The Abdus Salam
International Centre
for Theoretical Physics

The International Union of Geodesy and Geophysics

2373-5

Workshop on Geophysical Data Analysis and Assimilation

29 October - 3 November, 2012

Data Assimilation

Rodolfo Guzzi
Italian Space Agency
Italy

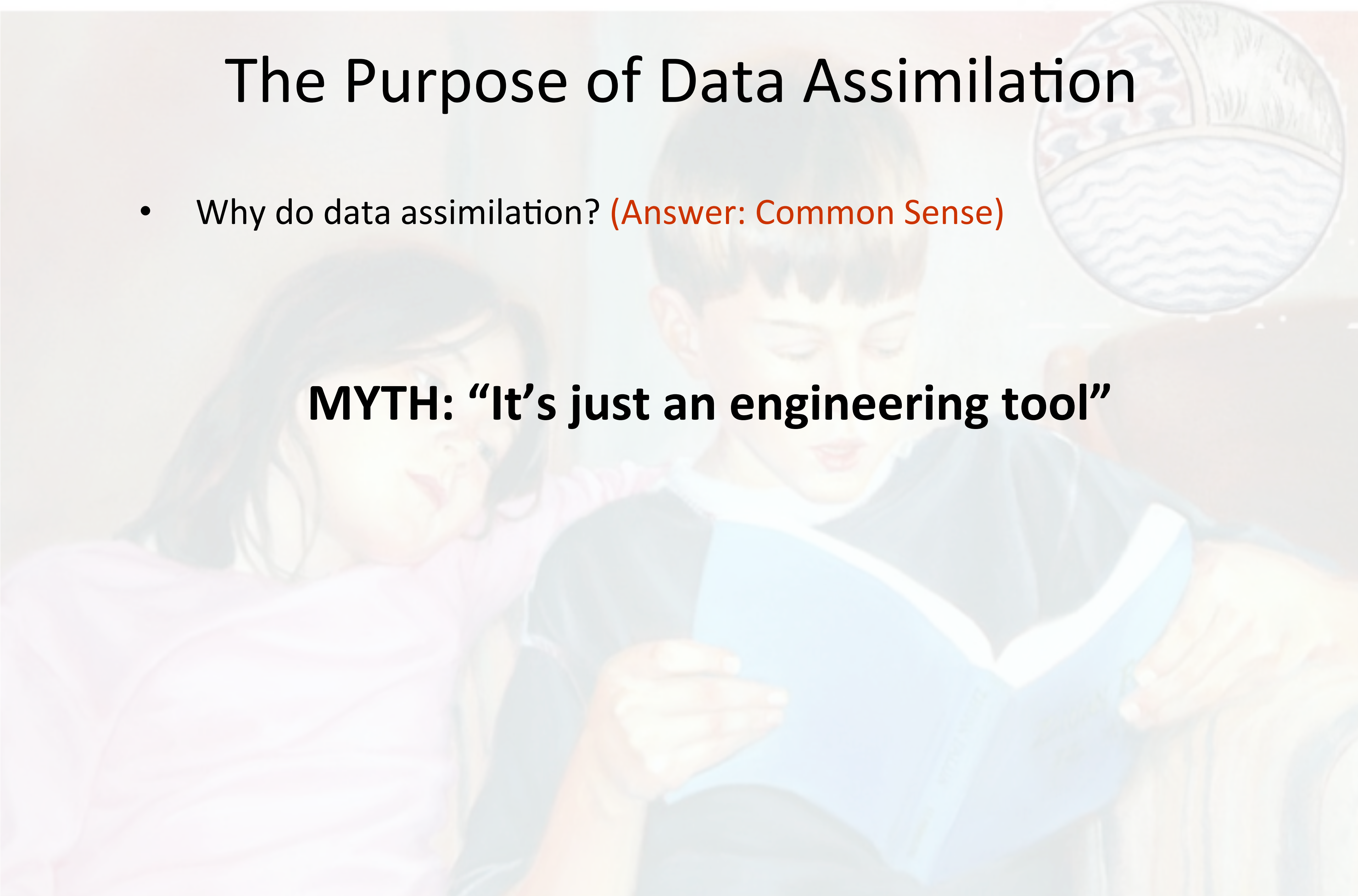
Data Assimilation

Rodolfo Guzzi
Italian Space Agency

Data Assimilation

Outline

- Why Do Data Assimilation?
- Who and What
- Important Concepts
- Definitions
- History by examples with the Gaussian approach
- Data Assimilation: the deep meaning
- Data Assimilation: technical approach
- Data Assimilation as a Inverse Problem
- Forecast vs Nowcast
- Optimal Least Square methods
- Variational methods and 3D-Var
- 4D-Var
- Conclusions
- Back up slides



The Purpose of Data Assimilation

- Why do data assimilation? (Answer: Common Sense)

MYTH: “It’s just an engineering tool”

The Purpose of Data Assimilation

- Why do data assimilation? (Answer: Common Sense)

~~**MYTH: “It’s just an engineering tool”**~~

If Truth matters,

“It’s our most important science tool”

The Purpose of Data Assimilation

- Why do data assimilation?
  1. I want better model initial conditions for better model forecasts
  2. I want better calibration and validation (cal/val)
  3. I want better acquisition guidance
  4. I want better scientific understanding of
     - Model errors (and their probability distributions)
     - Data errors (and their probability distributions)
     - Combined Model/Data correlations
     - DA methodologies (minimization, computational optimizations, representation methods, various method approximations)
     - Physical process interactions

The Purpose of Data Assimilation

- Why do data assimilation?
  1. I want better model initial conditions for better model forecasts
  2. I want better calibration and validation (cal/val)
  3. I want better acquisition guidance
  4. I want better scientific understanding of
     - Model errors (and their probability distributions)
     - Data errors (and their probability distributions)
     - Combined Model/Data correlations
     - DA methodologies (minimization, computational optimizations, representation methods, various method approximations)
     - **Physical process interactions (i.e., sensitivities and feedbacks)**
       Leads toward better future models

The Purpose of Data Assimilation

- Why do data assimilation?
  1. I want better model initial conditions for better model forecasts
  2. I want better calibration and validation (cal/val)
  3. I want better acquisition guidance
  4. I want better scientific understanding of
     - Model errors (and their probability distributions)
     - Data errors (and their probability distributions)
     - Combined Model/Data correlations
     - DA methodologies (minimization, computational optimizations, representation methods, various method approximations)
     - Physical process interactions (i.e., sensitivities and feedbacks)
       Leads toward better future models

VIRTUOUS CYCLE

The Data Assimilation Community

- **Who** is involved in data assimilation?
  - NWP Data Assimilation Experts
  - NWP Modelers
  - Application and Observation Specialists
  - Cloud Physicists / PBL Experts / NWP Parameterization Specialists
  - Physical Scientists (Physical Algorithm Specialists)
  - Radiative Transfer Specialists
  - Applied Mathematicians / Control Theory Experts
  - Computer Scientists
  - Science Program Management (NWP and Science Disciplines)
  - Forecasters
  - Users and Customers

The Data Assimilation Community

- **What** skills are needed by each involved group?
  - NWP Data Assimilation Experts **(DA system methodology)**
  - NWP Modelers **(Model + Physics + DA system)**
  - Application and Observation Specialists **(Instrument capabilities)**
  - Physical Scientists **(Instrument + Physics + DA system)**
  - Radiative Transfer Specialists **(Instrument config. specifications)**
  - Applied Mathematicians **(Control theory methodology)**
  - Computer Scientists **(DA system + OPS time requirements)**
  - Science Program Management **(Everything + $$ + Good People)**
  - Forecasters **(Everything + OPS time reqs. + Easy/fast access)**
  - Users and Customers **(Could be a wide variety of responses)**
    e.g., NWS / Army / USAF / Navy / NASA / NSF / DOE / ECMWF

The Data Assimilation Community

- **Are you part of this community?**
  - **Yes,** you just may not know it yet.
- **Who knows *all* about data assimilation?**
  - **No one knows it all, it takes *many* experts**
- **How large are these systems?**
  - Typically, the DA systems are “medium”-sized projects using software industry standards
    - Medium = multi-year coding effort by several individuals (e.g., RAMDAS is ~230K lines of code, ~3500 pages of code)
    - Satellite “processing systems” tend to be larger still
  - Our CIRA Mesoscale 4DVAR system was built over ~7-8 years with heritage from the ETA 4DVAR system

DA History by examples

- Mathematician and astronomers of seventeenth and eighteen centuries who made use of Newton laws to calculate the orbit of comets were the first assimilators
- Newton was among them and discussed the problem in *Principia* (Book III, Prop. XLI): “This being a problem of very difficulty I tried many methods of resolving it”.

DA History (continue)

- The task of finding the path of comets relied on the coupled set of non linear differential equation that described its paths under the assuntion of two body celestial mechanics. The motion was controlled by the gravitational attraction of the comet to the Sun
- Looking a celestial object in heavens one can express the position from its angular measurements by azimuth and elevation. We are ignorant of its distance from us and then we are unable to estimate its velocity from successive observations.
- It was clear that the observation of celestial bodies that were available to us could not easily translate into “standard” initial conditions, velocity and position.

DA History (continue)

- In order to take into account velocity it became necessary to obtain additional observation at other times.
- In the interval of time before the second observation is made, the Earth will have moved and the observed body will have gone to another place in its orbit. The second oservation simply determines another line on which the body is located at another date (that is the “difficulty”)

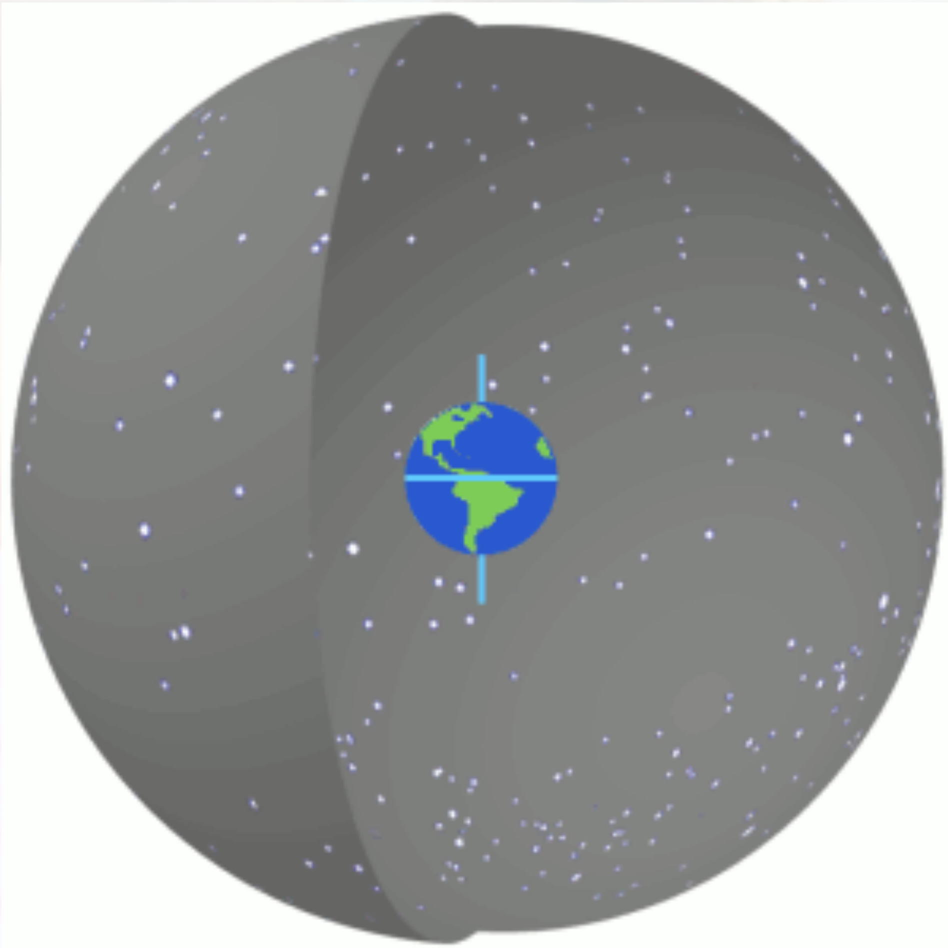
DA History

- At each epoch one could obtain two observation angles right ascension and declination.
- The six constants arising from the governing differential equations should be determined from three complete observations (i.e. two angular measurements at each of the three instants in time)

DA History

- Supposing we made three observations at the time t1, t2, t3 and the angular measurement be right ascension and declination.
- **Right ascension** ($\boldsymbol{\alpha}$) is the astronomical term for one of the two direction coordinates of a point on the celestial spheres in the equatorial coordinates system, usually combined with **Declination** $\boldsymbol{\delta}$. Right ascension's angular distance is measured eastward along the celestial equator celestial equator from the vernal equinox to the hour circle of the point in question.

DA History



DA History

$$
\begin{cases}
\alpha_1 = \psi(\Omega, i, \omega, a, e, T; t_1) \\
\alpha_2 = \psi(\Omega, i, \omega, a, e, T; t_2) \\
\alpha_3 = \psi(\Omega, i, \omega, a, e, T; t_3) \\
\delta_1 = \phi(\Omega, i, \omega, a, e, T; t_1) \\
\delta_2 = \phi(\Omega, i, \omega, a, e, T; t_2) \\
\delta_3 = \phi(\Omega, i, \omega, a, e, T; t_3)
\end{cases}
$$

Φ and Ψ are highly trascendental functions and involve the elements in a very complicated fashion that prevents to have direct solution of equation by ordinary processes

There are six elements, which are independent functions of these constants. They are:

1. The position of the plane of the orbit defined by:
Ω= longitude of ascending node, and
i = inclination to plane of the ecliptic

2. a = major semi-axis, which defines the size of the orbit and the period of revolution.
3. e = the eccentricity, which defines the shape of the orbit.
4. The orientation of the orbit in its plane defined by:
ω= longitude of the perihelion point measured from the node,
or π = longitude of the perihelion. $\pi=\Omega+\omega$
5. T = time of perihelion passage, defining, with the other elements, the position of the body in its orbit at any time.

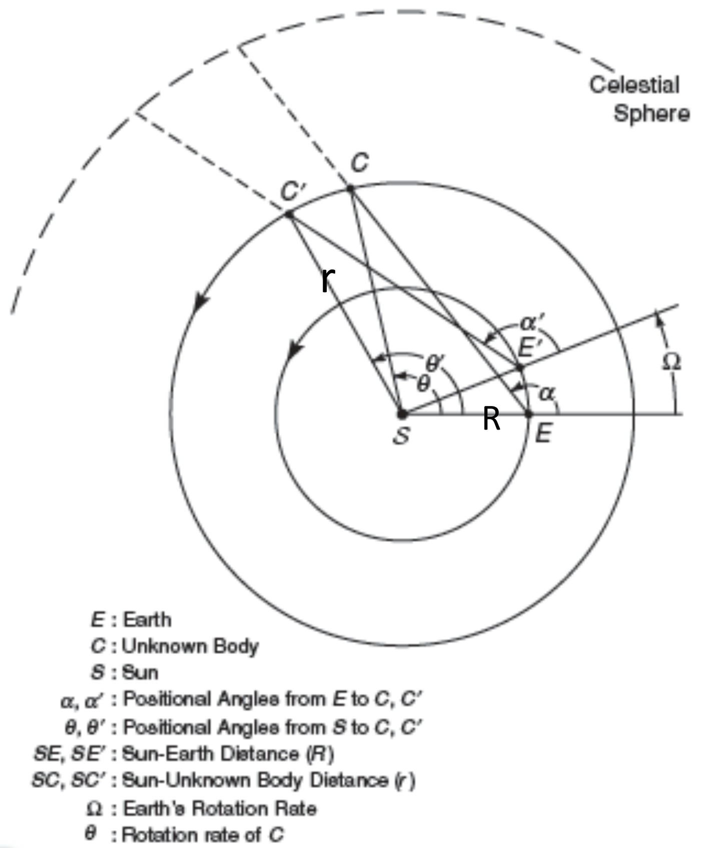
DA History: Gaussian approach

- Semplified assumptions that allow us to solve the problem meaningfully yet more simply than the problem discussed by Gauss in *Theoria*
- 1° step Solve the problem with the minimum set of observations: reduction of unknowns
- 2° step outline the method of solution in presence of more than the minimum requisite set

DA History: Gaussian approach

1.We assume a co-planar circular orbit of Earth E, Sun S, Planet C

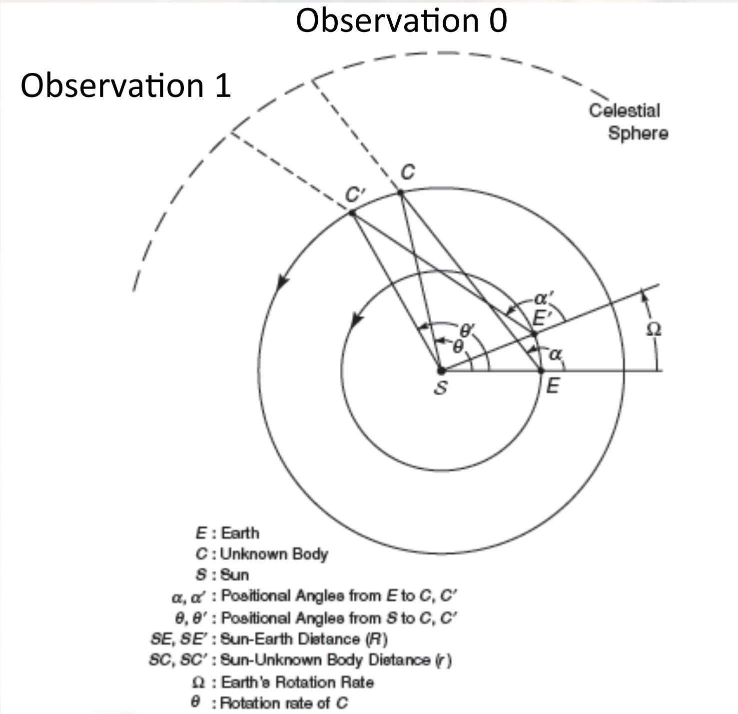
2. The observation are made by E



Measurement symbols associated with tracking of a planetoid (C) from earth (E).

DA History: Gaussian approach

Derive a formula for finding r and $\dot{\theta}$
radius and rotation rate of C
from the measurements of α



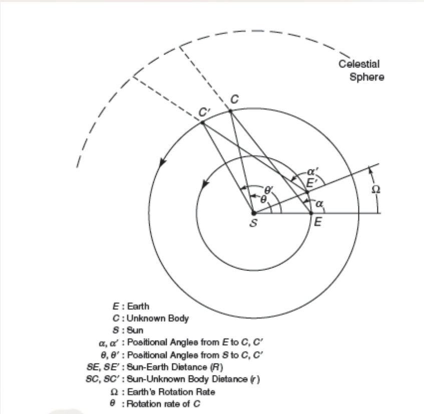
Measurement symbols associated with tracking of a planetoid (C) from earth (E).

$$\alpha(t) = \tan^{-1}\frac{r\sin(\theta+\dot{\theta}t) - R\sin\Omega t}{r\cos(\theta+\dot{\theta}t) - R\cos\Omega t}$$

$$\begin{aligned} x(t) &= -R\cos\Omega t + r\cos(\theta+\dot{\theta}t) \\ x(t) &= -R\cos\Omega t + r\cos(\theta+\dot{\theta}t) \\ \tan\alpha &= \frac{y}{x} \end{aligned}$$

Taking the derivative of α with respect to time

$$\dot{\alpha} = \frac{\dot{y} - \dot{x}\tan\alpha}{x\sec^2\alpha} \tag{1}$$



Measurement symbols associated with tracking of a planetoid (C) from earth (E).

Now at $t = 0$ (where θ is the unknown initial angle) we have

$$\tan\alpha = \frac{r\sin\theta}{r\cos\theta - R} \tag{2}$$

Evaluating $\dot{\alpha}$ at $t = 0$ we get

$$\dot{\alpha}\Big|_{t=0} = \left[\frac{\dot{y}(0) - \dot{x}(0)\tan\alpha}{x(0)\sec^2\alpha}\right]_{t=0} \tag{3}$$

From the Kepler's 3rd law we also know

$$\dot{\theta}^2 = \frac{4\pi^2}{r^3} \tag{4}$$

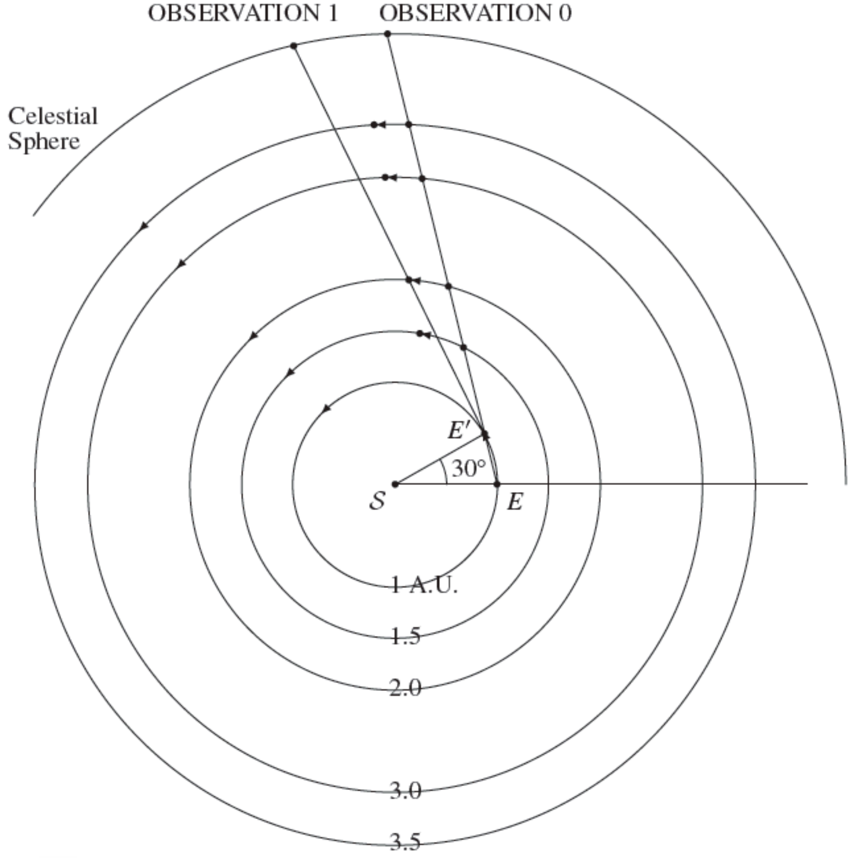
where

$$\begin{aligned} x(0) &= -R + r\cos\theta & (5)\\ y(0) &= r\sin\theta & (6)\\ \dot{x}(0) &= -r\dot{\theta}\sin\theta & (7)\\ \dot{y}(0) &= r\dot{\theta}\cos\theta - R\Omega & (8) \end{aligned}$$

They are not easily solved
We need an iterative method

Remembering that at $\alpha, \dot{\alpha}\,|_{t=0}$ R and Ω are kwown
equations 2,3,4 are three equations in terms of the three unknowns r, θ and $\dot{\theta}$

DA History: Gaussian approach



Rotation rates determined from Kepler's 3rd Law. An object at unknown distance from earth is observed at two times – "0" and "1" (earth at E and E', respectively). Use of Kepler's law yields its distance from the Sun (2.0 A.U.).

1. We start our iteration by "guessing" at r, a valu $\hat{r} = 3$
2. We measured α so from eq. 2 will give us a guess at θ, $\hat{\theta}$
3. If we substitute from eq 4 into eq. 3 eliminating $d\theta/dt$ then eq. 3 contains θ and r.
4. We have measured $d\alpha/dt|_{t=0}$ and we can solve for an improvedvalue of r= $\hat{r} + p$ linearizing eq. 3 as a function only of $p, \hat{\theta}$ and then solve for p.
5. We then return to eq 2 to get the new estimate of θ using $\hat{r} + p$ in place r.
6. Continue to iterate until p become vanishing small (Newton Raphson method)

DA History: Gaussian approach

- Assume now the observations of angle $\alpha(t)$ are subject to error. How would be accomodate more than two oservations?
- For example assume we have three measurements denote by:

$$\tilde{\alpha}_0(t), \tilde{\alpha}_1(t), \tilde{\alpha}_2(t)$$

- We furthermore assume we know the time that each observation is made t_0, t_1, t_2
- Gauss assumed to build up a measure of the fit of the model (Kepler's law) to the observations by using the least squares approach.

$$J = (\alpha_0 - \tilde{\alpha}_0)^2 + (\alpha_1 - \tilde{\alpha}_1)^2 + (\alpha_2 - \tilde{\alpha}_2)^2$$

DA History: Gaussian Approach

- In essence we find the set $r, \dot{\theta}, \theta$ the gives us the α's that minimize J, i.e. the least squares fit between the model derived state and the observations
- We can start assuming that the initial guesses are:

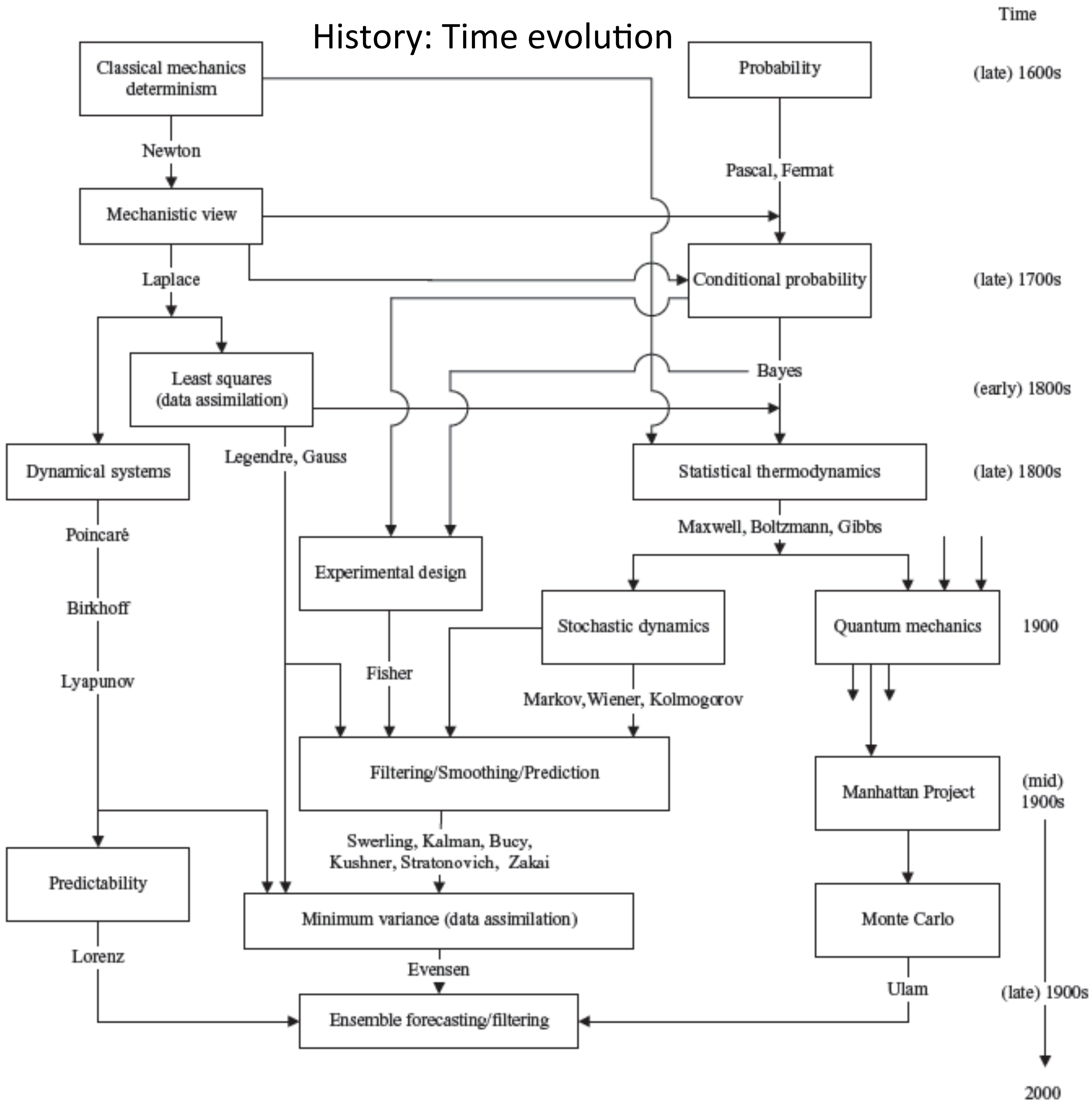
$$\alpha_0^{(1)} = \tilde{\alpha}_0$$
$$r^{(1)} = 3$$

and $\theta_0^{(1)}$ the solution to eq.1 where α and r are given by the guesses.

DA History: Gaussian approach

- Since the coordinates x and y are known as function of $r, \dot{\theta}, \theta$ we have a forecast equation to get values of the α's at the various times downstream.
- Is our initial guess a minimum?Generally not but if we know the fisrt derivatives at our operating point (the initial guess) we can open to find an improved estimate by moving along the direction of the negative gradient of J

History: Time evolution



DA Meaning

- What do we mean by data assimilation?
- Assimilation is an analysis in which the information is accumulated in the state of a dynamic model, exploiting the consistency of constraints inherent in the physics laws and time processes, combining the observations distributed in time with the dynamic model itself.

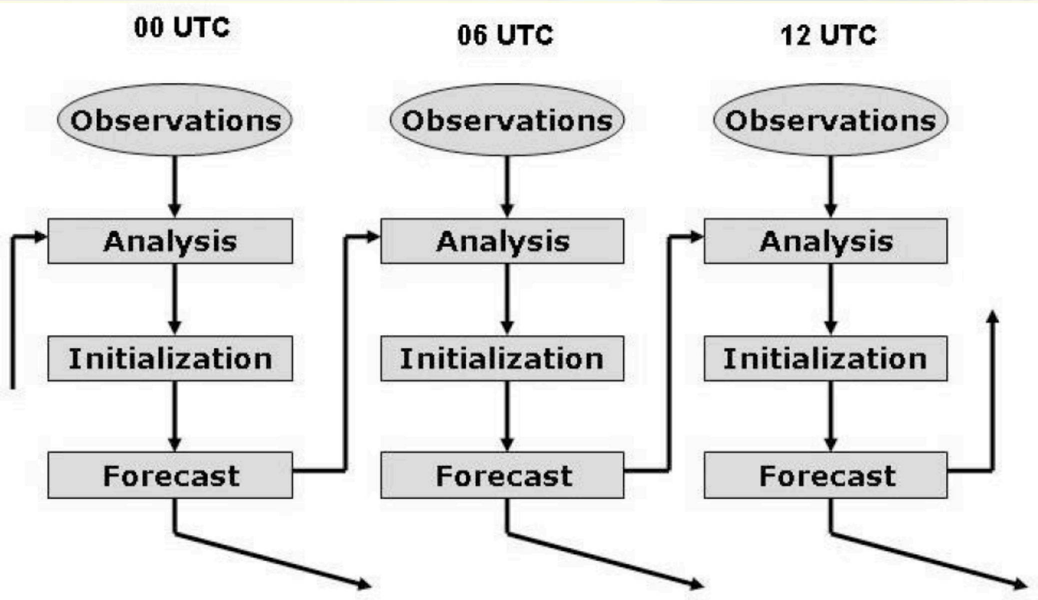
DA meaning

- The process analysis corresponds to various degrees:
  (1) as an approximation of the true state of a physics system at a given time;
  (2) as including diagnosis and self consisting of a physics system;
  (3) as a reference by which to make a test of the quality of the observation
  (4) as an input data useful for another operation, such as the initial state of a predictive model.

DA Meaning

- The more usual case is to use the assimilation to make a time prediction.
- Such approach implies that errors due to initial conditions must be reduced as far is possible only leaving to the model the possibility to generate the errors and then to proceed in a more realistic direction.
- The assimilation combines the observation data with the data produced by the model to reproduce an "Optimal" estimate of the evolving state of the system.
- The model provides consistency to the observed data allowing also to interpolate or extrapolate data into regions of space and time in which these are lacking.
- Furthermore the observed data adjust the trajectory of a model through the state space of the model, keeping in line in a loop prediction-osservation-correction.

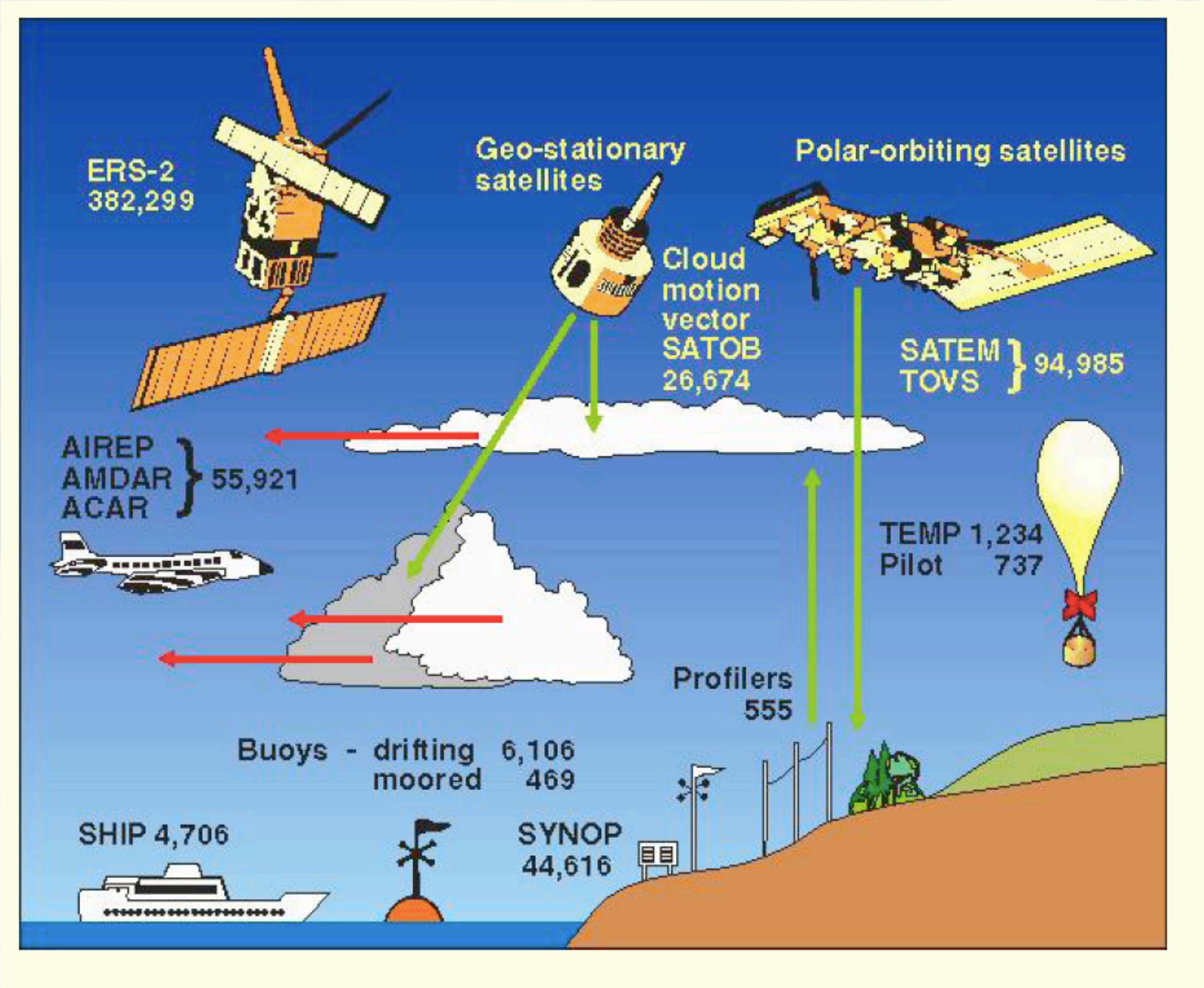
DA Approach



Typical 6-hour analysis cycle.

- Bayes interpretation: a forecast (the "prior"), is combined with the new observations, to create the Analysis (IC) (the "posterior")

DA Approach


ERS-2
382,299
Geo-stationary
satellites
Polar-orbiting satellites
Cloud
motion
vector
SATOB
26,674
SATEM
TOVS
94,985
AIREP
AMDAR
ACAR
55,921
TEMP 1,234
Pilot 737
Profilers
555
Buoys - drifting 6,106
moored 469
SHIP 4,706
SYNOP
44,616

DA toy model

- We want to measure the temperature in this room, and we have two thermometers that measure with errors:

$$T_1 = T_t + \epsilon_1$$

$$T_2 = T_t + \epsilon_2$$

- We assume that the errors are unbiased:

$$\overline{\epsilon_1} = \overline{\epsilon_2} = 0$$

that we know their variances $\overline{\epsilon_1^2} = \sigma_1 \quad \overline{\epsilon_2^2} = \sigma_2$

- and the errors of the two thermometers are uncorrelated: $\overline{\epsilon_1 \epsilon_2} = 0$

- **The question is: how can we estimate the true temperature optimally? We call this optimal estimate the "analysis of the temperature"**

DA: toy model

- We try to estimate the analysis from a linear combination of the observations:

$$T_a = a_1 T_1 + a_2 T_2$$

and assume that the analysis errors are unbiased:

$$\bar{T}_a = \bar{T}_t$$

- This implies that

$$a_1 + a_2 = 1$$

- T_a will be the *best estimate if the coefficients* a_1, a_2 are chosen to minimize the mean squared error of T_a:

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

DA toy model

- the minimization σ_a^2 of with respect a_1 to gives

$$\frac{\partial \sigma_a^2}{\partial a_1} = 0 \longrightarrow a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

- Or

$$a_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} \quad a_2 = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

- The first formula says that the weight of obs 1 is given by the variance of obs 2 divided by the total error.
- The second formula says that the weights of the observations are proportional to the "precision" or accuracy of the measurements(defined as the inverse of the variances of the observational errors).

DA as a inverse problem

- A forecast and an observation optimally combined (analysis):
- ***If the statistics of the errors are exact, and if the coefficients*** *are optimal, then the "precision" of the analysis (defined as the inverse of the variance) is the sum of the precisions of the measurements.*
- The importance of these toy examples is that the equations are identical to those obtained with big models and many observations.

DA as a inverse problem

- Let us now see why DA is an inverse problem
- A boundary value problem in mathematical physics is said to be well-posed in the senso of Hadamard if it satisfies the following three conditions

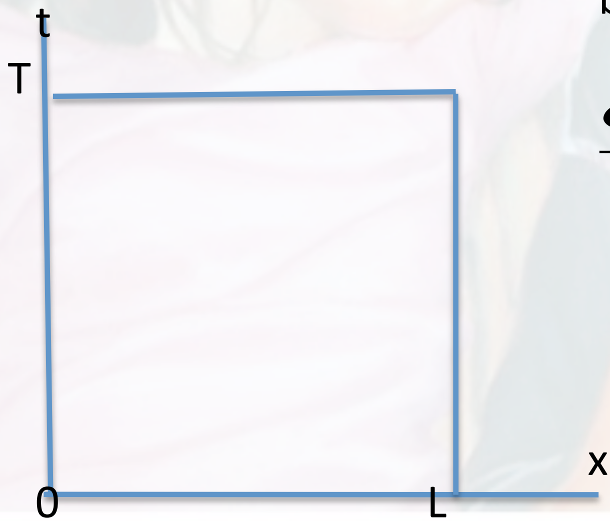
1. The solution exists
2. The solution is unique
3. The solution depends continuosly on the data

DA Inverse Problem: Ocean basin toy model

- Let us use a simple toy model of the ocean:
- Define an "ocean basin" given in the interval

$0 \leq x \leq L$ while the time of interest is

$0 \leq t \leq T$



The ocean dynamics are espressed by the differential equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = F$$

c is a known, constant, positive phase speed

F=F(x,t) is a specified forcing field

specified

An Initial Condition is $u(x,0) = I(x)$

A boundary Condition is $u(0,t) = B(t)$

Is the solution unique?

- In order to determine the uniqueness of solution let u1 and u2, two solutions for the same choiches of F, I and B. Define the difference $v \equiv u_1 - u_2$
- Then $$\frac{\partial v}{\partial t} + c\frac{\partial v}{\partial x} = 0$$
- with $v(x,0) = 0 \quad v(0,t) = 0$
- The solution is $v(x,t) = 0$ then

$$u_1(x,t) = u_2(x,t)$$

Yes it is unique

Does the solution exist?
We may construct the solution

- Using the Green's function $G = G(x, t, \zeta, \tau)$
- Our equation becomes:

$$-\frac{\partial G}{ds} - c\frac{\partial G}{dx} = \delta(x - \zeta)\delta(t - \tau)$$

- The inizialization and boundary conditions are:

$$G(L, t, \zeta, \tau) = 0 \qquad G(x, T, \zeta, \tau) = 0$$

- The solution is:

$$u(x,t) = \int_0^T d\tau \int_0^L d\zeta G(\zeta, \tau, x, t) F(\zeta, \tau) + \int_0^L d\zeta G(\zeta, 0, x, t) I(\zeta) + \int_0^T d\tau G(0, \tau, x, t) B(\tau)$$

The solution depends continuously on changes to I,F and B . We conclude eq. is WELL POSED

Inverse problem

- Let's now introduce additional information about the toy ocean circulation field u(x,t). This information will consist of imperfect observations of u at isolated point in space and time. Then the forward model become overdetermined and must be regarded as ill posed problem.
- Let's assume to collect M measurements defined by:

$$y_i = u(x_i, t_i) + \epsilon_i$$

- The equation of the ocean circulation becomes

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = F + f$$

- With the conditions

$$u(0,t) = B(t) + b(t)$$
$$u(x,0) = I(x) + i(x)$$

Limiting choice of weights: weak and strong constraints

Inverse problem

- We have established that for any choice of F+f, I+i and B+b there is a unique solution for u.
- We have only the M data values to guide us
- The error field f,i and b are undetermined while the errors on data are unknown
- We shall seek the minimum of the quadratic or *cost functional J.*

$$\mathcal{J} = \mathcal{J}[u] = W_f \int_0^T dt \int_0^L f(x,t)^2 dx + W_i \int_0^L i(x)^2 dx + W_b \int_0^T b(t)^2 dt + W_{ob} \sum_{i=1}^M \epsilon_i^2$$

- Rewriting explicitely the dependence on F,I,B and ϵ_i

$$\mathcal{J}(u) = W_f \int_0^T dt \int_0^L \{\frac{\partial u}{dt} + c\frac{\partial u}{\partial x} - F\}^2 dx + W_i \int_0^L \{u(x,0) - I(x)\}^2 dx + W_b \int_0^T \{u(0,t) - B(t)\}^2 dt + W_{ob} \sum_{m=1}^M \{u(x_i,t_i) - \mathbf{y}_i\}^2$$