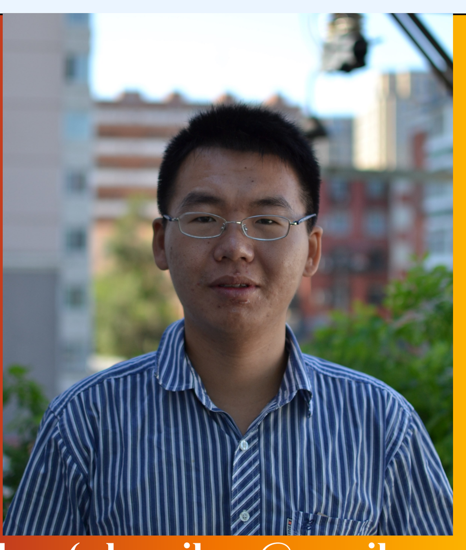
Dimension-Reduced Projection 4DVar with Nonlinear Correction

Si Shen[1,3], Juanjuan Liu[1] and Bin Wang[1,2]

1. LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
2. Ministry of Education Key Laboratory for Earth System Modeling, Center of Earth System Science, Tsinghua University, Beijing, China
3. College of Earth Science, University of Chinese Academy of Sciences, Beijing, China



Si Shen (shensilasg@gmail.com)
PhD Student on Data Assimilation

1. Introduction

One of the important strategies of improving the operational implementation of 4DVar is to develop an ensemble-based 4DVar (En4DVar), which uses an ensemble method similar to the one used in the ensemble Kalman filter (EnKF) for obtaining a solution for the minimization of 4DVar. The dimension-reduced projection 4DVar (DRP-4DVar) proposed by Wang et al. (2010) is one representative approach of the En4DVar family.

The DRP-4DVar solves the 4DVar problem in the subspace defined by an ensemble. This approach conserves the advantages of the 4DVar while avoiding the use of the adjoint model. Thus, DRP-4DVar is a much more computationally economical way to implement 4DVar. Besides, similar to the EnKF, DRP-4DVar uses real-time ensemble to estimate the background error covariance outside the assimilation window, which ensures its global flow-dependence. Plenty of studies on DRP-4DVar have shown good performance both with simple models like Lorenz96 and realistic models like MM5, WRF, AREM etc.

However, DRP-4DVar has a potential problem. The DRP-4DVar replaces the tangent linear model of 4DVar with a linear statistic relation to forecast perturbations (see 2. Method). In this way, DRP-4DVar avoids the use of adjoint model and its cost function is simplified as an explicit quadratic polynomial. As a result, the performance of DRP-4DVar may be limited by **the linear approximation**. This work focuses on this potential problem. **We verify the problem by Lorenz96 model and present an iterative nonlinear correction process to alleviate the problem.**

2. Method

i) DRP-4DVar

For 4DVar, its cost function can be written as follow:

$$\begin{cases} J(\mathbf{x}_0)=\frac{1}{2}\{(\mathbf{x}_0-\mathbf{x}_0{}^b)^T\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}_0{}^b)+[\mathbf{Y}^o-\mathbf{Y}(\mathbf{x}_0)]^T\mathbf{O}^{-1}[\mathbf{Y}^o-\mathbf{Y}(\mathbf{x}_0)]\} \\ \mathbf{Y}^o=[\mathbf{y}_1{}^o,\mathbf{y}_2{}^o,...,\mathbf{y}_N{}^o]^T;\mathbf{Y}(\mathbf{x}_0)=[\mathbf{y}_1(\mathbf{x}_0),\mathbf{y}_2(\mathbf{x}_0),...,\mathbf{y}_N(\mathbf{x}_0)]^T \\ \mathbf{y}_i(\mathbf{x}_0)=H_i\circ M_{t0->ti}(\mathbf{x}_0) \end{cases}$$

4DVar

where N is the total number of observation times within the assimilation window, $[t_1, t_2, \ldots, t_N]$ are the times of observations.

In DRP-4DVar, we have an ensemble of initial condition perturbations $\boldsymbol{P_x}$ (perturbations based on $\boldsymbol{x_b}$). If we assume the analysis increment can be expressed as a linear combination of the ensemble members (i.e. $\boldsymbol{\delta x_0}^a=\boldsymbol{P_x\alpha}$), we have the cost function transformed as a function of $\boldsymbol{\alpha}$:

$$J(\boldsymbol{\alpha})=\frac{1}{2}\{\boldsymbol{\alpha}^T\mathbf{P}_x{}^T\mathbf{B}^{-1}\mathbf{P}_x\boldsymbol{\alpha}+[\mathbf{Y}^o{}'-\mathbf{Y}'(\mathbf{P}_x\boldsymbol{\alpha})]^T\mathbf{O}^{-1}[\mathbf{Y}^o{}'-\mathbf{Y}'(\mathbf{P}_x\boldsymbol{\alpha})]\}$$

Nonlinear DRP-4DVar

where $\mathbf{Y}^o{}'=\mathbf{Y}^o-\mathbf{Y}(\mathbf{x}_0{}^b)$ and $\mathbf{Y}'(\mathbf{P}_x\boldsymbol{\alpha})=\mathbf{Y}(\mathbf{x}_0{}^b+\mathbf{P}_x\boldsymbol{\alpha})-\mathbf{Y}(\mathbf{x}_0{}^b)$ denote the observation increment and the simulated observation increment respectively.

A linear approximation could be made in the vicinity of from Taylor expansion:

$$\mathbf{Y}'(\mathbf{P}_x\boldsymbol{\alpha})=\mathbf{Y}'(\sum\alpha_i\mathbf{x}_i{}')\approx\sum\alpha_i\mathbf{Y}'(\mathbf{x}_i{}')=\mathbf{P}_Y\boldsymbol{\alpha}$$

where $\boldsymbol{P_Y}$ is the ensemble of the simulated observation increments corresponding to $\boldsymbol{P_x}$. Thus, the model term $\boldsymbol{Y'(P_x\alpha)}$ is replaced with the linear statistic relation $\boldsymbol{P_Y\alpha}$ and the cost function is simplified as a quadratic polynomial:

$$J(\boldsymbol{\alpha})=\frac{1}{2}\{\boldsymbol{\alpha}^T\mathbf{B}_\alpha{}^{-1}\boldsymbol{\alpha}+[\mathbf{Y}^o{}'-\mathbf{P}_Y\boldsymbol{\alpha}]^T\mathbf{O}^{-1}[\mathbf{Y}^o{}'-\mathbf{P}_Y\boldsymbol{\alpha}]\}$$

DRP-4DVar

Like EnKF, **B** can be estimated by the ensemble, $\mathbf{B}\approx\mathbf{P}_x\mathbf{B}_\alpha\mathbf{P}_x{}^T\approx\frac{1}{N-1}\sum(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T$

If $\mathbf{x}_b=\bar{\mathbf{x}}$ (which is the case in the following experiments), $\mathbf{B}_\alpha\approx\frac{1}{N-1}\mathbf{I}$

As the cost function of DRP-4DVar is explicit and quadratic, and usually the dimension of the ensemble is very small (~100) unless being localized, we can simply solve the linear equations $\nabla J=0$ and get the analysis $\boldsymbol{x_a}=\boldsymbol{x_b}+\boldsymbol{P_x\alpha}^*$

ii) DRP-4DVar with Nonlinear Correction

As shown in last subsection, a linear approximation is made. The approximation simplifies the solution very effectively, however it may cause our analysis suboptimal, especially when the observation operator or the model has strong nonlinearity or the analysis increment is no longer small. To alleviate this problem, we adopt the concept of "inner loop/outer loop" from the incremental approach (Courtier et al. 1994) and extends the DRP-4DVar with a nonlinear correction process, forming the NC-DRP-4DVar. In this extension, the original DRP-4DVar, which is solved directly and easily but with linear approximation, is treated as a special inner loop. In the outer loop, the background and the ensemble are updated by the nonlinear model to refine the results from the inner loop iteratively.

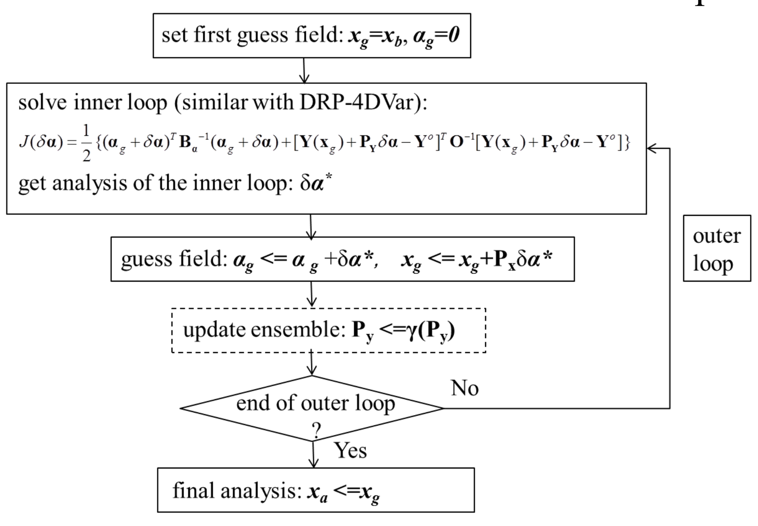


Fig1. schematic diagram of NC-DRP-4DVar

From Fig.1, we can see the two points of NC-DRP-4DVar:

a. The original DRP-4DVar is a special case of NC-DRP-4DVar with one outer loop;

b. The strategy to update ensemble $\mathbf{P_y}$ is undetermined. Two strategies are tested here:

1> **No Updating strategy**: $\gamma(\mathbf{P_y})=\mathbf{P_y}$, no actual updating is performed, so it's economical but empirical

2> **Re-Integrating strategy**: update $\mathbf{P_y}$ by re-integrate $\mathbf{P_x}$, it's theoretical correct but very time-consuming !

3. Experiments basic settings

- observing system simulation experiment (OSSE)
- Lorenz96 model

$$\frac{dx_j}{dt}=(x_{j+1}-x_{j-2})\,x_{j-1}-x_j+F$$

(j=1,2,..,M) M=40; F=8.0;
Fourth-order Runge-Kutta scheme; dt=0.05（~6h）

- Assimilation window: 4 steps (~24h), i.e. t=0,1,2,3
- observations: perturbed the truth with uncorrelated Gaussian noises (variance=0.16); available at all grids in t=0/3 (begin /end of the window)
- Ensemble number: ns=100 (the ensemble number is enough so as to isolate the error caused by linear approximation from the sampling error and truncation error in the dimension-reduction projection)
- No localization (as the ensemble number is sufficient to estimate **B** correctly)

- Assimilation Schemes

4DVar (incremental approach + adjoint model; **inner loop**: conjugate gradient method, terminate: $\|\nabla J\|$ reduces more than 90% , max iterations =12; **outer loop** iterations =5

DRP-4DVar (directly solve $\nabla J=0$)

NCDRP-NU(NCDRP-4DVar with 5 outer loop iterations, use 'No-Updating' strategy)

NCDRP-RI(NCDRP-4DVar with 5 outer loop iterations, use 'Re-Integrating' strategy)

- 2 sets of experiments, both with 30-window

CTRLXB: xb using a control run

CYCLEDA: cycling assimilation

4. Experiments — CTRLXB

Design: a set of 30-window assimilation experiments; background is from a control run (starts with the initial background and integrates freely for 30 windows), so assimilations among different windows are independent; **B** of 4DVar is an identity matrix multiplies an inflator: $\mathbf{B}=\lambda\mathbf{I}$; $\mathbf{P_x}$ of DRP-4DVar related schemes is randomly generated with Gaussian distribution (with zero mean and covariance as **I**), independently among all windows; $\mathbf{B}_\alpha$ of DRP-4DVar related schemes is also inflated by λ; λ is from 'perfect' inflation: λ=RMSE($\mathbf{x_b}$), so the **B** or $\mathbf{P_xB_\alpha P_x}^T$ can approximate the variance of $\mathbf{x_b}$. To further reduce computational cost, an EOF decomposition is performed on $\mathbf{P_y}$ before assimilation and the number of EOF modes selected is set to 40.

Purpose: in this design, $\mathbf{P_x}$ is generated independently and $\mathbf{P_xB_\alpha P_x}^T$ is approximately the same with 4DVar's fixed B, so no global flow-dependence can be shown for DRP-4DVar and DRP-4DVar can not be better than 4DVar. 4DVar is a reference to show how DRP-4DVar is affected by the linear approximation and how NC-DRP-4DVar schemes can improve its performance.

Results:

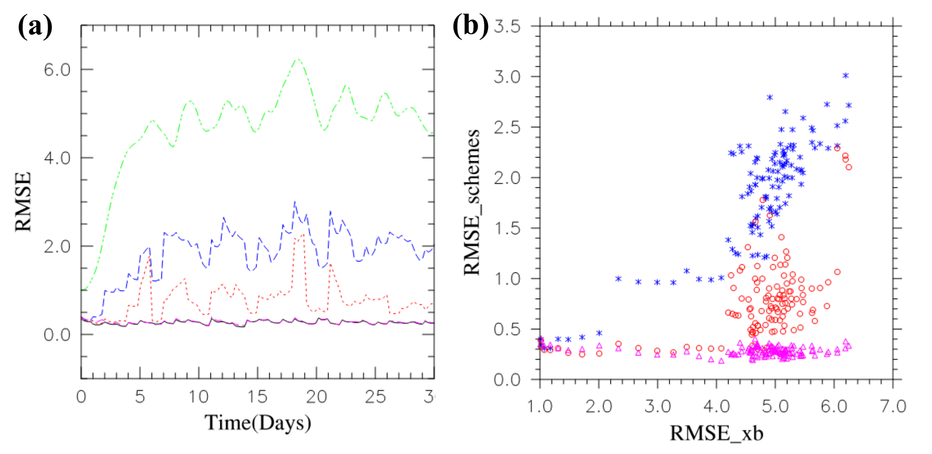


Fig.2 RMSE of
4DVar (black, only in (a));
DRP-4DVar(blue);
NC-DRP-NU(red);
NC-DRP-RI (pink);
Xb (green, only in (a)).
(a)in time-series view
(b)Scatter of (xb_rmse, xa_rmse)

Points:

1. Compared with 4DVar, DRP-4DVar's analysis RMSE is much larger and it increases with the increase of xb RMSE(see Fig.2(b)). This can be attributed to the **linearity approximation** in its solution. If the background is inadequately provided, the analysis increments $\delta\boldsymbol{x}_a$ tend to be larger in magnitude so the effect of nonlinearity tends to be more significant and poor performances are more likely to emerge
2. Both the two NCDRP schemes **reduce the RMSE significantly**. Especially for the NCDRP-RI, its RMSE almost achieves the same level with that of the 4DVar, also suggesting the correctness of our theorem.
3. Considering the **computational cost**: given $\mathbf{P_x/P_y/x_b/y_b}$, DRP-4DVar need no model runs, NCDRP-NU nees 4 model runs (each outer loop needs one model run), NCDRP-RI needs about 4*40 model runs (each outer loop needs update the whole ensemble), 4DVar needs about 30~50 tangent linear model runs & ajoint model runs in its inner loops and 4 nonlinear model runs in its outer loops.

A mixture between 'No Updating' & 'Re-Integrating'

To further improve the performance of NCDRP-NU and reduce the computational cost of NCDRP-RI, a mixed strategy is proposed, i.e., use 'Re-Integrating' strategy in the first X outer loops while use 'No Updating' in the rest outer loops, namely '**NCDRP-RIX**'.

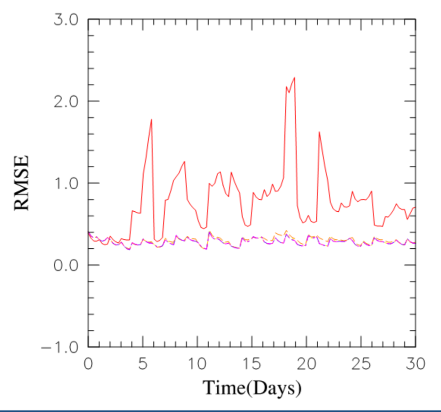


Fig.3 RMSE of
NCDRP-NU(red);
NCDRP-RI1(orange);
NCDRP-RI2(yellow);
NCDRP-RI3(purple);
NC-DRP-RI (pink);

As in Fig. 3, NCDRP-RI1 gets a similar performance with NCDRP-RI while it only updates the ensemble once and reduces 75% of the computation of NCDRP-RI.

5. Experiments — CYCLEDA

Unlike CTRLXB, CYCLEDA uses real-time updated background and ensemble, so the advantages of the DRP-4DVar related schemes, i.e. the global flow-dependent covariance, could be shown. As a comparison, 4DVar is also conducted, with the fixed **B** is the time average of ensemble-based covariances of DRP-4DVar, multiplied with a tuned inflator. As the background is of good quality in the cycling experiments, the nonlinear effects are small, only the 'No Updating' strategy is applied.

time-mean RMSE	4DVar	DRP-4DVar	NCDRP-NU
Basic settings	0.21	0.12	0.11
Nonlinear obs. $y=x^2$	0.08	0.05	0.02
Long window 8 steps (48h)	0.2	0.15	0.13

DRP-4DVar has a better performance over 4DVar under different settings. This can be attributed as the effect of global flow-dependence of ensemble methods. In spite of the good performance of DRP-4DVar, the nonlinear correction process can further reduce the RMSE and shorten the spin-up time (figure not shown) .

6. Summary

In this study, we prove that the DRP-4DVar is limited by its linear approximation and therefore we extend it with a nonlinear correction process, forming the NCDRP-4DVar. In the implementation of the NCDRP-4DVar, two strategies to update the ensemble are introduced: the economical but empirical 'No-Updating' strategy and the theoretically correct but time consuming "Re-Integrating" strategy. The latter needs to integrate all members of the ensemble for each outer loop iteration, while the former needs no extra computation.

We use the Lorenz96 model to test our methods. Two kinds of OSSE experiments are carried out, CTRLXB and CYCLEDA. In CTRLXB both two strategies for NCDRP-4DVar show significant improvements due to the correction of nonlinear effect, especially for the 'Re-Integrating' strategy, its RMSE is almost the same with that of 4DVar, which proves our theorem correctness. To reduce the computational cost of 'Re-Integrating' strategy, we suggest a mixed strategy 'RIX', in which 'Re-Integrating' strategy is only used in first X outer loops while 'No-Updating' is used in the rest. Perfect performance is shown with strategy 'RI1', which costs only 25% computations of that of NCDRP-RI. In CYCLEDA, which is also the operational case, the ensemble integrates so that the global flow-dependence of DRP-4DVar is retained. As the result, DRP-4DVar outperforms 4DVar under all the different settings. NCDRP-4DVar also has a stable improvement over DRP-4DVar.

Main References:

1. Wang, B., J. Liu, S. Wang, W. Cheng, L. Juan, C. Liu, Q. Xiao, and Y.-H. Kuo, 2010: An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **27**, 715–727
2. Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol. Soc., 120, 1367–1387
3. Shen, S., J. Liu, B. Wang, 2014: An Extension of the Dimension-Reduced Projection 4DVar , Atmos. Oceanic Sci. Lett. (Accepted)