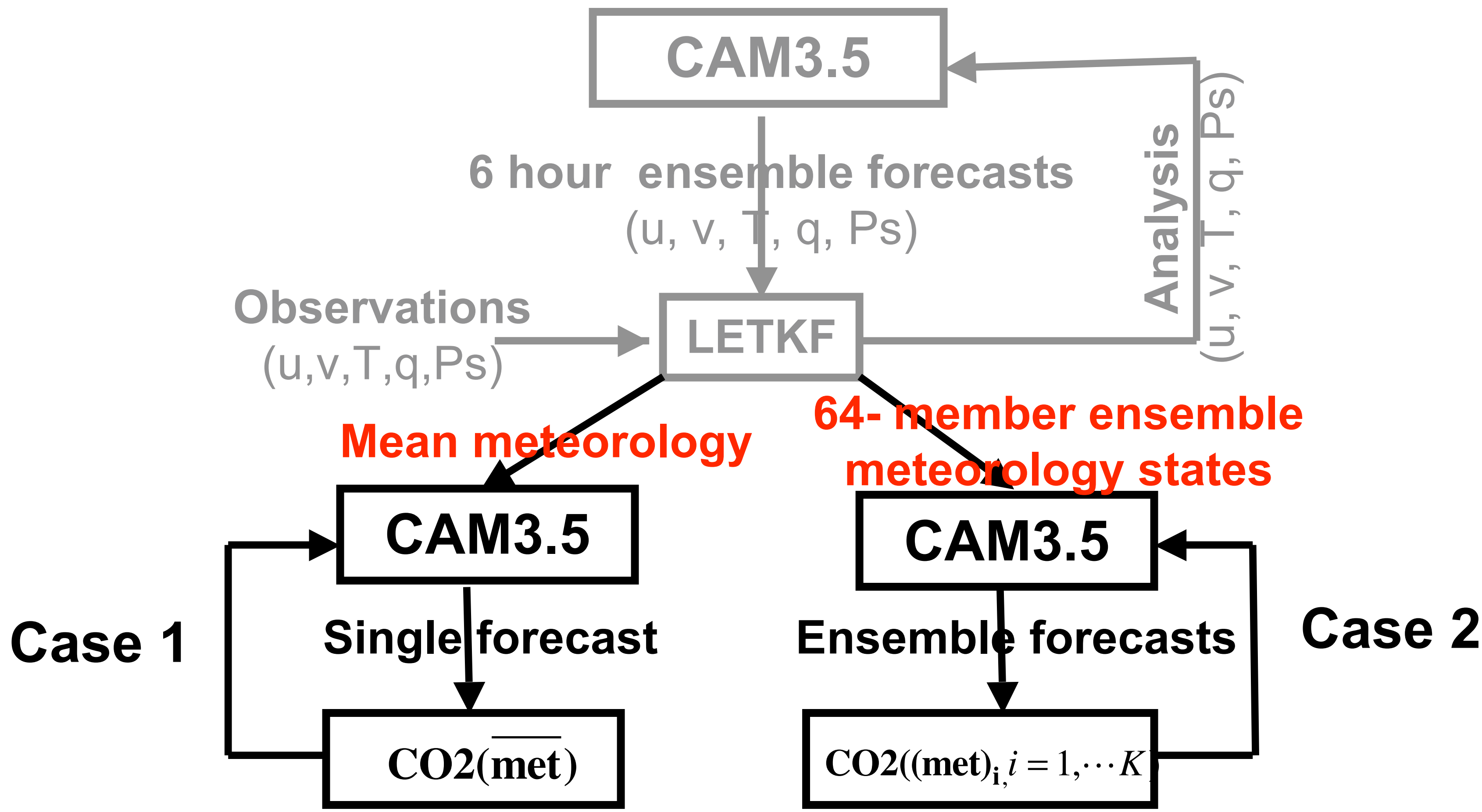
LETKF: Local Ensemble Transform Kalman Filter (Hunt et al., 2007)

- **Assimilate meteorological observations along with AIRS CO2**

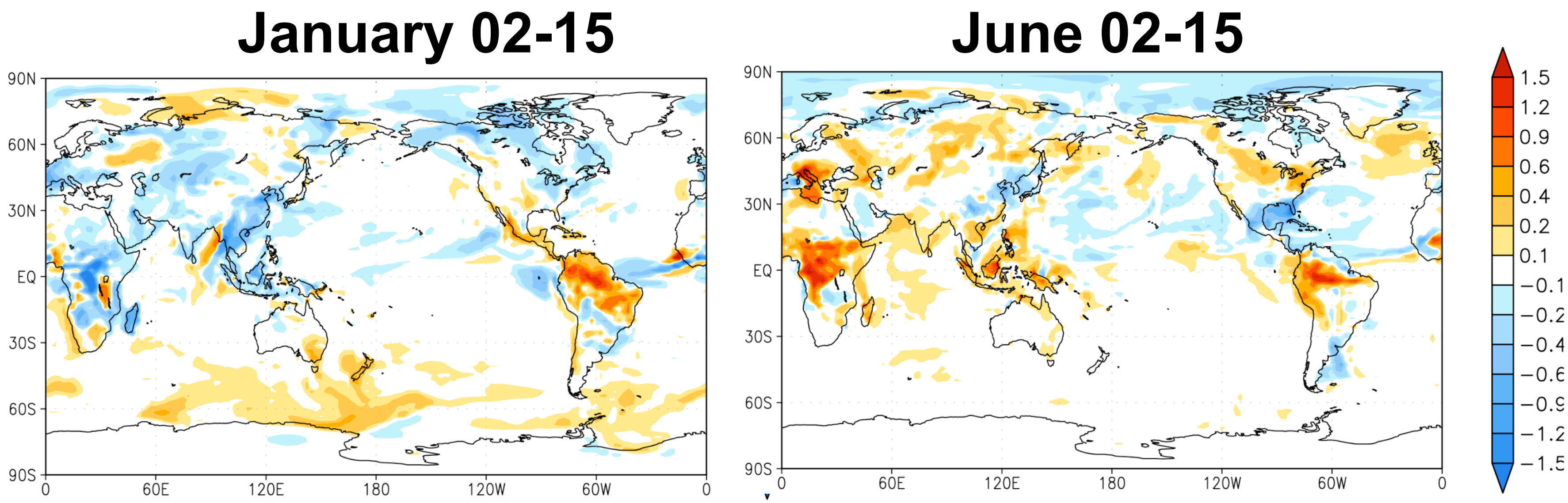
Ensemble data assimilation of meteorology observations



The impact of single and multiple meteorological states on CO2 forecast

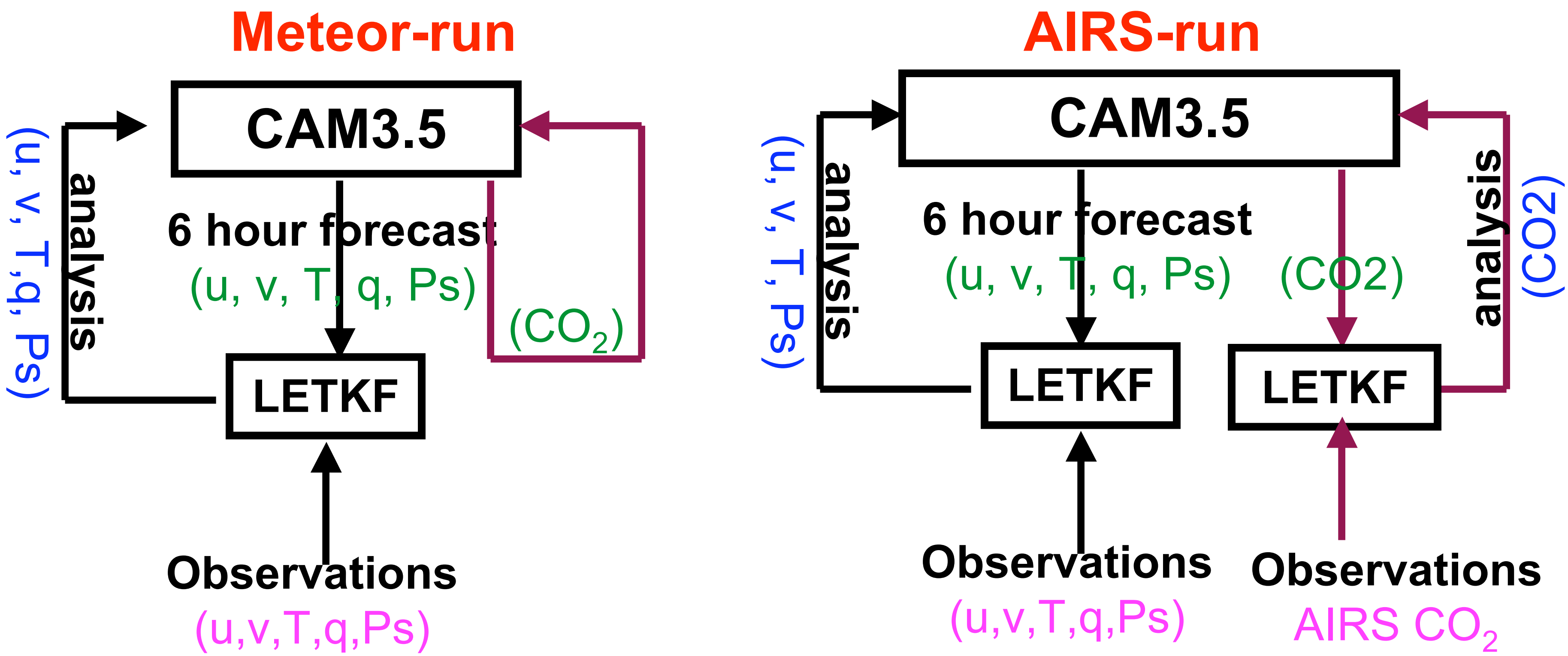


$$[\mathrm{CO2}(\overline{\mathrm{met}}) - \frac{1}{K}\sum_{i=1}^{K}\mathrm{CO2}((\mathrm{met})_{\mathrm{i}})] \text{ at surface}$$



- **Non-uniform in both space and time, with the difference as large as ±1.5ppm.**
- **The difference is due to the nonlinear interaction in the model alone.**

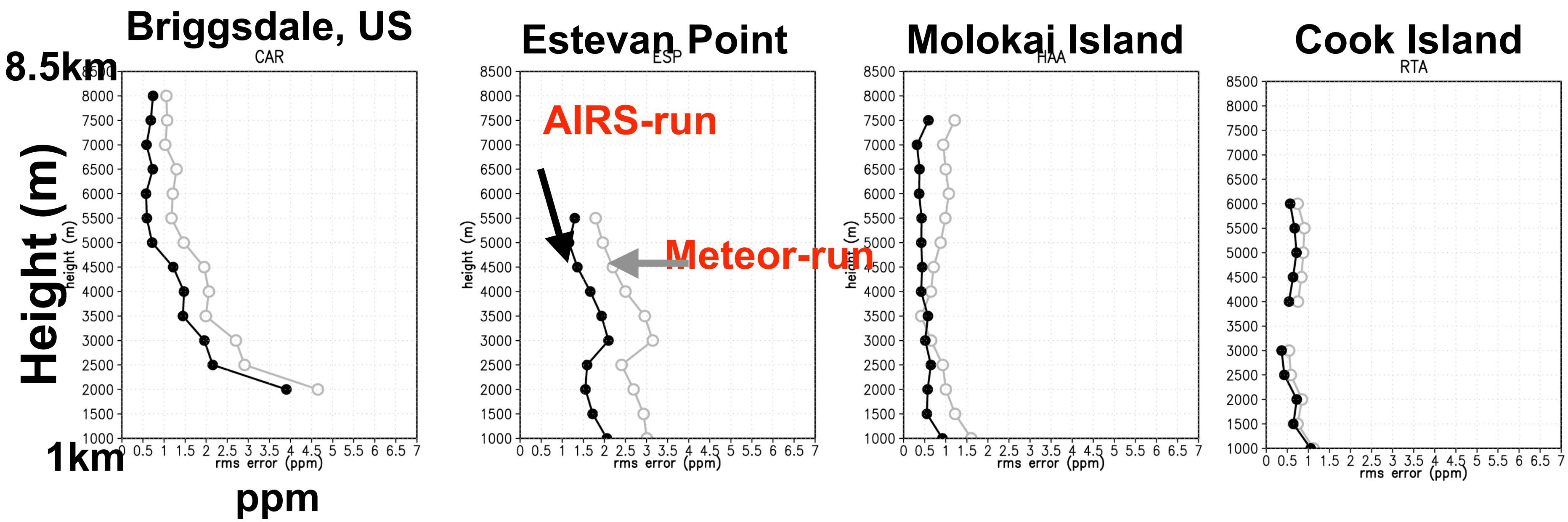
The impact of AIRS CO2 assimilation



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

- Assimilate meteorological observations along with AIRS CO2

Verified against independent aircraft CO2 (NOAA/ESRL)

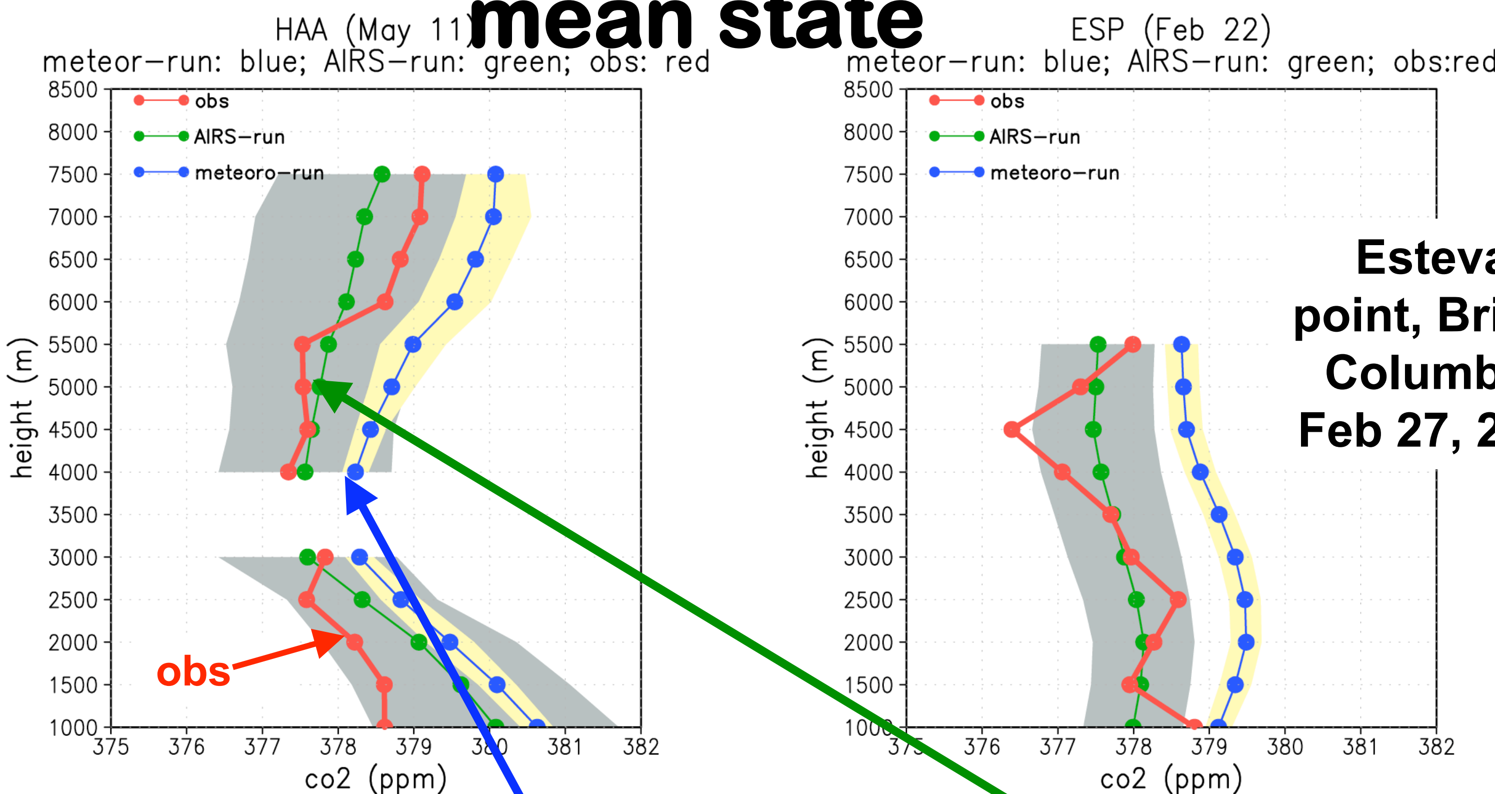


Time average of all the cases between 01Jan2003-30June2003

- **Grey: meteor-run; black: AIRS-run.**
- **CO2 vertical profiles from the AIRS-run can be about 1 ppm more accurate that those from the meteor-run.**

Analysis ensemble spread along with the mean state

Molokai Island, Hawaii. May 11, 2003



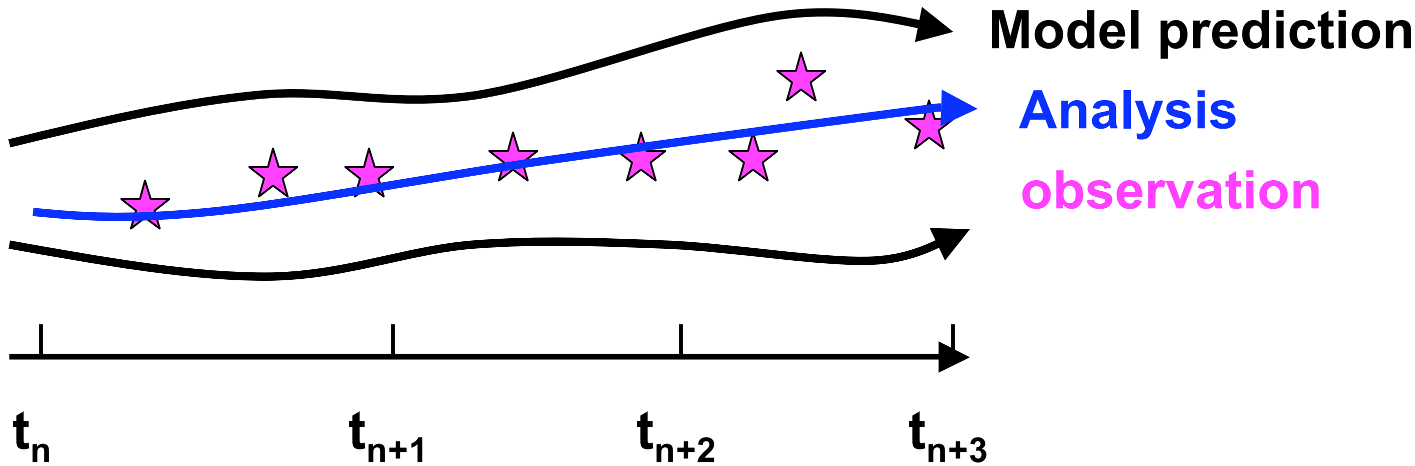
Estevan point, British Columbia, Feb 27, 2003

- Meteor-run: CO_2 tracer transported by 64-member ensemble meteorological analyses generated every 6hr --> "precision" of CO_2 "forecast" by model

•AIRS-run: CO_2 assimilated along with meteorological obs.

- Ensemble CO_2 analyses (grey shaded) bracket aircraft obs

Ongoing work: carbon flux estimation 4D-EnKF



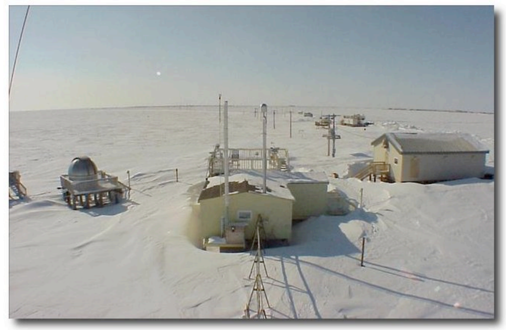
$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b \mathbf{A}^a \left(\sum_{l=1}^{P} (\mathbf{H}_l \mathbf{X}_l^b)^T \mathbf{R}_l^{-1} [\mathbf{y}_l^o - h_l(\bar{\mathbf{x}}_l^b)] \right)$$

- Assimilate observations at the right time;
- The analysis at time t_n is updated by all the observations within a time window based on the covariance.

(Hunt et al., 2004)

Application of 4D-EnKF on Carbon flux estimation

Carbon Source at A



CO2 obs at t_{n+1} at B

CO2 obs at t_{n+2} at C

CO2 observations at t_{n+1} at B and CO2 att$_{n+2}$ at C may all contain surface flux information at A.

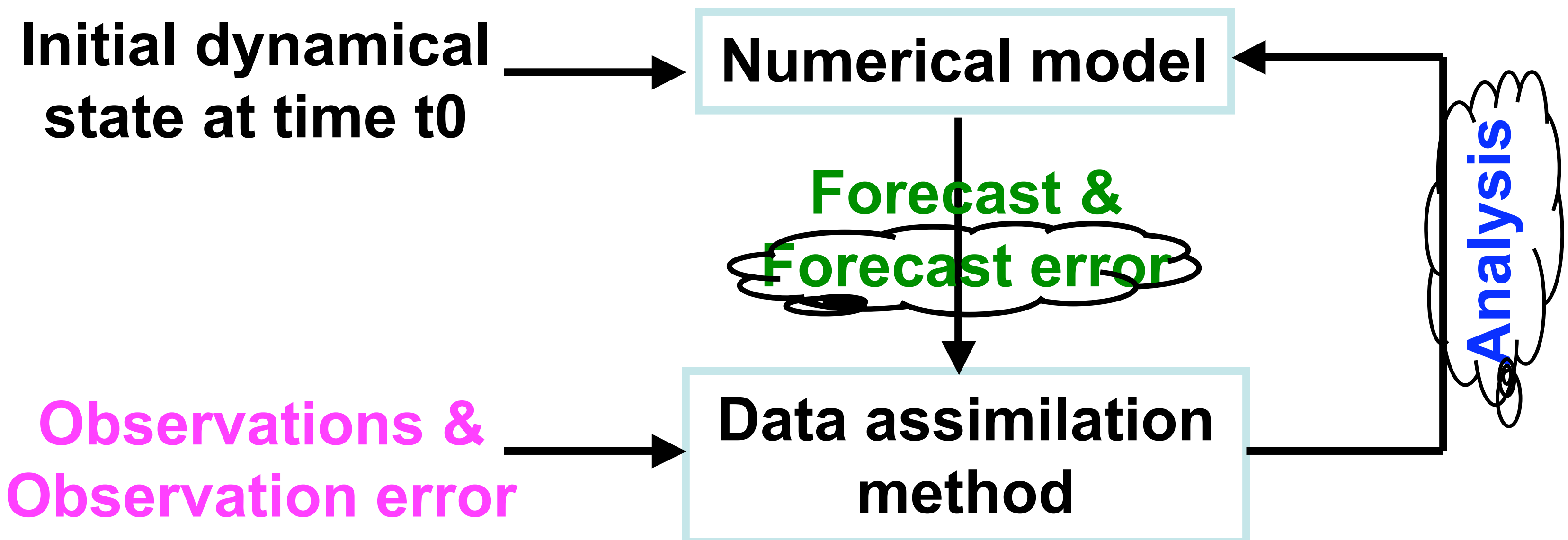
- With 4D-EnKF, naturally use the observations at time tn+1, tn+2, …. to estimate carbon flux at time tn

? Accurate estimate the covariance, the diffusivity of CO2

? How to decide the spatial and temporal range of the CO2 observations ?

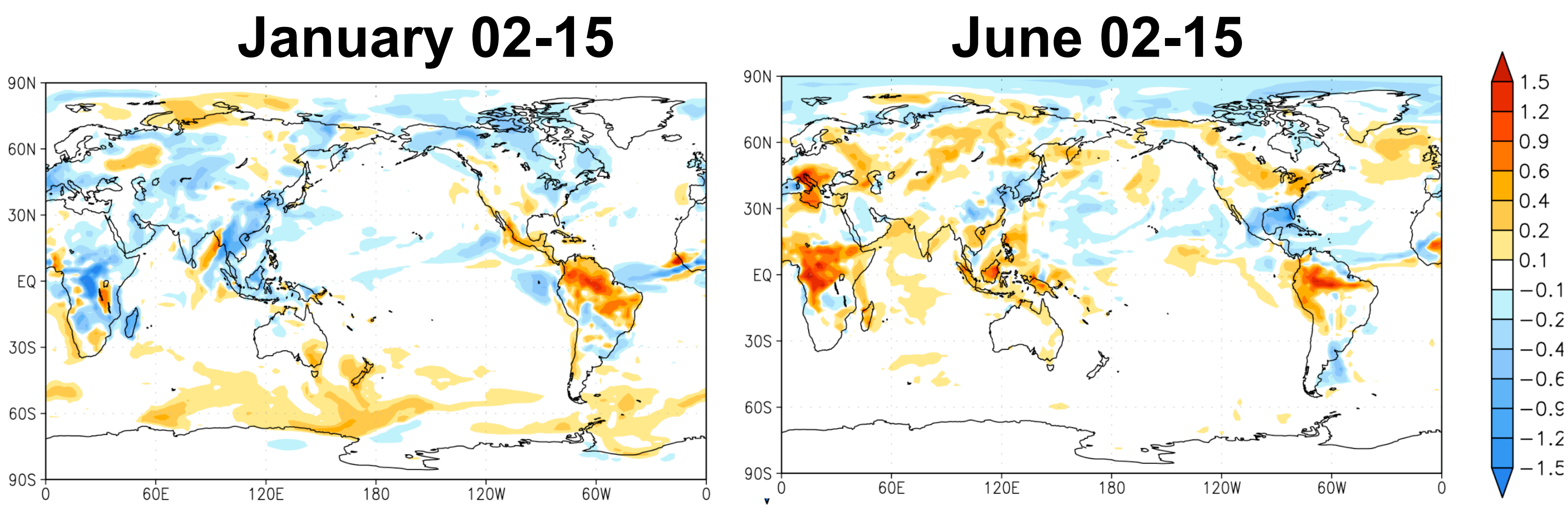
Summary (1)

- **Assimilation cycle and assimilation process.**



Summary (2)

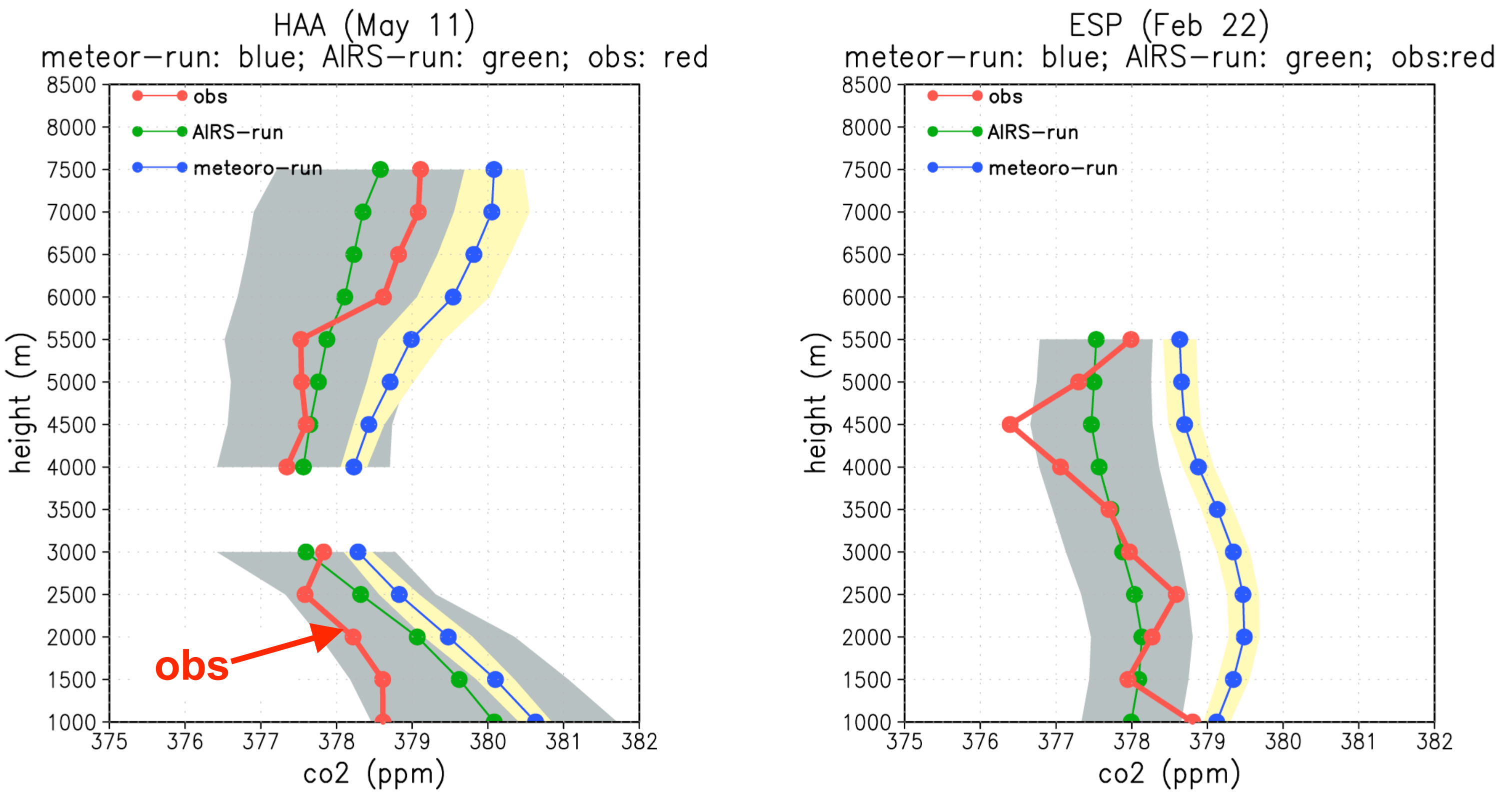
- **Non-uniform CO2 "bias" both spatially and temporally when transported by single meteorological state compared to ensemble meteorological states**



Summary (3)

Assimilation AIRS CO2 observations have improved the CO2 vertical profiles;

Ensemble Kalman filter estimates analysis uncertainty along with the mean.



Acknowledgement

- Thanks the guidance and collaboration from Dr. Inez Fung, Dr. Eugenia Kalnay, Dr. Moustafa T. Chahine, Dr. Edward T. Olsen.
- Thanks the help from Michael Wehner, Masao Kanamitsu, Kei Yoshimura, and Yu-heng Tseng
- Thanks the computational support from the National Energy Research Scientific Computing (NERSC) center.
- This project is supported by DOE.