Chapter 2

Data Assimilation Systems

2.1 Summary

Three kinds of major data assimilation systems for the analysis for atmospheric fields are operated at JMA: Global Analysis (GA), Meso-scale Analysis (MA) and Local Analysis (LA). Specifications of the JMA data assimilation systems are summarized in Table 2.1.1. All the analyses are performed by using the procedures shown in Figure 2.1.1.

The following is a brief description of the major components of the analysis systems.

1. Observational data are received from the GTS, Internet and dedicated network. They are decoded according to their code forms. If typhoons exist in the western North Pacific, typhoon bogus profiles are created.

2. Various pre-analysis procedures, such as quality control, data selection and bias correction, are applied to the decoded observational data. In the pre-analysis process, first guess fields retrieved from forecast models are used as a reference of the present atmospheric conditions.

3. The four-dimensional variational method is adopted in the global analysis and the meso-scale analysis. And the three-dimensional variational method is adopted in the local analysis. All the analyses are carried out on the grid of the corresponding forecast models except for the local analysis.

The atmospheric fields analyzed from the data assimilation systems are used as initial conditions of forecast models. First guess field and boundary conditions of data assimilation systems are provided from forecast models as shown in Figure 2.1.2.

Sea surface temperature field (see Section 5.2) and snow depth field are also analyzed every day. For climate monitoring, JMA Climate Data Assimilation System (JCDAS) is operated taking over the data assimilation cycle of the Japanese 25-year Reanalysis (JRA-25) (see Section 2.10).

2.2 Observation Data

2.2.1 Summary of Observation Data Used in the Analysis

A variety of observations has been utilized in the present NWP systems in JMA. Table 2.2.1 summarizes the utilized observation types and the parameters which are inputted into the objective analysis systems, as of 1 January 2013. The additional information for each observation type is described in the following subsection.
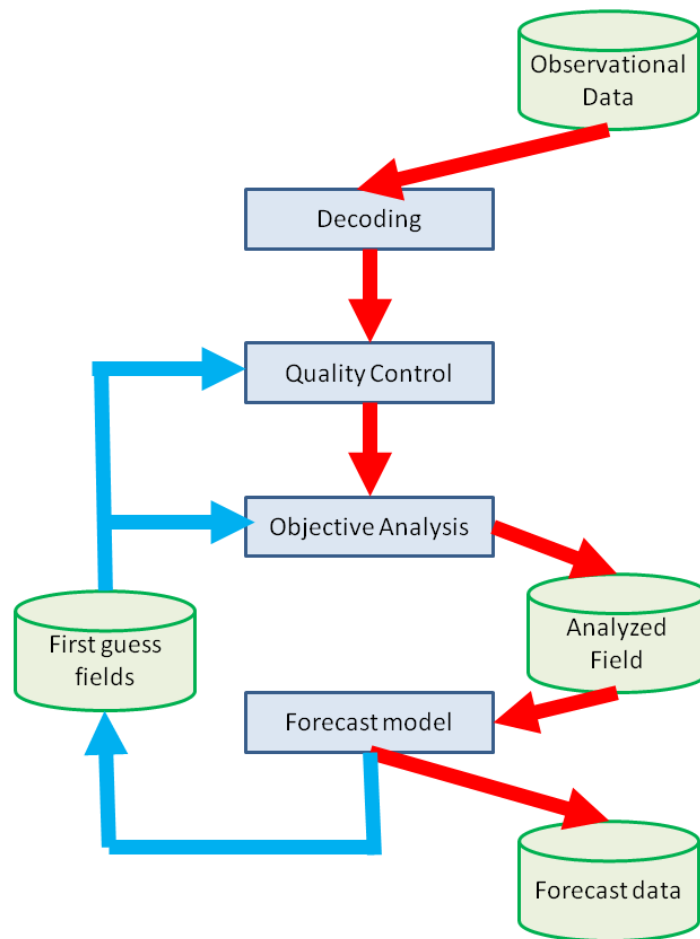
Figure 2.1.1: Major functional components and data flow in the JMA data assimilation system

Table 2.1.1: Specifications of the JMA data assimilation systems

Analysis Model	Global Analysis (GA)	Meso-scale Analysis (MA)	Local Analysis (LA)
Analysis time	00, 06, 12, 18 UTC	00, 03, 06, 09, 12, 15, 18, 21 UTC	00, 03, 06, 09, 12, 15, 18, 21 UTC
Data cut-off time	Early analysis: 2h20m (00,06,12,18UTC) Cycle analysis: 11h50m (00,12UTC) 7h50m (06,18UTC)	50m (00,03,06,09,12,15,18,21 UTC)	30m (00,03,06,09,12,15,18,21 UTC)
Horizontal Grid system	Reduced Gaussian grid	Lambert projection	Lambert projection
Horizontal resolution	TL959 (20km)	5km at 60°N and 30°N	5km at 60°N and 30°N
Number of grid points	$(1920-60) \times 960$	721×577	441×501
Horizontal resolution of inner model	TL319 (55km)	15km at 60°N and 30°N	-
Number of grid points of inner model	$(640-60) \times 320$	241×193	-
Vertical coordinate	σ-p hybrid	z-$z^{\star}$ hybrid	
Vertical levels	Surface +60 levels up to 0.1hPa	Surface +50 levels up to 21.8km	50 levels up to 21.8km
Analysis scheme	4-dimensional variational method		3-dimensional variational method

\# Global snow depth analysis is carried out everyday on $1° \times 1°$ longitude-latitude grids.

\#\# JMA Climate Data Assimilation System (JCDAS) is performed taking over the data assimilation
cycle of the Japanese 25-year Reanalysis (JRA-25)(see Section 2.10).
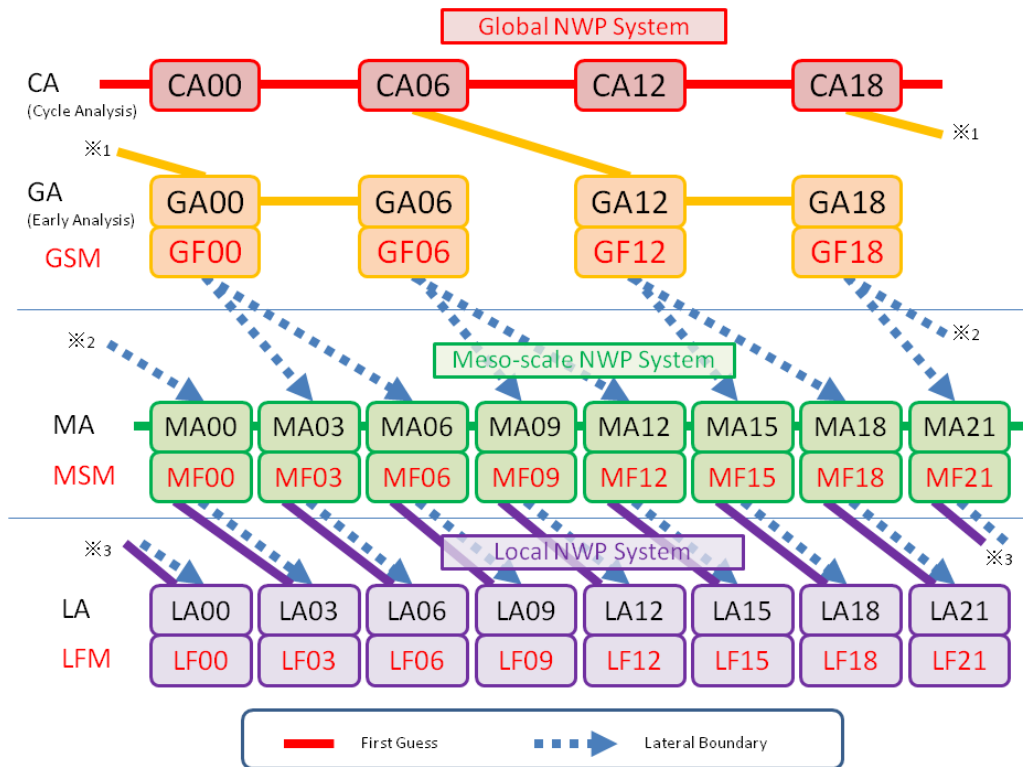

Figure 2.1.2: Major flow of the JMA data assimilation systems

Table 2.2.1: Summary of the observation types and the parameters used in the objective analysis. The meanings of the symbols in the third column are as follows. P: Surface pressure; u: zonal wind; v: meridional wind; T: Temperature; Rh: Relative humidity; T_B: Radiance in brightness temperature; R_1: Precipitation amount; Pwv: Precipitable water vapor; Vr: Radial velocity. The meanings of the symbols in the fourth columns are as follows. GA: Global Analysis; MA: Meso-scale Analysis; LA: Local Analysis.

Observation types (or the code name used for reporting the observation)	Short description	Parameters used in the analysis	The analyses which use the observation
SYNOP	Land surface observations at the world's weather stations	P, u, v, T, Rh	GA, MA, LA
AMeDAS	Land surface automated observation network in Japan	u, v, T	LA
SHIP	Sea surface observations by ships, oil rigs and moored buoys	P, u, v, T, Rh	GA, MA
BUOY	Sea surface observations by drifting buoys	P, u, v, T, Rh	GA, MA
TEMP	Upper-air observations by radiosondes	P, u, v, T, Rh	GA, MA
PILOT	Upper-air wind observations by rawins or pilot balloons	u, v	GA, MA
Aircraft	Upper-air observations by (mainly commercial) aircrafts.	u, v, T	GA, MA, LA
Wind Profiler	Upper-air wind profile observations in Japan, Hong Kong, Europe and the U.S.	u, v	GA, MA, LA
AMV	Atmospheric motion vector (AMV) wind data from geostationary (GEO) satellites and polar orbiting satellites	u, v	GA, MA
Scatterometer	Ocean surface wind vector (OSWV) data from scatterometers on low earth orbit (LEO) satellites	u, v	GA
MW Sounder	Radiance data from microwave (MW) sounders on polar orbiting satellites	T_B	GA, MA
MW Imager	Radiance data from MW imagers on LEO satellites and precipitation amount estimated from the MW imager radiance data	T_B, R_1	$GA (T_B$ only$), MA$
CSR	Clear sky radiance (CSR) data of water vapor channels on GEO satellites' infrared imagers	T_B	GA, MA
GNSS-RO	Refractivity profile data retrieved from radio occultation (RO) measurements of the global navigation satellite systems (GNSS) receivers on LEO satellites	refractivity	GA
GNSS-PWV	Precipitable water vapor data estimated from atmospheric signal delay measurements of the ground-based GNSS receivers	Pwv	MA, LA
Radar Reflectivity	Relative humidity data estimated from 3-dimensional reflectivity data of JMA weather (Doppler) radars.	Rh	MA
Radial Velocity	Radial velocity data from JMA weather Doppler radars (WDRs) and Doppler radars for airport weather (DRAWs).	Vr	MA, LA
R/A	Radar estimated precipitation amount calibrated by AMeDAS raingauge network data	R_1	MA
Typhoon Bogus	see Section 2.4	P, u, v	GA, MA

2.2.2 Supplemental Information for Used Observation

2.2.2.1 SYNOP

SYNOP is a numerical code name used for reporting surface observations at land stations. About 16,000 reports are available within every 6 hours.

2.2.2.2 AMeDAS

AMeDAS (Automated Meteorological Data Acquisition System) is a JMA land surface automated observation network. Near surface temperature and wind observations at about 840 stations are available in Japan with 10 minutes interval. For precipitation, about 1,300 raingauges are available in the AMeDAS network.

2.2.2.3 SHIP

SHIP is a numerical code name used for reporting surface observations at sea stations, such as ships, oil rigs and moored buoys anchored at fixed locations. About 4,500 reports are available within every 6 hours.

2.2.2.4 BUOY

BUOY is a numerical code name used for reporting surface observations by drifting buoys. About 7,500 reports are available within every 6 hours.

2.2.2.5 TEMP

TEMP is a numerical code name used for reporting upper-level pressure, temperature, humidity and wind observations by radiosondes. The upper air observations are usually taken at the same time each day (00 and/or 12 UTC). About 650 reports are available at these times.

2.2.2.6 PILOT

PILOT is a numerical code name used for reporting upper-level wind observations by rawins or pilot balloons. About 300, 200 and 100 reports are available at 00 and 12 UTC, 06 UTC and 18UTC, respectively.

2.2.2.7 Aircraft

Aircraft observations are reported by Aircraft Report (AIREP), Aircraft Meteorological Data Relay (AM-DAR), and Aircraft Communications Addressing and Reporting System (ACARS). Since numerous reports are received from the U.S., the reports are thinned to one-fiftieth over the continental U.S. in the pre-process of the analysis. Even after the pre-process, 30,000-50,000 reports are available within every 6 hours in the world. While vertical profile data can be obtained in the vicinity of airports, only flight level data are available along the airways apart from airports.

2.2.2.8 Wind Profiler

Upper air wind speeds and directions are observed by wind profilers at the ground. 33 wind profilers are operated by JMA and the data are available every 10 minutes. The detailed specifications are found in Ishihara *et al.* (2006). Wind profiler data from the U.S., Europe, and Hong Kong is also available.

2.2.2.9 AMV

Atmospheric motion vector (AMV) is the wind data derived by tracing the movement of individual cloud or water vapor patterns in successive satellite images. The AMVs from 5 geostationary (GEO) satellites (Meteosat-7, -9, GOES-13, -15 and MTSAT-2) and 2 polar orbiting satellites (Terra and Aqua) are used. AMVs from GEO satellites cover $60°N - 60°S$ and those from polar orbiting satellites cover the polar regions (higher latitudes than $60°$).

2.2.2.10 Scatterometer

Ocean surface wind vectors from scatterometers onboard polar orbiting satellites are used. Only ASCAT (advanced scatterometer) onboard Europe's polar orbiting satellite, Metop-A, is being used.

2.2.2.11 MW Sounder

Clear radiances from microwave (MW) sounders are used. The used sounders are AMSU-A (Advance Microwave Sounding Unit - A) onboard NOAA-15, -16, -18, -19, Metop-A and Aqua, and MHS (Microwave Humidity Sounder) onboard NOAA-18,-19 and Metop-A. The AMSU-A is a temperature sounder and the radiance is sensitive to the temperature profiles. MHS is a humidity sounder.

2.2.2.12 MW Imager

The less cloud/rain-affected radiances from MW imagers are used. The used imagers are SSMIS (Special Sensor Microwave Imager Sounder) onboard DMSP-F16, -F17, -F18 and TMI (TRMM (Tropical Rainfall Measuring Mission) Microwave Imager) onboard TRMM. The radiance is sensitive to the water vapor amount at the lower troposphere. Precipitation amounts estimated from the radiances by MSC method (Takeuchi and Kurino 1997) are also used in the Meso-scale Analysis.

2.2.2.13 CSR

Clear sky radiance (CSR) is the product name of the radiances averaged over cloud-free pixels on water vapor channels in GEO satellites' imagers. The CSRs from 5 GEO satellites (Meteosat-7, -9, GOES-13, -15 and MTSAT-2) are used. The CSR is sensitive to the water vapor amount at the upper troposphere.

2.2.2.14 GNSS-RO

GNSS-RO (Global Navigation Satellite Systems - Radio Occultation) is a technique for measuring atmospheric profiles. In this technique, a set of atmospheric time delay data of GNSS radio signals received by a low earth orbit (LEO) satellite is measured during each radio occultation event. Since the delay is a result of the atmospheric radio refraction along the propagation path of the signal, the vertical profiles of refractivity (or bending angle) of the atmosphere at the tangent point can be estimated with the set of delay data. As the refractivity is a function of temperature, humidity and pressure, it can be used for determining the profiles of these properties. The currently used LEO satellites and their GNSS receivers are IGOR (Integrated GPS Occultation Receiver) onboard COSMIC (Constellation Observing System for Meteorology, Ionosphere and Climate) satellites and TerraSAR-X, GRAS (GNSS Receiver for Atmospheric Sounding) onboard Metop-A, BlackJack onboard GRACE-A (Gravity Recovery and Climate Experiment) and CORISS (C/NOFS Occultation Receiver for Ionospheric Sensing and Specification) onboard C/NOFS (Communications/Navigation Outage Forecasting System).

2.2.2.15 GNSS-PWV

GNSS-PWV (GNSS - Precipitable Water Vapor) is a product name of the observations by ground-based GNSS receivers. Atmospheric time delays of GNSS radio signals are observed by the receivers and zenith total delay (ZTD) is estimated by averaging the delays of multiple GNSS satellite signals measured by one receiver. Since the ZTD depends on the amount of water vapor and temperature over the receiver, PWV amount can be estimated by using ZTD and supplemental temperature data. GEONET GNSS receiver data are used for GNSS-PWV analysis in JMA. The GEONET is the ground based GNSS receiver network operated by Geospatial Information Authority in Japan. About 1,200 receives are operated continuously over Japan.

2.2.2.16 Radar Reflectivity

Twenty C-band weather radars are operated by JMA and eighteen of them are equipped with Doppler function. Three-dimensional reflectivity data are obtained every five minutes. Relative humidity profiles are estimated from the reflectivity data and NWP grid point values by the technique based on Bayes' theorem (Caumont *et al.* 2010). The relative humidity data are produced within 200 km radius from each radar site below freezing level.

2.2.2.17 Radial Velocity

Eighteen C-band weather Doppler Radars (WDRs) and 9 Doppler Radars for Airport Weathers (DRAWs) are operated by JMA. Three-dimensional radial velocity data are available every five minutes within 150 km radius for WDRs and every six minutes within 120 km for DRAWs. The range resolution is 0.5 km and the azimuthal resolution is 0.703°.

2.2.2.18 R/A

Radar/Raingauge Analyzed Precipitation (R/A) data are a product name of composite precipitation data produced by JMA. The precipitation data are the accumulated precipitation estimation by weather radars with Z-R relationship ($Z = 200R^{1.6}$) calibrated by AMeDAS raingauge data in real time. The details are found in Subsection 4.4.1.

2.3 Quality Control and Relating Procedures

Quality control (QC) is a series of procedures by which "bad" observations are screened out. The QC is a vital component of the objective analysis system, because observations sometimes include large error and the erroneous data might degrade the quality of atmospheric analysis extremely. Such degradation leads to the worse forecast skill. The QC procedures in the JMA objective analysis systems are described in the following subsections.

2.3.1 SYNOP, AMeDAS, SHIP, BUOY, TEMP, PILOT, Aircraft and Wind Profiler

Direct observations (i.e. SYNOP, AMeDAS, SHIP, BUOY, TEMP, PILOT and aircraft) and wind profiler are the observations measuring prognostic variables in the NWP such as pressure, temperature, wind and humidity. The total QC system for these observations is composed by "Internal QC" and "External QC".

2.3.1.1 Internal QC

Internal QC is the procedures to check and correct observation values using the collocated data in the report and several external lists or tables. The check items are shown as follows.

1. Blacklist check: Blacklist is a list of problematic stations or data prepared in advance with non-real-time QC (see Section 2.9). The observations enlisted in the blacklist are rejected in this step.

2. Climatological check: Climatological reasonability is checked in this step. The criteria are defined based on WMO (1993) in advance.

3. Trajectory check: Consistency of consecutive locations is checked for the reports from moving stations such as SHIP, BUOY and aircrafts. The moving velocity and direction are checked in this step. It is also checked for SHIP and BUOY whether the location is in the ocean or not.

4. Inter-element consistency check: Temporal continuity of consecutive reports from surface stations are checked. Consistencies among observation elements within the report are also checked.

5. Vertical consistency check: The vertical consistency is checked for TEMP and PILOT data. The check items are (1) icing of instruments, (2) temperature lapse rate, (3) hydrostatic relationship, (4) consistency among data at standard pressure levels and those at significant levels and (5) vertical wind shear.

6. Bias correction: A bias correction is applied to the TEMP data which are reported without radiative heating correction or have apparent systematic biases. The bias correction constants are prepared with the one-month statistics in the previous month. Same bias correction approach is applied to the aircraft temperature data for the Global Analysis but not for the meso-scale and Local Analysis.

2.3.1.2 External QC

External QC is the procedures to check the observation values with comparing to the (external) first guess and neighboring observations. The check items are as follows.

1. Gross error check: The departure ($D \equiv O - B$) of the observed value (O) from the first guess (B) is calculated for all the observations. The absolute value of D is compared with the tolerance limits C_P (the criterion for "pass") and C_R (the criterion for "reject"). The datum with $|D| \leq C_P$ passes the QC and the datum with $|D| > C_R$ is rejected. The datum with $C_P < |D| \leq C_R$ is regarded as "suspected" and sent to the following spatial consistency check.

2. Spatial consistency check: The departure D of the suspected observations are compared with the departures interpolated by the optimum interpolation method (D_{OI}) using the neighboring observations. The absolute difference of the D and D_{OI} is compared with the tolerance limit C_S (the criterion for "suspect") for the final judgment and the datum with $|D - D_{OI}| \leq C_S$ is accepted.

    Where, the tolerance limits C_P, C_R, and C_S are variable according to the local atmospheric conditions in the first guess fields. The limits are made small if the time tendency and horizontal gradient are small in the fields, and *vice versa*. The scheme is called "Dynamic QC" (Onogi 1998).

3. Duplication check: Duplication of observation reports is frequently found for the data obtained through different communication lines. The most appropriate report is picked up from the duplicated reports after the above mentioned checks with considering the status.

2.3.2 AMV

The AMVs enlisted in the blacklist (Table 2.3.1) are rejected in the first step. Then the AMVs with the low quality indicator (QI, Holmlund 1998) are also rejected. The QI thresholds are defined for each satellite, domain, vertical level and type of image, respectively. It is followed by a thinning step. The thinning distance is 200 km. The following steps are the climatological check (see Subsection 2.3.1.1) and the external QC (see Subsection 2.3.1.2). The details of the QC for AMV are described in the NWP SAF AMV monitoring page[1].

2.3.3 Scatterometer

Level 2 ocean surface wind products are used in the Global Analysis. Low quality data over land or sea ice are rejected at the first step. The wind data with the speed larger than 15 m/s are also rejected for ASCAT because of the negative bias in the intense wind against the first guess. Then, the most likely wind directions are selected from the inherent ambiguity wind directions in the scatterometer measurements by both NWP nudging technique and median filter technique. The next step is the gross error check (see Subsection 2.3.1.2). In this step, correct wind data are occasionally rejected in and around severe weather systems such as cyclones and fronts where the wind direction and speed vary sharply. To avoid such undesirable rejection, a specialized quality control named "Group-QC" is applied. In the Group QC, spatial consistency among the wind vectors is checked in terms of smooth transition in wind direction and wind speed. The Group-QC-passed data are excluded from the rejection in the gross error check. The details of the QC for scatterometer are described in the NWP SAF scatterometer monitoring page[2].

---

[1] http://research.metoffice.gov.uk/research/interproj/nwpsaf/satwind_report/amvusage/jmamodel.html
[2] http://research.metoffice.gov.uk/research/interproj/nwpsaf/scatter_report/scatusage/jmamodel.html

Table 2.3.1: Summary of the blacklisting areas for AMV. The acronyms in the table are as follows. IR: infrared; WV: water vapor; CSWV: clear sky water vapor; NH: Northern Hemisphere; SH: Southern Hemisphere; Polar AMV: AMV from polar orbiting satellites; GEO AMV: AMV from geostationary satellite

Kind	Blacklisting area
Polar AMV (IR) at NH	above 300 hPa or below 900 hPa
Polar AMV (WV/CSWV) at NH	above 300 hPa or below 550 hPa
Polar AMV (IR/WV) at SH	above 300 hPa or below 550 hPa
Polar AMV (CSWV) at SH	above 350 hPa or below 550 hPa
Polar AMV (All)	poleward of 88°N or 88°S
GEO AMV (All)	above 175 hPa or below 975 hPa
GEO AMV (IR)	above 275 hPa at poleward of 20°N or 20°S
GEO AMV (WV)	above 225 hPa at poleward of 20°N or 20°S

Table 2.3.2: Summary of the used channel sets of microwave sounders under each condition

	AMSU-A	MHS
clear sky ocean	ch. 4–13	ch. 3–5
clear sky land/coast/sea-ice	ch. 6–13	ch. 3–5
cloudy ocean	ch. 7–13	ch. 3–5
rainy ocean	ch. 9–13	n/a

2.3.4 Satellite Radiance

Satellite radiance data are used in the Global and Meso-scale Analysis as a form of brightness temperature. A fast radiative transfer model RTTOV10 (Saunders *et al.* 2012) is employed for the radiance assimilation. The common QC procedures for the radiance data are blacklist check, thinning and external QC. The blacklist is the list for problematic instrument prepared in advance with non-real-time QC (see Section 2.9). The data enlisted in the blacklist are rejected in the first step. In the next step, the data are thinned spatially at each time slot of assimilation window (approximately one hour) to reduce the computational costs. The following external QC includes reduction of instrumental scan biases (except for CSR), cloud/rain contamination check, location check, channel selection and gross error check (see Subsection 2.3.1.2). The QC passed data are thinned again for the reduction of observation error correlation. The thinned data are outputted to be used in the data assimilation systems. In the Global Analysis, variational bias correction (VarBC, Derber and Wu 1998; Dee 2004) is used for the reduction of air-mass dependent biases. VarBC is an adaptive bias correction scheme where a linear regression formula to represent biases is embedded in the observation operator and the regression coefficients are set as analysis variables. The formulations are described in Subsection 2.5.7.4. In the Meso-scale Analysis, the air-mass dependent biases are removed in the pre-process by using the VarBC coefficients obtained in the latest Global Analysis. The used satellite radiance data are the data from MW sounder, MW imager and CSR. The specific procedures for each data are described in the following subsections.

2.3.4.1 MW sounder

The sets of used channels are defined according to each surface and atmospheric condition in advance. The sets are summarized in Table 2.3.2.

2.3.4.2 MW imager

The less cloud/rain-affected radiances of vertically polarized channels are assimilated over the ice-free ocean. In the Meso-scale Analysis, precipitation retrievals are also assimilated over the ocean surrounding Japan. The precipitation amount estimations are resampled onto the grid of inner model with spatial smoothing.

2.3.4.3 CSR

The CSR are thinned to every 2.0 degrees horizontally and every 2 hours temporally. The CSR having a low percentage of clear pixels and a large standard deviation of brightness temperature are excluded because these data have low representativeness of the area. For Meteosat-7, data in nearly local midnight are also excluded to avoid solar stray light contamination (Munro *et al.* 2004).

2.3.5 GNSS-RO

Refractivity data at the altitudes up to 30 km are used in the Global Analysis with the 500 m vertical intervals. The observation errors are defined as a function of height.

2.3.6 GNSS-PWV

The PWV data are used in the meso-scale and Local Analysis. Since there are steep mountains in Japan, large differences are found between the actual ground surface elevation and the model surface elevation especially in mountain area. In the Meso-scale Analysis, the stations with 500 m or higher elevation from the mean sea level are not used. The stations from which the absolute difference of the elevation to the model surface is larger than 200 m are also not used. The GNSS-PWV with the value smaller than 1 mm or larger than 90 mm is rejected in a climatorogical check. Then, the first guess PWV is interpolated or extrapolated to the actual terrain surface and compared to the GNSS-PWV. The data with the absolute difference from the first guess are larger than 8 mm are rejected in a gross error check. Since there are dense GNSS-PWV network for the analysis systems, the data are thinned by 30 km for the Meso-scale Analysis and 15 km for the Local Analysis.

2.3.7 Radar Reflectivity

To assimilate the radar reflectivity data in the Meso-scale Analysis, an indirect assimilation technique which is called 1D+4DVAR (Ikuta and Honda 2011) is employed. The technique is based on Caumont *et al.* (2010). In the 1D+4DVAR, radar reflectivity data are used for retrieving relative humidity (RH), and the RH retrievals are assimilated as conventional observation data by 4D-Var. In this system, only the RH retrievals below the melting layer are used because it is known the reflectivity can be only inappropriately simulated in ice phase with the operational MSM hydrometeors forecast and it causes large biases for the RH retrievals. In addition, the data around the height of 2000 m above sea level are also not used since the data used for making R/A, which are already assimilated in the Meso-scale Analysis in another form (surface rainfall, see Subsection 2.3.9). For the operation, reflectivity data from the JMA C-band radar network are used.

2.3.8 Radial Velocity

The hourly radial velocity data from the WDRs and the DRAWs are used in the Meso-scale Analysis. In the pre-process, the data are resampled into the 5 km range resolution and the 5.625° azimuthal resolution. The resampled data are checked with respect to the sampling data number, radial velocity variance and difference of the maximum and minimum velocity. High elevation angle data ($\geq 5.9°$) and the data close to the radar site (< 10km) are not used. The reason for the former procedure is to avoid the contamination of precipitation fall velocity, and the latter is to avoid the back scattering noise. The data with the wind speed less than 10 m/s is also not used to avoid ground clutter contamination.

2.3.9 R/A

Hourly R/A data are assimilated in the Meso-scale Analysis. Since the R/A data are the quality controlled product, the 1 km grid R/A data are simply resampled into the inner-model grid box (15 km) and inputted into the Meso-scale Analysis.

2.3.10 CDA: Feedback Data Base

All information concerning the quality of observational data obtained during the quality control procedure are archived in the Comprehensive Database for Assimilation (CDA). CDA is extensively used for both real-time and non real-time data monitoring activities. All information contained in CDA is managed by the form of integer 2 byte. The format of CDA is quite simple and is designed for flexible use so that any information concerning observation can be archived easily. CDA is so user-friendly that any information can be extracted easily. The CDA file size tends to become large but it can be remarkably compressed using utilities in UNIX.

2.4 Typhoon Bogussing

For tropical cyclones (TCs) over the western North Pacific, typhoon bogus data as a form of pseudo-observation data are generated and assimilated for a realistic TC structure analysis according to the model resolutions. They are made up of pressures at the mean sea level (P_{msl}) and vertical profiles of the wind (W_{prf}) around TC. The wind profiles are placed at 1000 hPa, 925 hPa, 850 hPa, 800 hPa, 700 hPa, 600 hPa, 500 hPa, 400 hPa and 300 hPa. The generated bogus has axially asymmetric structure in all the analyses.

Firstly, symmetric bogus profiles are generated automatically from the central pressure and the 15m/s wind speed radius of TC (R_{15}) analyzed by forecasters. The surface pressure profile is defined using Fujita's formula (Fujita 1952). The gradient wind balance is assumed to calculate the surface pressure profile meeting the requirement from the wind speed at the particular radius R_{15}. Upper geopotential profiles are defined by the empirical formula based on the TC analysis described in Frank (1977). It is assumed that the temperature anomaly has its maximum at 250 hPa. The wind field on each level is derived from the geopotential height profiles with the gradient wind balance. The surface wind field is also derived from the gradient wind balance but it is modified to include the effect of surface friction.

Secondly, asymmetric components are retrieved from the first guess fields and added to the symmetric bogus profile to generate the final asymmetric bogus structure. When the target area of bogussing is across the lateral boundary in the Meso-scale Analysis, asymmetric components are not added.

Finally, pseudo-observation data are generated from the resulting bogus structure at the analyzed TC center (P_{msl}), TC center on the first guess (P_{msl}), and several points surrounding the analyzed TC center (P_{msl} and W_{prf}). The configuration for the surrounding point distribution is adaptive to the typhoon track error on the first guess.

2.5 Global Analysis

2.5.1 Introduction

The 4-dimensional variational data assimilation (4D-Var) system for the JMA Global Spectral Model (GSM) has been in operation since February 2005 in place of the 3-dimensional variational data assimilation (3D-Var) system. The JMA global 4D-Var system is used for the global cycle analysis (00,06,12,18UTC) and the global early analysis for the GSM forecast (00,06,12,18UTC). The scheme has the following benefits over the 3D-Var scheme.

- The dynamics and physics of the forecast model are considered in assimilating data. As a result, observational data are optimally used in a meteorologically consistent way, so that the analysis increments become flow-dependent (Figure 2.5.1).

- The observations are assimilated at appropriate observation time.

- It can directly assimilate all observations data including precipitation amount that can be derived from model variables, although precipitation amount is not assimilated in the JMA global 4D-Var system.
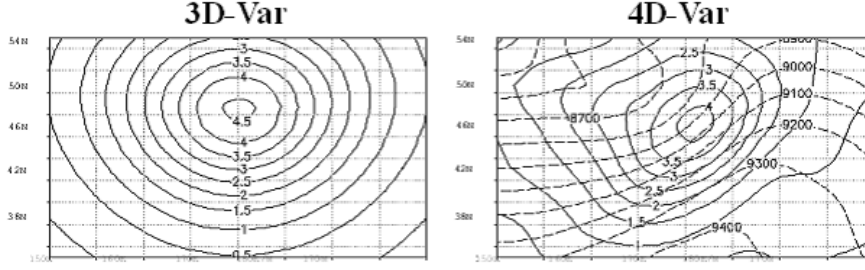
Figure 2.5.1: Analysis increments (solid line) by 3D-Var (left) and 4D-Var (right) when one pseudo observation height data (departure 5m) is assimilated (the 20th model level). The broken line indicates the first guess height field. 4D-Var analysis increments are flow-dependent in accordance with the first guess field.

2.5.2 Description of the Algorithm

The 4D-Var uses 3-9 hour forecast from GSM (TL959L60) as a first guess (background). All data within 3 hours from analysis time are assimilated at appropriate observation time with hourly assimilation slots. The cost-function measures the distance between the model trajectory and the observations over a 6-hour assimilation window.

An incremental method (Courtier *et al.* 1994) is adopted in the 4D-Var to save computer resources. In this method, analysis increment is computed at lower resolution (inner loop:TL319L60) and then is added to the high-resolution first guess (outer loop:TL959L60). The resolution of inner loop was upgraded from T159L60 to TL319L60 in October 2011 (Kadowaki and Yoshimoto 2012).

To obtain the analysis increment Δx_i, the minimization of the cost function J defined by Eq. (2.5.1) is performed in the inner loop.

$$J\left(\Delta x_0\right) = \frac{1}{2}\Delta x_0^{\mathrm{T}}\mathbf{B}^{-1}\Delta x_0 + \frac{1}{2}\sum_{i=0}^{n}\left(\mathbf{H}_i\Delta x_i - d_i\right)^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i - d_i\right) + J_C \qquad (2.5.1)$$

$$\Delta x_{i+1} = \mathbf{M}_i\Delta x_i = \mathbf{M}_i\mathbf{M}_{i-1}\mathbf{M}_{i-2}\ldots\mathbf{M}_0\mathbf{N}\Delta x_0 \qquad (2.5.2)$$

where subscript i indicates the time and n denotes the end of the assimilation window. Δx_0 is the low resolution increment at the initial time before the initialization, and Δx_i is the increment evolved according to the tangent linear model from the initial time to time i and $\mathbf{R}_i$ denotes the covariance matrix of observation errors at time i and $\mathbf{B}$ is the covariance matrix of background errors, which are described in detail in Subsection 2.5.6 and Subsection 2.5.7. $\mathbf{M}_i$ is the tangent linear (TL) model of the low resolution nonlinear (NL) forecast model M_i described in detail in Subsection 2.5.4. $\mathbf{N}$ is a nonlinear normal-mode initialization operator (Machenhauer 1977). $\mathbf{H}_i$ is the TL operator of the observation operator H_i. The innovation vector is given at each assimilation slot by $d_i = y_i^0 - H_i x_i^b$, where x_i^b is the background state evolved by the high resolution NL model, and y_i^0 is the observation data at time i. J_C is the penalty term to suppress the gravity wave described in Subsection 2.5.5.

To minimize the cost function J, the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Liu and Nocedal 1989) with Veersé's preconditioner (Veersé *et al.* 2000) is applied. Here, the gradient of the cost function ∇J is required. It is obtained from the following adjoint procedures Eq. (2.5.3)-Eq. (2.5.6), which is computed reverse in time.

$$p_{n+1} = 0 \qquad (2.5.3)$$

$$p_i = \mathbf{M}_i^{\mathrm{T}}p_{i+1} + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i - d_i\right) \qquad (i = n, \ldots, 1) \qquad (2.5.4)$$

$$p_0 = \mathbf{M}_0^{\mathrm{T}}p_1 + \mathbf{B}^{-1}\left(\Delta x_0\right) + \mathbf{H}_0^{\mathrm{T}}\mathbf{R}_0^{-1}\left(\mathbf{H}_i\Delta x_i - d_i\right) \qquad (2.5.5)$$

$$\nabla J\left(\Delta x_0\right) = p_0 \qquad\qquad (2.5.6)$$

where p_i is a dummy variable, $\mathbf{M}_i^{\mathrm{T}}$ is the adjoint (AD) model of the TL model $\mathbf{M}_i$, and $\mathbf{H}_i^{\mathrm{T}}$ is the AD operator of $\mathbf{H}_i$. Note that Eq. (2.5.3)-Eq. (2.5.6) should contain additional terms of the penalty term and the initialization in Eq. (2.5.1), which are neglected here for the simplicity.

The analyzed variables are the relative vorticity, divergence, temperature, surface pressure and the logarithm of specific humidity in the spectral space on the model layers (eta-coordinate). Observational data y_i^0 are wind vector, temperature, relative humidity, satellite radiances, etc.

The low resolution increment Δx_i obtained from the minimization of the cost function in the inner loop is interpolated to the high resolution analysis increment. By adding the increment to the first guess field, high resolution analysis field is derived.

2.5.3   Description of the Procedure

The JMA global 4D-Var system are performed 4 times a day (00,06,12,18UTC). The observations within 3 hours from each analysis time (within assimilation window) are assimilated. The flow of 4D-Var is shown in Figure 2.5.2 for the case of 12UTC analysis time. It is the same for the cycle and early analyses.
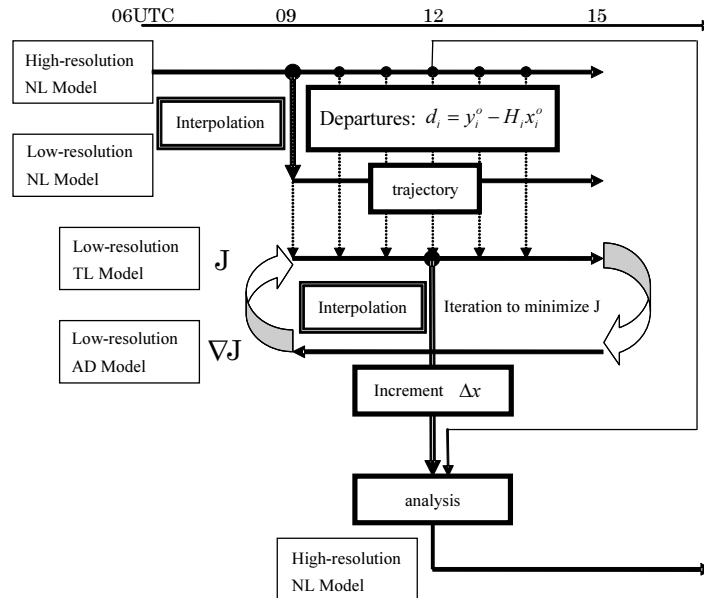

Figure 2.5.2: Flow of 4D-Var procedure for the case of 12UTC analysis time

The procedure is as follows:

1. 9-hour forecast (09UTC-15UTC) of the high resolution outer NL model (same as GSM with the resolution of TL959L60) from previous analysis is used as a first guess (background). The departures between the model trajectory and observations $d_i = y_i^0 - H_i x_i^b$ over a 6-hour assimilation window (09UTC-15UTC) are measured. Observations are organized in six time-slots. The time intervals for the first and last slots are 0.5 hour and 1.5 hour and the others are 1 hour (Figure 2.5.3). All observations in each time slot are regarded as observed in each representative time.

2. The 3 hour forecast field (valid at 09UTC) of the first guess is interpolated into the field with the resolution of the inner model (TL319L60). The interpolation is performed not only horizontally but also vertically to consider the difference of the topography between TL959 and TL319.

3. The inner NL model is performed from the interpolated field to calculate the background state in the low resolution model space.

4. The TL model and AD model are performed to calculate the cost function J and its gradient ∇J with the innovation vector $d_i = y_i^0 - H_i x_i^b$. These processes are iterated to minimize the cost function J. The iteration is performed up to about 70 times. The background trajectory is not updated in our system.

5. After the minimization of J, the field of 3 hour forecast (valid at 12UTC) of the TL model is chosen to be the analysis increment. It is interpolated horizontally and vertically into the field with the resolution of the first guess field (TL959L60). Finally the analysis increment is added to the first guess field (valid at 12UTC) to obtain the final product.
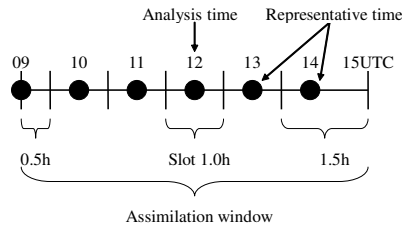

Figure 2.5.3: Schematic diagram of time slots for the analysis time 12UTC. The black circles indicate the representative time of each time slot.

2.5.4 Inner Model

The inner NL model is basically based on the JMA GSM, but moisture processes (convection scheme and cloud scheme) are replaced with those of the older GSM (GSM0103; JMA 2002), mainly for the stability of the inner TL model integration. In addition, the nonlinear normal-mode initialization (Machenhauer 1977) is added.

The inner TL model includes the following simple processes and most of those are based on the inner NL model.

1. **Initialization:** To control the gravity wave, nonlinear normal-mode initialization is adopted.

2. **Horizontal Diffusion:** Horizontal diffusion is enhanced over that of the inner NL model, according to Buizza (1998).

3. **Surface Turbulent Fluxes:** The surface turbulent fluxes are formulated as the Monin-Obukhov bulk formulae based on the inner NL model. The sensible and latent heat flux are perturbed only over the sea.

4. **Vertical Turbulent Diffusion:** The vertical turbulent diffusion of momentum, heat and moisture is formulated as the level 2 turbulence closure scheme of Mellor and Yamada (1974) based on the inner NL model. The diffusion coefficients are not perturbed.

5. **Gravity Wave Drag:** The parameterization for the orographic gravity wave drag consists of two components: one for long waves (wavelength > 100km) and the other for short waves (wavelength $\approx$ 10km) based on the inner NL model. The Richardson number is not perturbed in some part for the long waves for the stability of the inner TL model integration.

22

6. **Long-wave Radiation:** Two kinds of the long-wave radiation are included in the TL model. One is based on Mahfouf (1999). The tendency of the perturbed temperature T' is given by

$$\frac{\partial T'}{\partial t} = -\alpha \frac{g}{C_p} \frac{\partial}{\partial p}\left(4FT'/T\right) \tag{2.5.7}$$

where $\alpha = 1/\{1 + (P_r/P)^{10}\}$, $p_r = 300$hPa and F is the net radiation fluxes calculated in the inner NL model, g and C_p denote the gravitational constant and the isobaric specific heat respectively. The other is related to the difference between the surface ground temperature T_G and the bottom level temperature T_1. The tendency of the temperature T for the nonlinear model is:

$$\frac{\partial T}{\partial t} = \text{const} + \beta\left(T_G - T_1\right) \tag{2.5.8}$$

where β is a coefficient given by a function of the position. The tendency of the perturbed temperature T' is given by:

$$\frac{\partial T'}{\partial t} = -\beta T_1' \tag{2.5.9}$$

7. **Clouds and Large-scale Precipitation:** Clouds and large-scale precipitation are based on the inner NL model. In the inner NL model, clouds are prognostically determined in a similar fashion to that of Smith (1990). A simple statistical approach proposed by Sommeria and Deardorff (1977) is employed to compute the cloud amount and the cloud water content. The parameterization of the conversion rate from cloud ice to precipitation follows the scheme proposed by Sundqvist (1978). They are much simplified in the TL model. The cloud fraction, the amount of dropping cloud ice, and the dependence on the water vapor of the isobaric specific heat are not perturbed. Only some variables are perturbed in computing the conversion from cloud water to precipitation and in computing the evaporation of the precipitation.

8. **Cumulus Convection:** Cumulus convection is formulated as prognostic Arakawa - Schubert scheme (Arakawa and Schubert 1974) based on the inner NL model, but much simplified. The vertical wind shear and the planetary mixing length are not perturbed. The magnitude of perturbation of mass-flux is set bound for the stability of the inner TL model integration and it causes that the inner "TL" model is not exactly linear.

2.5.5 Penalty Term

The penalty term, which is the third term of Eq. (2.5.1), is given by

$$J_C = \alpha \left(\left|N_G \Delta x_0\right|^2 + \sum_{i=2}^{\text{maxslot}} \left|N_G \Delta x_i\right|^2\right) \tag{2.5.10}$$

where N_G denotes an operator to calculate the tendency of the gravity wave mode based on Machenhauer (1977). Δx_0 is the increment at the initial time before the initialization, and Δx_i is the increment evolved according to the tangent linear model from the initial time to the representative time of the i-th time slot after the initialization and the summation is from the second time slot ($i = 2$) to the last time slot ($i = \text{maxslot}$). α is a constant $3.0 \times 10^{-2} [\text{s}^4/\text{m}^2]$, determined empirically. Though this penalty term is introduced to suppress the gravity wave in the increment Δx_i, it is also effective to stabilize the calculation.

2.5.6 Background Term

The background term, which is the first term of Eq. (2.5.1), dominates how the 4D-Var analysis procedure converts the difference between the observation data and first guess into corrections to the first guess. The multivariate couplings in the analysis variables are based on the geostrophic linear balance between mass

and wind. To reduce the correlations among the analysis variables, control variables are introduced. In the algorithm some additional statistical relations are also considered such as the less geostrophic balance in the smaller horizontal and vertical scales, virtually no geostrophic balance near the equator, the dependency of the geostrophy on the vertical level, a weak coupling between divergence and vorticity, as well as between divergence and mass.

The control variables in the 4D-Var are the relative vorticity ζ, unbalanced divergence D_U, unbalanced temperature and surface pressure $(T, P_s)_U$ and the logarithm of specific humidity $\ln q$ in the spectral space on the model layers. Autocovariances of the control variables are assumed to be homogeneous and isotropic. The correlation structures do not depend on the geographical location, but vertical correlations depend on horizontal scale. The unbalanced variables D_U and $(T, P_s)_U$ are defined as

$$\Delta D_U \equiv \Delta D - P \Delta \phi_B \qquad (\phi_B = \phi_B(\zeta)) \tag{2.5.11}$$

$$\begin{pmatrix} \Delta T \\ \Delta p_s \end{pmatrix}_U \equiv \begin{pmatrix} \Delta T \\ \Delta p_s \end{pmatrix} - Q \Delta \phi_B - R \Delta D_U \tag{2.5.12}$$

where P, Q, R are regression coefficients, ϕ_B is a modified balance mass variable derived from relative vorticity described as follows. Δ denotes the deviation from the first guess. This formulation is similar to that used in ECMWF before (Derber and Bouttier 1999), they call the regression coefficients as the balance operator. The regression coefficients are computed statistically using the NMC method (Parrish and Derber 1992) with 24/48-hour forecast differences to estimate the total covariances for each total spectral coefficient.

In the following subsections, modified balance mass variable is described in Subsection 2.5.6.1, the regression coefficients are described in Subsection 2.5.6.2, the covariance matrix of background errors are described in Subsection 2.5.6.3 and Subsection 2.5.6.4, and conversions from control variables to analysis variables are described in Subsection 2.5.6.5.

2.5.6.1 Modified Balance Mass Variable

The geostrophic balance is well kept at midlevels in the troposphere in extratropics. In other areas the balance is weak. To consider these relationships a modified balance mass variable is introduced. The statistical relationships among relative vorticity, divergence and temperature and surface pressure are calculated. First, the singular value decomposition of the linear balance operator L[3] is conducted.

$$\Delta \tilde{\phi}_B = L \Delta \zeta = U W V^{\mathrm{T}} \Delta \zeta \tag{2.5.13}$$

where $\tilde{\phi}_B$ is the original balance mass variable, W is a positive semi-definite diagonal matrix, U and V are orthogonal matrices. The decomposed modes depend on latitude: a singular mode with a small singular value has large amplitude in low latitude. Second, the regression coefficients between mass variables[4], derived from temperature and surface pressure, and balance mass variables are calculated as follows:

$$D_n = \frac{\left\langle \left(U^{\mathrm{T}} \Delta \Phi \right)_n^m \left(U^{\mathrm{T}} \Delta \tilde{\phi}_B \right)_n^m \right\rangle}{\left\langle \left[\left(U^{\mathrm{T}} \Delta \tilde{\phi}_B \right)_n^m \right]^2 \right\rangle} \tag{2.5.14}$$

where $\langle \ \rangle$, D_n and n denote statistical, zonal-wavenumber and vertical-level mean, a positive definite diagonal matrix and index of singular vectors in latitudinal wave number respectively. The regression coefficients D_n $(0 - 1)$ indicates how much the geostrophic balance is satisfied. Then the modified balance mass variables are constructed as follows;

---

[3]Each wave number components of L is denoted as

$$\delta \tilde{\phi}_{Bn}^{\,m} = c_n^m \delta \zeta_{n-1}^m + c_{n+1}^m \delta \zeta_{n+1}^m \quad ((n, m) \neq (0, 0), \ n = m, m+1, \dots, N) , \qquad c_n^m = -\frac{2\Omega a^2}{n^2} \sqrt{\frac{n^2 - m^2}{4n^2 - 1}} , \qquad \delta \tilde{\phi}_{B0}^{\,0} = 0$$

where Ω is angular velocity of the Earth, a Earth radius, n total wavenumber, m zonal wavenumber.

[4]The mass variable Φ_k on the k-th model level is defined by $\Phi_k = \phi_k + R_d \bar{T}_k \ln p_k$.

where Φ_k is the geopotential height, $\bar{T}_k$ is the reference (global mean) temperature, and p_k is the pressure on the k-th level, and R_d is the dry gas constant.

$$\Delta\phi_B = UDU^{\mathrm{T}}\Delta\tilde{\phi}_B = UDWV^{\mathrm{T}}\Delta\zeta = \tilde{L}\Delta\zeta \tag{2.5.15}$$

Note that the modified balance operator $\tilde{L}$ consists of 1) the conversion from the spectral space to the singular vector space, 2) the product of the regression coefficients D, and 3) the conversion from the singular vector space to the spectral space. The correlation between the modified mass variables and unbalanced mass variables (i.e. original mass variables – modified balance mass variables) could be neglected in all regions including the tropics.

2.5.6.2 Regression Coefficients

The regression coefficient matrices P, Q, and R are calculated for each total wavenumber n as follows:

$$P_n = \left\langle \Delta D_n^m \left(\Delta \tilde{\phi}_{B_n}^m \right)^{\mathrm{T}} \right\rangle \left\langle \Delta \tilde{\phi}_{B_n}^m \left(\Delta \tilde{\phi}_{B_n}^m \right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.16}$$

$$Q_n = \left\langle \begin{pmatrix} \Delta T_n^m \\ \Delta (p_s)_n^m \end{pmatrix} \left(\Delta \tilde{\phi}_{B_n}^m \right)^{\mathrm{T}} \right\rangle \left\langle \Delta \tilde{\phi}_{B_n}^m \left(\Delta \tilde{\phi}_{B_n}^m \right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.17}$$

$$R_n = \left\langle \left[\begin{pmatrix} \Delta T_n^m \\ \Delta (p_s)_n^m \end{pmatrix} - Q_n \Delta \tilde{\phi}_{B_n}^m \right] ((\Delta D_U)_n^m)^{\mathrm{T}} \right\rangle \left\langle ((\Delta D_U)_n^m) ((\Delta D_U)_n^m)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.18}$$

where $\langle\ \rangle$ denotes statistical and zonal-wavenumber mean.

2.5.6.3 Background Error Covariance Matrix

The background error covariance matrices of the control variables are calculated for each total wavenumber n and the matrix size is equivalent to the number of vertical levels for ζ, D_U and $\ln q$ or the number of vertical levels $+1$ for $(T, P_s)_U$.

$$B_{\zeta n} = \left\langle \Delta\zeta_n^m \overline{\Delta\zeta_n^m}^{\mathrm{T}} \right\rangle, \qquad B_{D_U n} = \left\langle (\Delta D_U)_n^m \overline{(\Delta D_U)_n^m}^{\mathrm{T}} \right\rangle \tag{2.5.19}$$

$$B_{\begin{pmatrix} T \\ p_s \end{pmatrix}_U n} = \left\langle \Delta \begin{pmatrix} T \\ p_s \end{pmatrix}_{Un}^m \overline{\Delta \begin{pmatrix} T \\ p_s \end{pmatrix}_{Un}^m}^{\mathrm{T}} \right\rangle, \qquad B_{\ln q n} = \left\langle \Delta \ln q_n^m \overline{\Delta \ln q_n^m}^{\mathrm{T}} \right\rangle \tag{2.5.20}$$

where $\langle\ \rangle$ denotes statistical and zonal-wavenumber mean. Total variances of the control variables are rescaled by a factor of 0.81.

2.5.6.4 Cholesky Decomposition of Background Error Covariance Matrix

The background error covariance matrix mentioned above is decomposed by the Cholesky decomposition. It gives the independent and normalized (i.e. preconditioned) control variables Δy_n^m as follows:

$$J_n^{(x)} = \sum_{m=-n}^{n} \frac{1}{2} (\Delta \bar{x}_n^m)^{\mathrm{T}} B_n^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} (\Delta \bar{x}_n^m)^{\mathrm{T}} \left(L_n L_n^{\mathrm{T}} \right)^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} (\Delta \bar{y}_n^m)^{\mathrm{T}} \Delta y_n^m \tag{2.5.21}$$

$$\Delta y_n^m \equiv L_n^{-1} \Delta x_n^m \tag{2.5.22}$$

where $J_n^{(x)}$ is a background error term for a control variable x at total wavenumber n, B_n is a background covariance matrix for x, L_n is a lower triangular matrix.

In summary, normalized control variables $\Delta y_n^m(k)$ are independent completely, and normalized by the background error variance. The background term of the cost function is simplified as a summation of square of the normalized control variables.

2.5.6.5 Conversions from Preconditioned Control Variables to Analysis Variables

The conversions from the preconditioned control variables to the analysis variables are performed by the following procedures:

$$
\Delta y \xrightarrow{XCT2QD}
\begin{pmatrix}
\Delta \zeta \\
\Delta D_U \\
\Delta (p_S, T)_U \\
\Delta \ln q
\end{pmatrix}
\xrightarrow{RO2BMA}
\begin{pmatrix}
\Delta \zeta \\
\Delta D_U \\
\Delta (p_S, T)_U \\
\Delta \ln q \\
\Delta \phi_B
\end{pmatrix}
\xrightarrow[\substack{UDI2DI \\ LNQ2Q}]{UPT2PT}
\begin{pmatrix}
\Delta \zeta \\
\Delta D \\
\Delta (p_S, T) \\
\Delta q
\end{pmatrix}
\tag{2.5.23}
$$

First, the control variables (relative vorticity, unbalanced divergence, unbalanced temperature and surface pressure, the logarithm of specific humidity) on each model level in the spectral space are reconstructed by using Eq. (2.5.22) (XCT2QD in Eq. (2.5.23)). Then, the modified balance mass variable $\Delta \phi_B$ is calculated from the relative vorticity $\Delta \zeta$ by Eq. (2.5.15). The temperature and the surface pressure and the divergence are calculated by Eq. (2.5.11)-Eq. (2.5.12) from the unbalanced variables (UPT2PT and UDI2DI). The logarithm of specific humidity is converted to specific humidity (LNQ2Q). If the specific humidity of the first guess in the grid space is negative due to the wave-to-grid transform, Δq is set to be zero.

2.5.7 Observation Terms

2.5.7.1 Observation Data

Assimilated observation types are shown in Table 2.2.1. Brief explanation for each data type and the quality control procedures are found in Section 2.2 and Section 2.3.

Observational data and the departures (observation minus first guess) are given with the location and time through pre-analysis procedure. Reported surface pressure data at the station height and sea surface pressure data of surface observation are assimilated after converted onto the model surface height in prior to assimilation. Scatterometer data are assimilated as the wind data at the lowest model level, although the data are considered as wind data at 10 m height above the sea level. Satellite radiance data from MW sounders, MW imagers and CSRs are directly assimilated using the K matrix model of RTTOV10 (Saunders *et al.* 2012). GNSS-RO data are assimilated in the form of refractivity at the tangent point.

2.5.7.2 Observation Error

Observation errors (the diagonal part of the observation error covariance matrix) are estimated based on the innovation statistics (Desroziers *et al.* 2005). The observation errors are summarized in Table 2.5.1. The error at arbitrary reported pressure level is linearly interpolated in the logarithm of pressure ($\log(p)$). The cross correlations of the observation errors (off diagonal part of the observation error covariance matrix) are not considered explicitly in the 4D-Var. To ignore the cross correlation term in the cost function, (horizontally or vertically) dense observations are thinned spatially in the pre-analysis procedure and the observation errors are inflated with the predefined factors.

2.5.7.3 Observation Operator

In the 4D-Var, observations at the given location and time are simulated by using the forecast variables at the surrounding grids in the nearest forecast hour with the spatial inter/extrapolation and the variable conversion. Observation operator consists of these consecutive procedures. Linear vertical interpolation is performed first with respect to logarithm of pressure. Extrapolation under the model ground surface is also performed. Then linear horizontal interpolation and extrapolation to the North and South poles are carried out.Then the variable conversions such as u, v wind components to wind speed are performed.

Table 2.5.1: The observation error tables used in the operational Global Analysis for (a) direct observations, (b) AMV, (c) AMSU-A, (d) MHS, (e) SSMIS, (f) TMI and (g) CSR from five geostationary satellites . P_s, u, v, T, RH and T_B denote surface pressure, (u, v wind components), temperature, relative humidity and brightness temperature, respectively. "x" in (c)-(g) denotes that the channel is not used.

(a) conventional observation						(b) AMV	
element level(hPa)	P_s(hPa)	u, v(m/s)	T(K)	RH(%)		element level(hPa)	u, v(m/s)
1000		2.3	1.7	6.4		1000	4.5
850		2.4	1.2	15.9		850	4.5
700		2.5	1.0	19.8		700	4.5
500		2.5	0.8	31.5		500	4.5
300		2.7	0.9	31.7		300	5.3
200		2.8	1.1	24.1		200	5.8
100	0.7	3.1	1.2	3.8		100	6.8
50		3.0	1.4	1.4		50	7.0
30		3.0	1.5	1.3		30	7.2
10		3.9	2.5	1.3		10	7.6
1		4.6	5.4	1.3		1	9.1
0.4		7.7	7.6	1.3		0.4	10.6
0.1		7.7	7.6	1.3		0.1	10.6

(c) AMSU-A T_B (K)						
satellite channel	Aqua	Metop-A	NOAA-15	NOAA-16	NOAA-18	NOAA-19
4	x	0.45	0.45	x	0.45	0.45
5	x	0.3	0.3	0.3	0.3	0.3
6	0.45	0.3	0.3	0.3	0.3	0.3
7	x	x	0.3	x	0.3	0.3
8	0.3	0.3	0.3	x	0.3	x
9	0.3	0.3	0.3	0.3	0.3	0.3
10	0.3	0.3	0.3	0.3	0.3	0.3
11	0.3	0.45	x	0.45	0.45	0.45
12	0.64	0.64	0.64	0.64	0.64	0.64
13	1.02	1.02	0.85	0.85	1.02	1.02

(d) MHS T_B (K)				(e) SSMIS T_B (K)				(f) TMI T_B (K)	
satellite channel	Metop-A	NOAA-18	NOAA-19	satellite channel	DMSP-F16	DMSP-F17	DMSP-F18	satellite channel	TRMM
3	18	18	x	13	7.6	7.6	7.6	3	7.2
4	13.5	13.5	13.5	14	10	10	10	5	10
5	9	9	9	16	8	8	8	6	7.2
				17	8.8	8.8	8.8	8	8.8

(g) CSR T_B (K)					
satellite	GOES-13	GOES-15	Meteosat-7	Meteosat-9	MTSAT-2
	1.5	1.5	1.5	1.5	1.5

27

2.5.7.4 Variational Bias Correction

As mentioned in the Subsection 2.3.4, biases of satellite radiance data are corrected by variational bias correction (VarBC). In the 4D-Var with VarBC, the observation operators are extended to include bias correction terms and the control (analysis) variables are extended to include bias correction (regression) coefficients. The coefficients are optimized as control variables in each analysis.

The extended form of the cost function Eq. (2.5.1) is defined as follows.

$$
J(\Delta z_0) = \frac{1}{2} \Delta x_0^{\mathrm{T}} \mathbf{B}^{-1} \Delta x_0 + \frac{1}{2} \Delta \beta^{\mathrm{T}} \mathbf{B}_\beta^{-1} \Delta \beta + \frac{1}{2} \sum_{i=0}^{n} \left(\mathbf{H}_i \Delta x_i + \sum_{j=0}^{m} \Delta \beta_j p_{i,j} - d_i \right)^{\mathrm{T}} \mathbf{R}_i^{-1} \left(\mathbf{H}_i \Delta x_i + \sum_{j=0}^{m} \Delta \beta_j p_{i,j} - d_i \right) + J_C
$$

(2.5.24)

where,

$$
\Delta z_0 \equiv \left[\Delta x_0^{\mathrm{T}}, \Delta \beta^{\mathrm{T}} \right]^{\mathrm{T}}, \quad \mathbf{B}_\beta \equiv diag \left(\frac{F_{inf}^2}{N_{var}}, \ldots, \frac{F_{inf}^2}{N_{var}} \right), \quad N_{var} \equiv \begin{cases} \frac{N}{\log_{10} \frac{N}{N_0} + 1} & (N \geq N_0) \\ N_0 & (N < N_0) \end{cases}
$$

Δz_0 is the extended increments which consists of the low resolution model variables' increments Δx_0 and the bias correction coefficients' increments $\Delta \beta$, $\mathbf{B}_\beta$ is the background error covariance matrix for the bias correction coefficients β, $p_{i,j}$ is the predictors for the bias correction, m is the number of all the predictors for all the radiance observation types, F_{inf} is the inflation factor defined arbitrarily, N is the number of data and N_0 is the threshold for valid number of data.

The second term of the right hand side of Eq. (2.5.24) is the background term for the bias correction coefficients and the term $\sum_{j=0}^{m} \Delta \beta_j p_{i,j}$ is the bias correction term. This equation is used in the 4D-Var instead of the Eq. (2.5.1).

2.6 Meso-scale Analysis

2.6.1 Introduction

The Meso-scale Analysis (MA) produces initial conditions for the Meso-scale Model (MSM, Section 3.5) every three hours, aimed at incorporating information from observations into the model to assist in better forecasting weather phenomena, with emphasis on high impact events.

In March 2002, a 4-dimensional variational (4D-Var) scheme was introduced as the data assimilation scheme of the MA (Ishikawa and Koizumi 2002) in place of a 3-dimensional optimal interpolation (3D-OI) scheme. This is the first operational limited-area 4D-Var system in the world. After the MSM forecast model was upgraded to a non-hydrostatic model (JMA-NHM; Saito *et al.* 2006) in September 2004, a non-hydrostatic model-based 4D-Var, the "JMA Nonhydrostatic model"-based Variational Analysis Data Assimilation (JNoVA; Honda *et al.* 2005), replaced the former hydrostatic 4D-Var in April 2009, allowing the MA to produce initial conditions more consistent with the revised MSM forecast model.

Various observations are used to contribute to improve accuracy of predictions for meso-scale weather events, including those from weather radars, satellite observations, and ground-based GNSS. Assimilating these observations with an advanced data assimilation scheme of 4D-Var, the MA is designed to produce highly-balanced initial conditions consistent with the model equations.

2.6.2 Operational System

The MA adopts the JNoVA system and produces initial conditions for the MSM forecasts every 3 hours (00, 03, 06, 09, 12, 15, 18 and 21UTC). Figure 2.6.1 shows a schematic depiction of the MA process. This process is carried out as follows (ordinal numbers correspond to those in Figure 2.6.1):

1. Initialized with the previous MA, run the high-resolution (5km) forecast model within the data assimilation window (0 to 3-hours) to obtain the first guess.

2. Perform quality-control of observations (see Section 2.3 for details) and calculate deviations of the observations from the first guess.

3. Execute the JNoVA to assimilate the observations on a low-resolution (15km) space.

4. Add analysis increments (on the low-resolution space) to the (high-resolution) first guess through an interpolation process, and make an initial condition for the next step.

5. With the initial condition made in the previous step, run the high-resolution (5km) forecast model within the data assimilation window to obtain an initial condition for the MSM.

In the MA, the first and the last steps in which the high-resolution forecast model is run are called "outer loop", and the step in which the JNoVA on the low-resolution space is executed is called "inner loop". The forecast model used in the two outer loops is identical to the model used in the MSM, namely, JMA-NHM. The analysis domain is shown in Figure 2.6.2, with a topographic map at a 5-km resolution used in the MA. Lateral boundary conditions are given by the Global Spectral Model (GSM) forecasts, while the initial conditions of the first guess are taken over from the previous MA (3-hour forecast in the last outer loop). In other words, the MA frames a cycle analysis being nested into the GSM.

The data assimilation window is set to 3 hours, and the end of the window corresponds to an analysis time. The cut-off time of the inputted observation data for the MA is 50 minutes after each analysis time. The received observational data by the cut-off time are distributed to 4 time-slots by rounding off the observation time to hours (as described by 4 starry shapes under curly braces in Figure 2.6.1). Therefore, the data observed within the period from 3.5 hours before to 0.5 hours after the analysis time are assimilated in the inner loop.
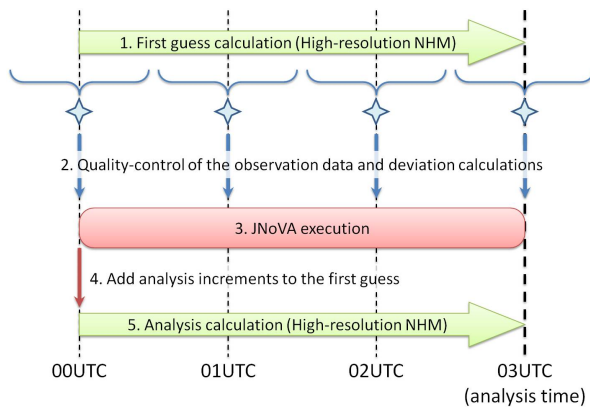

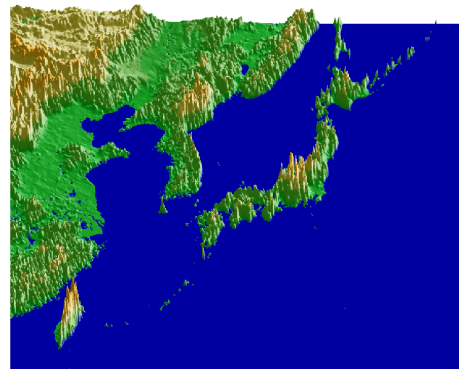Figure 2.6.1: Schematic procedure of the MA (an example of 03UTC analysis)

Figure 2.6.2: Analysis domain and topography of the MA

The JNoVA in the inner loop, as mentioned above, is a data assimilation system based on the 4-dimensional variational (4D-Var) method, detailed in the following Subsection 2.6.3. The variational method is based on the maximum likelihood estimation, and the optimal values (i.e. analysis fields) are determined by minimizing the "cost function" (see Subsection 2.6.3.1 for details). In this minimizing procedure, the cost function and its gradient require to be calculated iteratively (about 32 times on average) and it needs considerable computational costs. For the purpose of reducing these costs, the operational JNoVA adopts the incremental approach (Courtier *et al.* 1994). In this approach, a low-resolution model relative to the model used in the outer loop (equivalent to the MSM) is used in the minimization of the cost function. Operational formulation of minimizing the cost function with this incremental approach is explained in Subsection 2.6.3.1. The model used in the minimization process in the JNoVA is called "inner model", and its specifications are described in Subsection 2.6.3.3. As for the model resolution, we use the horizontal grid spacing of 5km (721×577 grid points) with 50 vertical layers in the outer loops. On the other hand, larger horizontal grid spacing of 15km (241×193 grid

points) with 40 vertical layers are used in the inner loop. In daily operation, the calculation time of the inner loop is about 21 minutes, while the outer loops are about 3 minutes each.

2.6.3 Basic Formulation

2.6.3.1 Cost Function

In the MA system, 4D-Var data assimilation is used to seek the model trajectory in a phase space by minimizing its deviation from observations and the first guess. The deviation is measured using a cost function J, defined as

$$J(x_0) = J_b + J_o + J_p = \frac{1}{2}\left(x_0 - x_0^b\right)^T \mathbf{B}^{-1}\left(x_0 - x_0^b\right) + \sum_{t=0}^{N} \frac{1}{2}\left(H(x_t) - y_t\right)^T \mathbf{R}^{-1}\left(H(x_t) - y_t\right) + J_p , \qquad (2.6.1)$$

where subscript T indicates transpose.

The first and the second terms of Eq. (2.6.1) are called the background and the observation terms, which measure the deviation from the first guess and the observations, respectively. x_0 is the model state at the beginning of the data assimilation window (time level $t = 0$) to be optimized[5], x_0^b the first guess of the model state at $t = 0$, y_t a column vector consisting of observational data available at t ($t = 0, ..., N$). x_t is a model state at t which is forecasted from the initial condition x_0, that is

$$x_t = M_t(x_0) , \qquad (2.6.2)$$

where M_t denotes the forecast operator. H is an observation operator which converts a model state space to an observation space. The H typically consists of conversions from model variables to observed parameters and interpolations from model grid points to observation points. The error covariance matrixes, $\mathbf{B}$ and $\mathbf{R}$, specify error profiles (uncertainty and error correlation) of x_0^b and y_t, respectively(see Subsection 2.6.3.2 and Subsection 2.6.4.2).

The third term of Eq. (2.6.1) J_p is the penalty term based on a digital filter to suppress high frequency noises (mainly gravity wave noises)(Gauthier and Thépaut 2001). The penalty term is given as

$$J_p = \frac{\lambda}{2} \left\| \delta x_{N/2} - \delta \overline{x}_{N/2} \right\|_E^2 , \qquad (2.6.3)$$

where λ denotes a weighting factor, $\delta x_{N/2}$ the model state analysis increment at the center of the data assimilation window ($t = N/2$), $\delta \overline{x}_{N/2}$ digital-filtered analysis increment at $t = N/2$, $\| \cdots \|_E$ the moist total energy norm suggested by Ehrendorfer *et al.* (1999).

For a time series of the model states over the data assimilation window, $\{x_0, ..., x_N\}$, the digital-filtered state $\overline{x}$ at $t = N/2$ is given as

$$\overline{x}_{N/2} \quad = \quad \sum_{k=0}^{N} h_{N/2-k} W_k x_k , \qquad (2.6.4)$$

where

$$h_k \quad = \quad \frac{\sin k\theta_c}{k\pi} , \qquad (2.6.5)$$

denotes the low-pass filter that removes time oscillations exceeding a cutoff frequency θ_c. The Dolph-Chebyshev window function W_k (Lynch 1997) is also used to suppress the noise from the Fourier truncation (Gibss oscillation).

In the MA, the incremental approach is adopted to reduce the computational costs. Some implementations to reduce the computational costs of 4D-Var scheme are proposed by Courtier *et al.* (1994). In the MA, the

---

[5]The vector to be optimized, x_0, can also include lateral boundary conditions over the data assimilation window, but this is not adopted in the operational MA.

Remark5 of Section3 in the Courtier *et al.* (1994) is used[6]. In the incremental approach of MA, the optimization is performed using a inner model (see Subsection 2.6.3.3) to obtain an analysis increment in the low resolution model space (inner loop). A low resolution version of an analysis increment at $t = 0$, δw_0, and background error covariance $\mathbf{B}_W$ are given as

$$\delta w_0 = \mathbf{S}(x_0 - x_0^b) , \qquad (2.6.6)$$

$$\mathbf{B}_W \approx \mathbf{S}\mathbf{B}\mathbf{S}^T , \qquad (2.6.7)$$

where $\mathbf{S}$ denotes a conversion operator from the high resolution to the low resolution. As a result, the low resolution cost function can be expressed as

$$J(\delta w_0) = \frac{1}{2}\delta w_0^T \mathbf{B}_W^{-1} \delta w_0 + \frac{1}{2}\sum_t \left[G(w_t) - \hat{y}_t \right]^T \mathbf{R}^{-1} \left[G(w_t) - \hat{y}_t \right] + J_p , \qquad (2.6.8)$$

$$w_t = L_t(w_0) , \qquad (2.6.9)$$

where G denotes the low resolution observation operator, L_t the inner model forecast operator. The observation vector y_t is modified as

$$\hat{y}_t = y_t - H(x_t^b) + G(w_t^b) , \qquad (2.6.10)$$

to input the high resolution departures $y_t - H(x_t^b)$ to the inner loop. The high resolution analysis at $t = 0$ is given as

$$x_0 = x_0^b + \mathbf{S}^{-I}\delta w_0 , \qquad (2.6.11)$$

where $\mathbf{S}^{-I}$ denotes an interpolation operator from the low resolution to the high resolution model space. The final analysis x_N is obtained by running forecast with the high resolution model over the data assimilation window (outer loop).

$$x_N = M_N(x_0) . \qquad (2.6.12)$$

2.6.3.2 Background Error Covariance

As mentioned above, the background error covariance $\mathbf{B}$ specifies error profiles of the first guess (Subsection 2.6.3.1). However, calculations using explicit $\mathbf{B}$ are unfeasible, because of an extremely large dimension of the model state space. In practice, drastic simplifications are applied on $\mathbf{B}$ to make the problem tractable.

A group of parameters are defined as control variables, and their background errors are treated to be uncorrelated with each other. The control variables used in the MA are as follows.

- u: x-component of horizontal wind
- v: y-component of horizontal wind
- (θ, p_s): potential temperature and surface pressure
- $\tilde{q}_v = q_v/q_v^{bs}$: pseudo relative humidity (q_v: specific humidity, q_v^{bs}: saturation specific humidity of the first guess)

For each control variable (denoted by ϕ), spatial structure of the background error covariance $\mathbf{B}^\phi$ is modeled as

$$\mathbf{B}^\phi = \mathbf{B}_v^{\phi 1/2} \mathbf{C}_h^{\phi 1/2} \mathbf{C}_h^{\phi 1/2 T} \mathbf{B}_v^{\phi 1/2 T} . \qquad (2.6.13)$$

The $\mathbf{B}_v^{\phi 1/2}$ is a square root of the vertical background error covariance $\mathbf{B}_v^\phi (= \mathbf{B}_v^{\phi 1/2} \mathbf{B}_v^{\phi 1/2 T})$, which is diagonal with respect to horizontal locations, and is assumed to be homogeneous over the domain, that is, the matrix element corresponding to a pair of spatial points (i, j, k) and (i', j', k'), $B_{v(i,j,k)(i'j'k')}^\phi$, is expressed as

$$B_{v(i,j,k)(i'j'k')}^\phi = \delta_{ii'}\delta_{jj'}B_{vkk'}^{\phi\text{col}} . \qquad (2.6.14)$$

---

[6]This method approximates a propagation in time of the perturbation ($\delta x_0 = x_0 - x_0^b$) using the non-linear inner model in the lower resolution, by taking a finite difference of inner model forecasts of observables from perturbed and unperturbed states ($w_0 = w_0^b + \delta w_0$ and w_0^b), that is, by taking $G(w_t) - G(w_t^b)$ as seen from Eq. (2.6.9) and Eq. (2.6.10).

Therefore, taking a single vertical column, we discuss properties of $\mathbf{B}_v^{\phi\text{col}}$ for simplicity. The $\mathbf{B}_v^{\phi\text{col}}$ can be written as an eigenvalue decomposition

$$\mathbf{B}_v^{\phi\text{col}} = \mathbf{V}\mathbf{D}\mathbf{V}^T , \tag{2.6.15}$$

where columns of $\mathbf{V}$ are eigenvectors of $\mathbf{B}_v^{\phi\text{col}}$, and $\mathbf{D}$ is the diagonal matrix of eigenvalues. The form of the $\mathbf{B}_v^{\phi\text{col}1/2}$ used in the MA is given as

$$\mathbf{B}_v^{\phi\text{col}1/2} = \mathbf{V}\mathbf{D}^{1/2} . \tag{2.6.16}$$

Thus, the matrix element of $\mathbf{B}_v^{\phi 1/2}$ is given as

$$B_{v(i,j,k)(i'j'\kappa')}^{\phi 1/2} = \delta_{ii'}\delta_{jj'}B_{v k\kappa'}^{\phi\text{col}1/2} , \tag{2.6.17}$$

where κ' denotes an vertical eigenmode.

The $\mathbf{C}_h^{\phi 1/2}$ is a square root of the horizontal background error correlation, $\mathbf{C}_h^{\phi}(= \mathbf{C}_h^{\phi 1/2}\mathbf{C}_h^{\phi 1/2T})$, which is diagonal with respect to eigenmodes of $\mathbf{B}_v^{\phi\text{col}}$. The matrix element of $\mathbf{C}_h^{\phi 1/2}$ is expressed as

$$C_{h(i,j,\kappa)(i',j',\kappa')}^{\phi} = \delta_{\kappa,\kappa'}C_{h(i,j)(i',j')}^{\phi\kappa} , \tag{2.6.18}$$

where (i, j) and (i', j') denotes horizontal coordinates, and κ and κ' denote vertical eigenmodes. A Gaussian form is assumed for the horizontal correlation between points (i, j) and (i', j') for an eigenmode κ, that is, a matrix element $C_{h(i,j)(i',j')}^{\phi\kappa}$,

$$C_{h(i,j)(i',j')}^{\phi\kappa} = \exp\left[-\frac{1}{2}\left\{\left(\frac{i-i'}{\sigma_x^{\phi\kappa}}\right)^2 + \left(\frac{j-j'}{\sigma_y^{\phi\kappa}}\right)^2\right\}\right] = C_{hxii'}^{\phi\kappa}C_{hyjj'}^{\phi\kappa} . \tag{2.6.19}$$

The horizontal correlation length in x- and y-directions σ_x^{κ} and σ_y^{κ} are also assumed to be homogeneous over the domain. In the MA, $\mathbf{C}_h^{\phi 1/2}$ is given as

$$C_{h(i,j,\kappa)(i',j',\kappa')}^{\phi 1/2} = \delta_{\kappa,\kappa'}C_{h(i,j)(i',j')}^{\phi\kappa 1/2} = \delta_{\kappa,\kappa'}C_{hxii'}^{\phi\kappa 1/2}C_{hyjj'}^{\phi\kappa 1/2} , \tag{2.6.20}$$

where $\mathbf{C}_{hx}^{\phi\kappa 1/2}$ ($\mathbf{C}_{hy}^{\phi\kappa 1/2}$) is taken to be the Cholesky decomposition of $\mathbf{C}_{hx}^{\phi\kappa}$ ($\mathbf{C}_{hy}^{\phi\kappa}$), that is, a lower triangular matrix. The final background error covariance $\mathbf{B}$ in the model state space is given as

$$\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2T} , \tag{2.6.21}$$

$$\mathbf{B}^{1/2} = \mathbf{K}\,\text{diag}(\mathbf{B}^{u1/2}, \mathbf{B}^{v1/2}, \mathbf{B}^{(\theta,p_s)1/2}, \mathbf{B}^{\tilde{q}_v1/2}) , \tag{2.6.22}$$

where $\mathbf{K}$ is a linearized transform from the space of control variables to the model state space. In the transform $\mathbf{K}$, pressure is determined from the control variable (θ, p_s), assuming hydrostatic balance.

Based on this simplified $\mathbf{B}$, a variable transform is made from the analysis increment in the model state $\delta\boldsymbol{x}_0$ to a new variable $\boldsymbol{\lambda}$, which is related to $\delta\boldsymbol{x}_0$ by $\mathbf{B}^{1/2}$

$$\delta\boldsymbol{x}_0 = \boldsymbol{x}_0 - \boldsymbol{x}_0^b = \mathbf{B}^{1/2}\boldsymbol{\lambda} . \tag{2.6.23}$$

and the variational optimization is performed with respect to $\boldsymbol{\lambda}$. This transform, called as a preconditioning, simplifies the background term of the cost function J (see Eq. (2.6.1) and Eq. (2.6.9). For simplicity, the present discussion does not deal with J_p). The cost function and its gradient after the transform are given as [7]

$$J(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{\lambda} + \sum_{t=0}^{N}\frac{1}{2}\Big(H(M_t(\boldsymbol{x}_0^b + \mathbf{B}^{1/2}\boldsymbol{\lambda})) - \boldsymbol{y}_t\Big)^T\mathbf{R}^{-1}\Big(H(M_t(\boldsymbol{x}_0^b + \mathbf{B}^{1/2}\boldsymbol{\lambda})) - \boldsymbol{y}_t\Big), \tag{2.6.24}$$

---

[7]Here, the resolutions of the inner model is set to be the same with that of the outer model for simplicity. The actual operational implementation uses two different resolutions based on the incremental approach as discussed previously (see Eq. (2.6.9)). The formulation in line with the operational implementation is obtained by making replacements, $(\delta\boldsymbol{x}_0, \boldsymbol{x}_0, \boldsymbol{x}_0^b, \mathbf{B}^{1/2}, \boldsymbol{y}_t, M_t, H) \rightarrow (\delta\boldsymbol{w}_0, \boldsymbol{w}_0, \boldsymbol{w}_0^b, \mathbf{B}_w^{1/2}, \hat{\boldsymbol{y}}_t, L_t, G)$, in Eq. (2.6.23), Eq. (2.6.24), and Eq. (2.6.25).

$$\nabla_\lambda J = \lambda + \sum_{t=0}^{N} \mathbf{B}^{1/2T} \mathbf{M}_t^T \mathbf{H}^T \mathbf{R}^{-1} \Big(H(M_t(\boldsymbol{x}_0^b + \mathbf{B}^{1/2}\lambda)) - \boldsymbol{y}_t \Big), \qquad (2.6.25)$$

where $\mathbf{M}_t^T$ and $\mathbf{H}^T$ are the adjoint model and the adjoint of the observation operator.

The parameters that characterize the error profile, $\mathbf{B}_v^\phi$, $\sigma_x^{\phi\kappa}$, and $\sigma_y^{\phi\kappa}$, are estimated using the NMC method (Parrish and Derber 1992). Differences between 6h and 12h MSM forecasts valid at the same time are calculated for different cases, and used as statistical samples of the background error. The samples are generated for the first ten days of each month of 2005, two pairs of the MSM forecasts a day, with each pair using the same boundary condition. The background error statistics, $\mathbf{B}_v^\phi$, $\sigma_x^{\phi\kappa}$, and $\sigma_y^{\phi\kappa}$, are obtained by taking an average over all these samples. The MA uses constant error statistics throughout the year, without taking into account seasonal variation of the error profiles.

2.6.3.3 Inner Model

In the JNoVA, a simplified nonlinear version of the JMA-NHM (NLM, M_t in Subsection 2.6.3.1 and Subsection 2.6.3.2) is used in the inner loop to provide trajectories at every iteration instead of the tangent linear model (TLM) of the NLM due to discontinuity and nonlinearlity of the JMA-NHM. The adjoint model (ADM, $\mathbf{M}_t^T$ in Subsection 2.6.3.2) of the NLM provides gradient information Eq. (2.6.25). The TLM has been developed only for use in the process of the ADM development. In addition to the use of simplified model, the inner loop of the JNoVA is executed using the NLM and ADM with a lower resolution (15 km horizontal grid spacing and 40 vertical layers) to reduce computational cost. In this subsection, the specification of the inner model in the JNoVA is surveyed focusing on schemes different from those of the MSM and also briefly listed in Table 2.6.1.

The version of the JMA-NHM in the NLM and ADM is not same as the MSM. In order to develop the TLM and ADM, the version of the JMA-NHM was fixed at the time when the development started in 2002. After that, some improvements of the MSM were included in the JNoVA. However, only some of the improvements have been adopted because of the development cost of the TLM and ADM. The NLM in the JNoVA gives the basic fields of the ADM, and all the grid point values at each time step are saved within the memory at the operation. The forecast variables are momentum (3 components), potential temperature, pressure and mixing ratio of water vapor. Additionally, temperature and evaporation efficiency at the land surface are predicted (those at the sea surface are assumed to be constant). Prognostic variables of the ADM are the same as those of the NLM except for evaporation efficiency.

In the NLM and ADM, fully compressible elastic equations are only supported as governing equations. The vertical coordinate is z*-coordinate, which is different from hybrid coordinate of the MSM. For this reason, the vertical interpolation from the outer loop to the inner loop is necessary. Regarding to the advection scheme, flux form fourth-order centered difference, same as the MSM, is adopted, but the advection correction scheme is only used for the NLM. In the horizontally explicit, and vertically implicit (HE-VI) scheme of the NLM and ADM, gravity wave and sound wave are split and calculated in a smaller time step. The smaller time step interval divides one larger time step (40 sec.) to seven small steps. In the JNoVA, all prognostic variables of the smaller time step are reserved and used for the ADM integration. As for the nonlinear computational diffusion,

$$D_{NL} \propto \frac{\partial}{\partial x}\left(\left|\frac{\partial \phi}{\partial x}\right| \frac{\partial \phi}{\partial x}\right)$$

is used in MSM and NLM(Nakamura (1978)), however, the perturbation of $\left|\frac{\partial \phi}{\partial x}\right|$ is not considered in ADM. Targeted moisture diffusion is adopted in the NLM, but not available in the ADM. Except for the advection correction, the nonlinear computational diffusion and targeted moisture diffusion, the dynamical processes of NLM are strictly linearized in the ADM.

For the moist processes, the large-scale condensation (LSC) scheme for the grid scale precipitation are used in the NLM and ADM. As a sub-grid scale convective parameterization, the NLM adopts the modified Kain-Fritsch scheme and the ADM does not considered.

As a turbulence scheme, the diagnostic-type Deardorff scheme (Deardorff 1980) is used in the NLM and ADM. Perturbations of turbulent kinetic energy in the ADM is not considered, resulting in no perturbation for the diffusive coefficients.

For a surface process in the NLM and ADM, Louis *et al.* (1982) for the land surface and Kondo (1975) for the sea surface, which were previously used in the MSM, are adopted. In the both schemes of the ADM, the perturbations of the bulk coefficients are not considered.

The NLM and ADM also use the four-layer heat diffusive model for ground temperature same as the MSM, but the perturbation is only considered at the highest layer which is close to the lowest layer of the atmosphere. The evaporation efficiency is given by the climate value in the NLM and ADM for simplicity.

The scheme for the radiation process in the NLM is slightly different from that in the MSM. The cloud is diagnosed by not partial condensation scheme but relative humidity. A method to evaluate the effective radius of a cloud ice particle in the NLM is based on Ou and Liou (1995) without modification by McFarquhar *et al.* (2003) . Additionally, Räisänen (1998) have been adopted for a method to calculate long wave radiation at each cloud layer in the MSM, whereas, it has not been adopted in the NLM. In the ADM, the radiation process is not considered for simplicity.

Table 2.6.1: Specification of MSM and models employed in JNoVA

	MSM	NLM	ADM
Resolution	5km, 50layers	15km, 40layers	15km, 40layers
Horizontal advection	Flux form fourth-order with advection correction	Flux form fourth-order with advection correction	Flux form fourth-order
Solver of pressure equation	HE-VI	HE-VI	HE-VI
Targeted moisture diffusion	Considered	Considered	Not considered
Moist physics	3-ice bulk microphysics	LSC	LSC
Convection	modified Kain-Fritsch	modified Kain-Fritsch	None
Turbulence	Mellor-Yamada-Nakanishi-Niino level-three	Deardorff	Deardorff
Surface flux	Beljaars and Holtslag	Louis(land) and Kondo(sea)	Louis(land) and Kondo(sea)
Ground temperature	4-layer thermal diffusion	4-layer thermal diffusion	4-layer thermal diffusion
Radiation	Considered	Considered	Not considered

2.6.4   Observation Terms

2.6.4.1   Observation Data

Assimilated observation types are shown in Table 2.2.1. Brief explanation for each data type and the quality control procedures are found in Section 2.2 and Section 2.3.

2.6.4.2   Observation Error

Observation error covariance matrix **R** in Eq. (2.6.1) is assumed to be a diagonal matrix and the cross correlation between the different observations are not considered. The observation errors (diagonal components of **R**) are estimated based on the innovation statistics (Desroziers *et al.* 2005). The observation errors for conventional observations and wind profilers, and AMV are summarized in Table 2.6.2. The errors for satellite radiances are the same as the Global Analysis (See Table 2.5.1(c)-(g)). The errors for GNSS-PWV and radial velocity is 3 mm and 3.3 m/s, respectively. The error at an arbitrary reported pressure level is linearly interpolated in the logarithm of pressure ($\log(p)$). The cross correlations of the observation errors between different observations

Table 2.6.2: The observation error tables used in the operational Meso-scale Analysis for (a) direct observations and (b) AMV. P_s, u, v, T and RH denote surface pressure, (u, v wind components in MSM Lambert projection space), temperature and relative humidity respectively.

(a) conventional observation and wind profiler						(b) AMV		
element level(hPa)	P_s(hPa)	u(m/s)	v(m/s)	T(K)	RH(%)	element level(hPa)	u(m/s)	v(m/s)
1000		2.1	1.9	1.3	9.8	1000	4.1	3.3
925		2.0	1.9	0.9	10.3	850	2.9	2.3
850		2.0	2.0	0.9	12.7	700	3.2	2.6
700		2.0	1.9	0.9	12.8	500	3.7	3.0
500		1.9	1.9	0.7	12.9	300	4.6	3.7
400		2.2	2.2	0.7	13.3	200	3.8	4.9
300		2.6	2.6	0.9	13.5	100	4.4	6.0
250	0.7	2.7	2.6	1.0	14.4	50	3.5	5.1
200		2.7	2.6	1.1	13.7	30	5.1	6.2
150		2.6	2.6	1.1	16.6	10	6.2	7.2
100		3.2	3.0	1.5	15.1			
70		3.7	3.1	1.9	13.6			
50		3.2	2.8	1.9	12.1			
30		3.2	2.8	1.9	11.8			
10		3.2	2.8	1.9	12.2			

are not considered explicitly in the 4D-Var. To ignore the cross correlation terms in the cost function, dense observations are thinned spatially and the observation errors are inflated in the pre-analysis procedure.

2.6.4.3 Special Treatment for Precipitation Data

The observation terms of the cost function Eq. (2.6.1) assume that the probability density function (PDF) for observation errors is a Gaussian function. However, the PDF for precipitation amount data does not follow the Gaussian function. Therefore, the following observation term is used for one-hour precipitation amount data (Koizumi *et al.* 2005).

$$J_o{}^{PREC}(x) = \sum_{j \ (\text{where } r_j^o \geq 0.5)}^{n} \frac{\left(H_j(x) - r_j^o\right)^2}{2\sigma_o(r_j^o)^2} \tag{2.6.26}$$

where, $H_j(x)$ denotes the observation operator that converts the state variables x to one-hour accumulated precipitation at the j-th grid point, and r_j^o the observed precipitation at the grid point. n is the number of grid points in the inner model. $\sigma_o(r_j^o)$ is the observation error which is defined as follows.

$$\sigma_o(r_j^o) \equiv \begin{cases} C_{sat} \max(r_{min}, r_j^o) & \left(H_j(x) \leq r_j^o\right) \\ C_{sat} C_a \max(r_{min}, r_j^o) & \left(H_j(x) > r_j^o\right) \end{cases}, \begin{cases} C_a = 3, \ C_{sat} = 1 & \text{for R/A} \\ C_a = 5, \ C_{sat} = 2 & \text{for satellite retrievals} \end{cases}, \ r_{min} \equiv 1\text{mm/h} \tag{2.6.27}$$

where, C_{sat} is an observation error inflation factor for satellite retrievals, C_a is the tuning factor for the asymmetric structure of the departure frequency distribution around 0.

The observations of one-hour precipitation less than 0.5 mm are not assimilated, since the quality of such observations is rather poor in snowfall cases. The observation error of the satellite retrievals is considered to be larger than R/A because the retrieval is not from one-hour accumulated observations but from instantaneous observations.

2.6.4.4 Variational Quality Control

Variational quality control (VarQC, Andersson and Järvinen 1999) is applied in the 4D-Var for the conventional observations. With VarQC, PDF of the observation error is supposed to be not a Gaussian function but a summation of a Gaussian function and a positive constant in the certain range. The constant means the probability of "rough error".

The following observation term and its gradient are used for the conventional observations in the cost function Eq. (2.6.1) in the 4D-Var with VarQC.

$$j_o^{VarQC} = -\log\left(\frac{\gamma + \exp(-j_o)}{\gamma + 1}\right), \quad \gamma \equiv \frac{A\sqrt{2\pi}}{(1-A)\,2d} \tag{2.6.28}$$

$$\nabla j_o^{VarQC} = W^{VarQC}\nabla j_o, \quad W^{VarQC} \equiv 1 - \frac{\gamma}{\gamma + \exp(-j_o)} \tag{2.6.29}$$

where, A is a prior probability of rough error, d is the maximum standard deviations below which rough error is possible, j_o^{VarQC} is the observation term for a single observation component with VarQC and j_o is the term without VarQC

Eq. (2.6.29) means that ∇j_o^{VarQC} is almost the same (effective) as ∇j_o when j_o is small ($W^{VarQC} \approx 1$) and ∇j_o^{VarQC} is almost 0 (not effective) when j_o is large ($W^{VarQC} << 1$). The observations with $W^{VarQC} < 0.25$ are regarded to be rejected by the VarQC.

2.7 Local Analysis

2.7.1 Introduction

The Local Analysis (LA), in operation starting from August 2012, produces initial conditions for the Local Forecast Model (LFM) (see Subsection 3.6.1) at a horizontal resolution of 2km.

Aimed at providing initial conditions to the high-resolution forecast model targeting small-scale severe weather events, the LA is designed to allow rapid production and frequent update of the analysis at a resolution of 5km (Subsection 2.7.2). An analysis cycle with hourly 3-dimensional variational (3D-Var) data assimilations is executed every time for the last 3 hours to incorporate information from newly received observations in each case.

As in the Meso-scale Analysis (MA), high density remote sensing data including those from weather radars and ground-based GNSS are hourly assimilated in the LA as important sources of detailed information that can contribute to a better forecast of high-impact phenomena (Subsection 2.7.4). Capability of high-resolution NWP to capture small-scale variations of topography is expected to help to reduce representativeness errors in assimilation of surface observations. Based on this viewpoint, the LA also assimilates automated surface station (AMeDAS) data ahead of other operational data assimilation systems in lower resolutions, in order to appropriately reflect effects from local-scale environments near the surface.

With these features, the LA is characterized as a data assimilation system for a high-resolution and high-frequency NWP.

2.7.2 Operational System

The LA is performed 8 times(00,03,06,09,12,15,18,21 UTC) a day to produce initial conditions for the LFM. The domain of the LA covers Japan and its surrounding areas at a horizontal resolution of 5 km with 441×551 grid points and 50 vertical layers (shown with the solid line in Figure 2.7.1.) In order to help the LFM provide severe weather forecasting promptly, the system configuration of the LA has to complete calculations within limited computational time, especially in view of our next plan to upgrade the LA operation on an extended domain as higher frequency of 24 times a day. In addition, the 4D-Var used in the MA(Subsection 2.6.1) is more advanced than the three dimensional variational data assimilation(3D-Var), but the computational time of the 4D-Var is much longer than that of the 3D-Var. Therefore, the LA employs an analysis cycle based on

the 3D-Var scheme(see Subsection 2.7.3 in detail). The 3D-Var is a part of JMA Nonhydrostatic model-based Variational Data Assimilation (JNoVA,Honda *et al.* (2005)).

In this subsection, FT=0(hour) is each LFM initial time. The data cut-off time of the LA is set to FT=0.5. The LA waits for observations to arrive by the data cut-off time, and starts the analysis cycle to assimilate observations from FT=-3.5 to 0.5(Subsection 2.7.4).

The flow of the 3-h analysis cycle of the LA is shown in Figure 2.7.2. In Figure 2.7.2, (a), (c), (e) and (g) are 3D-Var analyses, and (b), (d) and (f) are 1-h JMA-NHM forecasts(LF1). LF1 is made with MSM forecasts as boundary conditions and the previous 3D-Var analysis, and is used as the first guess by the following analysis. The first 3D-Var analysis of the cycle((a) in Figure 2.7.2) uses the latest forecast of the MSM(Subsection 3.5.1) as the first guess. 3D-Var analyses and LF1s are repeated in turn for 3 hours((a)-(g) in Fig.2.7.2). Each analysis assimilates observations within 0.5 hours before and after the analysis time. Quality control is applied on observations before each analysis, using the first guess of the analysis as a reference, and only observations that passes the quality checks are used in each analysis. Hydrometeors are not updated but taken over from the MSM forecasts at the beginning of the analysis cycle and just propagated throughout the cycle. All the processes are accomplished in about 15 minutes.
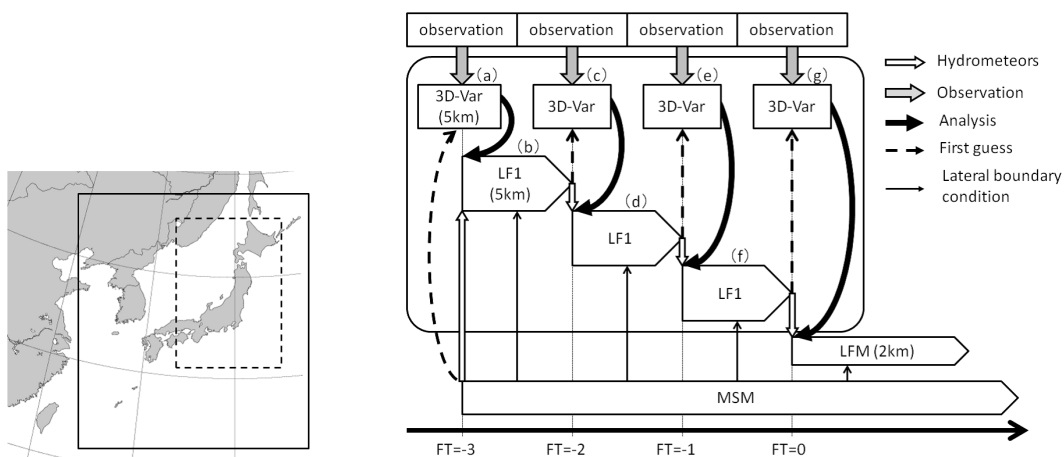

Figure 2.7.1: The domain of LA(bounded by the solid line), that of LFM (bounded by the broken line).

Figure 2.7.2: A schematic figure of the LA analysis cycle.

2.7.3 Basic Formulation

2.7.3.1 Cost Function

The LA uses a 3D-Var scheme for data assimilation. The 3D-Var optimizes the model state x to minimize its deviation from observations y and the first guess x_b, measured in terms of a cost function J,

$$J(x) = J_b + J_o = \frac{1}{2}\left(x - x^b\right)^T \mathbf{B}^{-1}\left(x - x^b\right) + \frac{1}{2}\left(H(x) - y\right)^T \mathbf{R}^{-1}\left(H(x) - y\right). \qquad (2.7.1)$$

The background term J_b and the observation term J_o measure deviations of the model state x from the first guess x^b and observations y, respectively. The observation operator H is a function from the model state space to observation space, and typically consists of collection of conversions from model variables to observed parameters and interpolations from model grid points to observation points. The background and observation error covariance matrices, $\mathbf{B}$ and $\mathbf{R}$, specify error profiles of the first guess and observations, respectively (see Subsection 2.7.3.2 and Subsection 2.7.4).

2.7.3.2 Background Error Covariance

Simplifications of **B** similar to those used in the MA are also applied in the LA. The control variables of the LA are

- u: x-component of horizontal wind
- v: y-component of horizontal wind
- (θ, p_s): potential temperature and surface pressure
- $\tilde{q}_v = q_v/q_v^{bs}$: pseudo relative humidity (q_v: specific humidity, q_v^{bs}: saturation specific humidity of the first guess)
- θ_g : ground potential temperature

In addition to the MA control variables, θ_g is also included, which is found to help appropriately assimilate surface temperature observations. The spatial structure of the background error covariance $\mathbf{B}^\phi$ of a control variable ϕ is modeled as

$$\mathbf{B}^\phi = \mathbf{U}\mathbf{B}_v^{\phi 1/2}\mathbf{C}_h^{\phi 1/2}\mathbf{C}_h^{\phi 1/2 T}\mathbf{B}_v^{\phi 1/2 T}\mathbf{U}^T \,, \tag{2.7.2}$$

where **U** denotes a vertical coordinate transformation, which is introduced to limit terrain effect of the vertical coordinate within the lower troposphere. As in the MA, assuming the vertical background error covariance, $\mathbf{B}_v^\phi$, is homogeneous over the domain, we introduce a single column vertical covariance $\mathbf{B}_v^{\phi\mathrm{col}}$ (see Eq. (2.6.14)). Using an eigenvalue decomposition of the $\mathbf{B}_v^{\phi\mathrm{col}}$ (see Eq. (2.6.15)), the square root matrix $\mathbf{B}_v^{\phi\mathrm{col}1/2}$ used in the LA is given as

$$\mathbf{B}_v^{\phi\mathrm{col}1/2} = \mathbf{V}\mathbf{D}^{1/2}\mathbf{V}^T \,, \tag{2.7.3}$$

which is taken to be a symmetric square root, compared to Eq. (2.6.16). Using this form of $\mathbf{B}_v^{\phi\mathrm{col}1/2}$, the horizontal background error correlation is specified for each vertical level, instead of each eigenmode. Thus, the horizontal correlation length in x- and y-directions σ_x^k and σ_y^k, corresponding to those in Eq. (2.6.19) are now defined on a vertical level k. In the LA, a recursive filter technique is used in calculation of $\mathbf{C}_{hx}^{\phi 1/2}$ and $\mathbf{C}_{hy}^{\phi 1/2}$ (Purser *et al.* 2003).

The background error covariance **B** in the model state space is similar with that in the MA (Eq. (2.6.21) and Eq. (2.6.22)), and is given as

$$\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2 T} \,, \tag{2.7.4}$$

$$\mathbf{B}^{1/2} = \mathbf{K}\,\mathrm{diag}(\mathbf{B}^{u 1/2}, \mathbf{B}^{v 1/2}, \mathbf{B}^{(\theta, p_s)1/2}, \mathbf{B}^{\tilde{q}_v 1/2}, \mathbf{B}^{\theta_g 1/2}) \,, \tag{2.7.5}$$

where **K** is a linearized transform from the space of control variables to the model state space.

The optimization is performed with respect to λ as in the MA (see Eq. (2.6.23), Eq. (2.6.24), and Eq. (2.6.25)),

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b = \mathbf{B}^{1/2}\lambda \,, \tag{2.7.6}$$

$$J(\lambda) = \frac{1}{2}\lambda^T\lambda + \frac{1}{2}\left(H(\mathbf{x}^b + \mathbf{B}^{1/2}\lambda) - \mathbf{y}\right)^T\mathbf{R}^{-1}\left(H(\mathbf{x}^b + \mathbf{B}^{1/2}\lambda) - \mathbf{y}\right) \,, \tag{2.7.7}$$

$$\nabla_\lambda J = \lambda + \mathbf{B}^{1/2 T}\mathbf{H}^T\mathbf{R}^{-1}\left(H(\mathbf{x}^b + \mathbf{B}^{1/2}\lambda) - \mathbf{y}\right) \,, \tag{2.7.8}$$

where $\mathbf{H}^T$ is the adjoint of the observation operator.

The background error statistics, $\mathbf{B}_v^\phi$, $\sigma_x^{\phi k}$ and $\sigma_y^{\phi k}$ are estimated applying the NMC method (Parrish and Derber 1992) to the same data set with that used in the MA. However, the error profiles in lower levels are modified to localize spatial correlations, so that surface observations are assimilated more appropriately. As in the MA, a seasonal variation of the background error statistics is not taken into account.

2.7.4 Observation Terms

2.7.4.1 Observation Data

Assimilated observation types are shown in Table 2.2.1. Brief explanation for each data type and the quality control procedures are found in Section 2.2 and Section 2.3.

2.7.4.2 Observation Error

Observation error covariance matrix $\mathbf{R}$ in Eq. (2.7.1) is assumed to be a diagonal matrix and the cross correlation between the different observations are not considered. The observation errors (diagonal components of $\mathbf{R}$) are the same as those for the Meso-scale Analysis (See Subsection 2.6.4.2).

2.7.4.3 Observation Operator

Observation operator for surface observations is based on the surface diagnostic scheme of the JMA-NHM described in Subsection 3.5.8. Observed parameters such as wind at 10 m height and temperature at 1.5 m height are diagnosed using transfer coefficients based on Beljaars and Holtslag (1991). In the adjoint operator, perturbations of the transfer coefficients are not considered in the diagnostic equations Eq. (3.5.120) and Eq. (3.5.121), but only those of u_a, θ_{va} and θ_{vs} are considered.

2.8 Snow Depth Analysis

The global snow depth analysis is executed every day separately from the global atmosphere analysis. Global snow depth with 1.0° latitude/longitude resolution is analyzed using SYNOP snow depth data on the day.

A two-dimensional optimum interpolation (OI) is employed for the analysis method. The first guess is prepared as follows,

$$G = C + \frac{1}{2} A_C \qquad (2.8.1)$$

where G, C and A_C are the first guess, the climatological value interpolated to the analysis day from monthly climatological data, and the analyzed anomaly from the climatological value on the previous day, respectively. The monthly climatological data used from September to June are the climatology compiled by USAF/ETAC (Foster and Davy 1988), and those for July and August are interpolated from the climatology for June and September.

Following type grids are the exception of this analysis; no probability of snow cover, evergreen broadleaves tree vegetation type, and land ice. The snow depth is 0 cm on the first and second type grids, and C is assigned on land ice grids. The snow cover probability is estimated from a past SSM/I observation statistic.

The analyzed snow depth with 1.0° latitude/longitude resolution is interpolated to the resolution of the GSM model grid (TL959), which is converted to snow water equivalent as an initial condition for the land-surface process (SiB) in GSM. In addition, snow water equivalents on Japan land girds are replaced with interpolated latest observation value of AMeDAS snow depth as well as Japanese SYNOP.

2.9 Non-real-time Quality Control

2.9.1 Operational Activities as a GDPFS RSMC

JMA is a Regional Specialized Meteorological Center (RSMC) of the World Meteorological Organization (WMO) Global Data-processing and Forecast System (GDPFS), namely RSMC Tokyo. In March 1991, the Commission for Basic Systems (CBS) of WMO designated RSMC Tokyo as a lead center for monitoring the quality of land surface observations in Region II (Asia). As a part of its operational activities, JMA produces a 6-monthly report containing a consolidated list of stations suspected of reporting low-quality observation data of station level pressure, mean sea level pressure and geopotential height during the 6-month periods ending June and December. The report on the quality of the land surface observations can be obtained from the website of JMA[8].

In addition, RSMC Tokyo produces monthly statistics on the quality of all observations received in time for use in its final global analyses. Copies of these reports are provided to major GDPFS centers and to the WMO Secretariat. The reports are also available in the website of JMA[9].

---

[8]http://qc.kishou.go.jp/clsf.html
[9]http://qc.kishou.go.jp/mmr.html

The data qualities are evaluated based on the differences between the observations and the first guess fields (3 to 9-hours forecasts) from the global model. These statistics are produced in accordance with the standards for the exchange of monitoring results recommended by WMO CBS.

2.9.2   Operational Activities as a PMOC

The objective of the Data Buoy Cooperation Panel (DBCP) which was established jointly by WMO and the Intergovernmental Oceanographic Commission (IOC) is to improve the quality of buoy data on the Global Telecommunication System (GTS) available to real-time users. DBCP requests some agencies or institutions to volunteer as a Principal Meteorological or Oceanographic Center (PMOC) responsible for controlling Argos GTS data on an operational basis for given variables. JMA as a PMOC undertakes the quality monitoring and makes recommendations on the recalibration and/or removal of buoy sensor data on GTS. The QC information is exchanged with DBCP and other PMOC centers through Internet in a standardized format.

2.9.3   Management of Blacklist

As mentioned in the Section 2.3, the low quality observations can lead the large forecast degradation. The cause of the low quality might be an instrumental failure and the failure can continue long time. Such observations should be excluded in the first step of QC. Blacklist is constructed to meet such a need. Blacklist management is one of the most important activities for QC. The quality for all the observations is evaluated based on the differences between the observations and the first guess fields from the global model (3 to 9-hours forecasts) and the meso-scale model (0 to 3-hours forecasts). The problematic observations are enlisted in the blacklist.

2.10   JMA Climate Data Assimilation System

JMA Climate Data Assimilation System (JCDAS) is an atmospheric global analysis for operational climate use and has been operational since March 2006. It was transitioned from Japanese 25-year Reanalysis (JRA-25) which covered 26 years from 1979 to 2004 and was produced by JMA and the Central Research Institute of Electric Power Industry (CRIEPI) (Onogi *et al.* 2007). JCDAS has been using the same data assimilation system as that of JRA-25. It has a spectral resolution of T106, equivalent to a horizontal grid size around 110 km, and 40 vertical layers with the top level at 0.4 hPa. The data assimilation method of JCDAS is the three dimensional variational (3D-var) assimilation which had been used in the operational global analysis until February 2005. The background error statistics were taken from the operational global analyses as of 2003 which were the latest statistics available at the time when JRA-25 production had started. For surface variables, surface pressure is assimilated simultaneously with upper air variables in the 3D-var; other surface variables of temperature, wind and relative humidity are assimilated separately with a uni-variate 2-dimensional optimal interpolation (2D-OI).

JRA-25 and JCDAS have many advantages. Firstly, predicted 6-hour global total precipitation distribution and amount are well reproduced both in space and time. Especially after July 1987, assimilating retrieved precipitable water from SSM/I radiance data contributed to the good performance. Furthermore, tropical cyclones (TCs) are properly analyzed owing to the assimilation of reconstructed wind profile around TCs.

JRA-25 and JCDAS jointly provide long-term consistent and high quality global analysis fields since 1979. JRA-25 and JCDAS are used to create a new climate normal value and is being used as a basic reference data for operational climate monitoring services. Reanalysis data, produced by the model whose characteristics are the same as the seasonal forecast model, can provide consistent initial field and verification data for the seasonal forecast and hindcast. Consequently the JRA-25 and JCDAS data greatly contributes to the development of the seasonal forecast model.

JMA is conducting the 2nd reanalysis named the Japanese 55-year Reanalysis (JRA-55) (Ebita *et al.* 2011). JRA-55 covers 55 years, extending back to 1958, with the four dimensional variational data assimilation (4D-var). It aims at providing a comprehensive atmospheric climate dataset by producing a more time-consistent dataset for a longer period than JRA-25. JRA-55 will be completed in the first half of 2013 and the current JRA-25 based JCDAS will be replaced with the JRA-55 version in 2014.