A Unified Framework for Four-Dimensional Ensemble-Variational Hybrid Data Assimilation:

Relationships among Ensemble Variational Algorithms with Full and Approximate Ensemble Covariance Localization

Chengsi Liu and Ming Xue

Center for Analysis and Prediction of Storms
University of Oklahoma

6th WMO Data Assimilation Symposium
October 2013

Liu and Xue (2013 QJ – under review)

Outline

- Background and Motivations
- 4D ensemble-variational (4D EnVar) algorithms
- Approximations and equivalence among the EnVar algorithms
- Single observation tests with a one-dimensional linear advection model
- Summary and on-going research

Background

- Lorenc (2003) proposed a so-called **alpha-control variable method** that allows the inclusion of localized ensemble-derived covariances within a variational framework. It was suggested that the method could be applied to a 4DVAR framework also.

- Clayton et al. (2012) implemented the algorithm within the U.K. Met Office 4DVAR system, and called it Hybrid En4DVAR. The system requires linear and adjoint models.

Background – continued

- Liu et al (2008, 2009) proposed an alternative 4D ensemble-variational formulation whose cost function projects the ensemble perturbations to the observation space so that the tangent linear model (TLM) and adjoint model (AJM) can be avoided.

- Buehner et al. (2010a, b) proposed a similar 4DEnVar without adjoint model and was implemented within a global spectral model

- The method is called En4DVar in original papers. It has been applied to an Antarctic cyclone case study and renamed **4DEnVar** (Liu and Xiao 2013) to better distinguish it with 4DVAR (that requires adjoint).

Motivations

- Extend Liu et al 4DEnVAR formulation to include static covariance and examine approximations needed to avoid TLM and adjoint models

- Understand relationship among different 4D ensemble-variational algorithms and the approximations involved.

- Establish a general framework for the algorithms

Hybrid En-4DVar (Lorenc 2003, Clayton et al 2012)

Static **B** part

$$\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$$

$$\delta \mathbf{x}_s = \mathbf{U}\mathbf{v}$$

Ensemble covariance part

Localization matrix $\mathbf{C} = \mathbf{C}'\mathbf{C}'^{\mathrm{T}}$

Alpha control variable $\boldsymbol{\alpha} = \begin{bmatrix} \mathbf{C}' & & 0 \\ & \ddots & \\ 0 & & \mathbf{C}' \end{bmatrix} \tilde{\boldsymbol{\alpha}}$

$$\delta \mathbf{x}_e = \sum_{i=1}^{N} (\mathbf{x}'_{bi} \circ \mathbf{C}'\tilde{\boldsymbol{\alpha}}_i)$$

Analysis increment and cost function

$$\beta_1^{\,2} + \beta_2^{\,2} = 1$$

$$\delta \mathbf{x}_0 = \beta_1 \delta \mathbf{x}_s + \beta_2 \delta \mathbf{x}_e = \beta_1 \mathbf{U}\mathbf{v} + \beta_2 \sum_{i=1}^{N} (\mathbf{x}'_{bi} \circ \mathbf{C}'\tilde{\boldsymbol{\alpha}}_i)$$

$$J(\mathbf{v}, \tilde{\boldsymbol{\alpha}}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\tilde{\boldsymbol{\alpha}} + \frac{1}{2}\sum_{t=1}^{I} [\mathbf{H}_t \mathbf{L}_{\mathrm{t}} \delta \mathbf{x}_0 + \mathbf{d}_t]^T \mathbf{R}^{-1} [\mathbf{H}_t \mathbf{L}_{\mathrm{t}} \delta \mathbf{x}_0 + \mathbf{d}_t]$$

innovation

4D ensemble-variational algorithms characteristics before including static B

	En4DVar	En4DVar-NPC	4DEnVar	4DEnVar-NPL
AJM	Needed	Not needed	Not needed	Not needed
TLM	Needed	Not needed	Needed	Not needed.
COST	Same order of magnitude as 4DVar, need initial 3D ensemble states	Same order of magnitude as 3DVar, need 4D ensemble trajectories	n*N ensemble forecast, need initial 3D ensemble states	N times 3DVar Plus ensemble forecast, need 4D ensemble trajectories
Additional approx.	none	NPC	None	NPL

AJM: adjoint model
TLM: tangent linear model
NPC: **N**on-**P**ropagation of alpha **C**ontrol variable
NPL: **N**o **P**ropagation of **L**ocalization

N: ensemble number
n: grid point number

En4DVar-**NPC** (**N**on-**P**ropagation of alpha **C**ontrol variable)

Two approximations

(1) Neglecting temporal propagation of alpha control variable by TLM

$$\mathbf{L}_t \delta \mathbf{x}_e = \mathbf{L}_t (\sum_{i=1}^{N} \mathbf{x}'_{bi} \circ \mathbf{C}' \tilde{\boldsymbol{\alpha}}_i) = \sum_{i=1}^{N} \mathbf{L}_t (\mathbf{x}'_{bi} \circ \mathbf{C}' \tilde{\boldsymbol{\alpha}}_i) \approx \sum_{i=1}^{N} (\mathbf{L}_t \mathbf{x}'_{bi}) \circ \mathbf{C}' \tilde{\boldsymbol{\alpha}}_i$$

$$\nabla_{\tilde{\boldsymbol{\alpha}}_i} J(\tilde{\boldsymbol{\alpha}}_i) = \tilde{\boldsymbol{\alpha}}_i + \sum_{t=1}^{m} \mathbf{C}'^{\mathrm{T}} \underline{(\mathbf{L}_{\mathrm{t}} \mathbf{x}'_{bi})} \circ \mathbf{H}_t{}^{\mathrm{T}} \mathrm{R}^{-1} [\mathbf{H}_t \sum_{i=1}^{N} \{(\mathbf{L}_{\mathrm{t}} \mathbf{x}'_{bi}) \circ \mathbf{C}' \tilde{\boldsymbol{\alpha}}_i\} + \mathbf{d}_t]$$

AJM avoided !!

(2) Use nonlinear model ensemble forecasts to replace the temporal propagation of perturbations by the TLM

$$\mathbf{L}_t \mathbf{x}'_{bi} \approx M_t(\mathbf{x}_{bi}) - \overline{M_t(\mathbf{x}_b)}$$

TLM avoided !!

4D ensemble-variational algorithms characteristics before including static B

	En4DVar	En4DVar-NPC	4DEnVar	4DEnVar-NPL
AJM	Needed	**Not needed**	Not needed	Not needed
TLM	Needed	**Not needed**	Needed	Not needed.
COST	Same order of magnitude as 4DVar, need initial 3D ensemble states	**Same order of magnitude as 3DVar, need 4D ensemble trajectories**	n*N ensemble forecast, need initial 3D ensemble states	N times 3DVar Plus ensemble forecast, need 4D ensemble trajectories
Additional approx.	none	**NPC**	None	NPL

AJM: adjoint model
TLM: tangent linear model
NPC: **N**on-**P**ropagation of alpha **C**ontrol variable
NPL: **N**o **P**ropagation of **L**ocalization

N: ensemble number
n: grid point number

Hybrid 4DEnVar

Static **B** part

$$\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$$

$$\delta \mathbf{x}_s = \mathbf{U}\mathbf{v}$$

Ensemble covariance part
(Liu et al 2008, 2009)

Localization matrix $\mathbf{C} = \mathbf{C}'\mathbf{C}'^{\mathrm{T}}$

perturbation matrix $\mathbf{S}_i' = \left(\mathbf{x}_{bi}' \quad \ldots \quad \mathbf{x}_{bi}' \right)$

Localized perturbation matrix

$$\mathbf{Z}_b' = [\mathbf{S}_{b1}' \circ \mathbf{C}' \quad \mathbf{S}_{b2}' \circ \mathbf{C}' \quad .... \quad \mathbf{S}_{bN}' \circ \mathbf{C}']$$

Project $\mathbf{Z}_b'$ to obs space $(\mathbf{HLZ}_b')$

$$J(\mathbf{v},\mathbf{w}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} +$$

(**w** is the same as $\tilde{\alpha}$ of En4DVar)

$$\frac{1}{2}\sum_{t=1}^{I}\left[\beta_1\mathbf{H}_t\mathbf{L}_t\delta\mathbf{x}_s + \beta_2(\mathbf{H}_t\mathbf{L}_t\mathbf{Z}_b')\mathbf{w} + \mathbf{d}_t\right]^{\mathrm{T}}\mathbf{R}^{-1}[(\beta_1\mathbf{H}_t\mathbf{L}_t\delta\mathbf{x}_s + \beta_2(\mathbf{H}_t\mathbf{L}_t\mathbf{Z}_b')\mathbf{w} + \mathbf{d}_t]$$

$$\nabla_{\mathbf{w}}J(\mathbf{w}) = \mathbf{w} + \sum_{t=1}^{I}\beta_1(\mathbf{H}_t\mathbf{L}_t\mathbf{Z}_b')^{\mathrm{T}}\mathbf{R}^{-1}[\beta_1\mathbf{H}_t\mathbf{L}_t\mathbf{U}\mathbf{v} + \beta_2\mathbf{H}_t\mathbf{L}_t\mathbf{Z}_b'\mathbf{w} + \mathbf{d}_t]$$

Avoid AJM

4D ensemble-variational algorithms characteristics before including static B

	En4DVar	En4DVar-NPC	4DEnVar	4DEnVar-NPL
AJM	Needed	Not needed	**Not needed**	Not needed
TLM	Needed	Not needed	**Needed**	Not needed.
COST	Same order of magnitude as 4DVar, need initial 3D ensemble states	Same order of magnitude as 3DVar, need 4D ensemble trajectories	**n*N ensemble forecast, need initial 3D ensemble states**	N times 3DVar Plus ensemble forecast, need 4D ensemble trajectories
Additional approx.	none	NPC	**None**	NPL

AJM: adjoint model
TLM: tangent linear model
NPC: **N**on-**P**ropagation of alpha **C**ontrol variable
NPL: **N**o **P**ropagation of **L**ocalization

N: ensemble number
n: grid point number

4DEnVar-**NPL** (**N**o **P**ropagation of **L**ocalization)

Two approximations

(1) The localization matrix propagation by TLM is neglected

$$\mathbf{L}_t[\mathbf{S}'_{b1} \circ \mathbf{C}' \quad \mathbf{S}'_{b2} \circ \mathbf{C}' \quad .... \quad \mathbf{S}'_{bN} \circ \mathbf{C}']$$
$$\approx [(\mathbf{L}_t \mathbf{S}'_{b1}) \circ \mathbf{C}' \quad (\mathbf{L}_t \mathbf{S}'_{b2}) \circ \mathbf{C}' \quad .... \quad (\mathbf{L}_t \mathbf{S}'_{bN}) \circ \mathbf{C}']$$

(2) Ensemble forecasts are used to replace the temporal propagation of perturbations by the TLM (as in En4DVar-NPC)

$$\mathbf{L}_t \mathbf{x}'_{bi} \approx M_t(\mathbf{x}_{bi}) - \overline{M_t(\mathbf{x}_b)}$$

TLM avoided !!

4D ensemble-variational algorithms characteristics before including static B

	En4DVar	En4DVar-NPC	4DEnVar	4DEnVar-NPL
AJM	Needed	Not needed	Not needed	Not needed
TLM	Needed	Not needed	Needed	Not needed.
COST	Same order of magnitude as 4DVar, need initial 3D ensemble states	Same order of magnitude as 3DVar, need 4D ensemble trajectories	n*N ensemble forecast, need initial 3D ensemble states	N times 3DVar Plus ensemble forecast, need 4D ensemble trajectories
Additional approx.	none	NPC	None	NPL

AJM: adjoint model
TLM: tangent linear model
NPC: **N**on-**P**ropagation of alpha **C**ontrol variable
NPL: **N**o **P**ropagation of **L**ocalization

N: ensemble number
n: grid point number

Equivalence of hybrid En4DVar and hybrid 4DEnVar

$$\mathbf{Z}_b^{'}\mathbf{w} = (\mathbf{S}'_{b1} \circ \mathbf{C}', \mathbf{S}'_{b2} \circ \mathbf{C}', \ldots., \mathbf{S}'_{bN} \circ \mathbf{C}') \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix} = \sum_{i=1}^{N} (\mathbf{S}'_{bi} \circ \mathbf{C}')\mathbf{w}_i = \sum_{i=1}^{N} \mathbf{x}'_{bi} \circ (\mathbf{C}'\mathbf{w}_i)$$

From Hybrid 4DEnVar

Substituting the above equation to the **Hybrid 4DEnVar** cost function

↓

Hybrid En4DVar Cost function ⟺ **Hybrid 4DEnVar Cost function**

Equivalence of En4DVar-NPC and 4DEnVar-NPL

$$\mathbf{T}_t'\mathbf{w} = [(\mathbf{L}_t\mathbf{S}_{b1}') \circ \mathbf{C}' \quad (\mathbf{L}_t\mathbf{S}_{b2}') \circ \mathbf{C}' \quad .... \quad (\mathbf{L}_t\mathbf{S}_{bN}') \circ \mathbf{C}'] \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix} = \sum_{i=1}^{N} \mathbf{L}_t\mathbf{x}_{bi}' \circ (\mathbf{C}'\mathbf{w}_i).$$

From Hybrid 4DEnVar-NPL

Substituting the above equation to
the **Hybrid 4DEnVar-NPL** cost function

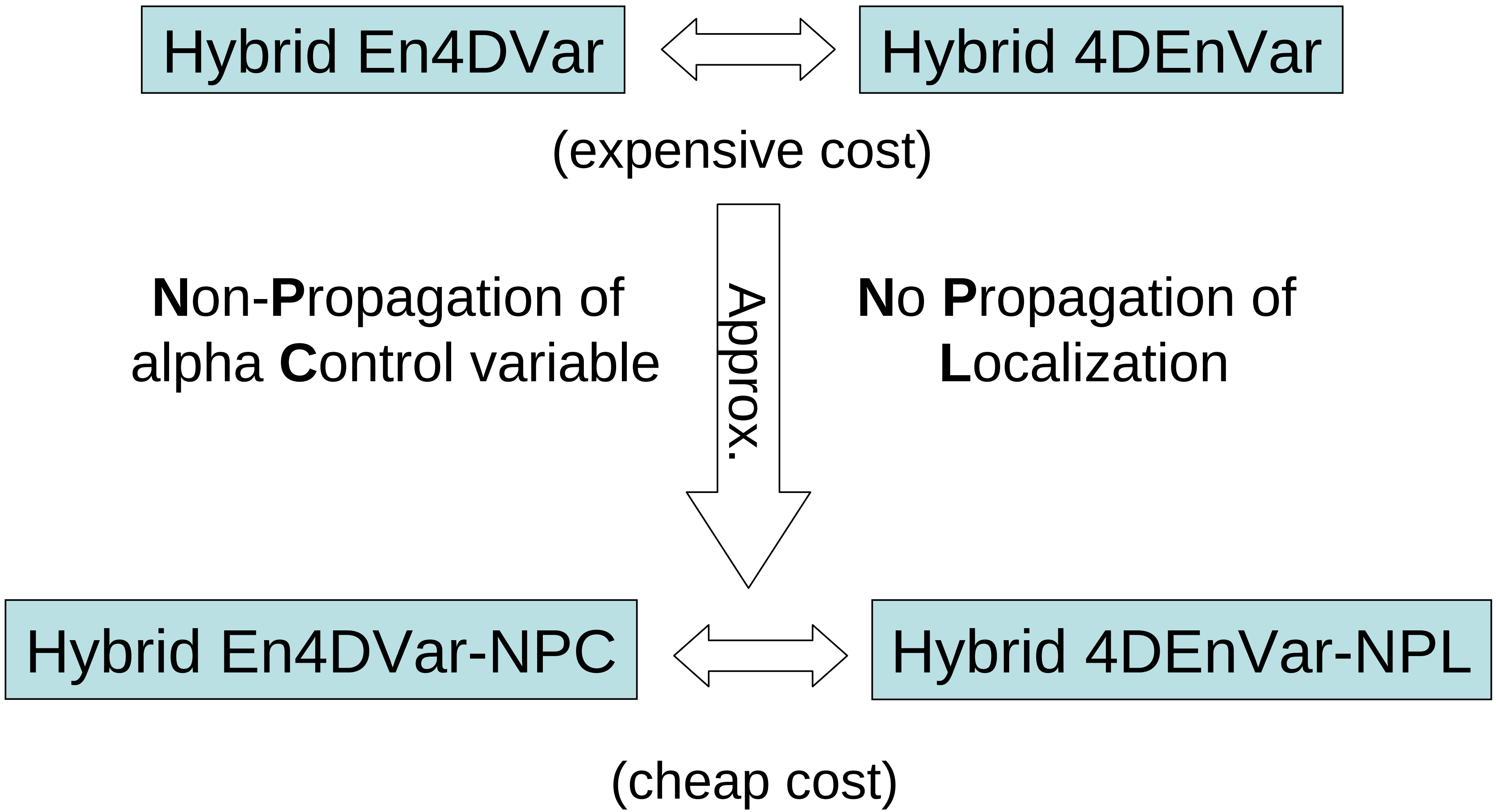
↓

Hybrid En4DVar-NPC
Cost function

⟺

Hybrid 4DEnVar-NPL
Cost function

The relationship of hybrid 4D ensemble variational algorithms



4D hybrid schemes with FGAT

- Although En4DVar-NPC, 4DEnVar and 4DEnVar-NPL avoid TLM and AJM, but when including static **B**, TLM and AJM are still needed

- Hybrid with FGAT so that TLM and AJM is completely avoided

4DVar

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\sum_{t=1}^{I}\left[\mathbf{H}_t\underline{\mathbf{L}_t\mathbf{U}}\mathbf{v} + \mathbf{d}_t\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{H}_t\mathbf{L}_t\mathbf{U}\mathbf{v} + \mathbf{d}_t\right]$$

FGAT (neglecting propagation of the static part of the control variable)

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\sum_{t=1}^{I}\left[\mathbf{H}_t\underline{\mathbf{U}}\mathbf{v} + \mathbf{d}_t\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{H}_t\mathbf{U}\mathbf{v} + \mathbf{d}_t\right]$$

Single observation tests with a 1D linear advection model

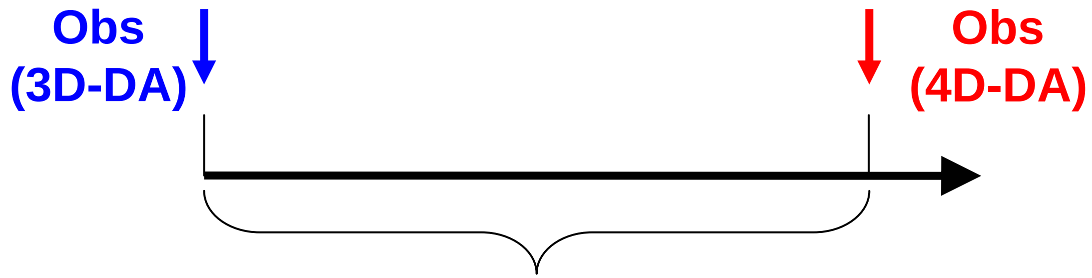
$$\frac{\partial u}{\partial t} + \mathrm{U}\frac{\partial u}{\partial x} = 0$$

The numerical result also proven:

En4DVar and 4DEnVar (with no approximation to the flow-following-localization) are the same

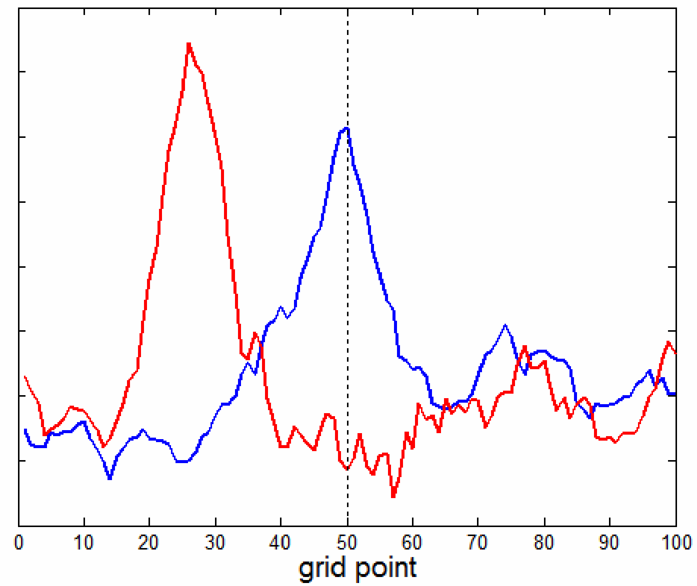
En4DVar-NPC and 4DEnVar-NPL (with non-flow-following localization approximation) are the same

Flow-following & non-flow-following localization



Assimilation window

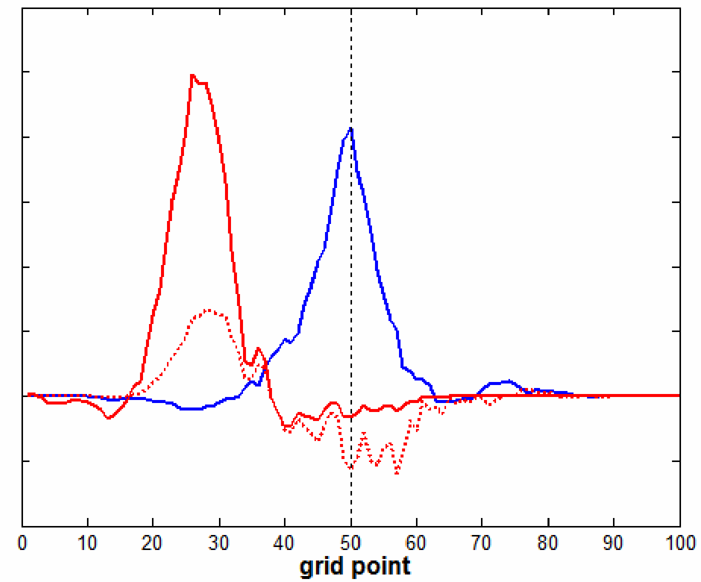
Ensemble DA **W/O localization**



(four algorithms are the same)

advection

Ensemble DA **W/ localization**



solid: flow-following localization (En4DVar/4DEnVar)

dot: non-flow-following localization (En4DVar-NPC/4DEnVar-NPL)

Two factors affecting approximations:
Advection speed & spatial de-correlation scale

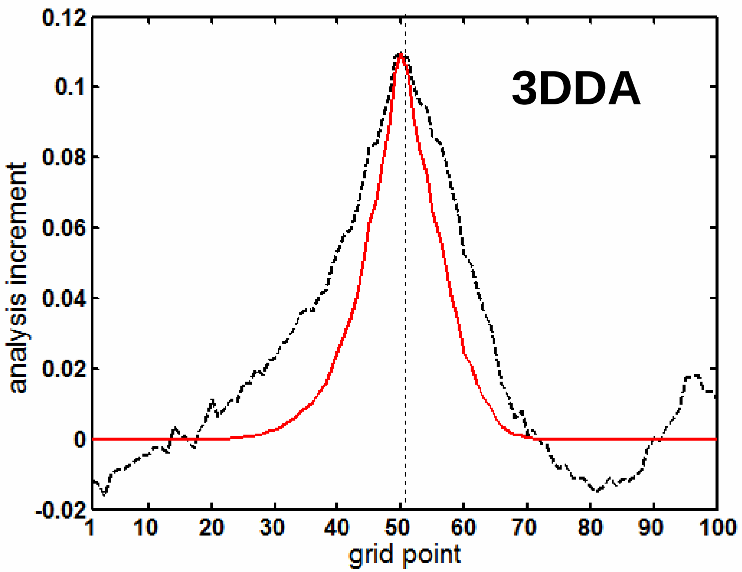
sensitivity experiments

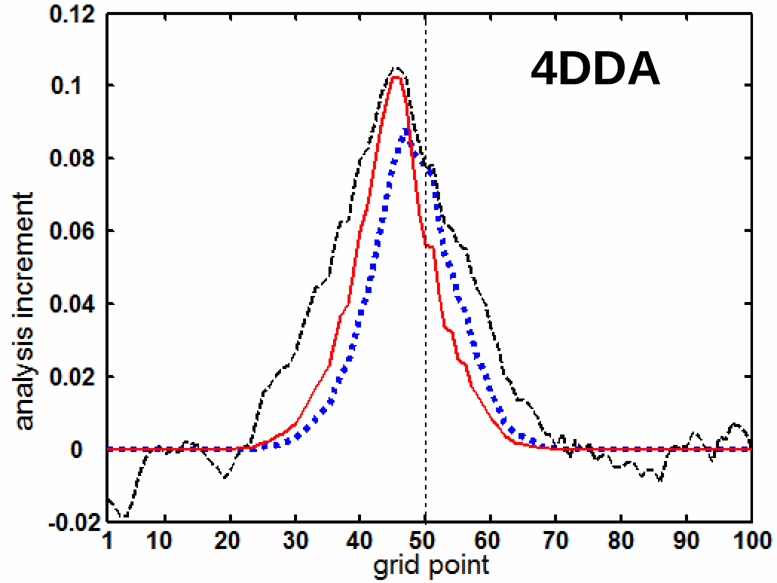
Exp1 (3DDA): obs at beginning of DA window

Exp2 (4DDA): obs at end of DA window
(by slow speed and large scale)

Exp3 (4DDA_2S): same as Exp2,
but doubling the advection speed

Exp4 (4DDA_L/2): same as Exp2,
but half of the error spatial de-correlation scale

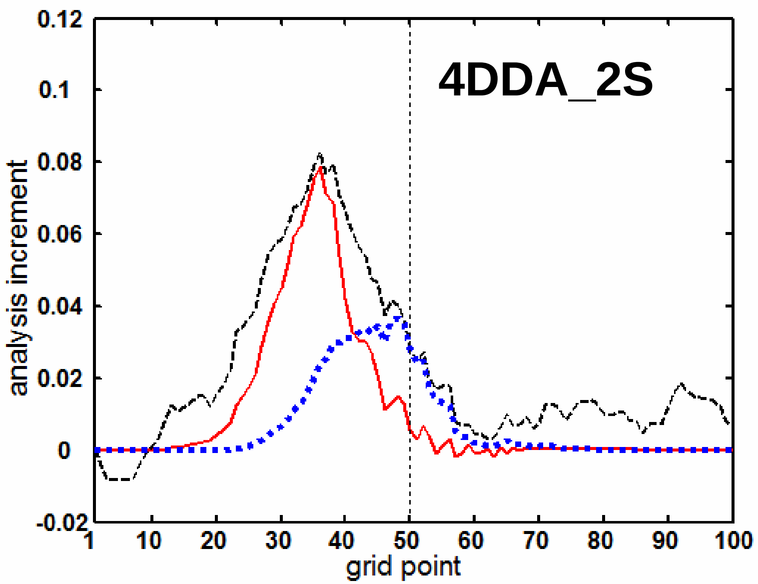


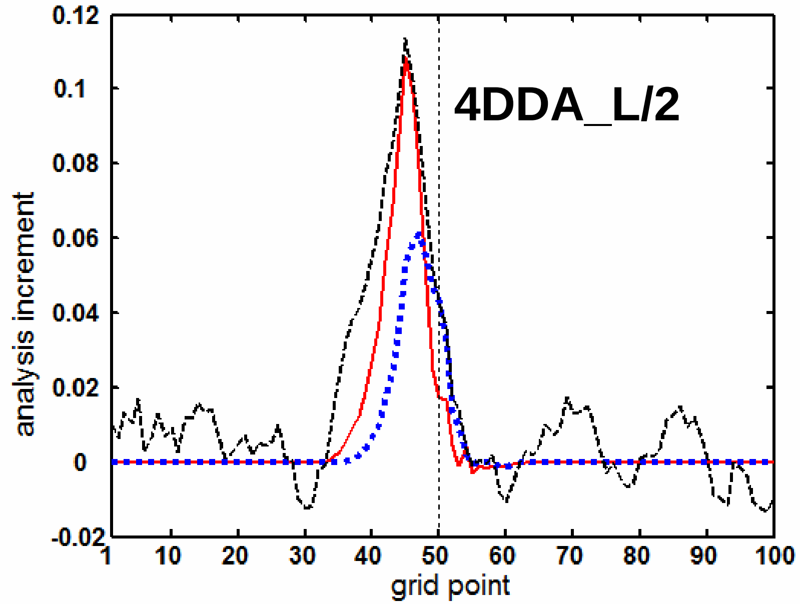


black no-localization

red flow-following localization (En4DVar/4DEnVar)

blue non-flow-following localization (En4DVar-NPC/4DEnVar-NPL)





- **How to deal with non-flow-following localization issue** is a key of implementing En4DVar-NPC/4DEnVar-NPL
  - short data assimilation window
  - Applied on slow flow
  - Use a large correlation scale localization parameter

Summary

- **En4DVar** based on the alpha control variable **and 4DEnVar** based on Liu et al. algorithm are mathematically equivalent. They use **flow-following ensemble covariance localization** implicitly or explicitly.

- **En4DVar-NPC and 4DEnVar-NPL** are approximate algorithms and both algorithm are equivalent. They use **non-flow-following localization** implicitly or explicitly

- Use **FGAT approximation** avoids AJM in hybrid algorithms while still allowing for the use of observations distributed over a time window in the static portion of En4DVar-NPC, 4DEnVar and 4DEnVar-NPL

- For relatively **slow signal propagation and/or relatively larger error correlation sales**, En4DVar-NPC and 4DEnVar-NPL are a good approximation to En4DVar and 4DEnVar

On-going Work

- A 4D hybrid EnVar DA system is being developed for ARPS then WRF, including capabilities for convective-scale radar DA (the En4DVar-NPC-FGAT scheme is adopted)

- Ensemble perturbations will be provided by 4DEnKF or by running an ensemble of the (perturbed observation) 4DEnVar

  – Combine benefits of 4DVAR and 4DEnKF
    - Can assimilate high-frequency data in 4D windows
    - Carry ensemble covariance through assimilation cycles
    - Does not require tangent linear or adjoint model