EWGLAM, 2008

Madrid

Review of Activities in Regional Data Assimilation

Bruce Macpherson

with thanks to Andrew Lorenc, Dale Barker, Stan Benjamin and more…...

Contents

This presentation covers the following areas

- current status and plans in Europe, US, Japan
- scientific issues
  - 4DVAR, EnKF, convective scale,
- forthcoming events

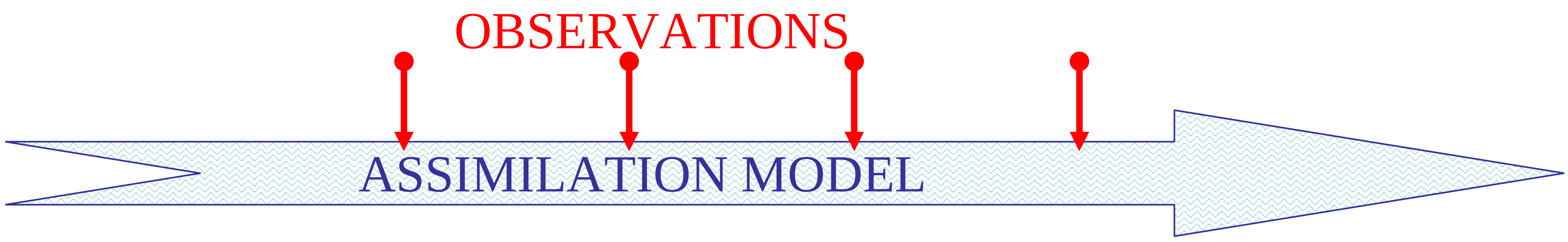
Met Office

Data Assimilation is the process of absorbing and incorporating observed information into a prognostic model.

OED "assimilate, v. t. … II: to absorb and incorporate."

This is normally done by integrating the model forward in time, adding observations.

- The model state summarises in an organised way the information from earlier observations.
- It is modified to incorporate new observations, by combining new & old information in a statistically optimal way.



Data Assimilation

Met Office

1. **Needs a good model:**
    - to carry information from past observations to current time;
    - to diagnose unobserved quantities via physical modelling relationships.
2. **Needs statistical-dynamical combination of information**.
    - Forecasts are generally more informative than latest observations, yet all the information from each observation should be extracted.
    - Observation networks are incomplete. Information on unobserved variables must be inferred (e.g. from satellite radiances).
    - It is impossible to properly sample error distributions – physical insight is needed to give:
        - a good model of observational variances and biases.
        - a good model of the structure of forecast errors.
3. **Advanced Data Assimilation methods also use models to predict the evolution of forecast errors.**

Regional Assimilation techniques operational in SRNWP (Nov 2007)

• None	(21 models)
• Surface only	(3 models)
• Nudging	(6 models)
• 3DVAR	(16 models)
• 3DVAR-FGAT	(4 models)
• 4DVAR	(1 model)
• EnKF	(0 models)

Regional Assimilation techniques in global modelling centres (WGNE)

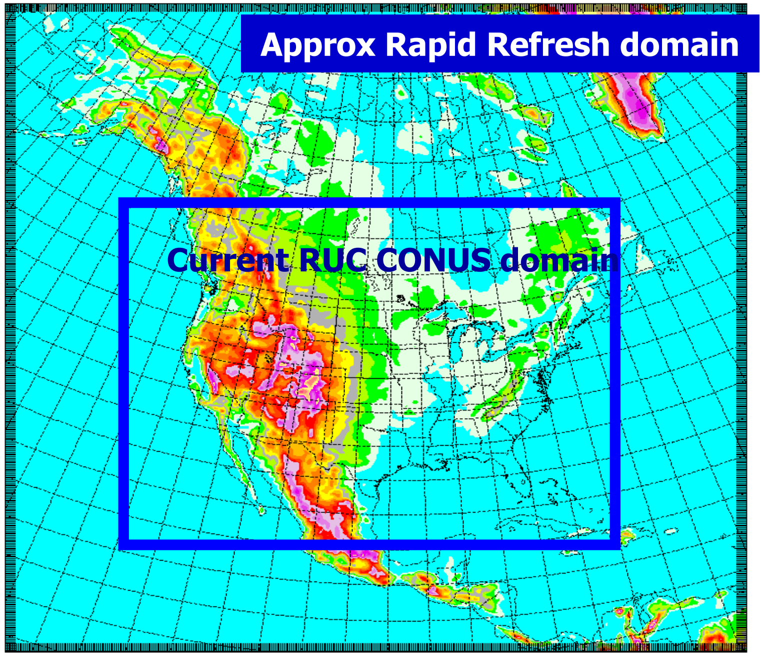
Met Office

Forecast Centre (Country)	2007	2008	2009	2010	2011	2012
ECMWF (Europe)	-	-	-	-	-	-
Met Office (UK)	4D-Var, 36 km 3D-Var, 4 km	4D-Var, 36 km 3D-Var, 4 km	4D-Var, 24 km 3D-Var, 1.5 km	4D-Var, 24 km 3D-Var, 1.5 km	4D-Var, 24 km 4D-Var, 1.5 km?	tbd
Météo France (France)	3D-Var; 9.5 km	3D-Var; 2.5 km	3D-Var; 2.5 km	4D-Var; 2.5 km	4D-Var; 2.5 km	tbd
DWD (Germany)	Nudging; 7 km Nudging; 2.8 km	Nudging; 7 km Nudging; 2.8 km	Ensemble based?	Ensemble based?	Ensemble based?	Ensemble based?
HMC (Russia)				3D-Var	3D-Var	tbd
NCEP (USA)	3D-Var; 12 km	Advanced-Var; 12 km	Advanced-Var; 12 km	Adv or 4D-Var; 8 km	Adv or 4D-Var; 8 km	Adv or 4D-Var; 5 km
Navy/FNMOC/NRL (USA)	3D-Var; 45/15/5 km	3D-Var; 27/9/3 km	3D-Var; 27/9/3 km	3D-Var; 27/9/3 km	4D-Var 9/3/1 km	4D-Var 9/3/1 km
CMC (Canada)	3D-Var; 10, 40 km; L58	4D-Var; 10, 40 km; L58	tbd	tbd	tbd	tbd
CPTEC/INPE (Brazil)	3D-Var; 30 km	LENKF; 20 km	LENKF; 20 km	LENKF; 20 km	LENKF; 10 km	LENKF; 10 km
JMA (Japan)	4D-Var, 10 km	4D-Var, 10 km	4D-Var, 10 km	4D-Var, 10 km	tbd	tbd
CMA (China)	GRAPES-3DVAR, 30 km	GRAPES-4DVAR, 30 km?	GRAPES-4DVAR, 20 km?	GRAPES-4DVAR, 20 km or EnKF?	GRAPES-4DVAR, 15 km or EnKF	GRAPES-4DVAR, 15 km or EnKF
KMA (Korea)	3D-Var; 10, 5 km	3D-Var 10, 5 km	3D-Var 10 km	4D-Var 10 km	4D-Var 10 km	4D-Var 10 km
NCMRWF (India	3D-Var	3D-Var	3D-Var	3D-Var	4D-Var?	tbd
BMRC (Australia)	3D-OI test Met Office 4D-Var (ACCESS)	Met Office 4D-Var (ACCESS)	Met Office 4D-Var (ACCESS)	tbd	tbd	tbd

NCEP Operational Model/Assimilation Systems

RUC – Rapid Update Cycle - http://ruc.noaa.gov

- 3DVAR Assimilation **update every 1h**
- Developed largely at ESRL (GSD)
- Forecasts out to 12h (72h in GSD experimental version)
- Rapid Refresh (RR)- early 2010- NCEP implement, 18h fcsts updated hourly



NAM (North American Mesoscale) –

- **6h update** frequency
- Forecasts out to 84 h

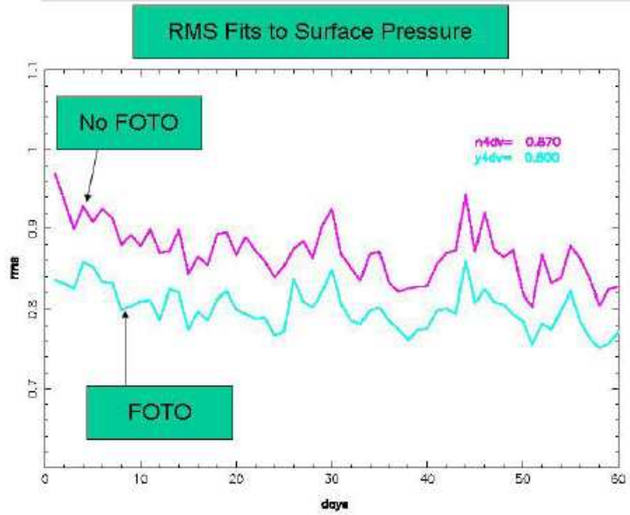
GFS (Global Forecast System) –

- **6h update** frequency
- Forecasts out to 240 h

NCEP Operational Model/Assimilation Systems

2. Advanced Data Assimilation Techniques

- **F**irst **O**rder **T**ime-extrapolation to **O**bservations (**FOTO**)
  - "Simplified 4d-var" technique
  - Will be implemented in next GFS upgrade (April 2008)
- **4d-var (more complex than FOTO)**
  - If resources can be obtained
    - Global and regional system available ~2013 (global) & ~2015 (regional)
  - Global application in time for NPOESS
  - NCEP/EMC collaboration with NASA/GMAO makes this effort possible

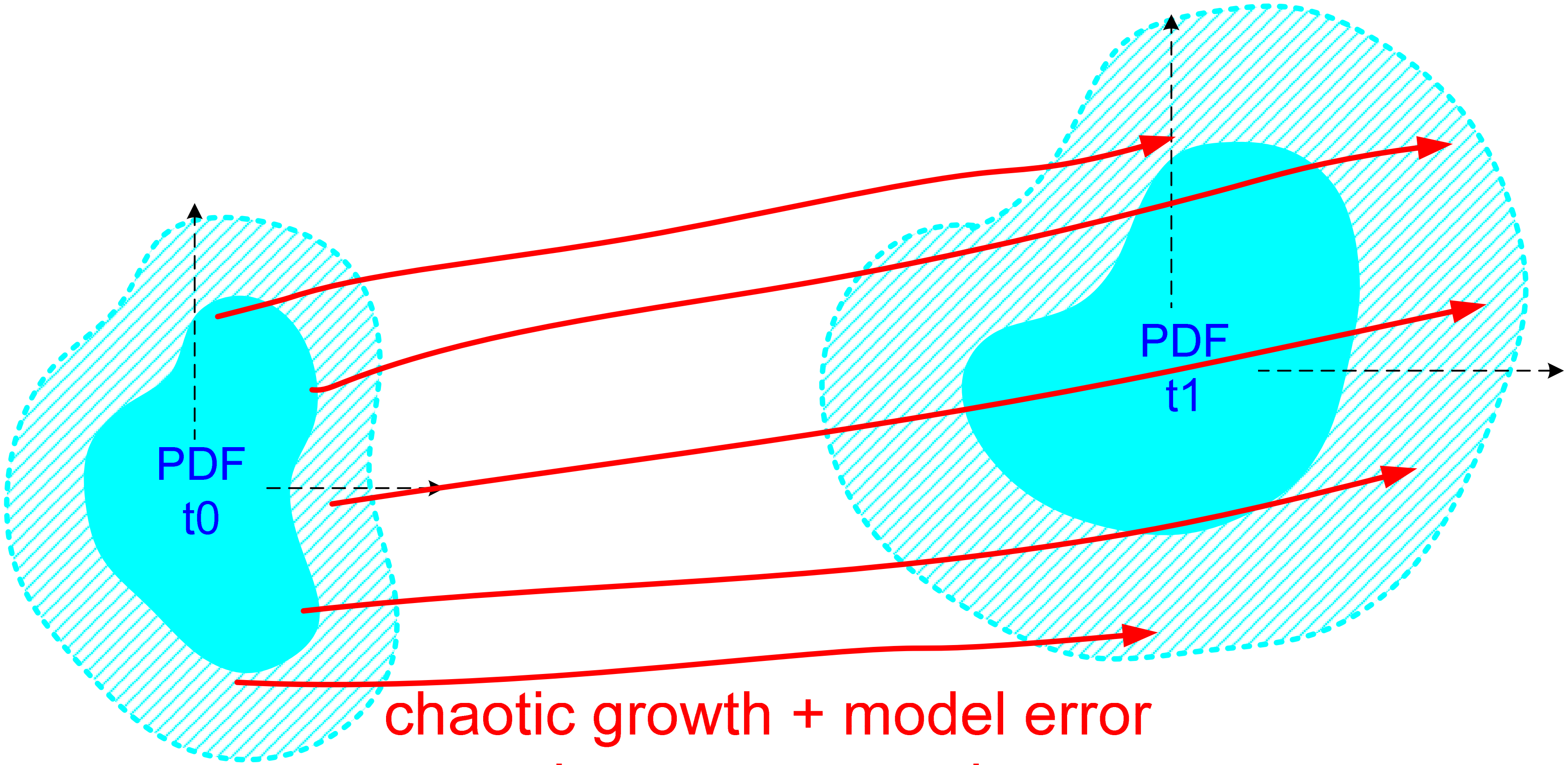


Regional Assimilation – cycle frequency in SRNWP

- 12-hourly (15 models)
    - mostly no DA
- 6-hourly (34 models)
- 3-hourly (2 models)
- 1-hourly (0 models)

NB 'cycle' means frequency at which forecast products are generated – assimilation cycle may be more frequent eg UK4 model

Fokker-Planck Equation



chaotic growth + model error
increase spread

Gaussian Probability Distribution Functions

- Easier to fit to sampled errors.
- Quadratic optimisation problems, with linear solution methods – much more efficient.
- The Kalman filter is optimal for linear models, but
  - it is not affordable for expensive models (despite the "easy" quadratic problem)
  - it is not optimal for nonlinear models.
- Advanced methods based on the Kalman filter can be made affordable:
  - Ensemble Kalman filter (EnKF, ETKF, ...)
  - Four-dimensional variational assimilation (4D-Var)

4D-Var vs Ensemble Kalman Filter

Met Office

Andrew Lorenc 2003:

The potential of the Ensemble Kalman filter for NWP - a comparison with 4D-Var.

Quart. J. Roy. Met. Soc., **129**, 3183-3203.

Eugenia Kalnay et al. 2007:

4-D-Var or ensemble Kalman filter?

Tellus A **59 (5)**, 758–773.

Nils Gustafsson 2007:

comments on above *Tellus A* **59 (5)**, 774–777

Eugenia Kalnay et al. 2007

response to comments! *Tellus A* **59 (5)**, 778–780

WWRP/THORPEX WORKSHOP on 4D-VAR and

ENSEMBLE KALMAN FILTER INTER-COMPARISONS

BUENOS AIRES - ARGENTINA, 10-13 NOVEMBER 2008

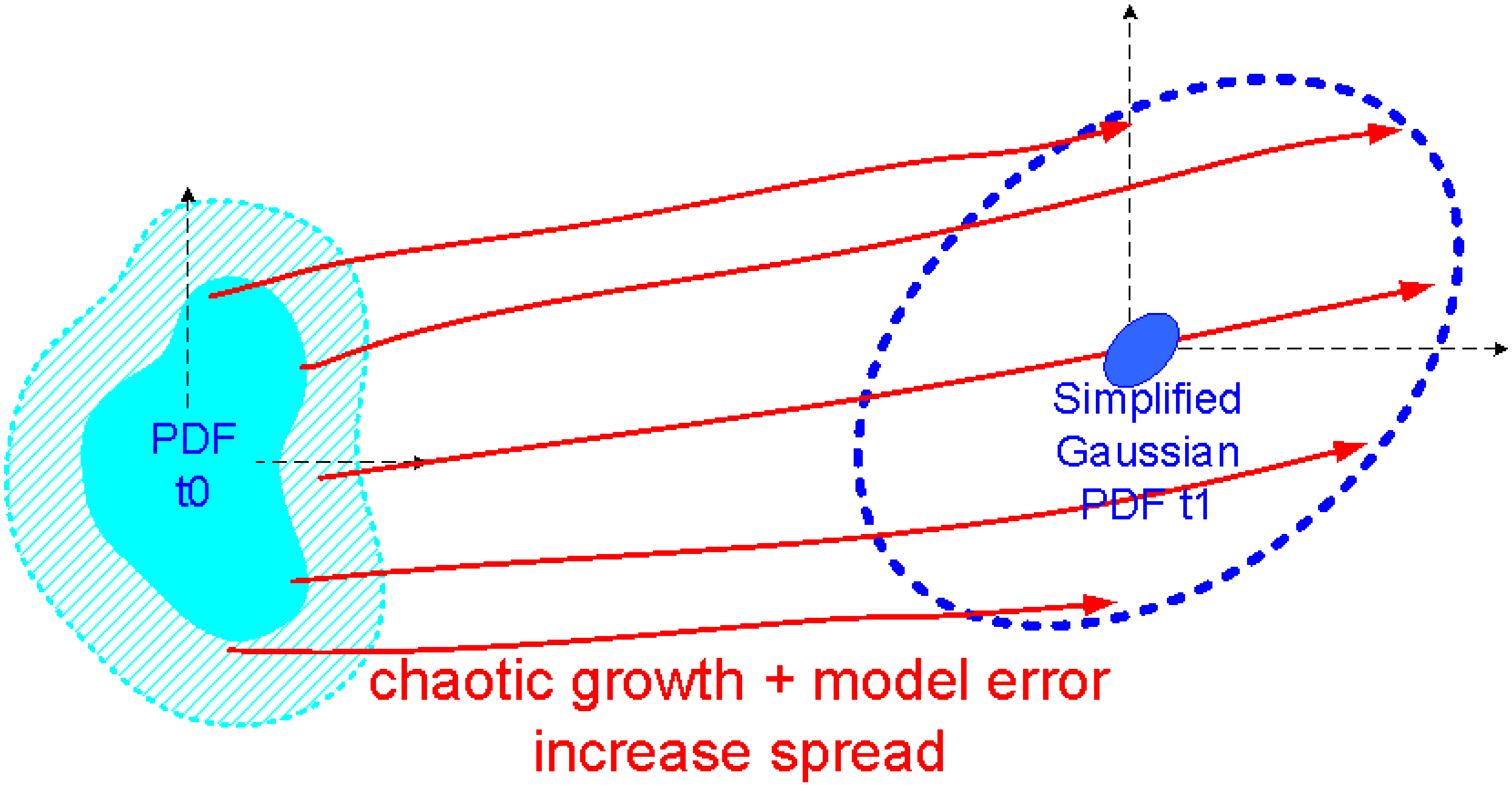
Hosted by the
School of Sciences of the University of Buenos Aires and the
Center for Atmospheric and Ocean Research CIMA/CONICET

ORGANIZING COMMITTEE

Dr. Luc Fillion, Environment Canada, CAN
Dr. Eugenia Kalnay, University of Maryland, USA
Dr. Ronald M. Errico, GEST/NASA, USA
Dr. Fuqing Zhang, Texas A&M University, USA
Dr. Kamal Puri, Bureau of Meteorology Research Centre, Australia

LOCAL ORGANIZING COMMITTEE

Ensemble Kalman filter



Fit Gaussian to forecast ensemble.

The Ensemble Kalman Filter (EnKF)

Construct an ensemble $\{\mathbf{x}_i^f\}, (i = 1, \ldots, N)$:

$$\mathbf{P}^f = \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T},$$

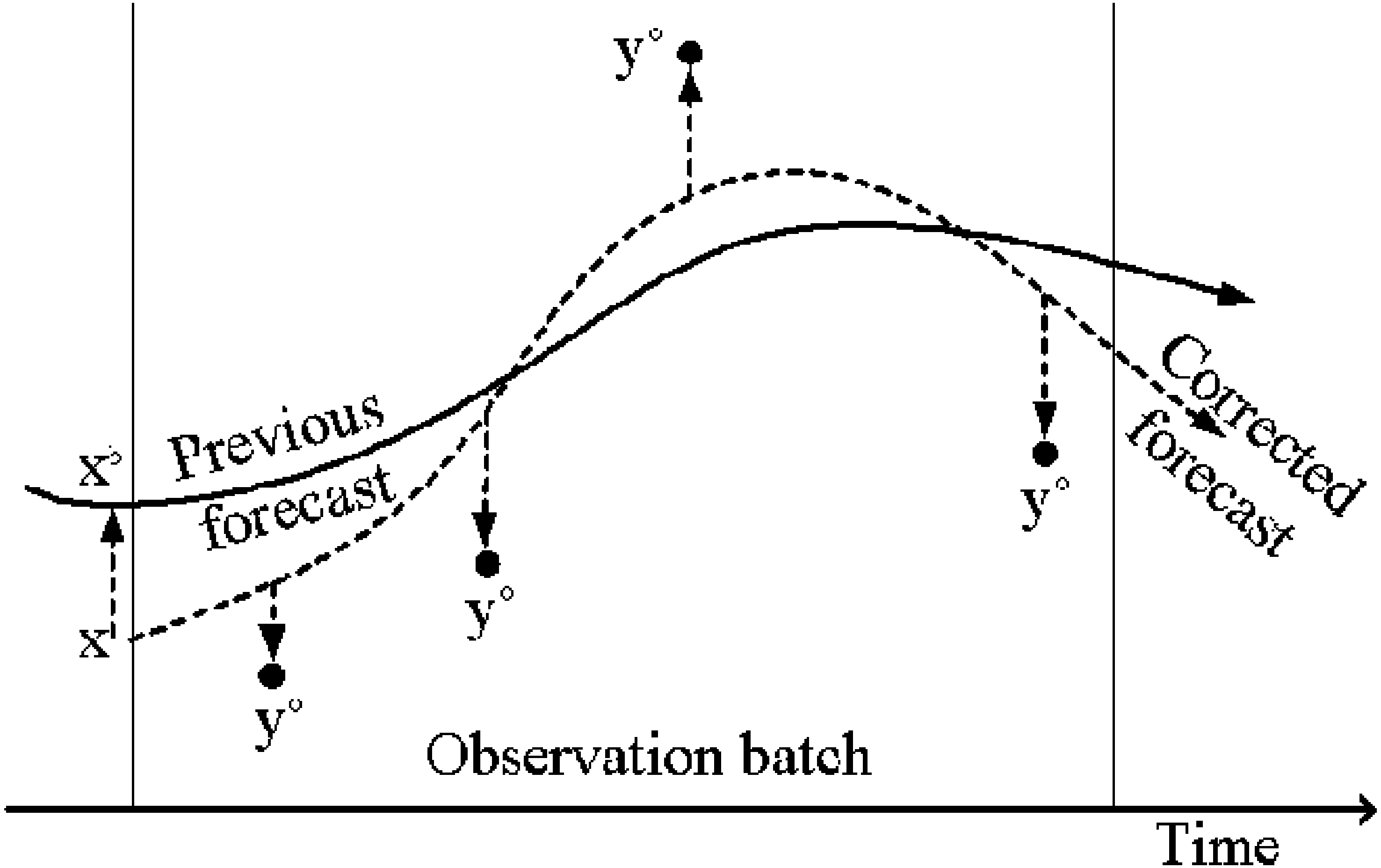
$$\mathbf{P}^f\mathbf{H}^T = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T},$$

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T = \overline{\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T}$$

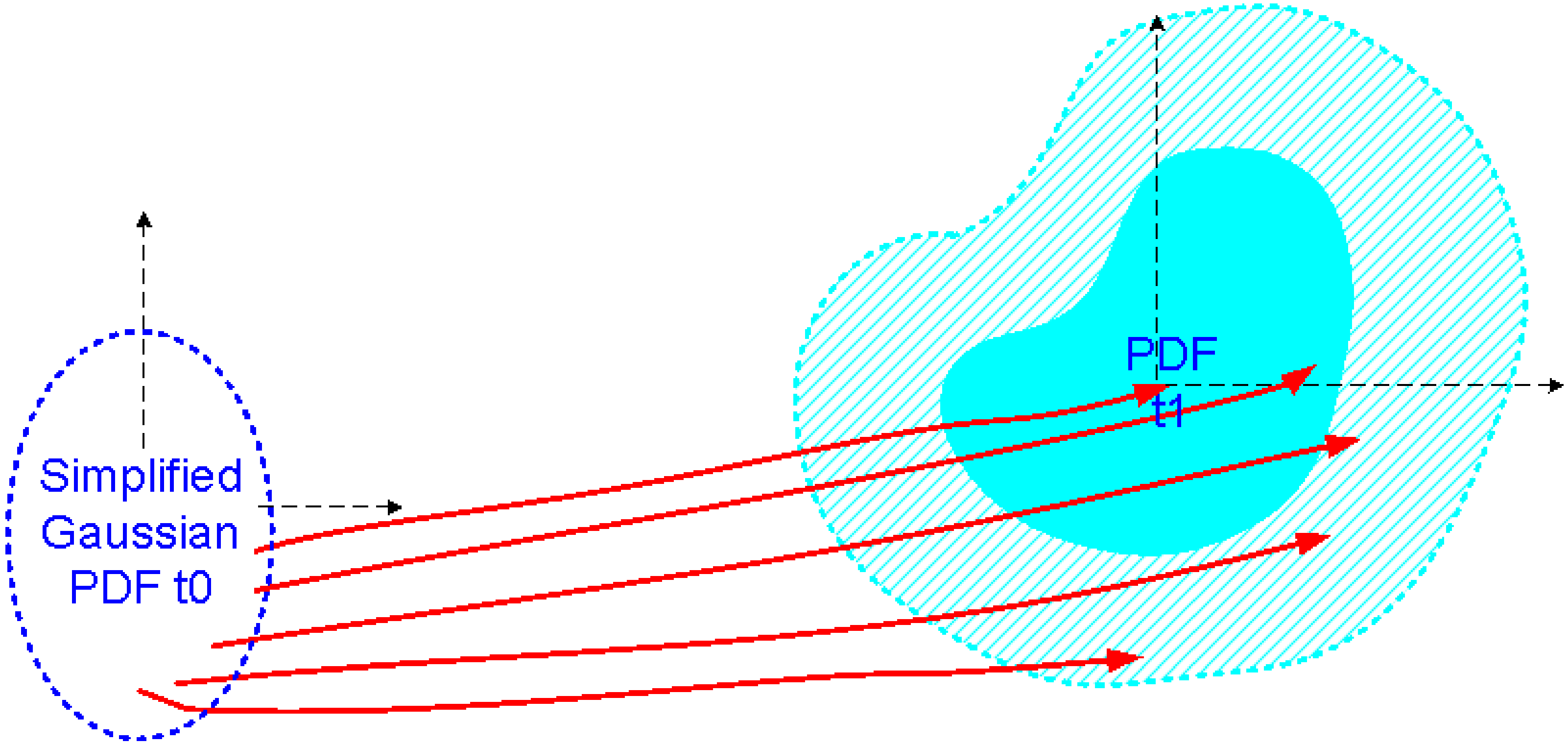
Use these in the standard KF equation to update the best estimate (ensemble mean):

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\overline{\mathbf{x}^f}\right)\right).$$

Simple 4D-Var, as a least-squares best fit of a deterministic model trajectory to observations



Deterministic 4D-Var



Initial PDF is approximated by a Gaussian.

Descent algorithm only explores a small part of the PDF, on the way to a local minimum.

Assumptions in deriving deterministic 4D-Var

Bayes Theorem - posterior PDF: $$P\left(x \middle| y^o\right) = P\left(y^o \middle| x\right) P(x) / P\left(y^o\right)$$

where the obs likelihood function is given by: $$P\left(y^o \middle| x\right) = f\left(y^o - y\right), \quad where \quad y = H(x)$$

Impossible to evaluate the integrals necessary to find "best".

Instead assume best *x* maximises PDF, and minimises -ln(PDF): $$J(x) = -\ln\left[P\left(y^o \middle| x\right)\right] - \ln\left[P(x)\right]$$

(Purser 1984, Lorenc 1986)

The deterministic 4D-Var equations

$$P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) \propto P(\mathbf{x})\,P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right)$$

Bayesian posterior pdf.

Assume Gaussians

$$P(\mathbf{x}) \propto \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right)\right)$$

$$P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right) = P\left(\underline{\mathbf{y}}^o\middle|\underline{\mathbf{y}}\right) \propto \exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right)$$

But nonlinear model makes pdf non-Gaussian:
full pdf is too complicated to be allowed for.

$$\underline{\mathbf{y}} = \underline{H}\left(\underline{M}(\mathbf{x})\right)$$

So seek mode of pdf by finding minimum of penalty function

$$J(\mathbf{x}) = \tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)$$

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \underline{\mathbf{M}}^*\underline{\mathbf{H}}^*\mathbf{R}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)$$

Unique Selling Points, for NWP

4D-Var

- Implicitly uses a complete 4-dimensional PDF, with time-evolution as accurate as perturbation model.

  $\Rightarrow$ Can make good use of time-distributed high-density incomplete observations such as satellite soundings.

- To date, 4D-Var has a better track record in good quality NWP systems

Ensemble Kalman Filter

- Gives the best available sample of background errors, at low cost if short-period ensemble forecasts are needed anyway.
- Easier to build

4D-Var vs Ensemble Kalman Filter

- Is the wrong question!
- What is the best data assimilation for NWP?
- 4D-Var and Ensemble Kalman Filter have different strengths and weaknesses

  $\Rightarrow$ combine them.

Ways to use Ensembles in VAR, reducing sampling error

1. Time average, to give mean covariances (Fisher *ECMWF*)
2. Use smoothed Errors Of The Day variances
3. Use EOTD scales, smoothed locally in wavelet covariances (Pannekoucke *MétéoFrance*)
4. Use EOTD modes, localised using Schur product, in VAR α-control variable (Barker & Lorenc *Met Office*)

Possible future system using Ensemble Perturbations

Met Office

Assume traditional transform or ensemble perturbations both model covariances:

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T \qquad \mathbf{P_e} = \mathbf{V_e}\mathbf{V_e}^T \circ \mathbf{C} \qquad \mathbf{B} \approx \mathbf{P_e}.$$

⇒traditional (**v**) & new ($\mathbf{v}_\alpha$) control variables can both represent most perturbations;

⇒Use both appropriately weighted:

$$\alpha = \mathbf{U}_v \mathbf{U}_h \mathbf{v}_\alpha$$

$$\beta_0^2 + \beta_1^2 = 1 \qquad \mathbf{v_p} = \beta_0 \mathbf{U}_v \mathbf{U}_h \mathbf{v} + \beta_1 \sum_{k=1}^{n} \alpha_k \circ \mathbf{v_{e}}_k$$

$$\mathbf{w}' = \mathbf{U}_p \mathbf{v_p}$$

$$\begin{aligned} J(\mathbf{w}', \alpha) &= J_b + J_e + J_o \\ &= \frac{1}{2}\mathbf{w}'^T \mathbf{B}^{-1} \mathbf{w}' + \frac{1}{2}\alpha^T \mathbf{C}^{-1} \alpha + J_o. \end{aligned}$$

$$J(\mathbf{v}, \mathbf{v}_\alpha) = \frac{1}{2}\mathbf{v}^T \mathbf{v} + \frac{1}{2}\mathbf{v}_\alpha{}^T \mathbf{v}_\alpha + J_o.$$

Benefits Of Hybrid Var/Ensemble DA

Met Office

Benefits for Var:

- Introduces flow-dependent initial PDF in 4D-Var.
- Explicit coupling between moisture/temp/wind fields (high-resolution).
- Easily incorporated in Var framework.
- Relatively cheap (if properly preconditioned).

Benefits over ensemble filters:

- Cost does not scale with observations.
- Can couple with nonlinear QC (serial filter can't do that by itself).
- Hybrids more robust for small ensemble sizes and large model error.

But NWP errors are not Gaussian!

Evidence for non-Gaussianity

Ensemble mean is not a likely state:

Deficient in power in small scales.

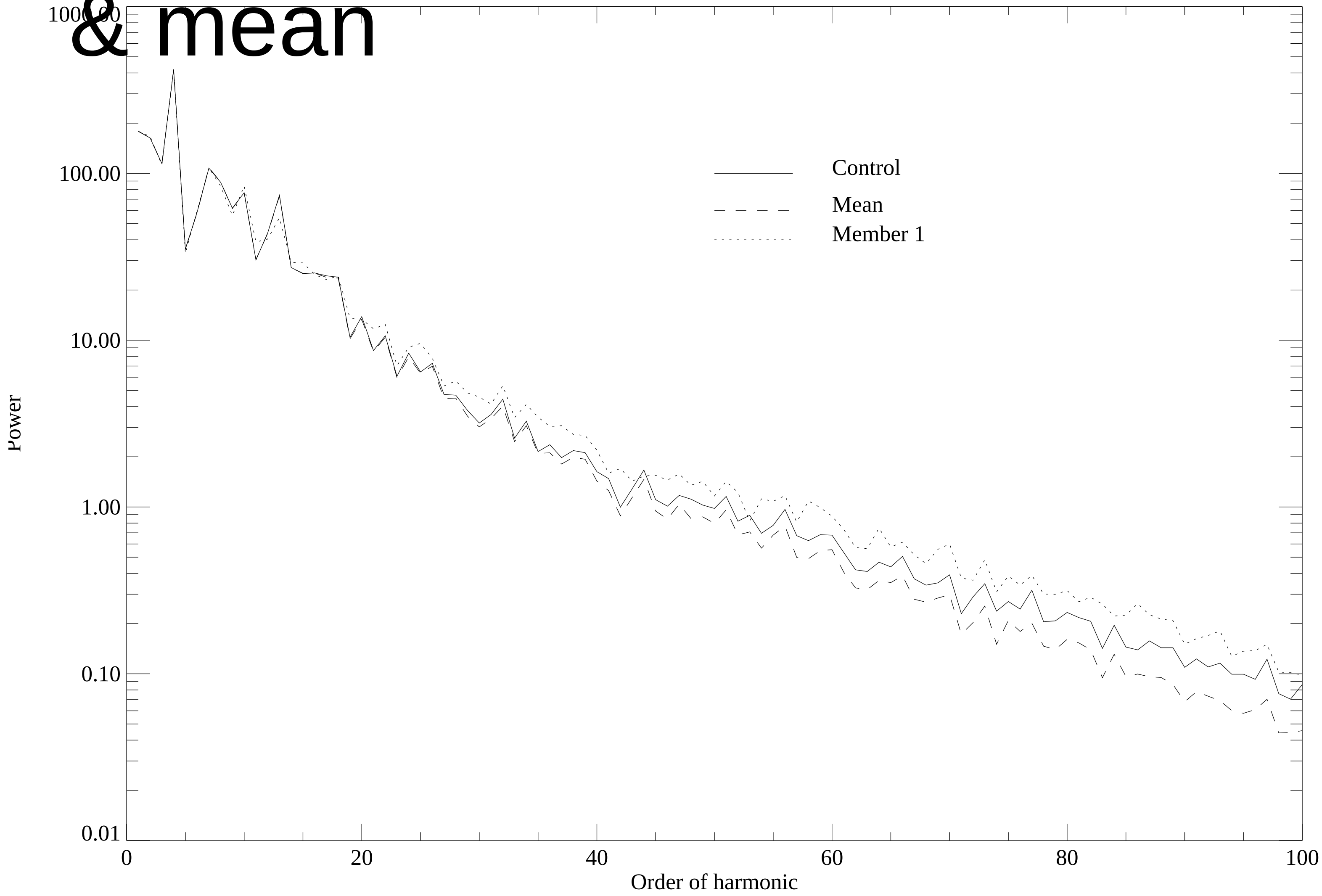
Structures are unrealistically smooth.

Short-period forecasts are deficient in precipitation.

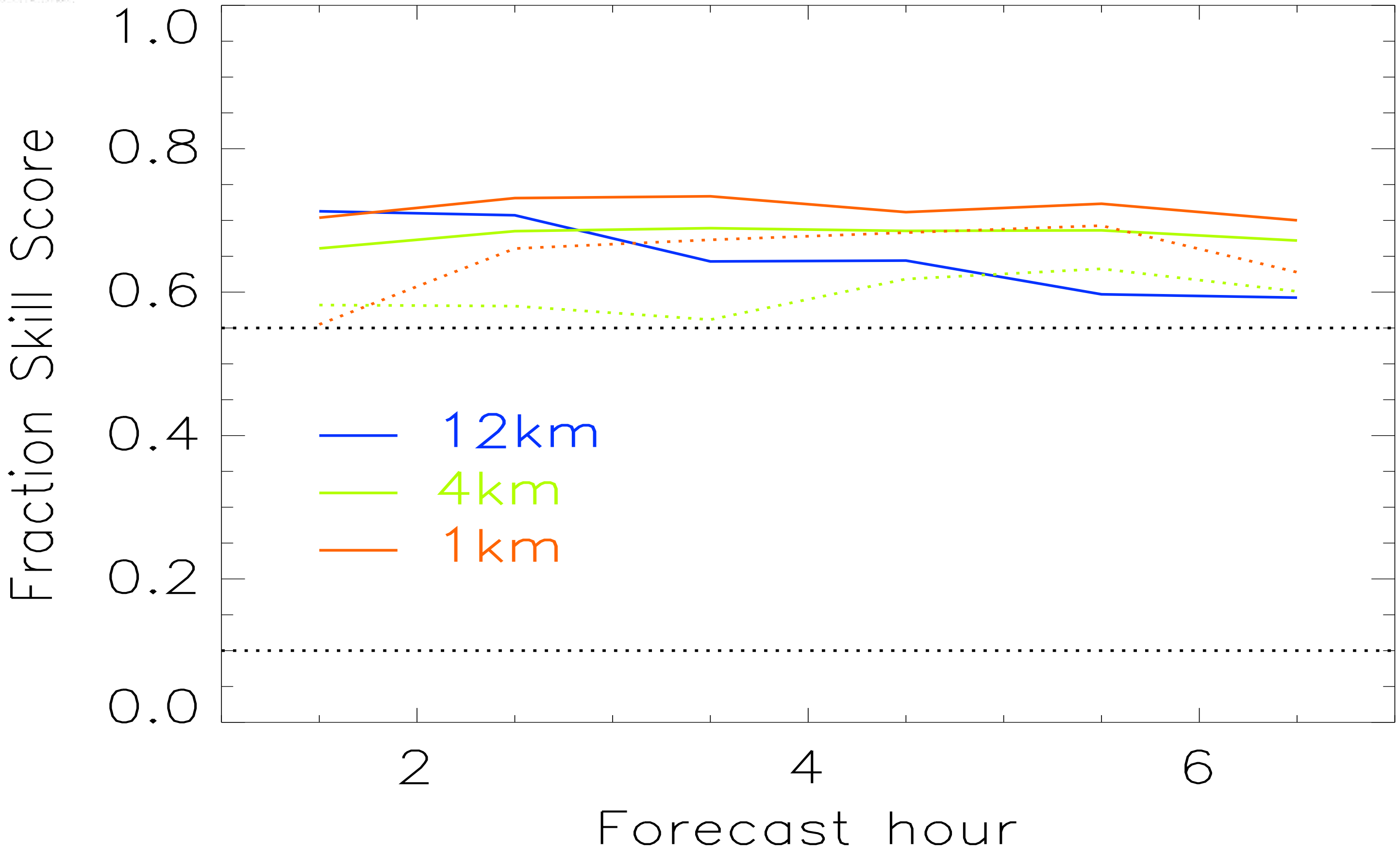
What is the “best” analysis?

Mainly a problem for short-period forecasts, since models spin-up realistic structures.

Power spectrum from MOGREPS 12-hour forecasts & mean



Spin-up of precipitation



Dotted lines show scores of high-resolution forecasts from initial conditions interpolated from 12km model.

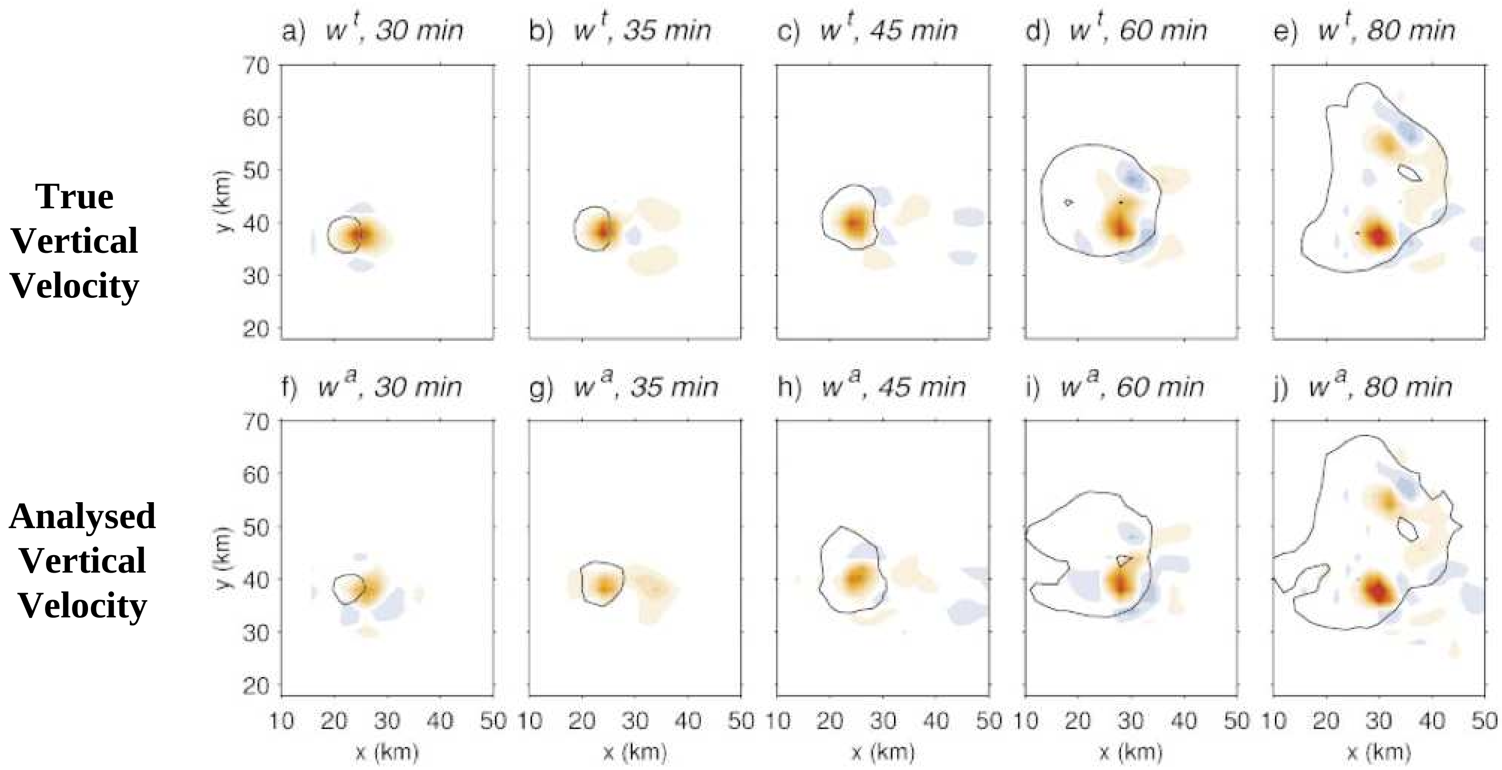
Lack of initial small scales ⇒ deficient precipitation.

Challenges For Convective-Scale DA

- Less experience with high-density observations
  - observation coverage poor
- More nonlinear
- Model errors (e.g. microphysics) large
- Covariances complex, Little diagnostic balance.
- Wide range of scales all significant
  - Lateral boundary conditions important
  - Downscaling of synoptic-scale often useful
  - >2km grid cannot resolve most convection
  - fast phenomena → frequent assimilation
- Expensive model, yet need fast delivery

Convective-Scale EnKF:Snyder and Zhang (2003)

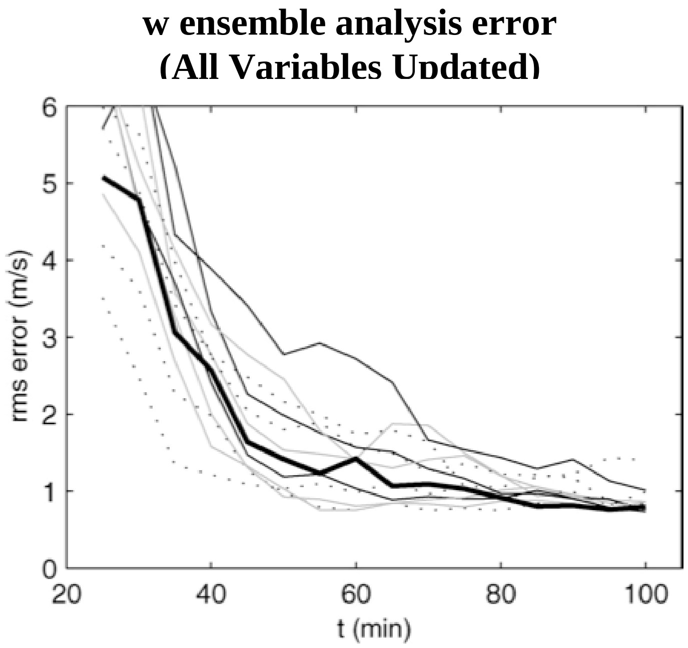
- Uses Sun and Crook (1997) nonhydrostatic model. 2km res.
- Assimilation radar radial velocity every 5 minutes. 50 members.

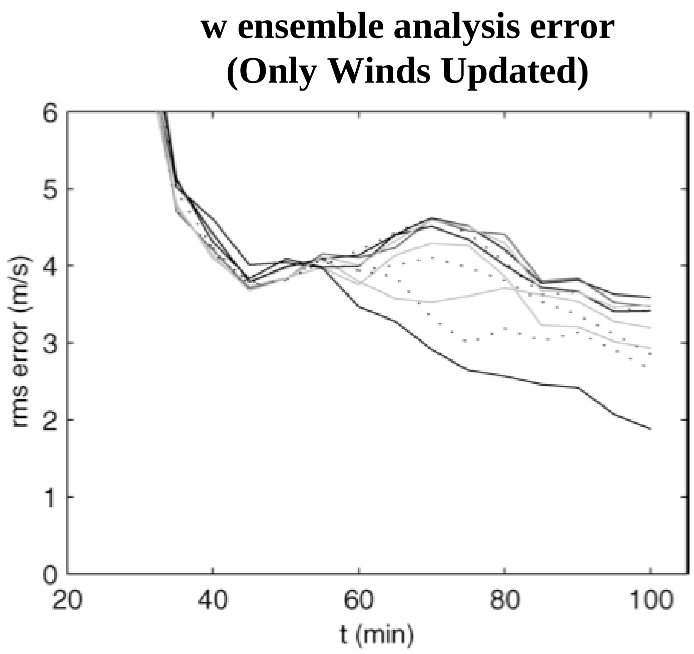


- Beware: OSSE study - perfect model!

Convective-Scale EnKF: Snyder and Zhang (2003)

- Interesting result: Importance of wind-mass covariances.
- Perform EnKF with and without updating unobserved fields:

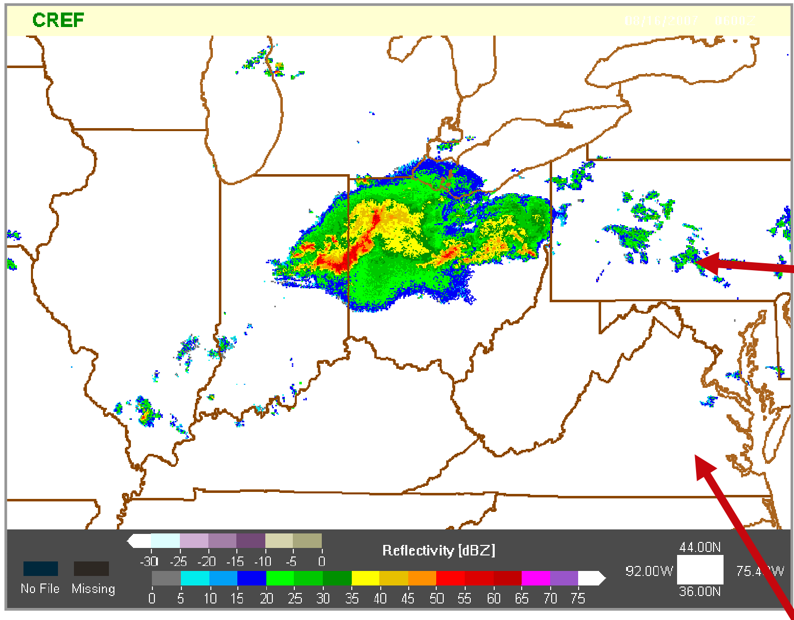




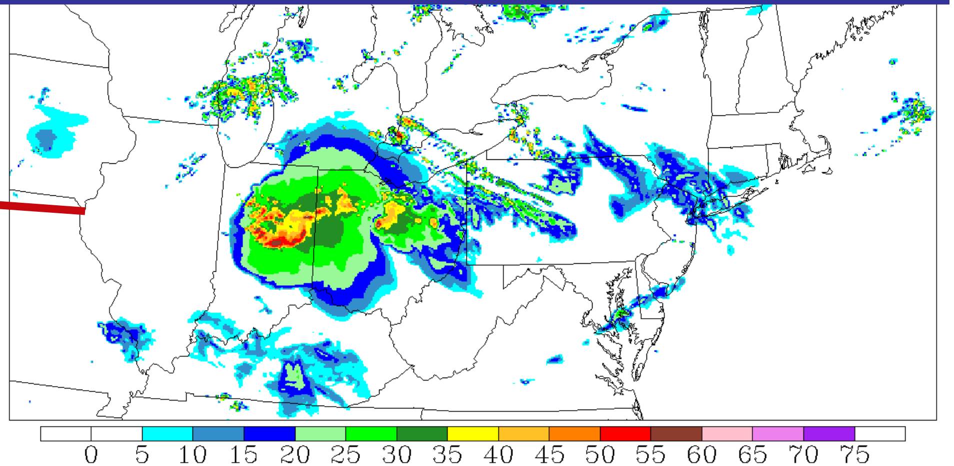
- Shows skill in the flow-dependent ensemble cross-covariances.

3km forecasts from Radar-Enhanced RUC

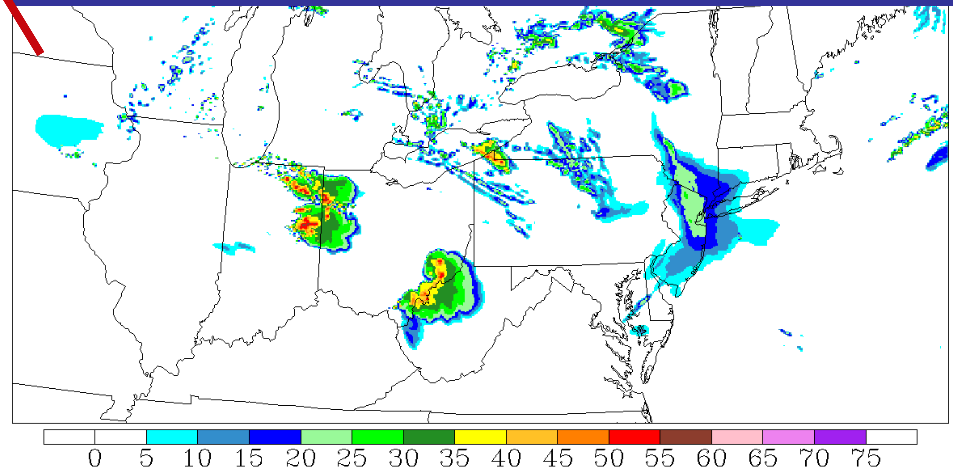
***Observed* radar**



6-h fcst - HRRR 3-km run initialized with radar-enhanced RUC



6-h fcst - HRRR but No-radar init - 3-km run



<u>Radar-enhanced</u> RUC essential for HRRR forecast success

6-h forecasts valid 00z 16 Aug 2007

Koizumi
(JMA)

Improvement of precipitation forecast by the assimilation of analyzed precipitation with 4D-VAR

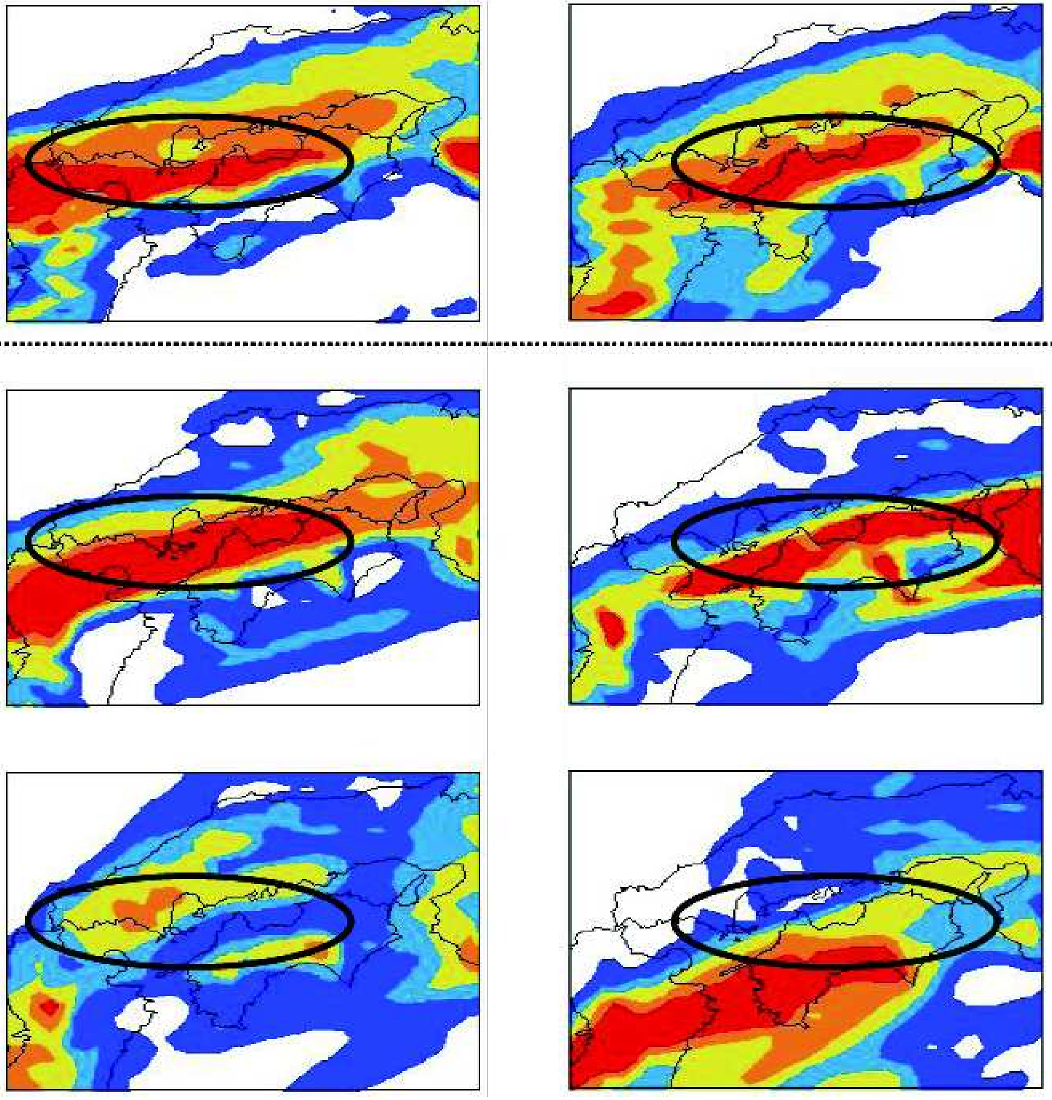
0-3 hour | **3-6 hour**

3 hr. precip. observation

Forecast from 4D-VAR analysis

(Initial Time: 12 UTC, 19 June 2001)

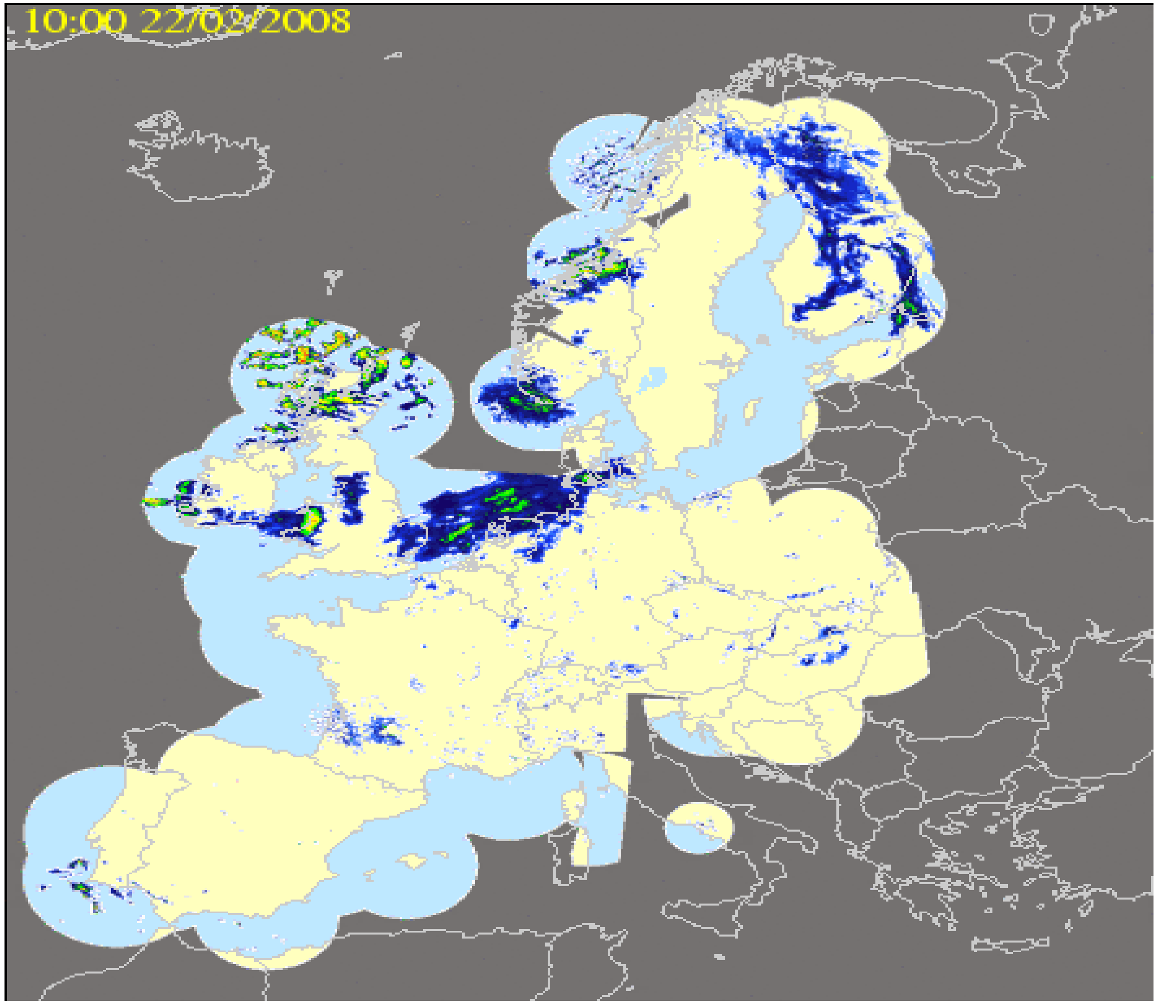
Forecast from OI analysis + PI



OPERA radar composite – pilot assimilation study (ERAD 2008)

- some benefit

- lots of quality issues

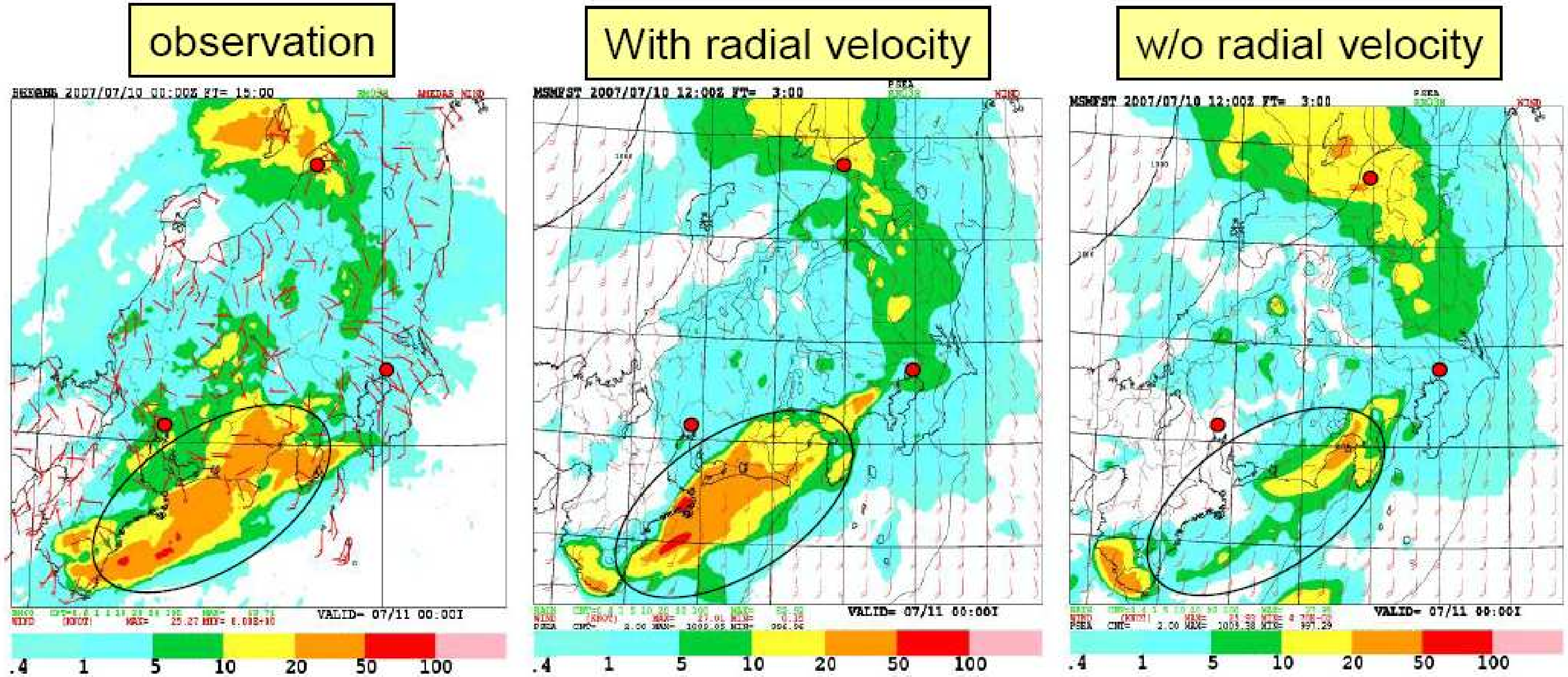


Koizumi (JMA)

Ongoing development

3. use of new observational data

Assimilation of Doppler radar radial velocity has positive impacts on moderate rain forecasts



Future events in DA….

8th International Workshop on Adjoint Model Applications in Dynamic Meteorology

18–22 May 2009,

Chateau Resort and Conference Center

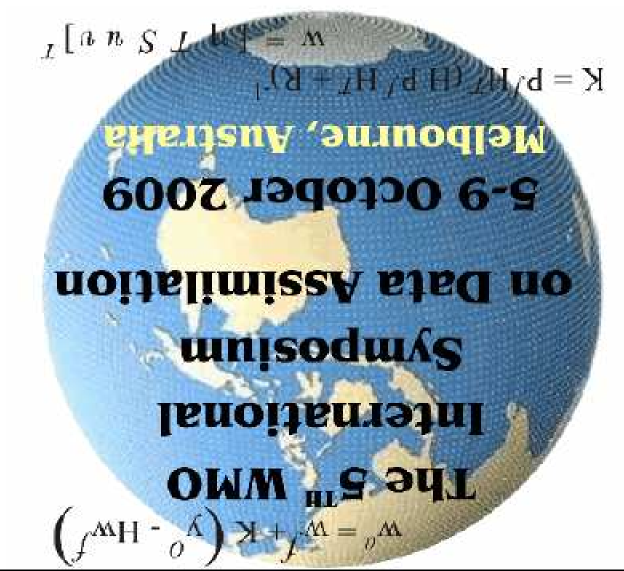
Tannersville, PA, USA

& this week next year….

Melbourne Exhibition Centre: Melbourne Australia

5-9 October 2009

THE 5TH WMO INTERNATIONAL SYMPOSIUM ON DATA ASSIMILATION



& this week next year….

THE 5TH WMO INTERNATIONAL SYMPOSIUM ON DATA ASSIMILATION

5-9 October 2009

Melbourne Exhibition Centre: Melbourne Australia

Part 2:
Review of observation impact studies
(Claude Fischer)