Satellite data assimilation for Numerical Weather Prediction (NWP)

Sean Healy
European Centre for Medium-range Weather Forecasts (ECMWF)

with contributions from many ECMWF colleagues

Special thanks to: Jean-Noël Thépaut (based on his lecture), Tony McNally, Niels Bormann, Stephen English, Peter Bauer, C. Cardinali …
Outline

1. Introduction to data assimilation for NWP
   - Data assimilation process
   - Observations used
   - Forecast impact of observations

2. Passive atmospheric sounding
   - What do satellite sensors measure?
   - Weighting functions

3. Retrieval algorithms
   - Forward versus inverse problem
   - Solutions to reduced problems
   - Optimal estimation/1DVAR methods with forecast background

4. Direct radiance assimilation and 4DVAR

5. Summary
How does NWP use observations?

1.) Introduction to data assimilation for NWP
Assimilation of measurements in NWP

- A cycling data assimilation system.

- A short-range forecast provides us with an 3D estimate of the atmospheric state (P,T, Q, U,V, Ozone).

- We combine the short-range forecast with observations in a statistically optimal way to produce the “analysis”.

- The analysis provides the initial conditions for the next forecast.
4D-Var assimilation

Its fancy least squares. Optimal estimation:

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y_m - H(x))^T R^{-1} (y_m - H(x))$$

Fit ~10 million values in a 12 hour window!
Example of conventional data coverage

SYNOP/SHIP observations
Total number of obs = 29131

BUOYS
Total number of obs = 5413

Radiosondes
Total number of obs = 589

Aircraft data
Total number of obs = 40926
Example of 6-hourly satellite data coverage

- LEO Sounders
- LEO Imagers
- Scatterometers
- GEO imagers
- Satellite Winds (AMVs)
- GPS Radio Occultation
Profilers
Radiosonde
Synop
Ship
Aircraft
Buoys

Composition
Ozone sondes
Air quality stations

Mass

Soil moisture
Rain gauge

Moisture

Wind

Aircraft
Buoy

Profilers

Radiosonde
Synop
Ship

Environment and Climatechange
Monitoring
Assessment
Forecasting

Earth Observation
Satellite

ESA Summer School 2014
Number of satellite data products actively assimilated at ECMWF

- POES
- COSMIC
- COSMIC-2
- Megha Tropiques
- Oceansat
- TERRA
- Sentinel 1
- Suomi-NPP
- DMSP
- CNOFS
- AQUA
- AURA
- HY-2A
- Meteosat
- Cryosat
- SMOS
- Metop
- GRACE
- FY-3A/B
- QuikSCAT
- GOES
- MTSAT
- JASON-1/2/3
- FY-2C/D
- GOSAT
- ERS-1/2
- ENVISAT
- TRMM
- ADM Aeolus
- EarthCARE
- ECMWF
Number of satellite data products monitored at ECMWF

- POES
- COSMIC
- TERRASAR-X
- FY3
- GOES
- EarthCare
- Suomi-NPP
- COSMIC-2
- GCOM-W/C
- QuikSCAT
- MTSAT
- ADM Aeolus
- DMSP
- CHAMP
- TRMM
- JASON-1/2/3
- FY-2C/D
- GOSAT
- Metop
- GRACE
- Megha Tropiques
- Oceansat
- TERRA
- Sentinel 1
- ERS-1/2
- CNOFS
- SAC-C
- AURA
- AURA
- SWISAT
- Cryosat
- Sentinel 3
- SMOS
- Sentinel 5p
Number of observations (1000s)
Anomaly correlation of 500hPa height forecasts

Operations

- Northern hemisphere
- Southern hemisphere

ERA-Interim

- Northern hemisphere
- Southern hemisphere

ECMWF
Combined impact of all satellite data

EUCOS Observing System Experiments (OSEs):

- 2007 ECMWF forecasting system,
- winter & summer season,
- different baseline systems:
  - no satellite data (NOSAT),
  - NOSAT + AMVs,
  - NOSAT + 1 AMSU-A,
- general impact of satellites,
- impact of individual systems,
- all conventional observations.

$\leftarrow 500$ hPa **geopotential height** anomaly correlation
Satellite data provide robustness to the global numerical forecasts.
Advanced diagnostics

Data assimilation:

State at initial time $\rightarrow$ NWP model $\rightarrow$ State at time $i$ $\rightarrow$ Observation operator $\rightarrow$ Observation simulations

State at initial time $\rightarrow$ Sensitivity of cost to change at initial time $\rightarrow$ AD of forecast model $\rightarrow$ Sensitivity of cost to change in state at time $i$ $\rightarrow$ AD of observation operator

Forecast sensitivity:

State at initial time $\rightarrow$ NWP model $\rightarrow$ State at time $i$ $\rightarrow$ Cost function $J$ $\rightarrow$ Analysis

State at analysis time $\rightarrow$ Sensitivity of cost to change at initial time $\rightarrow$ AD of forecast model $\rightarrow$ Sensitivity of cost to change in state at time $i$

Adjoints come up when minimizing functions

Take the derivative of the model $\frac{\partial M}{\partial x}$ to get a matrix and then transpose it. $\left(\frac{\partial M}{\partial x}\right)^T$. 

=max. 12 hours

Forecast sensitivity:

24 hours

Data assimilation:
Advanced diagnostics

Relative FC error reduction per system

Nadir sounders AMSU-A, AIRS, and IASI provide largest impact

(From C. Cardinali)
How do passive nadir sounders measure the atmosphere?

2.) Passive atmospheric sounding
What do satellite instruments measure?

They DO NOT measure TEMPERATURE.
They DO NOT measure HUMIDITY or OZONE.
They DO NOT measure WIND.

You might retrieve information about these quantities, but you don’t measure them directly!

- Satellite instruments measure the radiance $L$ that reaches the top of the atmosphere at a given frequency $\nu$.
- The measured radiance is related to geophysical atmospheric variables ($T,Q,O_3$, clouds etc…) by the radiative transfer equation.

$$L(\nu) = \int_0^\infty B(\nu, T(z)) \left[ \frac{d\tau(\nu)}{dz} \right] dz + \text{Surface emission} + \text{Surface reflection/scattering} + \text{Cloud/rain contribution} + \ldots$$
Atmospheric spectrum

- Depending on the wavelength, the radiation at the top of the atmosphere is sensitive to different atmospheric constituents.
Frequency selection

By selecting radiation at different frequencies or CHANNELS a satellite instrument can provide information on a range of geophysical variables.

In general, the channels currently used for NWP applications may be considered as one of two different types:

• Atmospheric sounding channels
• Surface sensing channels

In practice real satellite instruments have a combination of both atmospheric sounding and surface sensing channels.
Atmospheric sounding channels

These channels are located in parts of the infra-red and microwave spectrum for which the main contribution to the measured radiance is described by:

\[ L(\nu) = \int_0^\infty B(\nu, T(z)) \left[ \frac{d\tau(\nu)}{dz} \right] dz \]

That is they avoid frequencies for which surface radiation and cloud contributions are important. They are primarily used to obtain information about atmospheric temperature and humidity.

AMSUA-channel 5 (53GHz)  HIRS-channel 12 (6.7micron)
Surface sensing channels

These are located in window regions of the infra-red and microwave spectrum at frequencies where there is very little interaction with the atmosphere and the main contribution to the measured radiance is:

$$L(\nu) = \text{Surface emission} \left[ T_{\text{surf}}, \varepsilon(u,\nu) \right]$$

These are primarily used to obtain information on the surface temperature and quantities that influence the surface emissivity such as wind (ocean) and vegetation (land). They can also be used to obtain information on clouds/rain and cloud movements (to provide wind information) or total-column atmospheric quantities.

SSM/I channel 7 (89GHz)

HIRS channel 8 (11microns)
Atmospheric temperature sounding

Select sounding channels for which

\[ L(\nu) = \int_{0}^{\infty} B(\nu, T(z)) \left[ \frac{d\tau(\nu)}{dz} \right] dz \]

and the primary absorber is a well mixed gas (e.g. oxygen in MW or CO\textsubscript{2} in IR).

Then the measured radiance is essentially a weighted average of the atmospheric temperature profile:

\[ L(\nu) = \int_{0}^{\infty} B(\nu, T(z)) K(z) dz \]

with \( K(z) = \left[ \frac{d\tau}{dz} \right] \)

The function \( K(z) \) that defines this vertical average is known as a weighting function.
Ideal weighting functions

If the weighting function was a delta-function, this would mean that the measured radiance is sensitive to the temperature at a single level in the atmosphere.

If the weighting function was a box-car function, this would mean that the measured radiance was sensitive to the mean temperature between two atmospheric levels.
Atmospheric weighting functions

High in the atmosphere very little radiation is emitted, but most will reach the top of the atmosphere.

At some level there is an optimal balance between the amount of radiation emitted and the amount reaching the top of the atmosphere.

A lot of radiation is emitted from the dense lower atmosphere, but very little survives to the top of the atmosphere due to absorption.
Weighting functions continued

- The altitude at which the peak of the weighting function occurs depends on the strength of absorption for a given channel.

- Channels in parts of the spectrum where the absorption is strong (e.g. near the centre of CO₂ or O₂ lines) peak high in the atmosphere.

- Channels in parts of the spectrum where the absorption is weak (e.g. in the wings of CO₂ or O₂ lines) peak low in the atmosphere.

By selecting a number of channels with varying absorption strengths we sample the atmospheric temperature at different altitudes.
More weighting functions

- AMSUA: 15 channels
- HIRS: 19 channels
- AIRS: 2378 channels
- IASI: 8461 channels
**Important satellite instruments for NWP**

**AMSU-A:**
- Advanced *Microwave* Sounding Unit
- 15 *channels* (12 in 50-60 GHz region)
- 48 km field-of-view (nadir), 2074 km swath
- Primarily temperature-sounding
- On-board NOAA-15-19, Aqua, METOP-A

**AIRS:**
- Atmospheric *Infrared* Sounder
- 2378 *channels* covering 650 - 2700 cm\(^{-1}\) (3.7-15.4 μm)
- 13.5 km field-of-view (nadir), 2130 km swath
- Primarily temperature/humidity-sounding, trace gases
- On-board Aqua

**IASI:**
- *Infrared* Atmospheric Sounding Interferometer
- 8461 *channels* covering 645 - 2760 cm\(^{-1}\) (3.6-15.5 μm)
- 12 km field-of-view (nadir), 2132 km swath
- Primarily temperature/humidity-sounding, trace gases
- On-board METOP-A
How do we extract atmospheric information (e.g. temperature) from satellite radiances?

3.) Retrieval algorithms
If we know the entire atmospheric temperature profile $T(z)$ then we can compute the radiances a sounding instrument would measure using the *radiative transfer equation*. This is sometimes known as the **forward problem**.

In order to extract or **retrieve** the atmospheric temperature profile from a set of measured radiances we must solve what is known as the **inverse problem**.

Unfortunately as the weighting functions are generally broad and we have a finite number of channels, the inverse problem is **formally ill-posed** because an infinite number of different temperature profiles could give the same measured radiances !!!

A simple ill-posed problem

- Assume I measure
  \[ y = x_1 + x_2 \]
  Ignore any measurement error for the moment.

- But I want to know the individual components of \( x \) using \( y \).
  \[ x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]

- There are an infinite number of \((x_1 + x_2)\) combinations that are consistent with the measurement.

- We need more, a-priori, information to estimate both \( x_1 \) and \( x_2 \). The measurement alone cannot provide the solution.
  A priori: Perhaps I already have an estimate of \( x_1 \) and \( x_2 \).
Some history ... John Eyre (Met Office)

- Satellite data assimilation into NWP models began the 1970’s, but by the 1980’s it was clear that this was having a –ve impact.

- Early NWP systems designed to make use of sondes. Satellite products made to look like sondes.

- Satellite sounders and sondes have opposite strengths and weaknesses

- Treating satellite soundings as “poor-quality sondes” is flawed.

- Need to really consider the actual information content of the “raw” measurement!
Retrieval schemes for NWP

The **linear data assimilation schemes** used in the past at ECMWF such as Optimal Interpolation (OI) were unable to assimilate radiance observations directly (as they were nonlinearly related to the analysis variables) and the radiances had to be *explicitly converted to temperature products* before the analysis.

This conversion was achieved using a variety of **retrieval algorithms** that differed in the way they used prior information.

**All retrieval schemes** use some (either explicit of implicit) form of **prior information** to supplement the information of the measured radiances and solve the inverse problem!

Several different types of retrieval have been used in NWP:

Examples:
1. Regression / Neural Net (statistical) methods
2. Forecast background (1DVAR) methods
The difference between (retrievals/products) and observations

- This is an extremely important point! The retrieval is not the observation.

- Retrieval
  - $= (\text{something} \times a \text{ priori}) + (\text{something else} \times \text{observation})$

- Mathematically, a retrieval can be written as a matrix equation.
  $x_a = (I - KH)x_b + Ky_m$

- It is crucial to consider the role of the a-priori when using retrievals. What information has the observation really provided?
1. **Regression and Library search**

Using a sample of temperature profiles matched (collocated) with a sample of radiance observations/simulations, a *statistical relationship* is derived that predicts e.g. atmospheric temperature from the measured radiance.

These tend to be limited by the statistical characteristics of the training sample / profile library and will not produce physically important features if they are statistically rare in the training sample. **Furthermore, their assimilation can destroy sharp physical features in the analysis!**

The climatology used in the retrieval is a poorer estimate of the atmospheric state *(the weather)* than a short-range forecast!

We do not want to assimilate this information.
Late 1980s: problems - synoptically correlated biases

1DVAR retrievals and the cost function

It can be shown that the maximum likelihood approach to solving the inverse problem requires the minimization of a cost function $J$ which is a combination of two distinct terms (NOTE: The mathematics is the same as 3/4DVAR): 

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y_m - H[x])^T R^{-1} (y_m - H[x])$$

1D state or profile

Radiance vector

RT equation

Fit of the solution to the background estimate of the atmospheric state weighted inversely by the background error covariance $B$. 

Fit of the solution to the measured radiances ($y$) weighted inversely by the measurement error covariance $R$ (observation error + error in observation operator $H$).

***If background and observation errors are Gaussian, unbiased, uncorrelated with each other; all error covariances are correctly specified;
One simple linear form of the 1D-Var solution obtained by minimization of the cost function is given by the expression:

\[ x_a = x_b + [HB]^T [HBH^T + R]^{-1} (y - Hx_b) \]

Correction term, “increment”

The retrieved profile \(x_a\) is equal to the background profile \(x_b\) plus a correction term applied. Furthermore we can quantify the error covariance \(S_a\) of the 1D-Var retrieval which is needed for subsequent assimilation:

\[ S_a = B - [HB]^T [HBH^T + R]^{-1} HB \]

Improvement term

The retrieval being an improvement over the background information (assuming all parameters are correctly specified).
The magnitude of the improvement over the background clearly depends on a number of parameters, but one crucial factor is the number of channels and shape of the weighting functions implied by the radiative transfer operator $H$. 

**HIRS 19 channels**

**IASI 8461 channels**
Characteristics of 1DVAR retrievals

These have a number of advantages that make them more suitable for NWP assimilation than other retrieval methods:

• The prior information (short-range forecast) is very accurate (more than statistical climatology) which improves retrieval accuracy.

• The prior information contains information about physically important features such as fronts, inversions and the tropopause.

• The error covariance of the prior information and resulting retrieval is better known (crucial for the subsequent assimilation process).

• The 1DVAR may be considered an intermediate step towards the direct assimilation of radiances.

BUT the error characteristics of the 1DVAR retrieval may still be very complicated due to its correlation with the forecast background …

Direct radiance assimilation
But do we really need explicit retrievals for NWP?

4.) Direct radiance assimilation
**Direct assimilation of radiances**

Variational analysis methods such as 3DVAR and 4DVAR allow the direct assimilation of radiance observations (without the need for an explicit retrieval step).

This is because such methods do **NOT** require a linear relationship between the observed quantity and the analysis variables.

The retrieval is essentially incorporated within the main analysis by finding the 3D or 4D state of the atmosphere that minimizes

\[ J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y_m - H(x))^T R^{-1} (y_m - H(x)) \]

In direct radiance assimilation the forecast background still provides the prior information to supplement the radiances, **but it is not used twice** (as would be the case if 1D-Var retrievals were assimilated).
4DVAR data assimilation

Remember it is least squares! The forecast model is providing a curve that we try to fit through the observations, by adjusting the state at the start of the window.
4DVAR data assimilation

![Diagram showing the 4DVAR process]

- **Background**
- **Observation**
- **Background departure**

**Legend**
- **Tb [K]**: Temperature at a specific level
- **FG departure [K]**: Forecast error

**Images**
- Maps showing temperature and forecast error over different times.
Direct assimilation of radiances (II)

By the direct assimilation of radiances we avoid the problem of assimilating retrievals with complicated error structures.

BUT

There are still a number of significant problems that must be handled:

- Specifying the covariance ($B$) of background error statistics.
- Specifying the covariance ($R$) of radiance error statistics.
- Removing biases and ambiguities in the radiances / RT model.

Some of these issues are simplified by the direct assimilation of raw (unprocessed) radiance observations.
Direct assimilation of raw radiances

Further to the move away from retrievals to radiance data, most NWP centres are assimilating **raw radiances** (level-1b/1c).

- Avoid **complicated errors** (random and systematic) introduced by (unnecessary) pre-processing such as cloud clearing, angle (limb) adjustment and surface corrections.

- Avoid having to change (retune) our assimilation system when the **data provider changes the pre-processing**

- Faster **access to data** from new platforms (e.g. new data can be assimilated weeks after launch)

- Allows **consistent treatment of historical data** for re-analysis projects (ERA-40, ERA-interim) and other climate studies
Advantages of 4DVAR (or various flavours of it)

- Better use is made of observations far from the centre of the assimilation time window (particularly important for satellite data).
- The inversion of radiances is constrained by the background and its error covariance, but also by the forecast model’s physics and dynamics.
- Wind information can be retrieved from radiance data through tracing effects:
  - To fit the time and spatial evolution of humidity or ozone signals in the radiance data, the 4DVAR has the choice of creating constituents locally or advecting constituents from other areas. The latter is achieved with wind adjustments.
Wind adjustments from radiances in 4DVAR

- Assimilation of passive tracer information feeds back on wind field in a single analysis cycle. Small adjustments also visible in mean wind field.

Mean Analysis Difference

MET7-WV – CONTROL

Mean Analysis

(CONTROL)

200 hPa
Summary of key concepts

- Satellite data are extremely important in NWP.
- Data assimilation combines observations and a priori information in an optimal way and is analogous to the retrieval inverse problem.
- Passive nadir sounders have the largest impact on NWP forecast skill:
  - Nadir sounders measure radiance (not T,Q or wind).
  - Sounding radiances are broad vertical averages of the temperature profile (defined by the weighting functions).
  - The retrieval of atmospheric temperature from the radiances is ill-posed and all retrieval algorithms use some sort of prior information.
  - Most NWP centres assimilate raw radiances directly due to their simpler error characteristics. 4DVAR is now widely used (or alternative 4 dimensional techniques).
- 4DVAR is a least squares. Don’t be too scared of the maths.
Key concepts (continued)

- NWP misused satellite data for many years – problems identified in the 1980’s. Don’t make the same mistakes!

- Always be aware of the difference between a satellite product/retrieval and the actual measurement.
  - Satellites do not measure … P,T,Q …

- If the retrieval problem is ill-posed (e.g., \( y = x_1 + x_2 \)), then the solution must depend on other information or constraints (a priori).

- How do errors in the a priori affect your product? Could these errors be the dominant error source in your work?