The JMA Mesoscale 4D-Var System and Assimilation of Precipitation and Moisture Data

Tadashi Tsuyuki*, Ko Koizumi and Yoshihiro Ishikawa*****

** Meteorological Research Institute, Tsukuba, Japan*
*** Japan Meteorological Agency, Tokyo, Japan*
**** National Severe Storms Laboratory, Oklahoma, USA*

1. Introduction

In March 2001 the Japan Meteorological Agency (JMA) started the operation of the mesoscale model (MSM) to produce 18 hour forecasts four times a day (00, 06, 12 and 18 UTC initial) to assist forecasters in issuing warnings (JMA, 2002). Its products are provided within 1.5 hours from the time of initial conditions. MSM is a hydrostatic spectral model with a horizontal resolution of 10 km and 40 vertical levels up to 10 hPa. Figure 1 shows the forecast domain of MSM. The lateral boundary condition is provided by the regional spectral model (RSM) with a horizontal resolution of 20 km starting from initial conditions at 00 and 12 UTC.

The initial condition of MSM was at first prepared by a 1-hour cycle analysis system with optimum interpolation and physical initialization to assimilate 1-hour accumulated precipitation data from radar and rain gauge observations. This analysis system was executed for the 3-hour period just before the initial time with the first guess at the beginning of the period taken from the latest forecast of RSM. This analysis system is hereafter referred to as the pre-run system. The pre-run system was successfully replaced in March 2002 by a full forecast-analysis cycle with a 4-dimentional variational (4D-Var) method with 3-hour assimilation windows.

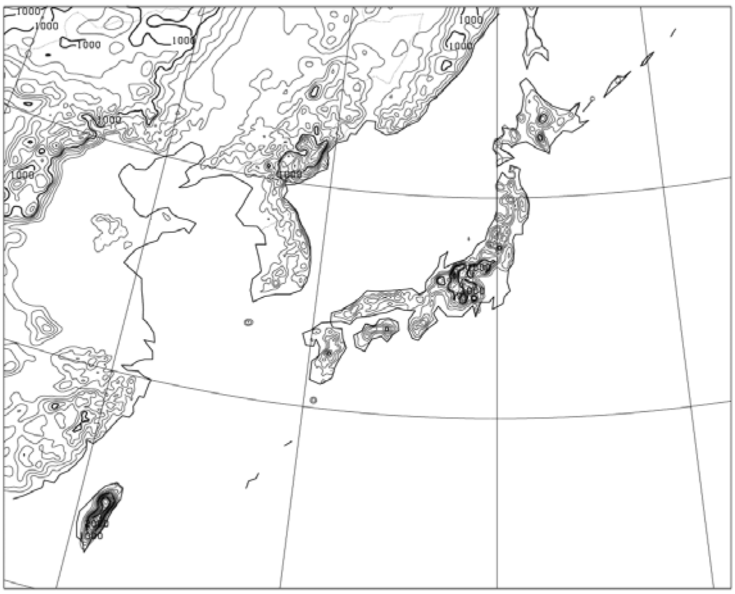

Figure 1 Forecast domain and topography of MSM. Contour interval is 200 m

This paper describes the outline of the mesoscale 4D-Var system and reports on results from parallel run tests conducted before the operational implementation of the system. Several observing system experiments (OSEs) with the 4D-Var system were also carried out to evaluate benefits of some observational data, and preliminary results from OSEs for the precipitation data and the GPS precipitable water data are presented in this paper.

2. Background error covariance

The state variables of MSM are wind (u, v), virtual temperature T_v, surface pressure p_s and specific humidity q in spectral space. The background error covariance matrix is constructed in grid space to easily handle the lateral boundary condition. The control variables of the 4D-Var system are unbalanced wind (u_U, v_U), T_v, p_s and q in grid space at the beginning of the assimilation window and the lateral boundary conditions at the beginning and end of the assimilation window. The lateral boundary conditions at the other time levels are obtained from linear interpolation. The unbalanced wind is defined as

$$\begin{pmatrix} u_U \\ v_U \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} u_g \\ v_g \end{pmatrix} \tag{1}$$

where (u_g, v_g) denote the geostrophic wind and $\{r_{ij}\}$ are regression coefficients. It is assumed that the control variables $\{u_U, v_U, (T_v, p_s), q\}$ are uncorrelated with each other.

The control variables normalized by their background error standard deviations are expanded in the eigenvectors of the background error vertical correlation matrices that are assumed to be homogeneous in the horizontal direction:

$$\begin{pmatrix} x_1/\sigma_1 \\ \vdots \\ x_K/\sigma_K \end{pmatrix} = \mathbf{U} \begin{pmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_K \end{pmatrix} \tag{2}$$

where x_k denotes one of the control variables at a grid point at the k-th level, σ_k the background error standard deviation of x_k, $\mathbf{U}$ the orthogonal matrix consisting of the eigenvectors, $\tilde{x}_k$ the k-th expansion coefficient, and K the number of vertical levels. It is assumed that the background error standard deviations are horizontally homogeneous.

The horizontal correlations of the expansion coefficients are also assumed to be horizontally homogeneous. Then the horizontal correlation matrix of the k-th coefficient is written as

$$\mathbf{C}^{(k)} = \begin{pmatrix} \mathbf{C}_1^{(k)} & \mathbf{C}_2^{(k)} & \cdots & \mathbf{C}_N^{(k)} \\ \mathbf{C}_2^{(k)} & \mathbf{C}_1^{(k)} & & \mathbf{C}_{N-1}^{(k)} \\ \vdots & & \ddots & \vdots \\ \mathbf{C}_N^{(k)} & \mathbf{C}_{N-1}^{(k)} & \cdots & \mathbf{C}_1^{(k)} \end{pmatrix} \tag{3}$$

where $\mathbf{C}_j^{(k)}$ is the $M \times M$ - correlation matrix between the k-th expansion coefficients at the grid points in two rows that lie in the zonal direction and are separated by $(j-1)\Delta y$ in the meridional direction. M and N denote the numbers of grid points in the zonal and meridional directions, respectively, and Δy the meridional grid interval.

Horizontal correlation functions are assumed to be of a Gaussian type. Then the sub correlation matrices $\{\mathbf{C}_j^{(k)}\}$ are proportional to $\mathbf{C}_1^{(k)}$:

$$\mathbf{C}_j^{(k)} = \exp\left[-\frac{1}{2} \left(\frac{(j-1)\Delta y}{\eta^{(k)}} \right)^2 \right] \mathbf{C}_1^{(k)} \tag{4}$$

where $\eta^{(k)}$ is the meridional correlation length of the k -th expansion coefficient. The Cholesky decomposition matrix of $\mathbf{C}^{(k)}$ is also written by the Cholesky decomposition matrix of $\mathbf{C}_1^{(k)}$:

$$\mathbf{C}^{(k)} = \mathbf{L}^{(k)}\mathbf{L}^{(k)\mathrm{T}},\ \mathbf{L}^{(k)} = \begin{pmatrix} a_{11}^{(k)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{N1}^{(k)} & \cdots & a_{NN}^{(k)} \end{pmatrix} \otimes \mathbf{L}_1^{(k)} \tag{5}$$

where $\left\{a_{ij}^{(k)}\right\}$ are scalars calculated from the scalar coefficient in Eq. (4), $\otimes$ the tensor product, and $\mathbf{L}_1^{(k)}$ the $M \times M$ -lower triangular matrix that satisfies the following equation.

$$\mathbf{C}_1^{(k)} = \mathbf{L}_1^{(k)}\mathbf{L}_1^{(k)\mathrm{T}} \tag{6}$$

The background error statistics were obtained from differences between 18- and 6-hour forecasts for the same valid time using the NMC method (Parrish and Derber, 1992). The elements of the correlation matrix $\mathbf{C}^{(k)}$ of which absolute values were less than 0.0001 were neglected to save computational time (Koizumi, 2001).

Figure 2 compares the analysis increments from a single pseudo observation of wind for the 3-dimensional variational (3D-Var) and 4D-Var methods using the same background error covariance. In the latter method the observation is put at the end of the 3-hour assimilation window, and the analysis increment at that time is shown. The observation site is located upstream of a wind shear line seen in the first guess, and the observed wind is set to be more westerly than the first guess. The analysis increment from 4D-Var indicates that the analyzed shear line is shifted eastward, and it seems more plausible than the increment from 3D-Var.

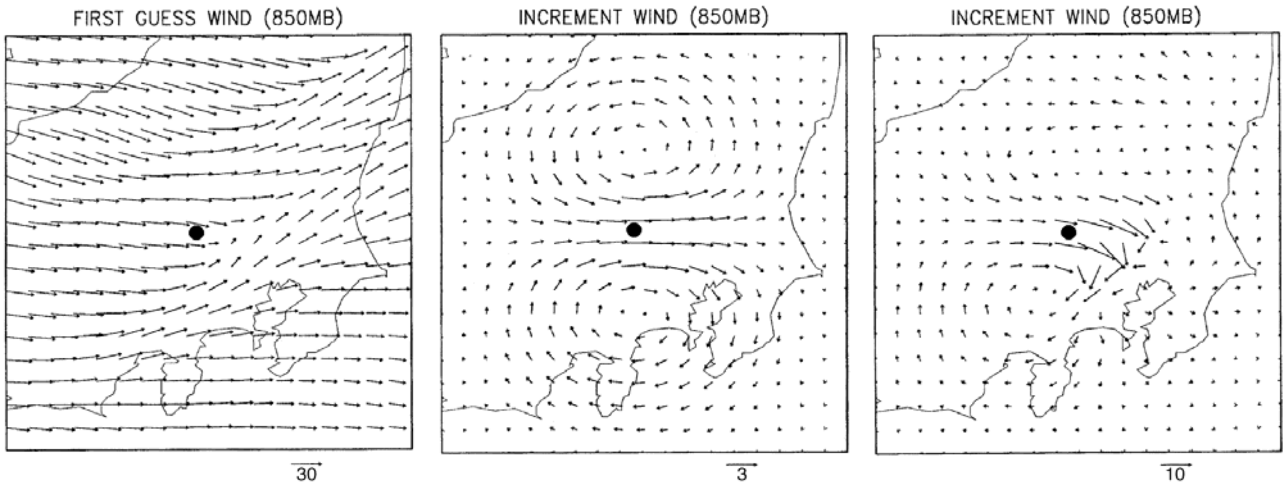

Figure 2 First guess of wind at 850 hPa (left) and its analysis increments for the 3D-Var (middle) and 4D-Var (right) from a single pseudo observation of wind at the dot point at 850 hPa.

3. Cost function

The cost function consists of the background term, the observation terms and the penalty term for suppressing high-frequency gravity wave noise. Observational data assimilated are taken from land surface stations, ship, buoys, radiosondes, pilot balloons, wind profilers, aircrafts, and 1-hour accumulated precipitation from radar and rain gauge observations. Temperature data from radiosonde observations are assimilated instead of geopotential height data. The above observational data are assimilated at 1-hour intervals in a 3-hour window.

The following observation term is used for assimilating the precipitation data that are interpolated to model grid points, taking into account a non-Gaussian character of the probability density function of precipitation data (Koizumi, 2001).

$$J_{\text{precip}}(\mathbf{x}) = \sum_{y_{o\,i} \geq 0.5\text{mm}} \frac{\left(H_i \mathbf{x} - y_{o\,i}\right)^2}{2\sigma_{o\,i}^{\;2}} \tag{7}$$

where H_i denotes the observation operator that converts model state variables $\mathbf{x}$ to 1-hour accumulated precipitation at the i-th grid point, and y_{oi} the observed precipitation at that grid point. The observation error standard deviation σ_{oi} is given by

$$\sigma_{o\,i} = \begin{cases} \sigma_i & (H_i\mathbf{x} \leq y_{o\,i}) \\ 3\sigma_i & (H_i\mathbf{x} > y_{o\,i}) \end{cases}, \qquad \sigma_i = \max\left(1 \ \text{ mm/h}, \ \ y_{o\,i}\right) \tag{8}$$

It is to be noted that observations of 1-hour precipitation less than 0.5 mm are not assimilated, since it was found that the quality of such observations was rather poor in snowfall cases.

4. Operational aspects

For high speed processing, an incremental approach (Courtier et al., 1994) is adopted with an inner-loop model with a horizontal resolution of 20 km. The full-physics nonlinear model and a reduced-physics adjoint model are used for inner-loop forward and inner-loop backward integrations, respectively, to effectively assimilate the precipitation data. As a result, the forward model and the backward model in the inner-loop are not consistent with each other.

The physics contained in the adjoint model are large-scale condensation, middle-level convection of a moist convective adjustment type, simplified vertical diffusion, simplified surface friction and simplified long wave radiation. It is to be noted that most of precipitation in MSM comes from the large-scale condensation although MSM contains a prognostic Arakawa-Shubert scheme to parameterize deep cumulus convection. Therefore, the absence of deep cumulus convection in the adjoint model may not cause a serious problem in the 4D-Var system.

Two consecutive 4D-Var assimilations with 3-hour windows are conducted just before prediction runs which start from the initial conditions at 00, 06, 12 and 18 UTC. For preparing the initial condition at 06 UTC, for example, the two assimilation windows cover 00-03 UTC and 03-06 UTC. The data cutoff times of the first and second assimilations are 3.5 hours and 50 minutes, respectively.

The L-BFGS method (Liu and Nocedal, 1989) is adopted for the minimization in the 4D-Var system. The number of iterations is limited by a wall clock time so that about 20 iterations are possible. Further details of the JMA mesoscale 4D-Var system are described in Ishikawa (2001) and JMA (2002).

5. Results from parallel run tests

Parallel run tests to compare the performance of the 4D-Var and pre-run systems were conducted for two one-month periods, June 2001 and September 2001. Both months are the rainy seasons in Japan. The numbers of samples are 120 for each month. Figure 3 compares the threat scores of 3-hour accumulated precipitation over Japan with a horizontal resolution of 40 km. It is found from the figure that MSM starting from the initial conditions prepared by the 4D-Var system showed significantly better skill of precipitation forecasts than MSM forecasts starting from initial conditions prepared by the pre-run system for both of the threshold values of 1 mm and 10 mm. Since the pre-run system assimilated the observational data during the 3-hour period just before the initial time while the 4D-Var system assimilated the observational data during the 6-hour period, the 4D-Var system had an advantage of using more observational data. However, large impacts similar to those seen in Fig.3 were also obtained from another parallel run test in which the 4D-Var system was run only for the 3-hour period just before the initial time with the first guess taken from the latest forecast from RSM as in the pre-run system. Therefore, the forecast improvement obtained from the parallel run tests may be primarily due to the difference in assimilation methods.

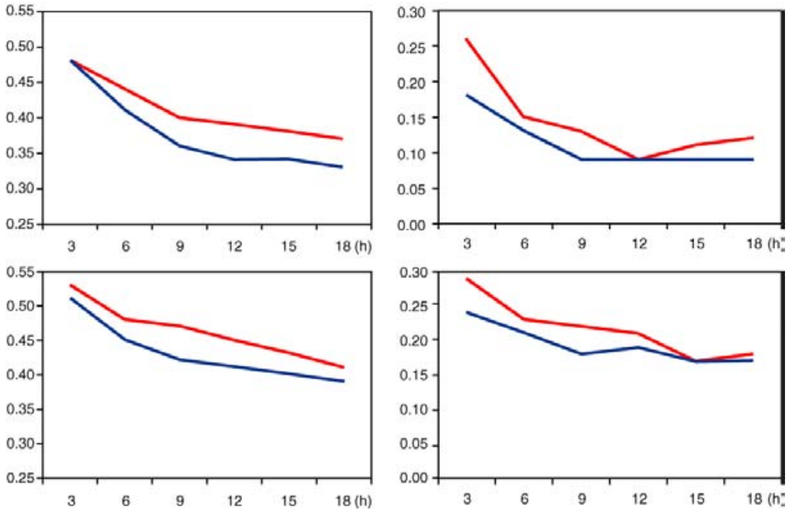

Figure 3 Threat score of 3-hour accumulated precipitation over Japan plotted against forecast time for June 2001 (top) and September 2001 (bottom). Threshold values are 1 mm (left) and 10 mm (right) with a horizontal resolution of 40 km. Solid lines are the 4D-Var and dashed lines the pre-run.

An example of precipitation forecasts are presented in Fig. 4. A band of heavy rain in western Japan was well predicted by MSM with the 4D-Var system, while MSM with the pre-run system failed to predict heavy rain for the first 3-hour period and the area of heavy rain predicted for the second 3-hour period was erroneously shifted southward in spite of the fact that the precipitation data were also assimilated.

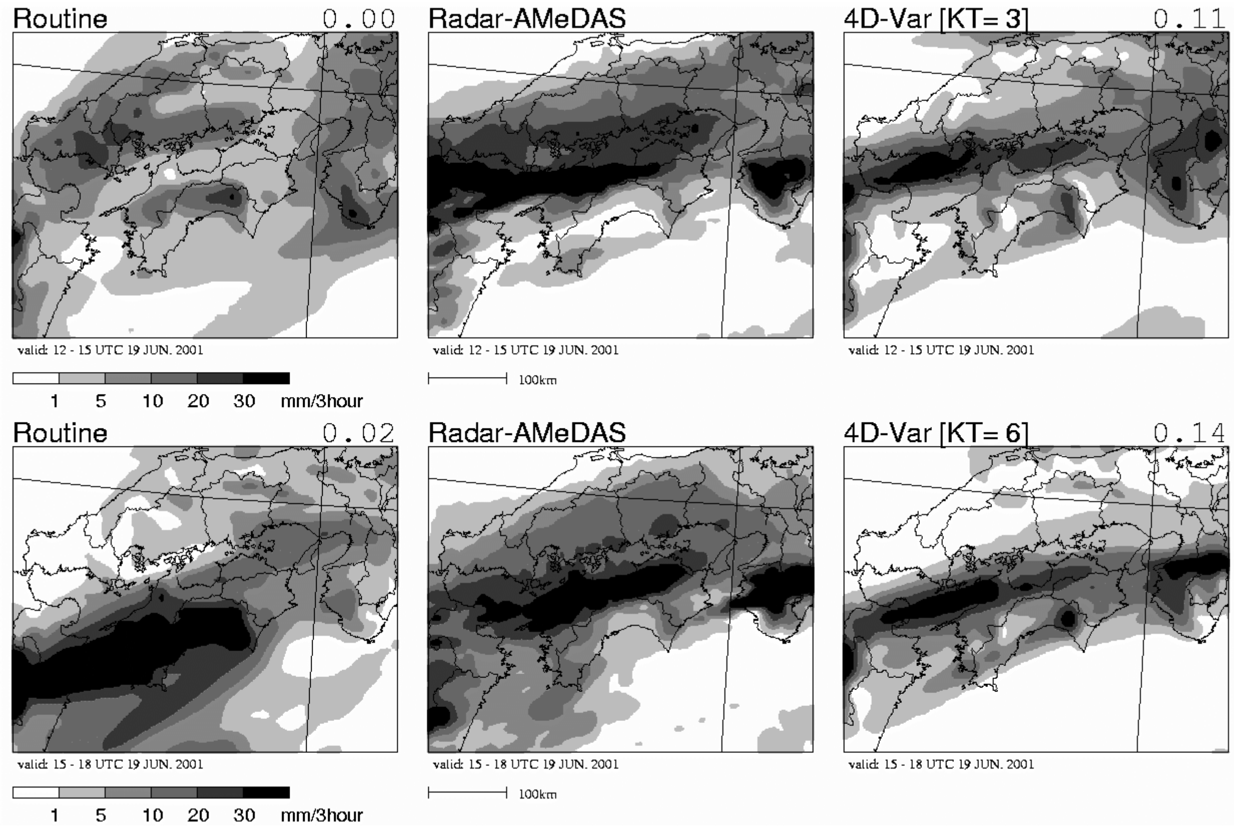

Figure 4 Verification of 3-hour accumulated precipitation forecasts starting from initial conditions at 12 UTC 19 June 2001. Left panel is the pre-run, middle panel the observation, and right panel the 4D-Var. Top panels are the 0-3 hour forecasts and bottom panels the 3-6 hour forecasts.

Figure 5 compares the root mean square error (RMSE) of zonal wind and temperature at 850 hPa verified against radiosonde observations in Japan. The 4D-Var system showed better forecast skill of zonal wind and temperature at that level. Similar improvements in forecast skill were also seen for wind and temperature at other levels. It is to be noted that large improvements in the RMSE of temperature at the initial time were primarily due to the fact that the 4D-Var system assimilated temperature data from radiosonde observations while the pre-run system assimilated geopotential height data.

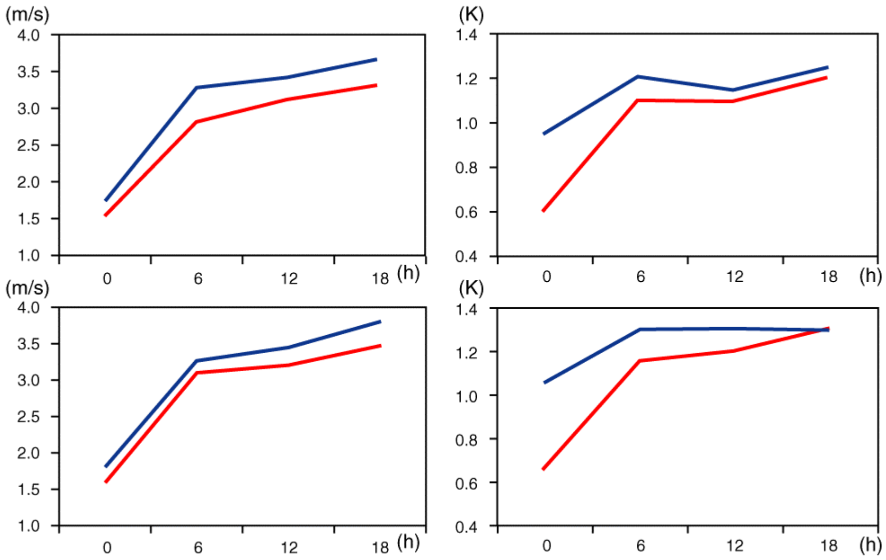

Figure 5 RMSE of zonal wind (left) and temperature (right) at 850 hPa verified against radiosonde observations in Japan plotted against forecast time for June 2001 (top) and September 2001 (bottom). Solid lines are the 4D-Var and dashed lines the pre-run.

Since three typhoons, T0115, T0116 and T0117, approached Japan in September 2001, the skill of track and intensity forecasts for those typhoons was also evaluated. The typhoon bogus data were not used in both of the 4D-Var and pre-run systems, while they were used for preparing the initial condition of RSM that provided MSM and the pre-run system with the lateral boundary condition and the first guess at the beginning of the 3-hour pre-run period. Although significant differences were not seen in typhoon track forecasts, the 4D-Var system showed much improvements in forecasts of the central pressure of T0115 and T0116 (Figure 6). MSM starting from initial conditions prepared by the pre-run system showed erroneously large time tendency of the central pressure of both typhoons. It is to be noted that the horizontal resolution of MSM is still not fine enough to realistically represent the observed central pressure of typhoons. Both systems made similar predictions for T0117, of which minimum central pressure was 975 hPa.

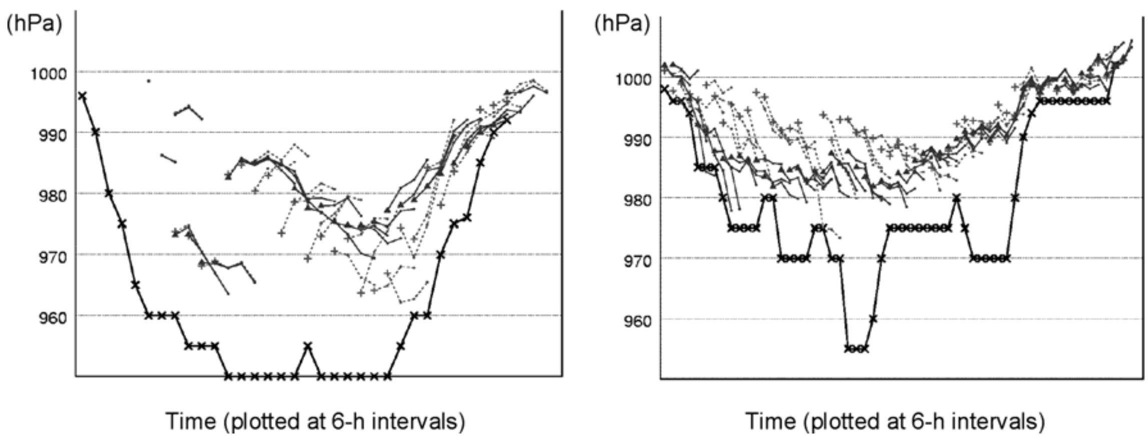

Figure 6 Verification of typhoon central pressure forecasts for T0115 (left) and T0116 (right). Thick solid lines are the observation, thin solid lines the 4D-Var and dashed lines the pre-run. Marks are plotted at 6-hour intervals.

6. Impacts of precipitation data and GPS precipitable water data

An OSE is currently being conducted for June 2001 to evaluate the impact of precipitation data on precipitation forecasts with the 4D-Var system. So far the OSE has been completed for the first two-week period. Figure 7 compares the threat scores and bias scores of 3-hour accumulated precipitation over Japan with a horizontal resolution of 10 km for the first half-month period. It is found from the figure that the precipitation data had a large positive impact on the threat scores for the first 3 hours, but that the scores were slightly degraded in later periods. The bias scores for the first 3 hours were also increased by assimilating the precipitation data. It is regarded as an improvement for heavy precipitation cases, while it is a deterioration of forecasts for weak precipitation cases. This deterioration may be partly due to the fact that observations of no precipitation were not assimilated in the 4D-Var system. The degradation of precipitation forecast skill in later forecast periods and the increase of the bias score for weak precipitation cases are problems with the current 4D-Var system.

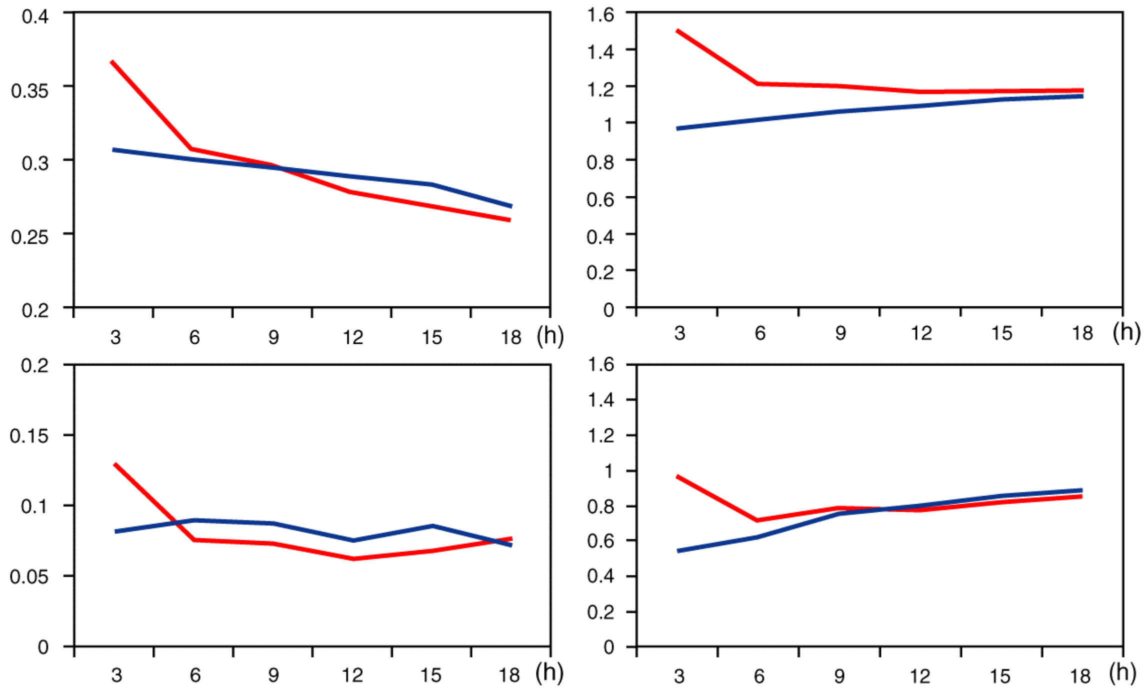

Figure 7 Threat score (left) and bias score (right) of 3-hour accumulated precipitation over Japan plotted against forecast time for the first half period of June 2001. Threshold values are 1 mm (top) and 10 mm (bottom) with a horizontal resolution of 10 km. Thick and thin lines are with and without assimilating the precipitation data, respectively.

Another problem is that the 4D-Var system sometimes failed to assimilate the precipitation data when the first guess was in a dry condition in spite of the fact that the full-physics nonlinear model is used as the inner-loop forward model. This is an inherent problem with 4D-Var assimilation of precipitation data using model physics that contain "on-off" switches. On the other hand, the problem may not be so serious for physical initialization, although there are a couple of shortcomings in physical initialization as was pointed out by Tsuyuki (1996). That problem may be ameliorated by assimilating moisture data from ground-based GPS observations and/or satellite microwave observations.

There are about 1000 GPS ground receivers deployed over Japan to monitor the crust movement. The precipitable water data from GPS observations are currently not assimilated at JMA. A preliminary OSE to evaluate the impact of GPS precipitable water data on precipitation forecasts was conducted for rainy days in June 2001. The target periods were 5-6, 13-14, 18-19 and 22-23 June 2001, so that the total number of cases is 24. The assimilated GPS data were taken from 170 GPS ground receivers over Japan at 1-hour intervals.

Precipitable water data from GPS ground receiver sites where the model topography was higher than the real topography were corrected by using the moisture profile of the first guess.

Results from the OSE showed that the GPS precipitable water data had positive impacts on moisture analysis while benefits from the data were not clear for skill of precipitation forecasts. Figure 8 presents an example of precipitation forecasts that demonstrated an improvement by assimilating the GPS data. The threat scores of 3-hour accumulated precipitation over Japan with a horizontal resolution of 10 km are shown in Fig.9, which suggests that the impact of the GPS data on precipitation forecasts is marginal. However, it is to be noted that the number of samples is not large enough to obtain a conclusive result. The direct assimilation of GPS line of sight data in slantwise directions may be preferable for extracting more information from GPS observations.

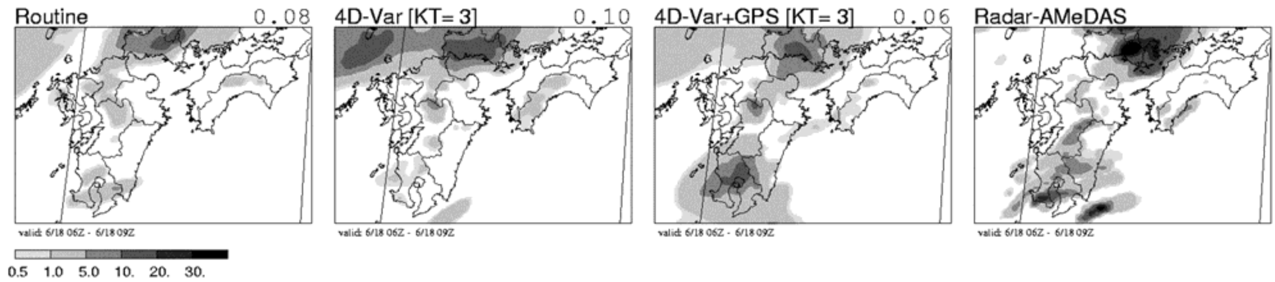

Figure 8 Verification of 3-hour accumulated precipitation forecasts starting from initial conditions at 06 UTC 18 June 2001. The right panel is the observation, and the other panels are, from left to right, the 0-3 hour forecasts from the pre-run without the GPS data, the 4D-Var without the GPS data and the 4D-Var with the GPS data.

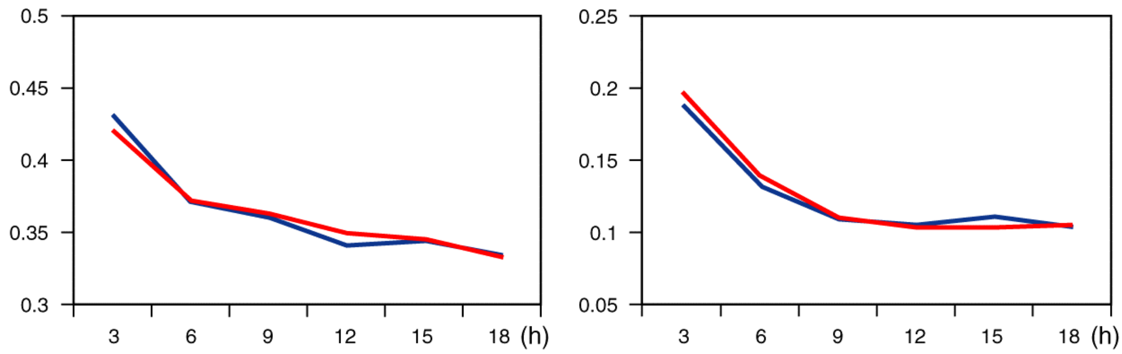

Figure 9 Threat score of 3-hour accumulated precipitation over Japan plotted against forecast time for 8 rainy days of June 2001. Threshold values are 1 mm (left) and 10 mm (right) with a horizontal resolution of 10 km. Thick and thin lines are with and without assimilating the GPS precipitable water data, respectively.

7. Conclusions

The JMA mesoscale 4D-Var system, which became operational in March 2002, significantly outperformed the former operational assimilation system in forecast skill of MSM in spite of rather small numbers of iterations imposed on from the operational requirements. MSM forecasts starting from initial conditions prepared by the 4D-Var system also well predicted the time trend of the central pressure of typhoons. However, positive impacts of assimilating the precipitation data on precipitation forecasts were seen only in the first 3-hour period and there was the degradation of forecast skill in later forecast periods. Furthermore, the precipitation data were sometimes not well assimilated in the 4D-Var system when the first guess was in a dry condition. Results from the preliminary OSE for the GPS precipitable water data showed that the impact of the GPS data on precipitation forecasts was marginal.

The JMA mesoscale 4D-Var system was successfully implemented in March 2002, and the skill for MSM precipitation forecasts was significantly improved after that. Currently, assimilation with the 4D-Var system of radial velocity data from Doppler radar observations and precipitation and moisture data from satellite microwave observations are being developed. Revisions to the functional form of the observation term for precipitation data and the background and observation error statistics are being considered in order to alleviate problems with precipitation data assimilation. JMA is planning to replace MSM by a nonhydrostatic grid model called NHM in a couple years. Although the initial condition of NHM will be at first prepared by using the MSM/4D-Var system described in this paper, it will be replaced by a NHM/4D-Var system in the near future. The development of the NHM/4D-Var system was started in 2002 in collaboration between the Numerical Prediction Division of JMA and the Forecast Research Department of the Meteorological Research Institute.

References

Courtier, P., J.-N. Thepaut and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.

Ishikawa, Y., 2001: Development of a mesoscale 4-dimensional variational data assimilation (4D-Var) system at JMA. *Proc. of the 81st Annual Meeting of AMS: Precipitation Extremes: Prediction, Impacts, and Responses, 14-19 January 2001, Albuquerque, New Mexico*, American Meteorological Society, P2.45.

JMA, 2002: *Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency*, 158pp. Available from the Japan Meteorological Agency.

Koizumi, K., 2001: The four-dimensional data assimilation system for the JMA mesoscale model. *Proc. of the 14th Conference on Numerical Weather Prediction, 30 July-2 August 2001, Fort Lauderdale, Florida*, American Meteorological Society, J96-J98.

Liu, D. C. and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, **45**, 503-528.

Parrish, D. and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3D model. *Mon. Wea. Rev.*, **124**, 2545-2561.