An Compensatory Approach of the Fixed Localization in the EnKF

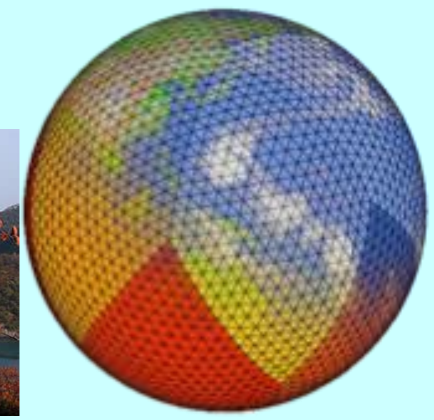
Zhao Juan

Data Assimilation

Observations

Data Assimilation

Numerical models



Data assimilation best combines observations and a model, and brings synergy.

3DVAR, 4DVAR, Ensemble Kalman Filter, Particle Filter , etc.

Ensemble Kalman Filter (EnKF)

Kalman Filter(KF)

$$\mathbf{x}_t^f = \mathbf{M}(\mathbf{x}_{t-1}^a)$$

$$\mathbf{P}_t^f = \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}_t$$

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

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}_t(\mathbf{x}_t^f))$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}\mathbf{H}_t)\mathbf{P}_t^f$$

Ensemble Kalman Filter(EnKF)

$$\mathbf{P}_t^f = \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}_t$$

$$\mathbf{P}_t^f \approx \frac{\delta\mathbf{X}(\delta\mathbf{X})^T}{N_m - 1}$$

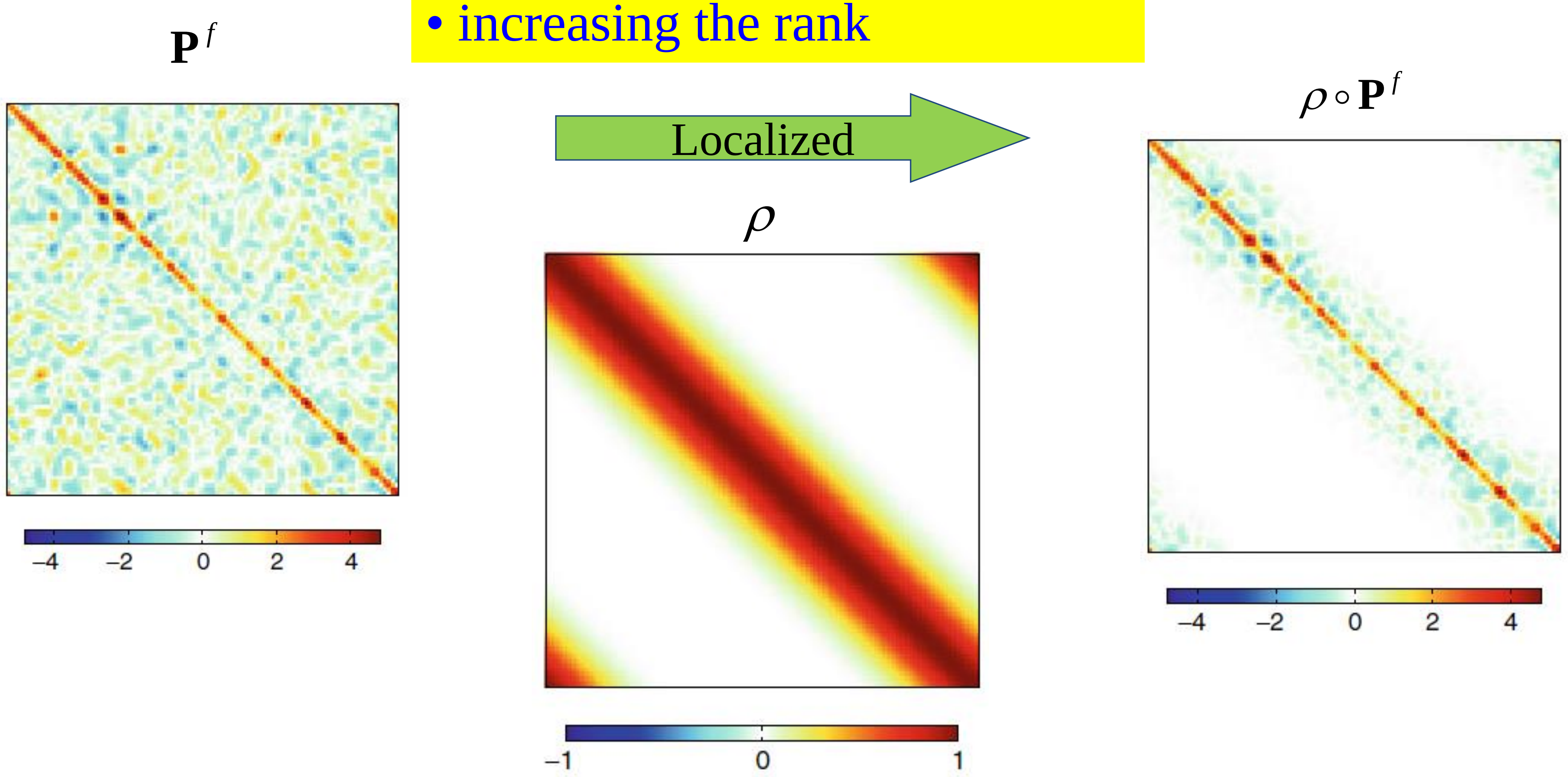
$$\delta\mathbf{X} = \mathbf{x}_{i,t}^f - \overline{\mathbf{x}}_t^f, i = 1, 2, \ldots, N_m$$

Ensemble Kalman Filter (EnKF)

- Spurious correlations
- Rank problem
- Filter divergence

Covariance Localization

Empirical treatment for…
- reducing sampling noise
- increasing the rank

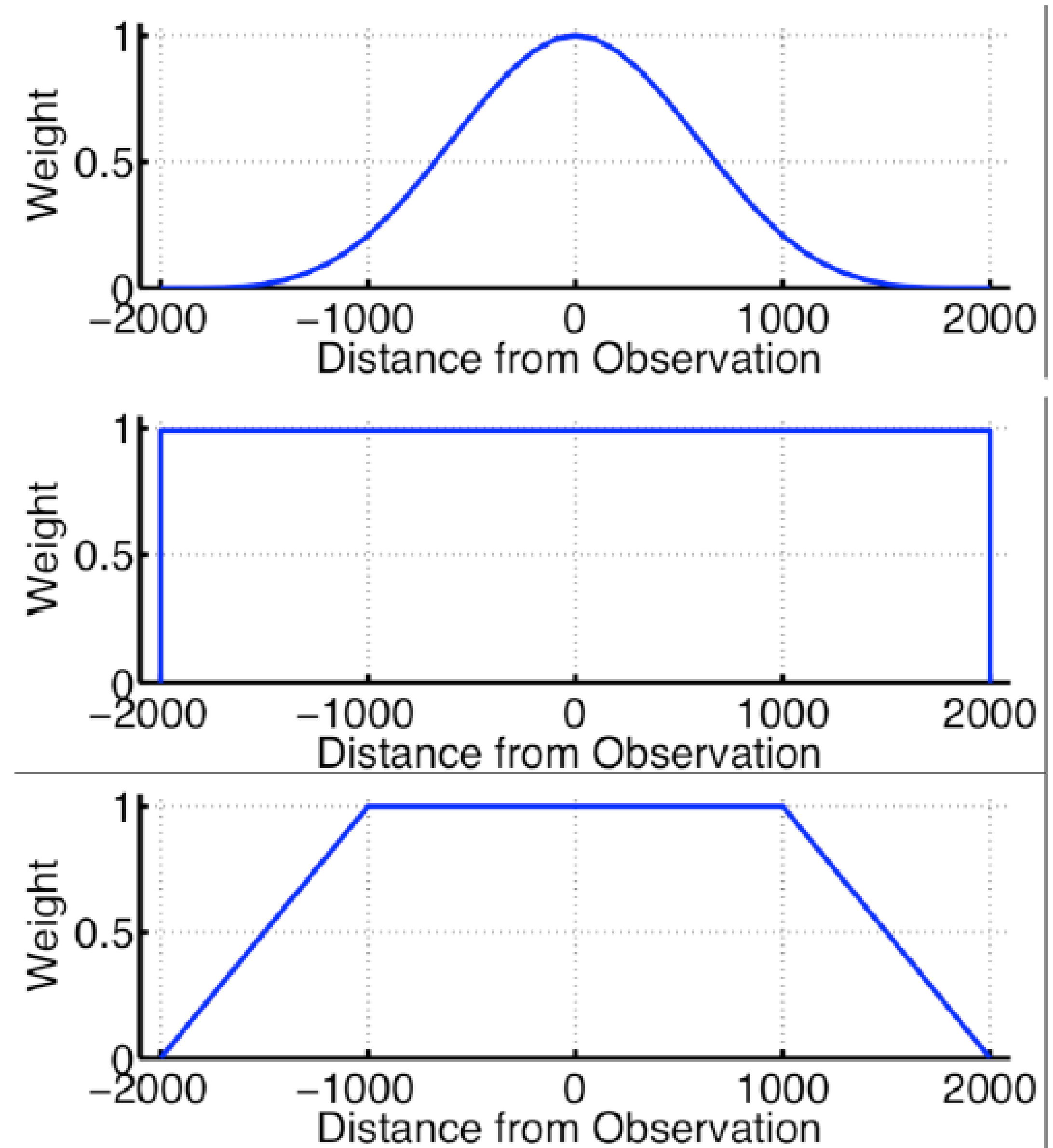


Localized Operator

1=> Gaspari-Cohn

2=> Boxcar

3=> Ramped Boxcar



Difficulties of localization

- How to define the halfwidth?
- How to define the distance?
- Flow dependent loss
- Imbalance

Difficulties of localization

- How to define the halfwidth?
- How to define the distance?
- Flow dependent loss
- Imbalance

EnKF-MSA

- Retrieve multiple-scale information by using multiple-scale analysis (MSA) technique to from the observational Residuals

- Update ensemble mean by adding the analysis fields to the ensemble mean of EnKF

Muliple-scale Analysis

- Multiple-scale analysis (MSA)

$$\boxed{\mathbf{x}_t^{res} = \mathbf{x}^t - \overline{\mathbf{x}}_{EnKF}}$$

True residual

$$\boxed{\mathbf{y}_t^{res} = \mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}_{EnKF}}$$

Observation residual

For overly large and overly small halfwidth values in fix localization, **observation residual** still contains some multiscale information of **true residual** to a certain extent.

Multiple-scale Analysis

- Under the framework of three-dimentional variational analysis

$$J^{(l)}(\delta\mathbf{x}^{(l)}) = \frac{1}{2}\delta\mathbf{x}^{(l)\mathrm{T}}(\mathbf{B}^{(l)})^{-1}\delta\mathbf{x}^{(l)} + \frac{1}{2}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)})^{\mathrm{T}}(\mathbf{R}_{\mathrm{MSA}}^{(l)})^{-1}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)}) + \frac{1}{2}\delta\mathbf{x}^{(l)\mathrm{T}}\mathbf{S}^{(l)}\delta\mathbf{x}^{(l)}, \quad l = 1, \ldots, L,$$

L : number of scale

$\delta\mathbf{x}^{(l)}$: the increment of the state vector $\mathbf{x}$

$\mathbf{L}^{(l)}$: the linear projection operator from the observation space to the state space

$\mathbf{B}^{(l)}$: the background error covariance

$\mathbf{R}^{(l)}$: the observation error covariance matrix

$\mathbf{d}^{(l)}$: the observation innovation vector

$\mathbf{S}^{(l)}$: the smoothing matrix

Muliple-scale Analysis

- For every level

$$\mathbf{d}^{(l)} = \begin{cases} \mathbf{y}^{o} - \mathbf{H}^{(0)}\mathbf{x}^{b}_{\mathrm{MSA}} & l = 1 \\ \mathbf{d}^{(l-1)} - \mathbf{H}^{(l-1)}\delta\mathbf{x}^{(l-1)}_{\mathrm{MSA}} & l = 2, \ldots, L \end{cases},$$

$$\delta\mathbf{x}^{(l)}_{\mathrm{MSA}} = [(\mathbf{B}^{(l)})^{-1} + \mathbf{S}^{(l)} + (\mathbf{R}^{(l)}_{\mathrm{MSA}})^{-1}]^{-1} \times (\mathbf{R}^{(l)}_{\mathrm{MSA}})^{-1}\mathbf{L}^{(l)}\mathbf{d}^{(l)}.$$

- Total increment of $\mathbf{x}$

$$\delta\mathbf{x}_{\mathrm{MSA}} = \sum_{l=1}^{L} \delta\mathbf{x}^{(l)}_{\mathrm{MSA}}.$$

Flowchart of the EnKF-MSA

$$
\begin{aligned}
J^{(l)}(\delta\mathbf{x}^{(l)}) = {} & \frac{1}{2}\delta\mathbf{x}^{(l)\mathrm{T}}(\mathbf{B}^{(l)})^{-1}\delta\mathbf{x}^{(l)} \\
& + \frac{1}{2}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)})^{\mathrm{T}}(\mathbf{R}_{\mathrm{MSA}}^{(l)})^{-1}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)}) \\
& + \frac{1}{2}\delta\mathbf{x}^{(l)\mathrm{T}}\mathbf{S}^{(l)}\delta\mathbf{x}^{(l)}, \quad l = 1, \ldots, L,
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
J^{(l)}(\delta\mathbf{x}^{(l)}) = {} & \frac{1}{2}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)})^{\mathrm{T}}(\delta\mathbf{x}^{(l)} - \mathbf{L}^{(l)}\mathbf{d}^{(l)}) \\
& + \frac{1}{2}\delta\mathbf{x}^{(l)\mathrm{T}}\mathbf{S}^{(l)}\delta\mathbf{x}^{(l)}, \quad l = 1, \ldots, L.
\end{aligned}
$$

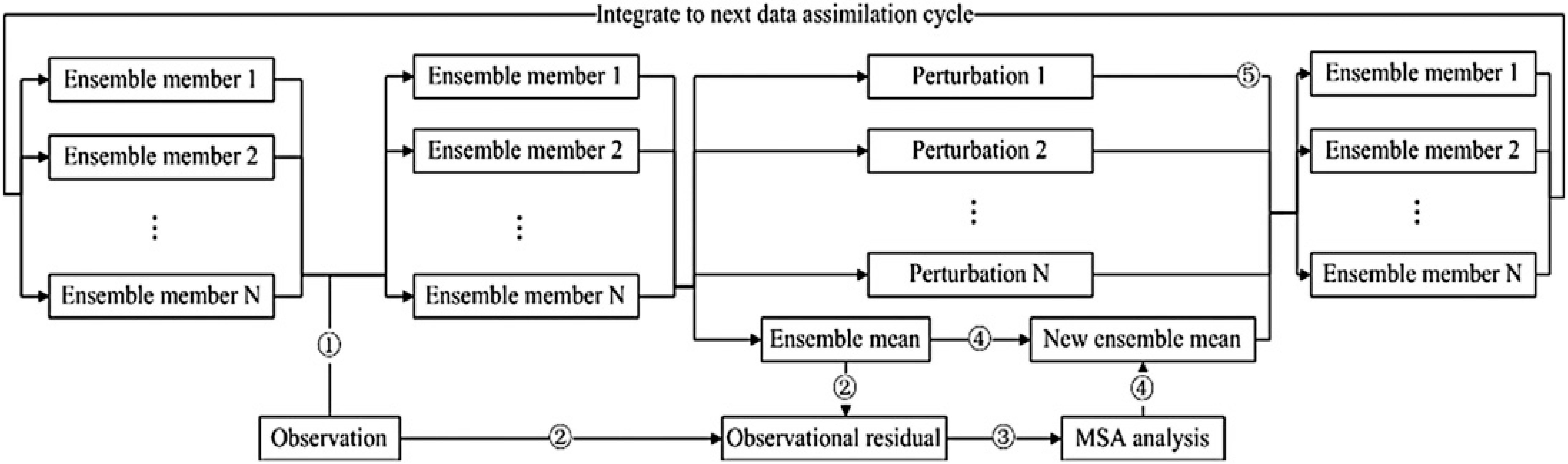
Update Ensemble mean

- Update Ensemble mean

$$
\begin{aligned}
\overline{\mathbf{x}}_{\mathrm{hybrid}} &= \overline{\mathbf{x}}_{\mathrm{EnKF}} + \delta\mathbf{x}_{\mathrm{MSA}} \\
&= \overline{\mathbf{x}}^{\mathrm{b}} + \hat{\mathbf{K}}(\mathbf{y}^{o} - \mathbf{H}\overline{\mathbf{x}}^{b}) + (\mathbf{S} + \mathbf{I})^{-1} \sum_{l=1}^{L} \mathbf{L}^{(l)}\mathbf{d}^{(l)} .
\end{aligned}
$$

a	$a^{(i)}$	No. of levels
125	125, 250, 500, 1000, 2000, 4000	6
250	250, 500, 1000, 2000, 4000	5
500	500, 1000, 2000, 4000	4
1000	1000, 2000, 4000	3
1500	1500, 3000, 4000	3
2000	2000, 4000	2
2500	2500, 4000	2

Flowchart of the EnKF-MSA



- **Step1:** **Adjusts the ensemble member using the observation with the standard EnKF**
- **Step2:** **Generate the observational residual**
- **Step3:** **Retrieves the multiscale information**
- **Step4:** **Adds the analysis field generated by MSA to the ensemble mean**
- **Step5:** **Generate the final ensemble members**

Biased twin-experiment setup

- Model : a global barotropical spectral model

$$\frac{d}{dt}\left(\frac{f+\zeta}{H}\right)=0,$$

- model error is assumed to arise from the uncertainty of the time filter coefficient(set the time filter coefficient value as 0.02 in the assimilation model and 0.01 in the true model)

- The state variables are spectral coefficients(the atmospheric streamfunction at the 64 (longitude)×54 (latitude) Gaussian grid points)

- Started from the streamfunction at 1200 UTC 1 January 1991

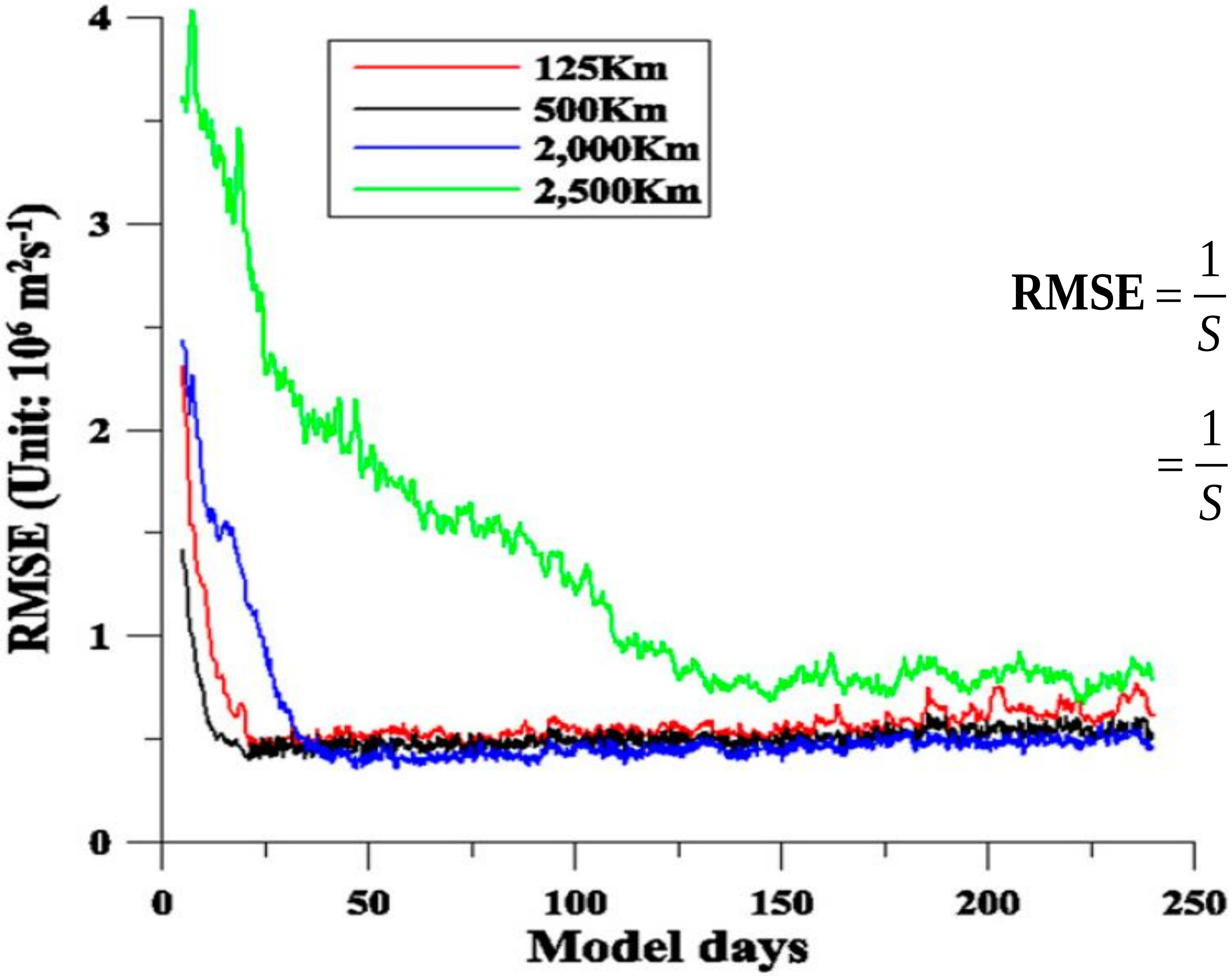
- The integration step size is a half-hour

Biased twin-experiment setup

- Data assimilation method : EnKF, EnKF-MSA , ensemble size is 20
- Spin up： 140 days
- True model states： true model run 240 days
- Observation : A Gaussian noise with the standard deviation of $10^6 m^2 s^{-1}$ is imposed to the truth streamfunction , all observations in NH is available, observations in SH on odd x-index and y-index grids are assumed to be available
- No variance inflation

Results

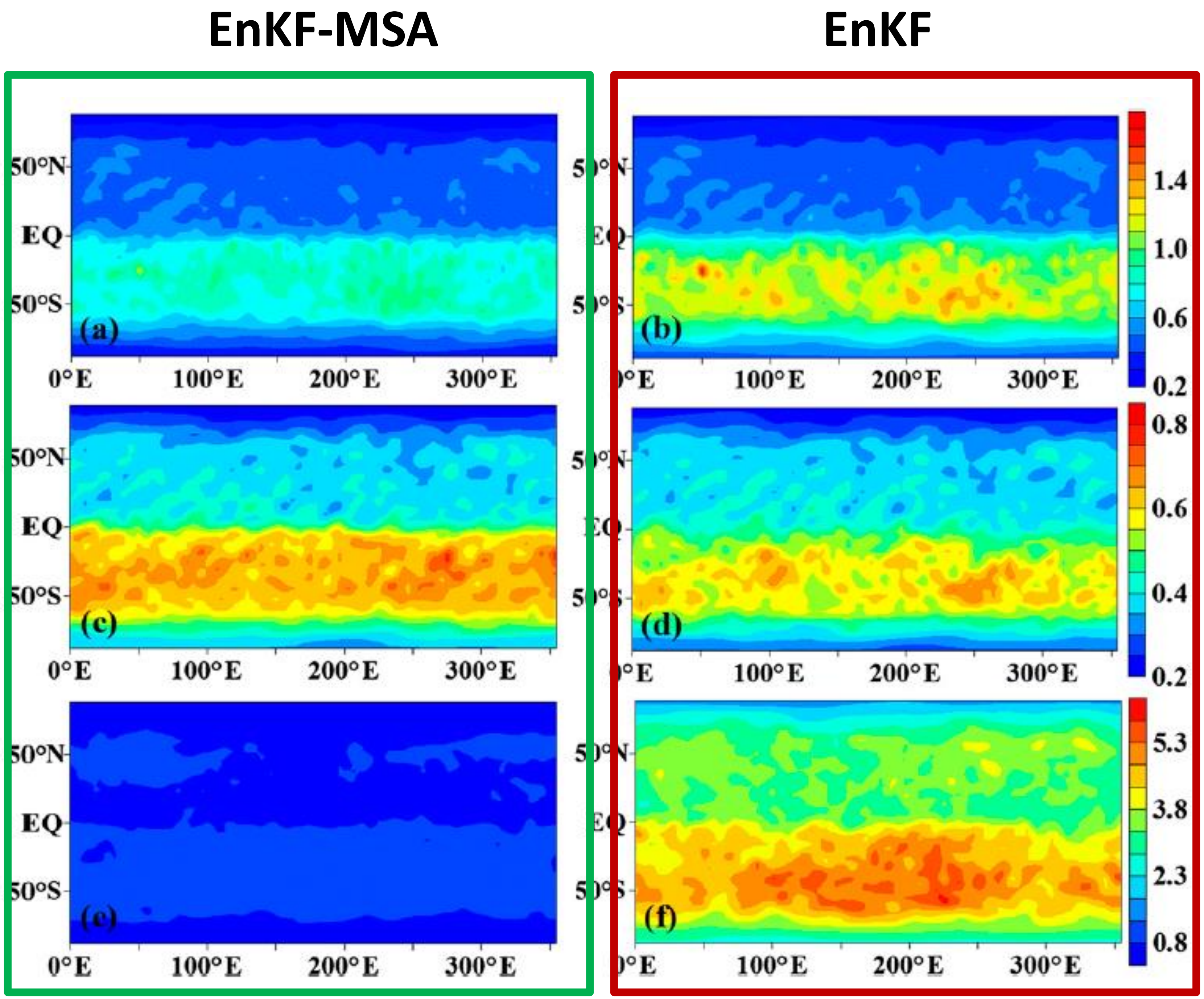
- Dependence on the GC localization half-width



$$\mathbf{RMSE} = \frac{1}{S}\sum_{s=1}^{S}\mathbf{RMSE}_s$$
$$= \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{im \times jm}\sum_{i=1}^{im}\sum_{j=1}^{jm}(\bar{\psi}_{i,j,s}^{prior} - \psi_{i,j,s}^{t})^2}$$

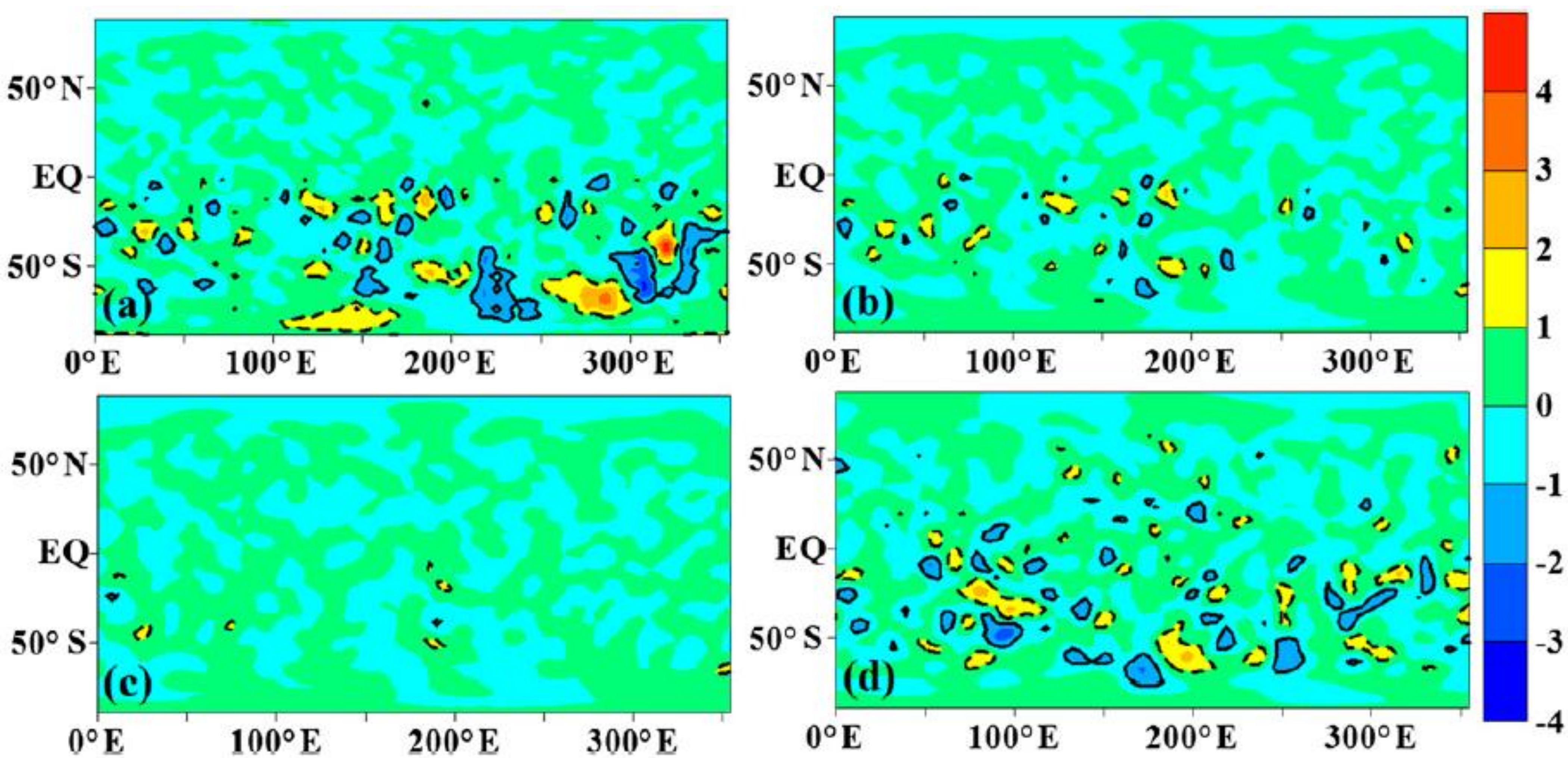
Results

- Spatial RMSEs



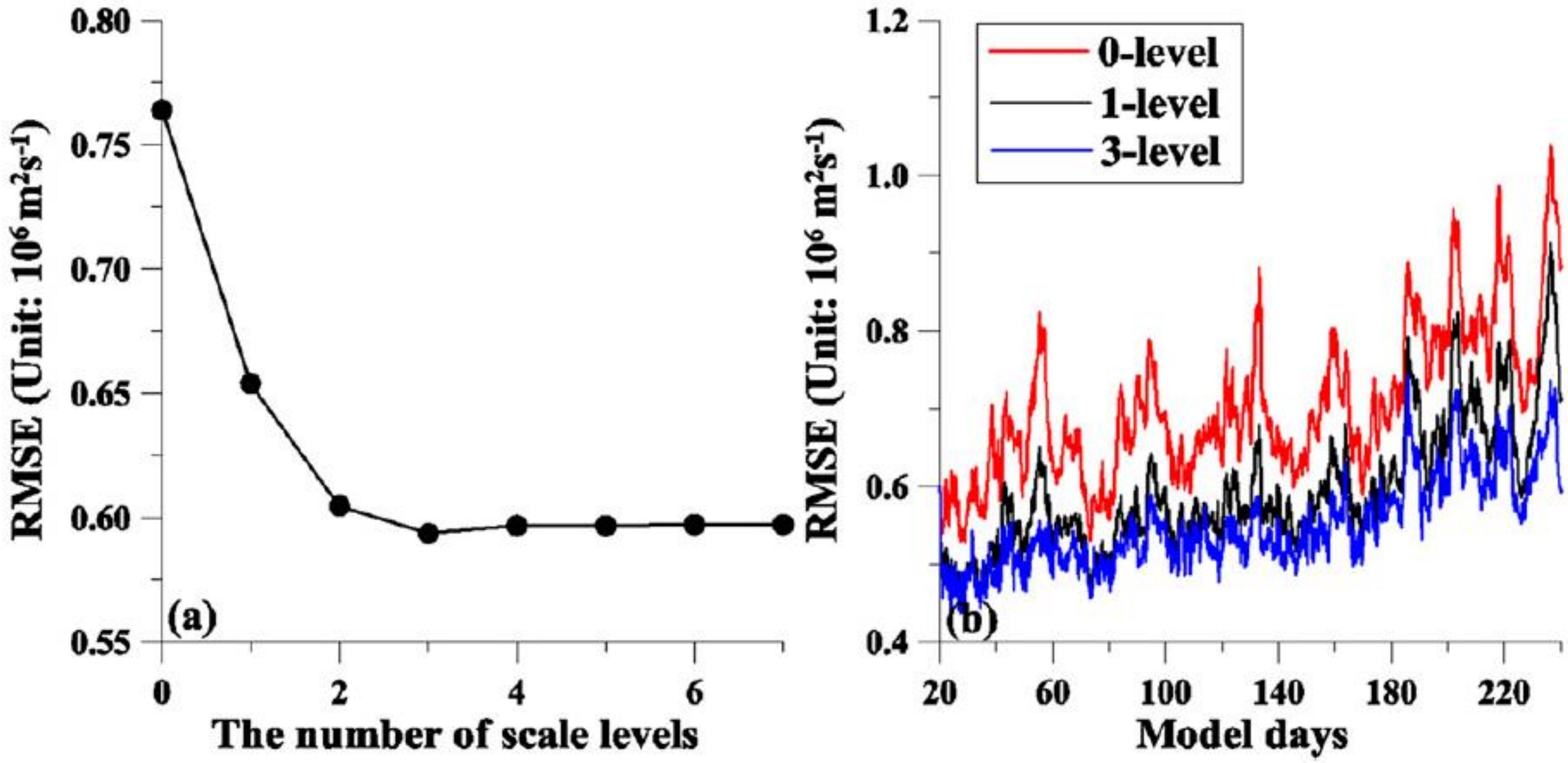
Compare with standard EnKF

- Truth minus analysis with different GC localization in standard EnKF



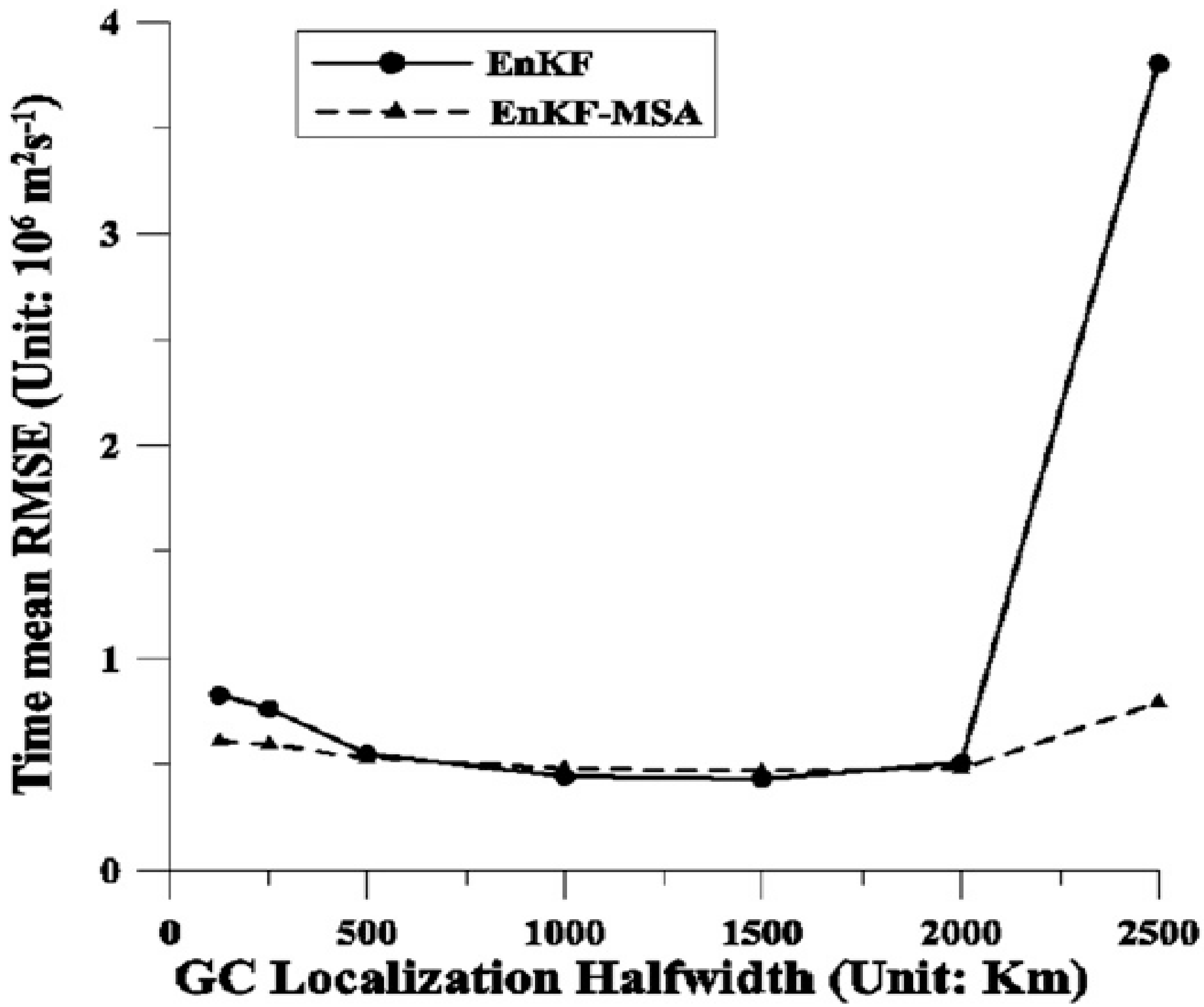
Results

- Dependences on the number of scale levels



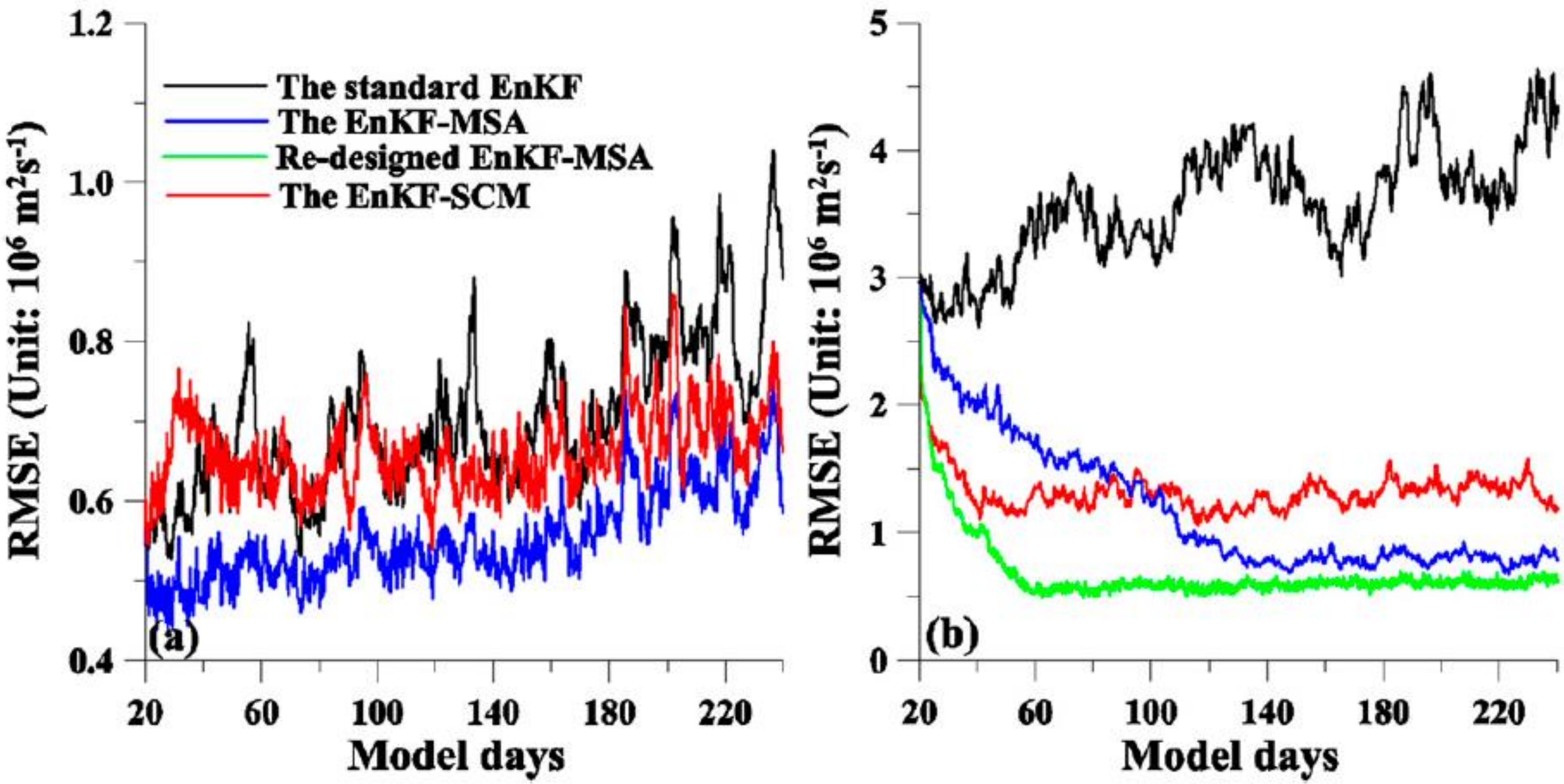
Compare with standard EnKF

- Sensitivity of RMSE respect to GC localization



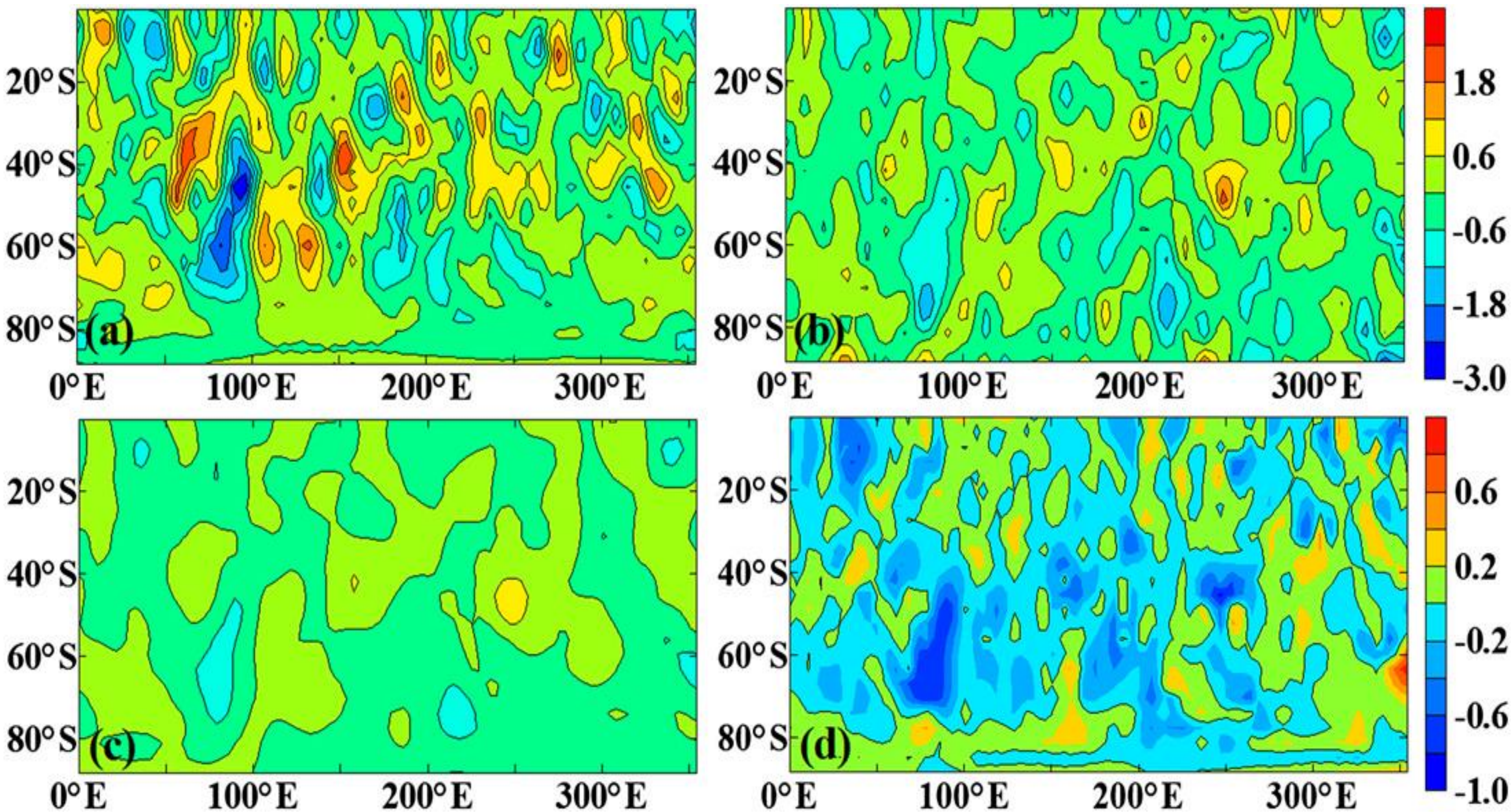
Compare with standard EnKF

- RMSE



Compare with standard EnKF

- The effectivity of retrieving multi-scale information



Compare with standard EnKF

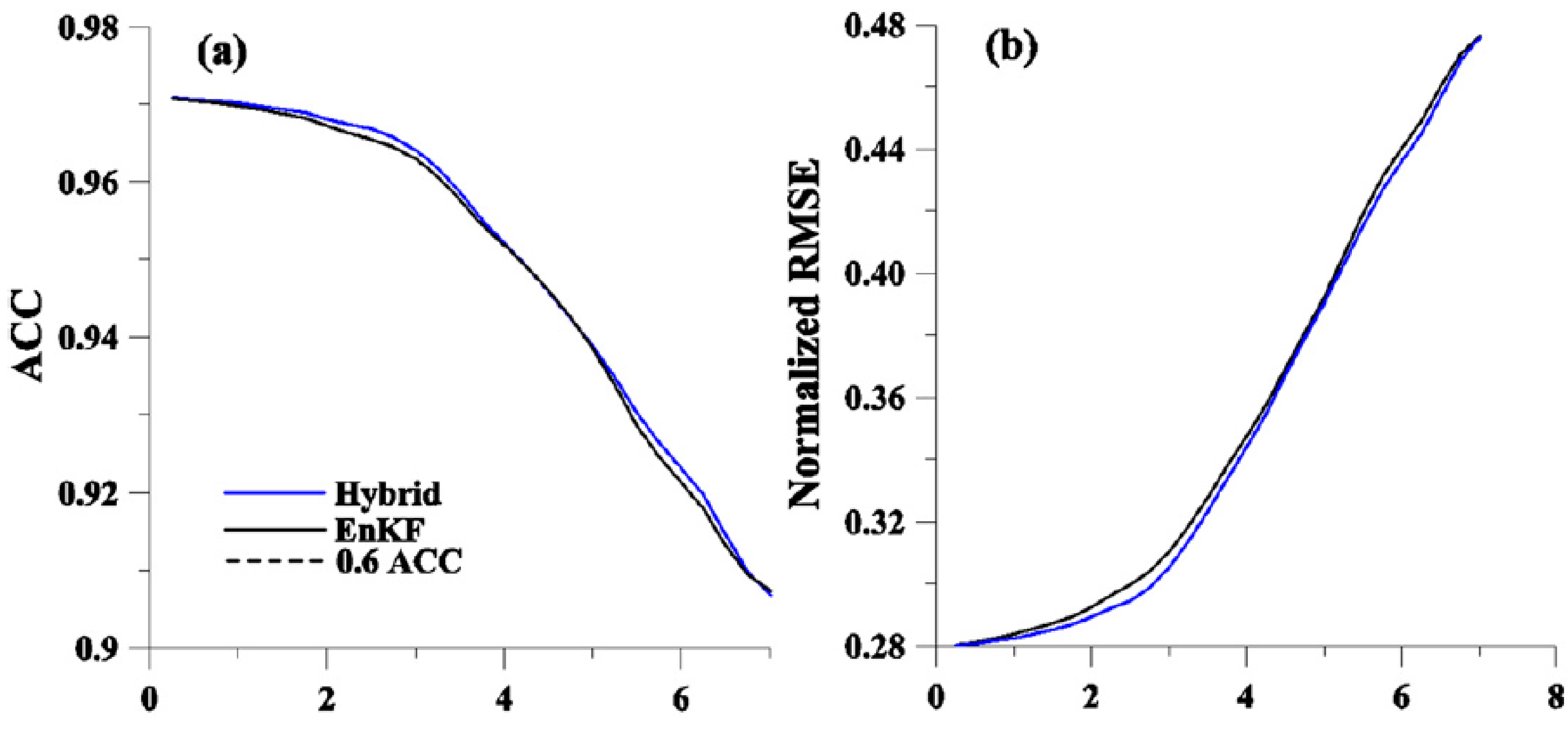
- Impact of the compensatory scheme on weather forecast

$$ACC_s = \frac{1}{R}\sum_{r=1}^{R}\frac{\sum_{i=1}^{im}\sum_{j=1}^{jm}(\overline{\psi'}^{f}_{i,j,r,s} - \overline{\overline{\psi'}}^{f}_{r,s})(\psi'^{t}_{i,j,r,s} - \overline{\psi'}^{t}_{r,s})}{\sqrt{\sum_{i=1}^{im}\sum_{j=1}^{jm}(\overline{\psi'}^{f}_{i,j,r,s} - \overline{\overline{\psi'}}^{f}_{r,s})^2}\sqrt{\sum_{i=1}^{im}\sum_{j=1}^{jm}(\psi'^{t}_{i,j,r,s} - \overline{\psi'}^{t}_{r,s})^2}}$$

$$\mathrm{RMSE}_s = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{im \times jm}\sum_{i=1}^{im}\sum_{j=1}^{jm}\left(\frac{\overline{\psi}^{f}_{i,j,r,s} - \psi^{t}_{i,j,r,s}}{\sigma^{\mathrm{clim}}_{i,j}}\right)^2}$$

Compare with standard EnKF

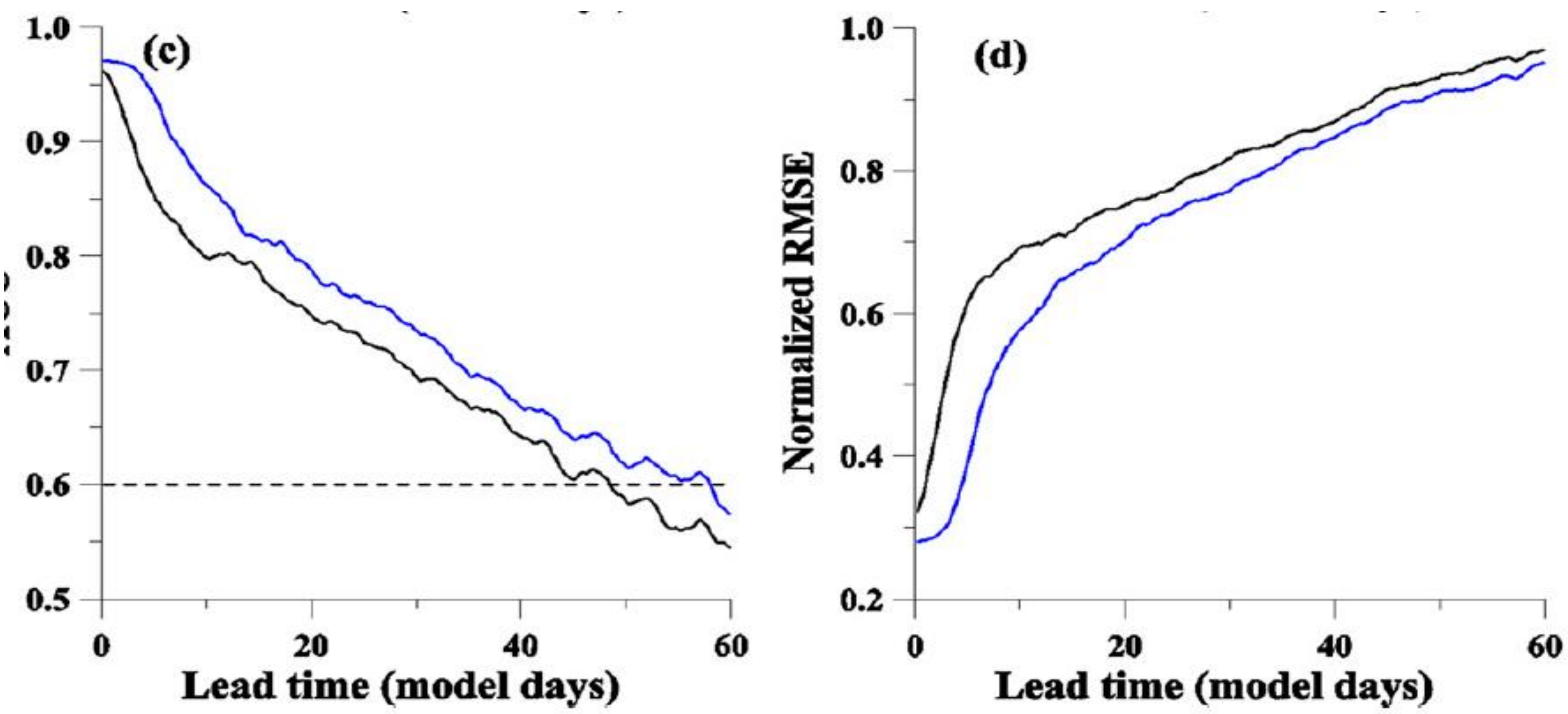
- Impact of the compensatory scheme on weather forecast



a =125km

Compare with standard EnKF

- Impact of the compensatory scheme on weather forecast



a =2500km

Conclusion

- Retrieve multi-scale information from observation effectively
- Without tuning an optimal cutoff distance
- EnKF-MSA is superior to a standard EnKF and it has less dependence on cutoff distances
- Enhance the short-term weather forecast skill

Challenges

- Observing network is rather homogeneous and sufficiently dense coverage
- Without comparison between the compensatory approach and the adaptive localization model
- Without variance inflation
- Forward operator is simple
- Without ensemble size sensitivity test

Further work

- How to choose level
- New method to retrieve the multi-scale information
- Ensemble size test
- Comparison with adaptive localization method

Thanks !