4D-Var data assimilation system for a limited-area model in JMA and the assimilation of precipitation amounts

Ko Koizumi

Forecast Research Department
Meteorological Research Institute/Japan Meteorological Agency

Symposium: 20 years of 4D-Var at ECMWF, Reading, U.K., 26 Jan., 2018

Background

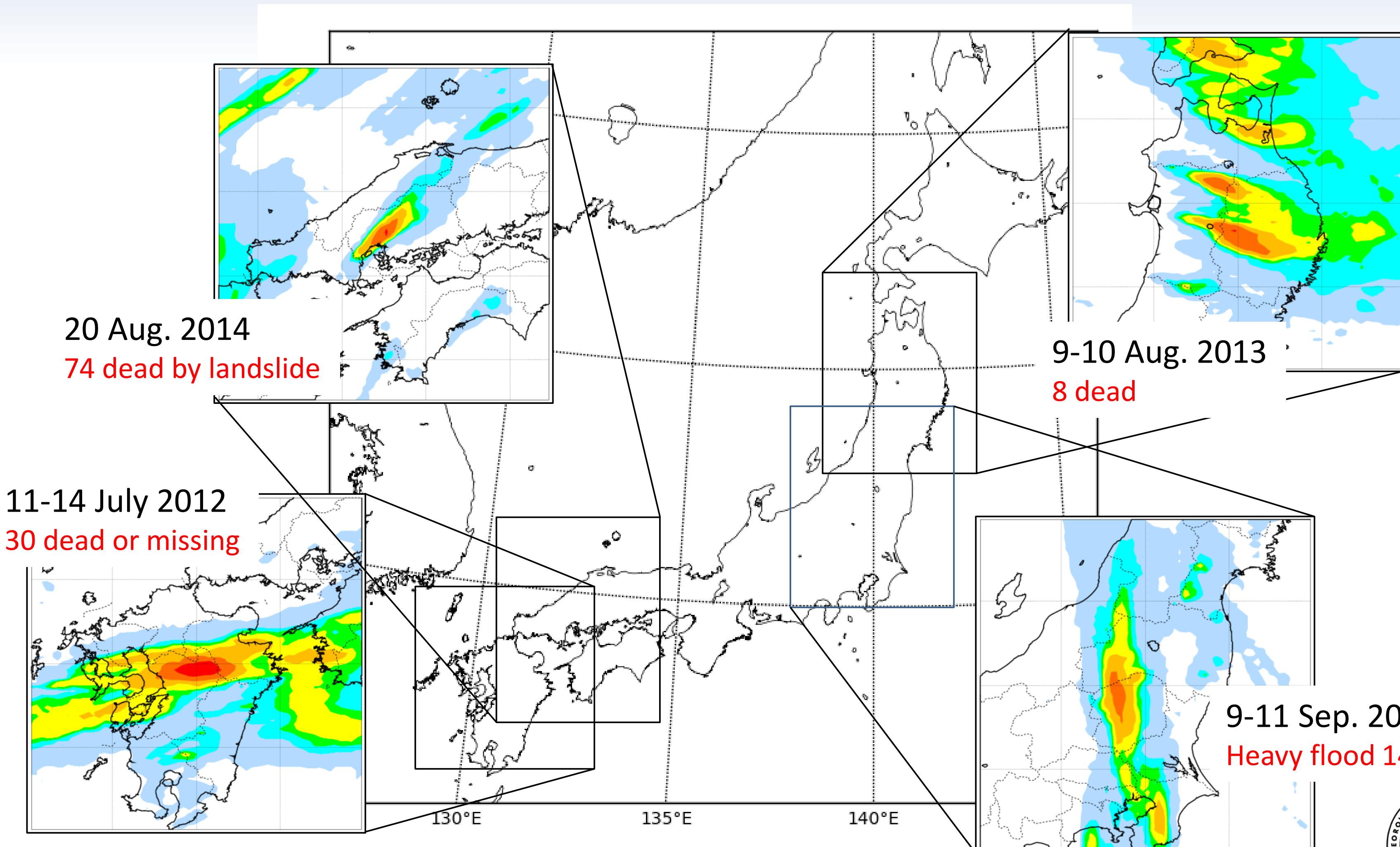
Various natural disasters in Japan

Pyroclastic flow from Mt. O
27 Sep. 2014 ©Ministry of Land

moto pref.

The 2011 off the Pacific coast of Tohoku Earthquake (M9.0)
11 March 2011 18456 dead or missing

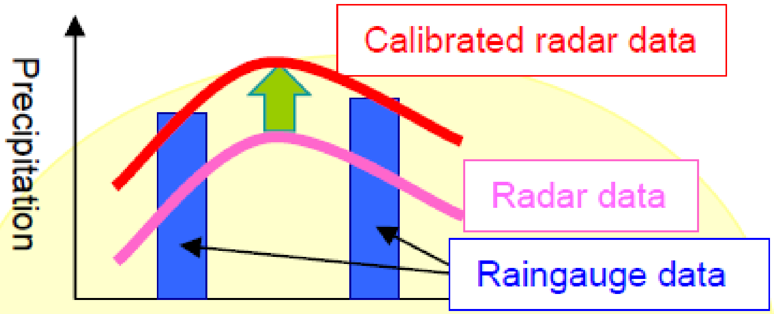
Deadly heavy rain events occur almost every year



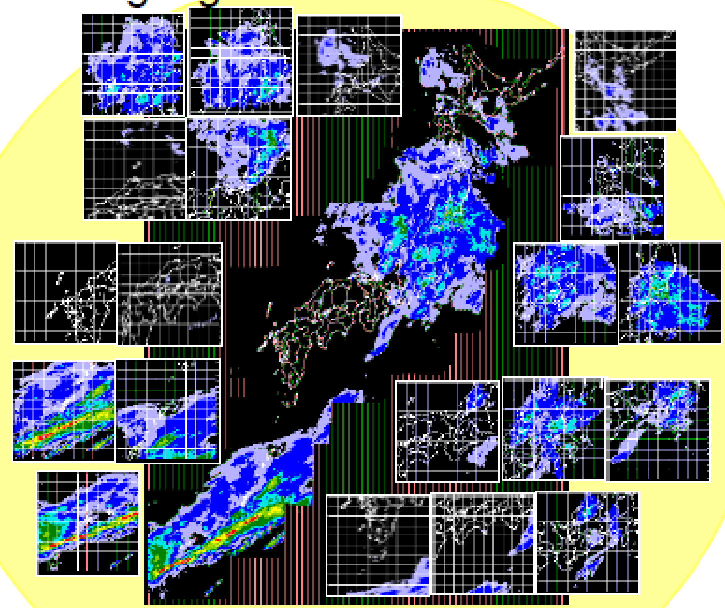
Watch and Prediction of heavy rainfall are crucial for disaster prevention and mitigation.

WATCH: precipitation analysis

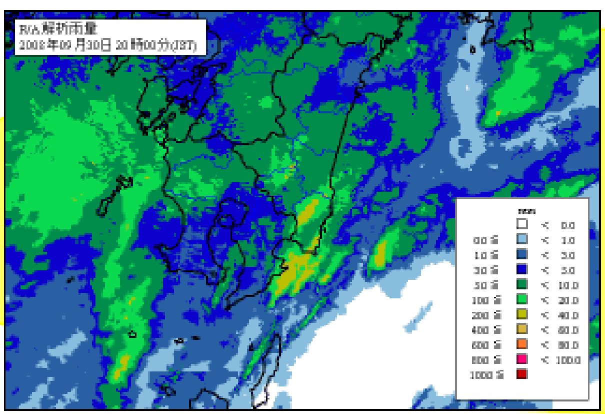
JMA has created an optimal mix of the advantages found in raingauge data and radar data.



Precipitation amounts observed by radar generally does not agree with those observed by raingauges, and radar data are therefore calibrated with raingauge data.



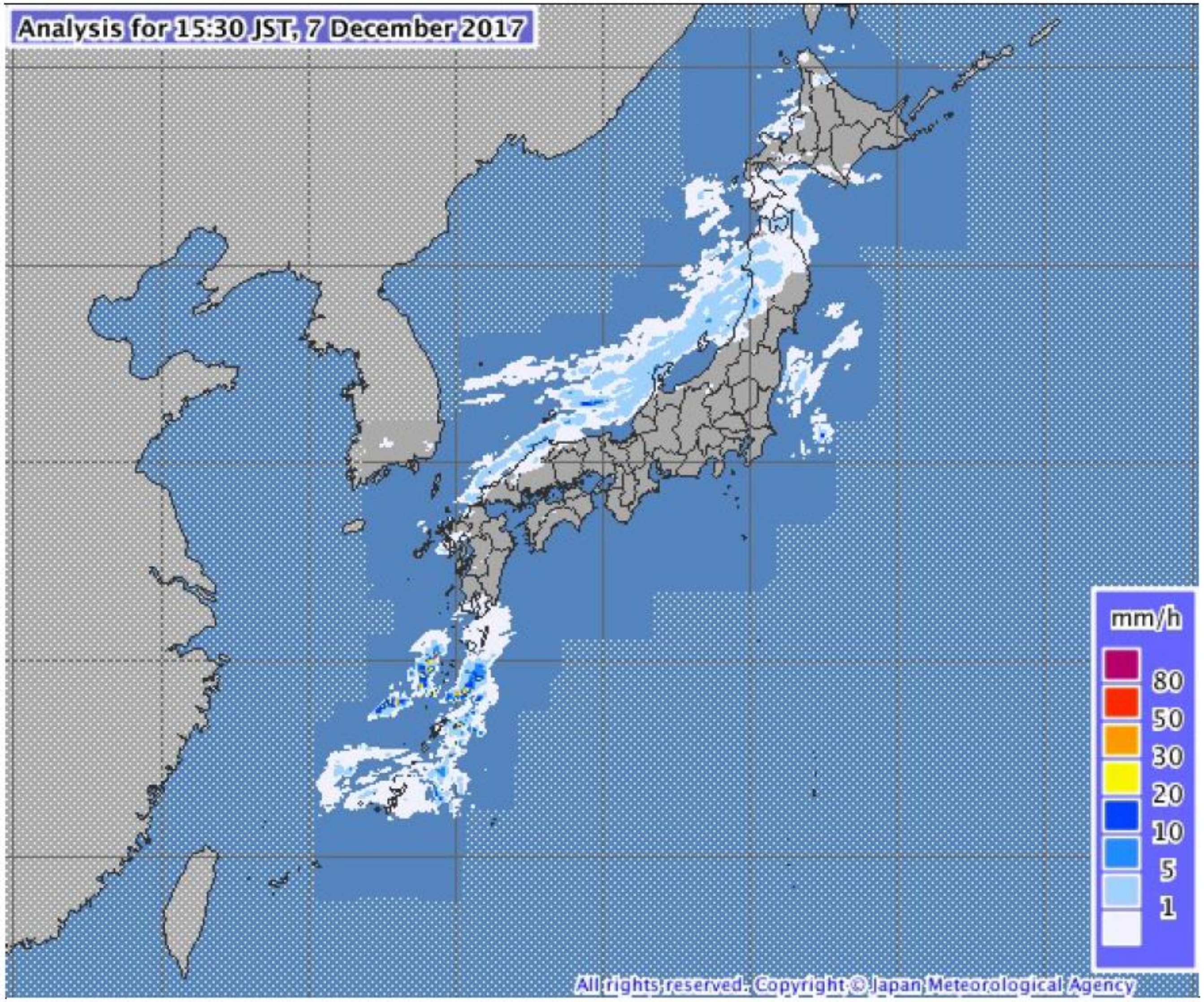
The calibrated radar data are then made into a single composite data set.



Radar/Raingauge-Analyzed Precipitation

Radar/Raingauge-Analyzed Precipitation data depicts hourly precipitation with high dimensional accuracy, and is issued every thirty minutes with a spatial resolution of 1 km.

'Watch' ⇨ 'Prediction'



Precipitation amounts (not just a synthesized radar echo)

⇩

If this quantitative value is assimilated to NWP, short-range forecasts by the model would improve … but How to assimilate?

JMA MesoScale Model (MSM)

- Operation started in 2001 as a hydrostatic spectral model with resolution of 10km, providing 18-hour forecasts four times a day
- Initial condition was provided by
  - Optimum Interpolation Method (for ordinary observation data)
  - Physical Initialization (for precipitation amounts)

    (if p

    - W
    - Sensible heat which is estimated from precipitation amounts is added to the column

PI tended to produce too much or false precipitation

A better assimilation method was required.

So we started development of 4D-Var assimilation system for MSM around year 2000.

Why 4D-Var?

Because the precipitation amounts have the dimension of time

Specification of the 4D-Var system

The "inner" model (used for iterative calculation)

- Grid distance: 20km (incremental approach)
- Forward: non-linear, backward: linear (the background field is updated every iteration)
- Some physical processes in the adjoint model were simplified or omitted

Assimilation window: Three hours before analysis time

Control variables: $\{u_U, v_U, (T_v, p_s), q\}$ in grid space including boundary

$$\begin{pmatrix} \Delta u_U \\ \Delta v_U \end{pmatrix} \equiv \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} - \begin{pmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{pmatrix} \begin{pmatrix} \Delta u_g \\ \Delta v_g \end{pmatrix}$$

Vertical correlation matrices are decomposed to eigenvectors

Some difficulties in the development

1. LARGE Dimension of background error covariance (B)

2. A strange (non-Gaussian) type of observation error probability distribution of precipitation amount

Size of B (horizontal correlation)

For a global model, assumption of homogeneity and isotropy of background errors reduces their horizontal covariance matrices to be diagonal in spectral space.

However…

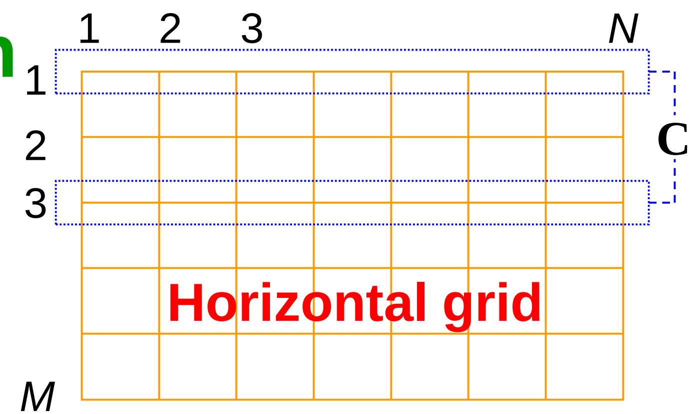
For a limited-area model, even assuming homogeneity and isotropy, the background error covariance matrix CANNOT be made diagonal (even in the spectral space).

Size of B (horizontal correlation) (cont.)

Assuming homogeneous Gaussian-type error correlation

$$\left[\mathbf{C}^{(k)}\right]_{ij;i'j'} = \exp\left[-\left(\frac{i-i'}{d_x^{(k)}}\right)^2 - \left(\frac{j-j'}{d_y^{(k)}}\right)^2\right] \rightarrow \mathbf{C}_j^{(k)} = \varepsilon_j^{(k)}\,\mathbf{C}_1^{(k)}$$

Then, Cholesky decomposition is applied



$$\mathbf{C}^{(k)} = \mathbf{L}^{(k)}\mathbf{L}^{(k)\mathrm{T}}, \quad \mathbf{L}^{(k)} = \begin{pmatrix} \mathbf{L}_{11}^{(k)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \mathbf{L}_{M1}^{(k)} & \cdots & \mathbf{L}_{MM}^{(k)} \end{pmatrix}$$

$$\mathbf{L}_{ij}^{(k)} = a_{ij}^{(k)}\mathbf{L}_{11}^{(k)}, \quad \mathbf{C}_1^{(k)} = \mathbf{L}_{11}^{(k)}\mathbf{L}_{11}^{(k)\mathrm{T}}$$

Size of B (horizontal correlation) (cont.)

- With Cholesky decomposition of B, an error covariance matrix of new variable u is identity matrix

$$\mathbf{B} = \mathbf{L}\mathbf{L}^{\mathrm{T}} \qquad \mathbf{x} = \mathbf{L}\mathbf{u}$$

$$J_b = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) = (\mathbf{u} - \mathbf{u}_b)^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}(\mathbf{L}\mathbf{L}^{\mathrm{T}})^{-1}\mathbf{L}(\mathbf{u} - \mathbf{u}_b)$$

$$= (\mathbf{u} - \mathbf{u}_b)^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}^{-\mathrm{T}}\mathbf{L}^{-1}\mathbf{L}(\mathbf{u} - \mathbf{u}_b) = (\mathbf{u} - \mathbf{u}_b)^{\mathrm{T}}(\mathbf{u} - \mathbf{u}_b)$$

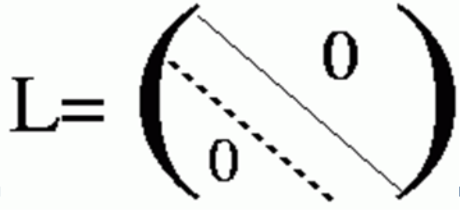
- Problem is, $\mathbf{x} = \mathbf{L}\mathbf{u}$ is necessary for each 4D-Var iteration, and L is still HUGE.

In our first implementation, $\mathbf{x} = \mathbf{L}\mathbf{u}$ took almost 80% of calculation time! It's not tolerable!

Size of B (horizontal correlation) (cont.)

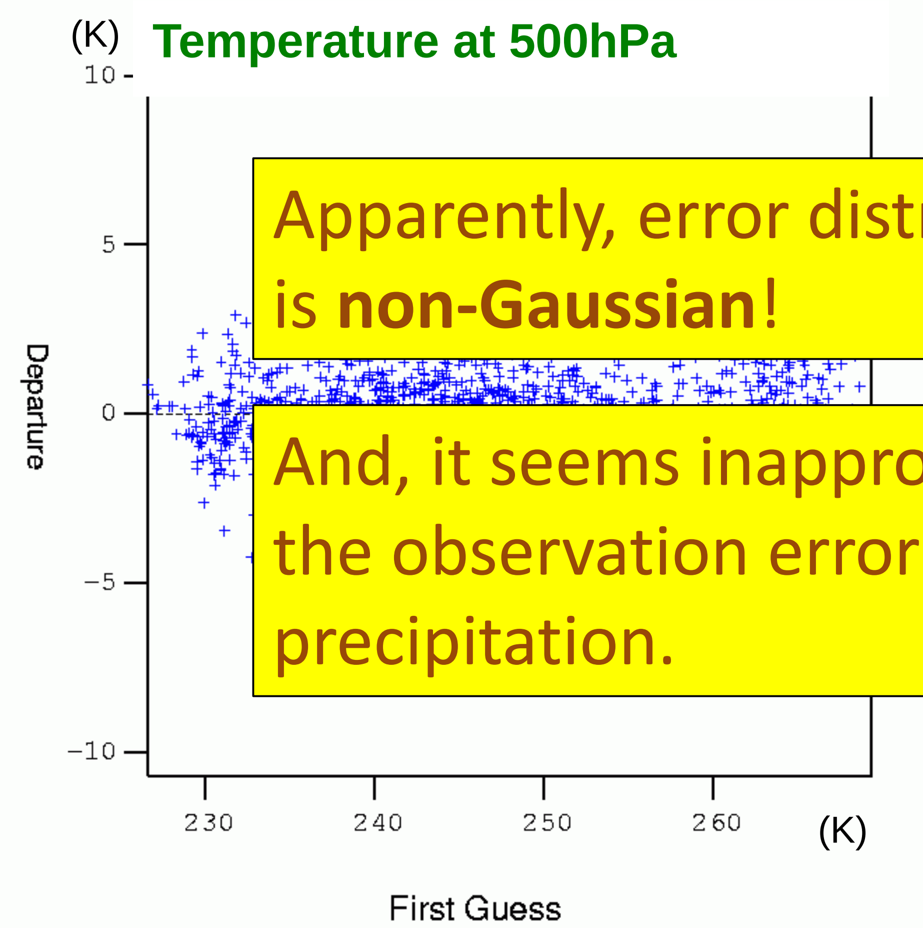
So, what did we do?

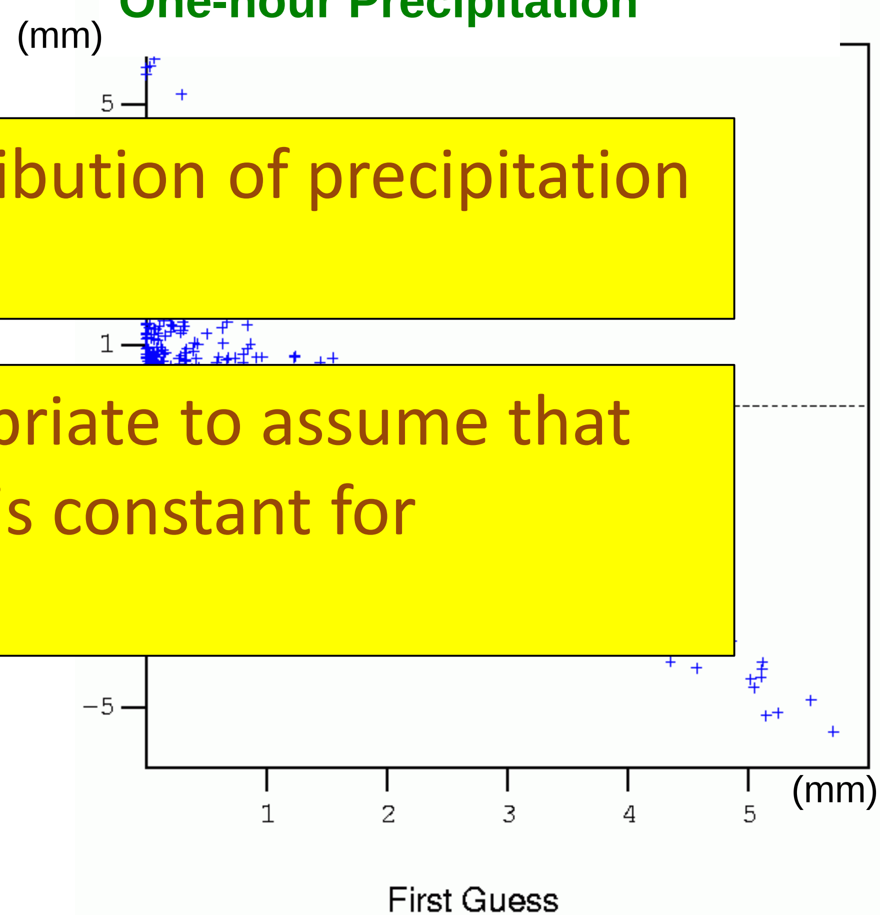
- As the horizontal correlation is assumed to be a Gaussian function of grid distance, correlation between distant grids might be negligible.
- In order to reduce computational time of x=Lu , once a term becomes smaller than a certain threshold value, the calculation of remaining rows are skipped.



Observation error of precipitation amounts

Scattering diagram of Observation Departure　(d = $\mathbf{y}_o - H\mathbf{x}_b$)





Apparently, error distribution of precipitation is **non-Gaussian**!

And, it seems inappropriate to assume that the observation error is constant for precipitation.

Observation error of precipitation amounts

Assuming exponential distribution for conditional PDF of precipitation

$$p_{\text{precip}}(y_o \mid y) = \frac{1}{y} e^{-\frac{y_o}{y}}, \qquad y = H\mathbf{x} \qquad (y_o, y > 0)$$

Then deriving observational term from the PDF

$$J_{\text{precip}}(y) = -\log p_{\text{precip}}(y_o \mid y) = \frac{y_o}{y} + \log y$$

$$= \frac{(y - y_o)^2}{2 y_o{}^2} + O\left[(y - y_o)^3\right] + 1 + \log y_o$$

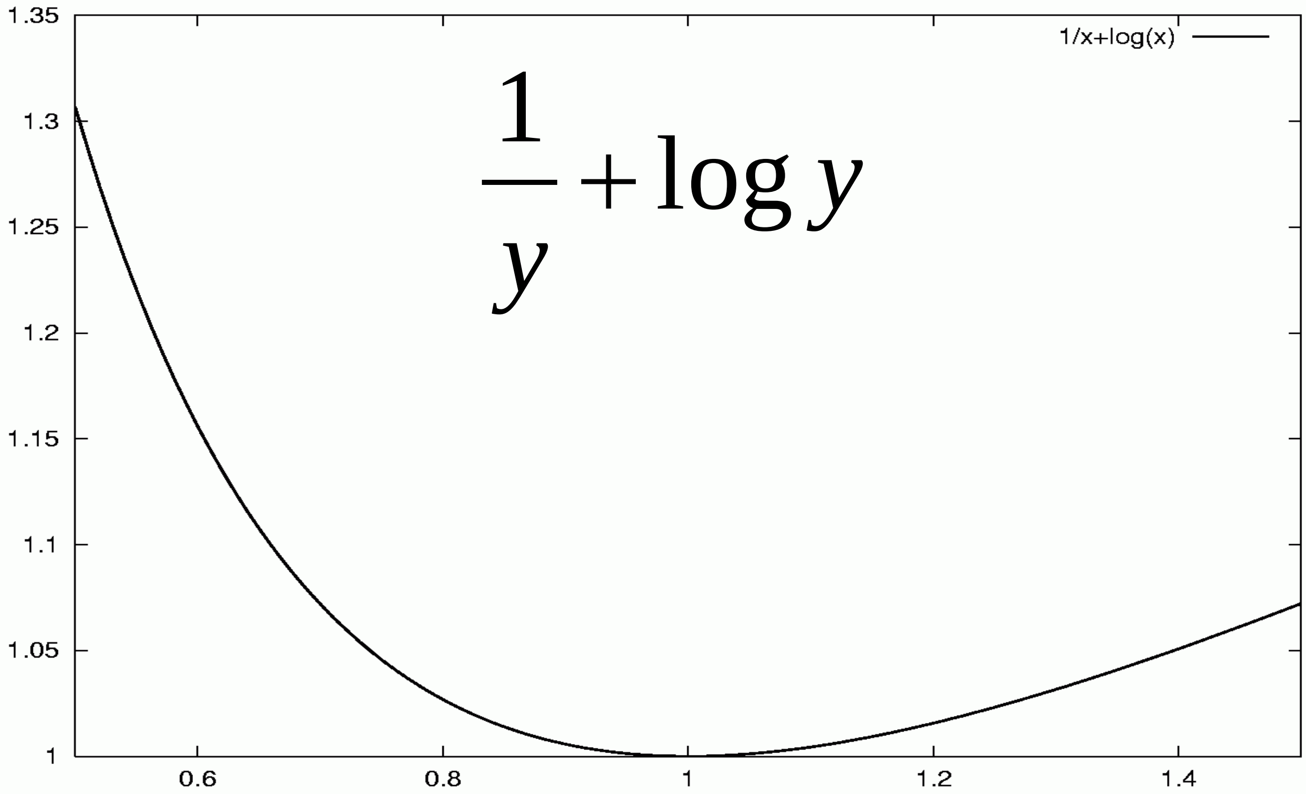
$$\approx \frac{(y - y_o)^2}{2 \sigma_o{}^2} + \text{const.}$$

Practically,

$$\sigma_o \equiv \begin{cases} \sigma & (y \le y_o) \\ 3\sigma & (y > y_o) \end{cases}, \qquad \sigma \equiv \begin{cases} 1\text{mm/h} & (y_o \le 1\text{mm/h}) \\ y_o & (y_o > 1\text{mm/h}) \end{cases}$$

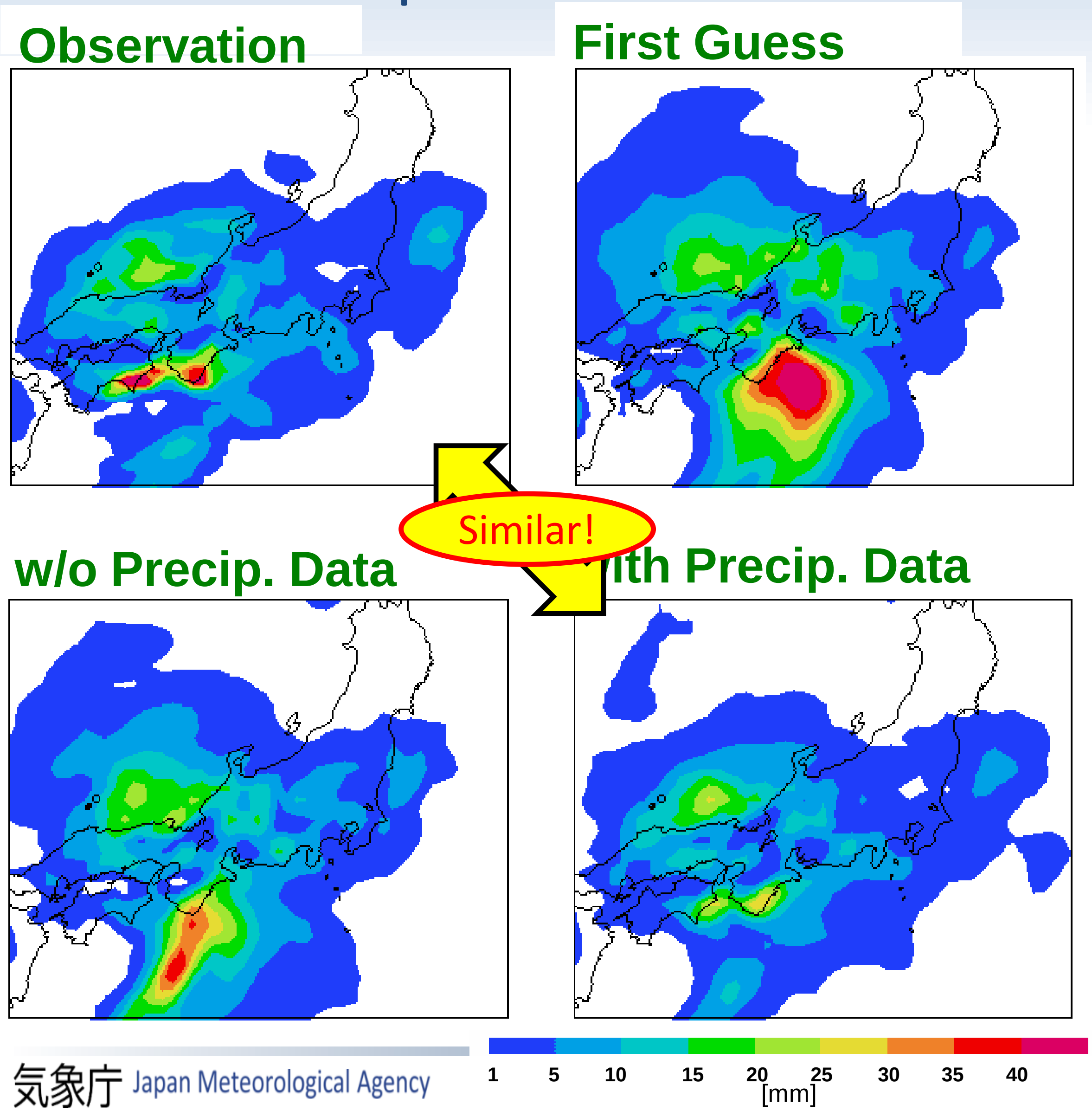
Koizumi, K., Y. Ishikawa and T. Tsuyuki, 2005: Assimilation of Precipitation Data to the JMA Mesoscale Model with a Four-dimensional Variational Method and its Impact on Precipitation Forecasts. SOLA, 1, 45-48

Log-likelihood of exponential distribution is asymmetric



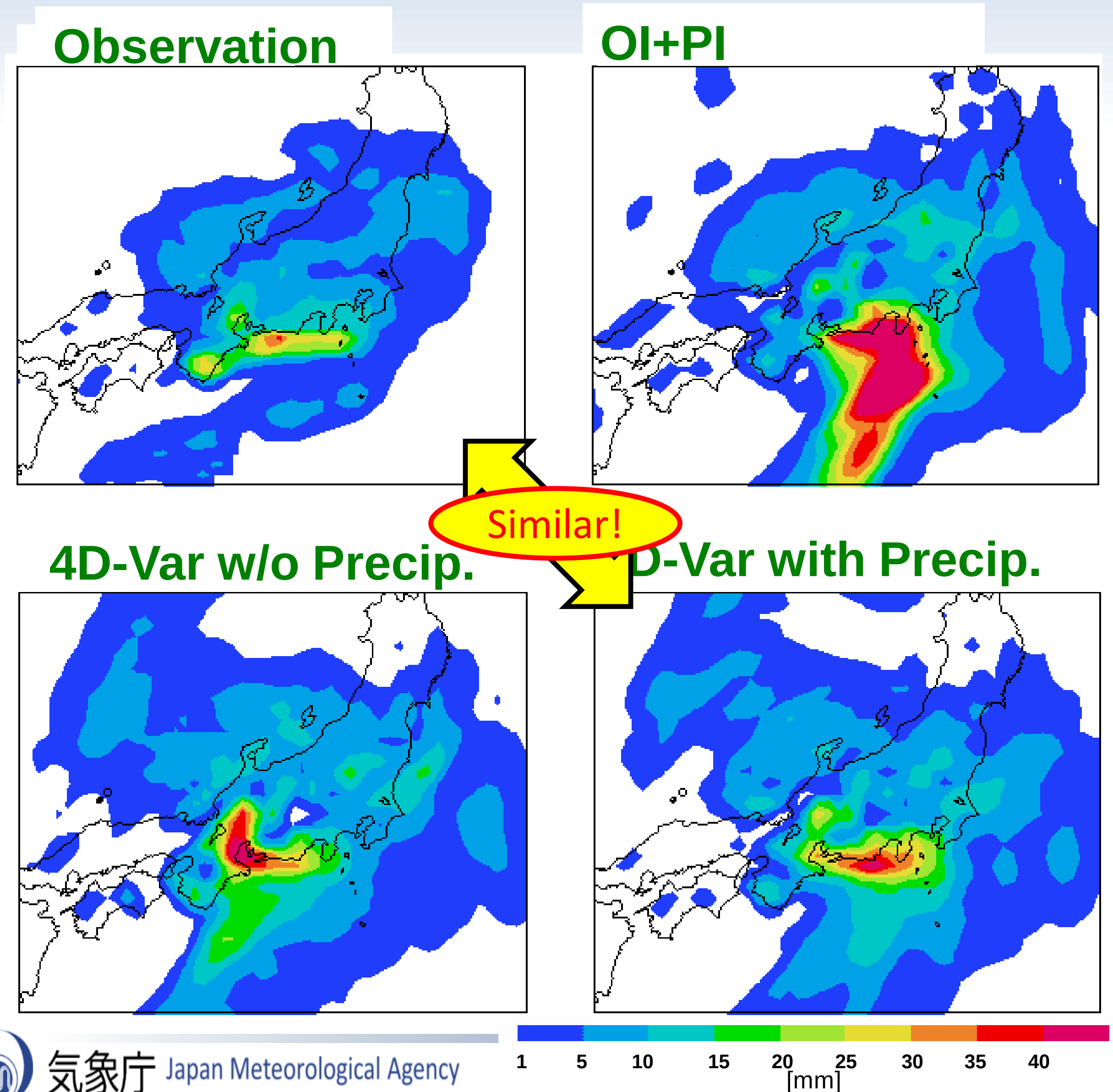
It has worked!

-Precipitation within assimilation window-



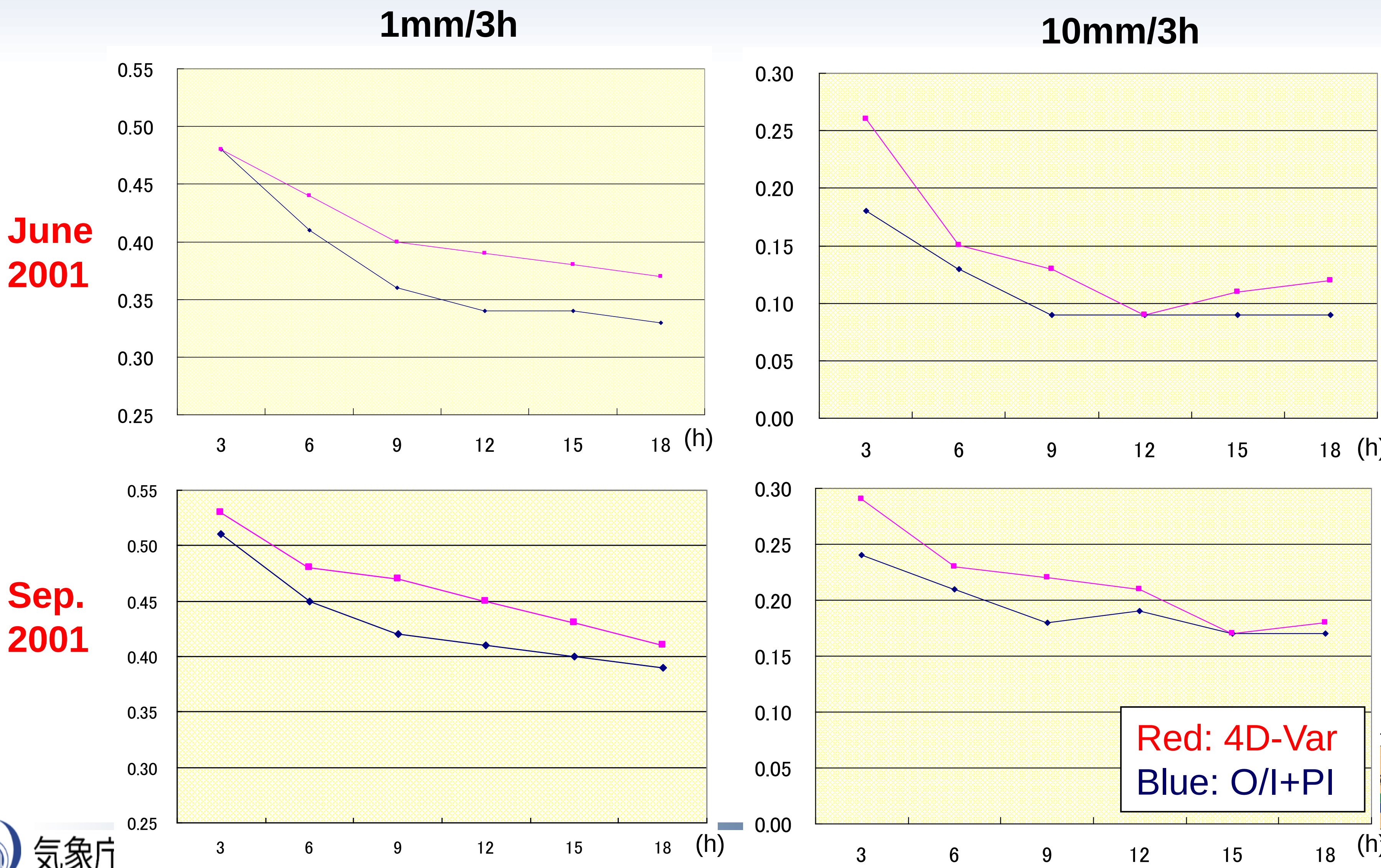
(21-24 UTC 15 March 2000)

Precipitation Forecasts (First 3-hour)



(0-3 UTC 16 March 2000)

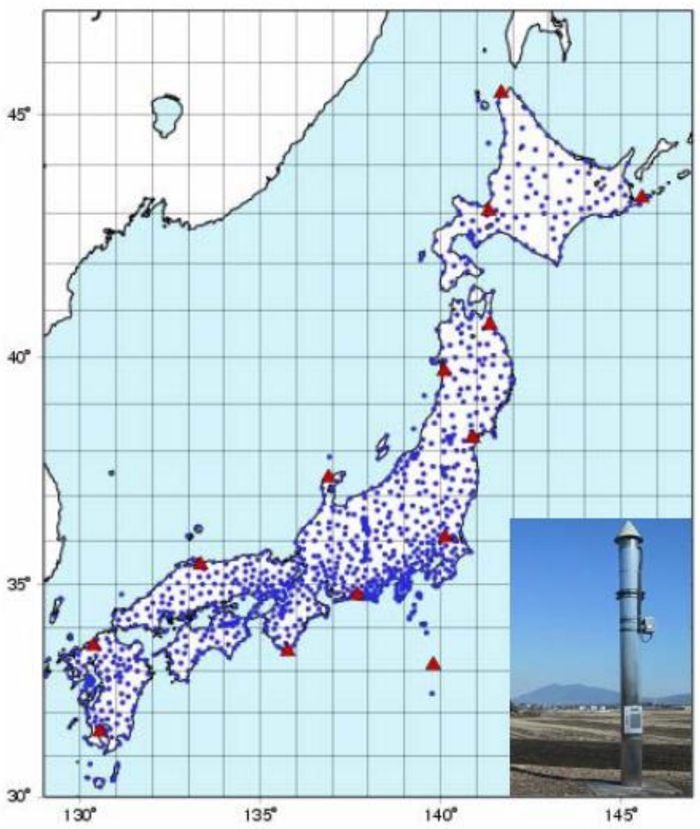
Verification of 3-hour precipitation forecasts on 40km meshes



Operation of 4D-Var data assimilation system for MSM started in 2002

World first 4D-Var for an operational limited-area NWP model, maybe

Moisture field is also crucial for precipitation forecasts

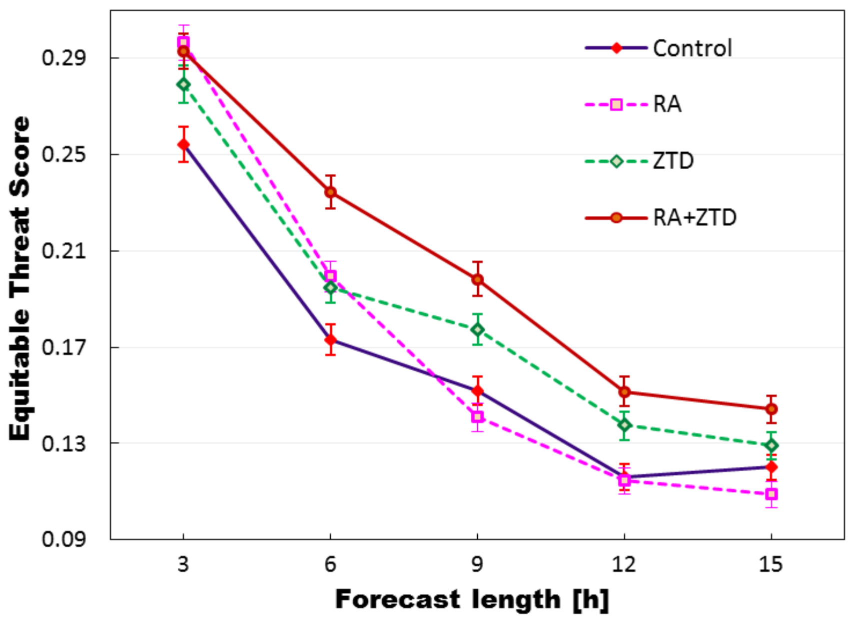


- JMA analyses Zenith Total Delay (ZTD) at over 1,000 ground-based GNSS receivers owned by Geographical Survey Institute.
- The hourly product is provided on real-time basis.

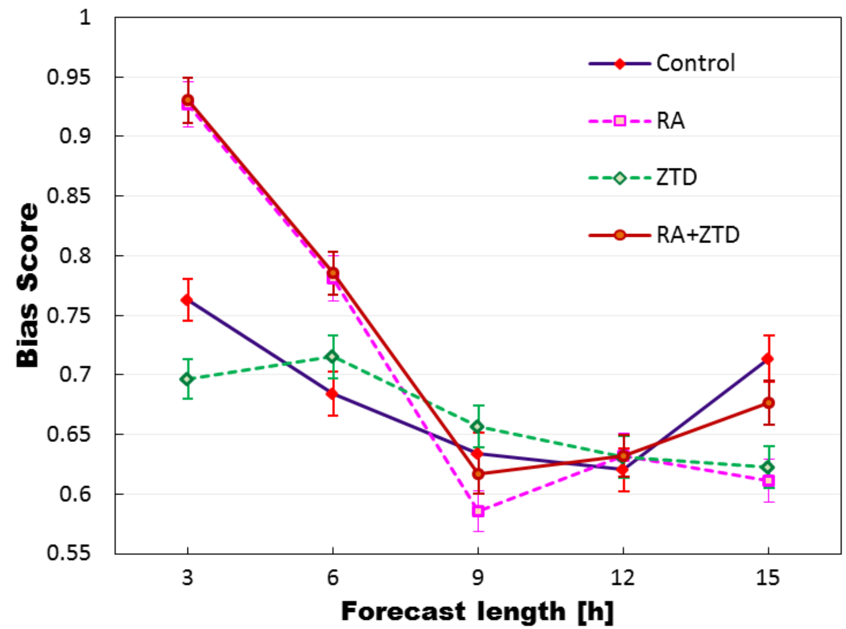
Assimilation of both precipitation amount and moisture provides better forecasts

Equitable threat scores and bias scores for 3 hour precipitation for 124 forecasts (15-hour forecast 4 times a day) for forecast-analysis cycles experiment during one month period of July 2006

ETS [Threshold 10mm/3hour]



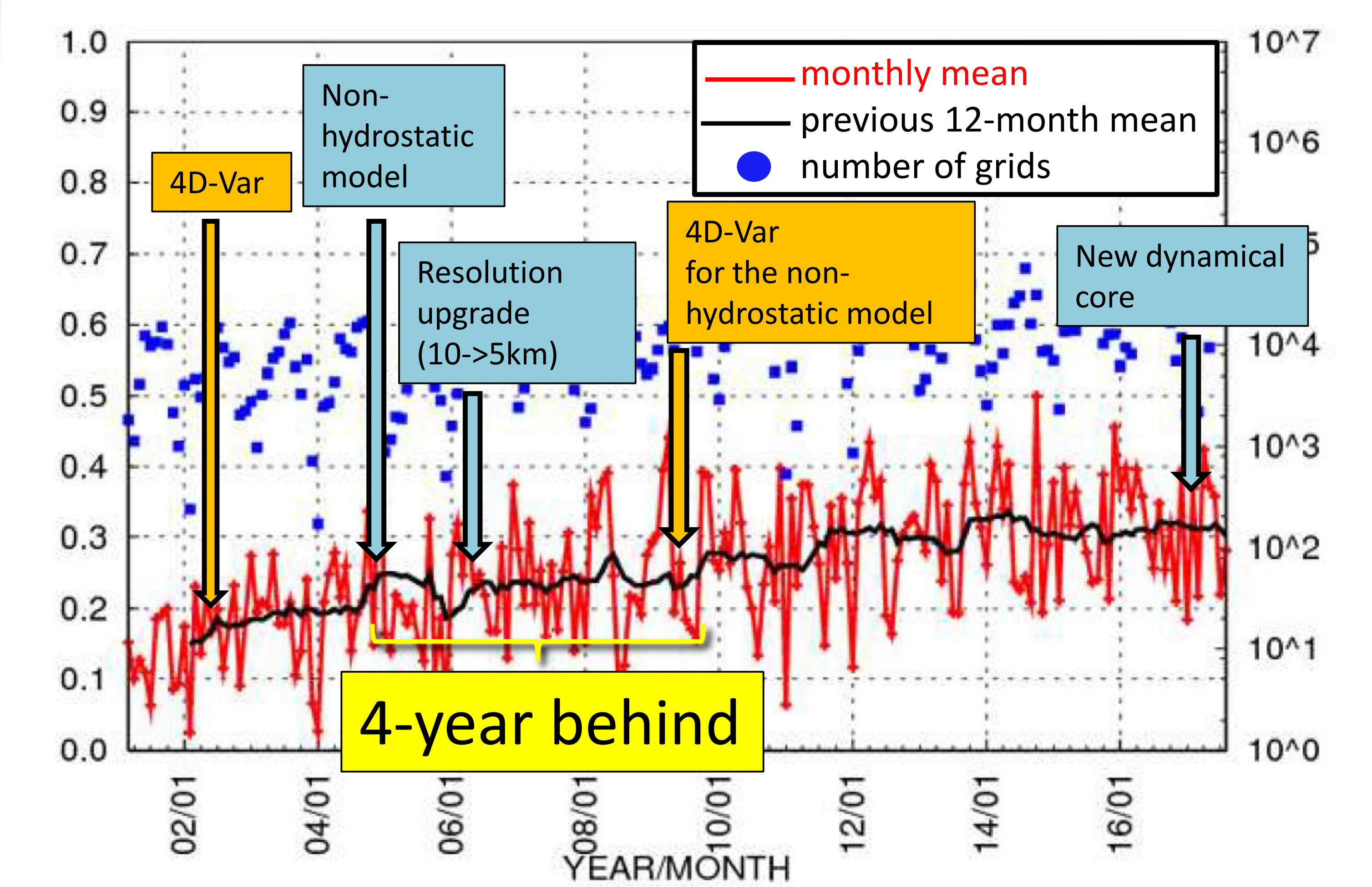
BIAS [Threshold 10mm/3hour]



Happy ever after? … NO!

- Development of tangent-linear codes of the NWP model and their adjoints is costly.
- That is the reason why upgrade of the assimilation system falls behind the model upgrade for several years.

Equitable Threat Score of MSM forecasts for 10mm/hour precipitation



For sustainable development

- Currently, developers of the latest model (asuca) also work for development of its TL/AD
  - Pros: they know the model well, so it is relatively easy for them to decide which parts in the model

We are still searching for a better way!

  - Cons: TL/AD development requires some extra programming techniques, meaning extra education is necessary for the developers.

Thank you!