Recent Developments in Data Assimilation

Florence Rabier & Jean-Noël Thépaut
October 2011
Introduction

- This presentation is a digest of two major events this year:
  - THORPEX DAOS
  - ECMWF annual seminar

- We will concentrate on:
  - Progress in data assimilation methods
  - Observation Usage and Impact
Data assimilation for atmosphere and ocean

- Data assimilation methods
- Observation related aspects
- Real data assimilation systems
- Efficient use of computer architectures

www.ecmwf.int
Historical Background:
What has been important for getting the best NWP forecast?  *(over last 3 decades)*

NWP systems are improving by 1 day of predictive skill per decade. This has been due to:

1. **Model improvements, especially resolution.**

2. **Careful use of forecast & observations, allowing for their information content and errors.** Achieved by variational assimilation e.g. of satellite radiances. *(Simmons & Hollingsworth 2002)*

3. **Advanced assimilation using forecast model: 4D-Var**

4. **Better observations.**
Anomaly correlation of 500hPa height forecasts

- Northern hemisphere
- Southern hemisphere

12% per decade

5% per decade
Statistical, incremental 4D-Var

PF model evolves any simplified perturbation, and hence covariance of PDF

Simplified Gaussian PDF t0

Full model evolves mean of PDF

optionally augmented by a model error correction term.
Background error (prior) covariance $B$ modelling assumptions

The first operational 3D multivariate statistical analysis method (Lorenc 1981) made the following assumptions about the $B$ which characterizes background errors, all of which are wrong!

- Stationary – time & flow invariant
- Balanced – predefined multivariate relationships exist
- Homogeneous – same everywhere
- Isotropic – same in all directions
- 3D separable – horizontal correlation independent of vertical levels or structure & vice versa.

Since then many valiant attempts have been made to address them individually, but with limited success because of the errors remaining in the others. The most attractive ways of addressing them all are long-window 4D-Var or hybrid ensemble-VAR.

Andrew Lorenc
# Hybrid Var/EnKF - best of both worlds?

<table>
<thead>
<tr>
<th>Features from EnKF</th>
<th>Features from VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra flow-dependence in $P^b$</td>
<td>Localization done correctly (in model space)</td>
</tr>
<tr>
<td>More flexible treatment of model error (can be treated in ensemble)</td>
<td>Reduction in sampling error in time-lagged covariances (full rank evolution of $P^b$ in assimilation window in 4DVar).</td>
</tr>
<tr>
<td>Automatic initialization of ensemble forecasts, propagation of covariance info from one cycle to the next.</td>
<td>Ease of adding extra constraints to cost function</td>
</tr>
</tbody>
</table>

--: covariance inflation, covariance localization  

--: scalability, static B, maintenance cost
Hybrid methods

**Hybrid method**: Use flow-dependent state error estimates (from an EnKF/EDA system) in the deterministic 3/4D-Var analysis system:

1) Integrate flow-dependent state error covariance information into the “static” variational analysis
2) Keep the full rank representation of $B$ and its implicit evolution that 3/4D-Var provide inside the assimilation window
3) More robust than pure EnKF for limited ensemble sizes and large model errors
4) Allow for flow-dependent QC of observations
Operational hybrid methods

In operational use (or in an advanced testing), there are currently two main approaches to doing an hybrid DA in a variational context:

1. **Alpha control variable** method (Met Office, NCEP/GMAO, CMC)
2. **Ensemble of Data Assimilations** method (Météo-France, ECMWF)
Comparison of 4D-Var/EnKF (Buehner et al 2010)

***Verifying analyses from 4D-Var with Bnmc***

Comparison of 4D-Var/EnKF (Buehner et al 2010)

Conclusion: Combined 4D-Var + EnKF covariances ->~10hrs SH skill

ECMWF
Hybrid methods: $\alpha$ control variable

1. Alpha control variable method (Met Office, NCEP/GMAO)

Conceptually **add a flow-dependent term** to the climatological $B$ matrix:

$$B = \beta_c^2 B_c + \beta_e^2 P_e \circ C_{loc}$$

$B_c$ is the static, climatological covariance

$P_e \circ C_{loc}$ is the localised ensemble covariance

In practice this is done through augmentation of control variable:

$$\delta x = \beta_c B_c^{1/2} v + \beta_e X' \circ \alpha$$

and introducing an additional term in the cost function:

$$J = \frac{1}{2} v^T v + \frac{1}{2} \alpha^T C_{loc}^{-1} \alpha + J_o + J_c$$

from: A. Clayton, MetOffice
2. The **Ensemble of Data Assimilations** (EDA, Raynaud et al., 2010, Isaksen et al. 2010) can be considered a flow-dependent extension of the way the **climatological background error matrix** is estimated (Fisher, 2003).

![Diagram of EDA process]
Scalability – exploiting massively parallel computers

- 4D-Var as usually implemented requires *sequential* running of a reduced resolution linear PF model and its adjoint. It will be difficult to exploit computers with more (but not faster) processors to make 4D-Var run as fast at higher resolution.

- Improved current 4D-Var algorithms *postpone* the problem a few years, but it will probably return, hitting 4D-Var before the high-resolution forecast models.

- 4DCV 4D-Var can be *parallelised* over each CV segment, but is difficult to precondition well.

- Ensemble DA methods run a similar number of model integrations in *parallel*. It is attractive to replace the iterated running of the PF model by precalculated ensemble trajectories: *4D–Ensemble-Var*. Other advantages of VAR can be retained.
Incremental 4D-Ensemble-Var

4D Gaussian PDF

Trajectories of perturbations from ensemble mean

Full model evolves mean of PDF
Localised trajectories define 4D PDF of possible increments

4D analysis is a (localised) linear combination of nonlinear trajectories. It is not itself a trajectory.
Long window weak constraint 4D-Var

Suppose we extend the window by a few hours:

We expect very little change in the analysis for the first sub-window:

Parallelisation over sub-windows

Mike Fisher
Other important aspects

- Diagnostics for specifying observation error covariances in the assimilation
  - Desroziers, Lonnberg & Hollingsworth, etc.
  - Effort in all centres to better characterize structure and amplitude
Other important aspects

- Enhanced diagnostics of assimilation and forecast performance (obs, R, B)
- The invisible world: pre- and post-processing in Data Assimilation

Carla Cardinali

Florence Rabier
Other important aspects: DA in stratosphere

What are the challenges in stratospheric and mesospheric data assimilation?

- Separation of model and observation error biases
  - Add more low-bias obs with vertical structure information (more limb data needed)

- Vertical spreading of information through covariances
  - Are background error covariances appropriately defined in the upper stratosphere given the poor vertical resolution provided by the observing system?
  - Ad hoc measures prevent spurious increments from contaminating mesosphere.

- Lack of wind information in tropics
  - Without clear mass-wind balance, temperature information of limited use.
  - Solution: new obs such as ADM or SWIFT? 4D-var and tracer assimilation?

Saroja Polavarapu
Improving the stratosphere improves 5-day forecasts in the troposphere

On June 22, 2009 Canadian Meteorological Centre implemented operationally a global stratospheric model (0.1 hPa) for medium range weather forecasts.

Polavarapu et al (2011)

A good stratosphere impacts troposphere forecasts as much as 4D-Var

Winter

(75 cases)
Other important aspects: Ocean Data assimilation

Summary

• Ocean DA is diverse and mature
• Many basic challenges still exist:
  - expansion of control vector (B?)
  - tracer assimilation
  - initialization shock & filtering
  - vertical projection of satellite obs
  - covariance models
  - biogeochemical data assimilation
  - model error
  - internal tides
  - quality control & bias correction
  - air-sea coupling at all scales
• Sub-mesoscale and deep ocean are poorly observed (and poorly constrained)

Andy Moore
Other important aspects: Ocean/atmosphere coupled data assimilation

Impact of coupling
Typhoon intensity forecast for Typhoon Morakot

- Too strong in operational GSM (green)
- Coupling weaken the intensity (red)

Keith Haines
Other important aspects

- Regional aspects
  - High resolution data assimilation, hydrometeors

- Challenge of satellite data assimilation

- Assimilation of the hydrological cycle

- Ocean/atmosphere coupled data assimilation

- Nonlinear data assimilation
  - Particle filters, etc.