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

Initial dynamical state at time $t_0$ → Numerical model → Forecast & Forecast error statistics → Analysis
Observations & Observation error → Data assimilation method
- 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

Initial dynamical state at time $t_0$ → Numerical model

Observations & Observation error → Data assimilation method

Forecast & Forecast error statistics

Analysis
Ensemble Kalman Filter (EnKF) process

\[
\begin{align*}
\bar{x}^a & = \bar{x}^b + K \left( y^o - h(\bar{x}^b) \right) ; K = F(\mathbf{R}, \mathbf{P}_b) \\
\end{align*}
\]

\[
\begin{align*}
\bar{x}^a & = \bar{x}^b + \left( \bar{P}^a \right)^{1/2} ; \bar{P}^a = F(\mathbf{R}, \mathbf{P}_b) \\
\end{align*}
\]
Ensemble Kalman Filter (EnKF) process

\[
\bar{x}^a = \bar{x}^b + K \left( y^o - h(\bar{x}^b) \right); K = F(\frac{R}{P_f}, P_f^b)
\]

\[
\tilde{\chi}^a = \tilde{\chi}^b + \frac{1}{2} \tilde{\chi}^a ; \tilde{\chi}^a = F(\frac{R}{P_f}, P_f^b)
\]

- 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

Initial dynamical state at time $t_0$ → Numerical model

Observations & Observation error → Data assimilation method

Forecast & Forecast error statistics → Analysis

Observations & Observation error → Data assimilation method

Forecast & Forecast error statistics → Numerical model

Analysis → Initial dynamical state at time $t_0$
Observation error statistics

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

$$
(y^o - y^a)(y^o - y^b)^T = R
$$

- $R$: 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:
  \[ P^b \approx \frac{1}{K-1} \sum_{i=1}^{K} (x^b_i - \bar{x}^b)(x^b_i - \bar{x}^b)^T \]
  \[ = \frac{1}{K-1} X^b X^{bT} \]

- Estimate error variance (diagonal) along with covariance (off-diagonal term)
Using ensemble method to estimate:

\[
P^b \approx \frac{1}{K-1} \sum_{i=1}^{K} (x_i^b - \bar{x}^b)(x_i^b - \bar{x}^b)^T
\]

\[
= \frac{1}{K-1} X^b X^{bT}
\]

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 CO2 ensemble spread from AIRS CO2 assimilation

06Z18Feb2003

12Z19Feb2003

Unit: ppm

- EnKF estimates time-changing forecast error
- CO2 is a tracer, so affected by the time-changing dynamical flow
The data assimilation process

Data assimilation method

- Compare and reject;
- Interpolate and adjust

Observations & Observation error statistics

Forecast (background) & Forecast error statistics

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

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 $\text{abs}(y^o - y^b)$ 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 $h(\cdot)$: convert the model forecast (e.g. $x^b=\text{CO2}(x,y,z,t)$) to observation space e.g. $\text{CO2}(x,y, t)$

- Assimilate AIRS CO2 retrievals:

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

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

\[
(x^a - x^b) \text{ (contour)}; (y^o - y^b) \text{ (shaded)}
\]

\[
\overline{x}^a = \overline{x}^b + K \left( y^o - y^b \right)
\]

CO2 column integrated analysis increment and observation increment (AIRS CO2- model forecast “AIRS”) agree with each other.
Error statistics are the bases for interpolation

\[
\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y}_o - \mathbf{y}_o^{\text{model forecast "obs"}})
\]

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

CO2 forecast ensemble spread at AIRS CO2 space
Example 2: vertical interpolation

- 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

- Zonal wind (u) analysis increments from assimilating T observation satisfy geostrophic balance
Data assimilation cycle

Initial dynamical state at time $t_0$ → Numerical model → Forecast & Forecast error statistics → Analysis

Observations & Observation error → Data assimilation method
Propagate information through forecast

\[ t_0=00Z01\text{Jan2003} \quad t_1=06Z01\text{Jan2003} \]

Vertically

Lon=175°E, lat=18°N

**AIRS CO2**

NO CO2

\[
\begin{align*}
\text{exp1} & \quad \text{exp2} \\
\text{analysis at } t_0: & \quad \text{meteo+CO}_2 & - & \text{meteo} & = \text{black} \\
& \quad \downarrow & & \downarrow \\
\text{forecast at } t_1: & \quad (\text{CO}_2)_1 & - & (\text{CO}_2)_2 & = \text{red} \\
\text{analysis at } t_1: & \quad \text{meteo+CO}_2 & - & \text{meteo} & \downarrow & \downarrow \\
\text{forecast at } t_2: & \quad (\text{CO}_2)_1 & - & (\text{CO}_2)_2 & = \text{green}
\end{align*}
\]
Propagate information through forecast

Vertically

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

Black: t0; green: t2

• Propagate observation information both vertically and horizontally
Data assimilation cycle

Initial dynamical state at time $t_0$ → Numerical model → Forecast & Forecast error statistics → Data assimilation method → Analysis

Observations & Observation error
Analysis error statistics in EnKF

\[ \mathbf{P}^a \approx \frac{1}{K-1} \sum_{i=1}^{K} (\mathbf{x}_i^a - \bar{\mathbf{x}}^a)(\mathbf{x}_i^a - \bar{\mathbf{x}}^a)^T \]

analysis error covariance

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

analysis ensemble member

analysis mean
Relationship between analysis ensemble spread and observation coverage

- Analysis ensemble spread is anti-correlated with the CO2 observation coverage
The impact of AIRS CO2 assimilation

**Meteor-run**

- CAM3.5
- LETKF
- Observations (u,v,T,q,Ps)
- 6 hour forecast (u, v, T, q, Ps)

**AIRS-run**

- CAM3.5
- LETKF
- Observations (u,v,T,q,Ps)
- Observations AIRS CO2
- 6 hour forecast (u, v, T, q, Ps) (CO2)

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

- **CAM3.5**
  - 6 hour ensemble forecasts: \((u, v, T, q, Ps)\)
  - Observations: \((u,v,T,q,Ps)\)

- **LETKF**
  - Analysis: \((u, v, T, q, Ps)\)
The impact of single and multiple meteorological states on CO2 forecast

CAM3.5

6 hour ensemble forecasts
(u, v, T, q, Ps)

Observations
(u, v, T, q, Ps)

LETKF

Mean meteorology

CAM3.5

CO2(met)

Case 1

Single forecast

64-member ensemble meteorology states

Case 2

Ensemble forecasts

CO2((met)i, i = 1, ... K)
\[
[\text{CO}_2(\text{met}) - \frac{1}{K} \sum_{i=1}^{K} \text{CO}_2((\text{met})_i)] \text{ at surface}
\]

- Non-uniform in both space and time, with the difference as large as \( \pm 1.5\text{ppm} \).
- The difference is due to the nonlinear interaction in the model alone.
The impact of AIRS CO2 assimilation

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

• Assimilate meteorological observations along with AIRS CO2
Verified against independent aircraft CO2 (NOAA/ESRL)

- CO2 vertical profiles from the AIRS-run can be about 1 ppm more accurate than those from the meteor-run.

Time average of all the cases between 01Jan2003-30June2003
Analysis ensemble spread along with the mean state

- Meteor-run: CO2 tracer transported by 64-member ensemble meteorological analyses generated every 6hr --> “precision” of CO2 “forecast” by model

- AIRS-run: CO2 assimilated along with meteorological obs.

- Ensemble CO2 analyses (grey shaded) bracket aircraft obs
Ongoing work: carbon flux estimation

4D-EnKF

\[
\bar{x}_n^a = \bar{x}_n^b + X_n^b A^a \left( \sum_{l=1}^{P} (H_l X_l^b)^T R_l^{-1} [y_l^o - h_l(\bar{x}_l^b)] \right)
\]

- 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)
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 at $t_{n+2}$ 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.

Diagram:
- Initial dynamical state at time $t_0$ → Numerical model → Forecast & Forecast error → Data assimilation method → Analysis
- Observations & Observation error → Data assimilation method
Summary (2)

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

**January 02-15**

**June 02-15**
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.