Data Assimilation – An Overview

Outline:

- What is data assimilation?
- What types of data might we want to assimilate?
- Optimum Interpolation
- Simple examples of OI
- Variational Analysis a generalization of OI
- Bayesian methods (simple example)
- Integration
- Common problems

• Rarely is one method used exclusively

Quote--

- **Lao Tzu, Chinese Philosopher**

"He who knows does not predict.
He who predicts does not know."

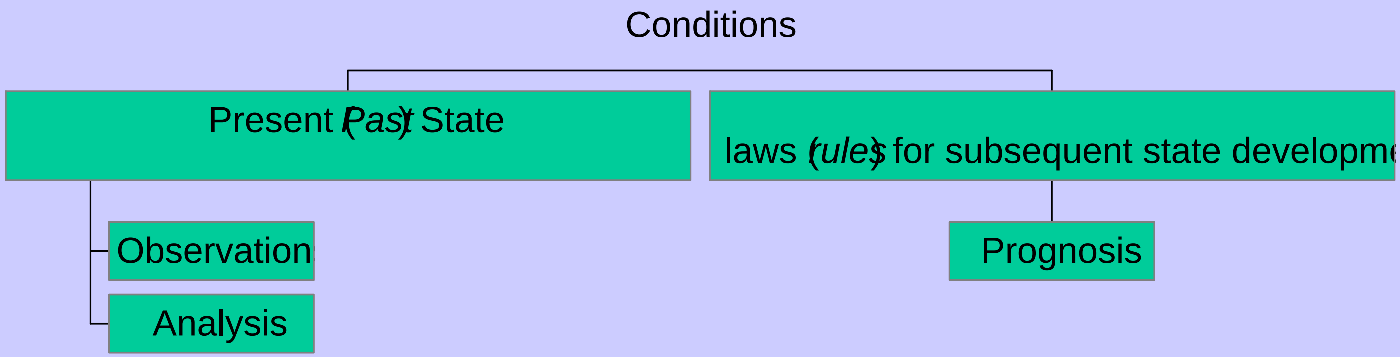
The Purpose of Data Assimilation

- To combine measurements and observations with our knowledge of statistics and the physical system's behavior as modeled to produce a "best" estimate of current conditions.
- The analysis has great diagnostic value and is the basis for numerical prediction.
- It allows to control of model error growth

Why use a (forecast) model to generate background field?

- Dynamical consistency between mass and motion
- Advection of information into data-sparse regions
- Improvement over persistence and climatology
- Temporal continuity

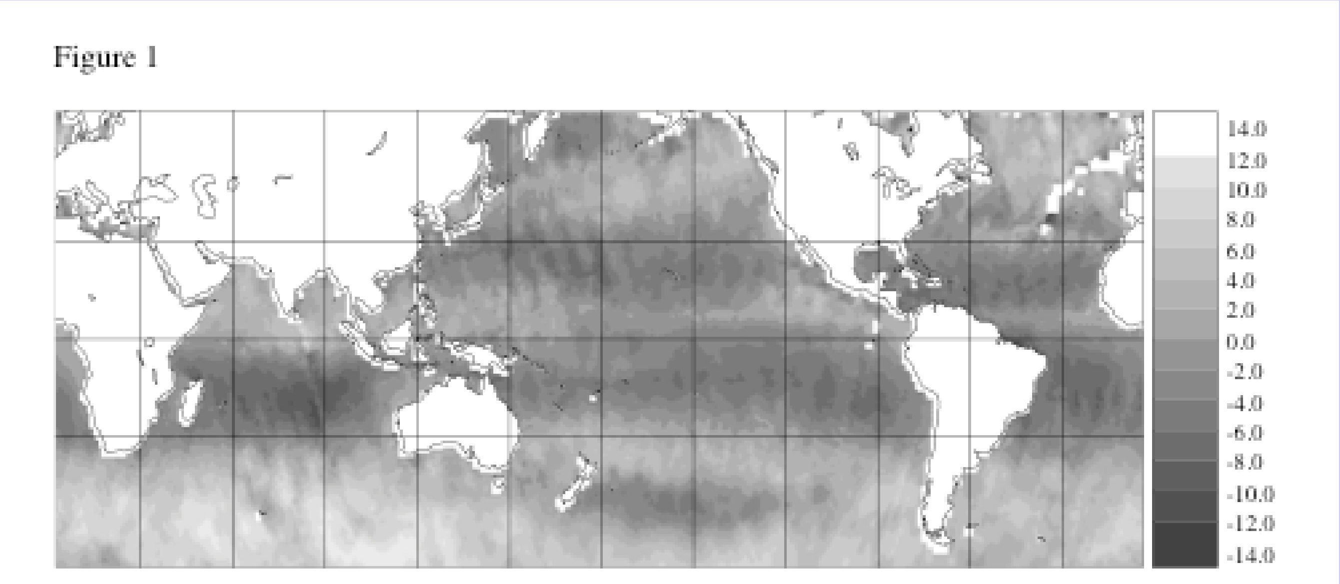
Necessary Conditions to Predict



The Data Assimilation Cycle

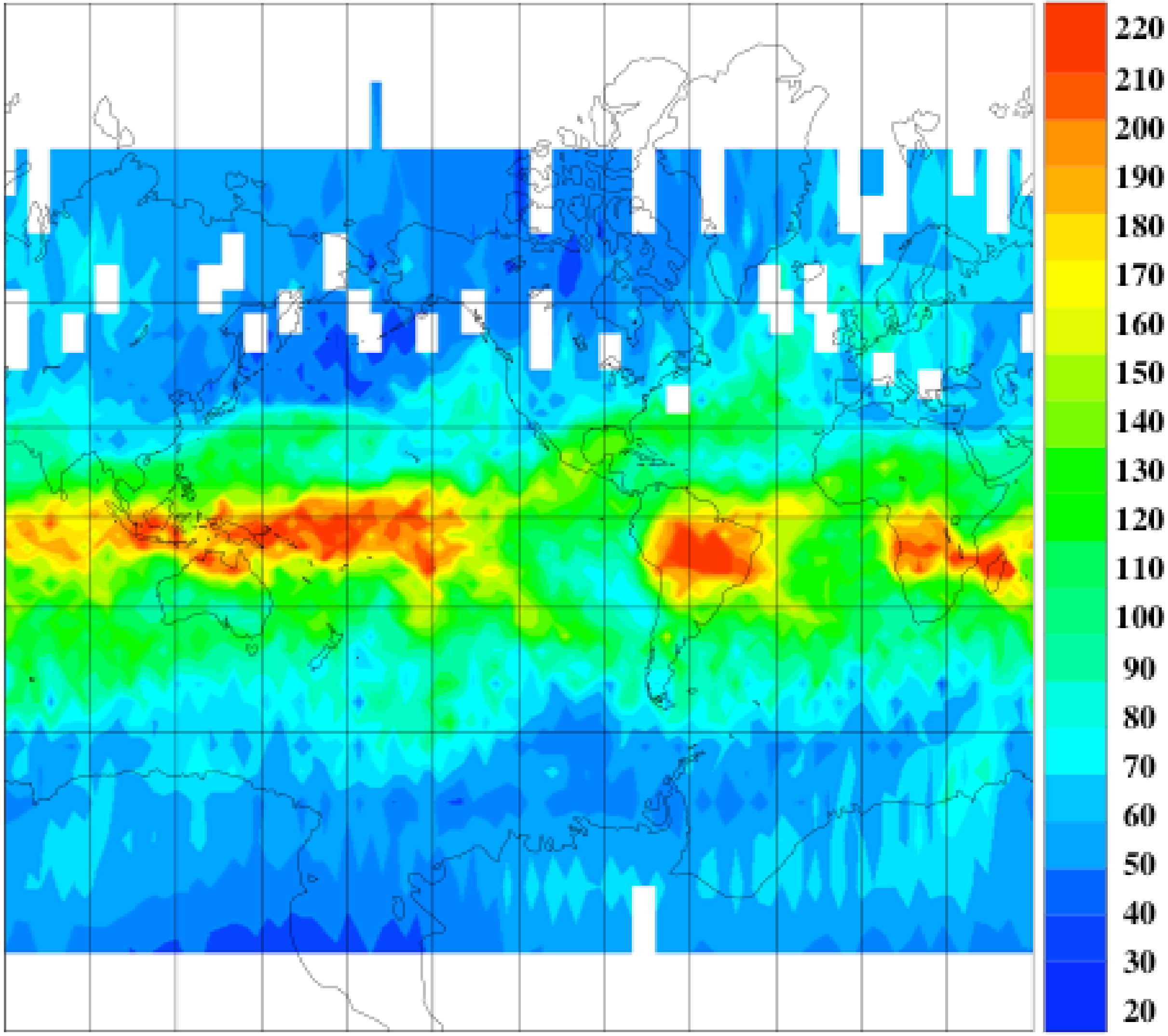
- Quality Control
- Objective Analysis (estimating parameters at points where no observations are available, usually on a grid)
- Initialization
- Short forecast prepared as next background field

Discussion break: what are unique challenges and advantages of assimilating remote-sensor data?



- Satellite sampling, coverage, representativeness and variable tradeoffs

Figure 2


220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20

Data Assimilation Cycle

- Data Assimilation: the fitting of a prediction model to observed data (reality check, Daley text).

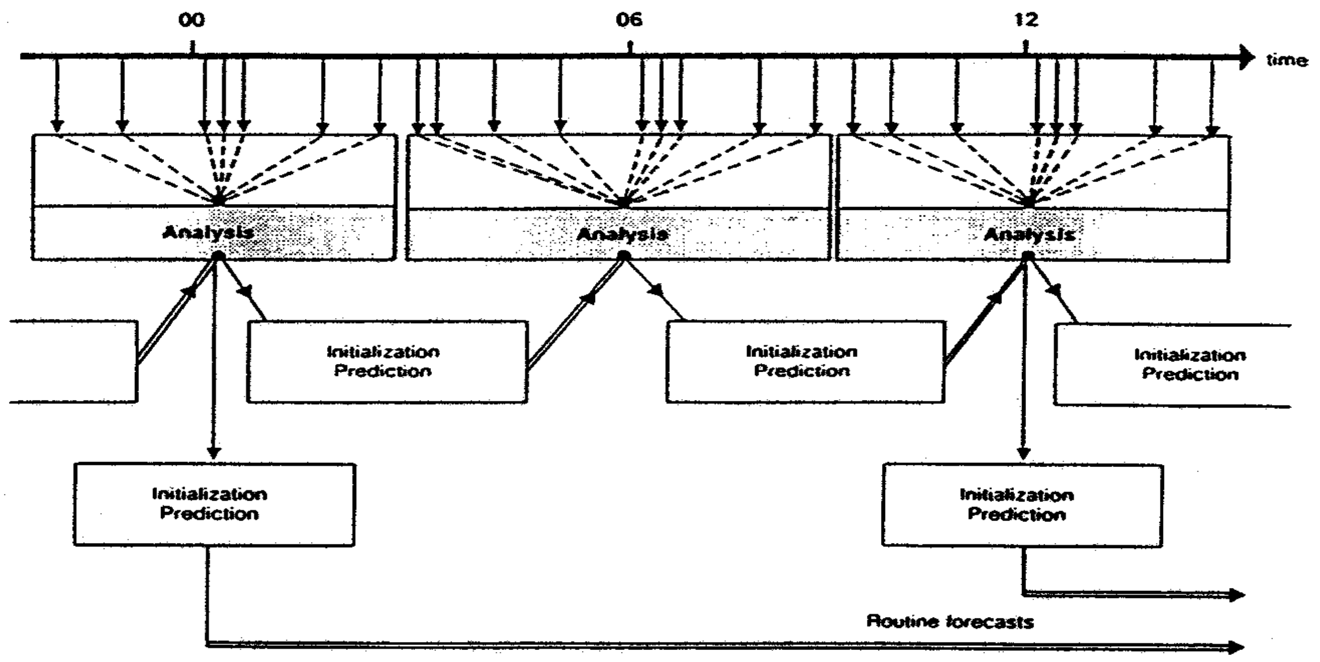


Figure 1.11 Schematic diagram of intermittent data assimilation cycle.

Objective Analysis Methods

– Surface Fitting
  - ordinary/weighted least squares fitting
  - principal components (EOF) improve numerical stability
  - spline functions control smoothness
  - cross-validation controls smoothness and sharpness of filters
– Empirical linear interpolation
  - successive correction techniques
– Optimal Interpolation

Above still widely used in research, less so in operations

– Bayesian Approach
– Newer hybrid techniques (variants of OI)
  - Adaptive filtering (Kalman filter)
  - Spectral Statistical Interpolation
  - Nudging and 4D Variational analysis

A simple problem

- Estimate an unknown quantity x from two collocated measurements y_1 & y_2 subject to errors

ε_1 & ε_2 :

$$y_1 = x + \varepsilon_1 \qquad\qquad y_2 = x + \varepsilon_2$$

Assume: $\mathrm{E}(\varepsilon_1) = \mathrm{E}(\varepsilon_2) = 0$; $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0$

errors are random & uncorrelated

$\mathrm{E}(\cdot)$ is the statistical mean

Now define the variance: $\sigma_1^2 = \mathrm{E}(\varepsilon_1^2)$; $\sigma_2^2 = \mathrm{E}(\varepsilon_2^2)$

Form a linear estimate of x : $x' = a_1 y_1 + a_2 y_2$

as unbiased estimate ($E(x' - x) = 0$; long term error of estimate $= 0$)

Simple Problem (continued)

- Constraints imply:

$$\text{Sum of the weights:} \qquad a_1 + a_2 = 1$$

$$\text{Finally, minimize the variance of the error of estimates:}$$

$$\sigma^2 = E[(x' - x)^2]$$

Solution for a_1, a_2 is a function of σ^2

Equivalent Problem

- Find an estimate ξ of x that is close to the observations.
- Do this by minimizing the “distance” (penalty/cost function) between ξ and the observations

$$J(\xi) = \frac{(\xi - y_1)^2}{\sigma_1^2} + \frac{(\xi - y_2)^2}{\sigma_2^2}$$

In the above, observed error variances act as weights (the larger the error associated with an obs., the smaller its weight in the estimate.

The ξ that minimizes J is the same as the estimate x'.

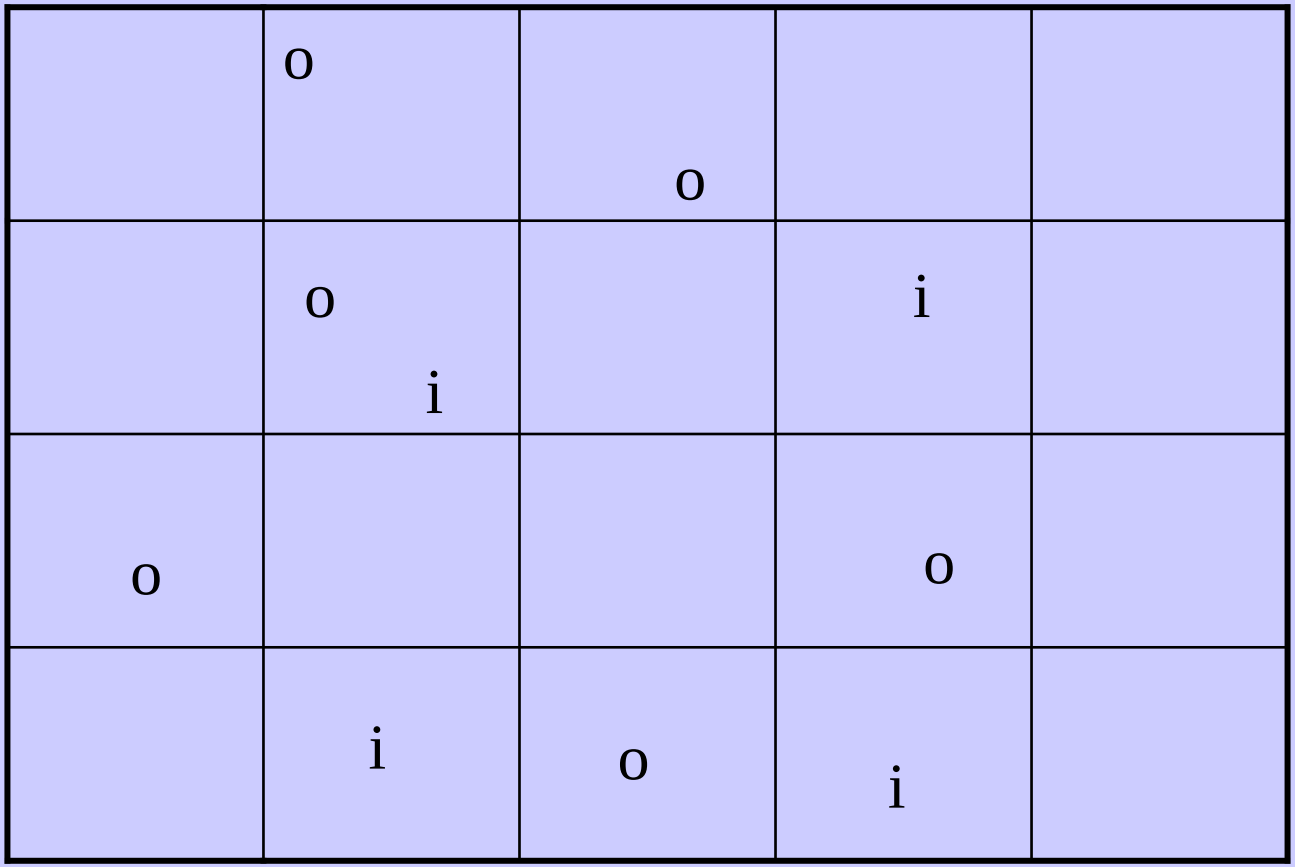
Optimum Interpolation

- A generalization of the previous problem in that that the observations may or may not coincide with points where estimates are desired.
- Interpret Obs. broadly to also include model predictions, so we have:
  - An a priori estimate on a grid provided by model
  - measurements

For now we do not distinguish between the two kinds of observations.

Optimum Interpolation

$$y = Hx + \varepsilon$$



X- true state on grid; y - observations; ε - error in going from x to y

Optimum Interpolation Formulation

$y = Hx + \varepsilon$ ← Error in going from true state to

↑

obs. Interpolation operator True state

y and x are vectors; H is an $(n \times m)$ matrix.

Assumptions about the errors:

$E(\varepsilon) = 0$ (no long term drift);

$E(\varepsilon\varepsilon^T) = \Sigma$ (errors can be correlated and corr

is known)

OI (continued, ignore details)

A linear, unbiased estimate of the true state :

$$x' = Ay \quad \text{subject to} : E(x' - x) = 0 \quad \text{(unbiased)}$$

$\uparrow$

weights

Above implies : $AH = I_m$ (identity matrix)

Finally, minimize the diagonal elements of the analysis

error covariance : $P = E[(x' - x)(x' - x)^T]$

The diagonal elements are the error variances of the estimates :

$$\sigma^2 = E[(x' - x)^2]$$

Off - diagonal elements relate the errors made at one point in one variable to errors made at another point to another variable.

OI (continued)

- The solutions to A & P involve only Σ, the error covariance matrix (non-trivial), and the H matrix. *This is the basis of OI*.

Rewrite the Equivalent Problem:

Find an estimate ξ of x that minimizes:

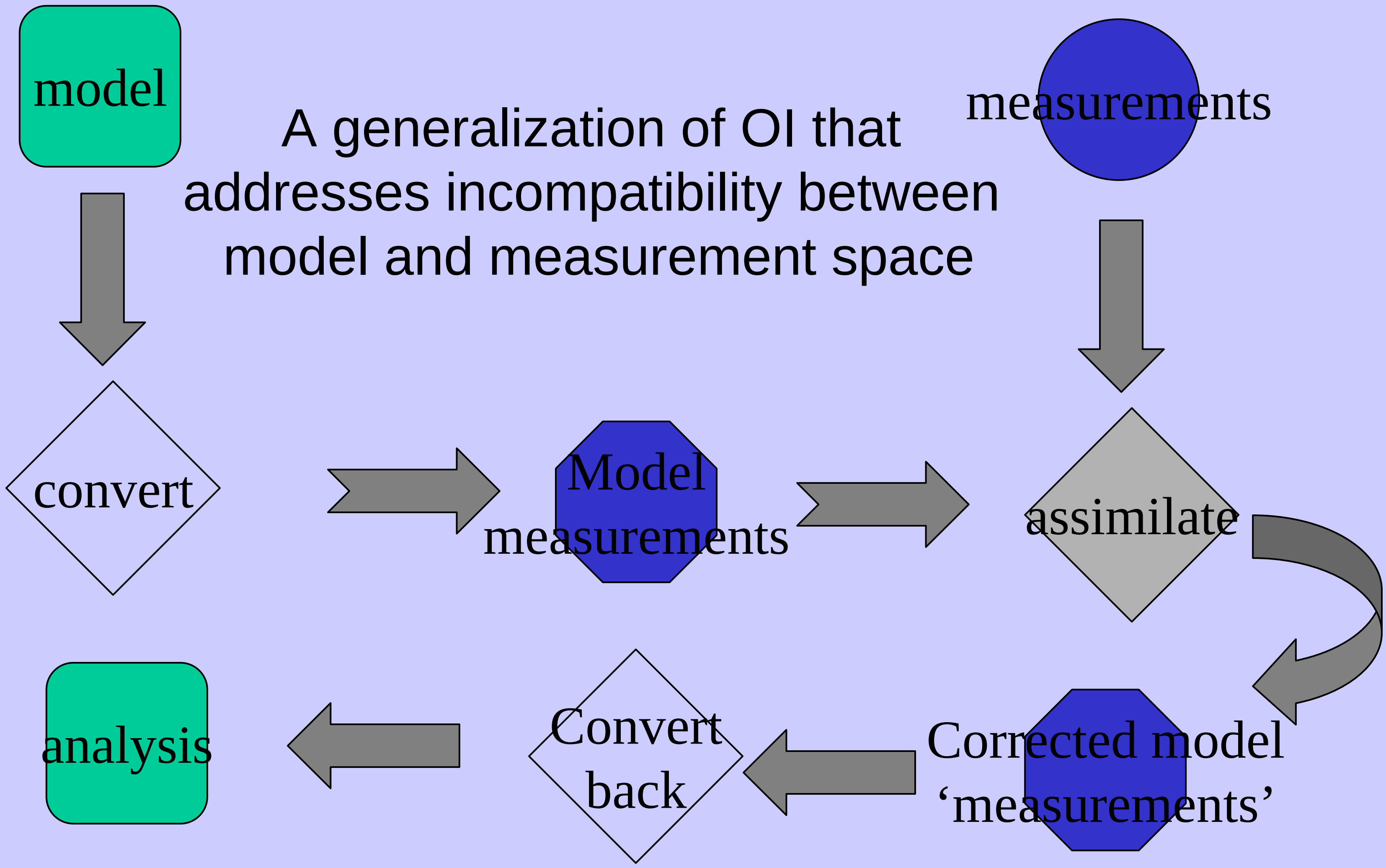
$$J(\xi) = (H\xi - y)^T \Sigma^{-1} (H\xi - y) \quad [\text{recall: } y = Hx + \varepsilon]$$

The ξ that minimizes $J(\xi)$ is the same as x' that solve

the least square minimization problem discuss

Variational Data Assimilation Methods

A generalization of OI that addresses incompatibility between model and measurement space



Variational Assimilation

Consider the cost function: $J(\xi) = (H\xi - y)^T \Sigma^{-1} (H\xi - y)$

Partition observations (separate 'model' from 'obs.' measurements

$y = \begin{Bmatrix} x_y \\ z \end{Bmatrix}$ x_y is a grid point (model) estimate; z is a measurement

$x_y = x + \varepsilon_f$ ε_f is a forecast error

$z = Kx + \varepsilon_o$ ε_o is the error in going from true state to observa

Then we can also partition: $H = \begin{pmatrix} I_m & 0 \\ 0 & K \end{pmatrix}$ as well as Σ, the error covarianc

matrix that has all the information we know about model and

Variational Analysis Summary

- A vector of observations is related to the true state through a function (possibly non-linear) that includes:
  - Interpolation of variables to observation points
  - Conversion of state to observed variables
- Observations differ from true value due to:
  - Measurement and background (model) errors
  - Errors of representativeness (sub-grid variability)
- An analyzed state close to observations and model is found by minimizing a cost function that separates the different errors and includes their covariances.

Summary of OI

- A widely used statistical approach
- A good framework for multivariate analysis that allows incorporation of physical constraints
- Weights determined as function of distribution and accuracy of data and models *(RS discussion)*

- Can be computationally demanding
- Scale-dependent correlation models require long history for accurate estimates of empirical coefficients *(RS discussion)*
- Underperforms in extreme events

Bayesian Methods

General characteristics

The Bayesian approach allows one to combine information from different sources to estimate unknown parameters.

Basic principles:

- Both data and external information (prior) are used.
- Computations are based on the Bayes theorem.
- Parameters are defined as random variables.

Bayesian Theory - historical perspective

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. Roy. Soc. London, 53, 370-418.

This 247 y.o. paper is the basis of the cutting edge methodology of data assimilation and forecasting (as well as other fields).

Basic Probability Notions

- **Basic probability notions needed for applying Bayesian methods:**
  i. **Joint probability.**
  ii. **Conditional probability.**
  iii. **Marginal probability.**
  iv. **Bayes theorem.**

Consider two random variables *A* and *B* representing two possible events.

Marginal (prior) probability

A and *B* are mutually exclussive events.

- Marginal probability of A = probability of event *A*
- Marginal probability of B = probability of event *B*
- Notation: P(A), P(B).

Joint probability

Joint probability = probability of event *A* and event *B*.

- Notation: P(AB) or P(A, B).

Conditional probability

Conditional probability = probability of event *B* given event *A*.

- Notation: P(B | A).

Bayes theorem

Bayes’ theorem allows one to relate P(B | A) to P(B) and P(A|B).

$$P(B \mid A) = P(A \mid B)\, P(B) / P(A)$$

This theorem can be used to calculate the probability of event B given event A.

In practice, A is an *observation* and B is an *unknown quantity of interest*.

How to use Bayes theorem?

"I figure there's a 40% chance of showers, and a 10% chance we know what we're talking about."

Guyatt, G. H et al. BMJ 2008;336:924-926

Bayes Theorem Example

A planned IO & SCS workshop at the SCSIO Nov. 17 -19. Climatology is of six days of rain in Guanzhou in November. Long term model forecast is of no rain for the 17th. When it actually is sunny, it is correctly forecasted 85% of the time.

Probability that it will rain?

A = weather in Day Bay on Nov. 30 (A1: «it rains », A2: « it is sunny»).

B = The model predicts sunny weather

Bayes Theorem Example

In terms of probabilities, we know:

$P(A_1) = 6/30 =0.2$ [It rains 6 days in November.]

$P(A_2) = 24/30 = 0.8$ [It does not rain 24 days in November.]

$P(B \mid A_1) = 0.15$ [When it rains, model predicts sun 15% of the time.]

$P(B \mid A_2) = 0.85$ [When it doesn’t rain, the model correctly predicts sun 85% of the time.]

We want to know $P(A_1 \mid B)$, the probability it will rain, given a sunny forecast.

$$P(A_1 \mid B) = P(A_1)\, P(B \mid A_1) / [P(A_1)\, P(B \mid A_1) + P(A_2)\, P(B \mid A_2)]$$

$$P(A_1 \mid B) = (0.2)\,(0.15)/[(0.2)\,(0.15) + (0.8)\,(0.85)]$$

$$P(A_1 \mid B) = 0.042$$

Forecasting Exercise

- Produce a probability forecast for Monday
- Propose a metric to evaluate your forecast

Possible sources for climatological and model forecasts for Taiwan:

http://www.cwb.gov.tw/V6e/statistics/monthlyMean/

http://www.cwb.gov.tw/V6e/index.htm

http://www.cma.gov.cn/english/climate.php

Evaluating (scoring) Probability forcast

Compare the forecast probability of an event p_i to the observed occurrence o_i, which has a value of 1 if the event occurred and 0 if it did not occur.

$$BS = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - o_i\right)^2$$

$$BSS = 1 - \frac{BS}{BS_{reference}}$$

The BS measures the mean squared probability error over N events (RPS can be used for a multi-category probabilistic forecast.

Bayesian data assimilation

What is the multidimensional probability of a particular state given a numerical forecast (first guess) and a set of observations

θ: vector of model parameters.

***y*: vector of observations**

P(θ): prior distribution of the parameter values.

P($y|\theta$): likelihood function.

P(θ | y): posterior distribution of the parameter values.

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_{\theta} P(y|\theta)P(\theta)}$$

often difficult to compute

Hybrid Methods and Recent Trends

- Kalman Filter: a generalization of OI where the error statistics evolve through (i) a linear prediction model, and (ii) the observations.
- Spectral Statistical Interpolation (SSI): an extension of OI to 3-D in spectral (global) space.
- Adjoint method (4DVAR): Extension of Bayesian approach to 4D. Given a succession of observations over time, find a succession of model states
- Nudging: A model is “nudged” towards observations by including a time dependent term.

Common Problems in Data Assimilation

- The optimal estimate is not a realizable state
  - Will produce noise if used to initialize
- The observed variables do not match the model variables
- The distribution of observations is highly non-uniform
  - Engineering and science specs clash
  - Over and under sampling due to orbit
- The observations are subject to errors
  - How closely should one fit the observations
  - Establishing errors and error covariances is not trivial…
- Very large dimensionality of problem
- Nonlinear transformations and functions

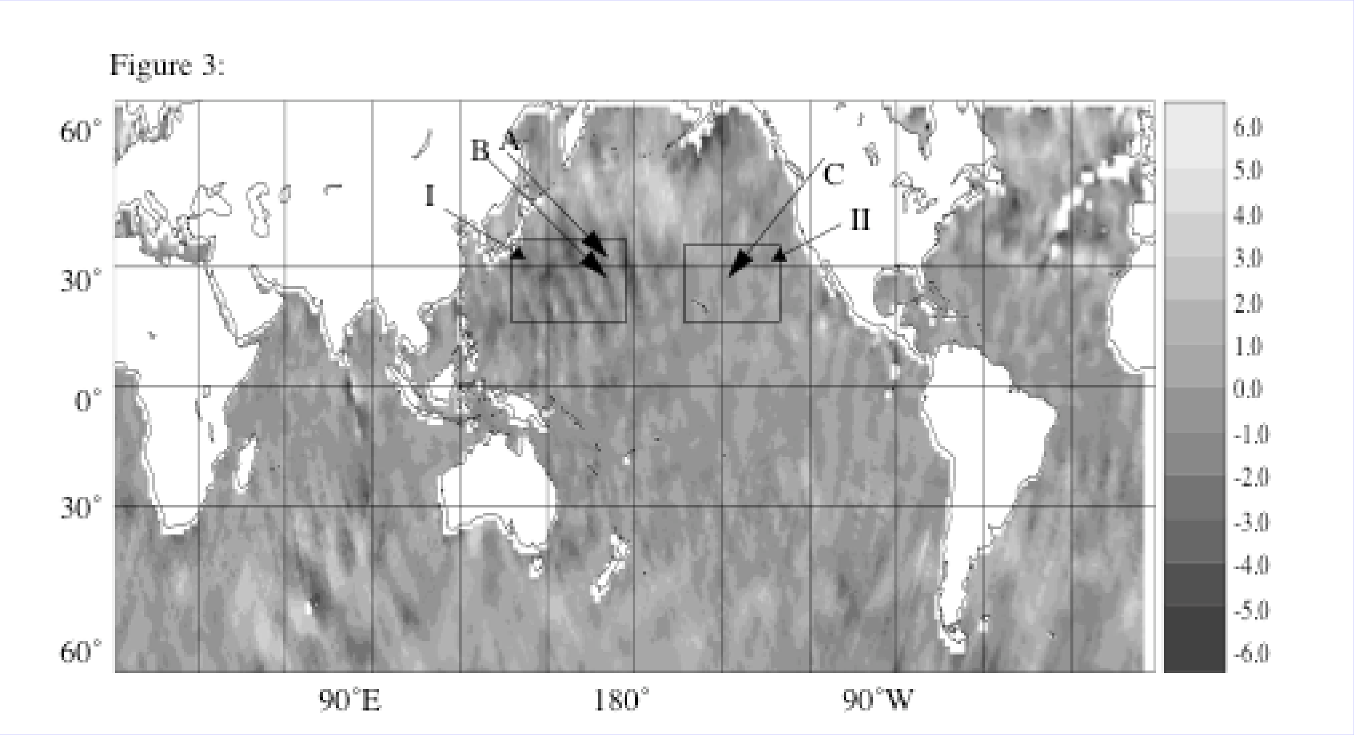


Figure 3:
60°
30°
0°
30°
60°
90°E
180°
90°W
A
B
I
C
II
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0

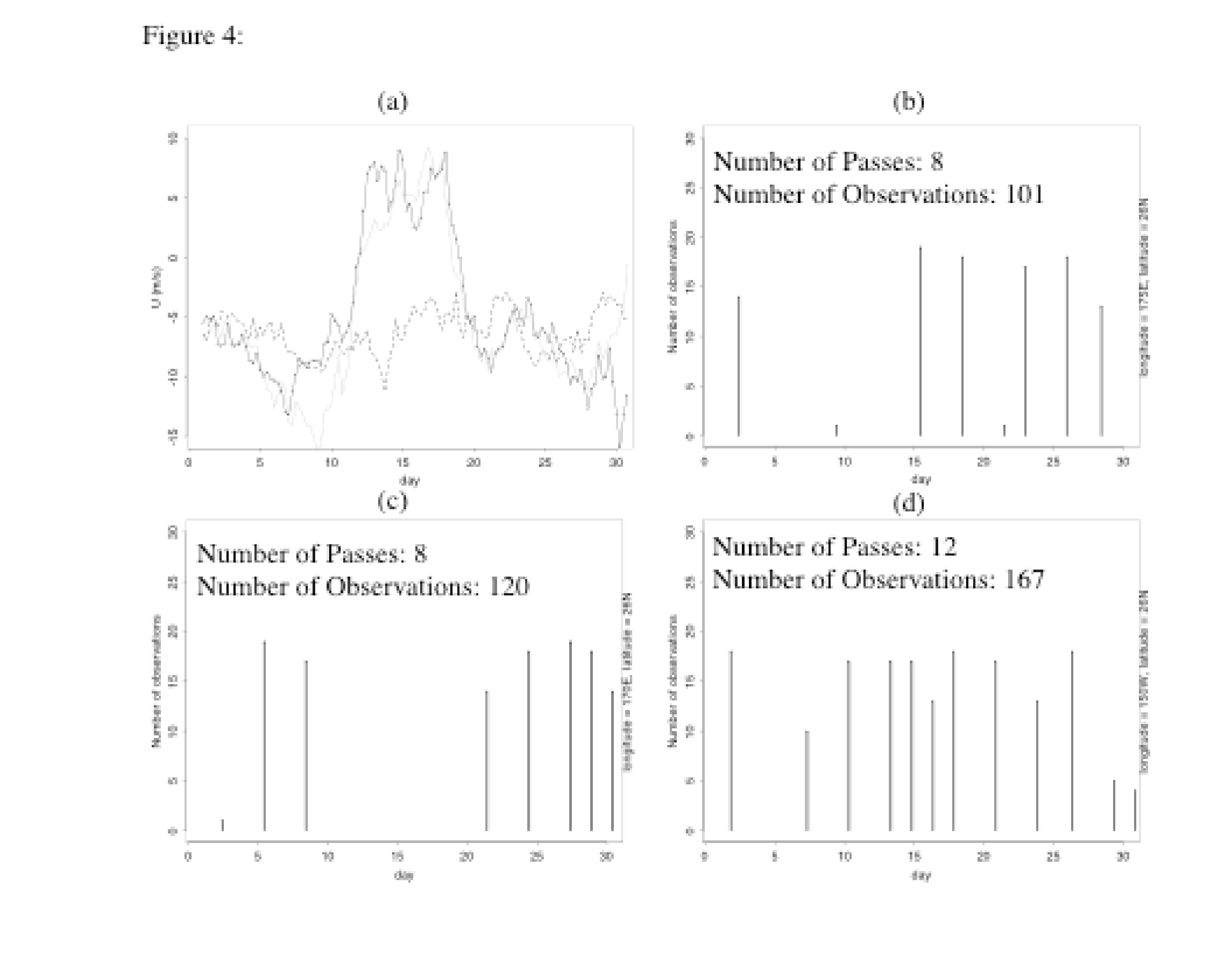


Figure 4:
(a)
(b)
Number of Passes: 8
Number of Observations: 101
(c)
Number of Passes: 8
Number of Observations: 120
(d)
Number of Passes: 12
Number of Observations: 167
day

References (textbooks)

- Atmospheric Data Analysis by Roger Daley. Published by Cambridge University Press, 1992. ISBN: 0521458250, 472 pp.
- Atmospheric Modeling, Data Assimilation, and Predictability by Eugenia Kalnay. Published by Cambridge University Press, 2003. ISBN 0521796296, 9780521796293. 341 pages

謝謝