GSI Hybrid/4DEnVar Data Assimilation

Daryl Kleist
NOAA/NWS/NCEP/EMC

2017 Joint DTC-EMC-JCSDA GSI-EnKF Tutorial

NCEP
College Park, MD, July 11-14, 2017

Many people have contributed to the GSI Hybrid/EnVar developments over the past several years. In particular, I would like to acknowledge Dave Parrish (GSI EnVar) and Jeff Whitaker (EnKF) for their significant contributions.

Perspectives of Data Assimilation

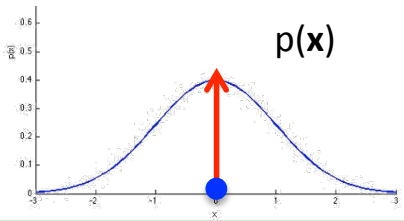
- Two main perspectives of practical data assimilation & hybrid approach

Variational Approach:

<u>Least square estimation</u>

[maximum likelihood]

- 3D-Var (3 dim in space)
- 4D-Var (4th dim is time)



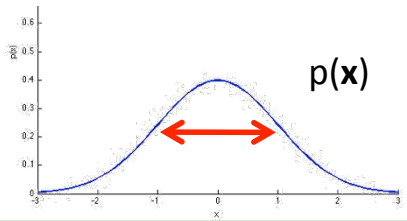
HYBRID

Sequential (KF) Approach:

<u>Minimum Variance estimate</u>

[least uncertainty]

- Optimal Interpolation (OI)
- (Extended / Ensemble)
  Kalman **F**ilter



Courtesy: Kayo Ide

Kalman Filter (Linear)

Forecast Step

$$\begin{cases} \mathbf{x}_k^{\mathrm{b}} = \mathbf{M}_k(\mathbf{x}_{k-1}^{\mathrm{a}}) \\ \mathbf{B}_k = \mathbf{M}_k \mathbf{A}_{k-1} \mathbf{M}_k^{\mathrm{T}} + \mathbf{Q}_k \end{cases}$$

Analysis

$$\begin{cases} \mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{b}} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^{\mathrm{b}}) \\ \mathbf{K}_k = \mathbf{B}_k \mathbf{H}_k^{\mathrm{T}} (\mathbf{R}_k + \mathbf{H}_k \mathbf{B}_k \mathbf{H}_k^{\mathrm{T}})^{-1} \\ \mathbf{A}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{B}_k \end{cases}$$

- Complete set of equations for DA cycling:
  - State ***and*** error covariances are propagated forward in time, and updated with observations at time *k*
  - Under assumptions of linearity (M, H), KF produces optimal set of analysis states
  - Analysis is the minimum variance estimate of the state

Kalman Filter for Large Dimensions

- Kalman filters (and EKF) are impractical for large system like NWP models
  - For present day NWP, the state size (N) can be > $O(10^8)$

- However, a variety of Kalman Filters have been developed for large dimensional systems
  - All of these rely on ***Low-Rank*** Approximations of the background and analysis error covariance matrices

- Assume that $\mathbf{B}_k$ has rank $M \ll N$, so that we can write the error covariance as a function of $\mathbf{X}^b$ ($N x M$), where M can be ~100

$$\boxed{\mathbf{B}_k = \mathbf{X}_k^b (\mathbf{X}_k^b)^T}$$

Ensemble Approach to Represent $p(\mathbf{x})$

Courtesy: Kayo Ide

- Ensemble
  - Members $\mathbf{X} = \{\mathbf{x}^{(m)}\} \quad = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}\}$
  - Spread $\Delta\mathbf{X} = \{\mathbf{x}^{(m)} - \overline{\mathbf{x}}\} = \{\mathbf{x}^{(1)} - \overline{\mathbf{x}}, \ldots, \mathbf{x}^{(M)} - \overline{\mathbf{x}}\}$
  - Mean

$$\overline{x}_n = \frac{1}{M}\sum_{m=1}^{M} x_n^{(m)}$$

$$\overline{\mathbf{x}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{x}^{(m)}$$

  - Covariance

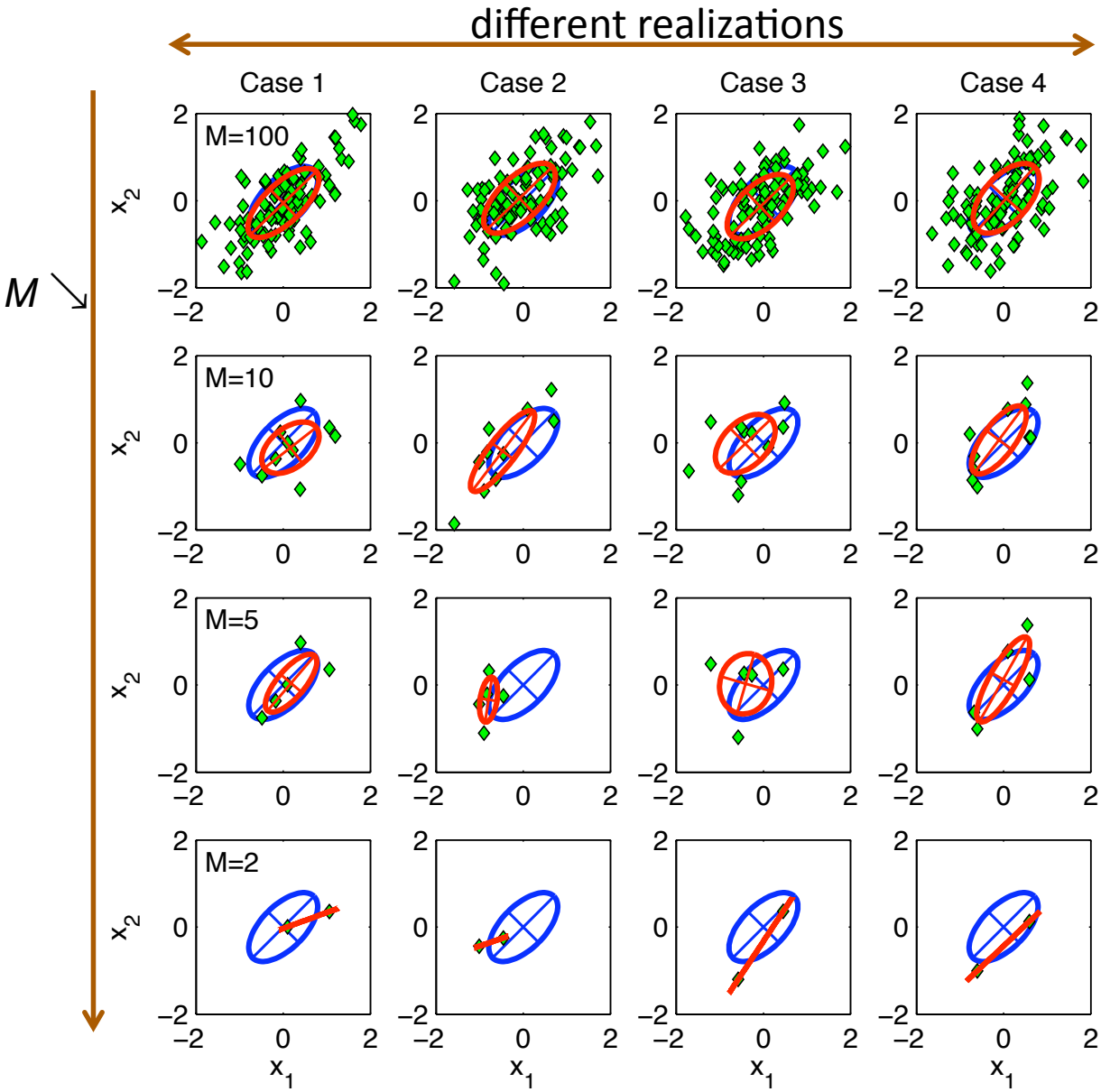
$$P_{nn} = \frac{1}{M-1}\sum_{m=1}^{M}(x_n^{(m)} - \overline{x}_n)^2$$

$$P_{in} = P_{ni} = \frac{1}{M-1}\sum_{m=1}^{M}(x_i^{(m)} - \overline{x}_i)(x_n^{(m)} - \overline{x}_n)$$

$$\mathbf{P} = \frac{1}{M-1}(\Delta\mathbf{X})(\Delta\mathbf{X})^T$$

- Issues
  - Sampling of by ensemble can be poor, especially for
    - Small M
    - Small P_{in}
  - Rank of $\mathbf{P}$ is at most M-1
  - There infinitely many $\Delta\mathbf{X}$ that have the same $\mathbf{P}=(1/M\text{-}1)\Delta\mathbf{X}(\Delta\mathbf{X})^T$

$p(\mathbf{x})$ Sampling & Reconstruction by Ensemble: 2D



Assuming Gaussian pdfs

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\,|\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\mathbf{P}^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right)$$

$$p(\mathbf{x}) \Rightarrow \{\mathbf{x}_m\}$$

— orig. $p(\mathbf{x})$ by $(\bar{\mathbf{x}},\mathbf{P})$

◆ M sample from $p(\mathbf{x})$

— Reconstructed $p^*(\mathbf{x})$ by $(\bar{\mathbf{x}}^*,\mathbf{P}^*)$

$$\{\mathbf{x}_m\} \Rightarrow p^*(\mathbf{x})$$

- If sampling is well-done, then $p^*(x) \sim p(x)$.
- 'Fitness' of $p^*(x)$ to $p^*(x)$ vary case by case particularly for small M.
- All cases, N≤M.

Courtesy: Kayo Ide

Ensemble Kalman Filters
More on these later today….

- Ensemble Kalman Filters (EnKF) are Monte Carlo approximations/implementations, using sample covariances from an ensemble (over bar represents ensemble mean):

$$\overline{\mathbf{x}}_k^{\mathrm{b}} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{x}_{k,m}^{\mathrm{b}}) \qquad \mathbf{B}_k \approx \mathbf{B}_k^{\mathrm{e}} = \mathbf{X}_k^{\mathrm{b}}(\mathbf{X}_k^{\mathrm{b}})^{\mathrm{T}} = \frac{1}{M-1}\sum_{m=1}^{M}(\mathbf{x}_{k,m}^{\mathrm{b}} - \overline{\mathbf{x}}_k^{\mathrm{b}})(\mathbf{x}_{k,m}^{\mathrm{b}} - \overline{\mathbf{x}}_k^{\mathrm{b}})^{\mathrm{T}}$$

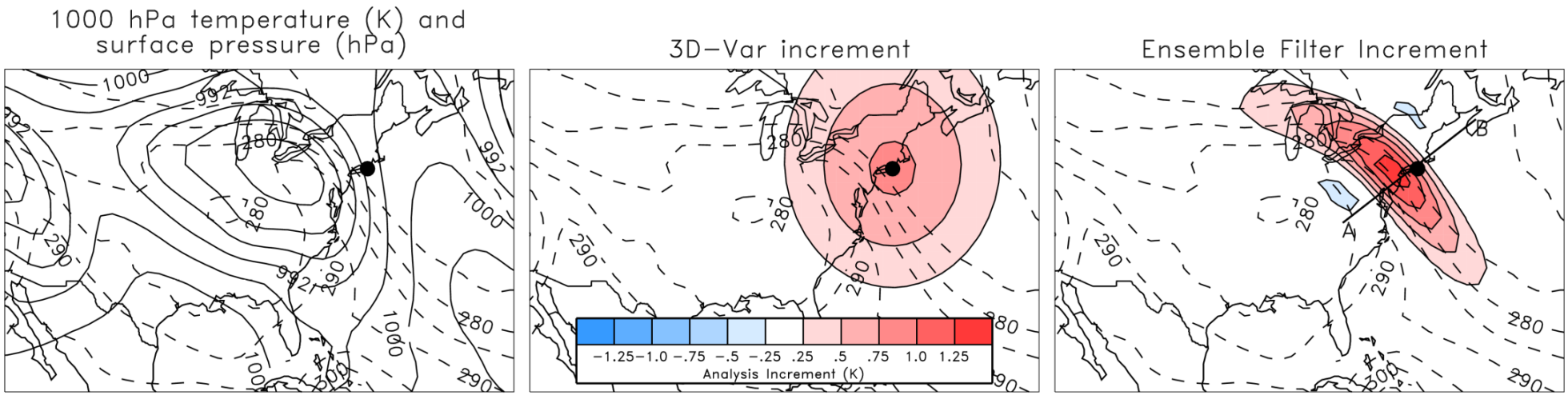
- Where $\mathbf{X}^{\mathrm{b}}{}_k$ is a matrix (NxM) of ensemble forecast perturbations:

$$\mathbf{X}_k^{\mathrm{b}} = \frac{1}{\sqrt{M-1}}\left((\mathbf{x}_{k,1}^{\mathrm{b}} - \overline{\mathbf{x}}_k^{\mathrm{b}})\right),\left((\mathbf{x}_{k,2}^{\mathrm{b}} - \overline{\mathbf{x}}_k^{\mathrm{b}})\right)\ldots,\left((\mathbf{x}_{k,m}^{\mathrm{b}} - \overline{\mathbf{x}}_k^{\mathrm{b}})\right)$$

- And the full $\mathbf{B}^{\mathrm{e}}$ is never explicitly computed! Instead, we represent it in the subspace of the M x M ensemble space.

What does $\mathbf{B}_e$ gain us?
Flow Dependence / Errors of the Day

Temperature observation near warm front

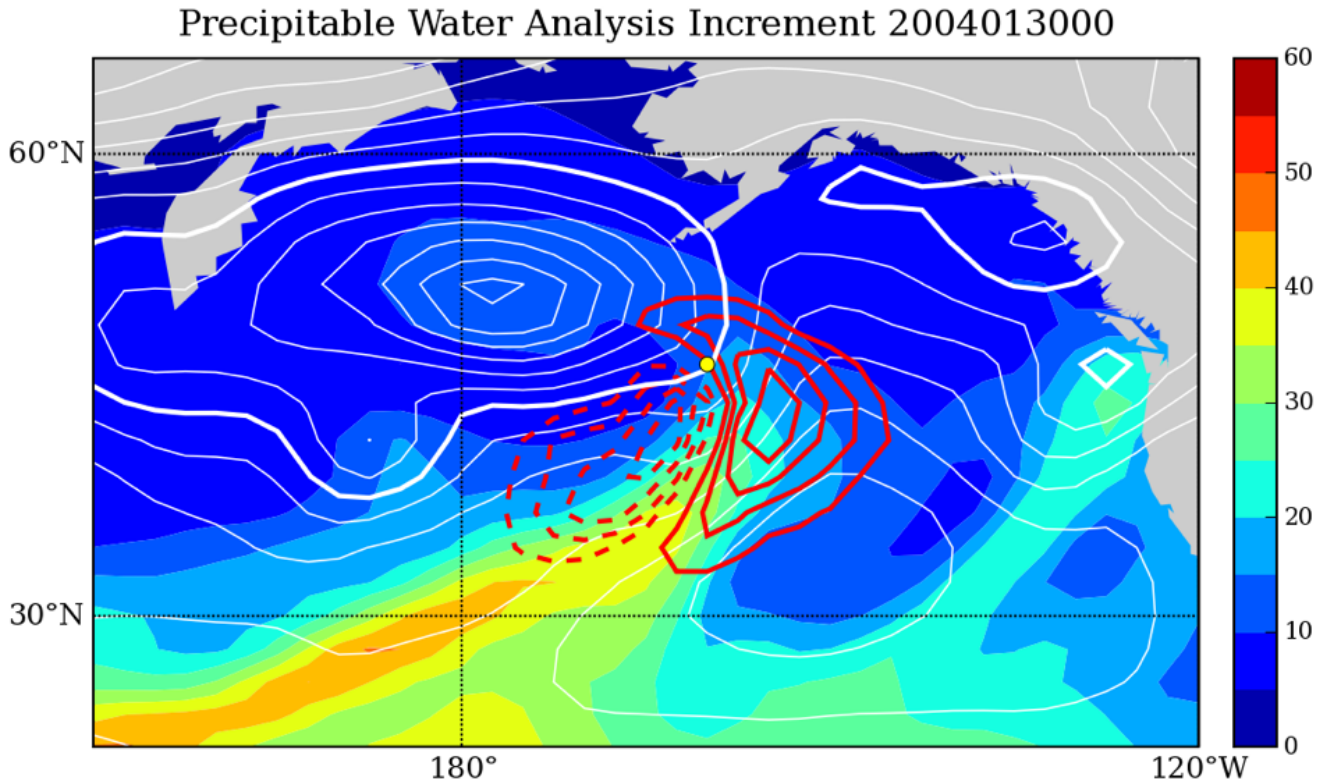


$\mathbf{B}_f$ $\mathbf{B}_e$

Courtesy: Jeff Whitaker

What does $\mathbf{B}_e$ gain us?
Multivariate Correlations

Surface pressure observation near "atmospheric river"



First guess surface pressure (white contours) and precipitable water increment (A-G, red contours) after assimilating a single surface pressure observation (yellow dot) using $\mathbf{B}_e$.

Courtesy: Jeff Whitaker

Use of $\mathbf{B}^e$ in Var (GSI)

- If we had it, we could substitute ensemble estimate of error covariance

$$J_{\text{EnKF}}(\mathbf{x}'_k) = \frac{1}{2}(\mathbf{x}'_k)^{\mathrm{T}}(\mathbf{B}^{\mathrm{e}})^{-1}(\mathbf{x}'_k) + \frac{1}{2}(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'_k)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'_k)$$

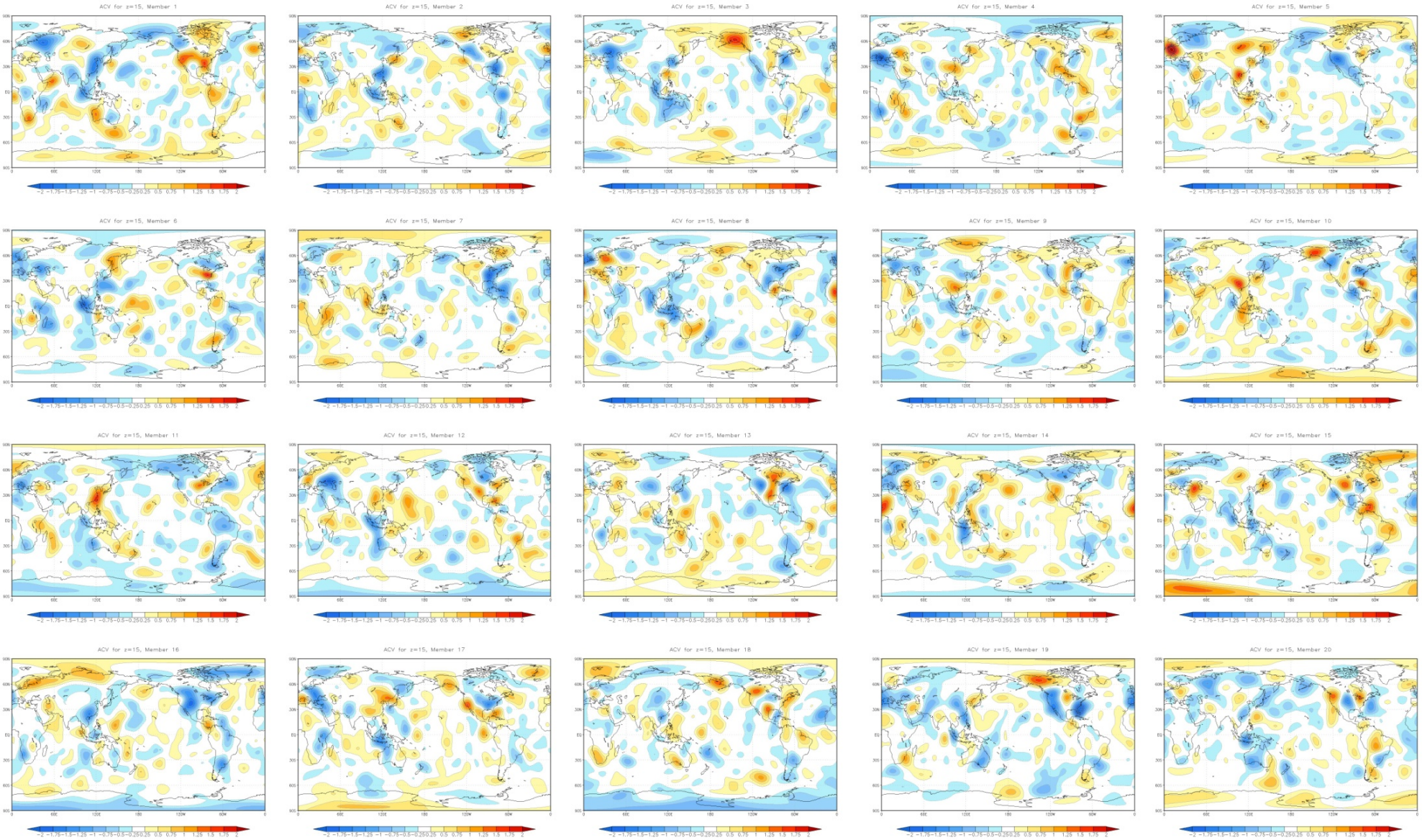
- This is in the full physical space, which we can work around by introducing a new control variable:

$$J_{\text{EnKF}}(\mathbf{x}'_k) = \frac{1}{2}(\alpha)^{\mathrm{T}}\mathbf{L}^{-1}(\alpha) + \frac{1}{2}(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'_k)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'_k)$$

$$\boxed{\mathbf{x}'_k = \sum_{m=1}^{M}\left(\alpha^m \circ \mathbf{x}_{\mathrm{e}}^m\right)}$$

- Where α is the local weight for the ensemble members
- **L** is the ***localization*** on the extended control variable
- $\mathbf{x}_{\mathrm{e}}$ are the ensemble perturbations that represent $\mathbf{B}^{\mathrm{e}}$ (as in EnKF)

Control Variable Example
(2012012212, z=15, 20 members)



Hybrid DA

- Linearly combine full rank (static) and flow-dependent (ensemble) background error covariance estimates

$$\mathbf{B}^{h}=(1-\beta)\mathbf{B}^{e}+\beta\mathbf{B}^{c}$$

- Solution in Var: Add a second background term (one for ensemble, and one for static). Here, we'll denote the climatological (c) and ensemble (e) contributions

$$J_{\text{Hyb}}(\mathbf{x}_{c},\alpha)=\beta_{c}\frac{1}{2}(\mathbf{x}_{c}^{'})^{T}\mathbf{B}^{c}(\mathbf{x}_{c}^{'})+\beta_{e}\frac{1}{2}(\alpha)^{T}\mathbf{L}^{-1}(\alpha)+\frac{1}{2}(\mathbf{y}_{o}^{'}-\mathbf{H}\mathbf{x}_{t}^{'})^{T}\mathbf{R}^{-1}(\mathbf{y}_{o}^{'}-\mathbf{H}\mathbf{x}_{t}^{'})$$

$$\mathbf{x}'_{t}=\mathbf{x}'_{f}+\sum_{m=1}^{M}(\alpha^{m}\circ\mathbf{x}_{e}^{m})$$

Hybrid EnVar (GSI)

$$J_{\text{Hyb}}(\mathbf{x}_c,\alpha)=\boxed{\beta_c\frac{1}{2}(\mathbf{x}_c^{'})^{\mathrm{T}}\mathbf{B}^{\mathbf{c}}(\mathbf{x}_c^{'})}+\boxed{\beta_e\frac{1}{2}(\alpha)^{\mathrm{T}}\mathbf{L}^{-1}(\alpha)}+\underbrace{\frac{1}{2}(\mathbf{y}_o^{'}-\mathbf{H}\mathbf{x}_t^{'})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_o^{'}-\mathbf{H}\mathbf{x}_t^{'})}_{\mathbf{x}'_t=\mathbf{x}'_f+\sum_{m=1}^{M}(\alpha^m\circ\mathbf{x}_e^m)}$$

β_c & β_e: weighting coefficients for clim. (var) and ensemble covariance respectively

$\mathbf{x}_t'$: (total increment) sum of increment from fixed/static **B** ($\mathbf{x}_c'$) and ensemble **B**

$\boldsymbol{\alpha}_k$: extended control variable; $\mathbf{x}_e^m$:ensemble perturbations

- analogous to the weights in the LETKF formulation

$$(\mathbf{w}_{k,m}=(\mathbf{Y}_{k,m}^{\mathrm{b}})^{\mathrm{T}}[\mathbf{Y}_{k,m}^{\mathrm{b}}(\mathbf{Y}_{k,m}^{\mathrm{b}})^{\mathrm{T}}+\mathbf{R}]^{-1}\mathbf{d}_k)$$

L: correlation matrix [effectively the localization of ensemble perturbations]

With Preconditioning

$$\nu^m = \beta_e \mathbf{L}^{-1}\alpha^m \qquad \mathbf{z} = \beta_c \mathbf{B}_c^{-1}\mathbf{x}'_c$$

$$J(\mathbf{z},\nu) = \frac{1}{2}\left(\mathbf{x}'_c\right)^T \mathbf{z} + \frac{1}{2}\alpha^T \nu + J_o$$

$$\mathbf{x}'_c = (\beta_c)^{-1}\mathbf{B}_c \mathbf{z} \qquad \alpha = (\beta_e)^{-1}\mathbf{L}\nu$$

For the double Conjugate Gradient (GSI default), inverses of **B** and **L** not need and the solution is pre-conditioned by full **B**.

This formulation differs from the UKMO and Canadians, who use a square root formulation. Also, the weights can be applied to the increments themselves:

$$\mathbf{x}'_t = \beta_c \mathbf{x}'_c + \beta_e \sum_{m=1}^{M} (\alpha^m \circ \mathbf{x}_e^m)$$

"Hybrid" Methods – Nomenclature

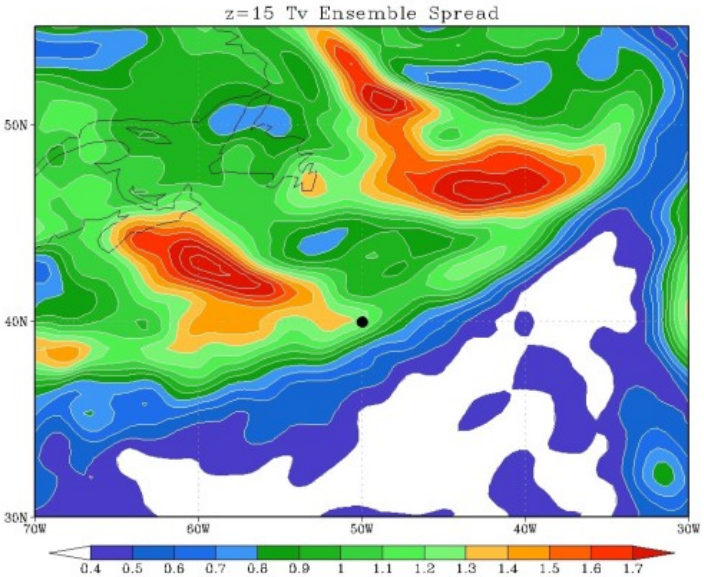
Hybrid: Variational methods that combine **static** and **ensemble** covariances.

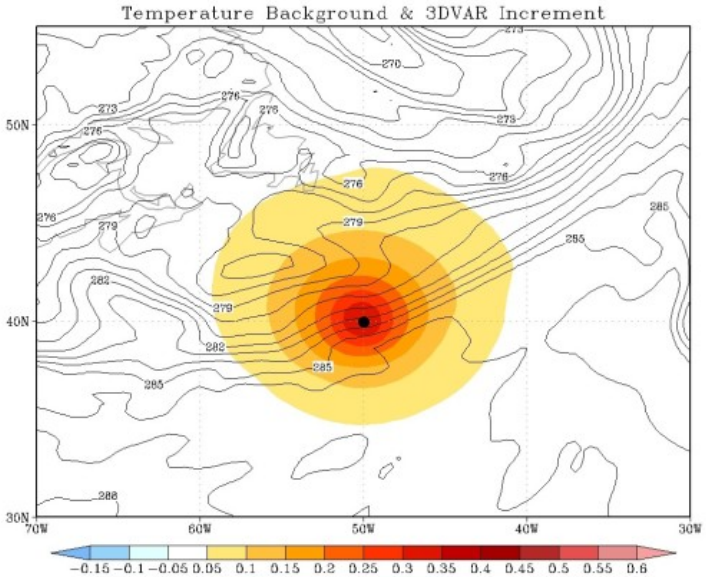
EnVar: Variational methods using ensemble covariances

Hybrid 4DVar: Variational method using a combination of **static** and **ensemble** covariances at the beginning of the window, but using a **tangent-linear** and **adjoint** model

Hybrid 4DEnVar: Variational method using a combination of **static** and **ensemble** covariances at the all times in the window, ***without*** the need of a **tangent-linear** or **adjoint**

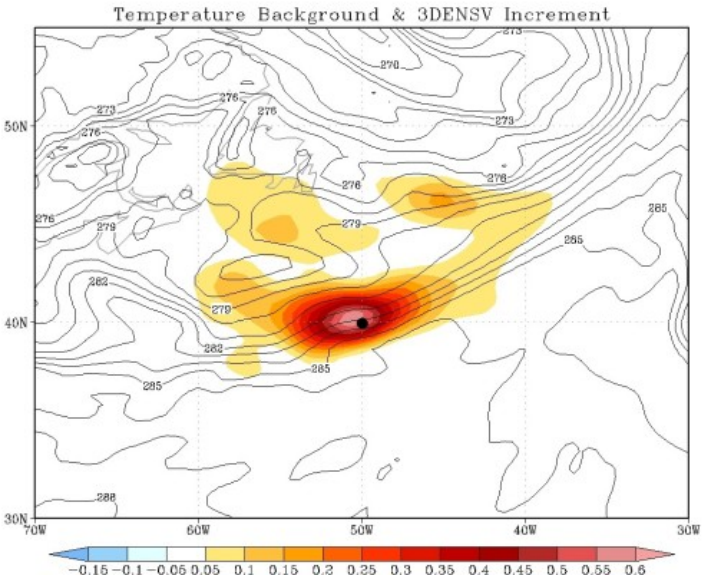
Single Temperature Observation





3DVAR

EnVar



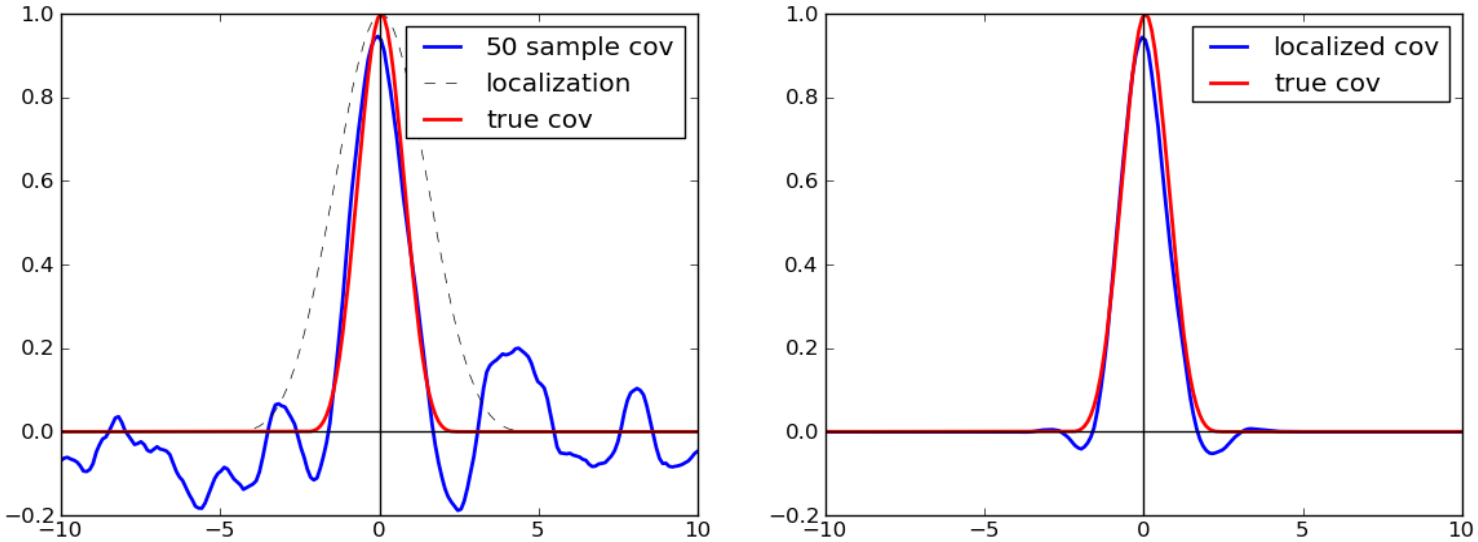
So what's the catch?

- Most configurations of hybrid DA systems require the development and maintenance of two DA systems
  - EnKF + Var

- Still need to deal with ***localization*** and other sampling-related issues (though somewhat mitigated by use of full rank $\mathbf{B}_c$)
  - Rank deficiency: Using small ensemble size to represent something of much larger dimension

- Even more parameters to explore
  - Trade off between ensemble size, resolution, hybrid weights, etc.

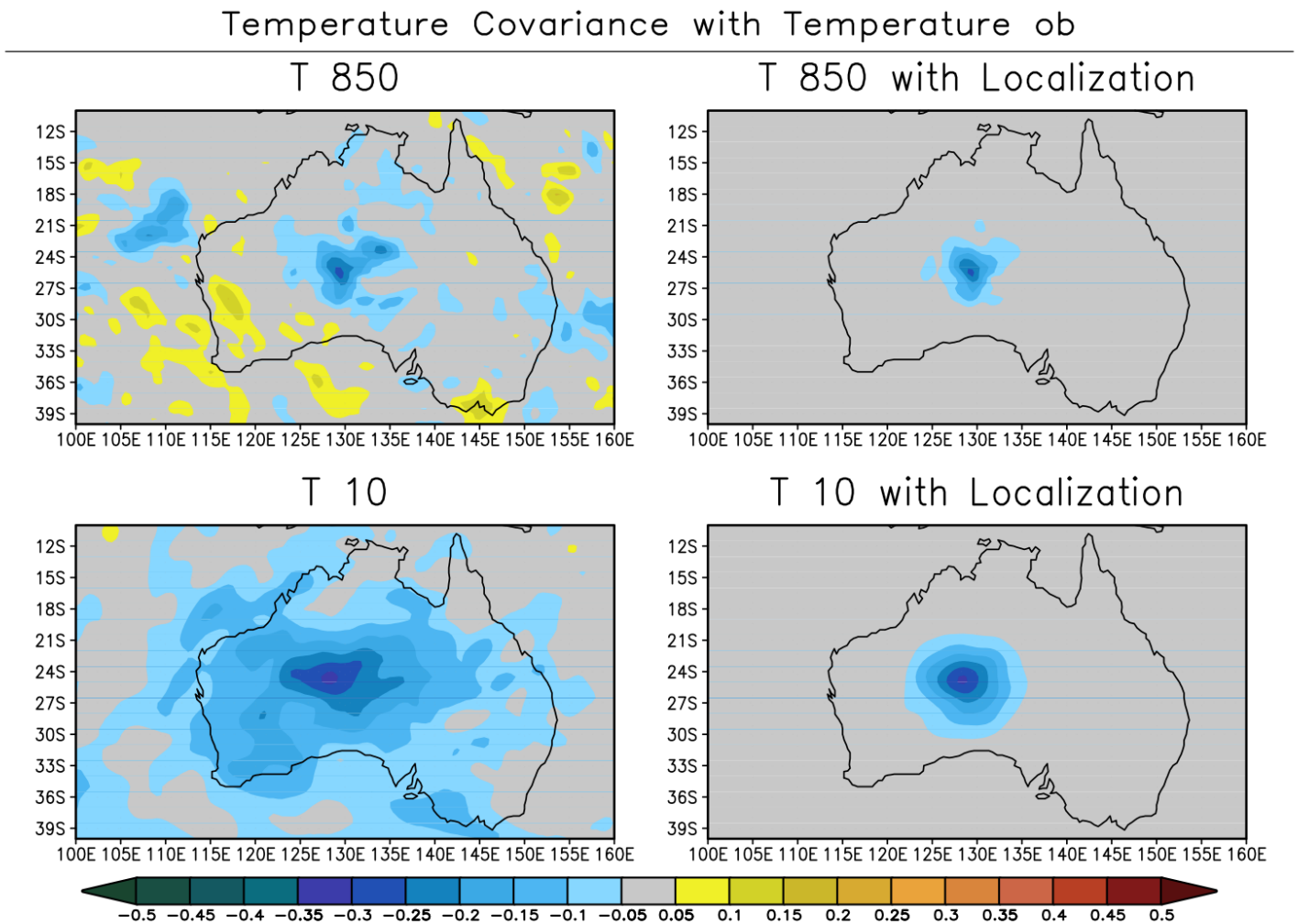
Example of Covariance Localization



Estimates of covariances from a small ensemble will be noisy, with signal to noise small especially when covariance is small

Courtesy: Jeff Whitaker

NWP Localization Example



Courtesy: Jeff Whitaker

GSI Hybrid Configuration

- Hybrid related parameters controlled via GSI namelist **&hybrid_ensemble**
- Logical to turn on/off hybrid ensemble option (**l_hyb_ens**)
- Ensemble size (**n_ens**), resolution (**jcap_ens**, **nlat_ens**, **nlon_ens**)
- Source of ensemble: GFS spectral, native model, etc. (**regional_ensemble_option**)
- Weighting factor for static contribution to increment (**beta_s0**)
  - Option to specify different beta weights as a function of vertical level (**readin_beta**)
- Horizontal and vertical distances for localization, via **L** on augmented control variable (**s_ens_h**, **s_ens_v**)
  - Localization distances are the same for all variables since operating on α. i.e. no variable localization
  - Option to specify different localization distances as a function of vertical level (**readin_localization**)
- Other regional options related to resolution, pseudo ensemble, etc.

Notes on some GSI Hybrid Options

- **beta_s0**
  - If weights set this way, code will assume that ensemble beta will be 1-beta_s0
  - If set via readin_beta, the ensemble + climatological beta weights do not have to sum to one for a given level (error tuning)
- **s_ens_h**
  - Horizontal decorrelation distance in km (Gaussian half-width, not zero-distance)
- **s_ens_v**
  - If > 0, vertical decorrelation given in units of model layers
  - If < 0, vertical decorrelation given in units of d ln p (scale height)
    - i.e. s_ens_v=-0.5 actually means 0.5 in units of dlnp
- **filename**
  - If using global (GFS) ensemble, will generally follow sigfHH_ens_memNNN
    - HH = forecast hour
    - NNN = integer ensemble member number
- **Regional specific**
  - Regional_ensemble_option, grid_ratio_ens, more….

Example: global_hybens_info.l64.txt

64				# of levels			
350.0	-0.5	0.1250	0.8750	s_ens_h(1)	s_ens_v(1)	beta_s0(1)	beta_e(1)
350.0	-0.5	0.1250	0.8750	s_ens_h(2)	s_ens_v(2)	beta_s0(2)	beta_e(2)
350.0	-0.5	0.1250	0.8750	s_ens_h(3)	s_ens_v(3)	beta_s0(3)	beta_e(3)
350.0	-0.5	0.1250	0.8750	s_ens_h(4)	s_ens_v(4)	beta_s0(4)	beta_e(4)
350.0	-0.5	0.1250	0.8750	s_ens_h(5)	s_ens_v(5)	beta_s0(5)	beta_e(5)
350.0	-0.5	0.1250	0.8750	s_ens_h(6)	s_ens_v(6)	beta_s0(6)	beta_e(6)
.				.			
.				.			
.				.			
1300.0	-0.5	0.1250	0.8750	s_ens_h(63)	s_ens_v(63)	beta_s0(63)	beta_e(63)
1300.0	-0.5	0.1250	0.8750	s_ens_h(64)	s_ens_v(64)	beta_s0(64)	beta_e(64)

Hybrid GSI details

- **Localization**:
  - Horizontal: Spectral operator to apply Gaussian function. Localization distances function of model vertical level
    - Option for RF-based horizontal localization (for regional applications)
  - Vertical: Recursive filter (separable from above)

- **Ensemble variables**: Can be different than standard control variable for **B**c (Default u, v, T, ps, q_{wv}, q_{oz}, q_{cw})
  - ensctl2state and state2ensctl does mapping between state space and ensemble variables

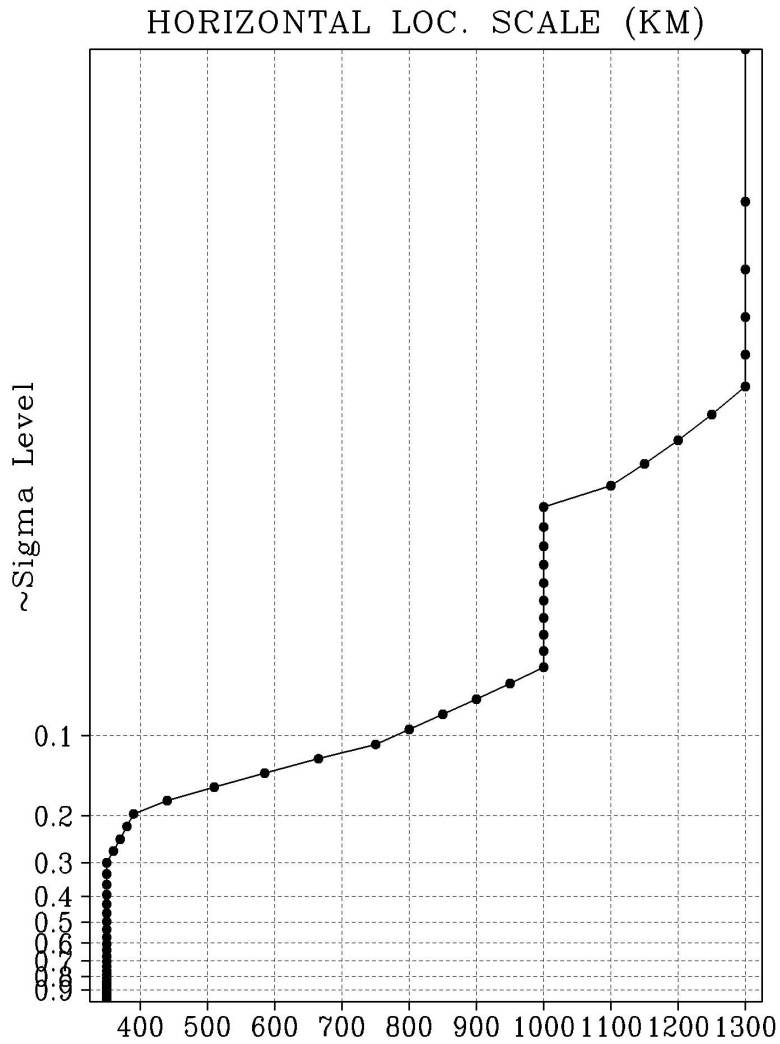
- **Dual Resolution**: If ensemble at lower resolution, interpolation (and adjoint) between ensemble/analysis grids* [**T**]

- **Initialization**: Option for Tangent Linear Normal Mode Constraint [C]

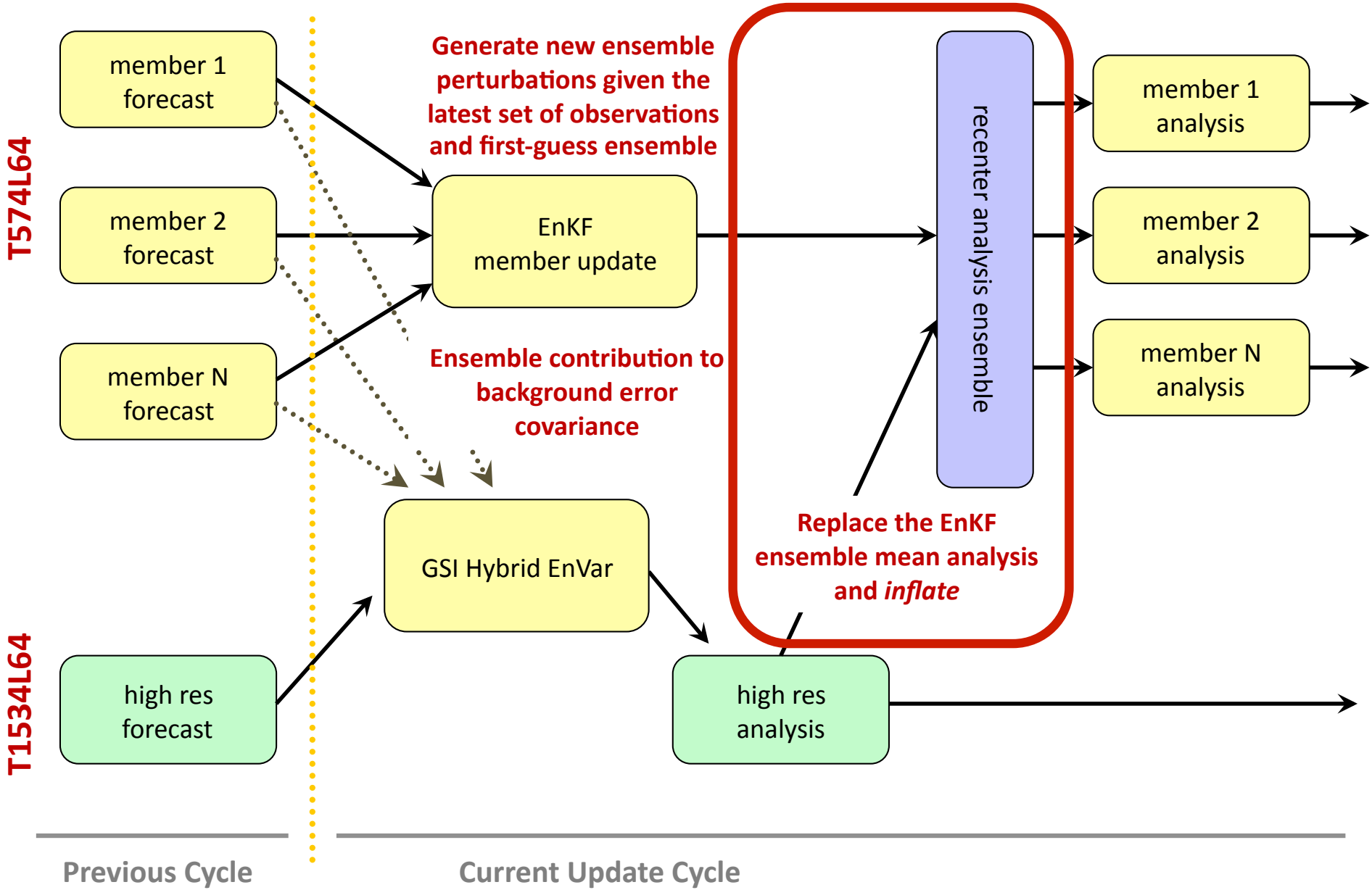
$$\mathbf{x}'_{\mathrm{t}} = \mathbf{C}\left[\mathbf{x}'_{\mathrm{f}} + \mathbf{T}\sum_{m=1}^{M}(\alpha^{m} \circ \mathbf{x}_{\mathrm{e}}^{m})\right]$$

Current Hybrid for GDAS/GFS

- T1534L64 Deterministic (SL dynamics)
  - T574L64 EnSRF, 80 members, Stoch. Physics, Hourly Output
- 87.5% ensemble, 12.5% climatological for hybrid increment
- Level dependent horizontal localization (divide by 0.38 to convert to GC zero distance)
  - 0.5 scale heights in vertical



Dual-Res Coupled Hybrid Var/EnKF Cycling



What if I am not running an EnKF?

- In principle any ensemble can be used;
  - so long as the ensemble represents the forecast errors well.

- GSI can ingest GFS global ensemble to update regional models
  - WRF ARW/NMM, NAM, RR, HWRF

- 80 member GFS/EnKF 6h ensemble forecasts are archived at NCEP since May 2012.
  - Real time ensemble is also publicly available.

Ensemble of Data Assimilations

- Much like the stochastic EnKF, ECMWF and Meteo-France use an ensemble of data assimilations instead of an EnKF
  - Perturb the observations and model
  - Designed to represent and estimate the uncertainty in their deterministic 4DVAR

- This provides flow-dependent estimates of analysis error for their EPS
- Also provides flow-dependent estimates of background error for use in DA (either as $\mathbf{B}_0$ or in hybrid)

- Can be hugely expensive, given that a variational (4DVAR) update has to be executed for each ensemble member!

4D Hybrids

The Hybrid EnVar cost function can be easily extended to 4D and include a static contribution (ignore preconditioning)

$$J(\mathbf{x}'_{\mathrm{c}},\alpha)=\beta_{\mathrm{c}}\frac{1}{2}(\mathbf{x}')^{\mathrm{T}}\mathbf{B}_{\mathrm{c}}^{-1}(\mathbf{x}')+\beta_{\mathrm{e}}\frac{1}{2}\alpha^{\mathrm{T}}\mathbf{L}^{-1}\alpha+\underbrace{\frac{1}{2}\sum_{k=1}^{K}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{x}'_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{x}'_k)}_{\text{Jo term divided into observation “bins” as in 4DVAR}}$$

Where the 4D increment is prescribed through linear combinations of the 4D ensemble perturbations plus static contribution, i.e. it is not itself a model trajectory

$$\boxed{\mathbf{x}'_k=\mathbf{C}_k\,[\mathbf{x}'_{\mathrm{c}}+\sum_{m=1}^{M}(\alpha^m\circ(\mathbf{x}_{\mathrm{e}})_k^m)]}$$

Here, static contribution is time invariant. **C** represents TLNMC balance operator. No TL/AD in Jo term (**M** and $\mathbf{M}^{\mathrm{T}}$)

GSI – Hybrid En-4DVar

Wang and Lei (2014); Kleist and Ide (2015)

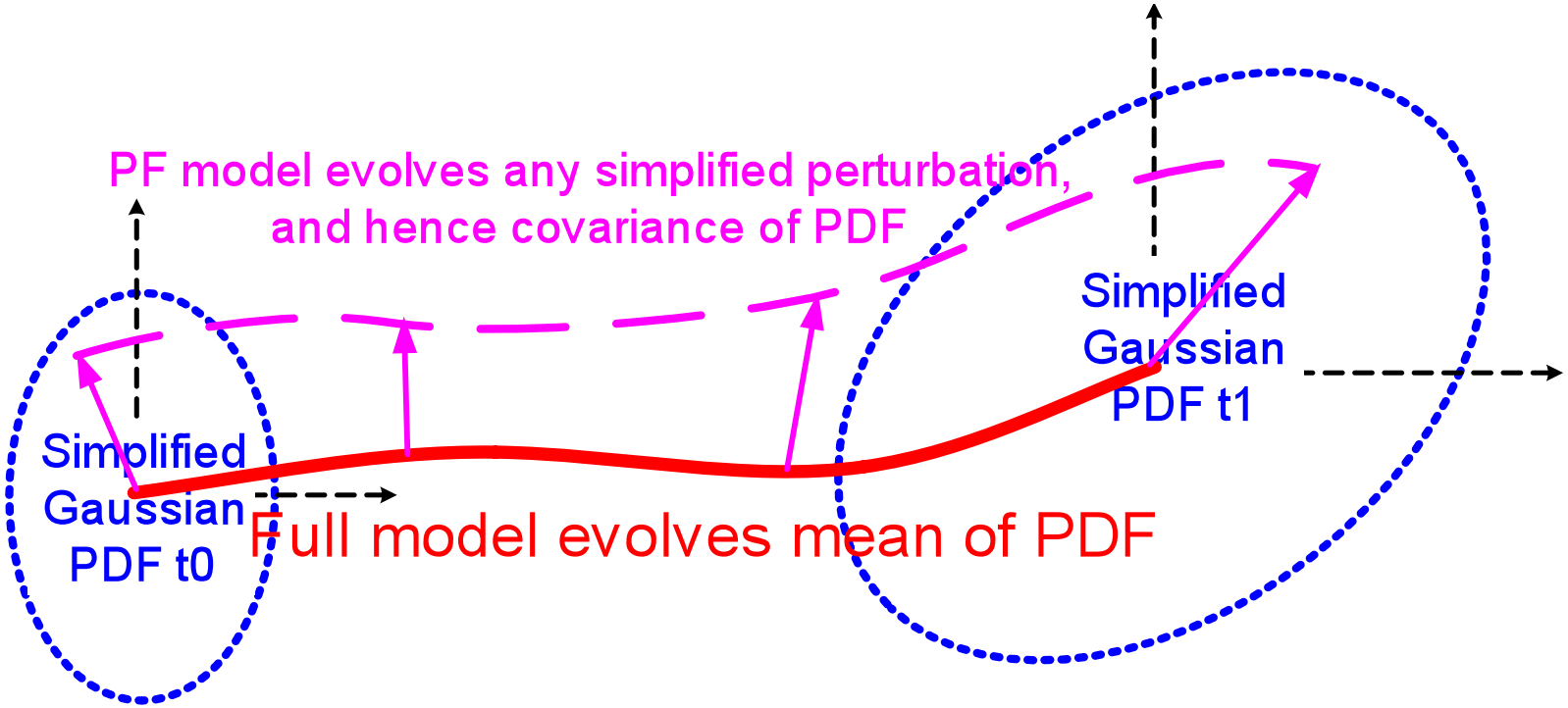
The traditional 4DVar cost function can be manipulated to use an ensemble to help prescribe the error covariance at the beginning of the window

$$J(\mathbf{x}'_c,\alpha)=\beta_c\frac{1}{2}(\mathbf{x}')^T\mathbf{B}_c^{-1}(\mathbf{x}')+\beta_e\frac{1}{2}\alpha^T\mathbf{L}^{-1}\alpha+\underbrace{\frac{1}{2}\sum_{k=1}^{K}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{M}_k\mathbf{x}'_0)^T\mathbf{R}_k^{-1}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{M}_k\mathbf{x}'_0)}_{\text{Jo term divided into observation “bins” as in 4DVAR}}$$

$$\mathbf{x}'_0=[\mathbf{x}'_c+\sum_{m=1}^{M}(\alpha^m\circ(\mathbf{x}_e)^m)]$$

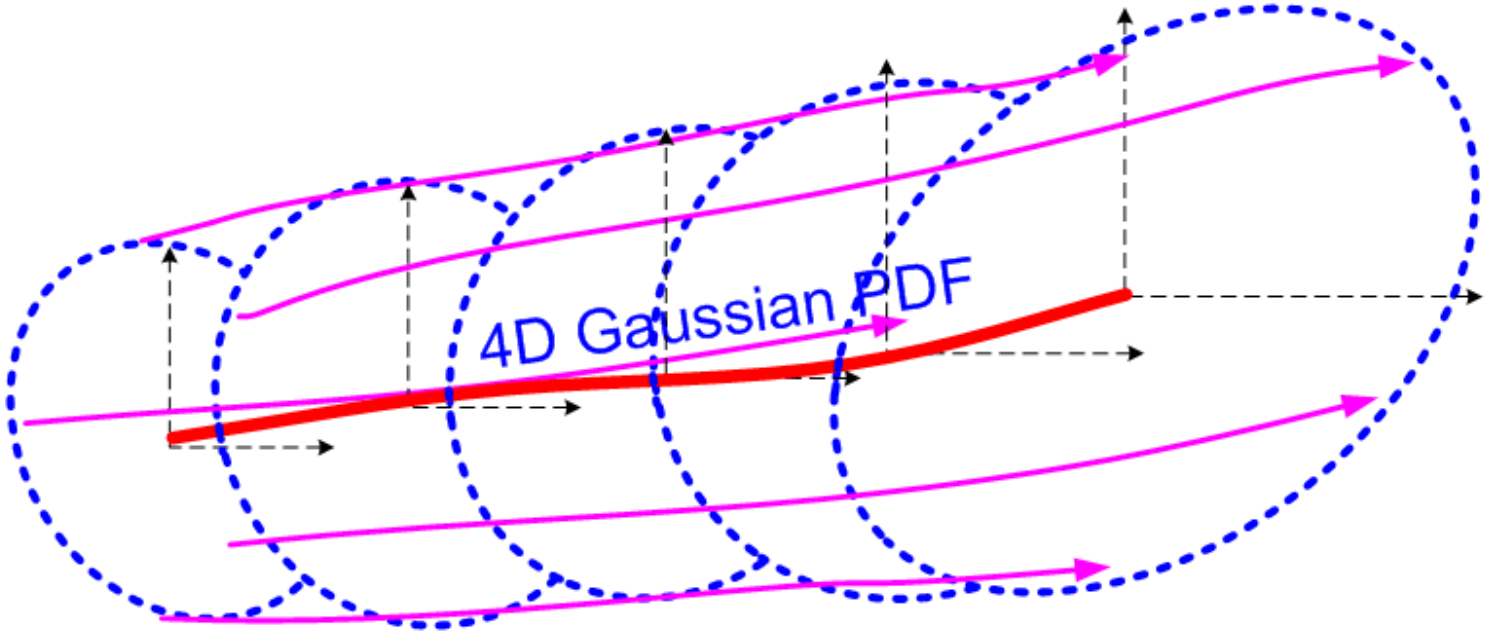
Here, the hybrid error covariance is applied at the beginning of the window, and the TL/AD propagate within observation window (**M** and $\mathbf{M}^T$) in Jo term

4DVAR



Lorenc & Payne 2007

4D EnVar



Trajectories of perturbations from ensemble mean

Full model evolves mean of PDF

Localised trajectories define 4D PDF of possible increments

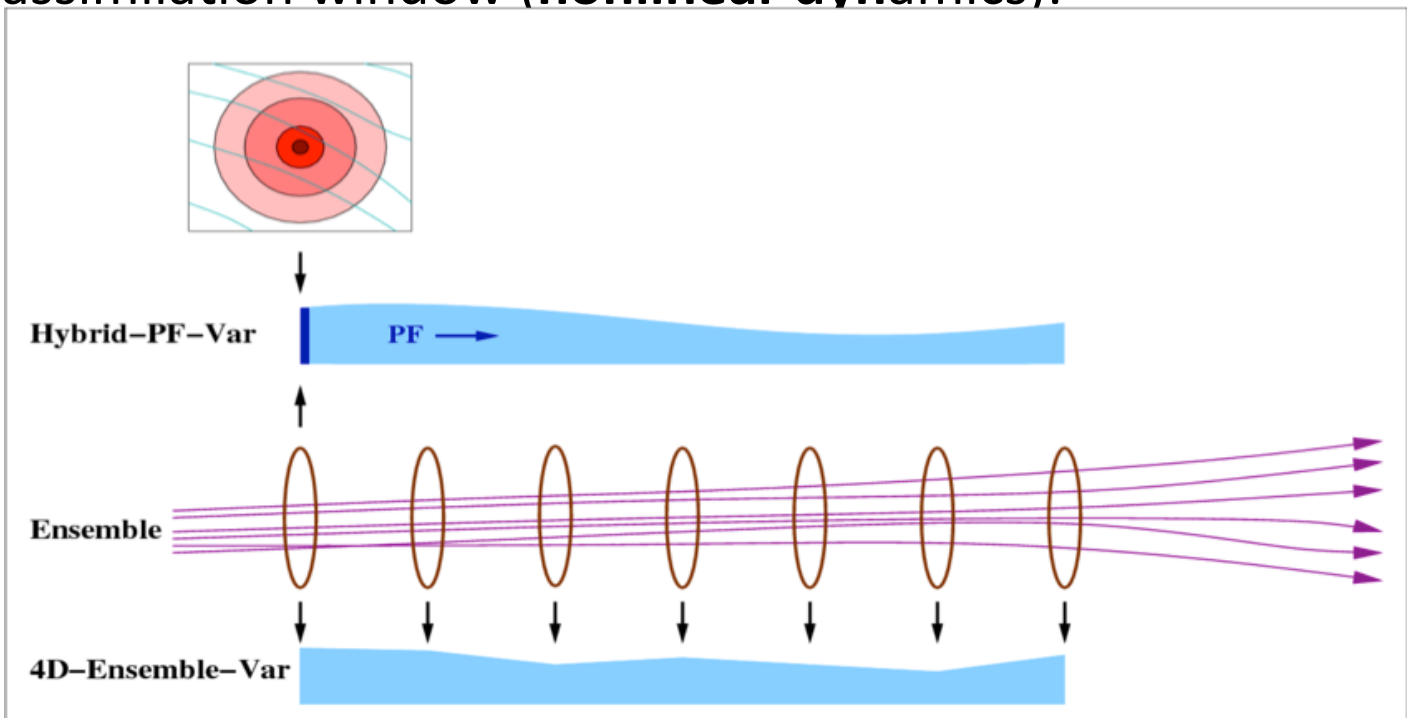
4D analysis is a (localised) linear combination of nonlinear trajectories. It is not itself a trajectory.

Courtesy: Andrew Lorenc

4D Hybrids

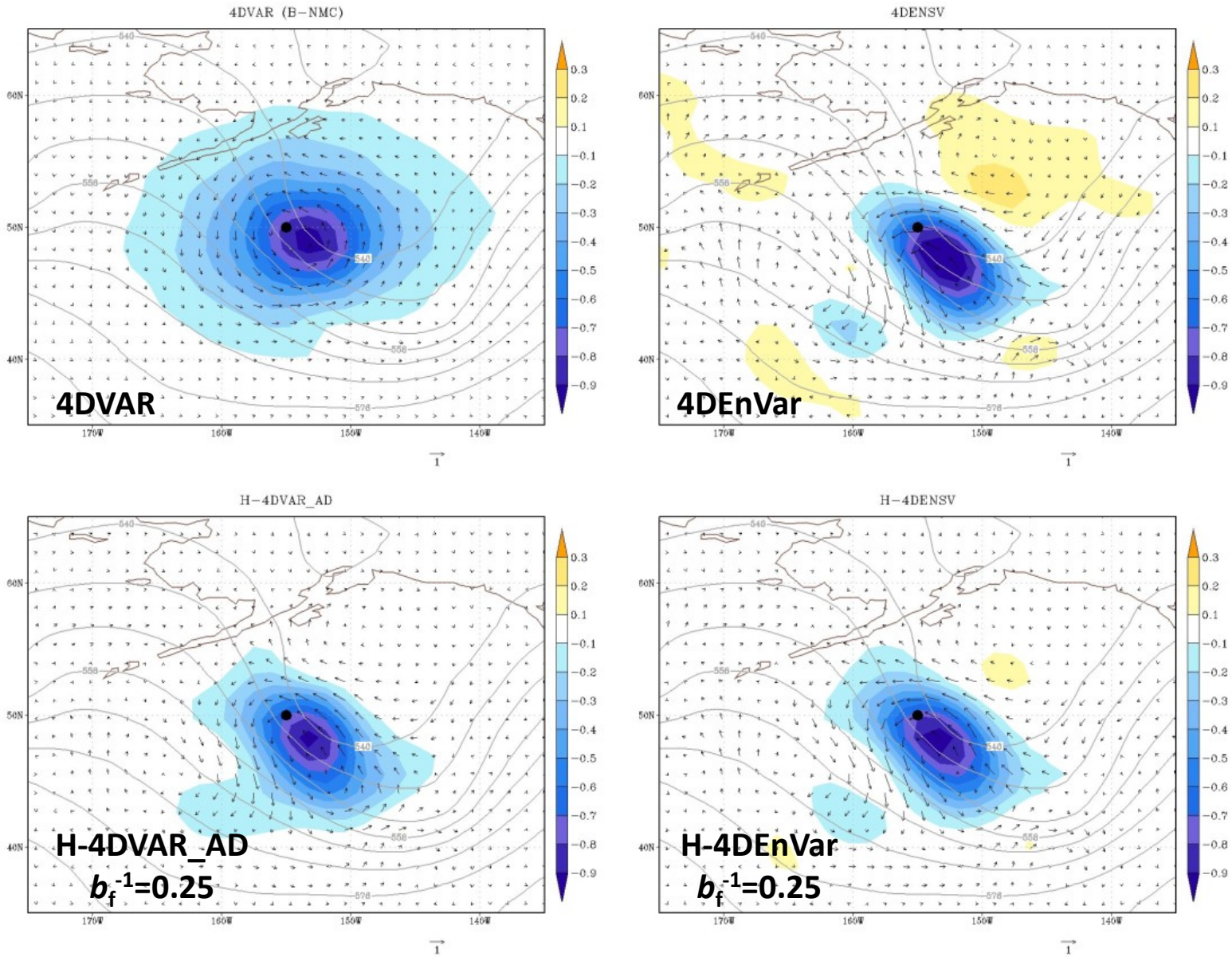
In the alpha control variable method one uses the ensemble perturbations to estimate $\mathbf{P}^b$ only at the start of the 4DVar assimilation window: the evolution of $\mathbf{P}^b$ inside the window is due to the **tangent linear dynamics** ($\mathbf{P}^b(t) \approx \mathbf{M}\mathbf{P}^b\mathbf{M}^T$)

In 4D-EnVar $\mathbf{P}^b$ is sampled from ensemble trajectories throughout the assimilation window (**nonlinear dyn**amics):



from: D. Barker & A. Clayton (UKMO)

Single Observation (-3h) Example for 4D Variants



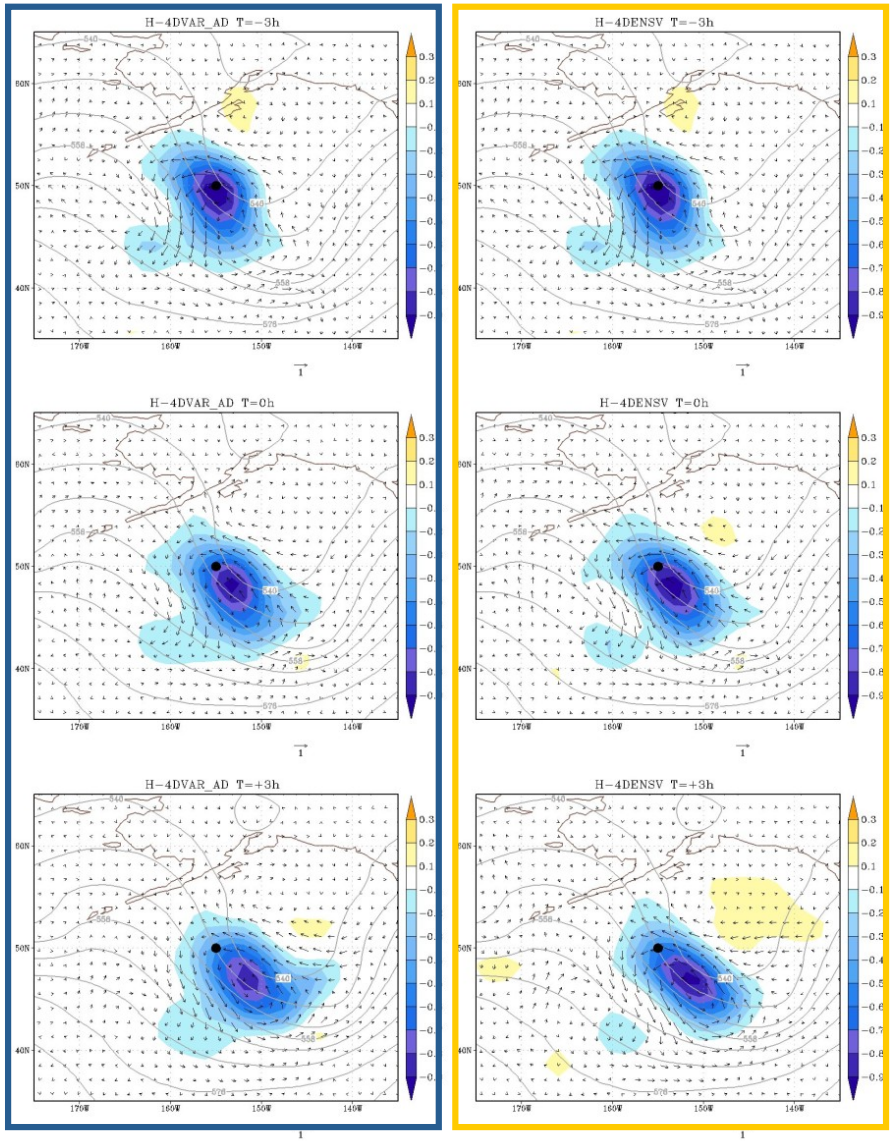
Time Evolution of Increment

H-4DVAR_AD **H-4DEnVar**

t=-3h

t=0h

t=+3h



Solution at beginning of window same to within round-off (because observation is taken at that time, and same weighting parameters used)

Evolution of increment qualitatively similar between dynamic and ensemble specification

GSI Hybrid 4DEnVar and En-4DVar Options

- **l4densvar**
  - Logical switch to turn on 4DEnVar
- **ens_nstarthr**
  - First forecast hour for ensemble valid in assimilation window. For the GDAS 06 hour cadence and window, this is "03", i.e. to use sigf03-sigf09 ensemble
- **nhr_obsbin**
  - Integer width for observation windows. GDAS uses hourly ("01").
- **lwrite4danl**
  - Option to write out 4D analysis (guess+increment at interval of nhr_obsbin).
- **thin4d**
  - Modify observation thinning to be a function of space only (no preferential treatment of observations near center of window)
- **filename**
  - Same as previously noted, but with multiple time levels consistent with above
- ***l4dvar***
  - ***Logical switch to turn on 4DVar (using TL/AD). User needs to compile with TL/AD interfaced to solver.***
  - ***Can run hybrid En-4DVar using this option plus hybrid options (*not* l4densvar).***

Comments

- 4DEnVar is more comparable in cost to 3DVar/3DEnVar since it does not involve TL/AD
  - 4DVar can be at least 10x more expensive depending on configuration
  - Lots of cost in ensemble, much more IO (4D), etc.

- Standard practice to use middle-loop option (relinearization), but work to be done in exploring use of outer loop as in 4DVar

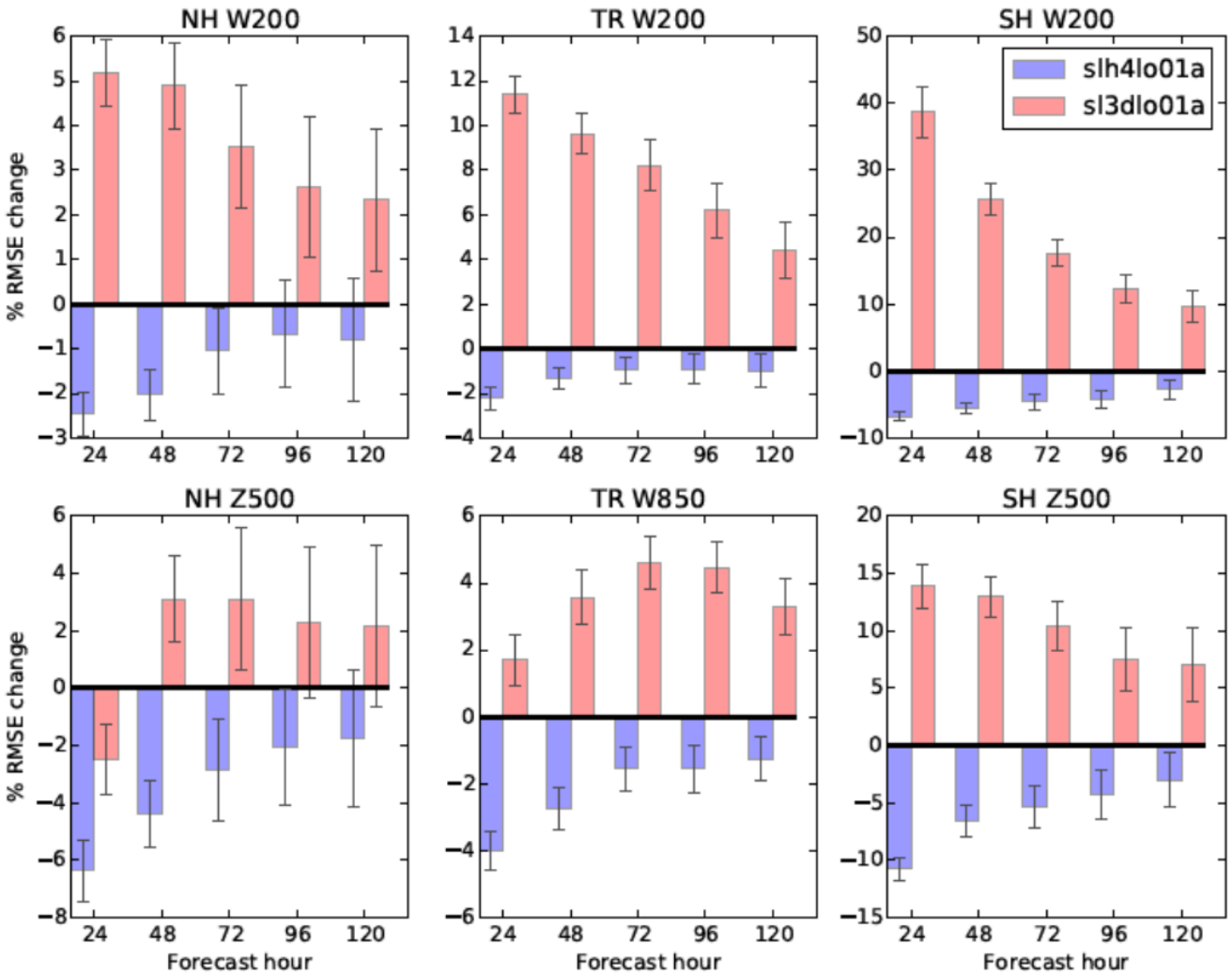
- Lots of work to do on multi-scale and scale-dependence
  - Localization
  - Hybrid weighting

Global 4D Hybrid at Major NWP Centers

- Hybrid 4D EnVar
  - CMC (Canada)
    - Replaced 4DVAR
  - NCEP
    - Extension of hybrid 3DEnVar

- Hybrid En-4DVAR (Operational or in Testing)
  - UKMO
  - ECMWF*
  - Meteo-France*
  - US Navgem
  - JMA

3DVar / H3DEnVar / H4DEnVar



- Move from 3D Hybrid (current operations) to Hybrid 4D-EnVar yields improvement that is about 75% in amplitude in comparison from going to 3D Hybrid from 3DVAR.

Some Hybrid References

- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational setting. *Quart. J. Roy. Meteor. Soc.,* **131**, 1013-1043.
- Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlin. Processes Geophys.,* **20**, 669-682, doi:10.5194/npg-20-669-2013, 2013.
- Clayton A. M, A. C. Lorenc and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.,* **139**, 1445-1461.
- Hamill, T. H., and C. Snyder., 2000. A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. Mon. Wea. Rev. 128, 2905–2919.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based Evaluation of Hybrid Variational-Ensemble Data Assimilation for the NCEP GFS, Part I: System Description and 3D-Hybrid Results, Mon. Wea. Rev.,143, 433-451.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based Evaluation of Hybrid Variational-Ensemble Data Assimilation for the NCEP GFS, Part II: 4D EnVar and Hybrid Variants, Mon. Wea. Rev., 143, 452-470.
- Kretschmer, M., B. R. Hunt, and E. Ott, 2015: Data assimilation using climatologically augmented local ensemble transform Kalman Filter, Tellus, 67, 9 pp.
- Liu C., Q. Xiao, and B. Wang, 2008: An ensemble-based four dimensional variational data assimilation scheme. Part I: technique formulation and preliminary test. *Mon. Wea. Rev*., **136,** 3363-3373.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc*., **129**, 3183-3203.
- Lorenc, A.C. 2013: Recommended nomenclature for EnVar data assimilation methods. *Research Activities in Atmospheric and Oceanic Modeling*. WGNE, 2pp, URL: http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01_Lorenc_Andrew_EnVar_nomenclature.pdf)
- Penny, S., 2014: A hybrid local ensemble transform Kalman filter. *Mon. Wea. Rev.* 142, 2139-2149..

Some Hybrid References

- Wang, X., C. Snyder, and T. M. Hamill, 2007a: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev*., **135**, 222-227.
- Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation (GSI) variational minimization: a mathematical framework. *Mon. Wea. Rev*., **138**, 2990-2995.
- Zhang F., M. Zhang, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited area weather prediction model and comparison to E4DVar. *Mon. Wea. Rev.*, **141,** 900–917.
- Zhang, M. and F. Zhang, 2012: E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Wea. Rev.*, **140,** 587–600.

Questions?