Observations for Carbon Data Assimilation

Scott Doney
Woods Hole Oceanographic Institution

Where does the “data” come from for “data assimilation”?

**Atmospheric CO\textsubscript{2} data**
- initial conditions
- innovation terms
- error covariance terms

**Land and ocean carbon data**
- flux estimates for priors
- process/mechanistic information

Data is NOT a black box!!
Atmospheric Inverse Modeling of $CO_2$

Concentration (observed samples) + Transport (modeled) = Sources & Sinks (solved for)
How data enters the problem (1)

Variational Assimilation
Adjust model state “x” (atmospheric CO$_2$ field) to minimize cost function $J$:

$$ J(x) = \frac{1}{2} \left\{ (x - x_b)^T B^{-1} (x - x_b) + \left[ y_o - H(x) \right]^T R^{-1} \left[ y_o - H(x) \right] \right\} $$

So where do we get:

- $y_0(t)$ data for innovation (model-data misfit)
- $R$ data error covariance
- $B$ model error covariance
**Some Issues to Ponder**

**Atmosphere CO\(_2\) drastically under-sampled**
- original design for marine background air
- mostly discrete surface samples
- NWP deg. freedom \(O(10^7)\); 6 hourly observations \(O(10^4-10^5)\)
- \(CO_2\) weekly observations \(O(10^2-10^3)\)

**Small signal on large background variability**
- surface North-south difference \(~2.5 \text{ ppmv}\)
- zonal continent to land contrast \(<1 \text{ ppmv}\)
- measurement precision
- accuracy in time & across stations and networks

reduce systematic biases!!

\[
y_{obs} = y_{true} + \varepsilon_{obs}
\]

\[
\varepsilon_{obs} = \varepsilon_{random} + \varepsilon_{systematic}
\]

\[
E[\varepsilon_{obs}^1, \varepsilon_{obs}^2] \neq 0
\]
Some Issues to Ponder

Representativeness of data \( y_0 \)

• “footprint” of observation & mismatch with model grid
• local heterogenity or point sources
• aliasing of unresolved frequencies/wavenumbers (e.g., diurnal cycle)
• data selection (i.e., exclude “unrepresentative” observations)

\[
R = R_{\text{instrument}} + R_{\text{representativeness}}
\]
CO₂ Concentration in the Outer Damon Room, NCAR Mesa Lab, 2/7 - 2/9/06
Multiple Time/Space Scales

- Century
  - 1-D BGC
- Decade
- Year
  - Tower Flux
  - Tower Footprint
- Day
  - Landscape
  - Region
  - Aircraft Flux
- Hour
  - Plot
  - Continent
  - Globe
  - GCM Carbon
  - Satellite GPP, NPP
  - Orbiting Carbon Observatory
In-situ Atmospheric Observing Network

- Discrete surface flasks (~weekly)
- Continuous surface (hourly) observatories
- Tall towers continuous (hourly)
- Aircraft profiles (~weekly)
Surface Observatory

Graph showing CO₂ measurements from a Surface Observatory in Barrow, Alaska. The graph illustrates the carbon cycle with CO₂ concentrations from 2002 to 2006. The data includes hourly averaged values from the semi-continuous measurement system and measurements thought to be representative of baseline conditions. The data is represented by different symbols and colors to indicate the reliability and origin of the measurements.
Seasonal $CO_2$: Continental vs Oceanic Sites

- $CO_2$ seasonal cycle attenuate, but still coherent, far away from source/sink region
- Peak-trough amplitude of seasonal cycle ~ 30 ppmv (~10%)
Weather Cycles: CO₂ is variable

Modeled: one day in July

On DT~5 days, Dx ~ 1000 km
CO₂ variability in the boundary layer ~ 10 ppmv (3 %)

[TURC/NDVI Biosphere; Takahashi Ocean; EDGAR Fossil Fuel [U. Karstens and M. Heimann, 2001]

[LSCOP, 2002]
Diurnal CO$_2$: Highly variable in boundary layer

- Diurnal cycle of photosynthesis and respiration
  - $> 60$ ppmv (20%) diurnal cycle near surface
- Varying heights of the planetary boundary layer (varying mixing volumes)
Vertical Profiles (free troposphere)
Aircraft Campaigns

COBRA 2000
Surface Fluxes => Atmospheric CO2
Orbiting Carbon Observatory
(Planned Fall 2008 launch)

- Estimated accuracy for single column ~1.6 ppmv
- 1 x 1.5 km IFOV
- 10 pixel wide swath
- 105 minute polar orbit
- 26° spacing in longitude between swaths
- 16-day return time
How data enters the problem (2)

Separating transport, initial conditions & surface fluxes

\[ x^{i+1}_b = M(x^i_a) \]  Analysis at time \( i \) \( \Rightarrow \) forecast at time \( i+1 \)

\[ x^{i+1}_b = \Phi(x^i) + G(u^i) \]

transport fluxes

\[ x^0 \neq x^0_{\text{prior}} \]  \( \text{4D Variational methods: adjust initial conditions to better match future data} \)

\[ J(x) = \frac{1}{2} \left\{ (x^0 - x^0_{\text{prior}})^T B^{-1} (x^0 - x^0_{\text{prior}}) + [y_o - H(x)]^T R^{-1} [y_o - H(x)] + (u - u_{\text{prior}})^T P^{-1} (u - u_{\text{prior}}) \right\} \]

Deviation of initial conditions from “prior”  Deviation of \( x \) from “observations”  Deviation of fluxes from “prior”
$\text{NEE} = \text{GPP} - R_{\text{eco}}$

NEE is measured at the tower

Ecosystem Respiration typically based on nighttime NEE & air temperature & ?

$R_{\text{eco}} = R_{\text{hetero}} + R_{\text{auto}}$

GPP Photosynthesis (Daytime only)

CO$_2$

Adapted from Gilmanov et al.
Eddy-Flux Towers

Vertical velocity (mean and anomaly)
\[ w = \bar{w} + w' \]

CO2 concentration (mean and anomaly)
\[ c = \bar{c} + c' \]

Vertical CO2 flux
\[ wc = (\bar{w} + w')(\bar{c} + c') \]
\[ = \bar{w}\bar{c} + w'\bar{c} + \bar{w}c' + w'c' \]

\[ \bar{wc} = \bar{w}\bar{c} + w'c' \]
FluxNet Tower Sites
Diurnal and Seasonal Cycle
1. Variability is at a maximum on the strongly forced time scales
2. They have an annual sum of ~0
3. Modeling the carbon storage time scales is much more difficult
The Flux Footprint:

- What Part of the Ecosystem does the Flux Sensor ‘see’?
- Is that Part Representative of the Ecosystem? (answer varies over time)
- If yes: use data; if not: reject data

e.g.: Schmid (2002, Ag. For. Met., 113, 159-184)
Forest Inventory Analysis: Slow Process Observations

- Plot-scale measurement of carbon storage, age structure, growth rates: 170,000 plots in US!
- Allows assessment of decadal trends in carbon storage
Air-Sea $\text{CO}_2$ Flux Estimates

Takahashi

-undersampling & aliasing of surface water $\Delta \text{pCO}_2$

-transfer velocity $k$ empirically derived from wind speed relationships

$F = k_s (p\text{CO}_2^\text{air} - p\text{CO}_2^\text{water}) = ks\Delta \text{pCO}_2$

Kinetics

transfer velocity

Thermodynamics
Generating a Global Flux Map
In-Situ Sensors and Autonomous Platforms

Moorings/drifters ($D_\text{pCO}_2$, pH, DIC, NO$_3$)

Profiling ARGO floats (also AUVs, caballed observatories)
JGOFS/WOCE global survey (1980s and 1990s)

- Global baseline (hydrography, transient tracers, nutrients, carbonate system)
- Improved analytical techniques for inorganic carbon and alkalinity (±1-3 mmol/kg or 0.05 to 0.15%)
- Certified Reference Materials
- Data management, quality control, & public data access
Ocean Inversion Method

- Ocean is divided into n regions \( (n = 30, \text{ aggregated to 23}) \)
- Basis functions for ocean transport created by injecting dye tracer at surface in numerical models
Simple Inversion Technique

Basis functions are model simulated footprints of unit emissions from a number of fixed regions.

- Estimate linear combination of basis functions that fits observations in a least squares sense.

 Premultiply both sides by inverse of $A$

\[
A\phi = c \\
\hat{\phi} = A^{-1}c
\]

® Inversion is analogous to linear regression
INVERSION: AIR-SEA CO₂ FLUX COMPONENTS

Gruber et al. (2006)