Earth Observation and Data Assimilation
QUEST ES4 Spring School, Sept. 2006

Ross Bannister*

*Data Assimilation Research Centre (DARC),
NERC National Centre for Earth Observation,
Dept. of Meteorology,
Univ. of Reading,
Reading,
RG6 6BB

r.n.bannister@reading.ac.uk

Colour slides available (PDF) via www.met.rdg.ac.uk/~ross/DARC/DataAssim.html
Thanks to DARC colleagues: Stefano Migliorini, William Lahoz
1. INTRODUCTION

The data assimilation problem

State vector, $\tilde{x}$

A-priori information, $\tilde{x}_B$, and errors
- Background state
- Best guess
- Forecast

Assimilation algorithm ("inverse model")
- Optimal Interpolation (OI)
- Variational data assimilation (Var.)
- Kalman filter

Observations, $\tilde{y}$, and errors
- Sondes
- Surface stations
- Ships
- Satellites

Models ("forward models")
- Linking model state to observations
  \[ \tilde{y} = \tilde{h}[\tilde{x}] + \tilde{e} \]
Representation of data

The 'state vector', $\tilde{x}$

- Values of all variables and at all grid points are assembled in this vector.
- The system's state may be represented as a point in the model's $(5 \times n \times m \times L)$-dimensional phase space.

The 'observation vector', $\tilde{y}$

- Every measurement to be assimilated is assembled in this vector.
- The observation type, location and time needs to be associated with each observation.

*These vector structures allow them to be used in matrix equations (later).*
Data assimilation as an inexact and underconstrained inverse problem

- The 'inverse model' approach to data assimilation can deal with 'direct' (in-situ) and 'indirect' (remotely sensed) observations.
- The data assimilation problem is termed 'inexact' because all quantities have errors which must be accounted for.
- The data assimilation problem is termed 'under constrained' because the state vector is not fully observed.

All models are wrong! All observations are inaccurate!
Combining observational data: 1 unknown, 2 direct observations

Aim: to estimate the value of a scalar, \( x \), and its uncertainty.

Information to use: two unbiased direct measurements of \( x \) from different instruments.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
<th>Error*</th>
<th>Std. dev.†</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>'truth'</td>
<td>( x_t )</td>
<td>0</td>
<td>n/a</td>
<td>Abstract, as ( x_t ) can never be known precisely</td>
</tr>
<tr>
<td>obs. 1</td>
<td>( y_1 )</td>
<td>( \varepsilon_1 )</td>
<td>( \sigma_1 )</td>
<td>( \sigma_1 ) is the precision of inst. 1</td>
</tr>
<tr>
<td>obs. 2</td>
<td>( y_2 )</td>
<td>( \varepsilon_2 )</td>
<td>( \sigma_2 )</td>
<td>( \sigma_2 ) is the precision of inst. 2</td>
</tr>
<tr>
<td>best est. of 'truth'</td>
<td>( x_a )</td>
<td>( \varepsilon_a )</td>
<td>( \sigma_a )</td>
<td>( \sigma_a ) is a fn. of ( \sigma_1 ) and ( \sigma_2 ). ( a = \text{'analysis'} )</td>
</tr>
</tbody>
</table>

*Deviation from 'truth', \( y_n = x_t + \varepsilon_n \), \( n = 1,2 \). \( \varepsilon_n \) are not known, only their 'stats'†.

†Width of the probability density function (PDF), \( \sigma_n \equiv \langle (y_n - x_t)^2 \rangle^{1/2} = \langle \varepsilon_n^2 \rangle^{1/2} \).

Unbiased: means that repeated measurements are centred about the 'truth', \( \langle \varepsilon_n \rangle = 0 \), ie \( \langle y_n \rangle = x_t \).

\[
\begin{align*}
x_a &= \left( \frac{y_1}{\sigma_1^2} + \frac{y_2}{\sigma_2^2} \right) \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}, \\
\sigma_a &= \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1/2}
\end{align*}
\]

- This is simple data assimilation.
- The larger the '\( \sigma \)' of a measurement, the smaller its importance.
- Use (i) the 'method of least squares' and (ii) normal (a.k.a. Gaussian) PDFs (see later).
- Beware: the term 'error' is often used to indicate \( \sigma \). Should use the term 'error statistics'.
Combining observational data: 6 unknowns, >6 indirect observations - orbital determination

Aim: to estimate the six orbital parameters of Venus, $\mathbf{x}$, and their uncertainty.
Information to use: many indirect measurements.

$$\mathbf{\hat{x}} = \begin{pmatrix} a \\ e \\ i \\ \Omega \\ \omega \\ \varepsilon \end{pmatrix}, \quad \mathbf{\hat{y}} = \begin{pmatrix} \text{alt (1)} \\ \text{azi (1)} \\ \text{alt (2)} \\ \text{azi (2)} \\ \ldots \end{pmatrix}$$

$$\mathbf{\hat{y}} = \mathbf{h} [\mathbf{\hat{x}}] + \mathbf{\hat{e}}$$
\[ x = (0.7210, 0.0201, 4.23, 88.9, 110.0, 176.6) \]
\[ \sigma = (0.0020, 0.0078, 0.70, 8.1, 50.3, 6.8) \]
\[ x_t = (0.7233, 0.0067, 3.39, 76.7, 131.5, 182.0) \]

(a) Assimilation period

(b) Forecast period
Applications of data assimilation

- Keeping dynamical systems 'in touch' with reality.
- Initial conditions for weather or ocean forecasting.
- Reanalysis for scientific studies of climate (e.g. NCEP/NCAR, ERA).
- Inferring information that is difficult or impossible to measure directly, or using data from remote sensing instruments (e.g. satellite sounding, surface carbon flux estimation, solar dynamics).
- Model and observation system evaluation.
- Systems control (e.g. landing a rocket on the moon, shooting a moving target).
1. Introduction
2. Observations
3. Models
4. Data assimilation fundamentals
5. Applications of and problems with data assimilation
6. Further reading
2. OBSERVATIONS
Types of instrument

Measurements from instruments assimilated routinely (not exhaustive)

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Quantities measured</th>
<th>Coverage</th>
<th>Temporal</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-situ instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiosondes</td>
<td>$u, v, T, p, q, (O_3)$</td>
<td>Cont'l N.H., t'sphere</td>
<td>6 hourly</td>
<td>point</td>
</tr>
<tr>
<td>Surface stations</td>
<td>$u, v, T, p, q$</td>
<td>Cont'l, surface</td>
<td>6 hourly</td>
<td>point</td>
</tr>
<tr>
<td>Aircraft</td>
<td>$u, v, T, p, q$</td>
<td>Flight paths, airports</td>
<td>In flight</td>
<td>point</td>
</tr>
<tr>
<td>Drifting buoys</td>
<td>$u, v, T, p$</td>
<td>Drift paths, sea lev.</td>
<td>Hourly</td>
<td>n/a</td>
</tr>
<tr>
<td>Remote sensing instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geostationary sat.</td>
<td>Rad: MW, IR, Vis</td>
<td>Global</td>
<td>15-30 mins</td>
<td>&gt; 1 km</td>
</tr>
<tr>
<td>Polar orbiting sat. (nadir)</td>
<td>Rad: MW, IR, Vis</td>
<td>Global</td>
<td>Continuous</td>
<td>&gt; 1 km</td>
</tr>
<tr>
<td>Polar orbiting sat. (limb)</td>
<td>Rad: MW, IR, Vis</td>
<td>Global</td>
<td>Continuous</td>
<td>100s km</td>
</tr>
<tr>
<td>Scatterometer</td>
<td>Radar backscatter</td>
<td>Oceans</td>
<td>Continuous</td>
<td>1-2 km</td>
</tr>
</tbody>
</table>

'Rad'=radiances, 'MW'=microwave, 'IR'=infrared, 'Vis'=visible

In operational global weather forecasting there are $\sim 10^6$ observations assimilated per cycle.
Coverage

Locations of four example observation types (courtesy Met Office (c) Crown copyright)
Volumes of data and quality control

ECMWF stats. (one cycle in June '03)
Total No. obs.: ~ 70,000,000
Total No. assimilated: ~ 3,500,000 (only 5%!)

Why are some observations rejected?
- Observation 'too far' from forecast (large systematic, human, or instrument error),
- Observation did not reach centre in time,
- Satellite radiance data - complications due to radiation from land, clouds or precipitation.
Satellite borne instruments

Orbit configurations

**Polar orbiter** (courtesy WAL)

- Quasi-global coverage.
- Non-continuous sampling of a given location.
- Often used for sounders (e.g. on board EnviSat, EOS Aura, etc).

**Geostationary orbit** (courtesy NASDA)

- 35 786 km above sea level, latitude 0.0°.
- View 1/4 of Earth's surface (60S-60N).
- Continuous sampling of a given location.
- Often used for imagers (e.g. on board MeteoSat, etc).
- Horiz. resolution degrades poleward.
Satellite borne instruments

Viewing geometries

Limb (left) and nadir (right) viewing geometries

**Limb**
- Good vertical resolution possible (~1km).
- Poor horizontal resolution.
- Difficulties in constructing observation operator.
- Used mainly in research.

**Nadir**
- Good horizontal resolution possible.
- Poor vertical resolution (several km).
- Used mainly in operational weather forecasting.
## Satellite borne instruments

*(not comprehensive!)*

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Expanded name</th>
<th>Platform</th>
<th>Geometry</th>
<th>Orbit</th>
<th>Measures</th>
<th>Pass/Act</th>
<th>Sensitive to</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIRDLS</td>
<td>High Resolution Dynamics Limb Sounder</td>
<td>EOS Aura</td>
<td>Limb</td>
<td>Polar</td>
<td>?</td>
<td>Passive</td>
<td>$T, q, O_3, etc$</td>
</tr>
<tr>
<td>OMI</td>
<td>Ozone Monitoring Experiment</td>
<td>EOS Aura</td>
<td>Nadir</td>
<td>Polar</td>
<td>Vis/UV</td>
<td>Passive</td>
<td>$O_3, TCO_3, etc.$</td>
</tr>
<tr>
<td>TES</td>
<td>Tropospheric Emission Spectrometer</td>
<td>EOS Aura</td>
<td>Limb/Nadir</td>
<td>Polar</td>
<td>IR</td>
<td>Passive</td>
<td>$T, q, O_3, etc$</td>
</tr>
<tr>
<td>MLS</td>
<td>Microwave Limb Sounder</td>
<td>EOS Aura</td>
<td>Limb</td>
<td>Polar</td>
<td>MW</td>
<td>Passive</td>
<td>$T, q, O_3, etc$</td>
</tr>
<tr>
<td>SSM/I</td>
<td>Special Sensor Microwave Imager</td>
<td>DMSP</td>
<td>Nadir</td>
<td>Polar</td>
<td>MW</td>
<td>Passive</td>
<td>TCWV, cloud, precip, surface wind, snow, sea ice</td>
</tr>
<tr>
<td>HIRS</td>
<td>High resolution InfraRed Sounder</td>
<td>NOAA</td>
<td>Nadir</td>
<td>Polar</td>
<td>IR</td>
<td>Passive</td>
<td>$T, q, O_3, etc$</td>
</tr>
<tr>
<td>AMSU</td>
<td>Advanced Microwave Sounding Unit</td>
<td>NOAA</td>
<td>Nadir</td>
<td>Polar</td>
<td>MW</td>
<td>Passive</td>
<td>$T, q, etc$</td>
</tr>
<tr>
<td>AIRS</td>
<td>Advanced InfraRed Sounder</td>
<td>EOS Aqua</td>
<td>Nadir</td>
<td>Polar</td>
<td>IR/MW/Vis</td>
<td>Passive</td>
<td>$T, q, etc$</td>
</tr>
<tr>
<td>SBUV</td>
<td>Satellite Backscattered UltraViolet</td>
<td>NOAA</td>
<td>Nadir</td>
<td>Polar</td>
<td>UV</td>
<td>Passive</td>
<td>$O_3$</td>
</tr>
<tr>
<td>MIPAS</td>
<td>Michelson Interferometer for Passive Atmospheric Sounding</td>
<td>EnviSat</td>
<td>Limb</td>
<td>Polar</td>
<td>IR/MW</td>
<td>Passive</td>
<td>$T, q, O_3, etc$</td>
</tr>
<tr>
<td>GOME</td>
<td>Global Ozone Monitoring Experiment</td>
<td>ERS-2, METOP</td>
<td>Nadir</td>
<td>Polar</td>
<td>UV</td>
<td>Passive</td>
<td>$O_3$</td>
</tr>
<tr>
<td>SCIAMACHY</td>
<td>SCanning Imaging Absorption spectroMeter for Atmospheric Cartography</td>
<td>EnviSat</td>
<td>Limb/Nadir</td>
<td>Polar</td>
<td>IR</td>
<td>Passive</td>
<td>$O_3, q, clouds, etc$</td>
</tr>
<tr>
<td>MVIRI</td>
<td>Meteosat Visible and InfraRed Imager</td>
<td>MeteoSat</td>
<td>Nadir</td>
<td>Geost.</td>
<td>Vis/IR/WV</td>
<td>Passive</td>
<td>Cloud, surface, motion vectors</td>
</tr>
<tr>
<td>SEVIRI</td>
<td>Spinning Enhanced Visible and InfraRed Imager</td>
<td>MSG</td>
<td>Nadir</td>
<td>Geost.</td>
<td>Vis/IR/WV</td>
<td>Passive</td>
<td>Cloud, surface, motion vectors</td>
</tr>
<tr>
<td>GERB</td>
<td>Geostationary Earth Radiation Experiment</td>
<td>MSG</td>
<td>Nadir</td>
<td>Geost.</td>
<td>LW/SW</td>
<td>Passive</td>
<td></td>
</tr>
<tr>
<td>AVHRR</td>
<td>Advanced Very High Resolution Radiometer</td>
<td>NOAA</td>
<td>Nadir</td>
<td>Polar</td>
<td>Vis/IR/WV</td>
<td>Passive</td>
<td>Cloud, surface, motion vectors</td>
</tr>
<tr>
<td>ATSR</td>
<td>Along Track Scanning Radiometer</td>
<td>ERS-1, 2</td>
<td>Nadir</td>
<td>Polar</td>
<td>Vis/IR/WV</td>
<td>Passive</td>
<td>SST, surface, clouds, cryosphere</td>
</tr>
<tr>
<td>SMOS</td>
<td>Soil Moisture Ocean Salinity</td>
<td>Earth explorer</td>
<td>Nadir</td>
<td>Polar</td>
<td>L-band (1.4GHz)</td>
<td>Passive</td>
<td>Soil moisture, ocean salinity</td>
</tr>
<tr>
<td>SCAT</td>
<td>Scatterometer</td>
<td>ERS-1, 2</td>
<td>QuasiNadir</td>
<td>Polar</td>
<td>C-band (6GHz)</td>
<td>Active</td>
<td>Surface wind</td>
</tr>
<tr>
<td>PR</td>
<td>Precipitation Radar</td>
<td>TRMM</td>
<td>Nadir</td>
<td>NEO</td>
<td>Radar</td>
<td>Active</td>
<td>Precipitation</td>
</tr>
<tr>
<td>GPS/GLONASS</td>
<td>Global Positioning System</td>
<td>Limb</td>
<td>Refractive index</td>
<td>Active</td>
<td>$T, q, p$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'Vis'=visible, 'UV'=ultra violet, "IR"=infrared, 'MW'=microwave, TC03=total column ozone, TCWV=total column water vapour, Geost.=geostationary, NEO=near equator orbit
Deriving information from satellite soundings

A one-dimensional example - to show the need for adequate consideration of errors
Rodgers (2000)

Make \( m \) nadir radiance measurements

Forward model (radiative transfer equation)

\[
L_i(\nu_i) = \int_0^\infty B(\nu, T(z)) K_i(z) \, dz
\]

What is \( B(\nu, T(z)) \) given a set of measurements?

Choose a basis of \( m \) polynomials to represent \( B \),

\[
B(\nu, T(z)) = \sum_{j=1}^{m} w_j z^{j-1}
\]

An inappropriate means of computing the \( w_j \) (and hence \( B \), and hence \( T(z) \)),

\[
L_i(\nu_i) = \sum_{j=1}^{m} C_{ij} w_j \quad C_{ij} = \int_0^\infty z^{j-1} K_i(z) \, dz
\]

\[
\hat{y} = C\hat{w} \quad \Rightarrow \quad \hat{w} = C^{-1}\hat{y}
\]
Results of 'exact' inverse problem

The 'C' operator is ill conditioned

'Exact' methods are inappropriate for real-world inverse problems

Need 'inexact' methods that properly account for errors - use the method of least squares - see later.

Courtesy, Rodgers (2000)
General principles for deriving information from remotely sensed observations

✓ Use of forward model (a.k.a. observation operator).
  • Remotely sensed observations contain information about those model quantities that the operators are sensitive to (e.g. temperature).
✓ Account for error statistics (data are inexact).
✓ Need a-priori information (first guess) - observations may not constrain all unknowns (under constrained).

✗ Exact inversion.

The 'method of least squares' (later) can be used to solve the inexact, ill-conditioned, underconstrained inverse problem.
Deriving chemical species from satellite data

Courtesy Jean Noel Thepaut, ECMWF

Courtesy NASA Goddard
Alternatives for assimilating satellite derived data

- Have hinted that it is possible to derive geophysical information from satellites in a 1d vertical column (called 'retrievals').
- There are a number of options to assimilate satellite data with large 3d weather forecasting models.

<table>
<thead>
<tr>
<th>'L0' Data</th>
<th>'L1' Data</th>
<th>'L2' Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photons (counts)</td>
<td>Direct radiance assimilation.</td>
<td>Assimilate columns as though</td>
</tr>
<tr>
<td></td>
<td>Need radiance operator in large</td>
<td>radiosonde data.</td>
</tr>
<tr>
<td></td>
<td>assimilation problem.</td>
<td>Suboptimal.</td>
</tr>
<tr>
<td></td>
<td>1st choice ← Radiances ( P/(λAΩ) )</td>
<td>(solve small inexact ill-posed</td>
</tr>
<tr>
<td></td>
<td>retrieval algorithm</td>
<td>inverse problems)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd choice ← Columns of geophysical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quantities (vertical 'retrieval'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>profiles)</td>
</tr>
</tbody>
</table>
3. MODELS

DYNAMICAL CORE
(primitive equations)

PARAMETRISATIONS (e.g. ATMOS)
Cloud
SW and LW radiation
Boundary layer
Precipitation
Convection
Surface hydrology
Vertical diffusion
Gravity wave drag
Vegetation
Chemistry (e.g. ozone)

BOUNDARY CONDITIONS
Sea surface temperature
Sea ice
Solar insolation

Met Office "New Dynamics" Unified Model
Semi-Lagrangian advection
Typical res: $0.8^\circ \times 0.5^\circ \times 50$ levs (60km mid-lats)
Typical timestep: ~ 15 minutes

"Charney Phillips" vertical grid
"Arakawa C" horizontal grid

- Coupled atmosphere/ocean models exist, but no coupled data assimilation systems exist.
- Component specific models (e.g. carbon cycle) exist.

Summary of observations and models

• Wealth of obs for use in data assimilation.
• Broadly two types of observation:
  • in-situ (geophysical quantities),
  • remotely sensed (e.g. radiances).
• In-situ obs are straightforward to assimilate:
  • good resolution,
  • poor coverage.
• Remotely sensed obs are complicated to deal with:
  • limited resolution,
  • good coverage.
• Geophysical quantities can be derived from remotely sensed observations:
  • off-line retrieval (1d vertical column) or
  • (e.g.) direct assimilation of radiances.
  • 'forward models' predict observations from geophysical quantities).
• Satellite instruments:
  • orbit types (geostationary, polar, sun synchronous),
  • viewing geometry types (nadir, limb),
  • techniques (e.g. passive, active).
• Only a small fraction of observations survives quality control.
• Observation uncertainty is very important.
• Models are an integral part of data assimilation:
  • models (i) predict obs, (ii) provide a-priori information and (iii) make forecasts.