ENSEMBLE KALMAN FILTER IN THE PRESENCE OF MODEL ERRORS

Hong Li[1], Eugenia Kalnay[1], Takemasa Miyoshi[2] and Christopher M. Danforth[1]

[1]University of Maryland

[2]Japan Meteorological Agency

AMS annual meeting

January 2007

Problems

- Errors in numerical forecasts arise due to errors in the initial conditions and the model deficiencies.

- A large effort has been made to deal with the IC problem through the process data assimilation (DA)

  —3DVAR, 4DVAR, Kalman Filter. With time, errors in the IC have been much reduced.

- Accounting for model deficiencies has become crucial for data assimilation and ensemble forecasting.

Model error estimation schemes (1)

1. Covariance inflation

$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{P}_{i-1}^a \mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q}$$ (Ideal KF)

$$\mathbf{P}_i^f = \frac{1}{k-1} \sum_{i=1}^{K} (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$ (EnKF)

$$\widetilde{\mathbf{P}}_i^f = (1+\Delta) * \mathbf{P}_i^f = \mathbf{P}_i^f + \Delta \mathbf{P}_i^f$$

$\Delta \mathbf{P}_i^f \rightarrow \mathbf{Q}$

Model error estimation schemes (2)

2. Dee and daSilva bias estimation scheme (1998)

Do data assimilation twice:

first for model error

$$b_t^f = \mu b_{t-1}^a$$

$$b^a = b^f - L[y^o - (Hx^f) - Hb^f)]$$

$$L = P^{bias} H^T (HP^{bias} H^T + HP^f H^T + R)^{-1}$$

then for model state (expensive)

$$\tilde{x}^f = x^f - b^a$$

$$x^a = \tilde{x}^f + K[y^o - H\tilde{x}^f]$$

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

$$P^{bias} = \alpha * P^f$$

$$\mathbf{P}^f = \frac{1}{k-1} \sum_{i=1}^{K} (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$

$$0 < \mu, \alpha \le 1$$

and need to be tuned

Model error estimation schemes (3)

3. Low-order scheme (Danforth et al, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev*)

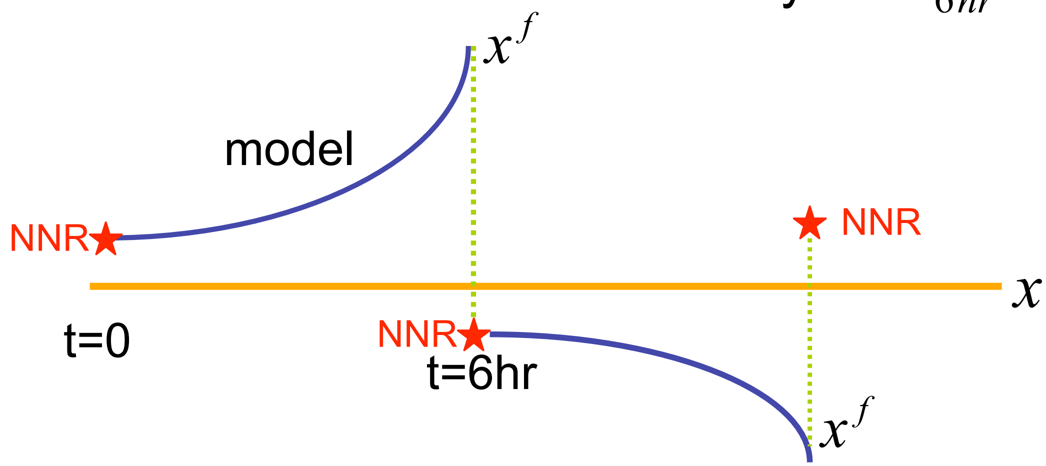
- Generate a long time series of model forecast minus reanalysis x^e_{6hr} from the training period

$$\mathbf{b} = \overline{x^e_{6hr}}$$

$$x^e_{6hr}{}' = x^e_{6hr} - \overline{x^e_{6hr}}$$

$$\mathbf{e}_l = EOF(x^e_{6hr}{}')$$

$$\mathbf{f}_m = SVD[(x^e_{6hr}{}'),(x^f{}')]$$



✓ Danforth et al 2007 did not compute the IC errors. Here we are concerned with both the IC error and the model error:

$$\varepsilon^f_{n+1} = \mathbf{x}^f_{n+1} - \mathbf{x}^t_{n+1} = \boxed{M(\mathbf{x}^a_n) - M(\mathbf{x}^t_n)} + \boxed{\mathbf{b} + \sum_{l=1}^{L} \beta_{n,l}\mathbf{e}_l + \sum_{m=1}^{M} \gamma_{n,m}\mathbf{f}_m}$$

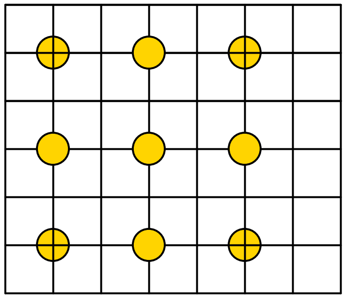
Forecast error due to error in IC | Time-mean model bias | Diurnal model error | State dependent model error

SPEEDY MODEL (Molteni 2003)

- T30L7 global spectral model
- total 96x48 grid points on each level
- State variables u,v,T,Ps,q

Data Assimilation: Local ensemble transform Kalman filter (LETKF, Hunt 2006)

Dense Observation



OBSERVATIONS

- Generated from the NCEP reanalysis plus “random errors”

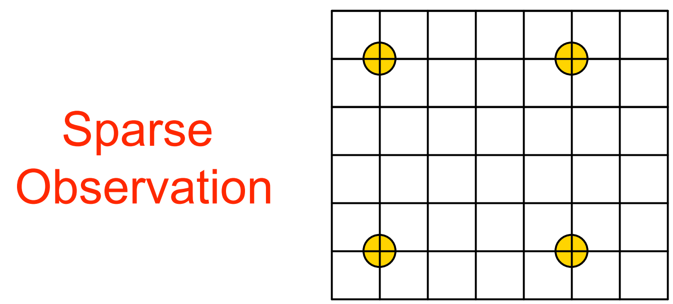
  - assume NCEP reanalysis approximates the real atmosphere, whereas the SPEEDY has its own biased climatology.

- Dense observation network: 1 every 2 grid points in both x and y direction:

SPEEDY MODEL (Molteni 2003)

- T30L7 global spectral model
- total 96x48 grid points on each level
- State variables u,v,T,Ps,q

Data Assimilation: Local ensemble transform Kalman filter (LETKF, Hunt et al. 2006)



OBSERVATIONS

- Generated from the NCEP reanalysis plus “random errors”

  - assume NCEP reanalysis approximates the real atmosphere, whereas the SPEEDY has its own biased climatology.

- Sparse observation network: 1 every 4 grid points in both x and y direction:

Experimental Design:

➢Experimental period : 1987 Jan & Feb

➢ LETKF with 20 ensemble members

➢Control run: Assimilate observations created from NCEP reanalysis with LETKF but without estimating the model errors.

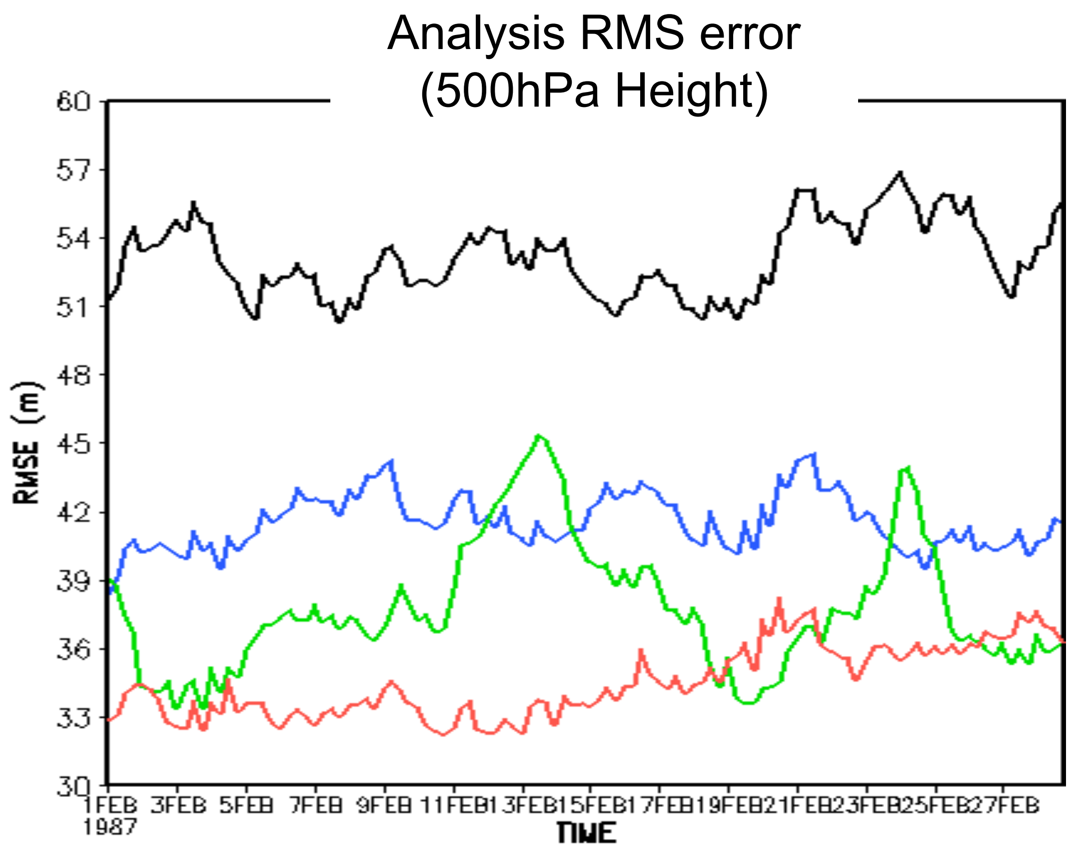
➢Model error correction Experiments: Apply different model error correction methods at each analysis cycle (6-hour)

1. Inflation
2. Dee&daSilva (tune parameters)
3. Low-order

$$\mathrm{mod}\,el\_error = \mathbf{b} + \sum_{l=1}^{L} \beta_{n,l}\mathbf{e}_l + \sum_{m=1}^{M} \gamma_{n,m}\mathbf{f}_m$$

- First test: only correct the time-mean bias
- Training period: One month prior to the experiment period

Dense Observation



Control run

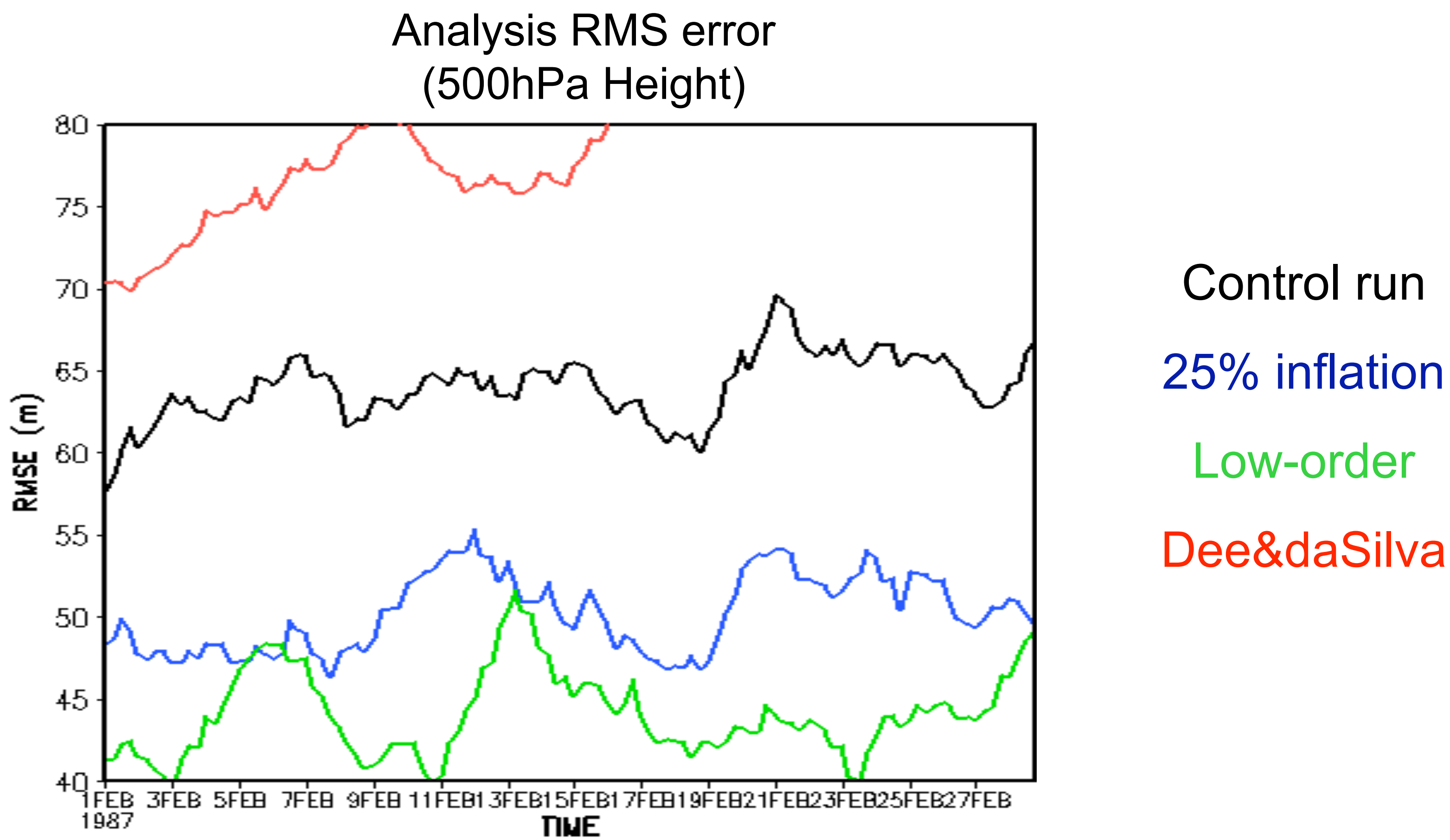
25% inflation

Low-order

Dee&daSilva

Dense observation network: All schemes are better than the control run, Dee&daSilva gives best results (but it is expensive)

Sparse Observation



Sparse observation network: Dee&daSilva makes the filter divergent
low-order gives the best results.

Further explore the Low-order scheme:

Correct the Diurnal and the state-dependent model errors:

$$model\_error = \mathbf{b} + \sum_{l=1}^{10} \beta_{n,l}\mathbf{e}_l + \sum_{m=1}^{10} \gamma_{n,m}\mathbf{f}_m$$

Time-mean model bias

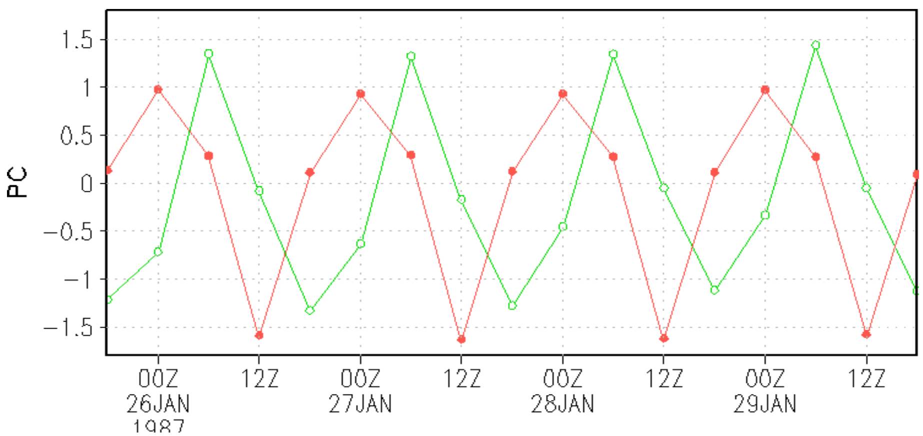
Diurnal model errors

- Generate the leading EoFs from the forecast error anomalies fields for temperature.

$$x^{e\,'}_{6hr(i)} = x^{e}_{6hr} - \overline{x^{e}_{6hr}}$$

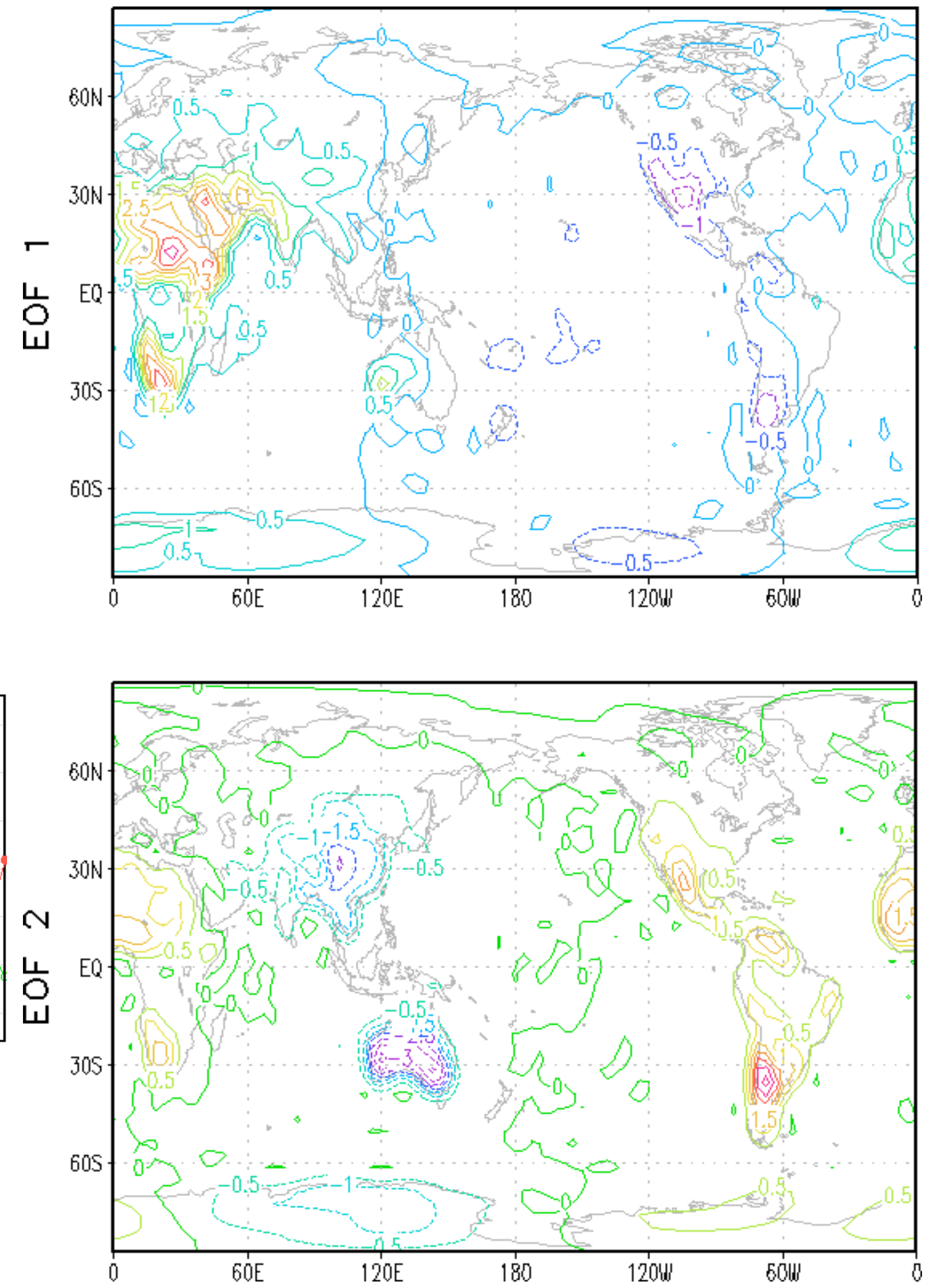
$$EOF[x^{e\,'}_{6hr(i)}]$$

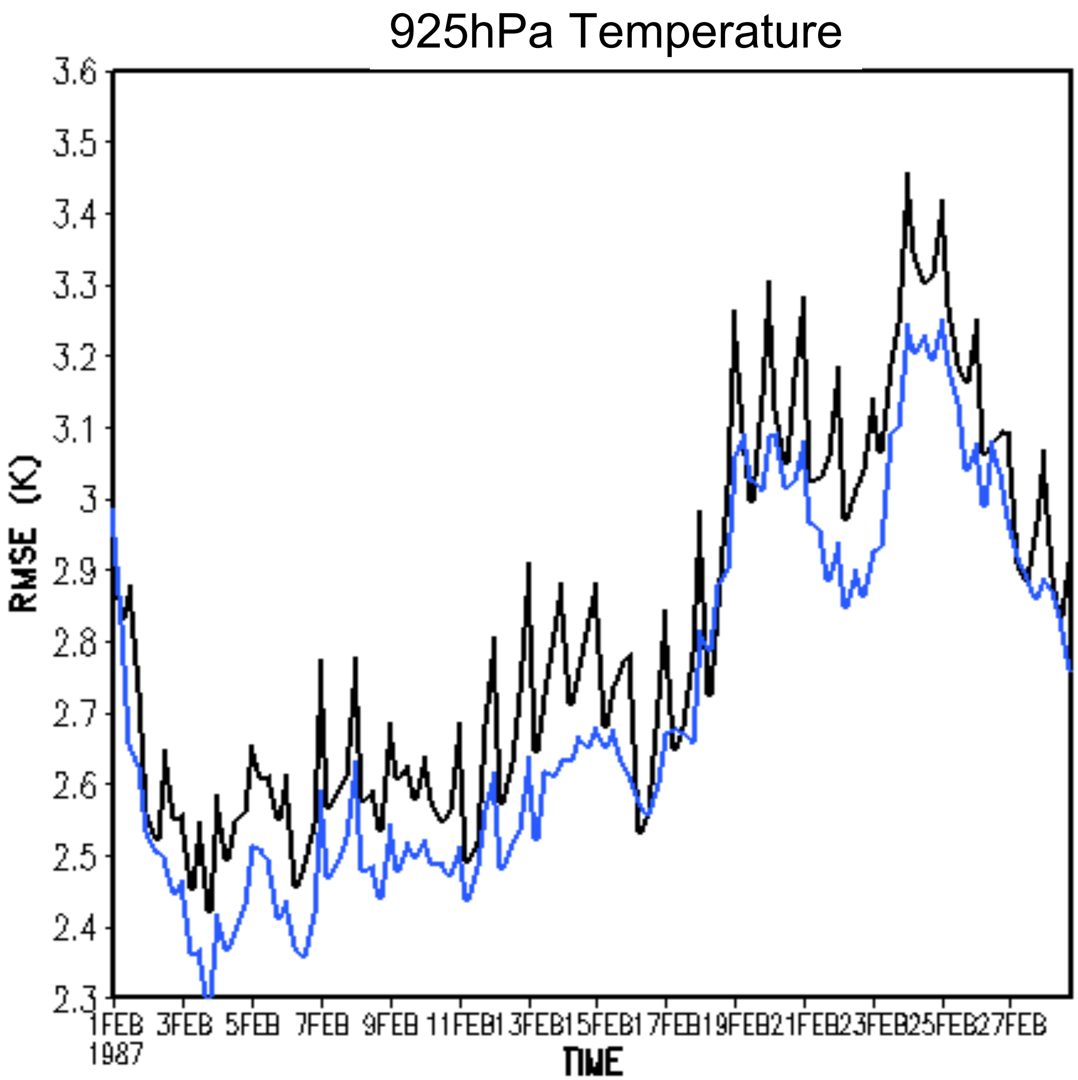
pc1
pc2



➢Lack of diurnal forcing generates wavenumber 1 structure

Leading EOFs for 925 mb TEMP





Black line:

$$\tilde{x}^f = x^f - \mathbf{b}$$

Blue line:

$$\tilde{x}^f = x^f - (\mathbf{b} + \sum_{l=1}^{10} \beta_{n,l} \mathbf{e}_l)$$

State-dependent model errors

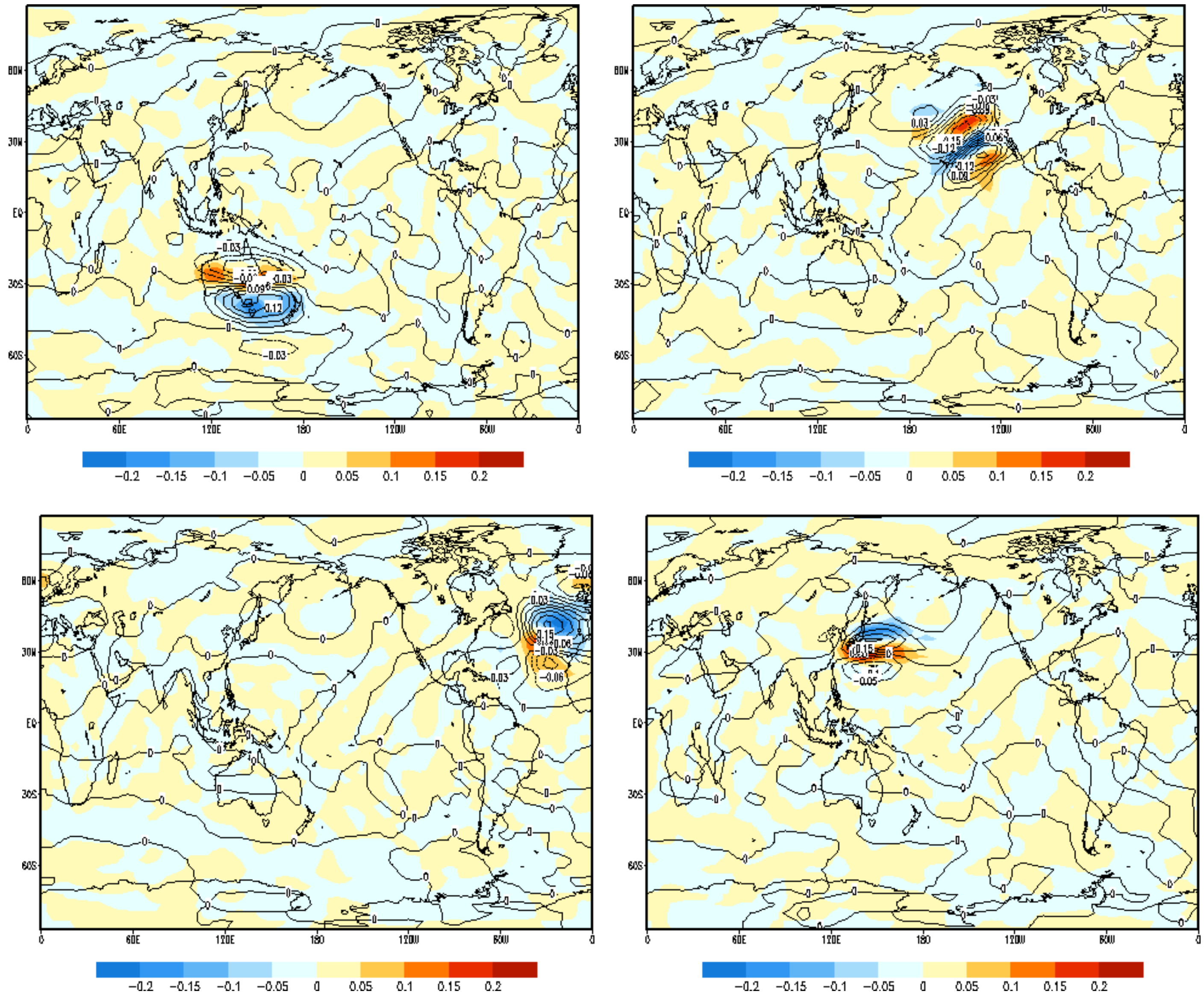
the local state anomalies (Contour) and the forecast error anomalies (Color)

SVD1

SVD2

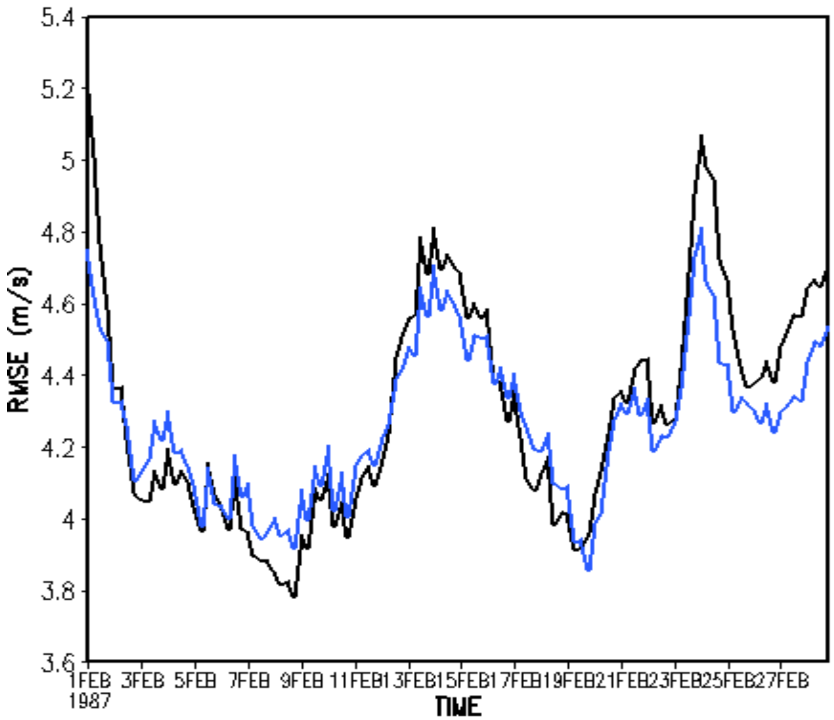
SVD3

SVD4

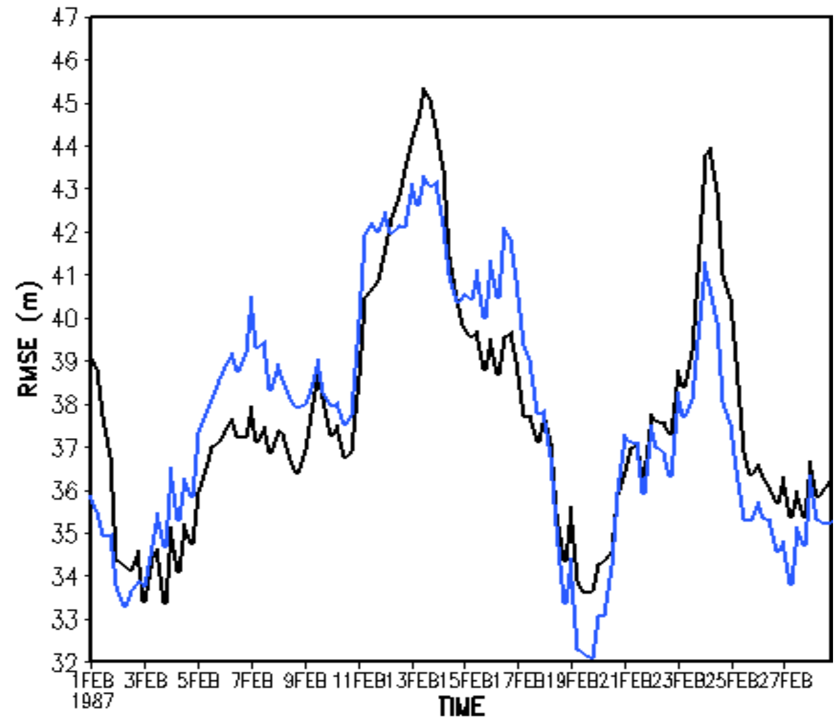


Correct state-dependent model errors

500hPa Uwind



500hPa Height



Black line: $\tilde{x}^f = x^f - \mathbf{b}$

Blue line: $\tilde{x}^f = x^f - [\mathbf{b} + \sum_{m=1}^{10} \gamma_{n,m} \mathbf{f}_m]$

Univariate SVD (not account for the relations between different variables)

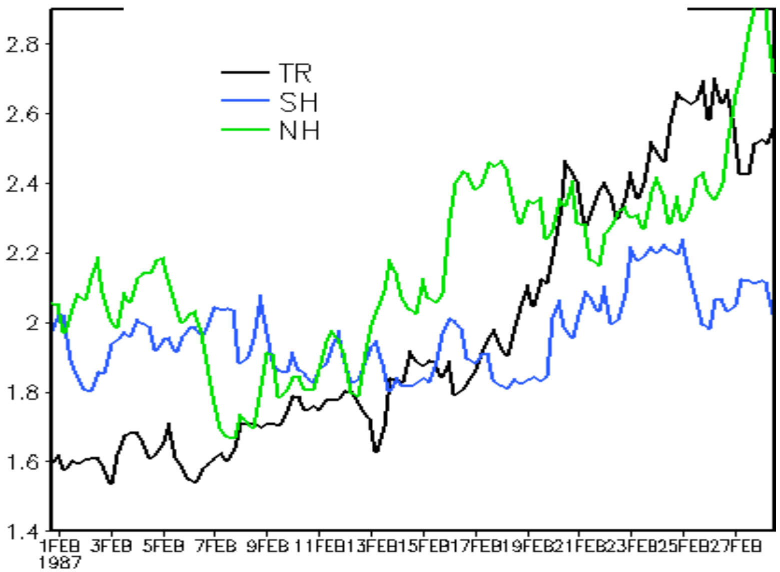
Summary

- For dense observations, all of the methods work well. Dee&daSilva is better than the other two (but expensive).
- However, for sparse observation, Dee&daSilva makes the filter diverge. By simply subtracting the constant mean bias from the background fields Low-order method still works well.
- For Low-order method, correcting the diurnal and state-dependent model errors further improves the analysis accuracy.

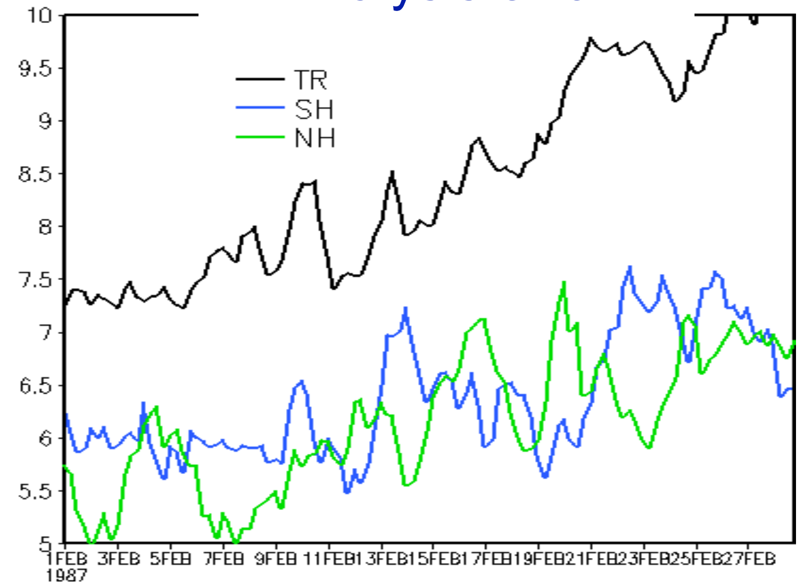
Back-up

500hPa Uwind

Estimated model bias

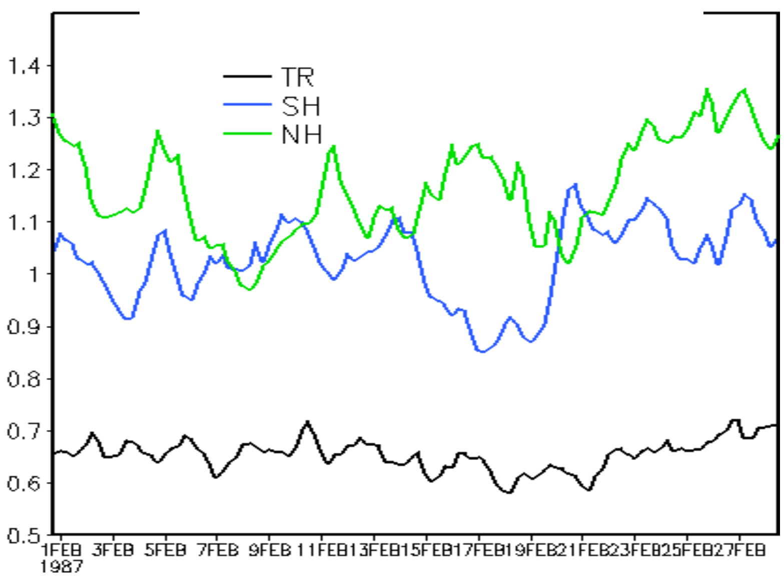


Analysis error



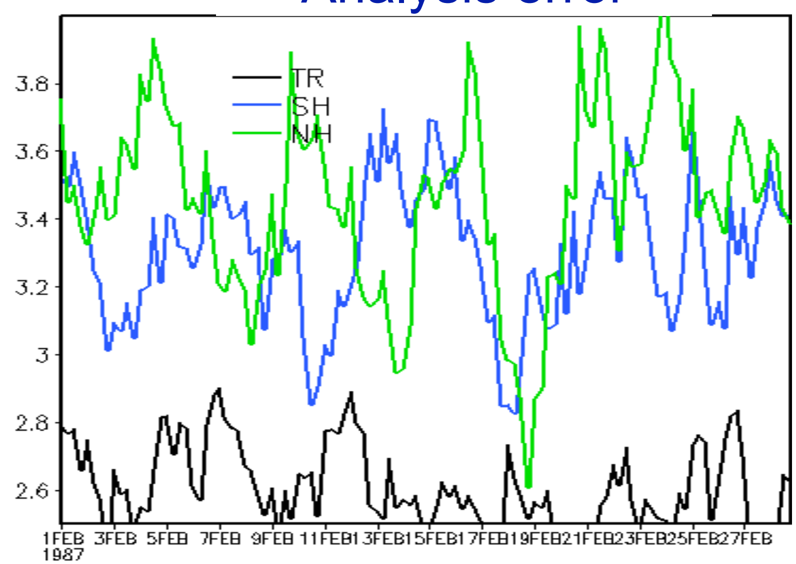
Sparse

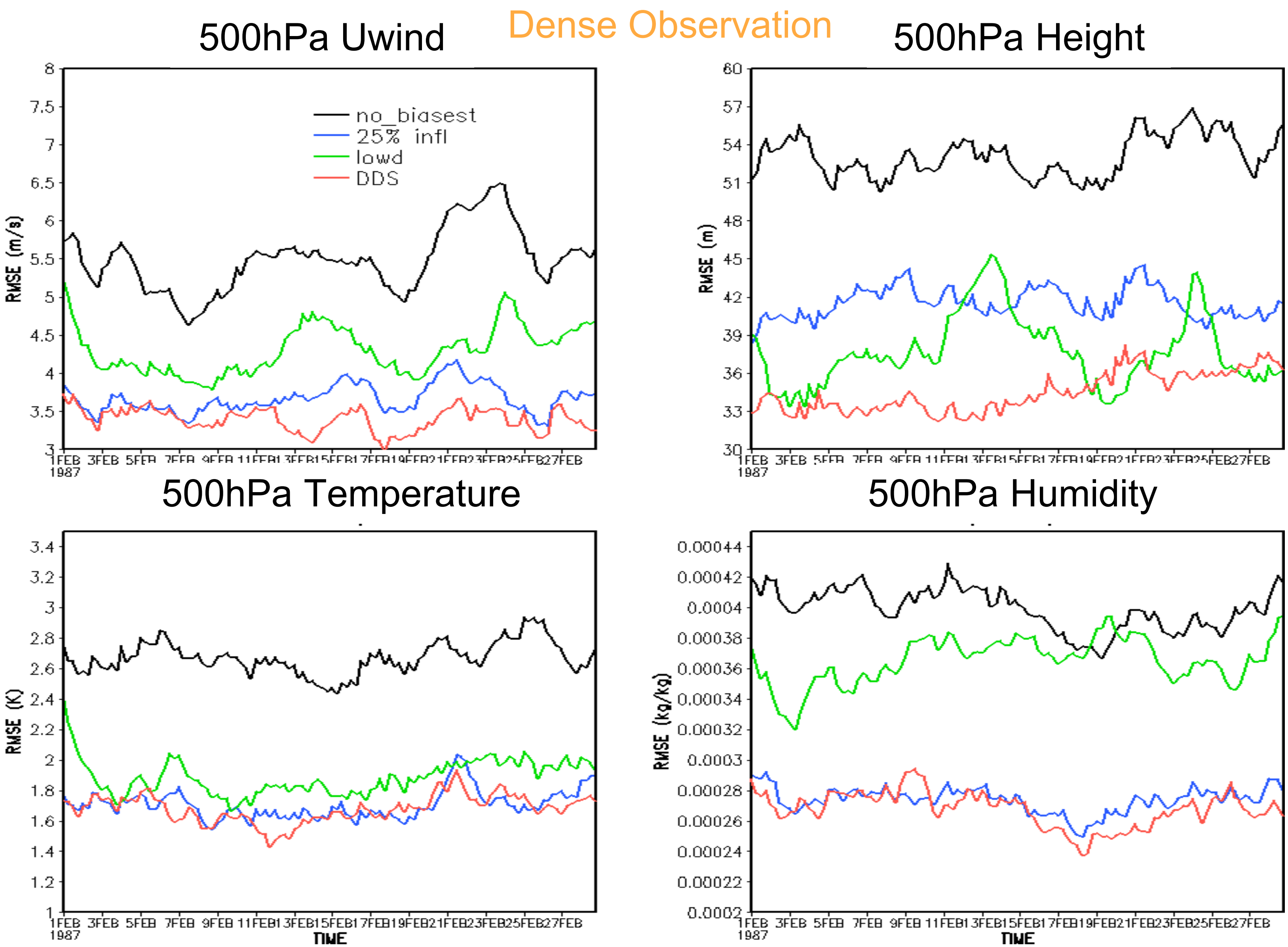
Estimated model bias



Dense

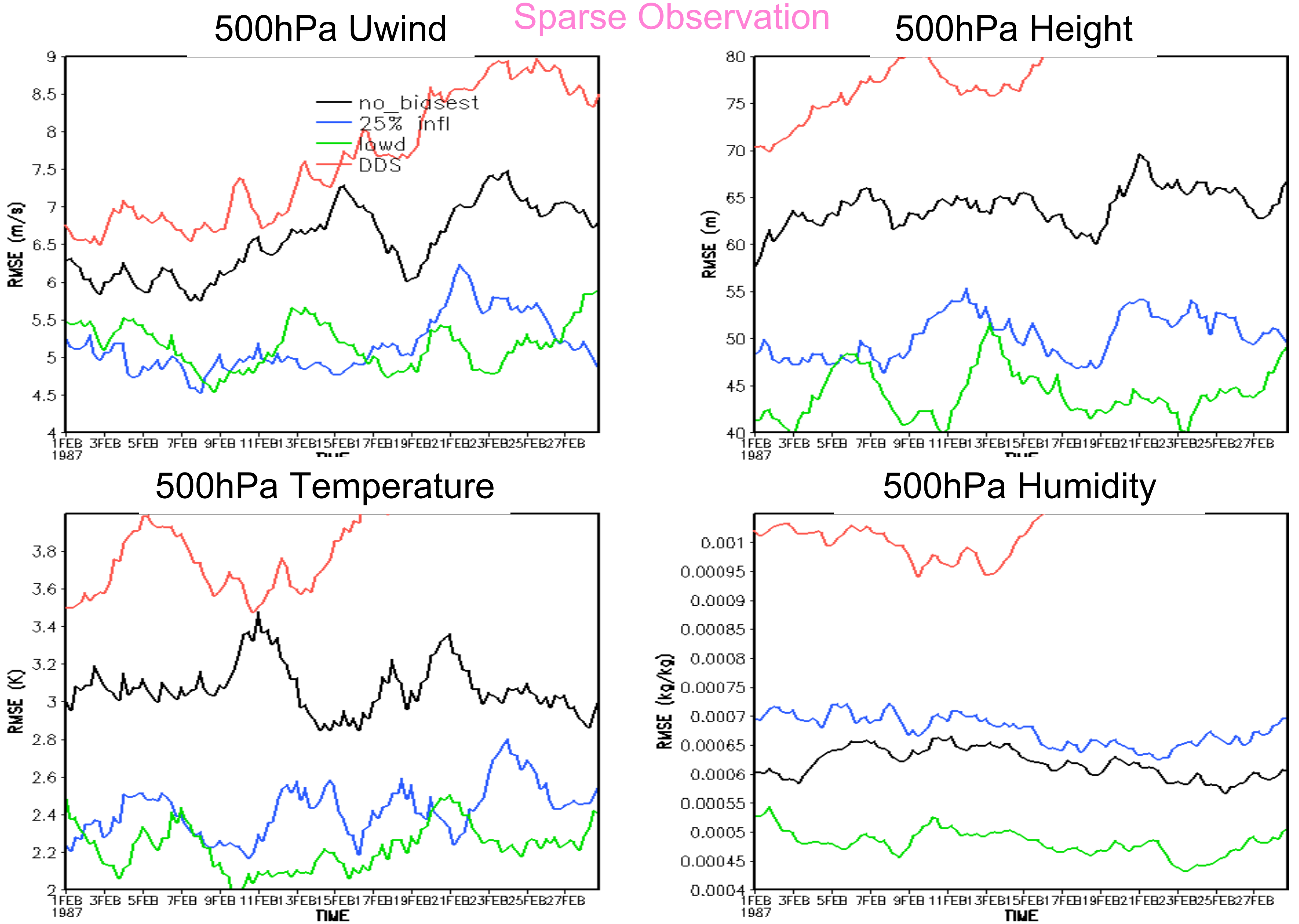
Analysis error





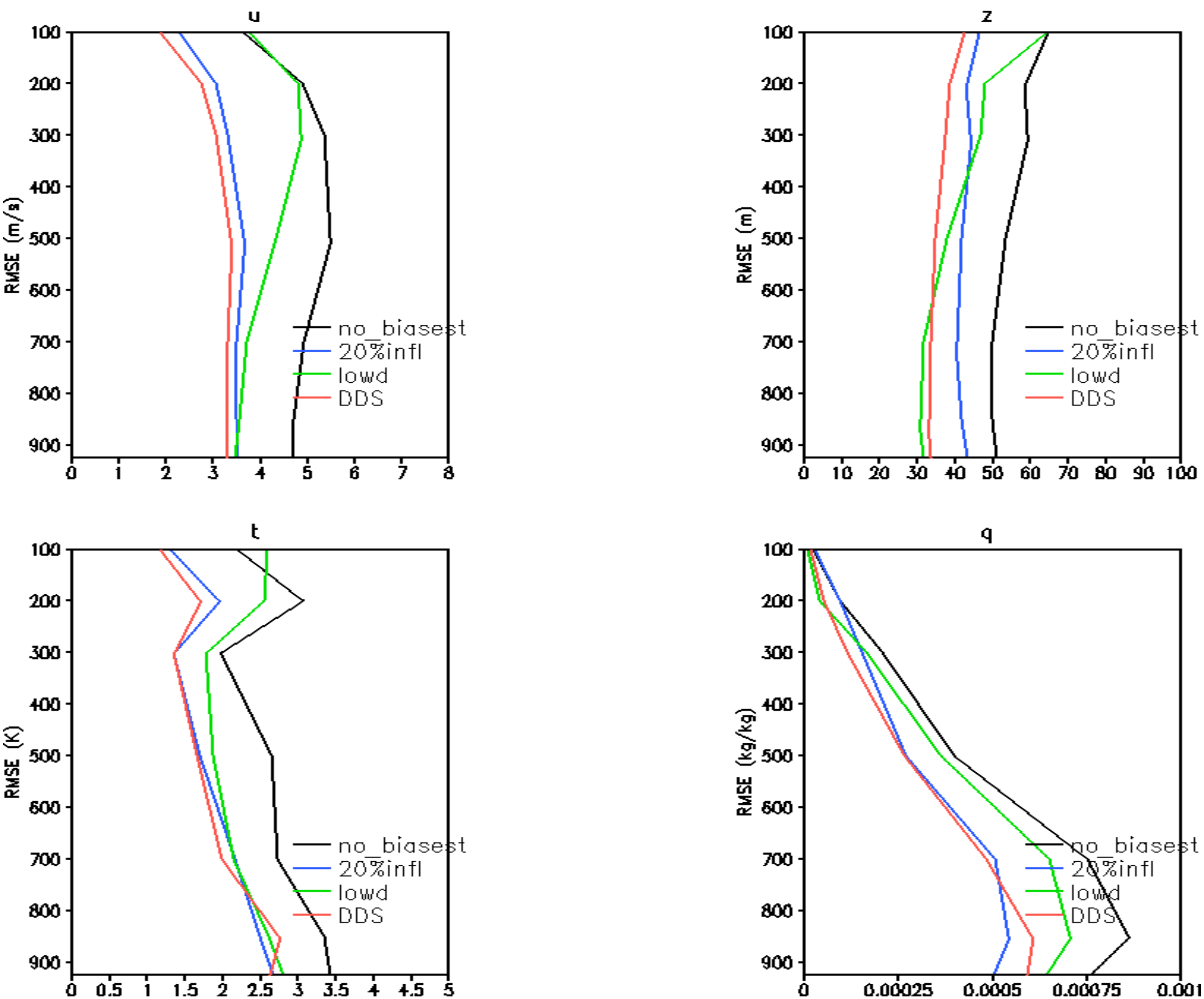
Dense observation network: All schemes are better than the control run, DdS gives best results (but it is expensive)

Sparse Observation



Sparse observation network: DdS makes the filter divergent

low-order gives best results.


u
RMSE (m/s)
100
200
300
400
500
600
700
800
900
0
1
2
3
4
5
6
7
8
no_biasest
20%infl
lowd
DDS
z
RMSE (m)
0
10
20
30
40
50
60
70
80
90
100
t
RMSE (K)
0.5
1.5
2.5
3.5
4.5
q
RMSE (kg/kg)
0.00025
0.0005
0.00075
0.001

Impact of model errors

Analysis rms error (SPEEDY model)

(perfect model exp .vs. imperfect model exp)