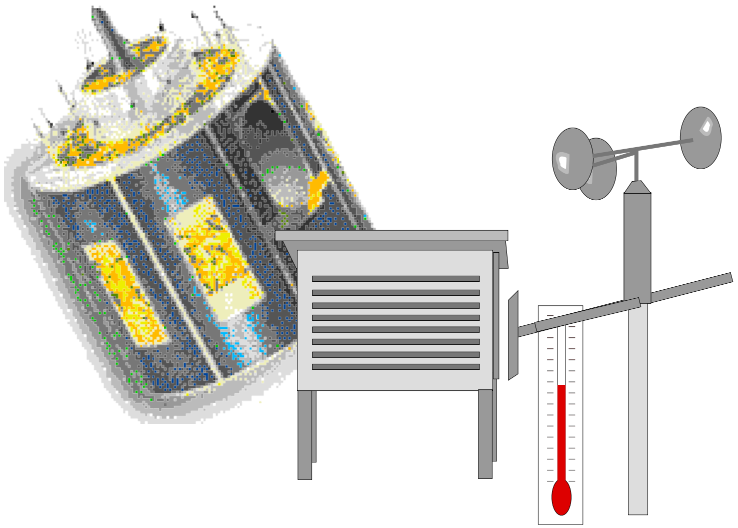
Earth Observation and Data Assimilation
QUEST ES4 Spring School, Sept. 2006

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

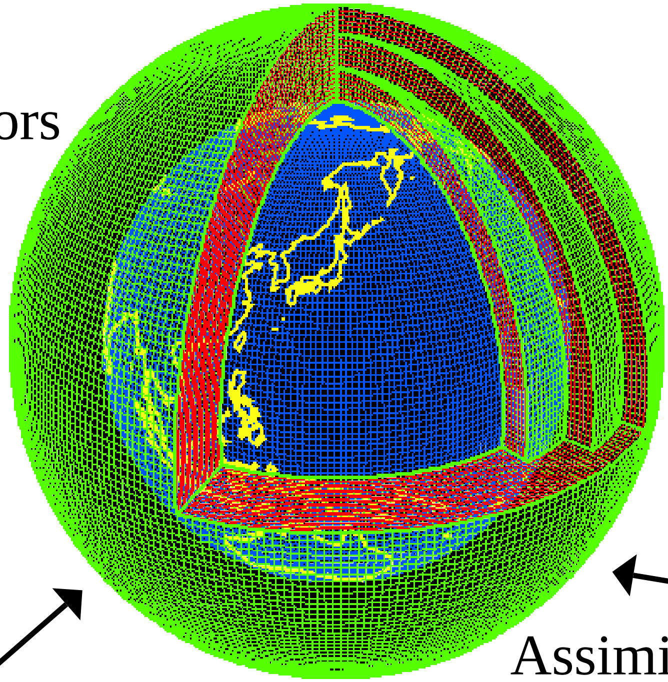


Colour slides available (PDF) via `www.met.rdg.ac.uk/~ross/DARC/DataAssim.html`

Thanks to DARC colleagues: Stefano Migliorini, William Lahoz

1. INTRODUCTION
The data assimilation problem

State vector, $\vec{x}$



Observations, $\vec{y}$, and errors

- Sondes
- Surface stations
- Ships
- Satellites

A-priori information, $\vec{x}_B$, and errors

- Background state
- Best guess
- Forecast

Models ("forward models")

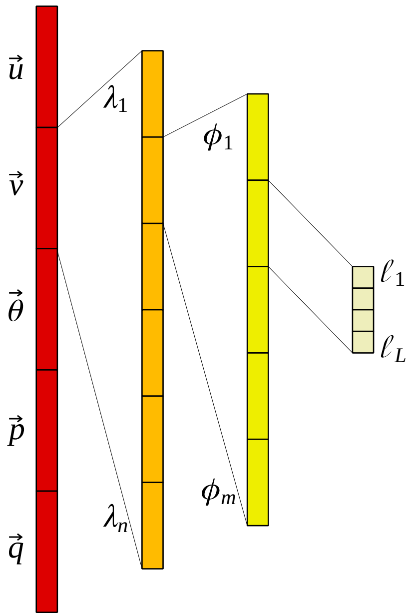
- Linking model state to observations
- $\vec{y} = \vec{h}[\vec{x}] + \vec{\varepsilon}$

Assimilation algorithm ("inverse model")

- Optimal Interpolation (OI)
- Variational data assimilation (Var.)
- Kalman filter

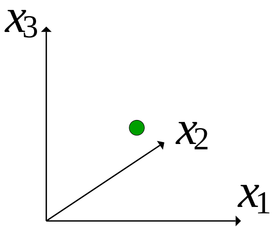
Representation of data

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

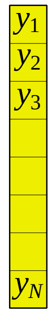


$\vec{u}$ zonal wind field
$\vec{v}$ meridional wind field
$\vec{\theta}$ potential temperature
$\vec{p}$ pressure
$\vec{q}$ specific humidity
λ longitude
ϕ latitude
ℓ vertical level

- 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.



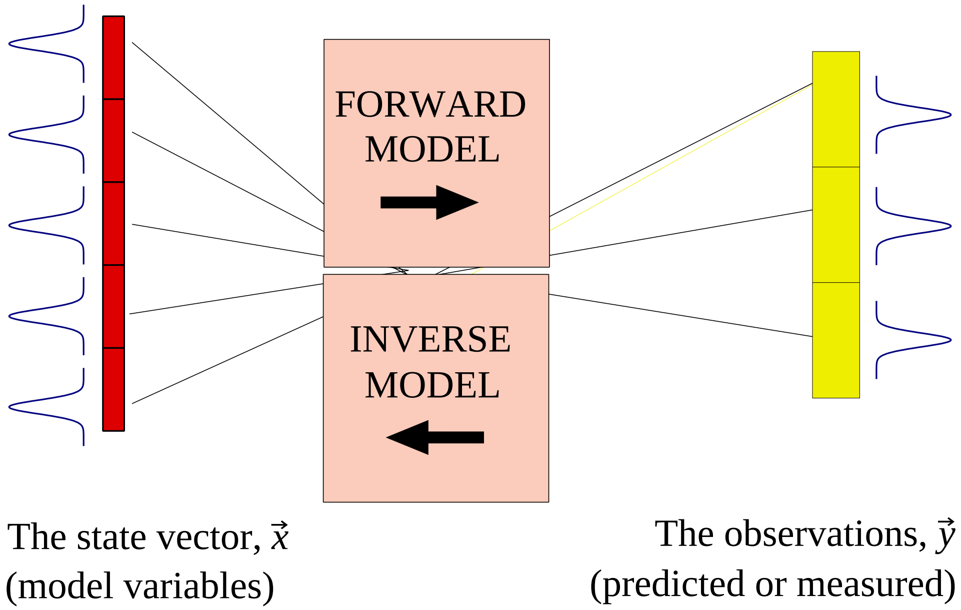
The 'observation vector', $\vec{y}$

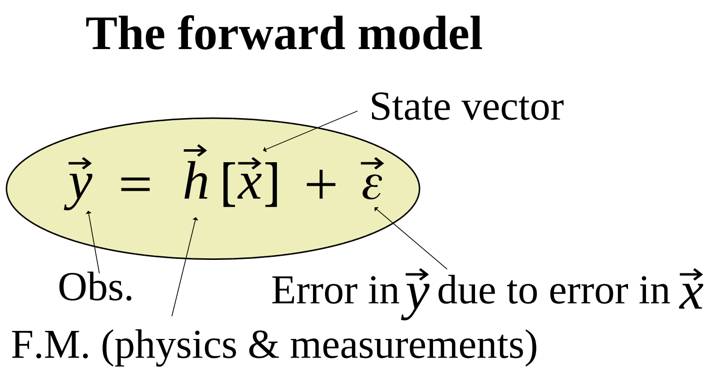


- Every measurement to be assimilated is assembled in this vector.
- The observation type, location and time needs to associated with each observation.

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.

Quantity	Value	Error*	Std. dev.†	Notes
'truth'	x_t	0	n/a	Abstract, as x_t can never be known precisely
obs. 1	y_1	ε_1	σ_1	σ_1 is the precision of inst. 1
obs. 2	y_2	ε_2	σ_2	σ_2 is the precision of inst. 2
best est. of 'truth'	x_a	ε_a	σ_a	σ_a is a fn. of σ_1 and σ_2. a = 'analysis'

*Deviation from 'truth', $y_n = x_t + \varepsilon_n$, $n = 1,2$. ε_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 $\langle y_n \rangle = x_t$.

$$x_a = \left(\frac{y_1}{\sigma_1^2} + \frac{y_2}{\sigma_2^2} \right) \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}, \qquad \sigma_a = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1/2}$$

- This is simple data assimilation.
- The larger the 'σ' 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 σ. Should use the term 'error statistics'.

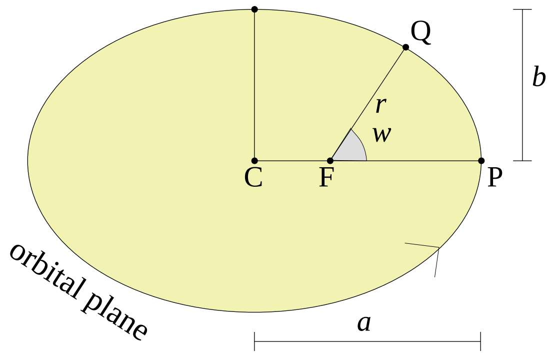
Combining observational data: 6 unknowns, >6 indirect observations - orbital determination

Aim: to estimate the six orbital parameters of Venus, $\vec{x}$, and their uncertainty.

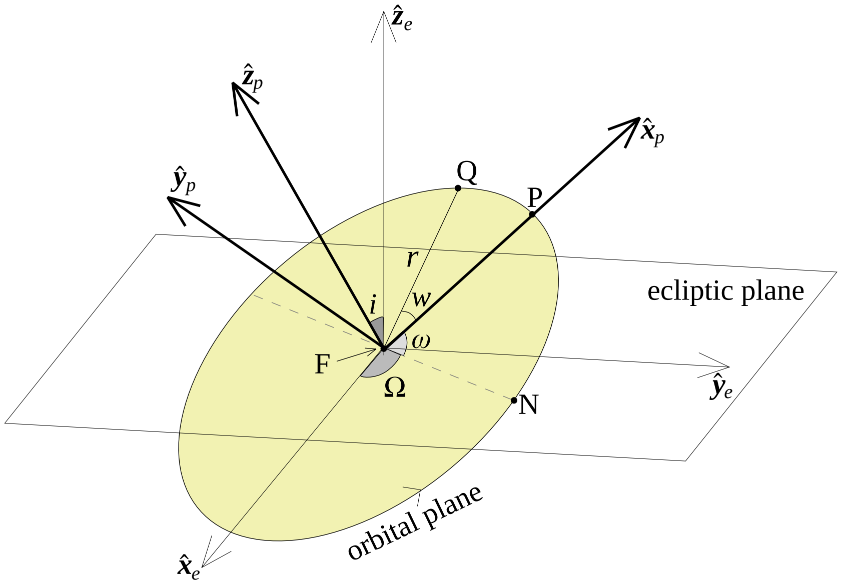
Information to use: many indirect measurements.

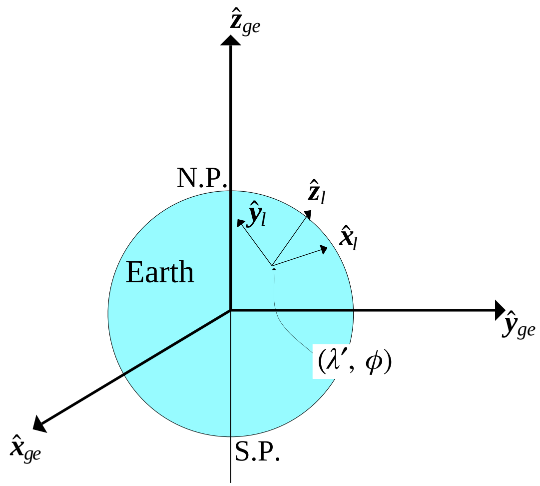
$$\vec{x} = \begin{pmatrix} a \\ e \\ i \\ \Omega \\ \varpi \\ \varepsilon \end{pmatrix}, \qquad \vec{y} = \begin{pmatrix} \text{alt}\,(1) \\ \text{azi}\,(1) \\ \text{alt}\,(2) \\ \text{azi}\,(2) \\ \cdots \\ \cdots \end{pmatrix}$$

$$\vec{y} = \vec{h}[\vec{x}] + \vec{\varepsilon}$$



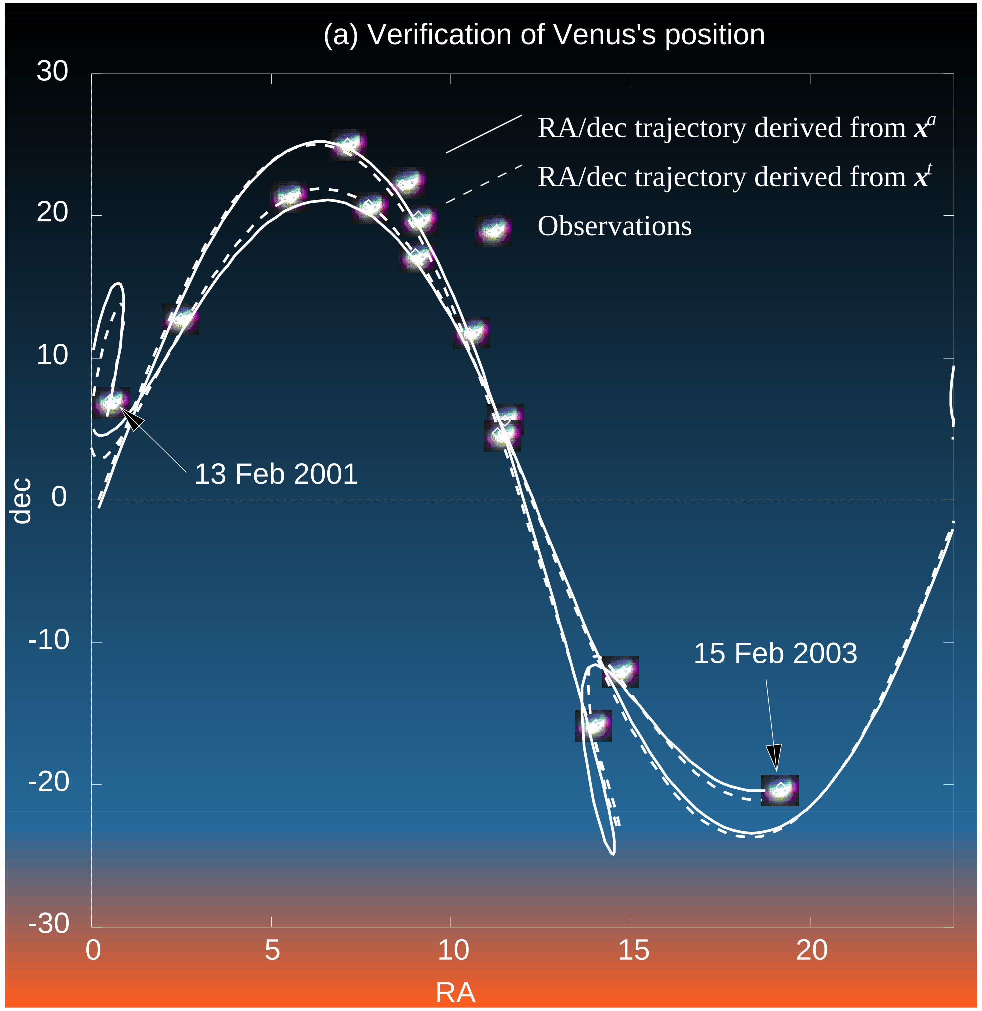
$$e^2 = 1 - \left(\frac{b}{a}\right)^2$$



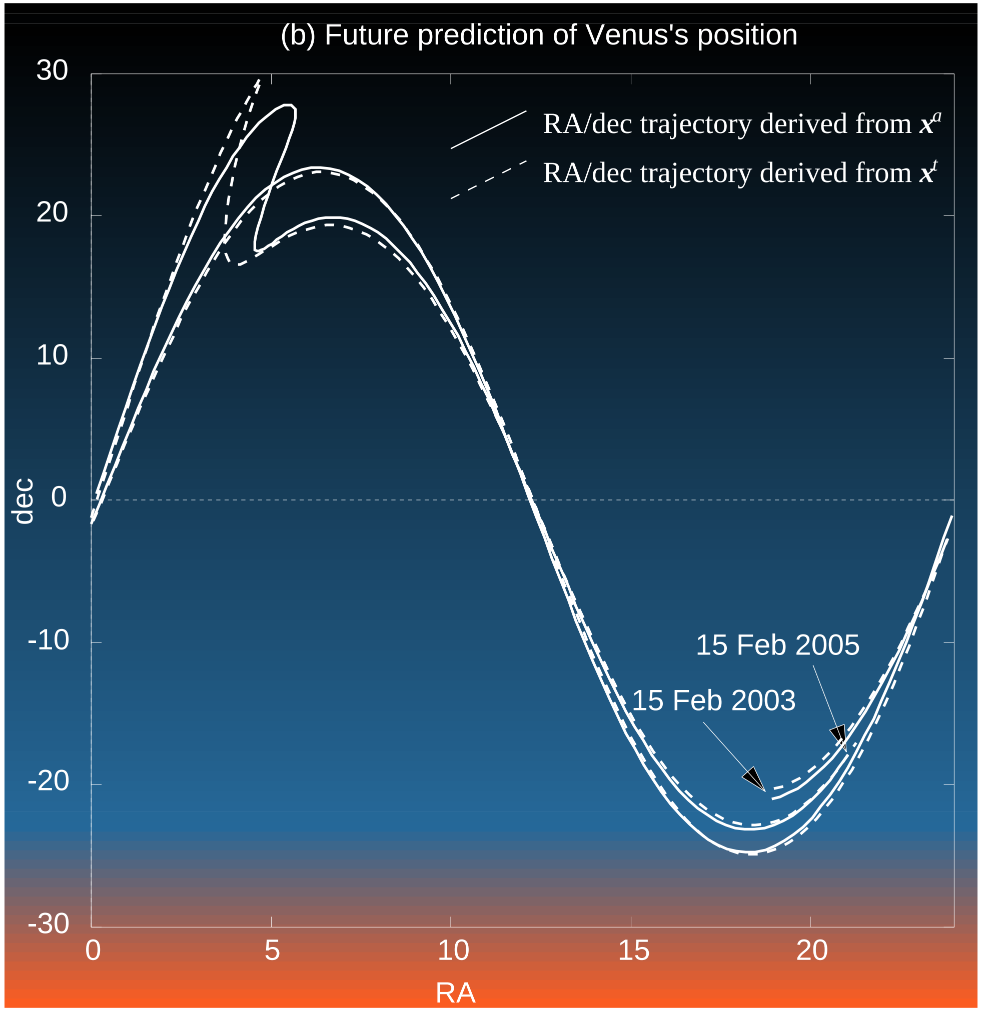


$$\boldsymbol{x} = (0.7210,\ 0.0201,\ 4.23,\ 88.9,\ 110.0,\ 176.6)$$
$$\sigma = (0.0020,\ 0.0078,\ 0.70,\ 8.1,\ 50.3,\ 6.8)$$
$$\boldsymbol{x}_t = (0.7233,\ 0.0067,\ 3.39,\ 76.7,\ 131.5,\ 182.0)$$

(a) Assimilation period

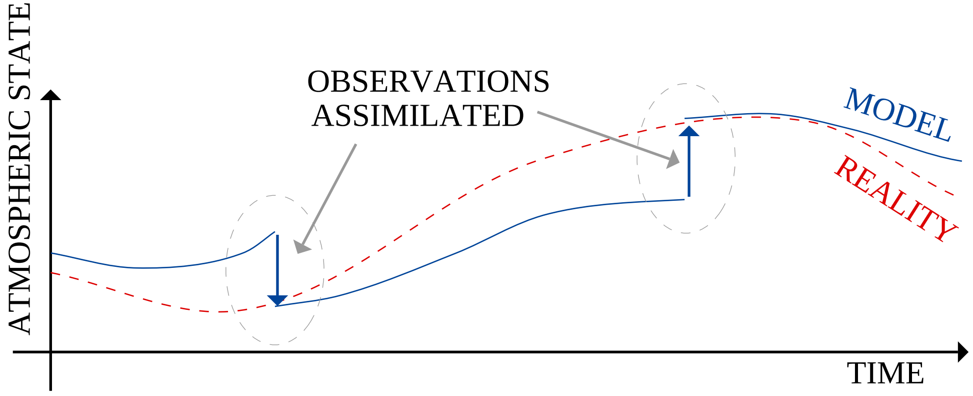


(b) Forecast period



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).

CONTENTS OF LECTURES

2. OBSERVATIONS
Types of instrument

Measurements from instruments assimilated routinely (not exhaustive)

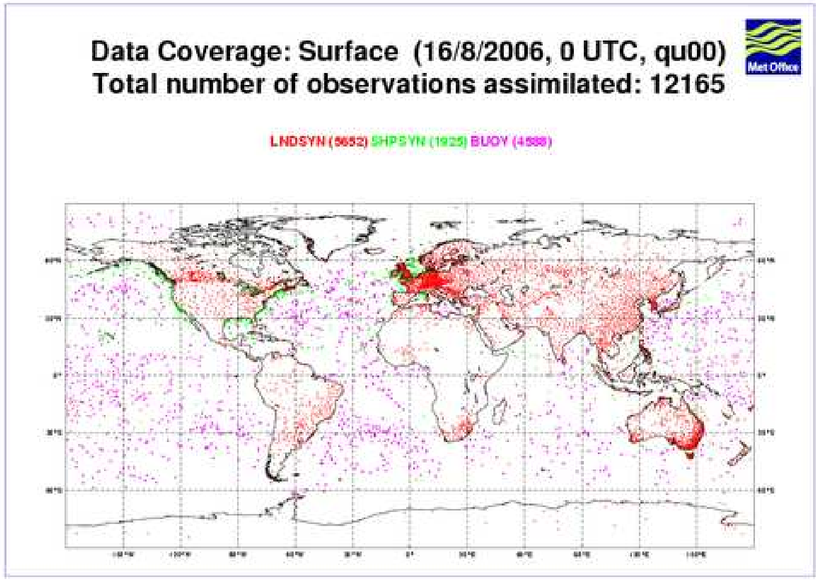
Instrument	Quantities measured	Coverage Spatial	Coverage Temporal	Resolution Horiz.	Resolution Vert.
In-situ instruments					
Radiosondes	u, v, T, p, q, (O_3)	Cont'l N.H., t'sphere	6 hourly	point	point
Surface stations	u, v, T, p, q	Cont'l, surface	6 hourly	point	n/a
Aircraft	u, v, T, p, q	Flight paths, airports	In flight	point	point
Drifting buoys	u, v, T, p	Drift paths, sea lev.	Hourly	point	n/a
Remote sensing instruments					
Geostationary sat.	Rad: MW, IR, Vis	Global	15-30 mins	> 1 km	kms
Polar orbiting sat. (nadir)	Rad: MW, IR, Vis	Global	Continuous	> 1 km	kms
Polar orbiting sat. (limb)	Rad: MW, IR, Vis	Global	Continuous	100s km	1-2 km
Scatterometer	Radar backscatter	Oceans	Continuous	50 km	n/a

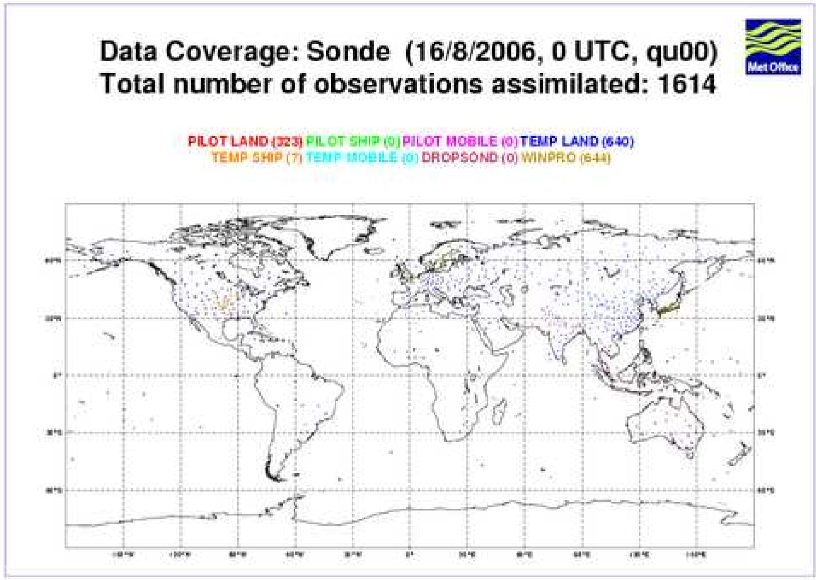
'Rad'=radiances, 'MW'=microwave, 'IR'=infrared, 'Vis'=visible

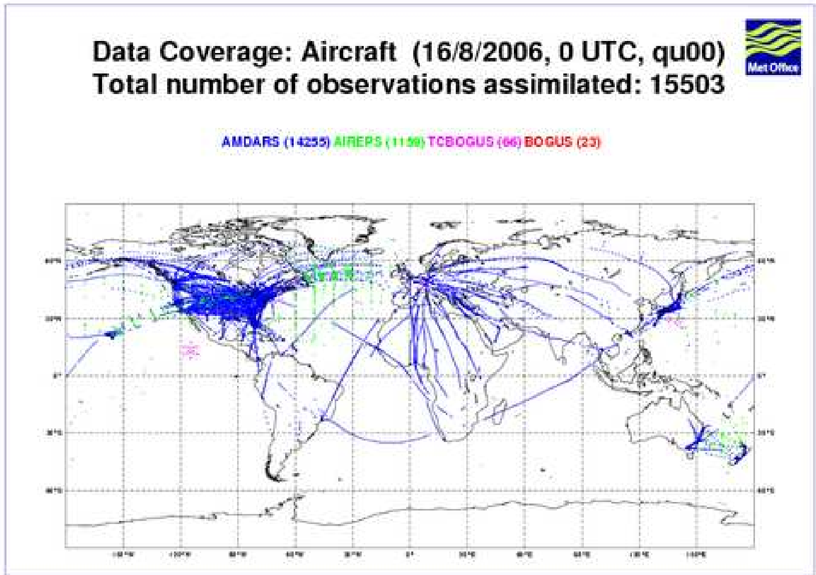
In operational global weather forecasting there are ~10^6 observations assimilated per cycle

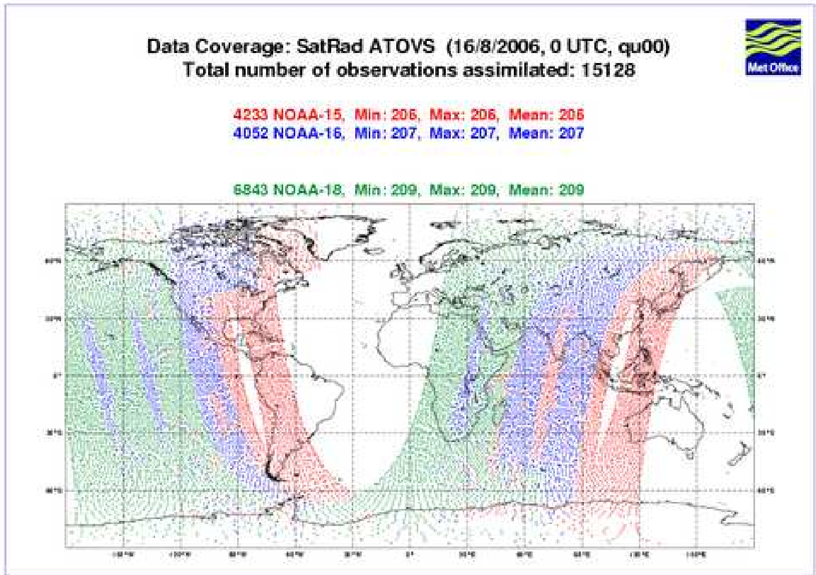
Coverage

Locations of four example observation types *(courtesy Met Office (c) Crown copyright)*

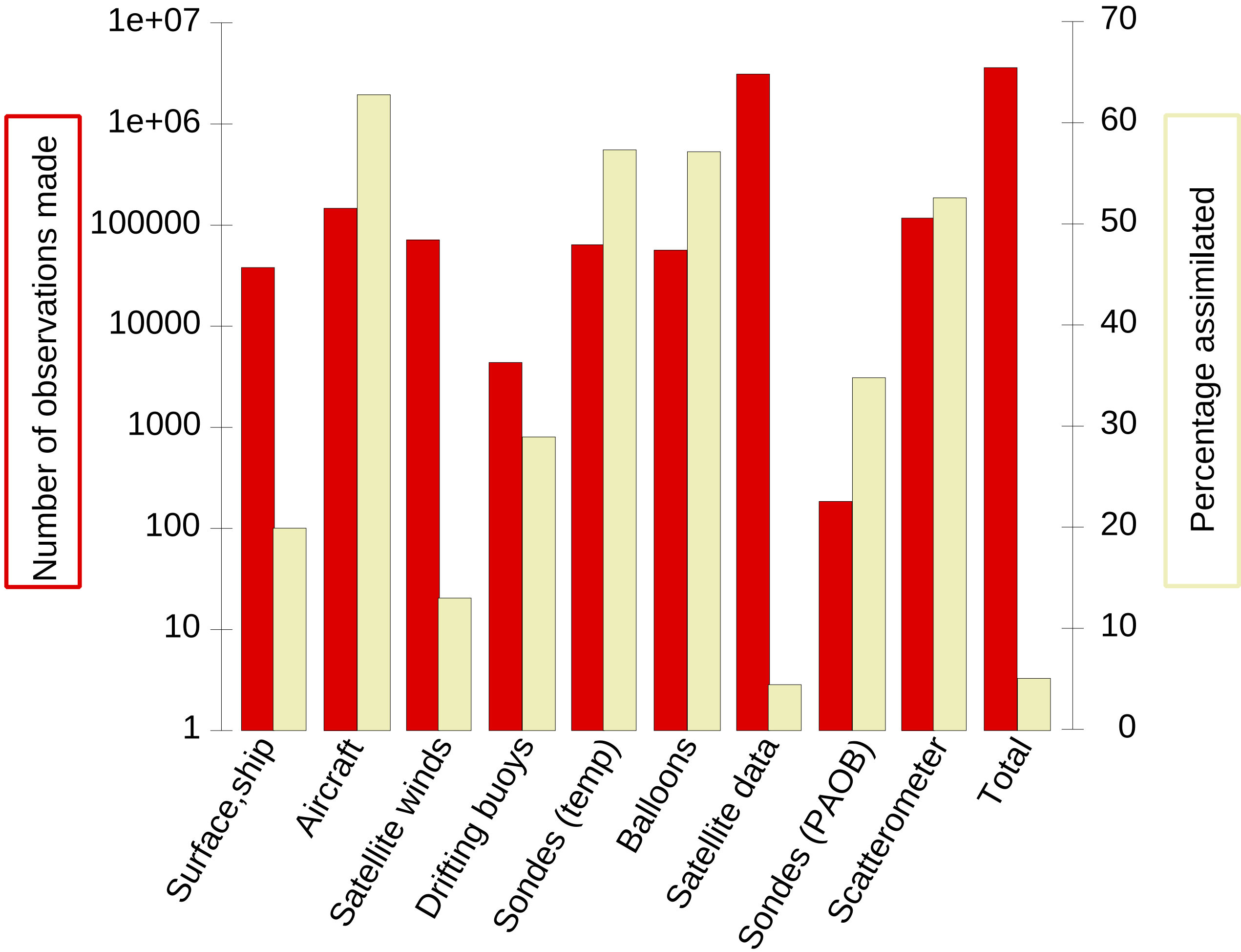








Volumes of data and quality control



ECMWF stats. (one cycle in June '03)

Total No. obs.: ~ 70,000,000

Total No. assimilated: ~ 3,500,000

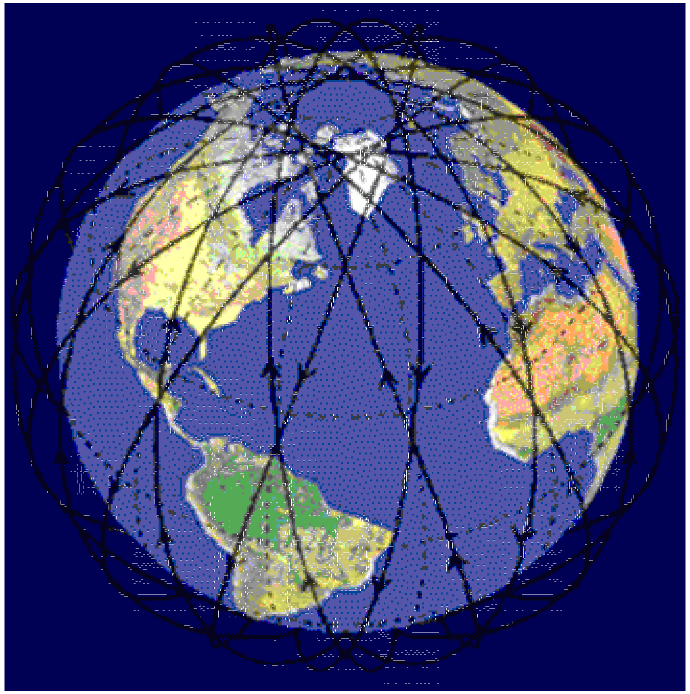
(only 5%!)

Why are some observations rejected?

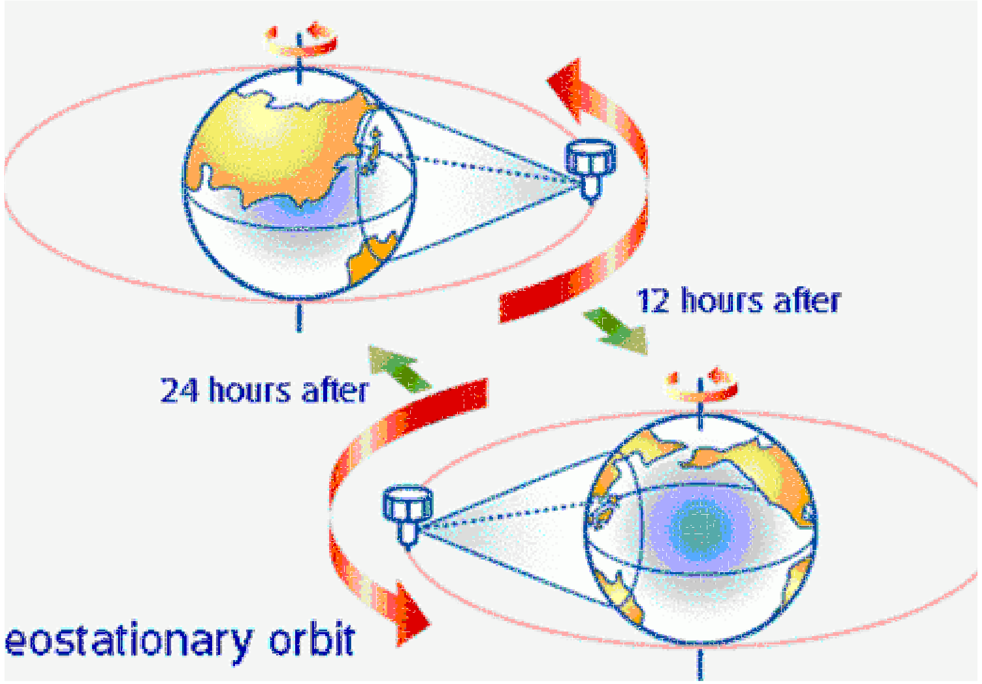
- Observation 'too far' from forecast (large systematic, human, or instrument error),
- Observation did not reach centre in time,
- Satellite radiance data - complications due to radiation from land, clouds or precipitation.

Satellite borne instruments

Orbit configurations



Polar orbiter (courtesy WAL)



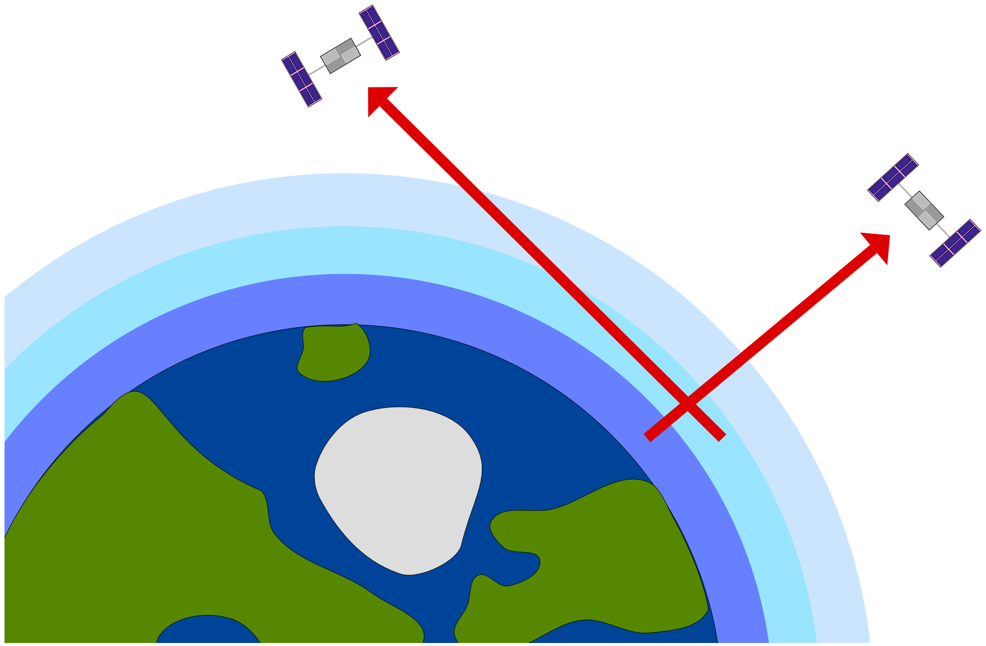
Geostationary orbit (courtesy NASDA)

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

- 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!)

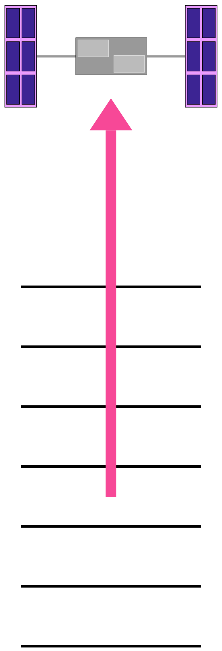
Instrument	Expanded name	Platform	Geometry	Orbit	Measures	Pass/Act	Sensitive to
HIRDLS	High Resolution Dynamics Limb Sounder	EOS Aura	Limb	Polar	?	Passive	*T*, *q*, O3, etc
OMI	Ozone Monitoring Experiment	EOS Aura	Nadir	Polar	Vis/UV	Passive	O3, TCO3, etc.
TES	Tropospheric Emission Spectrometer	EOS Aura	Limb/Nadir	Polar	IR	Passive	*T*, *q*, O3, etc
MLS	Microwave Limb Sounder	EOS Aura	Limb	Polar	MW	Passive	*T*, *q*, O3, etc
SSM/I	Special Sensor Microwave Imager	DMSP	Nadir	Polar	MW	Passive	TCWV, cloud, precip, surface wind, snow, sea ice
HIRS	High resolution InfraRed Sounder	NOAA	Nadir	Polar	IR	Passive	*T*, *q*, O3, etc
AMSU	Advanced Microwave Sounding Unit	NOAA	Nadir	Polar	MW	Passive	*T*, *q*, etc
AIRS	Advanced InfraRed Sounder	EOS Aqua	Nadir	Polar	IR/MW/Vis	Passive	*T*, *q*, etc
SBUV	Satellite Backscattered UltraViolet	NOAA	Nadir	Polar	UV	Passive	O3
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding	EnviSat	Limb	Polar	IR/MW	Passive	*T*, *q*, O3, etc
GOME	Global Ozone Monitoring Experiment	ERS-2,METOP	Nadir	Polar	UV	Passive	O3
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY	EnviSat	Limb/Nadir	Polar	IR	Passive	O3, *q*, clouds, etc
MVIRI	Meteosat Visible and InfraRed Imager	MeteoSat	Nadir	Geost.	Vis/IR/WV	Passive	Cloud, surface, motion vectors
SEVIRI	Spinning Enhanced Visible and InfraRed Imager	MSG	Nadir	Geost.	Vis/IR/WV	Passive	Cloud, surface, motion vectors
GERB	Geostationary Earth Radiation Experiment	MSG	Nadir	Geost.	LW/SW	Passive	
AVHRR	Advanced Very High Resolution Radiometer	NOAA	Nadir	Polar	Vis/IR/WV	Passive	Cloud, surface, motion vectors
ATSR	Along Track Scanning Radiometer	ERS-1, 2	Nadir	Polar	Vis/IR/WV	Passive	SST, surface, clouds, cryosphere
SMOS	Soil Moisture Ocean Salinity	Earth explorer	Nadir	Polar	L-band (1.4GHz)	Passive	Soil moisture, ocean salinity
SCAT	Scatterometer	ERS-1,2	QuasiNadir	Polar	C-band (6GHz)	Active	Surface wind
PR	Precipitation Radar	TRMM	Nadir	NEO	Radar	Active	Precipitation
GPS/GLONASS	Global Positioning System		Limb		Refractive index	Active	T, *q*, *p*

'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



$$\vec{y} = \begin{pmatrix} L(\nu_1) \\ L(\nu_2) \\ \dots \\ L(\nu_m) \end{pmatrix}$$

Forward model (radiative transfer equation)

$$L_i(\nu_i) = \int_0^\infty B(\langle\nu\rangle, T(z)) \; K_i(z) \, \mathrm{d}z$$

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

Choose a basis of m polynomials to represent B,

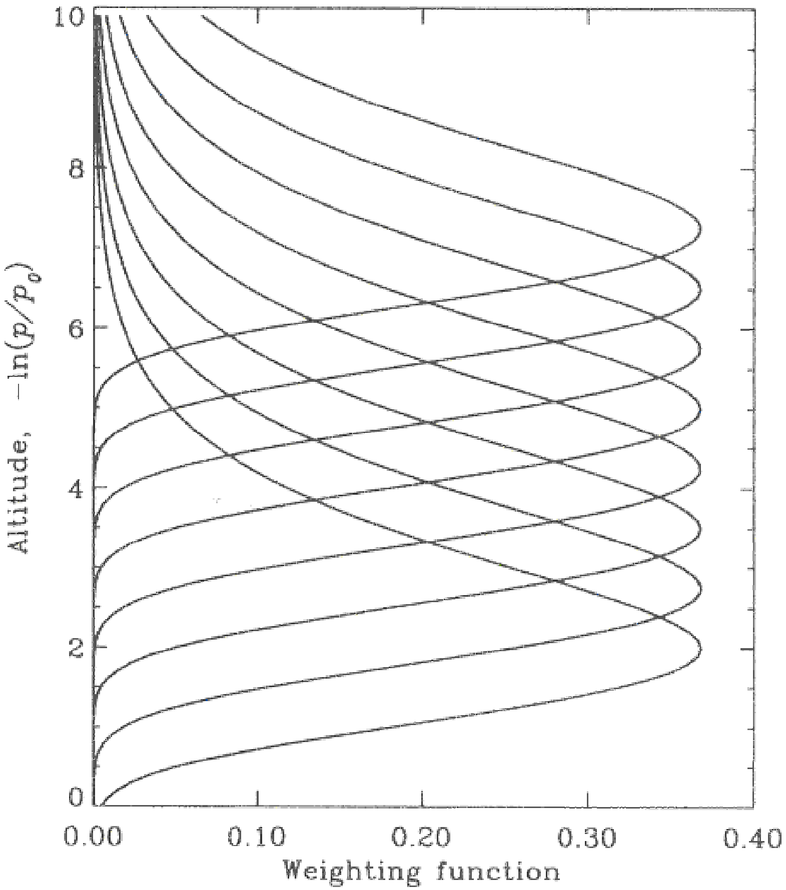
$$B(\langle\nu\rangle, T(z)) = \sum_{j=1}^{m} w_j z^{j-1}$$

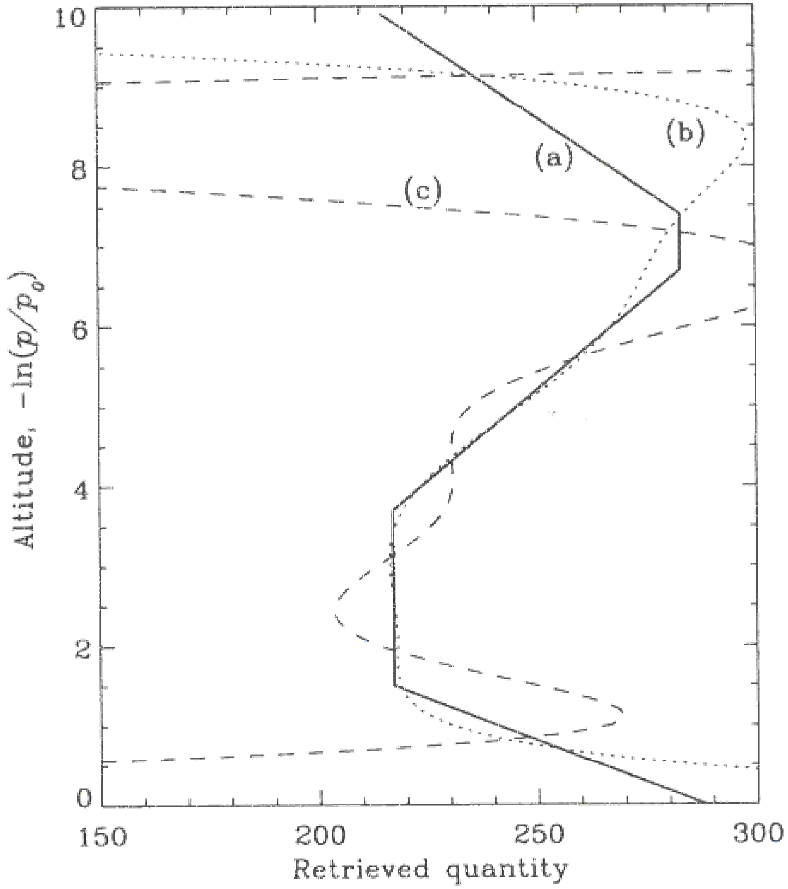
An inappropriate means of computing the w_j (and hence B, and hence $T(z)$),

$$L_i(\nu_i) = \sum_{j=1}^{m} C_{ij} w_j \qquad C_{ij} = \int_0^\infty z^{j-1} K_i(z) \, \mathrm{d}z$$

$$\vec{y} = \mathbf{C}\vec{w} \qquad \Rightarrow \qquad \vec{w} = \mathbf{C}^{-1}\vec{y}$$

Results of 'exact' inverse problem





Courtesy, Rodgers (2000)

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.

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).

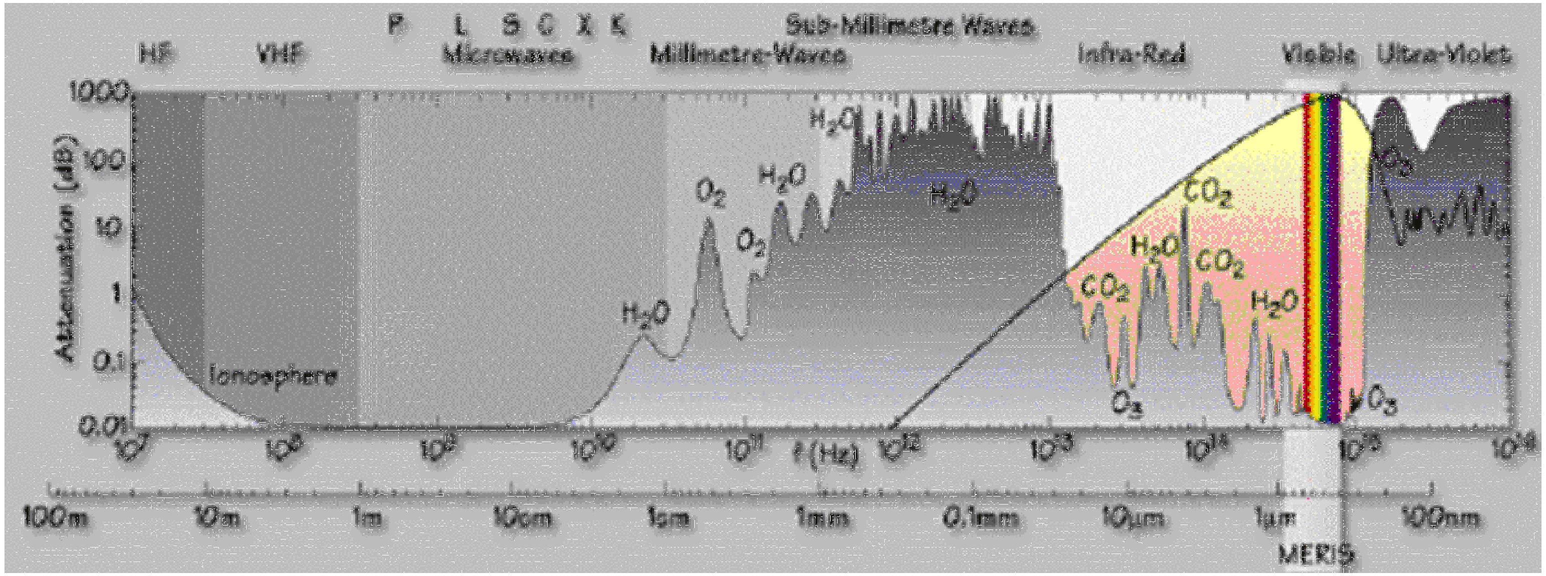
✓Account for error statistics (data are *inexact*).

✓Need a-priori information (first guess) - observations may not constrain all unknowns (*under constrained*).

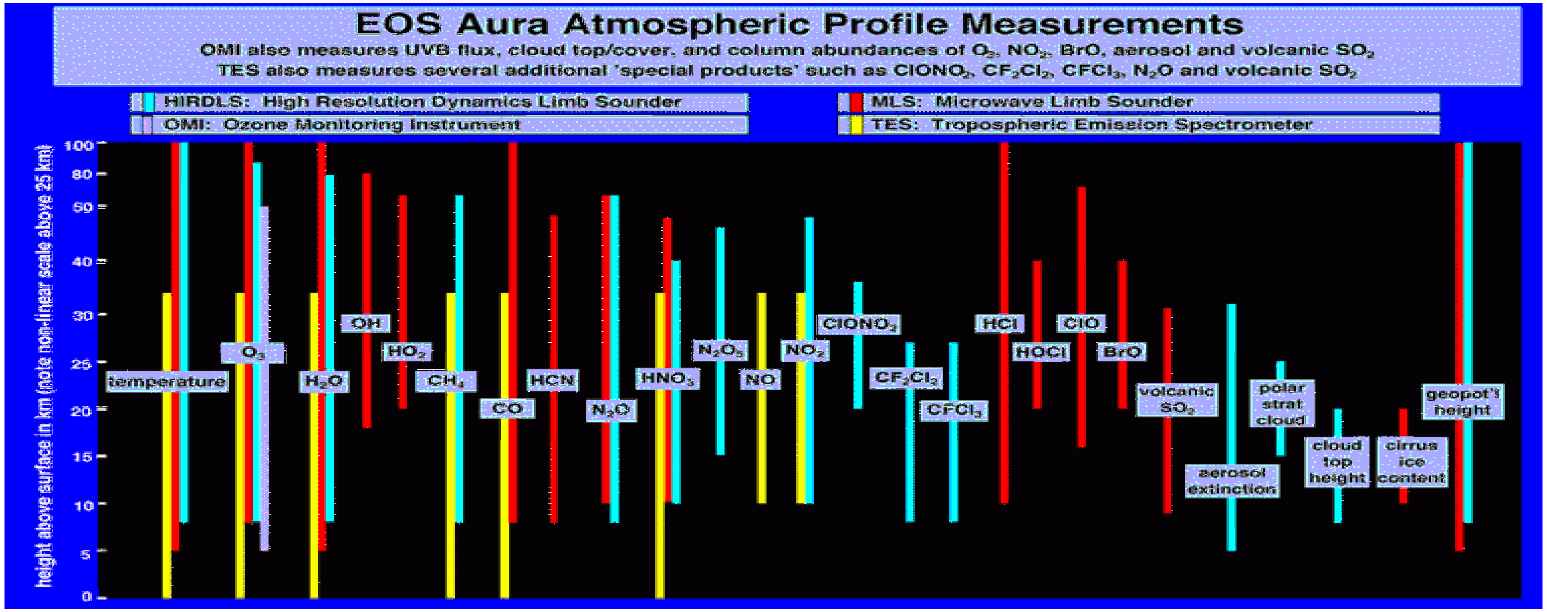
✗Exact inversion.

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.

<u>'L0' Data</u>

Photons

(counts)

↓ algorithm

<u>'L1' Data</u>

1st choice ← Radiances

$(P/(\lambda A \Omega))$

Direct radiance assimilation.

Need radiance operator in large assimilation problem.

↓ retrieval algorithm

(solve small inexact ill-posed inverse problems)

<u>'L2' Data</u>

2nd choice ← Columns of geophysical quantities

(vertical 'retrieval' profiles)

Assimilate columns as though radiosonde data.

Suboptimal.

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

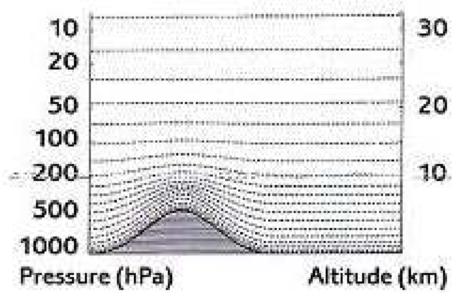
Met Office "New Dynamics" Unified Model

Semi-Lagrangian advection

Typical res: 0.8° × 0.5° × 50 levs (60km mid-lats)

Typical timestep: ~ 15 minutes

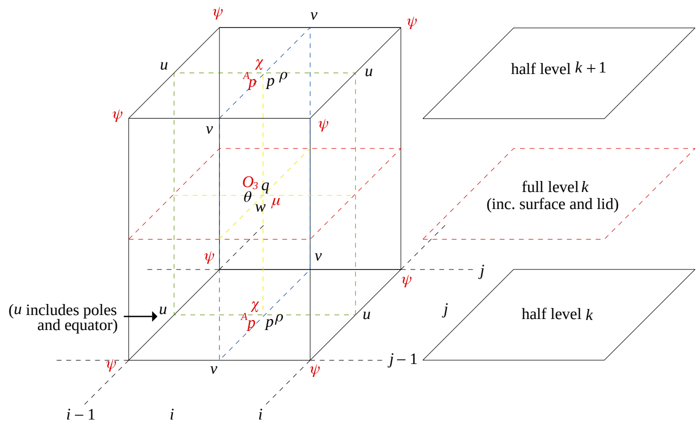
Courtesy Australian Bureau for Meteorology



Courtesy climateprediction.net

"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.