Estimation of linear observation impact and its applications

Toshiyuki Ishibashi
MRI/JMA, Japan
ishibasi[AT]mri-jma.go.jp

* MRI: Meteorological Research Institute.
**JMA: Japan Meteorological Agency.

Contents

1. Observation impact estimation

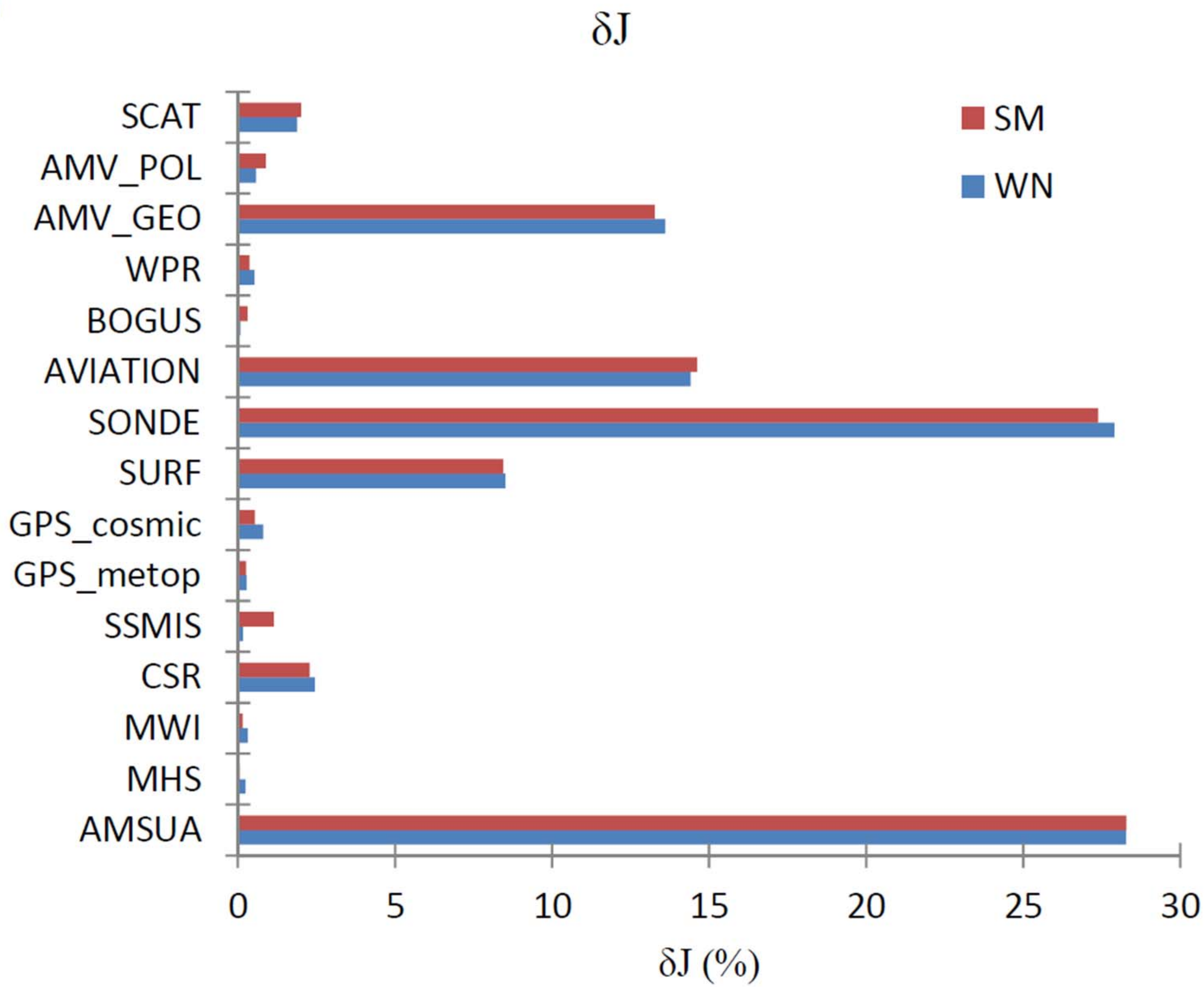
Two types of observation impact

- Basic definition
  - The observation impact is defined as "the variations of analyses and forecasts caused by changes of observation data".
- Non-linear observation impact
  - This is the observation impact that has no limitations on the changes of observation data, so the changes includes perturbations in observation data values, and additions of observation datasets.
  - Estimation methods:
    - ✓ OSE (observing system experiment).
- Linear observation impact
  - This is the observation impact that has an limitation on the changes of observation data, which is Kalman gain is invariant.
  - Estimation methods are
    - ✓ ADJ-based method (FSO)
      - Langland and Baker (2004), Errico (2007), Cardinali (2009), Tremolet (2008)
    - ✓ TL-based method
      - Ishibashi (2011)
    - ✓ DFS
      - Cardinali et al (2004), Desroziers et al (2005)

These two observation impacts are different quantities, so, in general, they cannot work as a proxy for each other.

ADJ-based estimation in JMA global 4D-Var

- JMA global 4D-Var
  - ✓ Using Low resolution system (Tl319/T106).
- Evaluation periods are:
  Summer: Aug 2010,
  Winter: Jan 2010.
  * 00UTC analyses only.
- Using dry total energy norm.
- Forecast error evaluation time is 15hours.



Formulation of TL-based method

- The analysis increment vector can be written as a superposition of partial increment vectors (PIVs).

$$\delta \mathbf{x} = \delta \mathbf{x}^{*P} + \delta \mathbf{x}^{*Q} + \cdots \qquad \delta x_i^{*P} \equiv \sum_{r \in P} K_{i,r} d_r \, ; \quad \delta x_i^{*Q} \equiv \sum_{r \in Q} K_{i,r} d_r$$

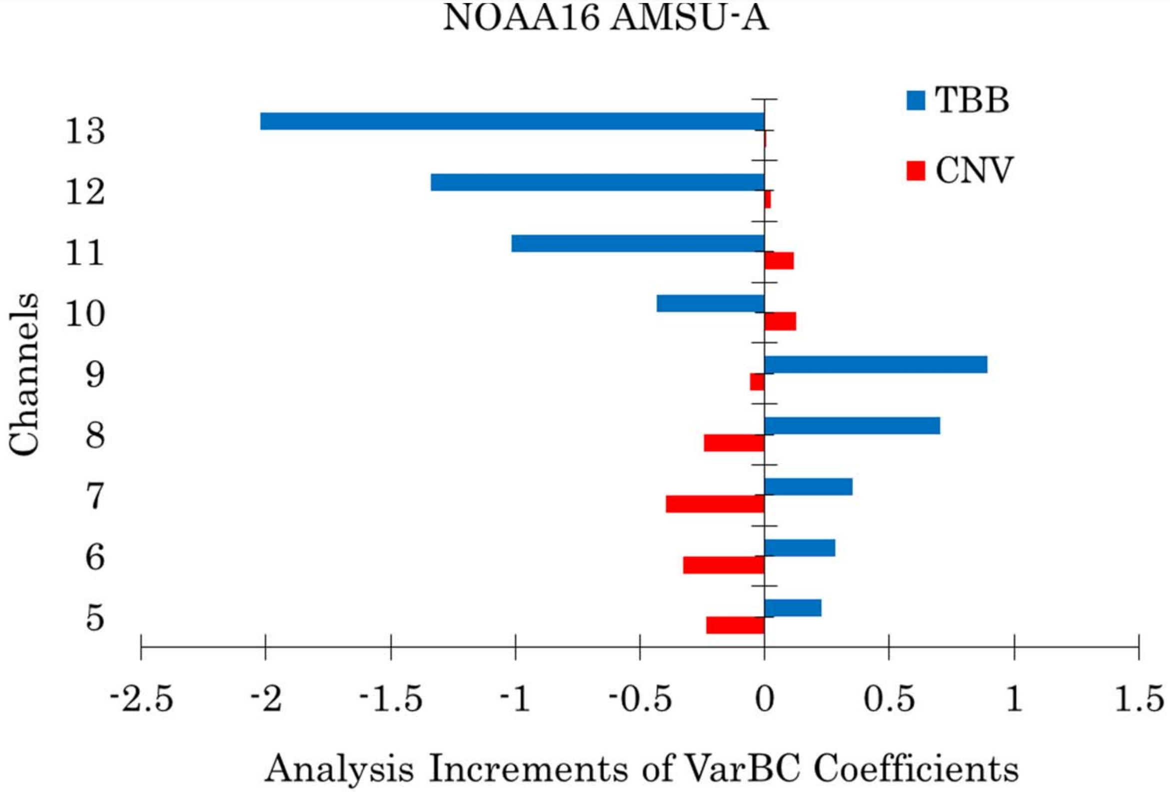
- The PIV represents a linear observation impact of each dataset.
- The departure vector (observations minus background) can be written a superposition of partial departure vectors (PDVs).

$$\mathbf{d} = \mathbf{d}^{*P} + \mathbf{d}^{*Q} + \cdots \qquad d_r^{*P} \equiv \begin{Bmatrix} d_r & r \in P \\ 0 & r \notin P \end{Bmatrix} ; \; d_r^{*Q} \equiv \begin{Bmatrix} d_r & r \in Q \\ 0 & r \notin Q \end{Bmatrix}$$

- PIVs can be written in terms of PDVs

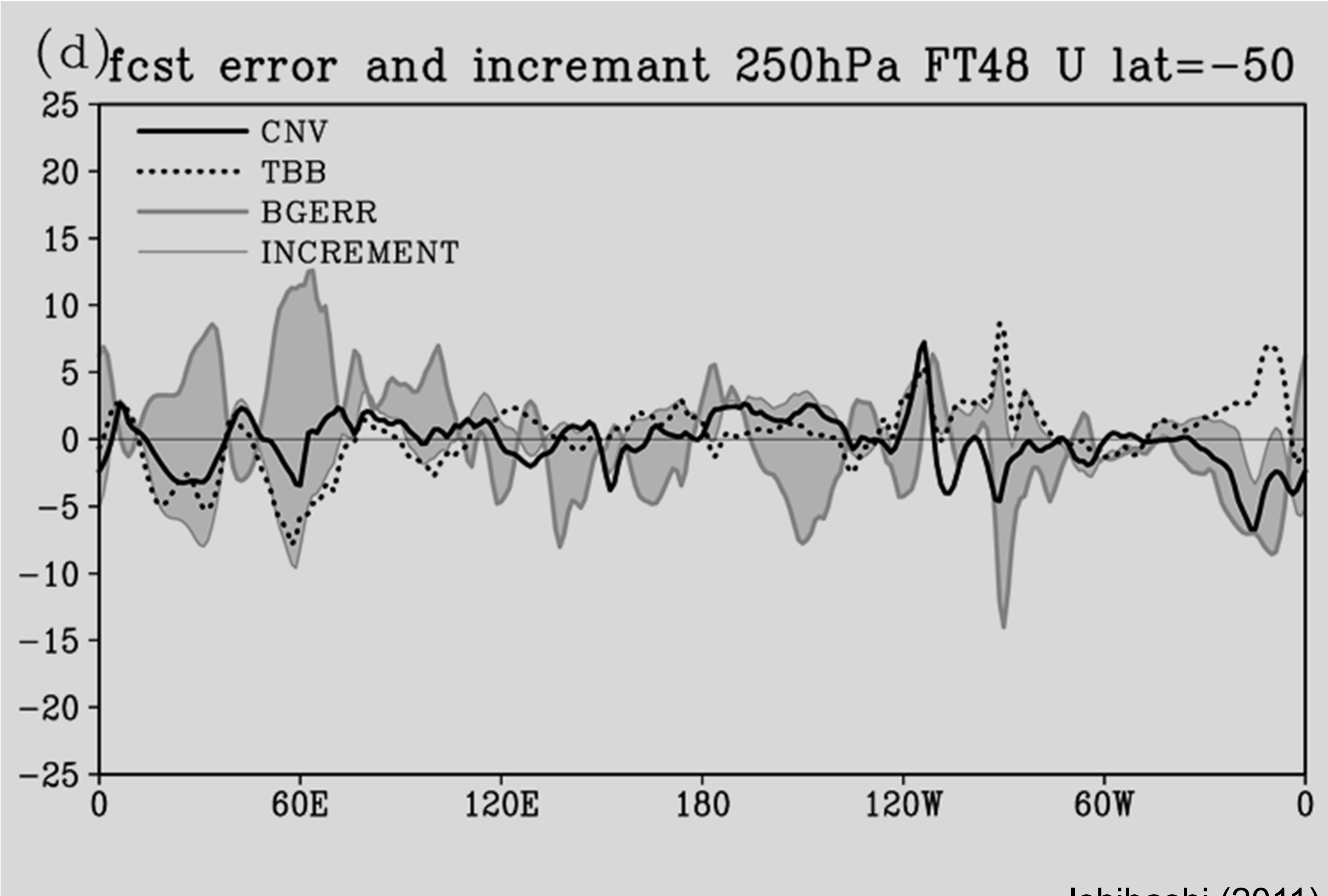
$$\delta \mathbf{x}^{*P} = \mathbf{K}\mathbf{d}^{*P}, \; \delta \mathbf{x}^{*Q} = \mathbf{K}\mathbf{d}^{*Q}, \; \cdots$$

TL-based method



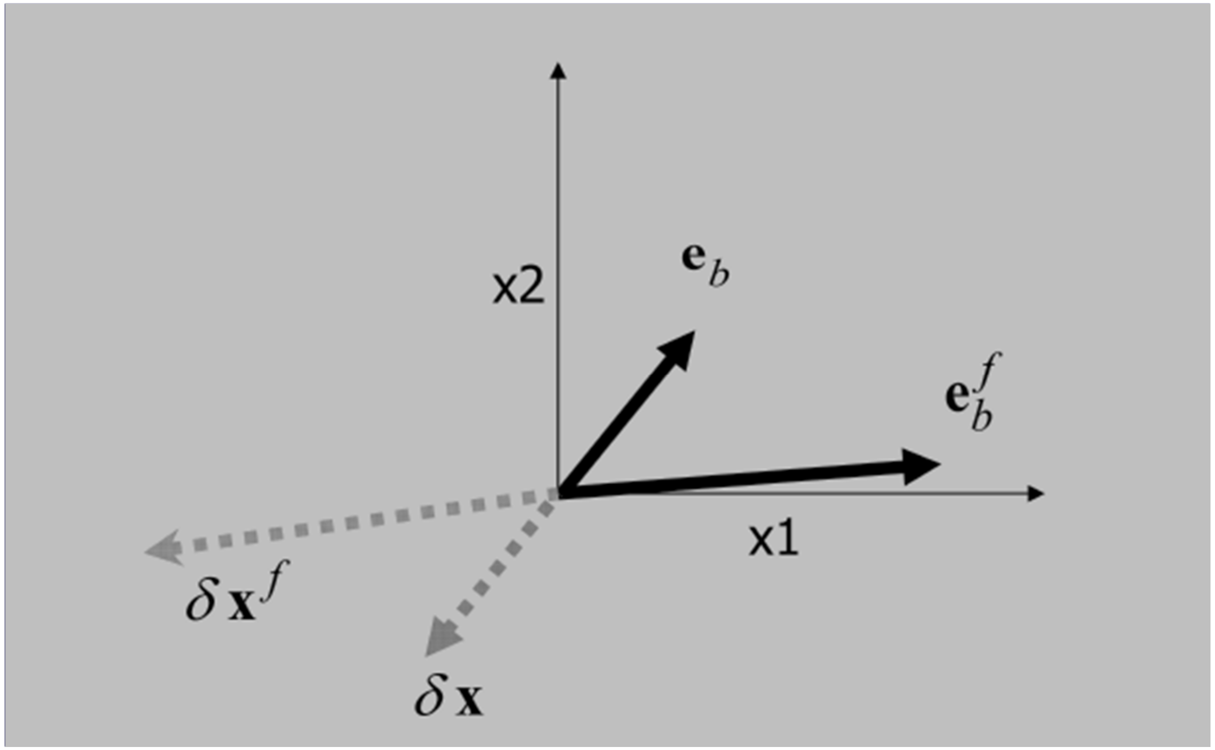
- This figure shows the CNV-PIV and the TBB-PIV for VarBC (variational bias correction) variables of the AMSU-A sensor of the NOAA16 satellite.
- We can find finite contribution from the CNV.
- This result suggests the existence of a stability effect of the CNV for the VarBC variables (Auligné et al., 2007) at least qualitatively.

TL-based method



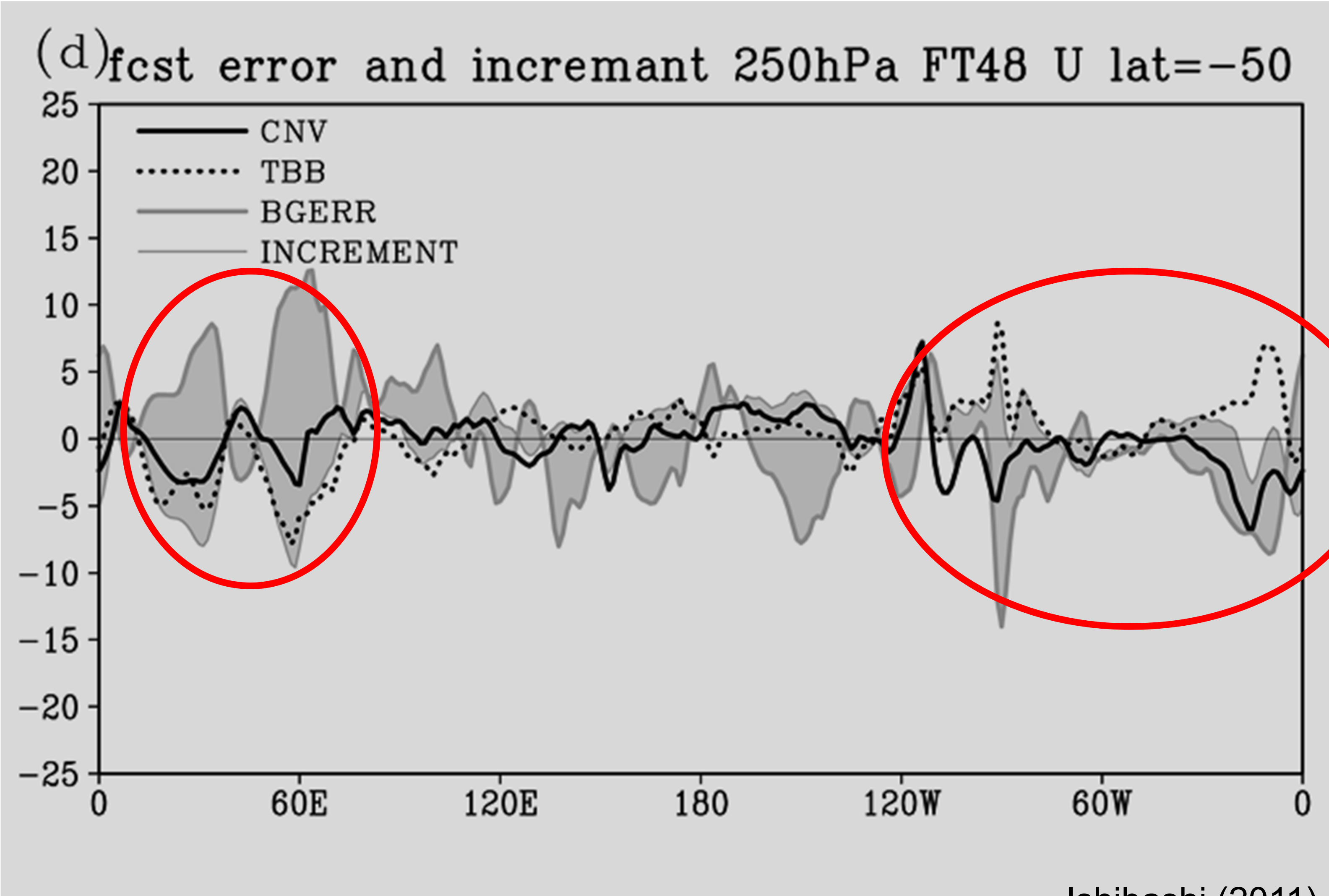
Ishibashi (2011) QJRMS

TL-based method



Ishibashi (2011) QJRMS

TL-based method



Ishibashi (2011) QJRMS

2. Covariance matrix optimization

Relationships between observation impact estimations and covariance optimizations

- Two types of error covariance matrix optimization methods.
  1. Expectation-based method
     - This method optimizes error covariance matrices based on the theoretical relation ships;

       $$2E[J_o] = Tr[\mathbf{I} - \mathbf{HK}] \qquad 2E[J_b] = Tr[\mathbf{KH}]$$

     - Desroziers and Ivanov (2001), Desroziers et al (2005), and Chapnik et al (2004, 2006)
  2. Sensitivity-based method
     - This method uses sensitivity of forecast errors with respect to covariance matrices;

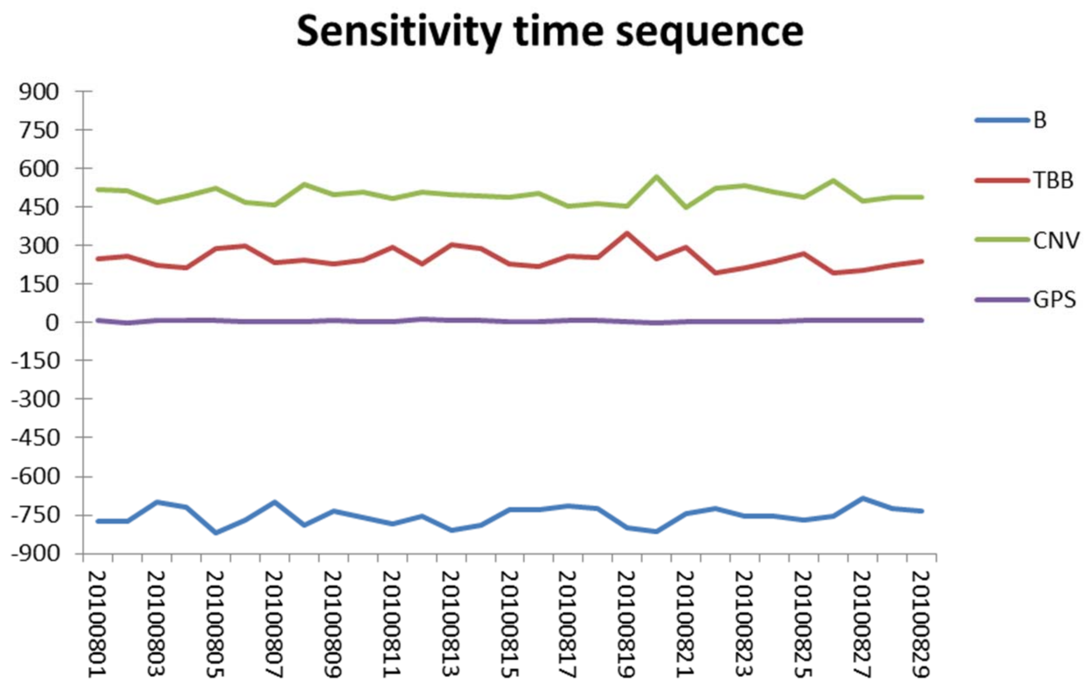
       $$\frac{\partial J}{\partial \mathbf{R}}, \frac{\partial J}{\partial \mathbf{B}}$$

     - Daescu (2008) , Daescu and Todling (2010).

- Each optimization method include a linear observation impact estimation.
  1. Expectation-based method includes DFS calculation.
  2. Sensitivity-based method includes ADJ-based estimation.

Here, Let's see the sensitivity-based method

Diagnoses of the JMA global 4D-Var

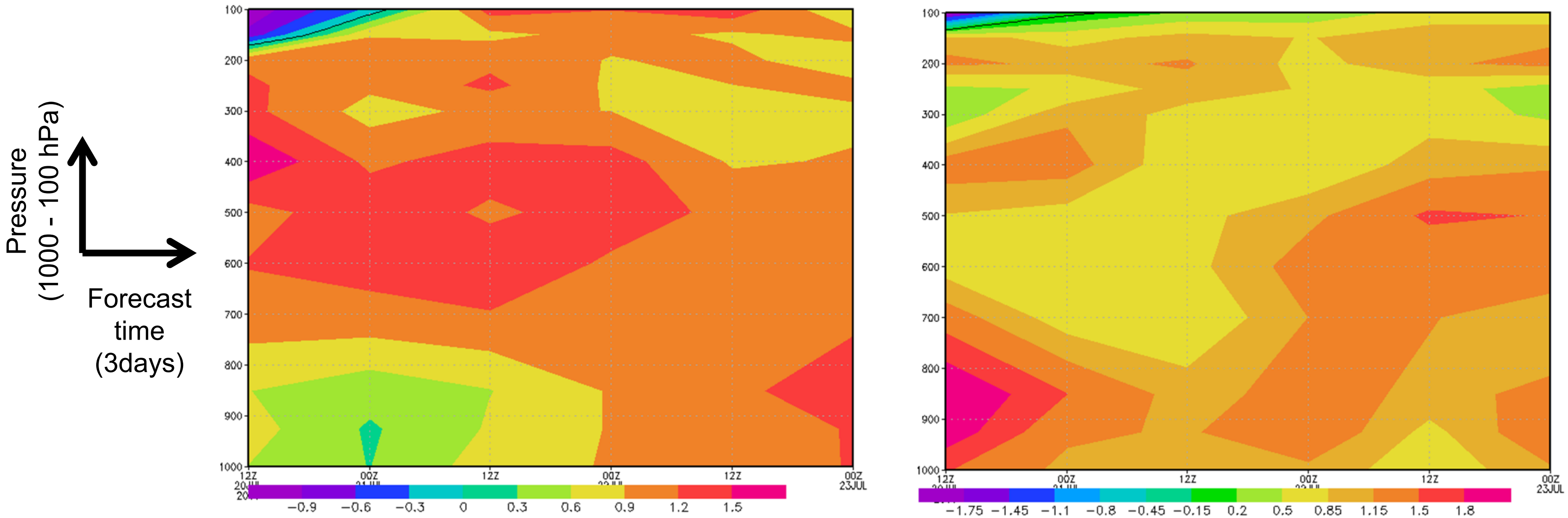


- Sensitivity calculation results in August 2010.
- Using dry total energy norm with 15hr forecasts.
- The results show that **B** is too small and **R** is too large in average.

Impact of error covariance optimization on forecast accuracy

Improvement rate: Zonal wind

Improvement rate: Temperature



Results of a single case experiment of covariance optimization using the sensitivity method.

- TEST uses optimized **R,** CNTL uses original **R** (operational setting).
- The figure shows normalized forecast RMSE differences between TEST and CNTL: (CNTL – TEST)/CNTL.
- Warm (cold) color areas are forecast error decrease (increase) areas.

3. Analysis error estimation

3. Analysis error estimation

Background

- We want to know analysis errors of a DAS because the analysis error information is useful to improve current DASs and to design future observational systems which can detect the analysis errors.
- Analysis error estimation is the same with construction of more accurate analysis than current DASs. Such analyses can be used as "pseudo truth".

Previous studies

- "Key analysis error" (Rabier et al 1996, Klinker et al 1998, Isaksen et al 2005) can generate more accurate forecasts than current DASs.
- However, there are inconsistency between key analysis errors and observation information. SOSE (Marseille 2007) can partly reduce this problem.

Our approach

1. We construct the pseudo truth based on the data assimilation theory.
2. We construct the pseudo truth based on the ADJ-based method.

Data assimilation theory based method

- Conditional PDF

$$P(\mathbf{x}|\mathbf{y},\mathbf{x}_b) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x}_b|\mathbf{x})$$

Ordinary 4D-Var

$$P(\mathbf{x}|\mathbf{y},\mathbf{x}_b,\mathbf{x}_{ref}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x}_b|\mathbf{x})P(\mathbf{x}_{ref}|\mathbf{x})$$

Extended 4D-Var with reference analyses information

- Add reference analysis fields information

$$J = J_{org} + 1/2\left(\mathbf{x}_{ref} - M(\mathbf{x}_b + \delta\mathbf{x})\right)^T \mathbf{A}^{-1}\left(\mathbf{x}_{ref} - M(\mathbf{x}_b + \delta\mathbf{x})\right)$$

$$= J_{org} + 1/2\left(\mathbf{e}^t + \mathbf{M}\delta\mathbf{x}\right)^T \mathbf{A}^{-1}\left(\mathbf{e}^t + \mathbf{M}\delta\mathbf{x}\right)$$

$$J_{org} \equiv 1/2\,\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + 1/2(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})$$

X: analysis,

Y: observations

X$_b$: background field

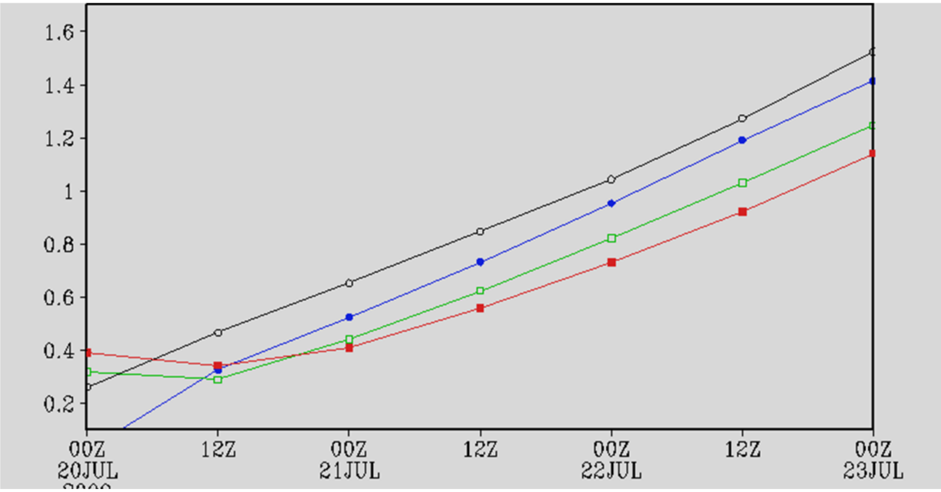
Xref: reference analyses

- Analytical solution has an error covariance matrix **A** of reference information in Kalman gain, and forecast error in input data, as follows,

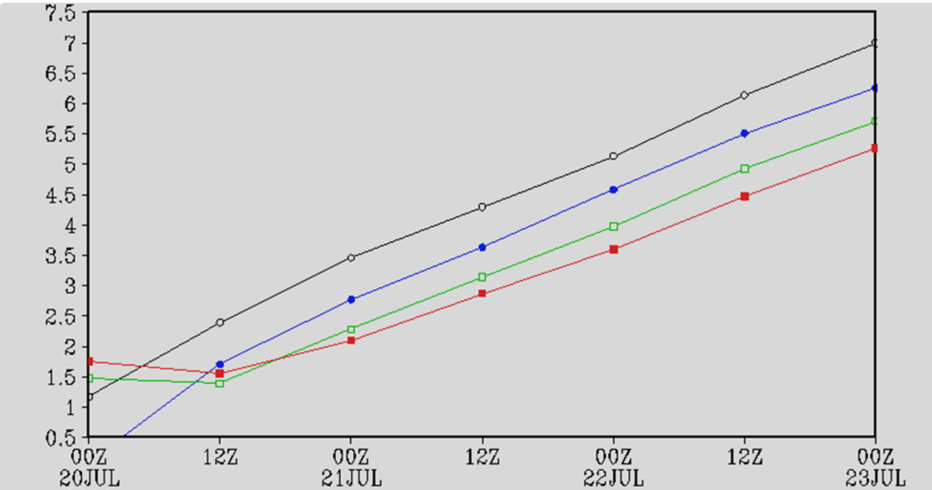
$$\delta\mathbf{x} = \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{M}^T\mathbf{A}^{-1}\mathbf{M}\right)^{-1}\left\{\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} - \mathbf{M}^T\mathbf{A}^{-1}\mathbf{e}^t\right\}$$

Accuracy of optimized forecasts

500hPa Temperature



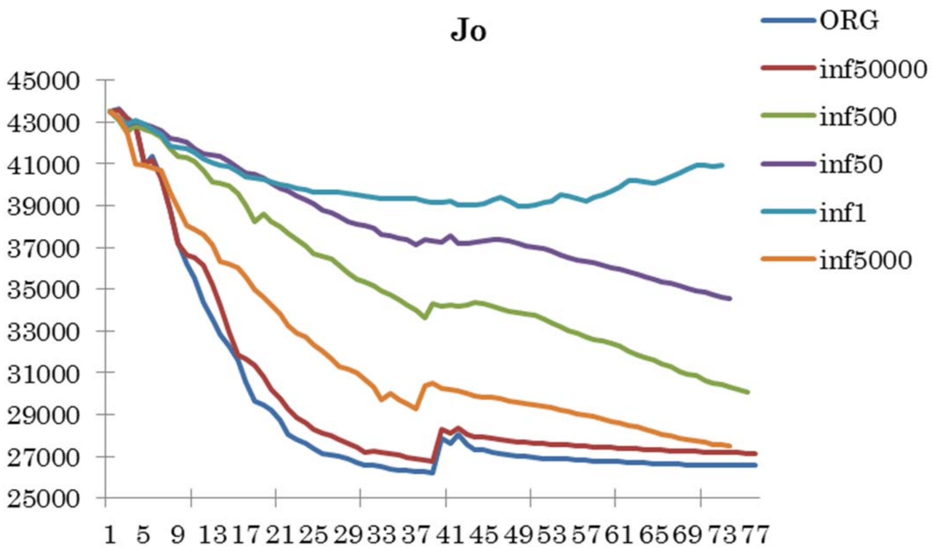
250hPa Zonal wind

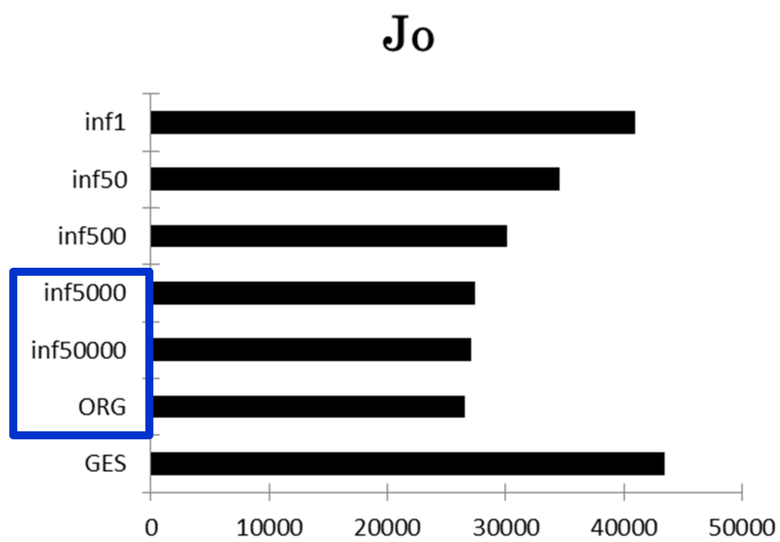


***Inflation factor is one.**

- Optimized forecast with four reference analyses of every 6hours.
- Optimized forecast with only two reference analyses
- Original forecast
- Original forecast from 6 hours after initial.

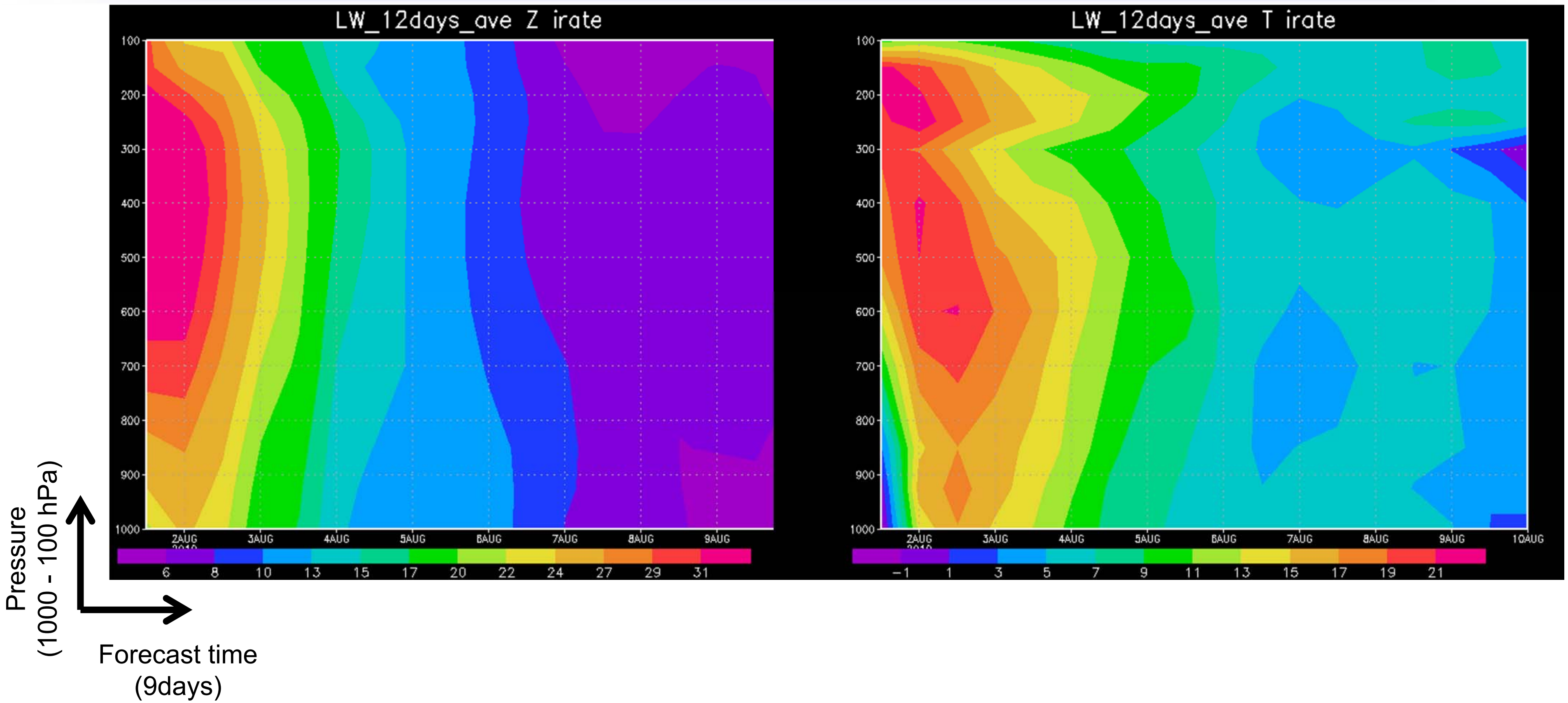
Fitting of the optimized analysis to observations





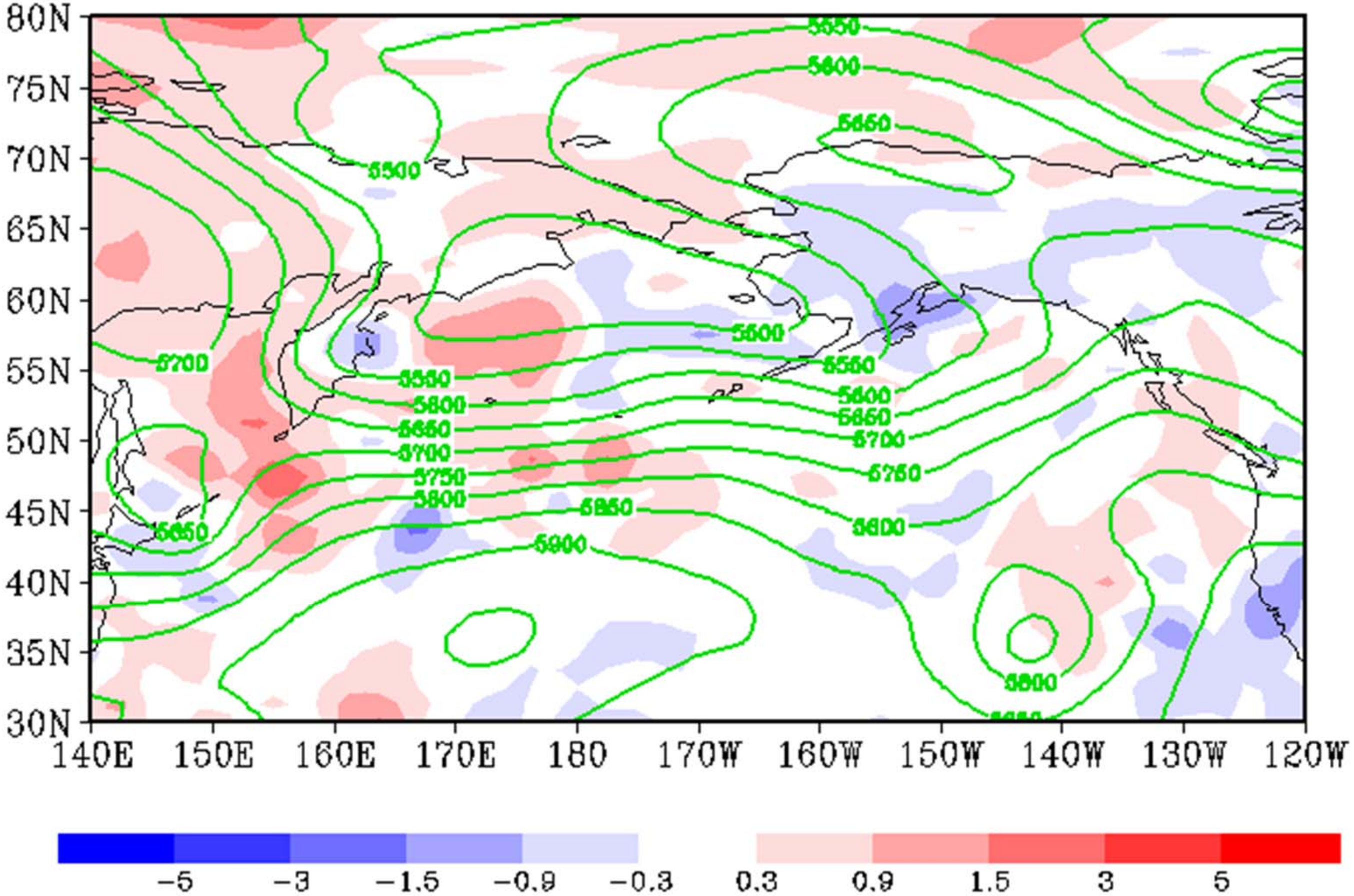
- The inflation factor dominates the fittings of analysis to observations.
- The inflation factors larger than 500 achieve good fitting to observations.

Two weeks statistics



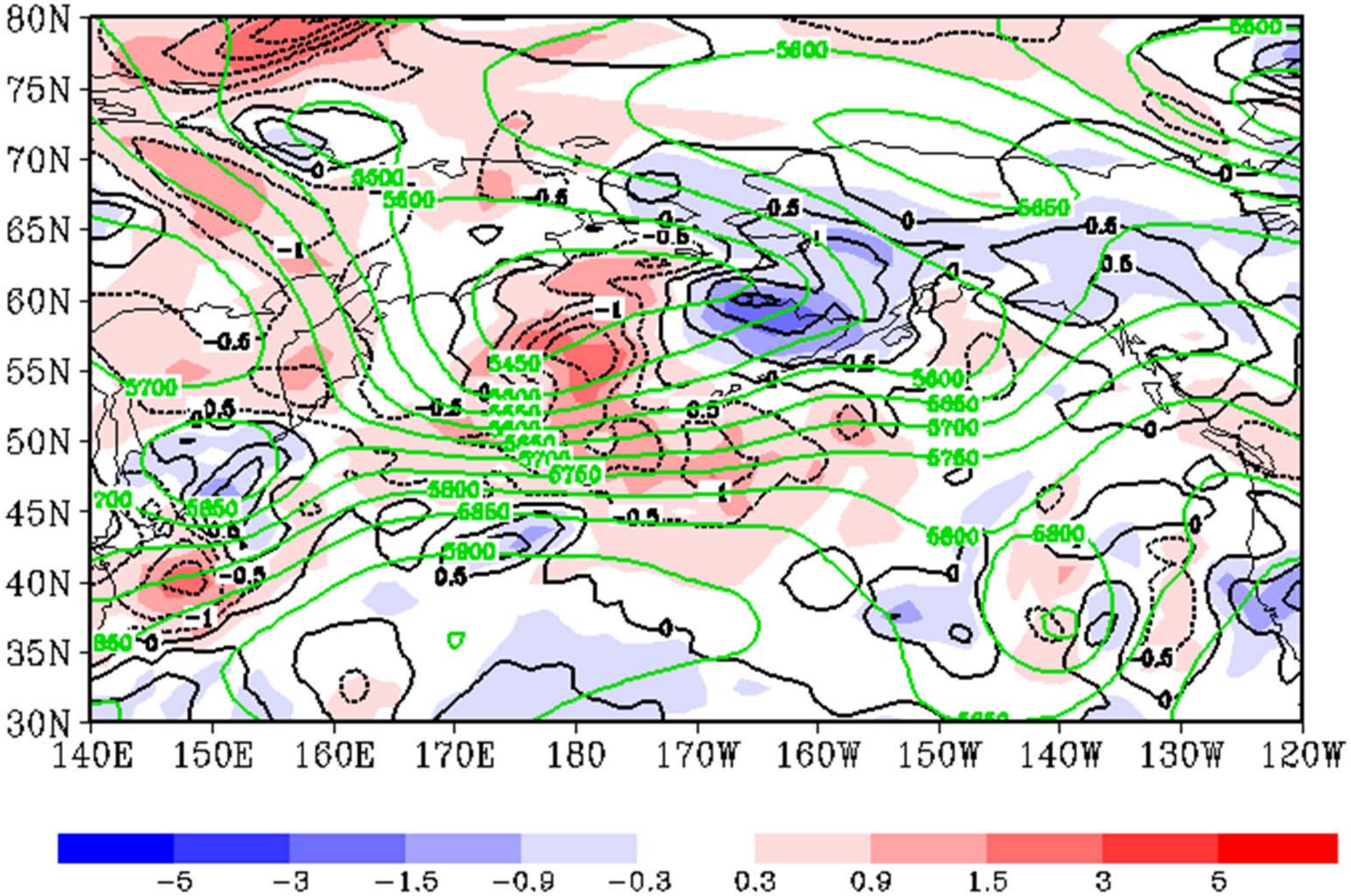
- Forecast accuracy improvement rate of the optimized forecasts against the original forecasts.
- Forecast accuracy are kept 9days with 95% statistical significance until 6 or 7days.
- The inflation factor is 5000.

Optimized analysis increments and background error 072000



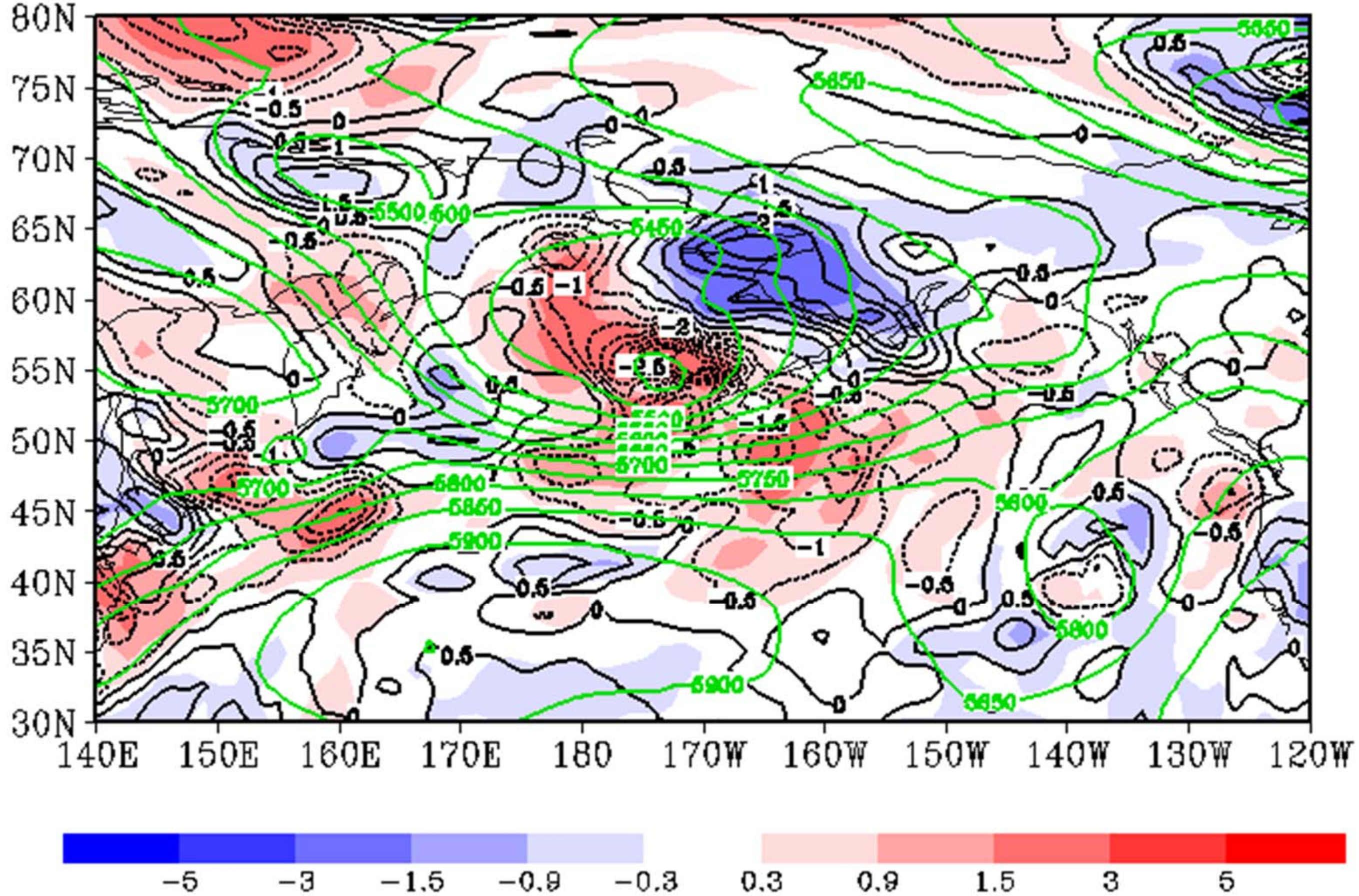
- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

Optimized analysis increments and background error 072012



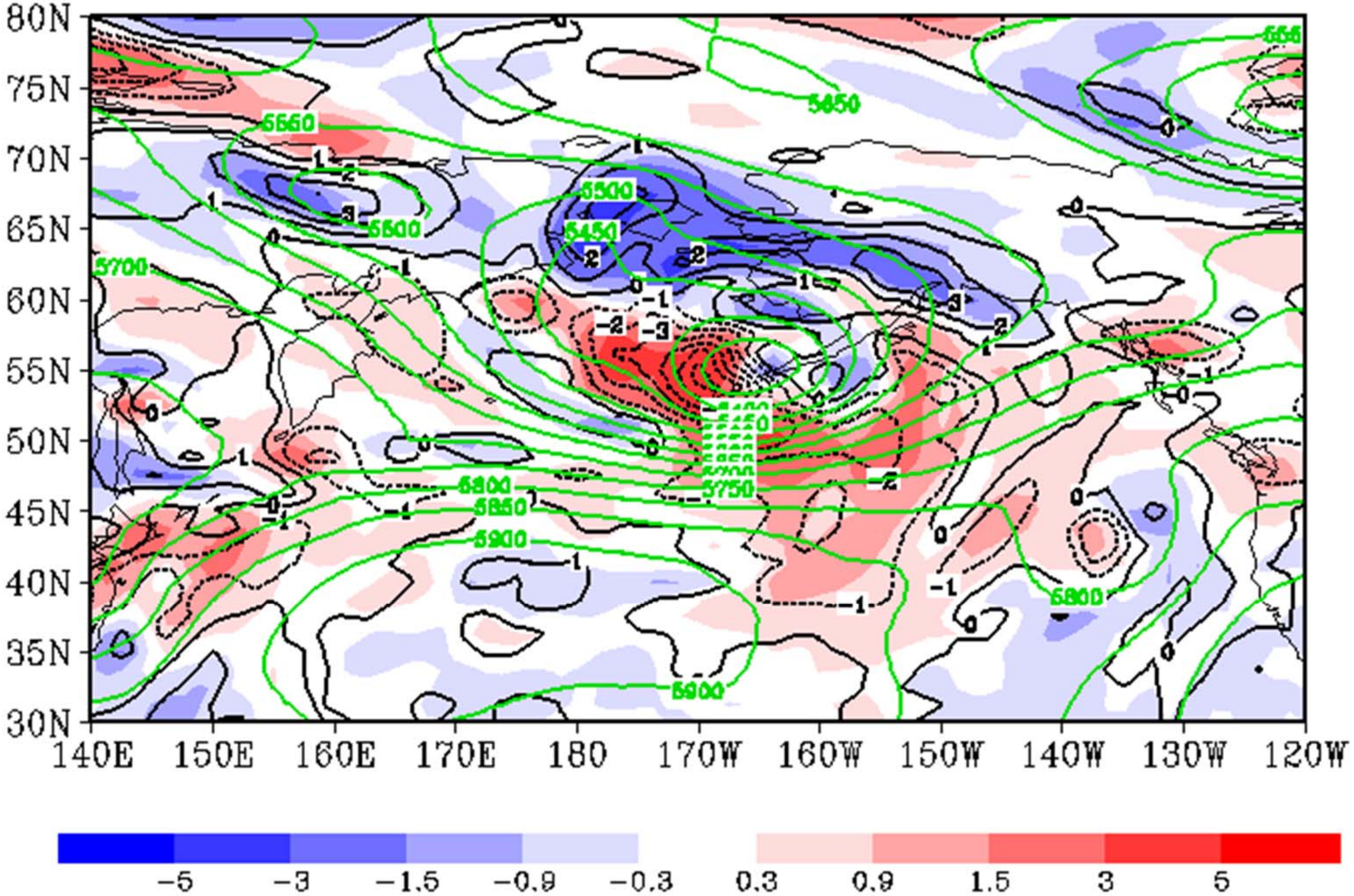
- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

Optimized analysis increments and background error 072100



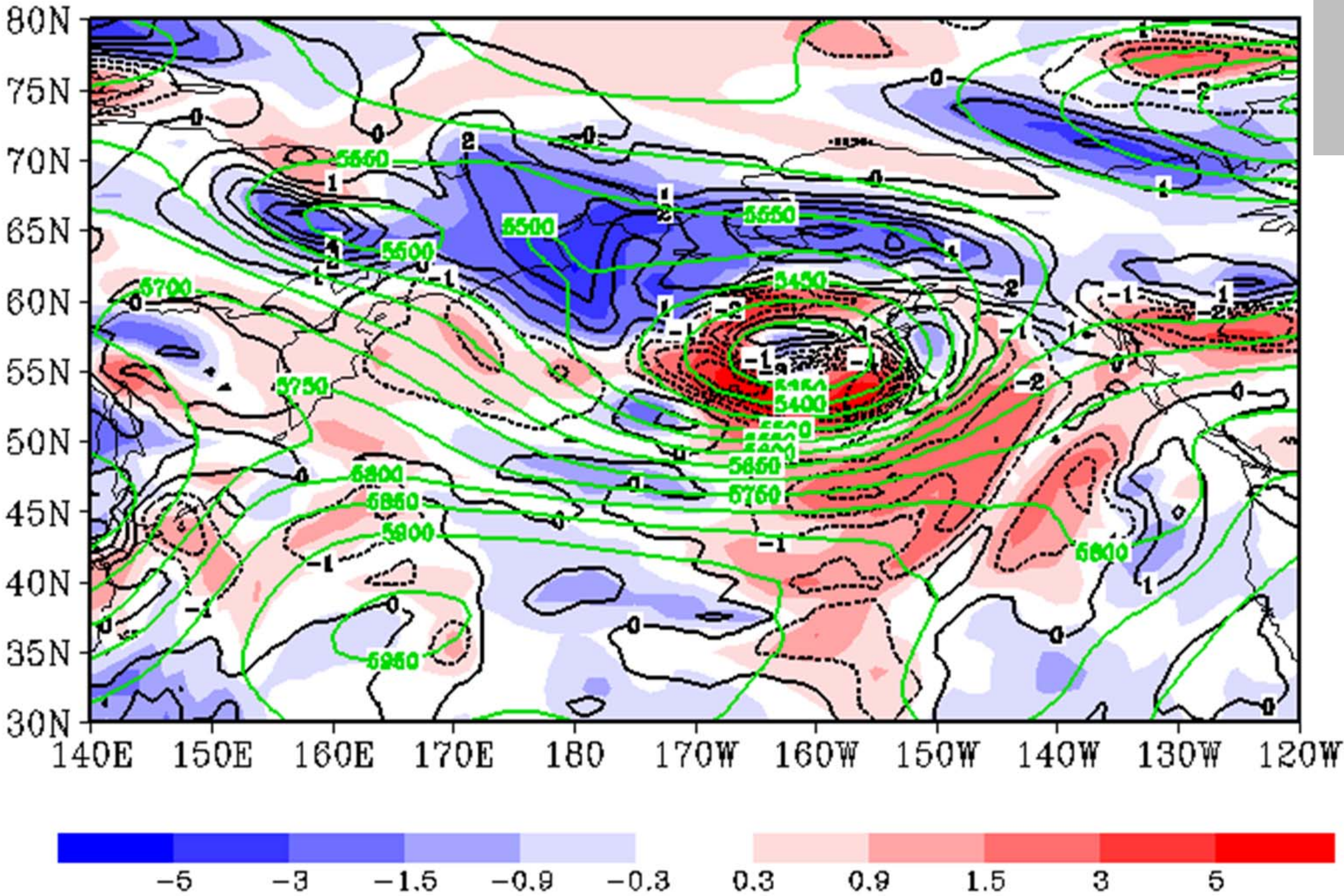
- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

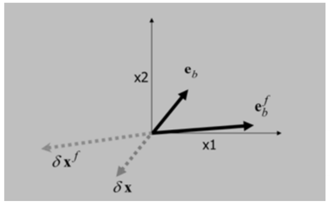
Optimized analysis increments and background error 072112



- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

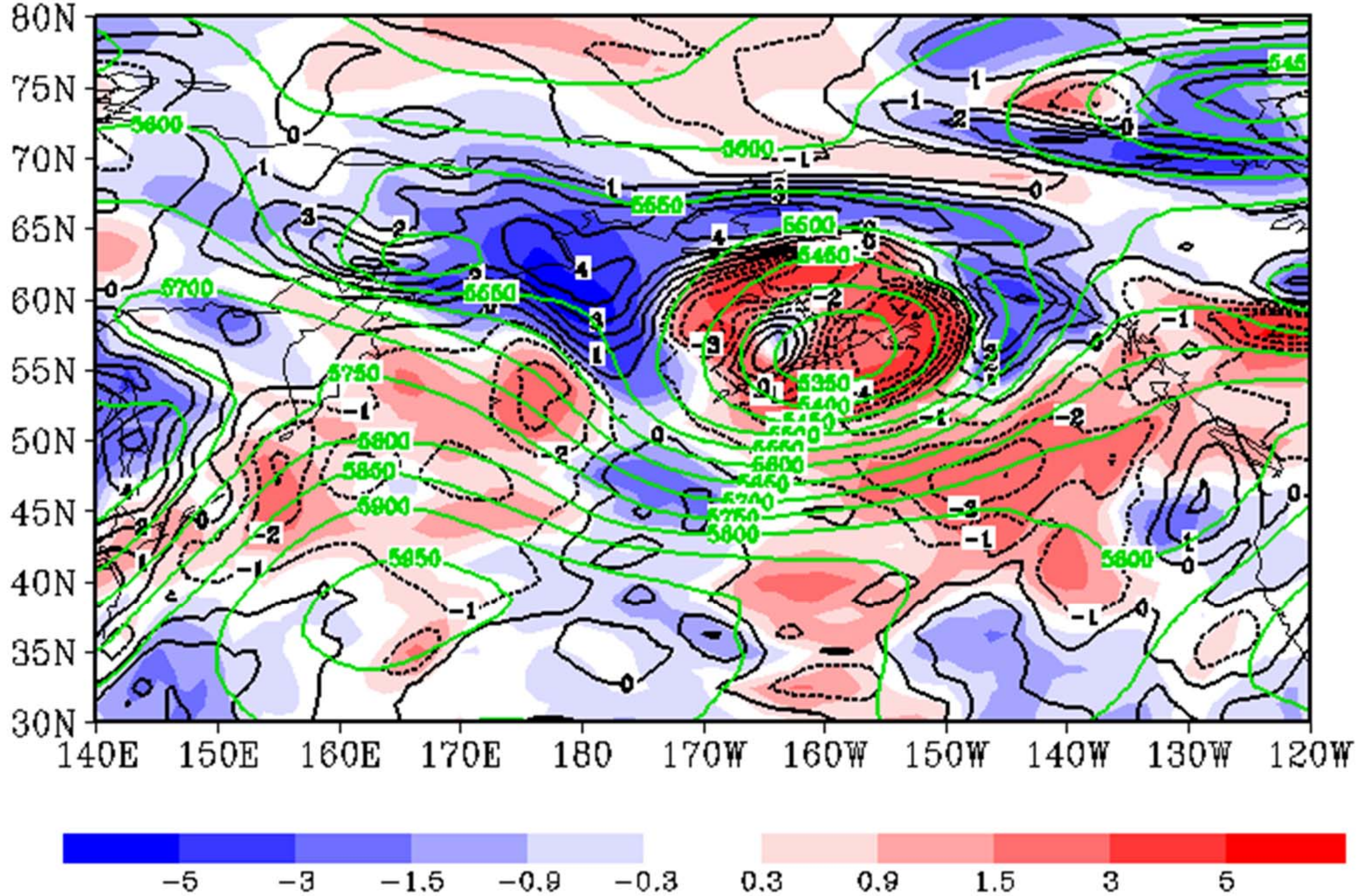
Optimized analysis increments and background error 072200





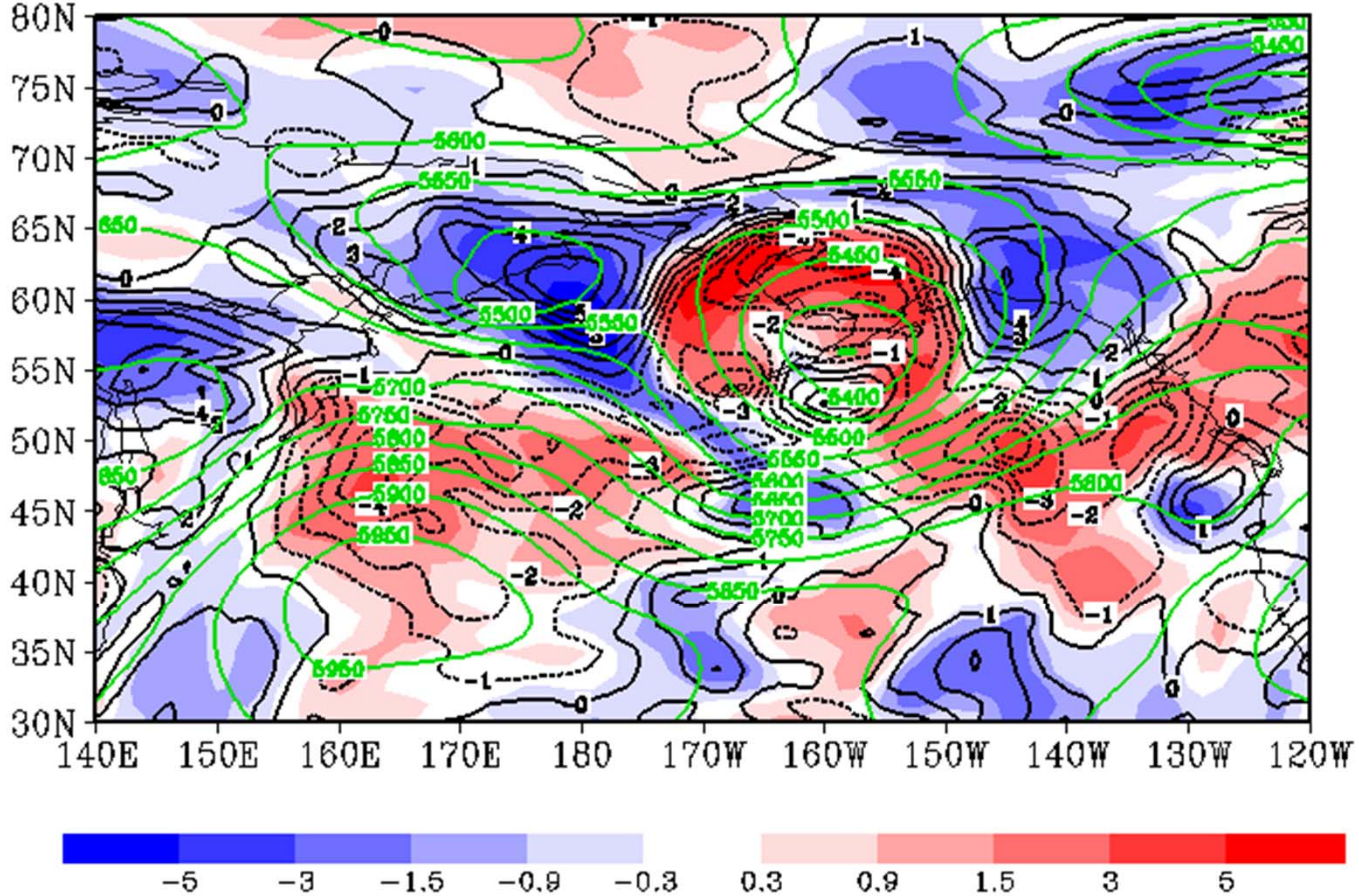
- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

Optimized analysis increments and background error
072212



- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

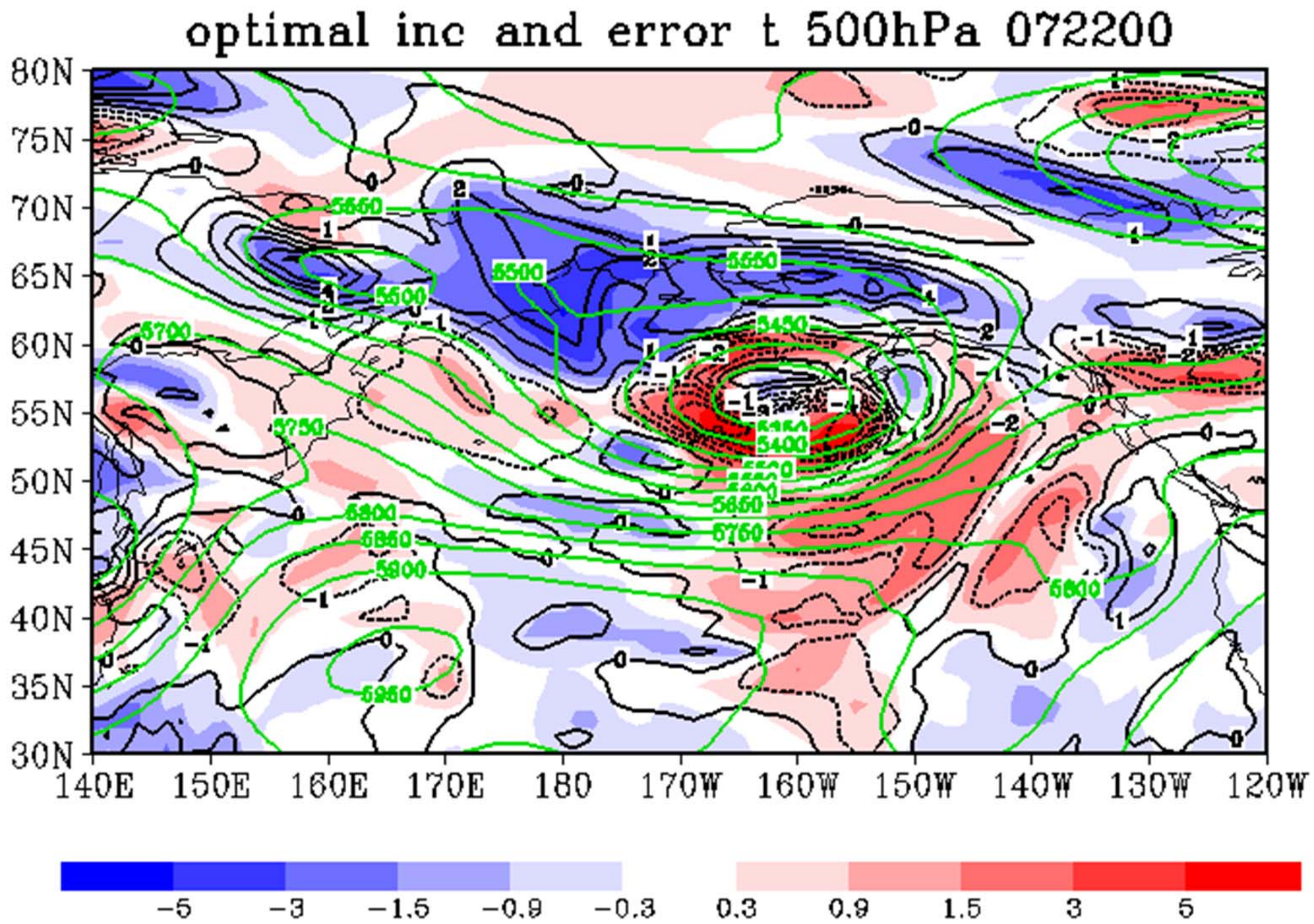
Optimized analysis increments and background error 072300



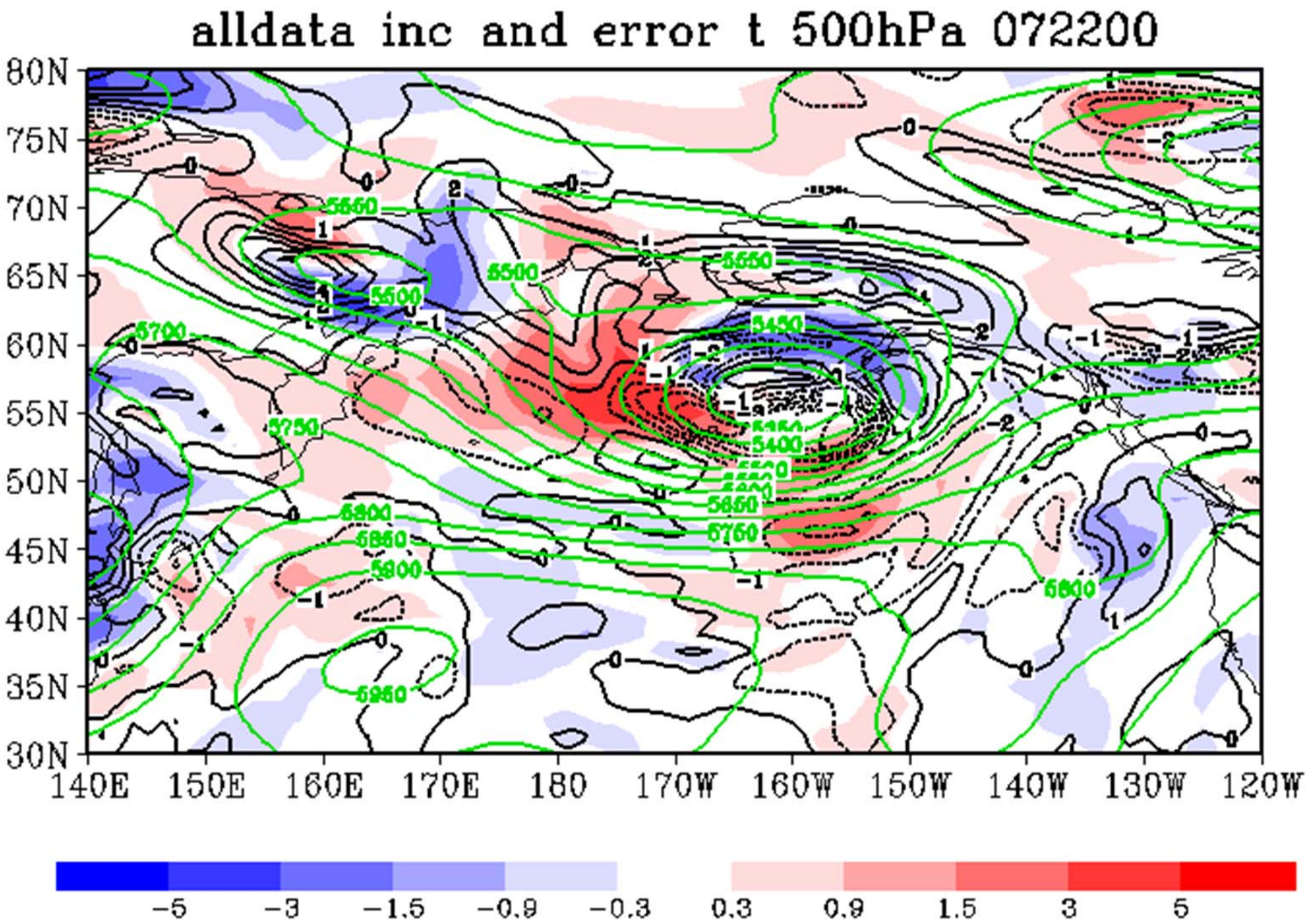
- Color shade: Optimized increments, red=plus, blue=minus.
- Green contour: 500hPa height.
- Black contour: Integrated background error, solid lines=plus, dotted lines=minus.

Comparison between original analysis and optimized analysis

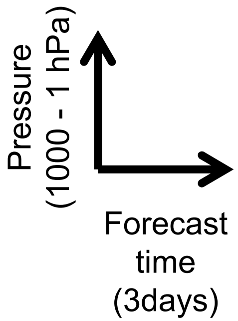
Optimized :FT48 500hPa T



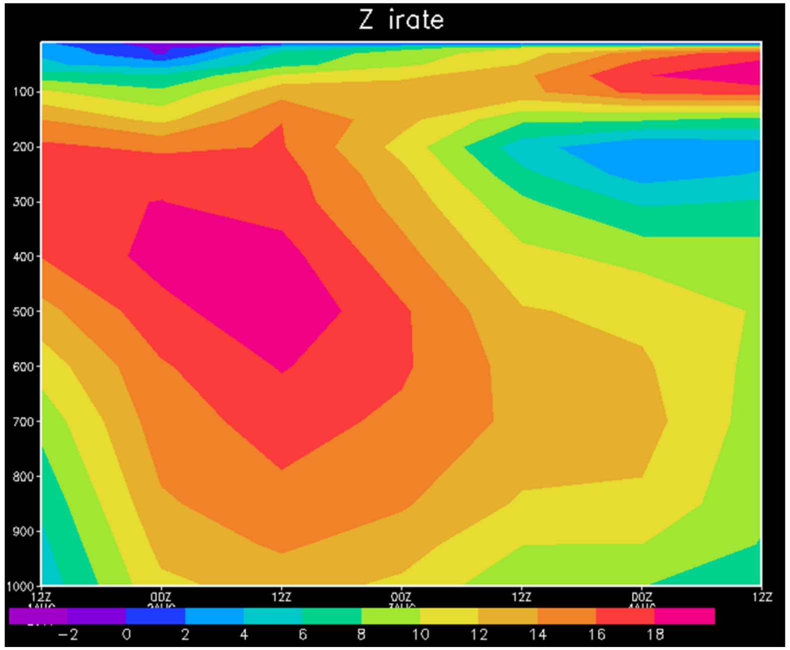
Original: FT48 500hPa T



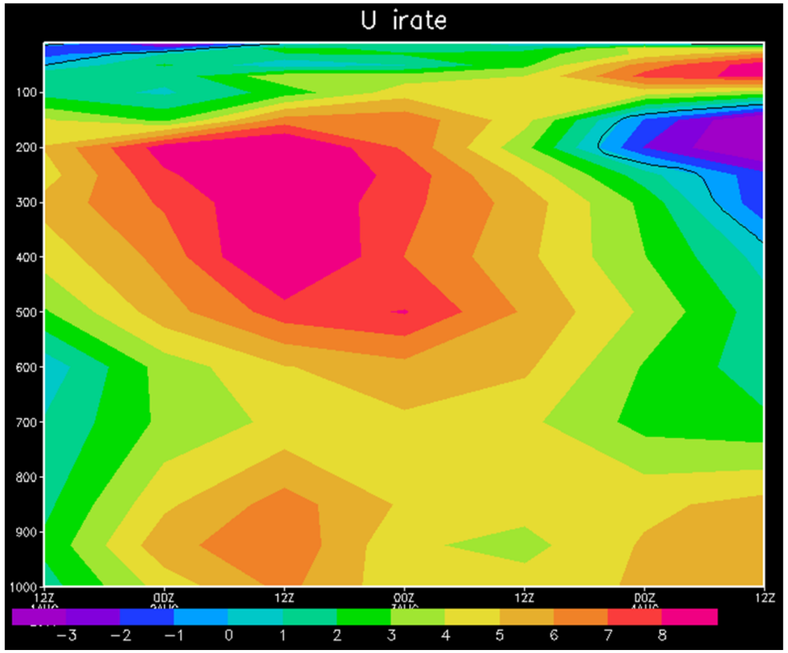
Pseudo truth with ADJ-based method

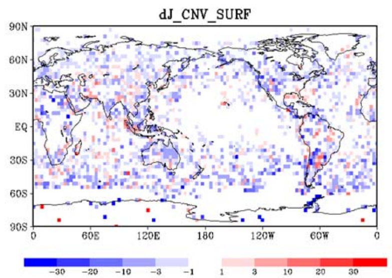


Improvement rate: Height

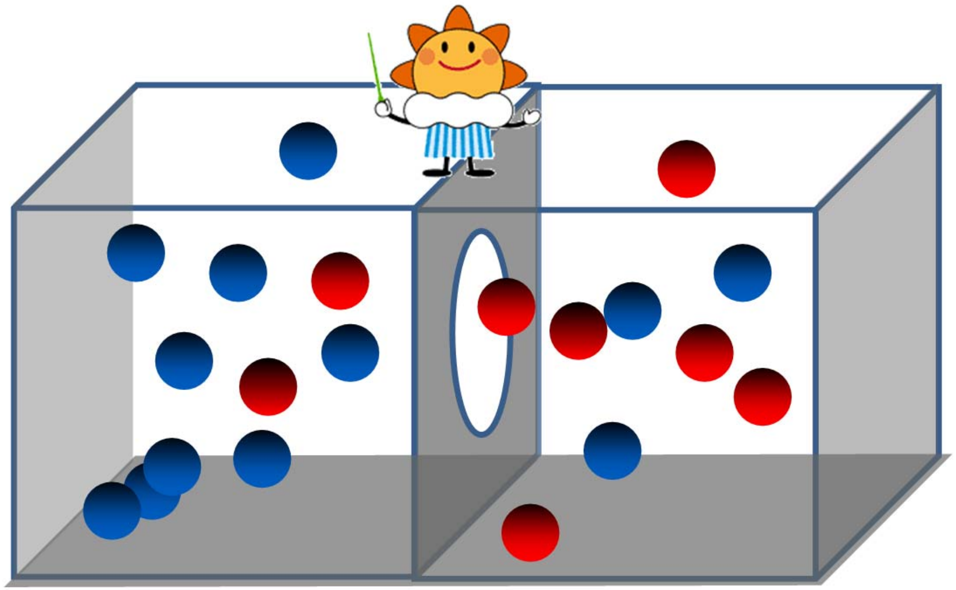


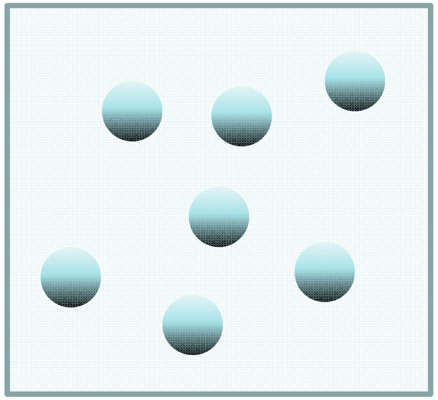
Improvement rate: Zonal wind



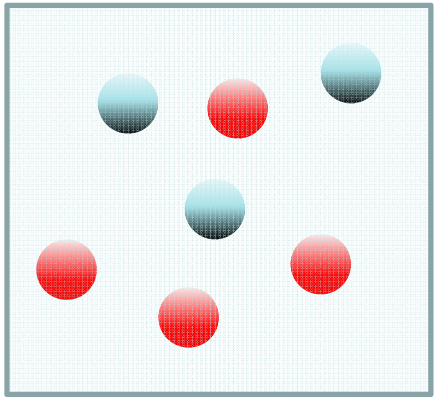


Maxwell's demon ?





We know only statistical property of data, R and B.



Lets think about a system on thermal equilibrium at temperature T. We know only statistical property of the system, temperature T. While, if one can know velocity of each particle, the one can get usable energy from this max entropy state, This is the Maxwell's demon.

We know property of each observation and can use this information.

Summary

Observation impact

- We defined two types of observation impact; the linear impact and the non-linear impact.
- Diagnoses of the JMA global 4D-Var shows almost all observation data types contribute forecast error reduction in monthly average.
- The diagnoses imply that it is possible to derive more information from radiance data by improving usage of these data and operators.
- The TL-based method was introduced.
- We can see time evolution and space distribution of linear observation impacts, and evaluate them by comparison with those of integrated background errors.

Covariance matrix optimization

- Optimization methods include observation impact estimations.
- Sensitivity based method diagnosed the JMA GDAS has too large (small) **R** (**B**).
- The single case experiment of optimization showed the explicit forecast error reductions.

Analysis error estimation

- We constructed new method based on data assimilation theory. The method assimilate reference analysis fields.
- The method reduce forecast error and also consistent with observations, if adequate inflation factor is given.
- ADJ-based method can be used to generate improved forecasts, so it may be possible to be used as pseudo truth.