Data Assimilation

Lars Isaksen
ECMWF
(European Centre for Medium-Range Weather Forecasts)

Special acknowledgements to Jean-Noël Thépaut
Further acknowledgements to Gianpaolo Balsamo, Niels Bormann, Carla Cardinali, Dick Dee, Patricia De Rosnay, Stephen English, John Eyre, Mike Fisher, Sean Healy, Andras Horanyi, Marta Janisková, Philippe Lopez, Paul Poli, Mike Rennie, Yannick Trémolet, and others…
NWP definition: Process by which “optimal” initial conditions for numerical forecasts are defined.

- The best analysis (initial conditions) is the analysis that leads to the best forecast
- Makes “quickly” the best out of all information available
Model Forecast (with errors)

Observations (with errors)

Computer (with a lot of CPUs)

People (with a lot of good ideas)

Analysis (with smaller – errors)
The forecast model is a very important part of the data assimilation system.

Most important physical processes in the ECMWF model.
Models and observation operators have become much more realistic and accurate.

ECMWF Fc 20121025 00UTC
Model simulated satellite images

GOES-13 IR10.8 20121025-20121030

model

observation
Accurate radiative transfer models allows comparison of model and observed radiances.
Models and observation operators have become much more realistic and accurate. Continual improvement of ECMWF short-range precipitation forecasts with respect to ground-based radar data. This widens opportunities to assimilate new data.
Methodologies

Over the past decades, operational DA techniques have evolved from:

- Cressman type methods (1960/1970s)

To:

- Hybrid methods exploiting the best of both variational and ensemble worlds

Choice remains application dependent

- Atmosphere, waves, sea-ice, ocean, composition, land...
“Observation – model” values are computed at the observation time at high resolution: 16 km

4D-Var finds the 12-hour forecast that take account of the observations in a dynamically consistent way

Based on a tangent linear and adjoint forecast models, used in the minimization process at lower resolution

80,000,000 model variables (surface pressure, temperature, wind, specific humidity and ozone) are adjusted

ECMWF use a 4D Variational (4D-Var) Data Assimilation method

Around 13,000,000 observations within a 12-hour period are used simultaneously in one global (iterative) estimation problem
Observations and model background have errors. It is important to specify them accurately.

Observed temperature: 8°C
Short-range model background temperature: 10°C
Analysis: \(x°C\)
The Balance Operator ensures height/wind field approx. balance is retained in the extra-tropics.

Wind increments at 300hPa

Wind increments 150 metre above surface

Wind increments obtained from a single surface pressure observation
4D-Var dynamically evolves increments

A useful property of 4D-Var is that it implicitly evolves the analysis increments over the length of the assimilation window (Thépaut et al., 1996) in accordance with the model dynamics.

Temperature analysis increments for a single temperature observation at the start of the assimilation window:

$$x^a(t) - x^b(t) \approx MBM^T H^T(y-Hx)/(\sigma_b^2 + \sigma_o^2)$$

- $t=+0$ hours
- $t=+3$ hours
- $t=+6$ hours
More methodologies

Overarching considerations include:

• Seamless quantification of uncertainty estimation (analysis to long-range forecast)
• Improved specification of a priori errors
  • Model, background, observations - systematic and random
  • Errors-of-the-day
• Covariance modeling
  • More variables (aerosols, trace gases, clouds)
  • Non gaussianity
  • Higher resolution
• Data Assimilation for a coupled earth system
Ensemble of Data Assimilations (EDA)

Run an ensemble of independent analyses with perturbed observations, model physics and Sea Surface Temperature fields.

25 EDA members plus a control at lower resolution.

Form differences between pairs of analyses (and short-range forecasts).

These differences estimates the statistical characteristics of analysis (and short-range forecast) error.

Yellow shading where the short-range forecast is uncertain → give observations more weight in these regions.
The EDA provides analysis and background uncertainty estimates:

- To improve the initial perturbations in the Ensemble Prediction
- To calculate static and seasonal background error statistics
- To estimate flow-dependent background error covariances in 4D-Var
- To improve QC decisions and improve the use of observations in 4D-Var

Hurricane Fanele, 20 January 2009

EDA based background error variance for Surface pressure

hPa
Why are flow-dependent covariances important?

Vertical correlations

2012-01-01 00Z

2012-02-01 00Z
Why is flow-dependent JB better?

Vertical correlations at (30N,140W) at 850hPa

The 12-day averaging allows JB to cater for flow-regime changes

The vertical correlations between 850hPa and the boundary layer change significantly from 10th of January 2012 to 10th of February 2012.
Land Data Assimilation

- Land surfaces: heterogeneities, range of spatial and time scales controlling the processes, reservoirs and fluxes.

- The Land Data Assimilation Systems (LDAS) make use of:
  - Processes and feedbacks represented with coupled land-atmosphere models (extension to carbon cycle available)
  - Data assimilation schemes, such as nudging, OI, EKF, EnKF, that update models states variables and/or surface parameters for NWP and climate applications
  - Routine Near Real Time observations with high information content about land surface variables (in-situ, SMOS, ASCAT, SMAP, etc.)
Snow in the ECMWF Data Assimilation System

2009  2010  2011  2012  2013  2014  ...

Snow Model
- Liquid Water
- Density
- Albedo
- Fraction

Snow Obs and DA
- Optimum Interpolation
- 4km IMS snow data
- Obs Quality Control
- IMS latency/acquisition

Additional in situ obs
WMO/SnowWatch action
IMS data assimilation
obs error revision

Snow Model & DA
- Multi-layer model
- Snow cover Fract
- BUFR SYNOP
- RT modelling
- Snow COST action

ECMWF Land Data Assimilation System:
https://software.ecmwf.int/wiki/display/LDAS/LDAS+Home
The Global Observing System

Maximizing the impact of new observations:

– Need to plan
– Need to assess
– Need to proof-of-concept
– Need to consolidate:
  • transfer from Research to Operations

Underpinning requirement: Full exploitation of new observations require sustained investments in model and data assimilation developments
Supported by field campaign experiments, Data targeting studies, etc.

**WMO Integrated Global Observing System**

*Courtesy: WMO*

**ECMWF Preview – DTS**
5.5 more instruments per year
Conventional observations used by ECMWF’s analysis

SYNOP/METAR/SHIP:
- MSLP, 10m-wind, 2m-Rel.Hum.

Radiosonde balloons (TEMP):
- Wind, Temperature, Spec. Humidity

Aircraft: Wind, Temperature

DRIBU: MSL Pressure, Wind-10m

PILOT/Profilers: Wind

Note: We only use a limited number of the observed variables; especially over land.
Satellite data sources used by ECMWF’s analysis

**Sounders:** NOAA AMSU-A/B, HIRS, AIRS, IASI, MHS

**Imagers:** SSMI, SSMIS, AMSR-E, TMI

**Scatterometer ocean low-level winds:** ASCAT

**Geostationary+MODIS:** IR and AMV

**GPS radio occultations**

**Ozone**
Observation Impact at ECMWF

Possible zoom on observation type, level, geographical area, etc.
Quality control of observations is very important

Data extraction
- Check out duplicate reports
- Ship tracks check
- Hydrostatic check

Thinning
- Some data is not used to avoid over-sampling and correlated errors
- Departures and flags are still calculated for further assessment

Blacklisting
- Data skipped due to systematic bad performance or due to different considerations (e.g. data being assessed in passive mode)
- Departures and flags available for all data for further assessment

Model/4D-Var dependent QC
- First guess based rejections
- VarQC rejections

Used data \(\rightarrow\) Increments

Analysis
Operational schedule
Delayed Cut Off and Early Delivery suites

02:00
12h 4D-Var, obs 09-21Z

18 UTC analysis
6h 4D-Var
3hFC

21-03Z
04:00

00 UTC analysis (DA)
T1279 10 day forecast
51*639/T399 EPS forecasts

05:00
06:00
06:35

Disseminate

03:30
12h 4D-Var, obs 09-21Z

14:00
12h 4D-Var, obs 21-09Z

06 UTC analysis
6h 4D-Var
3hFC

9-15Z
16:00

12 UTC analysis (DA)
T1279 10 day forecast
51*639/T399 EPS forecasts

17:00
18:00

Disseminate
Disseminate
Disseminate
Methodologies

More specific challenges and opportunities

- Preparing for novel observing systems
- Scalability
- Meso-scale Data Assimilation
- Climate Monitoring Applications
ADM-AEOLUS: A new perspective for wind distribution understanding and data assimilation

An ESA Earth-explorer mission

Doppler wind lidar
Measures Doppler shift (due to wind) of backscattered UV laser light from the atmosphere

Main application is to improve global analyses and forecasts
Profiles of horizontal line-of-sight (HLOS) wind components

Launch expected mid 2016

More wind profiles would greatly benefit the Global Observing System

Source: ESA/AOES Medialab

Courtesy: ESA
Aeolus measurement principle

12 hrs coverage, ~92K winds

Example of Aeolus vertical sampling
DA has to remain efficient on massively parallel computers

Long window weak-constraint 4D-Var (saddle point algorithm)

4D-en-Var (no TL/AD needed, ensembles run in parallel, I/O costs to be managed)

EnKF (“embarrassingly parallel”) and various hybrid EDA/VAR methods

Pre- and Post-processing of big data are part of the story!
Scalability of T1279 Forecast and 4D-Var

Speed-up

Operations
48 Nodes

10-day Forecast
4D-Var
Ideal

User Threads on IBM Power6

2000 3000 4000 5000 6000
Challenges with Meso-scale Data Assimilation

**General**
- Quick evolving processes
- Rapid updates requires (hourly or sub-hourly)
- Uncertainties and predictability

**Remote sensing observations**
- More timely use of information from GEO satellites
- Novel observations for convective scale DA
- Assimilate cloud-affected radiances
- Non-linear observation operators
- Accuracy and efficiency of radiative transfer in all-sky

**Covariance modeling**
- Traditional balance (e.g. geostrophic & hydrostatic) not applicable at high-resolution
- Impact on ensemble size
- Complex, non-linear, flow-dependent relationship between model variables
- Significant model error (in phase and amplitude)

Source: Thibaut Montmerle, Météo-France
Novel observations for convective-sale DA

Source: Tom Hamill and WMO DA symposium 2013

Radar reflectivities:

Meteosat Cloud Top Height:

TAMDAR Observations:
T, wind, RH, icing, turbulence, etc

Solar PV cells

High resolution water vapour from lidar:
Positioning Data Assimilation at the heart of weather modeling
Where are we today?

Short range forecast skill: We are still progressing

500hPa geopotential height forecast skill
Lead time of anomaly correlation reaching 99%

This is also the case for medium-range forecast skill!
Conclusions

• Prospects of reducing further initial condition errors are great (improved models and observations)
• Data assimilation is the natural vehicle to confront models and observations, and contribute to a seamless quantification of uncertainty estimation
• Full exploitation of the GOS needs:
  • Careful planning and coordination with data providers
  • Sustained investment in model and DA developments
• We expect a lot of NWP impact from the Aeolus DWL
• Efficiency on future HPCs will be a fundamental driver
• Specific challenges and opportunities for meso-scale data assimilation
Thank You!