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, Table 2.1.2 and Table 2.1.3. 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 WMO Information System (WIS) including Global Telecommunication System (GTS), Internet and dedicated networks. The data 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 Global Analysis and Meso-scale Analysis. To reduce the computational cost, the analysis increment is calculated using a coarser-resolution inner model in the four-dimensional variational method. The resolution of these analysis type is the same as that of the corresponding forecast models.

4. Local Analysis involves a three-hour cycle based on the three-dimensional variational method. Its resolution is coarser than that of the corresponding forecast model.

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

Sea surface temperature fields (see Section 2.7) and snow depth fields (see Section 2.8) are also analyzed every day.

The Japanese 55-year Reanalysis project has been conducted for the period from 1958 and continues today on a near-real-time basis (see Section 2.10).

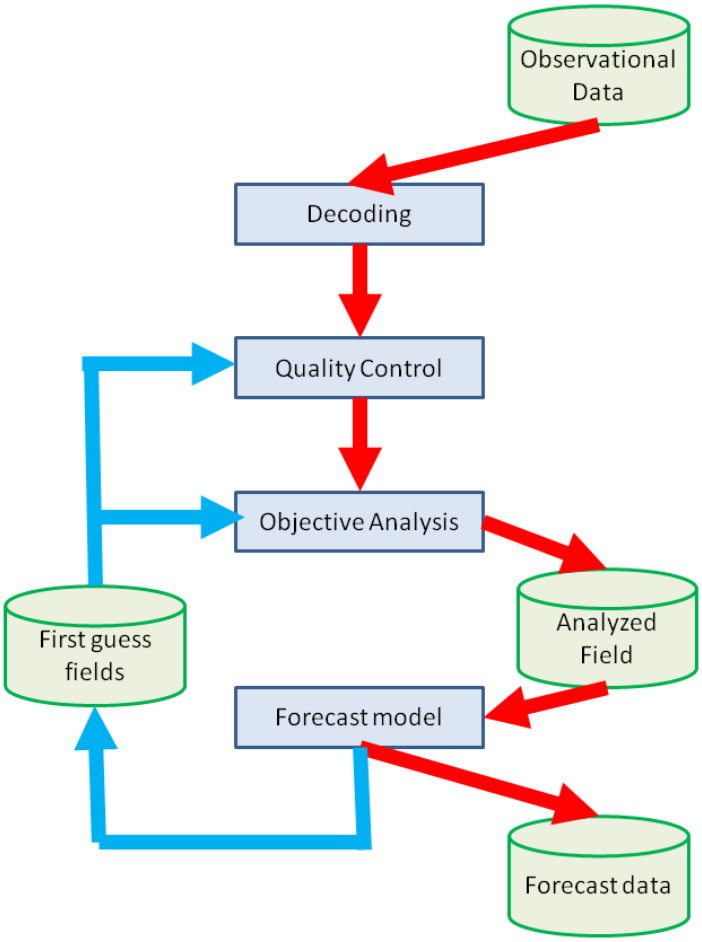

Figure 2.1.1: Major functional components and data flow in the JMA data assimilation system

Table 2.1.1: Specifications of the Global Analysis (GA)

Analysis time	00, 06, 12 and 18 UTC
Analysis scheme	Incremental 4D-Var
Data cut-off time	2 hours and 20 minutes for early run analysis at 00, 06, 12 and 18 UTC 11 hours and 50 minutes for cycle run analysis at 00 and 12 UTC 7 hours and 50 minutes for cycle run analysis at 06 and 18 UTC
First guess	6-hour forecast by the GSM
Domain configuration (Outer step)	Globe TL959,(Reduced Gaussian grid, roughly equivalent to 0.1875 ° (20 km) [1920 (tropic) – 60 (polar)] × 960
(Inner step)	TL319, Reduced Gaussian grid, roughly equivalent to 0.5625 ° (55 km) [640 (tropic) – 60 (polar)] × 320
Vertical coordinate	σ-p hybrid
Vertical levels	100 forecast model levels up to 0.01 hPa + surface
Analysis variables	Wind, surface pressure, specific humidity and temperature
Observation (as of 31 December 2017)	SYNOP, METAR, SHIP, BUOY, TEMP, PILOT, Wind Profiler, AIREP, AMDAR; atmospheric motion vectors (AMVs) from Himawari-8, GOES-[13, 15] and Meteosat-[8, 10]; MODIS polar AMVs from Terra and Aqua satellites; AVHRR polar AMVs from NOAA and Metop satellites; LEO-GEO AMVs; ocean surface wind from Metop-[A, B]/ASCAT; radiances from NOAA-[15, 18, 19]/ATOVS, Metop-[A, B]/ATOVS, Aqua/AMSU-A, DMSP-F[17, 18]/SSMIS, Suomi-NPP/ATMS, GCOM-W/AMSR2, GPM-core/GMI, Megha-Tropiques/SAPHIR, Aqua/AIRS, Metop-[A, B]/IASI and Suomi-NPP/CrIS; clear sky radiances from the water vapor channels (WV-CSRs) of Himawari-8, GOES-[13, 15] and Meteosat-[8, 10]; GNSS RO bending angle data from Metop-[A, B]/GRAS, COSMIC/IGOR, GRACE-[A, B]/Blackjack and TerraSAR-X/IGOR; zenith total delay data from gound-based GNSS
Assimilation window	6 hours

Table 2.1.2: Specifications of the Mesoscale Analysis (MA)

Analysis time	00, 03, 06, 09, 12, 15, 18 and 21 UTC
Analysis scheme	Incremental 4D-Var using a nonlinear forward model in the inner step with low resolution
Data cut-off time	50 minutes for analysis at 00, 03, 06, 09, 12, 15, 18 and 21 UTC
First guess	3-hour forecast produced by the JMA-NHM
Domain configuration (Outer step) (Inner step)	Japan and its surrounding area Lambert projection: 5 km at 60°N and 30°N, 817 × 661 Grid point (1, 1) is at the northwest corner of the domain. Grid point (565, 445) is at 140°E, 30°N Lambert projection: 15 km at 60°N and 30°N, 273 × 221 Grid point (1, 1) is at the northwest corner of the domain. Grid point (189, 149) is at 140°E, 30°N
Vertical coordinate	z-$z^\star$ hybrid
Vertical levels	(Outer step) 48 levels up to 22 km (Inner step) 38 levels up to 22 km
Analysis variables	Wind, potential temperature, surface pressure and pseudo-relative humidity
Observations (as of 31 December 2017)	SYNOP, SHIP, BUOY, TEMP, PILOT, Wind Profiler, Weather Doppler radar (radial velocity, reflectivity), AIREP, AMDAR; AMVs from Himawari-8; ocean surface wind from Metop-[A, B]/ASCAT; radiances from NOAA-[15, 18, 19]/ATOVS, Metop-[A, B]/ATOVS, Aqua/AMSU-A, DMSP-F[17, 18]/SSMIS, GCOM-W/AMSR2 and GPM-core/GMI; WV-CSRs of Himawari-8; Radar/Raingauge-Analyzed Precipitation; precipitation retrievals from DMSP-F[17, 18]/SSMIS, GCOM-W/AMSR2 and GPM-core/GMI; GPM-core/DPR; GNSS RO refractivity data from Metop-[A, B]/GRAS, COSMIC/IGOR, GRACE-[A, B]/Blackjack, TerraSAR-X/IGOR and TanDEM-X/IGOR; Total Precipitable Water Vapor from gound-based GNSS
Assimilation window	3 hours

Table 2.1.3: Specifications of the Local Analysis (LA)

Analysis time	00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23 UTC
Analysis scheme	The three-hour analysis cycle repeats hourly assimilation with 3D-Var and one-hour forecasts
Data cut-off time	30 minutes for analysis at 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23 UTC
First guess	Initial fields produced by the latest MSM
Domain configuration	Japan and its surrounding area Lambert projection: 5 km at 60°N and 30°N, 633 × 521 Grid point (1, 1) is at the northwest corner of the domain. Grid point (449, 361) is at 140°E, 30°N
Vertical coordinate	z-$z^\star$ hybrid
Vertical levels	48 levels up to 22 km
Analysis variables	Wind, potential temperature, surface pressure, pseudo-relative humidity, skin temperature, ground temperature and soil moisture
Observations (as of 31 December 2017)	SYNOP, SHIP, BUOY, AMeDAS, TEMP, PILOT, Wind Profiler, Weather Doppler radar (radial velocity, reflectivity), AIREP, AMDAR; AMVs from Himawari-8; radiances from NOAA-[15, 18, 19]/ATOVS, Metop-[A, B]/ATOVS, Aqua/AMSU-A, DMSP-F[17, 18]/SSMIS, GCOM-W/AMSR2 and GPM-core/GMI; WV-CSRs of Himawari-8; soil moisture from GCOM-W/AMSR2 and Metop-[A, B]/ASCAT; Total Precipitable Water Vapor from ground-based GNSS

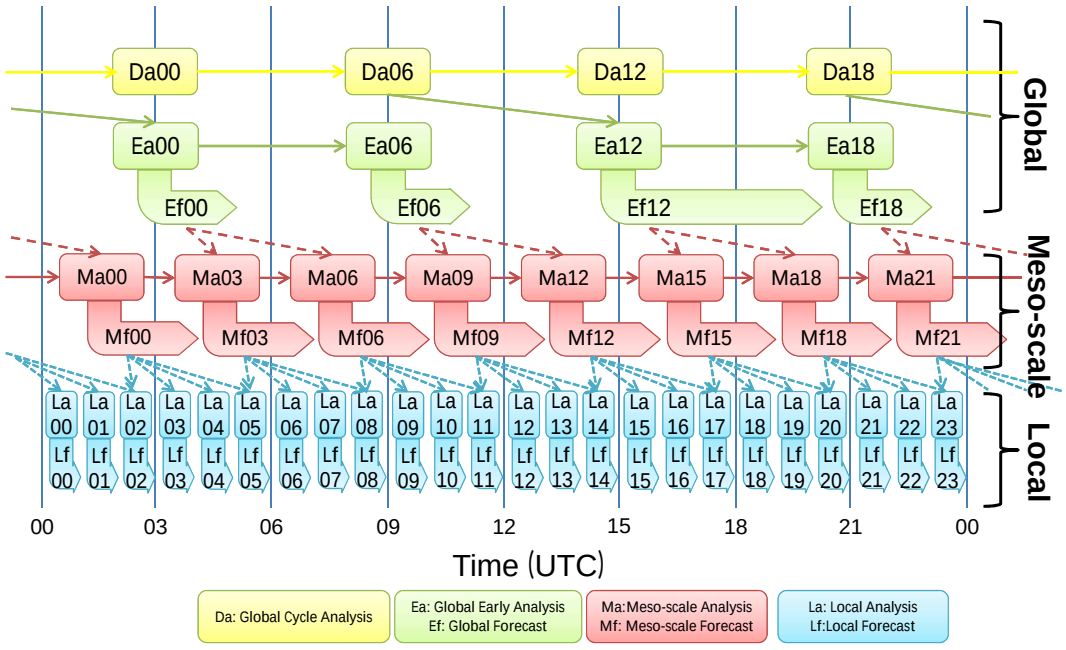

Figure 2.1.2: Main flow of JMA data assimilation systems. The first-guess and boundary conditions for Local Analysis are obtained from the latest MSM output.

2.2 Observation Data

2.2.1 Summary of Observation Data Used in Analysis

A variety of observations are utilized in JMA's current NWP systems. Table 2.2.1 summarizes the types used and the input parameters for the objective analysis systems, as of 1 November, 2018. Additional information on each observation type is provided in the following subsection.

2.2.2 Supplemental Information for Used Observation

2.2.2.1 SYNOP

SYNOP is a numerical code used for reporting surface observations at land stations. About 23,000 reports are produced every six hours.

2.2.2.2 AMeDAS

AMeDAS (the Automated Meteorological Data Acquisition System) is a JMA land surface automated network used to observe near surface temperature and wind at about 930 stations in Japan at 10 minute intervals. To monitor precipitation, around 1,300 raingauges are used in the network.

2.2.2.3 METAR

METAR is a numerical code used for reporting aerodrome weather information. Around 45,000 reports are produced every six hours.

2.2.2.4 SHIP

SHIP is a numerical code used for reporting surface observations performed at sea stations such as ships, oil rigs and moored buoys anchored at fixed locations. Around 6,400 reports are produced every six hours.

2.2.2.5 BUOY

BUOY is a numerical code used for reporting surface observations performed by drifting buoys. Around 10,000 reports are produced every six hours.

2.2.2.6 TEMP

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

2.2.2.7 PILOT

PILOT is a numerical code used for reporting upper-level wind observations performed by rawins or pilot balloons. Around 300, 200 and 100 reports produced at 00, 12, 06 and 18, respectively.

2.2.2.8 Aircraft

Aircraft observations are reported via Aircraft Report (AIREP) and Aircraft Meteorological Data Relay (AMDAR). The numerous reports received from the U.S. are thinned to 1/50th over the continental U.S. in analysis pre-processing. Even after this processing, 80,000-100,000 reports are produced every six hours covering areas around the world. While vertical profile data can be obtained at the vicinity of airports, only flight level data can be collected along the other airways.

Table 2.2.1: Summary of the observation types and parameters used in objective analysis. Third column: 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, Smc: soil moisture content. Fourth column: GA: global analysis; MA: meso-scale analysis; LA: local analysis.

Observation type (or code name used for reporting observation)	Brief description	Parameters used in analysis	Analysis type which observations are used
SYNOP	Land surface observations from world weather stations	P, u, v, T, Rh	GA, MA, LA
AMeDAS	Land surface automated observation network in Japan	u, v, T	LA
METAR	Routine weather report from aerodrome	P	GA
SHIP	Sea surface observations from ships, oil rigs and moored buoys	P, u, v, T, Rh	GA, MA, LA
BUOY	Sea surface observations from drifting buoys	P, u, v, T, Rh	GA, MA, LA
TEMP	Upper-air observations from radiosondes	P, u, v, T, Rh	GA, MA, LA
PILOT	Upper-air wind observations from rawins or pilot balloons	u, v	GA, MA, LA
Aircraft	Upper-air observations from aircraft (mainly commercial)	u, v, T	GA, MA, LA
Wind Profiler	Upper-air wind profile observations from Japan, Hong Kong and Europe	u, v	GA, MA, LA
AMV	Atmospheric motion vector (AMV) wind data from geostationary (GEO) satellites, low earth orbit (LEO) satellites and a combination of LEO and GEO	u, v	GA, MA, LA
Scatterometer	Ocean surface wind vector data from scatterometers on LEO satellites	u, v	GA, MA
MW Sounder	Radiance data from microwave (MW) sounders on LEO satellites	T_B	GA, MA, LA
MW Imager	Radiance data from MW imagers on LEO satellites and precipitation amounts estimated from MW imager radiance data	T_B, R_1(MA only)	GA, MA, LA
CSR	Clear sky radiance (CSR) data from water vapor channels on GEO satellite infrared imagers	T_B	GA, MA, LA
Hyperspectral IR Sounder	Radiance data from Infrared(IR) sounders on polar orbitting satellites	T_B	GA
GNSS-RO	Bending angle and Refractivity profile data retrieved from radio occultation (RO) measurements of global navigation satellite system (GNSS) receivers on LEO satellites	Bending Angle Refractivity	GA(Bending Angle), MA(Refractivity)
Ground-based GNSS	Zenith total delay (ZTD) data and precipitable water vapor (Pwv) data estimated from ground-based GNSS receivers	ZTD,Pwv	$GA(ZTD)$, $MA, LA(Pwv)$
Radar Reflectivity	Relative humidity data estimated using 3-dimensional reflectivity data from JMA weather (Doppler) radars and Dual-frequency Precipitation Radar (DPR) onboard GPM-core satellite.	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 amounts calibrated using AMeDAS raingauge network data	R_1	MA
Soil Moisture Contents	Soil moisture data retrieved from a microwave imager radiance and microwave scatterometer observations	Smc	LA
Typhoon Bogus	See Section 2.4.	P, u, v	GA, MA

2.2.2.9 Wind Profiler

Upper air wind speeds and directions are monitored by wind profilers on the ground. A total of 33 wind profilers operated by JMA produce data every 10 minutes. The specifications are detailed in Ishihara *et al.* (2006). Wind profiler data from Europe and Hong Kong are also available.

2.2.2.10 AMVs

Atmospheric motion vector (AMV) wind data are derived by tracing the movement of cloud or water vapor patterns in successive satellite images. AMVs from four geostationary (GEO) satellites (Meteosat-8, -11, GOES-15 and Himawari-8), low earth orbit (LEO) satellites (Terra, Aqua, NOAA and Metop) and LEOGEO-AMVs are used. LEOGEO-AMV data are derived using imagery from a combination of polar-orbiting and geostationary satellites for 60°N and 60°S latitude areas. AMVs from GEO satellites cover 60°N – 60°S and those from polar-orbiting satellites cover polar regions (i.e., latitudes higher than 60°).

2.2.2.11 Scatterometers

Ocean surface wind vectors from scatterometers onboard polar orbiting satellites are used. Data from ASCAT (the advanced scatterometer) onboard Europe's Metop-A, -B polar orbiting satellite are currently utilized.

2.2.2.12 MW Sounders

Clear radiance data from microwave (MW) sounders are used. The data adopted are from AMSU-A (Advance Microwave Sounding Unit A) on NOAA-15, -18, -19, Metop-A, -B and Aqua, MHS (Microwave Humidity Sounder) units on NOAA-18, -19, Metop-A, -B, SAPHIR (Sondeur Atmospherique du Profil d'Humidite Intertropicale par Radiometrie) on Megha-Tropiques, and ATMS (Advanced Technology Microwave Sounder) on Suomi-NPP. AMSU-A is a temperature sounder, and radiance is sensitive to temperature profiles. MHS and SAPHIR are humidity sounders. ATMS has temperature and humidity sounding channels.

2.2.2.13 MW Imagers

Radiances from MW imagers less affected by cloud/rain are used. The data adopted are from AMSR2 (Advanced Microwave Scanning Radiometer 2) on GCOM-W, SSMIS (Special Sensor Microwave Imager Sounder) on DMSP-F17, -F18 and GMI (GPM (Global Precipitation Measurement mission) Microwave Imager) on GPM. Radiance is sensitive to water vapor amounts in the lower troposphere. Precipitation amounts estimated from radiances using the MSC method (Takeuchi and Kurino 1997) are also used in meso-scale analysis.

2.2.2.14 CSR

Clear sky radiance (CSR) is a product providing averaged radiance over cloud-free pixels in GEO satellite imagers. CSR data from four GEO satellites (Meteosat-8, -11, GOES-15 and Himawari-8) are used. CSR data sensitive to water vapor amounts in the upper and middle troposphere are used.

2.2.2.15 Hyperspectral IR Sounders

Clear radiance data from hyperspectral IR sounders are used. The data adopted are from AIRS on Aqua, IASI on Metop and CrIS on Suomi-NPP. Channels located within a CO2 absoption band sensitive to temperature are used.

2.2.2.16 GNSS-RO

GNSS-RO (Global Navigation Satellite Systems - Radio Occultation) is a technique for measuring atmospheric profiles. With this approach, a set of atmospheric time delay data of GNSS radio signals received by a low earth orbit (LEO) satellite is obtained during each radio occultation event. Since the delay is a result of atmospheric radio refraction along the propagation path of the signal, the vertical profiles of refractivity (or the bending

angle) of the atmosphere at a tangent point can be estimated from the delay data set. As refractivity is a function of temperature, humidity and pressure, it can be used to determine 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 and -B.

2.2.2.17 Ground-based GNSS

Ground-based GNSS data are provided from atmospheric time delays of GNSS radio signals collected by ground-based GNSS receivers. JMA uses ground-based GNSS data collected from the global network along with GEONET GNSS receiver data. GEONET is a ground-based GNSS receiver network operated by the Geospatial Information Authority in Japan using around 1,200 receivers located throughout the country.

GNSS-ZTD (GNSS - zenith total delay) data are estimated by averaging the delays of multiple GNSS satellite signals monitored using a single receiver, and are used in global analysis.

GNSS-PWV (GNSS - Precipitable Water Vapor) data based on analysis of GEONET GNSS-ZTD data are used in meso-scale and local analysis.

2.2.2.18 Radar Reflectivity

A total of 20 C-band weather radars with Doppler functionality are operated by JMA. Three-dimensional reflectivity data are obtained every five minutes, and relative humidity profiles are estimated from reflectivity data and NWP grid point values using a technique based on Bayes' theorem (Caumont *et al.* 2010). Relative humidity data are produced for areas within a 200 km radius of each radar site below freezing level. Relative humidity profiles estimated from reflectivity data collected using the Dual-frequency Precipitation Radar (DPR) on the GPM-core satellite are used in meso-scale analysis.

2.2.2.19 Radial Velocity

A total of 20 C-band weather Doppler Radars (WDRs) and 9 Doppler Radars for Airport Weather (DRAWs) are operated by JMA. Three-dimensional radial velocity data are produced every five minutes within a 150 km radius for WDRs and every six minutes within a 120 km radius for DRAWs. The range resolution is 250 m and the azimuthal resolution is 0.703°.

2.2.2.20 R/A

Radar/Raingauge Analyzed Precipitation (R/A) is a product providing composite precipitation data produced by JMA. These data are cumulative precipitation estimations based on weather radar data with a Z-R relationship ($Z = 200R^{1.6}$) calibrated using AMeDAS raingauge data in real time. The details are found in Subsection 4.4.1.

2.2.2.21 Soil Moisture Content

Soil moisture content data estimated from microwave imager (AMSR2/GCOM-W) radiances and microwave scatterometer (ASCAT/Metop-A, -B) observations are used for areas over land in local analysis.

2.3 Quality Control and Related Procedures

Quality control (QC) is a series of procedures by which "bad" observations are screened out. It is a vital component of the objective analysis system because observations sometimes include large errors and erroneous data can significantly impair the quality of atmospheric analysis, leading to low levels of forecast skill. QC procedures in JMA's objective analysis systems are described in the following subsections.

2.3.1 SYNOP, AMeDAS, METAR, SHIP, BUOY, TEMP, PILOT, Aircraft and Wind Profilers

Direct observations (i.e. SYNOP, AMeDAS, METAR, SHIP, BUOY, TEMP, PILOT and aircraft) and wind profilers measures prognostic variables in NWP such as pressure, temperature, wind and humidity. The QC system for these observations consists of internal QC and external QC.

2.3.1.1 Internal QC

Internal QC involves procedures to check and correct observation values using collocated data in reports and several external lists or tables. The checks are outlined below.

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

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

3. Trajectory check: Consistency at consecutive locations is checked for reports from moving stations such as SHIP, BUOY and aircraft. The movement velocity and direction are checked in this step and checking is also performed to ensure that SHIP and BUOY locations are in the ocean.

4. Inter-element consistency check: The temporal continuity of consecutive reports from surface stations is checked along with consistency among observation elements within the report.

5. Vertical consistency check: Vertical consistency is checked in TEMP and PILOT data. The check items are (1) instrument icing, (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: Bias correction is applied to TEMP data reported without radiative heating correction or with apparent systematic biases. The correction constants are determined from one-month statistics for the previous month. The same bias correction approach is applied to aircraft temperature data for global analysis but not for meso-scale and local analysis.

2.3.1.2 External QC

External QC involves procedures to check observation values with comparison to (external) first guess and neighboring observations. The checks are outlined below.

1. Gross error check: The departure ($D \equiv O - B$) of the observed value (O) from the first guess (B) is calculated for all observations. The absolute value of D is compared with the tolerance limits C_P (the pass criterion) and C_R (the failure criterion). Data satisfying $|D| \leq C_P$ pass the QC, and those characterized by $|D| > C_R$ are rejected. Data characterized by $C_P < |D| \leq C_R$ are regarded as suspect and sent for spatial consistency checking.

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

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

3. Duplication check: Duplication is often found in observation reports with data obtained through different communication lines. The most appropriate report is picked out from among duplicates after performance of the above checks in consideration of status.

Table 2.3.1: Summary of blacklist areas for AMVs. IR: infrared; WV: water vapor; CSWV: clear sky water vapor; NH: Northern Hemisphere; SH: Southern Hemisphere; Polar AMV: AMVs from polar-orbiting satellites; GEO AMV: AMVs from geostationary satellites; LEOGEO AMV: AMVs from a combination of low earth orbiting and geostationary satellites.

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
LEOGEO AMV	above 300 hPa or below 900 hPa
LEOGEO AMV	between 600 hPa and 640 hPa
LEOGEO AMV	at poleward of 70°N or 70°S
LEOGEO AMV	between 640 hPa and 900 hPa at poleward of 60°S

2.3.2 AMV

Blacklisted AMVs (Table 2.3.1) are rejected in the first step, as are those with low quality indicators (QI, Holmlund 1998). QI thresholds are defined for each satellite, domain, vertical level and image type. And thinning is then performed based on a distance of 200 km. Climatological checking (see Subsection 2.3.1.1) and external QC (see Subsection 2.3.1.2) are then performed. The details of the QC and its detailed settings for AMVs are given on the NWP SAF AMV monitoring page[1].

2.3.3 Scatterometers

Level 2 ocean surface wind products are used in global analysis and meso-scale analysis. Low quality data from areas over land or sea ice are rejected at the first step. The most likely wind directions are then selected from the inherent ambiguity wind directions in scatterometer measurements using both the NWP nudging technique and the median filter technique. The next step is gross error checking (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, specialized quality control named Group-QC is applied. In this step, spatial consistency among wind vectors is checked in terms of smooth transition in wind direction and wind speed. Data that pass Group-QC are excluded from rejection in gross error checking. The details of scatterometer QC are given on the NWP SAF scatterometer monitoring page[2].

2.3.4 Satellite Radiance

Satellite radiance data are used in global and meso-scale analysis as a form of brightness temperature. The RTTOV-10 fast radiative transfer model (Saunders *et al.* 2012) is employed for radiance assimilation. The common QC procedures for radiance data are blacklist checking, thinning and external QC. The blacklist specifies problematic instruments, and is made in advance based on non-real-time QC (see Section 2.9). Blacklisted data are rejected in the first step. In the next step, data are thinned spatially in each time slot of the assimilation window (approximately one hour) to reduce computational costs. The subsequent external QC includes reduction of instrumental scan biases (except for CSR), cloud/rain contamination checking, location checking,

[1] `https://nwpsaf.eu/monitoring/amv/amvusage/jmamodel.html`
[2] `https://nwpsaf.eu/monitoring/scatter/scatusage/jmamodel.html`

Table 2.3.2: Summary of microwave sounder channel sets used for each condition

	AMSU-A	MHS	ATMS	SAPHIR	SSMIS (Water vapor sounding channels)
Clear sky ocean	Ch. 4–13	Ch. 3–5	Ch. 6–9, Ch. 18–22	Ch. 1–6	Ch. 9–11
Clear sky land/coast/sea-ice	Ch. 6–13	Ch. 3–5	Ch. 7–9, Ch. 18–22	n/a	n/a
Cloudy ocean	Ch. 7–13	Ch. 3–5	Ch. 8–9	n/a	n/a
Rainy ocean	Ch. 9–13	n/a	n/a	n/a	n/a

channel selection and gross error checking (see Subsection 2.3.1.2). Data passing this QC are thinned again to reduce the observation error correlation, and the thinned data are output for use in data assimilation systems. In global and local analysis, variational bias correction (VarBC, Derber and Wu 1998; Dee 2004) is used to reduce air-mass dependent biases. VarBC is an adaptive bias correction scheme in which a linear regression formula representing biases is embedded in the observation operator and regression coefficients are set as analysis variables. The formulations are described in Subsection 2.5.7.4. In meso-scale analysis, air-mass dependent biases are removed in pre-processing using the VarBC coefficients obtained in the latest global analysis. The satellite radiance data used are from MW sounders, MW imagers, CSR and hyperspectral IR sounders (GA only). The specific procedures for each data type are described in the following subsections.

2.3.4.1 MW sounders

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

2.3.4.2 MW imagers

Vertically polarized-channel radiances that are less affected by cloud/rain are assimilated over ice-free ocean areas. In meso-scale analysis, precipitation retrieval is also assimilated for areas over the ocean surrounding Japan. Precipitation amount estimations are resampled onto inner model grids with spatial smoothing.

2.3.4.3 CSR

CSR data are horizontally thinned to divisions of 220 km for global analysis and 45 km for meso-scale analysis and local analysis. Hourly CSR data are used in these analysis types. Values with a low percentage of clear pixels and a large standard deviation of brightness temperature are excluded due to their low representation of the area. CSR data from high-altitude areas (above 4,000 m) are not used. In calculation with RTTOV-10, emissivity atlas data and retrieved surface temperatures from window channel radiance are used for areas over land.

2.3.4.4 Hyperspectral IR sounders

Clear sky radiance data from hyperspectral IR sounders are used in global analysis. Data are horizontally thinned to divisions of 200 km, and cloud top height estimation/cloud screening are applied in quality control. The related methods are based on the CO2 slicing approach (Eyre and Menzel 1989).

2.3.5 GNSS-RO

Bending angle data for altitudes up to 60 km are used in global analysis with 500-m vertical intervals. Refractivity data are used in meso-scale analysis. No bias correction is applied for GNSS-RO data.

2.3.6 Ground-based GNSS

ZTD data are used in global analysis. Stations at an elevation of 5,000 m are not used, and those from which the absolute difference of elevation to the model surface exceeds 300 m are excluded. GNSS-ZTD values smaller than 1,000 mm or larger than 3,000 mm are rejected in climatological checking. Data with absolute differences of more than 50 mm are regarded as suspect. If the absolute difference of elevation to the model surface in values around suspect data exceeds 50 mm, the ZTD data are not used.

PWV data are used in meso-scale and local analysis. As Japan is characterized by steep mountainous terrain, large differences are found between actual ground surface elevations and model surface elevations especially in mountainous areas. In meso-scale analysis, stations at an elevation of 700 m or more above mean sea level are not used, and those from which the absolute difference of elevation to the model surface exceeds 200 m are excluded. GNSS-PWV values smaller than 1 mm or larger than 90 mm are rejected in climatological checking. The first guess PWV is then interpolated or extrapolated to the actual terrain surface and compared to the GNSS-PWV. Data with an absolute difference of more than 8 mm from the first guess are rejected in gross error checking. As there is a dense GNSS-PWV network for the analysis systems, data are thinned by 30 km for meso-scale analysis and 20 km for local analysis. GNSS-PWV data in rainy conditions are not used in meso-scale analysis.

2.3.7 Radar Reflectivity

To assimilate radar reflectivity data in meso-scale analysis and local analysis, an indirect assimilation technique called 1D+4DVAR (Ikuta and Honda 2011) is employed. This approach is based on Caumont *et al.* (2010). In 1D+4DVAR, radar reflectivity data are used to retrieve relative humidity (RH) values, which are assimilated as conventional observation data in 4D-Var. In this system, only retrieved RH values from below the melting layer are used because it is known that reflectivity inappropriately simulated in the ice phase with the operational MSM hydrometeor forecast, causing large biases in RH retrievals. In addition, data from around a height of 2000 m above sea level are also not used since these data are used for making R/A and are already assimilated in meso-scale analysis in another form (surface rainfall, see Subsection 2.3.9). For operation, reflectivity data from the JMA C-band radar network are used. Reflectivity data from space-based Dual frequency Precipitation Radar are used to retrieve RH values and assimilated in meso-scale analysis.

2.3.8 Radial Velocity

Hourly radial velocity data from WDRs and DRAWs are used in meso-scale analysis and local analysis. In pre-processing, the data are resampled into a 5 km range resolution and a 5.625° azimuthal resolution. The resampled data are checked with respect to the number of data sampled, radial velocity variance and the difference between maximum and minimum velocity. High elevation angle data ($\geq$ 5.9°) are not used to avoid the contamination of precipitation velocity values, and those from areas close to radar site ($<$ 10km) are not used to avoid the influence of back scatter noise. Data showing wind speeds of less than 5 m/s are also not used to avoid ground clutter contamination.

2.3.9 R/A

Hourly R/A data are assimilated in meso-scale analysis. As R/A data are quality controlled, 1 km grid values are simply resampled into inner-model grid boxes (15 km) and input for this type of analysis.

2.3.10 Soil Moisture Content

Variable transformation using the cumulative distribution function (CDF) matching method is applied for soil moisture content (SMC). The CDF matching method involves fitting the probability density function (PDF) of observation to the PDF of model variables. This pre-conditioning via CDF matching helps to minimize the cost function because the innovation of SMC becomes Gaussian after the CDF matching. A variational bias correction method is used for SMC in local analysis.

2.3.11 CDA: Feedback Data Base

All information concerning the quality of observational data obtained during the quality control procedure is archived in the Comprehensive Database for Assimilation (CDA), which is extensively used for both real-time and non real-time data monitoring activities. All information contained in the CDA is managed in the form of integer two byte data. The database's format is quite simple and designed for flexible use so that information on observations can be archived easily, and is also user-friendly to facilitate data retrieval.

2.4 Typhoon Bogussing

For tropical cyclones (TCs) over the western North Pacific, typhoon bogus data are generated as a form of pseudo-observation information and assimilated for realistic TC structure analysis based on model resolutions. The data consist of pressures values at the mean sea level (P_{msl}) and vertical profiles of the wind (W_{prf}) around TCs. Wind profiles are placed at 850 and 300 hPa in global analysis, 1000, 925, 850, 800, 700, 600, 500, 400 and 300 hPa in meso-scale analysis. The bogus generated has an axially asymmetric structure in types of the analysis.

Firstly, symmetric bogus profiles are generated automatically from central pressure values and the 15m/s wind speed radius of a TC (R_{15}) as analyzed by forecasters. The surface pressure profile is defined using Fujita's formula (Fujita 1952). Gradient wind balance is assumed for calculation of surface pressure profile meeting the requirements at hand from the wind speed at the particular radius of R_{15}. Upper geopotential profiles are defined using an empirical formula based on the TC analysis described by Frank (1977). It is assumed that the temperature anomaly has its maximum at 250 hPa. The wind field on each level is derived from geopotential height profiles with gradient wind balance. The surface wind field is also derived from gradient wind balance, but is modified to include the effects 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 TC center analyzed (P_{msl}), the TC center of the first guess (P_{msl}), and several points surrounding the TC center analyzed (P_{msl} and W_{prf}). The configuration for the surrounding point distribution is adaptable to the typhoon track error of the first guess.

2.5 Global Analysis

2.5.1 Introduction

A four-dimensional variational (4D-Var) data assimilation method is employed in analysis of the atmospheric state for the Global Spectral Model (GSM), and is performed at 00, 06, 12, 18 UTC. An early analysis with a short cut-off time is performed to prepare initial conditions for operational forecasting, and a cycle analysis with a long cut-off time is performed to maintain the quality of the global data assimilation system. In order to improve computational efficiency, an incremental method (Courtier *et al.* 1994) is adopted in which the analysis increment is evaluated first at a lower (inner) resolution (TL319L100: grid roughly equivalent to 0.5625° (55 km) and up to 0.01 hPa), and is then interpolated and added to the first-guess field at the original resolution (TL959L100: grid roughly equivalent to 0.1875° (20 km)). Specification of the JMA Global Analysis system is summarized in Section 2.1.

The three-dimensional variational (3D-Var) data assimilation system was operated as the first operational variational analysis system for GSM with the inner resolution of T106L40 (grid roughly equivalent to 110 km and up to 0.4 hPa) in September 2001 (Takeuchi and Tsuyuki 2002). In February 2005, the 4D-Var data assimilation system was operated with the inner resolution of T63L40 (grid roughly equivalent to 180 km) (Kadowaki 2005). The inner resolution was upgraded to T106L40 in March 2006 (Narui 2006), T159L60 (grid roughly equivalent to 75 km and up to 0.1 hPa) in November 2007, TL319L60 in October 2011 (Kadowaki and Yoshimoto 2012), and TL319L100 in March 2014.

2.5.2 Incremental 4D-Var Formulation

In 4D-Var, 3–9-hour forecasts from the GSM are used as a first guess (background), and all observations from within three hours of analysis time are organized in hourly time slots as described in the Subsection 2.5.3. The cost function is used to measure the distance between the model trajectory and observations over a six-hour assimilation window.

In an incremental method, the analysis increment is evaluated first at a low-resolution. To determine the low-resolution analysis increment Δx_i, minimization of the cost function J as defined by Eq. (2.5.1) is performed in the inner loop.

$$\begin{aligned} J\left(\Delta x_0\right) &= \frac{1}{2}\Delta x_0^{\mathrm{T}}\mathbf{B}^{-1}\Delta x_0 + \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{H}_i\Delta x_i - d_i\right)^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i - d_i\right) + J_C \\ \Delta x_{i+1} &= \mathbf{M}_i\Delta x_i \qquad (i = 0, \ldots, n-1) \end{aligned} \tag{2.5.1}$$

where the subscript $i(\geqq 1)$ is the index of time slot and n is the number of time slots. Δx_0 (Δx_1) is the low-resolution analysis increment at the initial time before (after) initialization, $\Delta x_{i\geqq 2}$ is the analysis increment evolved according to the tangent linear (TL) model to time of slot i. $\mathbf{M}_0$ is the TL operator of the nonlinear normal-mode initialization operator (Machenhauer 1977) and $\mathbf{M}_{i\geqq 1}$ is the TL model of the low-resolution nonlinear (NL) forecast model M_i as detailed in Subsection 2.5.4 for integration from time of slot i to that of slot $i+1$. $\mathbf{R}_i$ denotes the covariance matrix of observation errors at time of slot i, and $\mathbf{B}$ is the covariance matrix of background errors as detailed in Subsection 2.5.6 and Subsection 2.5.7. $\mathbf{H}_i$ is the TL operator of the observation operator H_i. The innovation vector is given for each assimilation slot by $d_i = y_i^0 - H_i x_i^b$, where x_i^b is the background state evolved using the high-resolution NL model, and y_i^0 is the observation data at time of slot i. J_C is a penalty term used to suppress gravity waves 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. This is determined via the adjoint procedures of Eq. (2.5.2)–Eq. (2.5.5) as computed in reverse time.

$$p_{n+1} = 0 \tag{2.5.2}$$

$$p_i = \mathbf{M}_i^{\mathrm{T}} p_{i+1} + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i - d_i\right) \qquad (i = n, \ldots, 1) \tag{2.5.3}$$

$$p_0 = \mathbf{M}_0^{\mathrm{T}} p_1 + \mathbf{B}^{-1}\Delta x_0 \tag{2.5.4}$$

$$\nabla J\left(\Delta x_0\right) = p_0 \tag{2.5.5}$$

where p_i is a auxiliary variable, $\mathbf{M}_i^{\mathrm{T}}$ represents 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. (2.5.2)–Eq. (2.5.5) should contain additional terms for the penalty term in Eq. (2.5.1), which are omitted here for simplicity.

The variables analyzed are relative vorticity ζ, divergence η, temperature T, surface pressure P_s ,and the logarithm of specific humidity $\ln q$ in the spectral space on the model layers (eta coordinates). The observational data y_i^0 include, but are not limited to, wind vectors, temperature, relative humidity, and satellite radiances.

The low-resolution analysis increment Δx_i determined from minimization of the cost function in the inner loop is interpolated to the high-resolution analysis increment. Adding this analysis increment to the first guess field produces high-resolution analysis.

2.5.3 Procedural Description

The flow of 4D-Var is shown in Figure 2.5.1 for the 12 UTC analysis time, and is the same for the cycle and early analyses.

The procedure is as follows:

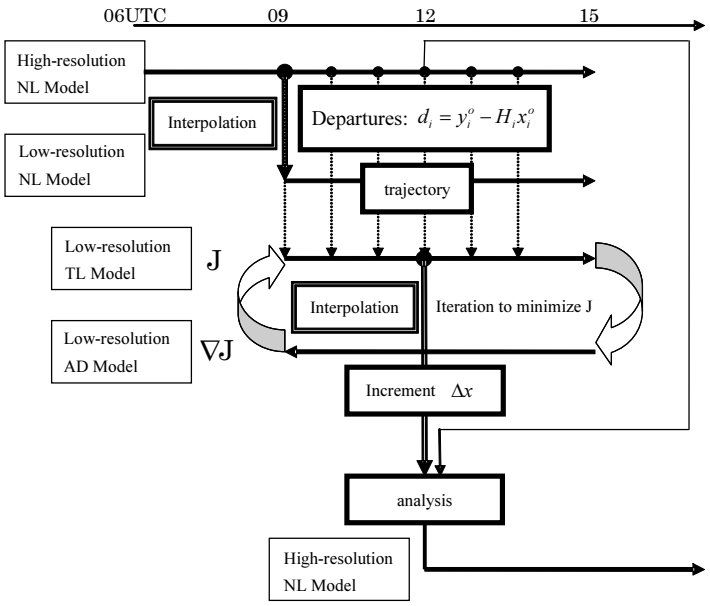

Figure 2.5.1: Flow of 4D-Var operation for the 12 UTC analysis time

1. The nine-hour forecast (09 – 15 UTC) of the high-resolution outer NL model (as per the GSM with a resolution of TL959L100) from previous analysis is used as a first guess (background), and departures between the model trajectory and observations $d_i = y_i^0 - H_i x_i^b$ over a six-hour assimilation window (09 – 15 UTC) are measured. Observations are organized in six time slots with intervals of 0.5 hours for the first slot, 1.5 hours for the last slot, and 1 hour for the others (Figure 2.5.2). All observations in each time slot are regarded as taking place at each representative time.

2. The three-hour forecast field (valid at 09 UTC) of the first guess is interpolated into the field with the resolution of the inner model (TL319L100). The interpolation is performed not only horizontally but also vertically in consideration of the topography difference between TL959 and TL319.

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

4. The TL and AD models are run 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, and the background trajectory is not updated in the system.

5. After the minimization of J, the field of the three-hour forecast (valid at 12 UTC) of the TL model is chosen as the analysis increment. It is interpolated horizontally and vertically into the field with the resolution of the first-guess field (TL959L100). Finally, the analysis increment is added to the first-guess field (valid at 12 UTC) to obtain the final product.

2.5.4 Inner Model

The inner NL model is based on the GSM, but many processes are based on those of the older GSM for many reasons. Especially, moisture processes (the convection and cloud schemes) are based on those of GSM0103 (JMA 2002), mainly for the stability of inner TL model integration. Nonlinear normal-mode initialization (Machenhauer 1977) is also added.

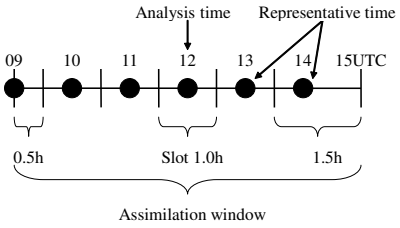

Figure 2.5.2: Schematic diagram of time slots for the analysis time 12 UTC. The black circles indicate the representative time of each time slot.

The inner TL model includes the following simple processes, most of which are based on the inner NL model:

1. **Initialization:** To control gravity waves, the TL version of the nonlinear normal-mode initialization is adopted.

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

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

4. **Vertical Turbulent Transports:** The vertical turbulent transports of momentum, heat, and moisture are formulated as the hybrid downgradient-type scheme based on the inner NL model (Subsection 3.2.7). Those diffusion coefficients are not perturbed in the inner NL model.

5. **Orographic Gravity Wave Drag:** The parameterization for orographic gravity wave drag consists of two components: one for long waves (wavelength > 100 km) and the other for short waves (wavelength $\approx$ 10 km) based on the inner NL model which is almost same as in the GSM (Subsection 3.2.8.1). The Richardson number is not perturbed in some parts for long waves for the stability of inner TL model integration.

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

$$\frac{\partial T'}{\partial t} = -\alpha \frac{g}{C_p} \frac{\partial}{\partial p}\left(4F\frac{T'}{T}\right) \tag{2.5.6}$$

   where $\alpha = 1/\{1 + (p_r/p)^{10}\}$, $p_r = 300$ hPa, F represents the net radiation fluxes calculated in the inner NL model, and g and C_p denote the gravitational constant and isobaric specific heat, respectively.

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

8. **Cumulus Convection:** Cumulus convection is formulated as the prognostic Arakawa-Schubert scheme (Arakawa and Schubert 1974; Moorthi and Suarez 1992; Randall and Pan 1993) based on the inner NL model, but is highly simplified. Vertical wind shear and the planetary mixing length are not perturbed. The magnitude of mass-flux perturbation is set bounds for the stability of inner TL model integration, and as a result this 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

$$J_C = \alpha \left(\left| \mathrm{N}_G \Delta x_0 \right|^2 + \sum_{i=2}^{\mathrm{n}} \left| \mathrm{N}_G \Delta x_i \right|^2 \right) \tag{2.5.7}$$

where N_G denotes an operator used to calculate the tendency of the gravity wave mode based on Machenhauer (1977). α is an empirically determined constant $3.0 \times 10^{-2}[\mathrm{s}^4/\mathrm{m}^2]$. Although this penalty term is primarily introduced to suppress gravity waves in the analysis increment Δx_i, it is also effective in stabilizing calculation.

2.5.6 Background Term

The background term, which is the first term on the right side of Eq. (2.5.1), dominates how the difference between observation data and the first guess is converted into correction for the first guess in the 4D-Var analysis procedure. The multivariate couplings in the analysis variables are based on the geostrophic linear balance between mass and wind. Control variables are introduced to reduce the correlations among the analysis variables, and additional statistical relations are considered in the algorithm. These include the lower geostrophic balance on smaller horizontal and vertical scales, the almost complete lack of geostrophic balance near the equator, the dependency of geostrophy on the vertical level, and the weak coupling between divergence and vorticity as well as between divergence and mass.

The control variables in 4D-Var are relative vorticity $\Delta\zeta$, unbalanced divergence $\Delta\eta_U$, unbalanced temperature and surface pressure (ΔT_U, ΔP_{sU}), and the logarithm of specific humidity $\Delta \ln q$ in the spectral space on the model layers. Δ denotes deviation from the first guess and subscript U means the term of "unbalanced". Autocovariances of the control variables are assumed to be homogeneous and isotropic. Correlation structures do not depend on geographical locations, but vertical correlations do depend on horizontal scale. The unbalanced variables $\Delta\eta_U$ and (ΔT_U, ΔP_{sU}) are defined as

$$\Delta\eta_U \equiv \Delta\eta - P\Delta\phi_B \qquad (\Delta\phi_B = \Delta\phi_B(\Delta\zeta)) \tag{2.5.8}$$

$$\begin{pmatrix} \Delta T_U \\ \Delta p_{s_U} \end{pmatrix} \equiv \begin{pmatrix} \Delta T \\ \Delta p_s \end{pmatrix} - Q\Delta\phi_B - R\Delta\eta_U \tag{2.5.9}$$

where P, Q, and R are regression coefficients, $\Delta\phi_B$ is a modified balance mass variable derived from relative vorticity as described in Subsection 2.5.6.1. This formulation is similar to that previously used in ECMWF (Derber and Bouttier 1999). The regression coefficients are computed statistically using the NMC method (Parrish and Derber 1992) with 24/48-hour forecast differences to enable estimation of the total covariances for each total spectral coefficient.

2.5.6.1 Modified Balance Mass Variable

The geostrophic balance is well kept at mid-levels in the troposphere in the extratropics. In other areas, the balance is weak. To incorporate consideration of these relationships, a modified balance mass variable is introduced. The statistical relationships linking relative vorticity, divergence, temperature, and surface pressure are calculated. First, singular value decomposition of the linear balance operator L is conducted.

$$\Delta\tilde{\phi}_B = L\Delta\zeta = UWV^{\mathrm{T}}\Delta\zeta \tag{2.5.10}$$

where $\Delta\tilde{\phi_B}$ is the original balance mass variable, W is a positive semi-definite diagonal matrix, and U and V are orthogonal matrices. The decomposed modes depend on latitude (i.e., a singular mode with a small singular value has a large amplitude in the low latitudes). Each wave number component of L is denoted as

$$\begin{aligned} \Delta\tilde{\phi}_{B_n}^{\ m} &= c_n^m \Delta\zeta_{n-1}^m + c_{n+1}^m \Delta\zeta_{n+1}^m \qquad ((n,\, m) \neq (0,\, 0),\, n = m,\, m+1, \ldots,\, N), \\ c_n^m &= -\frac{2\Omega a^2}{n^2}\sqrt{\frac{n^2-m^2}{4n^2-1}}, \qquad c_{N+1}^m = 0, \qquad \Delta\tilde{\phi}_{B_0}^{\ 0} = 0 \end{aligned} \tag{2.5.11}$$

here Ω is the angular velocity of the earth, a is the earth's radius, n is the total wavenumber, and m is the zonal wavenumber. Second, the coefficient of regression between mass variable $\Delta\Phi$ [3] (as derived from temperature and surface pressure) and balance mass variable is calculated as follows

$$D_j = \frac{\left\langle \left(U^{\mathrm{T}}\Delta\Phi\right)_j^m \left(U^{\mathrm{T}}\Delta\tilde{\phi_B}\right)_j^m \right\rangle}{\left\langle \left[\left(U^{\mathrm{T}}\Delta\tilde{\phi_B}\right)_j^m\right]^2 \right\rangle} \tag{2.5.12}$$

where $\langle\ \rangle$ denotes statistical, zonal-wavenumber, and vertical-level mean, D_j denotes a positive definite diagonal matrix, and j denotes an index of singular vectors in latitudinal wave numbers, respectively. The regression coefficients D_j $(0-1)$ indicate the extent to which geostrophic balance is satisfied. Modified balance mass variable $\Delta\phi_B$ is then constructed as follows:

$$\Delta\phi_B = UDU^{\mathrm{T}}\Delta\tilde{\phi_B} = UDWV^{\mathrm{T}}\Delta\zeta = \tilde{L}\Delta\zeta \tag{2.5.13}$$

Note that the modified balance operator $\tilde{L}$ is based on 1) conversion from the spectral space to the singular vector space, 2) the product of the regression coefficients D, and 3) conversion from the singular vector space to the spectral space. The correlation between the modified mass variable and the unbalanced mass variable (i.e. $\Delta\Phi - \Delta\phi_B$) can be neglected in all regions including the tropics.

2.5.6.2 Regression Coefficients for $\Delta\eta_U$ and $(\Delta T_U,\ \Delta P_{sU})$

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

$$P_n = \left\langle \Delta\eta_n^m \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle \left\langle \Delta\phi_{B_n}^{\ m} \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.14}$$

$$Q_n = \left\langle \begin{pmatrix} \Delta T_n^m \\ \Delta p_{s_n}^{\ m} \end{pmatrix} \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle \left\langle \Delta\phi_{B_n}^{\ m} \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.15}$$

$$R_n = \left\langle \left[\begin{pmatrix} \Delta T_n^m \\ \Delta p_{s_n}^{\ m} \end{pmatrix} - Q_n \Delta\phi_{B_n}^{\ m} \right] \left(\Delta\eta_{U_n}^{\ m}\right)^{\mathrm{T}} \right\rangle \left\langle \Delta\eta_{U_n}^{\ m} \left(\Delta\eta_{U_n}^{\ m}\right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.16}$$

where $\langle\ \rangle$ denotes the 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 $\Delta\zeta$, $\Delta\eta_U$, and $\Delta \ln q$ or the number of vertical levels +1 for $(\Delta T_U,\ \Delta P_{sU})$.

$$B_{\zeta\, n} = \left\langle \Delta\zeta_n^m \overline{\Delta\zeta_n^m}^{\mathrm{T}} \right\rangle, \qquad B_{\eta_U\, n} = \left\langle \Delta\eta_{U_n}^{\ m} \overline{\Delta\eta_{U_n}^{\ m}}^{\mathrm{T}} \right\rangle \tag{2.5.17}$$

[3] The mass variable $\Delta\Phi_k$ is defined as $\Delta\Phi_k = \Delta\phi_k + R_d\bar{T}_k\Delta p_k/\bar{p}$, where subscript k is the vertical level index, $\Delta\phi_k$ is the geopotential height, $\bar{T}_k$ is the reference (global mean) temperature, $\bar{p}$ is the reference (global mean) pressure at ground surface, Δp_k is the pressure, and R_d is the dry gas constant. In the caluculation of $\Delta\phi_k$, $\bar{T}_k$ and $\bar{p}$ are also used and some approximation is done.

$$B_{\begin{pmatrix} T_U \\ p_{sU} \end{pmatrix}_n} = \left\langle \begin{pmatrix} \Delta T_{U_n}^m \\ \Delta p_{sU_n}^m \end{pmatrix} \overline{\begin{pmatrix} \Delta T_{U_n}^m \\ \Delta p_{sU_n}^m \end{pmatrix}}^{\mathrm{T}} \right\rangle, \qquad B_{\ln q\, n} = \left\langle \Delta \ln q_n^m \overline{\Delta \ln q_n^m}^{\mathrm{T}} \right\rangle \tag{2.5.18}$$

where $\langle\ \rangle$ denotes the statistical and zonal-wavenumber mean, and overline means complex conjugate. The total variances of the control variables are rescaled by a factor of 0.8836.

2.5.6.4 Cholesky Decomposition of Background Error Covariance Matrix

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

$$J_n^{(x)} = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta x_n^m}\right)^{\mathrm{T}} B_n^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta x_n^m}\right)^{\mathrm{T}} \left(L_n L_n^{\mathrm{T}}\right)^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta y_n^m}\right)^{\mathrm{T}} \Delta y_n^m \tag{2.5.19}$$

$$\Delta y_n^m \equiv L_n^{-1} \Delta x_n^m \tag{2.5.20}$$

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

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

2.5.7 Observation Terms

2.5.7.1 Observation Data

The assimilated observation types are shown in Table 2.1.1, and brief explanations for each data type as well as the quality control procedures are found in Section 2.2 and Section 2.3.

Observational data and related departures (observation minus first guess) are given with the location and time via the pre-analysis procedure. Reported surface pressure data from the station height and sea surface pressure data from surface observation are assimilated after conversion to the model surface height prior to assimilation. Scatterometer data are assimilated as wind data at the lowest model level, but are taken as wind data at 10 m above sea level. The zenith total delay from GNSS data are assimilated over land. Satellite radiance data from MW sounders, MW imagers, Hyperspectral IR sounders and CSRs are directly assimilated using the K matrix model of RTTOV-10 (Saunders *et al.* 2012). GNSS-RO data are assimilated in the form of bending angle at the tangent point using ROPP8 (Culverwell *et al.* 2015).

2.5.7.2 Observation Error

Observation errors (the diagonal part of the observation error covariance matrix) are estimated based on innovation statistics (Desroziers *et al.* 2005), and are summarized in Table 2.5.1 and Table 2.5.2. For ocean surface wind data from scatterometers, such errors are defined as those with values of 4 or 6 m/s. For bending angle data, errors are defined as a function of height only. The observation error threshold is 1% of the observed bending angle above 10 km, and varies linearly from 20% at 0 km to 1% at 10 km. For ground-based GNSS-ZTD data, the observation error is 20 mm. The error at an arbitrary reported pressure level is linearly interpolated in the logarithm of pressure ($\log(p)$). The cross correlations of observation errors (off the diagonal part of the observation error covariance matrix) are not considered explicitly in 4D-Var. To eliminate the cross-correlation term in the cost function, horizontally or vertically dense observations are thinned spatially in pre-analysis, and observation errors are inflated with predefined factors.

Table 2.5.1: Observation error tables used in operational global analysis for (a) conventional observation, (b) AMV, (c) AMSU-A, (d) ATMS, (e) SAPHIR, (f) MHS, (g) SSMIS, (h) GMI, (i) AMSR2 and (j) CSR from four geostationary satellites. P_s, u, v, T, RH and T_B denote surface pressure, zonal and meridional wind components, temperature, relative humidity and brightness temperature, respectively. "x" in (c) - (f) denotes that the channel is not used.

(a) conventional observation				
Element Level(hPa)	P_s(hPa)	u, v(m/s)	T(K)	RH(%)
Surface	0.7			
1,000		2.3	1.7	6.4
850		2.4	1.2	15.9
700		2.5	1.0	19.8
500		2.5	0.8	31.5
300		2.7	0.9	31.7
200		2.8	1.1	24.1
100		3.1	1.2	3.8
50		3.0	1.4	1.4
30		3.0	1.5	1.3
10		3.9	2.5	1.3
1		4.6	5.4	1.3
0.4		7.7	7.6	1.3
0.1		7.7	7.6	1.3

(b) AMV	
Element Level(hPa)	u, v(m/s)
1,000	4.5
850	4.5
700	4.5
500	4.5
300	5.3
200	5.8
100	6.8
50	7.0
30	7.2
10	7.6
1	9.1
0.4	10.6
0.1	10.6

(c) AMSU-A T_B (K)						
Satellite Channel	Aqua	Metop-A	Metop-B	NOAA-15	NOAA-18	NOAA-19
4	x	0.45	x	0.45	0.45	0.45
5	x	0.3	x	0.3	x	0.3
6	x	0.3	0.3	x	0.3	0.3
7	x	x	0.3	0.3	0.3	0.3
8	0.3	x	0.3	0.3	x	x
9	0.3	0.3	0.3	0.3	x	0.3
10	0.3	0.3	0.3	0.45	0.3	0.3
11	0.3	0.45	0.45	x	0.45	0.45
12	0.64	0.64	0.64	0.64	0.64	0.64
13	x	1.02	1.02	0.85	1.02	1.02
14	x	2.83	2.83	x	2.63	2.63

(d) ATMS T_B (K)	
Satellite Channel	Suomi-NPP
6	0.3
7	0.3
8	0.3
9	0.3
18	9
19	11.25
20	13.5
21	15.75
22	18

(e) SAPHIR T_B (K)	
Satellite Channel	Megha-Tropiques
1	13.5
2	11.25
3	11.25
4	11.25
5	9
6	9

(f) MHS T_B (K)				
Satellite Channel	Metop-A	Metop-B	NOAA-18	NOAA-19
3	18	18	18	x
4	13.5	13.5	13.5	13.5
5	9	9	9	9

(g) SSMIS T_B (K)		
Satellite Channel	DMSP-F17	DMSP-F18
9	18	18
10	18	18
11	27	27
13	7.6	7.6
14	10	10
16	8	8
17	8.8	8.8

(h) GMI T_B (K)	
Satellite Channel	GPM
3	7.2
5	10
6	7.2
8	8.8
12	13.5
13	9

(i) AMSR2 T_B (K)	
Satellite Channel	GCOM-W
7	6.4
9	10
11	6.4
13	9.2

(j) CSR T_B (K)								
Satellite Central Wavelength (μm)	GOES-15	Satellite Central Wavelength (μm)	Meteosat-8	Satellite Central Wavelength (μm)	Meteosat-11	Satellite Central Wavelength (μm)	Himawari-8
6.55	1.5	6.25	1.5	6.25	1.5	6.25	1.5
		7.35	1.5	7.35	1.5	6.95	1.5
						7.35	1.5

Table 2.5.2: Observation error tables used in operational global analysis for hyperspectral IR sounders. (k) AIRS, (l) IASI and (m) CrIS.

(k) AIRS T_B (K)					
Ch.	Aqua	Ch.		Ch.	
6	1.59	144	1.29	226	1.02
7	1.62	145	1.38	227	0.96
15	1.59	150	1.38	232	1.02
20	1.56	151	1.38	239	0.93
21	1.59	156	1.26	250	1.02
22	1.56	157	1.32	251	0.96
27	1.56	162	1.29	252	0.99
28	1.56	168	1.26	253	0.96
39	1.53	169	1.17	256	1.08
40	1.53	173	1.14	257	1.05
51	1.50	174	1.20	261	1.11
68	1.44	175	1.20	262	1.08
69	1.47	179	1.14	267	1.11
71	1.44	180	1.14	272	1.11
92	1.44	185	1.17	295	1.23
93	1.44	186	1.17	299	1.02
98	1.44	190	1.11	1897	0.60
99	1.44	192	1.11	1901	0.54
104	1.47	193	1.08	1911	0.60
105	1.47	198	1.02	1917	0.54
110	1.47	201	1.05	1918	0.57
111	1.47	204	1.05	1921	0.72
116	1.47	207	1.11	1923	0.69
117	1.50	210	1.11	1924	0.66
123	1.41	213	1.02	1928	0.72
124	1.41	215	0.99		
128	1.47	216	1.02		
129	1.47	218	0.99		
138	1.32	221	0.99		
139	1.35	224	1.08		

(l) IASI T_B (K)								
Ch.	Metop-A	Metop-B	Ch.	Metop-A	Metop-B	Ch.	Metop-A	Metop-B
32	1.35	1.23	217	0.90	0.84	332	1.02	0.99
38	1.32	1.17	219	0.93	0.87	335	0.84	0.75
44	1.29	1.14	224	0.90	0.84	345	1.44	1.44
49	1.23	1.11	226	0.93	0.84	347	0.90	0.84
51	1.26	1.14	230	0.93	0.84	350	0.93	0.90
55	1.20	1.08	232	0.90	0.84	354	0.84	0.75
57	1.23	1.11	237	0.90	0.84	356	0.90	0.87
61	1.17	1.05	239	0.90	0.81	360	0.87	0.75
63	1.20	1.08	243	0.96	0.93	363	0.96	0.93
85	1.11	0.99	246	0.87	0.78			
109	1.11	0.99	249	0.90	0.81			
116	1.11	0.99	252	0.87	0.81			
122	1.08	0.96	259	0.84	0.78			
128	1.08	0.96	262	0.87	0.78			
135	1.05	0.93	265	0.84	0.78			
141	1.05	0.93	267	0.87	0.78			
146	0.99	0.87	269	0.87	0.75			
148	1.02	0.90	275	1.02	0.99			
154	1.02	0.90	279	0.87	0.84			
159	0.99	0.87	282	0.84	0.75			
161	0.99	0.87	285	0.90	0.84			
167	0.99	0.90	294	0.87	0.78			
173	0.99	0.90	296	0.90	0.81			
180	0.99	0.90	306	0.87	0.78			
185	0.96	0.87	309	0.90	0.84			
187	0.96	0.87	313	0.99	0.96			
193	0.99	0.87	320	1.02	0.99			
199	0.99	0.90	323	0.84	0.78			
205	0.96	0.87	326	1.02	0.96			
212	0.93	0.87	329	0.81	0.75			

(m) CrIS T_B (K)	
Ch.	Suomi-NPP
27	0.90
31	0.75
37	0.60
51	0.54
69	0.45
73	0.42
75	0.45
79	0.45
80	0.45
81	0.42
83	0.45
85	0.48
87	0.51
88	0.51
93	0.54
96	0.51
102	0.57
106	0.57
116	0.66
123	0.75
124	0.66
125	0.72
132	0.66
136	0.63
138	0.69
142	0.66
148	0.63

2.5.7.3 Observation Operator

In 4D-Var, observation data at a given location and time are simulated using forecast variables for the surrounding grids in the nearest forecast hour with spatial inter/extrapolation and variable conversion. Observation operator involves these consecutive procedures. A fast radiative transfer model known as RTTOV-10 is used as the observation operator for satellite radiance data assimilation. ROPP is used as the observation operator for the assimilation of GNSS-RO bending angle data. These operators are provided as external libraries from EUMETSAT NWP-SAF and ROM SAF, respectively.

2.5.7.4 Variational Bias Correction

As mentioned in Subsection 2.3.4, variational bias correction (VarBC) is applied to satellite radiance data biases. In 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. These coefficients are optimized as control variables.

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

$$J(\Delta z_0) = \frac{1}{2}\Delta x_0^{\mathrm{T}}\mathbf{B}^{-1}\Delta x_0 + \frac{1}{2}\Delta\beta^{\mathrm{T}}\mathbf{B}_\beta{}^{-1}\Delta\beta + \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{H}_i\Delta x_i + \sum_{j=0}^{m}\Delta\beta_j p_{i,j} - d_i\right)^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i + \sum_{j=0}^{m}\Delta\beta_j p_{i,j} - d_i\right) + J_C \tag{2.5.21}$$

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

The second term on the right hand side of Eq. (2.5.21) is the background term for the bias correction coefficients, and $\sum_{j=0}^{m}\Delta\beta_j p_{i,j}$ is the bias correction term. This equation is used instead of Eq. (2.5.1) in 4D-Var.

2.6 Meso-scale Analysis

2.6.1 Introduction

Meso-scale Analysis (MA) produces initial conditions for the Meso-Scale Model (MSM, Subsection 3.5.1) every three hours with incorporation of information from observations into the model for better forecasting of weather phenomena with emphasis on high-impact events.

In March 2002, a four-dimensional variational (4D-Var) scheme was introduced as the data assimilation approach for MA (Ishikawa and Koizumi 2002) in place of the previous three-dimensional optimal interpolation (3D-OI) scheme to create the world's first operational limited-area 4D-Var system. Following the upgrade of the MSM forecast model to a non-hydrostatic type (JMA-NHM; Saito *et al.* 2006, 2007) in September 2004, the previous hydrostatic 4D-Var was replaced by a non-hydrostatic model-based 4D-Var system (known as JMA Nonhydrostatic model-based Variational Analysis Data Assimilation (JNoVA; Honda *et al.* 2005)) in April 2009. This development enabled MA to produce initial conditions more consistent with the upgraded MSM forecast model. A further upgrade of the model also took place in February 2017, replacing JMA-NHM with a newly developed non-hydrostatic model called ASUCA (Ishida *et al.* 2009, 2010; Hara *et al.* 2012) as described in Subsection 3.5.1. A 4D-Var system based on ASUCA is currently under development (Aranami *et al.* 2015), and the current MA is still based on JNoVA.

The domain of MA is the same as that of the MSM, covering Japan and its surrounding areas. Operation covering the current 4,080 × 3,300 km domain, extended from the previous 3,600 × 2,880 km, was introduced in March 2013 (Subsection 3.5.1).

Various observational data are used to help improve the accuracy of prediction for meso-scale weather events, including information from weather radars, satellite observations and ground-based GNSS (see Table 2.1.2). Thanks to the advanced data assimilation scheme of 4D-Var utilized with these data, MA produces initial conditions highly consistent with the balance of model equations.

2.6.2 Operational System

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

1. With the initial condition produced in the previous MA, the high-resolution (5 km) forecast model is run within the 3-hour data assimilation window to obtain the first guess.
2. Quality-control on observations (Section 2.3) and the calculation of their deviations from the first guess are conducted for preparing inputs to the next step.
3. The JNoVA 4D-Var analysis is performed to obtain the optimal model state by assimilating the observations in a low-resolution (15 km) space.
4. The low-resolution analysis increment is added to the high-resolution first guess via an interpolation process to prepare the initial condition for the next step.
5. Based on the initial condition set in the previous step, the high-resolution (5 km) forecast model is run within the data assimilation window to obtain the initial condition for the MSM.

In MA, the first and last steps in which the high-resolution forecast model is run are called the outer steps, and the step in which the JNoVA in the low-resolution space is executed is called the inner step. The forecast model used in the two outer steps is the JMA-NHM[4]. The analysis domain is shown in Figure 2.6.2 with a topographic map at a 5-km resolution as used in MA. The lateral boundary conditions are given by Global Spectral Model (GSM) forecasts, while the initial conditions of the first guess are taken from the previous MA (the 3-hour forecast in the last outer step). In other words, MA frames cycle analysis nested into the GSM.

The data assimilation window is set to 3 hours, and the end of the window corresponds to the analysis time. The cut-off time of input observation data for MA is 50 minutes after each analysis time. Observation data received by this time are distributed into four time slots by rounding off the observation time to the nearest hour (as represented by four star shapes under the curly braces in Figure 2.6.1). Accordingly, data observed within the period from 3.5 hours before to 0.5 hours after the analysis time are assimilated in the inner step.

As described previously, the JNoVA in the inner step is a data assimilation system based on the four-dimensional variational (4D-Var) method as detailed in Subsection 2.6.3. This approach is based on maximum likelihood estimation, and optimal values (i.e. analysis fields) are determined by minimizing the cost function (see Subsection 2.6.3.1 for details). This minimization procedure requires iterative calculation of the cost function and its gradient (about 32 times on average), which takes considerable computational resources. To reduce this burden, the operational JNoVA involves an incremental approach (Courtier *et al.* 1994) in which a model with a low-resolution relative to that used in the outer step is adopted in cost function minimization. Operational formulation of this minimization with the incremental approach is detailed in Subsection 2.6.3.1. The model used in the minimization process for the JNoVA is called the inner model, and its specifications are described in Subsection 2.6.3.3. The horizontal grid spacing is 5 km (817 × 661 grid points) with 48 vertical layers in the outer steps, whereas larger horizontal grid spacing of 15 km (273 × 221 grid points) with 38 vertical layers is used in the inner step. In daily operation, the calculation time of the inner step is about 21 minutes, and those of the outer steps are about 3 minutes each.

[4]The JMA-NHM has not been used as the MSM forecast model since February 2017 (Section 3.5), but MA is still based on the JMA-NHM.

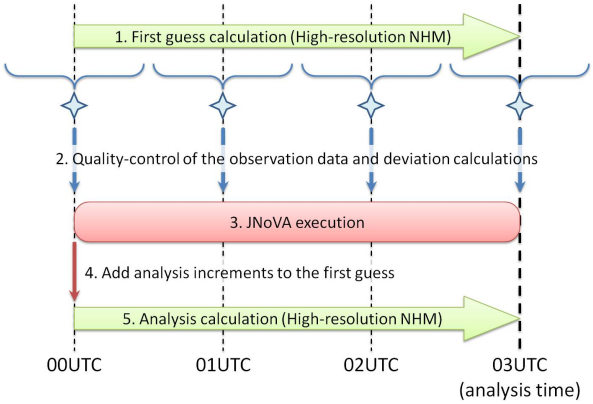

Figure 2.6.1: Schematic procedure of MA (example of 03 UTC analysis)

Figure 2.6.2: MA domain and topography

2.6.3 Basic Formulation

2.6.3.1 Cost Function

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

$$J(\boldsymbol{x}_0) = J_b + J_o + J_p = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathbf{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \sum_{t=0}^{N} \frac{1}{2}\left(H_t(\boldsymbol{x}_t) - \boldsymbol{y}_t\right)^T \mathbf{R}_t^{-1}\left(H_t(\boldsymbol{x}_t) - \boldsymbol{y}_t\right) + J_p \ , \tag{2.6.1}$$

where the superscript T stands for transpose.

The first and the second terms of Eq. (2.6.1) are called the background and observation terms, and represent deviations from the first guess and the observations, respectively. $\boldsymbol{x}_0$ is the model state at the beginning of the data assimilation window (time level $t = 0$) to be optimized[5], $\boldsymbol{x}_0^b$ is the first guess of the model state at $t = 0$, $\boldsymbol{y}_t$ is a column vector consisting of observational data available at t ($t = 0, ..., N$), and $\boldsymbol{x}_t$ is a model state at t as forecast from the initial condition $\boldsymbol{x}_0$ as

$$\boldsymbol{x}_t = M_t(\boldsymbol{x}_0) \ , \tag{2.6.2}$$

where M_t denotes the forecast operator. H_t is an observation operator used to convert model state variables to observations, typically consisting of conversion from model variables to observed parameters and interpolation from model grid points to observation points. The error covariance matrixes $\mathbf{B}$ and $\mathbf{R}_t$, specify the error profiles (uncertainty and error correlation) of $\boldsymbol{x}_0^b$ and $\boldsymbol{y}_t$, respectively(see Subsection 2.6.3.2 and Subsection 2.6.4.2).

The third term of Eq. (2.6.1) J_p is a penalty term based on a digital filter to suppress high-frequency noise mainly caused by gravity waves (Gauthier and Thépaut 2001). It is given as

$$J_p = \frac{\lambda}{2}\left\|\delta\boldsymbol{x}_{N/2} - \delta\overline{\boldsymbol{x}}_{N/2}\right\|_E^2 \ , \tag{2.6.3}$$

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

[5]Lateral boundary conditions over the data assimilation window can be included in the vector to be optimized, $\boldsymbol{x}_0$, but this is not adopted in the operational MA.

For a time series of model states over the data assimilation window $\{\boldsymbol{x}_0, ..., \boldsymbol{x}_N\}$, the digitally filtered state $\overline{\boldsymbol{x}}$ at $t = N/2$ is given as

$$\overline{\boldsymbol{x}}_{N/2} = \sum_{k=0}^{N} h_{N/2-k} W_k \boldsymbol{x}_k \ , \tag{2.6.4}$$

where

$$h_k = \frac{\sin k\theta_c}{k\pi} \ , \tag{2.6.5}$$

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

In MA, the incremental approach is adopted to reduce the computational cost. Some implementations to reduce the computational cost of the 4D-Var scheme were proposed by Courtier *et al.* (1994). In MA, the stipulation of *Remark5 in Section3* of Courtier *et al.* (1994) is used[6]. In the incremental approach for MA, optimization is performed using an inner model (see Subsection 2.6.3.3) to obtain an analysis increment for the low-resolution model space (inner step). A low-resolution version of the analysis increment at $t = 0$, $\delta \boldsymbol{w}_0$ and the background error covariance $\mathbf{B}_W$ are respectively given as

$$\delta \boldsymbol{w}_0 = \mathbf{S}(\boldsymbol{x}_0 - \boldsymbol{x}_0^b) \ , \tag{2.6.6}$$

$$\mathbf{B}_W \approx \mathbf{S}\mathbf{B}\mathbf{S}^T \ , \tag{2.6.7}$$

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

$$J(\delta \boldsymbol{w}_0) = \frac{1}{2}\delta \boldsymbol{w}_0^T \mathbf{B}_W^{-1} \delta \boldsymbol{w}_0 + \frac{1}{2}\sum_t \Big[G_t(\boldsymbol{w}_t) - \hat{\boldsymbol{y}}_t\Big]^T \mathbf{R}_t^{-1} \Big[G_t(\boldsymbol{w}_t) - \hat{\boldsymbol{y}}_t\Big] + J_p \ , \tag{2.6.8}$$

$$\boldsymbol{w}_t = L_t(\boldsymbol{w}_0) \ , \tag{2.6.9}$$

where G_t is the low-resolution observation operator and L_t is the inner-model forecast operator. The observation vector $\boldsymbol{y}_t$ is modified as

$$\hat{\boldsymbol{y}}_t = \boldsymbol{y}_t - H_t(\boldsymbol{x}_t^b) + G_t(\boldsymbol{w}_t^b) \ , \tag{2.6.10}$$

for input of the high-resolution departures $\boldsymbol{y}_t - H_t(\boldsymbol{x}_t^b)$ to the inner step. The high-resolution analysis at $t = 0$ is given as

$$\boldsymbol{x}_0 = \boldsymbol{x}_0^b + \mathbf{S}^{-I} \delta \boldsymbol{w}_0 \ , \tag{2.6.11}$$

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

$$\boldsymbol{x}_N = M_N(\boldsymbol{x}_0) \ . \tag{2.6.12}$$

2.6.3.2 Background Error Covariance

As detailed previously, the background error covariance $\mathbf{B}$ indicates the error profile of the first guess (Subsection 2.6.3.1). However, calculation using the complete form of $\mathbf{B}$ is impractical due to the extremely large dimensions of the model state space. In practice, drastic simplification is applied to $\mathbf{B}$ to make the problem tractable.

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

[6]This method involves the approximation of propagation in the time of perturbation ($\delta \boldsymbol{x}_0 = \boldsymbol{x}_0 - \boldsymbol{x}_0^b$) using the non-linear inner model in lower resolution by taking the finite difference of inner-model forecasts of observables from perturbed and unperturbed states ($\boldsymbol{w}_0 = \boldsymbol{w}_0^b + \delta \boldsymbol{w}_0$ and $\boldsymbol{w}_0^b$), or by taking $G_t(\boldsymbol{w}_t) - G_t(\boldsymbol{w}_t^b)$ as obtained from Eq. (2.6.9) and Eq. (2.6.10).

- 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 first guess)

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

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

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

$$B^{\phi}_{v(i,j,k)(i',j',k')} = \delta_{ii'}\delta_{jj'}B^{\phi\mathrm{col}}_{vkk'} \ , \tag{2.6.14}$$

where (i, j) and (i', j') denote horizontal coordinates, and k and k' denote vertical levels. Accordingly, taking a single vertical column, the properties of $\mathbf{B}_v^{\phi\mathrm{col}}$ are discussed here for simplicity. $\mathbf{B}_v^{\phi\mathrm{col}}$ can be written as an eigenvalue decomposition

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

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

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

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

$$B^{\phi 1/2}_{v(i,j,k)(i',j',k')} = \delta_{ii'}\delta_{jj'}B^{\phi\mathrm{col}1/2}_{vkk'} \ . \tag{2.6.17}$$

$\mathbf{C}_h^{\phi 1/2}$ is the 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 the vertical levels. The matrix element of $\mathbf{C}_h^{\phi 1/2}$ is expressed as

$$C^{\phi}_{h(i,j,k)(i',j',k')} = \delta_{k,k'}C^{\phi k}_{h(i,j)(i',j')} \ . \tag{2.6.18}$$

A Gaussian form is assumed for the horizontal correlation between points (i, j) and (i', j') for the vertical level k, i.e., the matrix element $C^{\phi k}_{h(i,j)(i',j')}$ is given as

$$C^{\phi k}_{h(i,j)(i',j')} = \exp\Big[-\frac{1}{2}\Big\{\Big(\frac{i-i'}{\sigma_x^{\phi k}}\Big)^2+\Big(\frac{j-j'}{\sigma_y^{\phi k}}\Big)^2\Big\}\Big] = C^{\phi k}_{hxii'}C^{\phi k}_{hyjj'} \ . \tag{2.6.19}$$

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

$$C^{\phi 1/2}_{h(i,j,k)(i',j',k')} = \delta_{k,k'}C^{\phi k1/2}_{h(i,j)(i',j')} = \delta_{k,k'}C^{\phi k1/2}_{hxii'}C^{\phi k1/2}_{hyjj'} \ , \tag{2.6.20}$$

where $\mathbf{C}_{hx}^{\phi k1/2}$ ($\mathbf{C}_{hy}^{\phi k1/2}$) is taken as the Cholesky decomposition of $\mathbf{C}_{hx}^{\phi k}$ ($\mathbf{C}_{hy}^{\phi k}$), i.e., a lower triangular matrix.

The final background error covariance $\mathbf{B}$ in the model state space is given as

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

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

where $\mathbf{K}$ is a linearized transform from the control variable space 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 form of $\mathbf{B}$, 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 $\mathbf{B}^{1/2}$

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

Variational optimization is performed with respect to $\boldsymbol{\lambda}$ thus obtained. This transform, called 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]

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

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

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

The parameters that characterize the error profile, $\mathbf{B}_v^{\phi}$, $\sigma_x^{\phi k}$ and $\sigma_y^{\phi k}$, 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 from August 2014 to July 2015 with two pairs of MSM forecasts a day. The background error statistics $\mathbf{B}_v^{\phi}$, $\sigma_x^{\phi k}$ and $\sigma_y^{\phi k}$ are taken as the averages over all these samples. MA involves the use of constant error statistics throughout the year without consideration of seasonal variations in 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 step to provide trajectories at every iteration instead of the tangent linear model (TLM) of the NLM due to the discontinuity and nonlinearity 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 a simplified model, the inner step of the JNoVA is executed using the NLM and ADM with a lower resolution (15 km horizontal grid spacing and 38 vertical layers) to reduce the computational cost. In this subsection, the specifications of the inner model in the JNoVA are surveyed focusing on schemes different from those of the JMA-NHM used in the outer step (outer-NLM). These are also briefly listed in Table 2.6.1.

The version of the JMA-NHM in the NLM and ADM differes from that in the outer-NLM. In order to develop the TLM and ADM, the version of the JMA-NHM was fixed when development started in 2002, and some improvements to the JMA-NHM were subsequently included in the JNoVA. However, only certain improvements were adopted due to the development costs 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 upon operation. The forecast variables are momentum (three components), potential temperature, pressure and mixing ratio of water vapor. Additionally, temperature and evaporation efficiency on the land surface are predicted (those on the sea surface are assumed to be constant). The prognostic variables of the ADM are the same as those of the NLM except for evaporation efficiency.

In the NLM and ADM, only fully compressible elastic equations are supported as governing equations. The vertical coordinate is z*-coordinate, which is different from hybrid coordinate of the outer-NLM. For this reason, the vertical interpolation from the outer step to the inner step is necessary. Regarding the advection scheme, the flux form fourth-order centered difference, same as the outer-NLM, 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 smaller time steps.

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

The smaller time step interval divides one larger time step (40 seconds) to seven small steps. In the JNoVA, all prognostic variables of the smaller time steps are reserved and used for the ADM integration. 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 the outer-NLM and NLM(Nakamura (1978)), however, the perturbation of $\left|\frac{\partial \phi}{\partial x}\right|$ is not considered in the 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 the NLM are strictly linearized in the ADM.

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

In regard to turbulence scheme, the diagnostic-type Deardorff scheme (Deardorff 1980) is used in the NLM and ADM. As perturbations of turbulent kinetic energy are not considered in the ADM, there is no perturbation for the diffusive coefficients.

For a surface process in the NLM and ADM, the approaches proposed by Louis *et al.* (1982) for the land surface and Kondo (1975) for the sea surface (as previously used in the JMA-NHM) 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 as with the outer-NLM, but the perturbation is considered only at the highest layer which is close to the lowest layer of the atmosphere. 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 outer-NLM. The cloud diagnosis is based on relative humideity rather than partial condensation scheme. The method to evaluate the effective radius of a cloud ice particle in the NLM is based on Ou and Liou (1995) without the modification proposed by McFarquhar *et al.* (2003) . Additionally, the approach of Räisänen (1998) is adopted to calculate long wave radiation at each cloud layer in the outer-NLM, but not in the NLM. In the ADM, consideration of the radiation process is omitted for simplicity.

Table 2.6.1: Specifications of the outer-NLM and the models employed in the JNoVA

	outer-NLM	NLM	ADM
Resolution	5km, 48layers	15km, 38layers	15km, 38layers
Horizontal advection	Flux form fourth-order with advection correction	Flux form fourth-order with advection correction	Flux form fourth-order
Pressure equation solver	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

Table 2.6.2: Observation error tables used in the operational meso-scale analysis for (a) conventional observation and wind profiler data and (b) AMV. P_s, u, v, T and RH denote surface pressure, x and y wind components in an MSM Lambert projection space, temperature and relative humidity respectively.

(a) Conventional observation and wind profiler data					
Element Level (hPa)	P_s (hPa)	u (m/s)	v (m/s)	T (K)	RH (%)
Surface	0.7				
	0.6 for SYNOP in Japan				
1,000		2.1	1.9	1.3	9.8
925		2.0	1.9	0.9	10.3
850		2.0	2.0	0.9	12.7
700		2.0	1.9	0.9	12.8
500		1.9	1.9	0.7	12.9
400		2.2	2.2	0.7	13.3
300		2.6	2.6	0.9	13.5
250		2.7	2.6	1.0	14.4
200		2.7	2.6	1.1	13.7
150		2.6	2.6	1.1	16.6
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

(b) AMV		
Element Level (hPa)	u (m/s)	v (m/s)
1,000	4.1	3.3
850	2.9	2.3
700	3.2	2.6
500	3.7	3.0
300	4.6	3.7
200	3.8	4.9
100	4.4	6.0
50	3.5	5.1
30	5.1	6.2
10	6.2	7.2

2.6.4 Observation Terms

2.6.4.1 Observation Data

Assimilated observation types are shown in Table 2.1.2, and brief outlines of each data type and related quality control procedures are given in Section 2.2 and Section 2.3.

2.6.4.2 Observation Error

The observation error covariance matrix **R** in Eq. (2.6.1) is assumed to be diagonal, and cross-correlation between different observations is not considered. Estimation of observation errors (diagonal components of **R**) is based on innovation statistics (Desroziers *et al.* 2005). Errors for conventional observations, wind profiler data and AMVs are summarized in Table 2.6.2. Errors for satellite radiance are the same as those in global analysis (See Table 2.5.1(c) - (j)). The errors for GNSS-PWV and radial velocity are 3 mm and 3.3 m/s, respectively. Errors for relative humidity data from DPR are 5%, and those from ground-based radar are approximately 30 %. Errors for ocean surface wind data from scatterometers are 3 m/s. For GNSS-RO refractivity data, observation errors are defined as a function of height only. The observation error is calculated using a linear interpolation to be 10% of refractivity at 0 km, 0.5% at 10 km, 0.18% at 20 km and 0.04% at 30 km. Errors for R/A are based on the precipitation amount (Koizumi *et al.* 2005). The error at an arbitrary reported pressure level is linearly interpolated in the logarithm of pressure ($\log(p)$). The cross-correlations of errors between different observations are not considered explicitly in 4D-Var. To eliminate consideration of cross-correlation terms in the cost function, dense observations are thinned spatially and observation errors are inflated in the pre-analysis procedure.

2.6.4.3 Observation Operator

The RTTOV-10 fast radiative transfer model is used as the observation operator for satellite radiance data assimilation. ROPP8 is used as the observation operator for the assimilation of refractivity data from GNSS-

RO. These observation operators are provided as external libraries from EUMETSAT NWP-SAF and ROM SAF, respectively.

2.6.4.4 Special Treatment for Precipitation Data

For the observation terms of the cost function Eq. (2.6.1), it is assumed that the probability density function (PDF) for observation errors is Gaussian. However, PDF for precipitation amount data does not follow a Gaussian function. Accordingly, 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, $H_j(x)$ is an observation operator used to convert the state variables x to one-hour cumulative precipitation values at the j-th grid point, r_j^o is the precipitation observed at the grid point and 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:

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

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

One-hour precipitation observation showing values of less than 0.5 mm are not assimilated, since the quality of such data is rather poor for snowfall. The observation error of satellite retrievals is considered to be larger than that of R/A because retrieval is from instantaneous observations rather than from one-hour cumulative observations.

2.6.4.5 Variational Quality Control

Variational quality control (VarQC, Andersson and Järvinen 1999) is applied in 4D-Var for conventional observations. With VarQC, the PDF of the observation error is assumed to be a summation of a Gaussian function and a positive constant value in the certain range. The constant value means the probability of rough error within the range.

The following observation term and its gradient are used for conventional observations in the cost function Eq. (2.6.1) in 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 the prior probability of rough error (e.g. $A = 0.05$ for SYNOP), d is the maximum standard deviations below which rough error is possible (e.g. $d = 9$ for SYNOP), 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) shows 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 ($W^{VarQC} << 1$). Observations satisfying $W^{VarQC} < 0.25$ are regarded to be rejected by the VarQC.

2.7 Local Analysis

2.7.1 Introduction

Local Analysis (LA) produces initial conditions for the Local Forecast Model (LFM) (Subsection 3.6.1) at a horizontal resolution of 2 km. Its operation started in August 2012, with eight runs per day on an area of Japan measuring 2,200 × 2,500 km to initialize LFM forecasts over a domain covering the eastern part of the country. An enhancement in the operation of LA, along with the LFM, was implemented in May 2013, extending its domain to cover Japan and the surrounding areas (3,160 × 2,600 km) and increasing its daily operations to 24 runs per day.

To provide initial conditions for this high-resolution forecast model targeting small-scale severe weather events, LA is designed to allow rapid production and frequent updating of analysis at a resolution of 5 km (Subsection 2.7.2). In each LA run, an analysis cycle with hourly three-dimensional variational (3D-Var) data assimilation is executed for the previous three hours to incorporate information from newly received observational data in each case. The analysis cycle was originally based on JMA-NHM (Saito *et al.* 2006, 2007) and JNoVA 3D-Var (the 3D-Var version of JNoVA (Honda *et al.* 2005)), which was replaced by the new-generation version based on ASUCA (Ishida *et al.* 2009, 2010; Hara *et al.* 2012) and ASUCA-3DVar in January 2015 (Aranami *et al.* 2015).

As with MA, high-density remote sensing data (including information from weather radars and ground-based GNSS) are assimilated on an hourly basis in LA as important sources of detailed information that can contribute to better forecasting of high-impact phenomena (see Table 2.1.3). LA was also adopted to make extensive use of satellite observations in January 2017, with application of the variational bias correction technique. The capacity of high-resolution NWP to capture small-scale variations in topography is expected to help to reduce representativeness errors in the assimilation of surface observations. In association, LA also assimilates automated surface station (AMeDAS) data ahead of other operational data assimilation systems with lower resolutions in order to appropriately reflect the effects of local-scale environments near the surface.

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

2.7.2 Operational System

To satisfy the requirements outlined in Subsection 2.7.1, operational LA incorporates an analysis cycle with 3D-Var, which can meet the relevant demand for prompt and frequent product updates within a limited time frame (all processes, including data quality control, are completed within around 15 minutes) with far fewer computational resource requirements than 4D-Var.

LA involves the running of an analysis cycle on a domain identical to that of the LFM (see Figure 2.7.1), following the flow chart shown in Figure 2.7.2. The cycle consists of four successive instances of 3D-Var analysis (3, 2, 1 and 0 hours prior to the initial time of the LFM; (a), (b), (c) and (d) in Figure 2.7.2), where quality controlled observation data rounded to the nearest hour are assimilated. After each instance of analysis except the last one ((d) in Figure 2.7.2), a one-hour forecast (LF1) with a horizontal resolution of 5 km is executed using the analysis for initial conditions, thereby providing the first guess for the next 3D-Var analysis. In the first 3D-Var ((a) in Figure 2.7.2), the first guess is supplied from the MSM (Subsection 3.6.2), which also provides lateral and upper boundary conditions for LF1 throughout the three-hour data assimilation period. The analyzed field in the last 3D-Var ((d) in Figure 2.7.2) is used as the initial condition of the LFM. Further specifications of LA are provided in Table 2.1.3.

2.7.3 Basic Formulation

2.7.3.1 Cost Function

Local Analysis involves four time slots for the assimilation of observations using 3D-Var. The related cost function in the i-th time slot is defined as

$$J(\delta \boldsymbol{x}_i) = \frac{1}{2} \delta \boldsymbol{x}_i^T \mathbf{B}^{-1} \delta \boldsymbol{x}_i \tag{2.7.1}$$

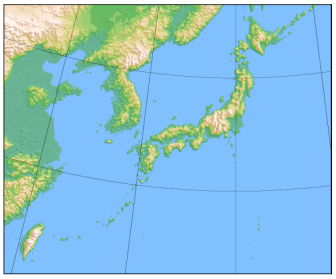

Figure 2.7.1: domains of LA and LFM.

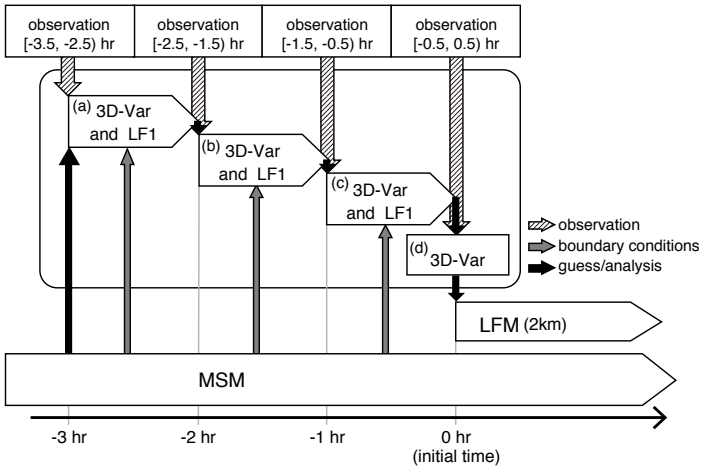

Figure 2.7.2: Schematic representation of LA analysis.

$$+\frac{1}{2}\left(\mathbf{H}\delta\boldsymbol{x}_i - \boldsymbol{d}_i + \mathbf{P}\delta\boldsymbol{b}_i\right)^T \mathbf{R}^{-1}\left(\mathbf{H}\delta\boldsymbol{x}_i - \boldsymbol{d}_i + \mathbf{P}\delta\boldsymbol{b}_i\right) \tag{2.7.2}$$

$$+\frac{1}{2}\delta\boldsymbol{b}_i^T \mathbf{S}_i^{-1}\delta\boldsymbol{b}_i \tag{2.7.3}$$

Here $\delta\boldsymbol{x}_i$ is the increment, $\delta\boldsymbol{b}_i$ is the increment of the observation bias vector, and the subscript i is the hourly time-index. The innovation vector $\boldsymbol{d}_i$ is given by

$$\boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}\left(\boldsymbol{x}_i^b\right) - \mathcal{P}\left(\boldsymbol{b}_i^b\right), \tag{2.7.4}$$

where $\boldsymbol{y}_i$ is the vector of observations, $\boldsymbol{x}_i^b$ is the background state and $\boldsymbol{b}_i^b$ is the background state of observation bias in the i-th time slot. The other symbols are as follows: $\mathcal{H}$ is the nonlinear observation operator, $\mathbf{H}$ is the tangent linearized observation operator, $\mathcal{P}$ is the nonlinear operator of predictors for VarBC, $\mathbf{P}$ is the matrix of predictors for VarBC , $\mathbf{B}$ is the covariance matrix of background error, $\mathbf{R}$ is the diagonal matrix of observation error and $\mathbf{S}_i$ is the diagonal matrix of parameters used to control the adaptivity of observation bias vector estimation in VarBC.

2.7.3.2 Solution Procedure

The analysis $\boldsymbol{x}^a$ at the initial time t of LFM, corresponding to $i = 4$, is calculated by repeating 3D-Var and one-hour forecasting. The one-hour forecast operator $\mathcal{M}_{i,i+1}$ is configured specifically for Local Analysis. In this configuration, the horizontal resolution is set to 5 km as in the MSM, but the physics schemes differ from those used in the MSM and the LFM.

The Local Analysis cycle is conducted as outlined below.

1. $\boldsymbol{x}_{i=1}^b$ valid at $t - 3$h is provided by the MSM, and $\boldsymbol{b}_{i=1}^b$ is equal to the analysis variables of observation bias $\boldsymbol{b}^a$ in the previous Local Analysis.

2. $\delta\boldsymbol{x}_i$ is optimized by minimization of the cost function, $J\left(\delta\boldsymbol{x}_i\right)$.

3. The i-th background state is updated with the one-hour forecast:

$$\boldsymbol{x}_{i+1}^b = \mathcal{M}_{i,i+1}\left(\boldsymbol{x}_i^b + \delta\boldsymbol{x}_i\right), \tag{2.7.5}$$

and the observation bias is given by

$$\boldsymbol{b}^b_{i+1} = \boldsymbol{b}^b_i + \delta \boldsymbol{b}_i. \tag{2.7.6}$$

4. Steps 2 and 3 are repeated three times.

5. For the initial time, the cost function $J\left(\delta \boldsymbol{x}_{i=4}\right)$ is minimized and the analysis increments are added to the background state. The low-resolution (5 km) analysis variables are given by

$$\boldsymbol{x}^a = \boldsymbol{x}^b_4 + \delta \boldsymbol{x}_4, \tag{2.7.7}$$

and the analysis variables of observation bias $\boldsymbol{b}^a$ are given by

$$\boldsymbol{b}^a = \boldsymbol{b}^b_4 + \delta \boldsymbol{b}_4. \tag{2.7.8}$$

Finally, the low-resolution analysis $\boldsymbol{x}^a$ is interpolated to the high resolution of 2 km with consideration of ancillaries (i.e. topography, soil type and land usage) to be used as the initial condition with the LFM.

2.7.3.3 Analysis variables

The analysis increment is defined as $\delta \boldsymbol{x} = \left(\delta u, \delta v, \left(\delta T_g, \delta p_s, \delta \theta\right), \left(\delta W_g, \delta \mu_p\right)\right)^T$.

- u: x-component of horizontal wind.
- v: y-component of horizontal wind.
- $\left(T_g, p_s, \theta\right)$: underground temperature and skin temperature, surface pressure and potential temperature.
- $\left(W_g, \mu_p\right)$: soil moisture and pseudo-relative humidity.

T_g is perturbed only on land grids, as that on ocean grids is not a prognostic variable in the LFM.

2.7.3.4 Background Error Covariance

$\delta \boldsymbol{x}$ is transformed from uncorrelated control variables χ, as follows

$$\delta \boldsymbol{x} = \mathbf{B}^{1/2} \chi, \tag{2.7.9}$$

$\mathbf{B}^{1/2}$ is the square root of $\mathbf{B}$ and has the form

$$\mathbf{B}^{1/2} = \mathbf{K}_p \mathbf{B}_s^{1/2} \tag{2.7.10}$$

$$= \mathbf{K}_p \operatorname{diag}\left(\mathbf{B}_{s,u}^{1/2}, \mathbf{B}_{s,v}^{1/2}, \mathbf{B}_{s,\left(T_g, p_s, \theta\right)}^{1/2}, \mathbf{B}_{s,\left(W_g, \mu_p\right)}^{1/2}\right) \tag{2.7.11}$$

Here $\mathbf{K}_p$ is a linearized parameter transform from the control variables to the model variables, $\mathbf{B}_s^{1/2}$ is a spatial transform and $\mathbf{B}_{s,\chi i}^{1/2}$ is the spatial transform for the i th sub-group χ_i of parameters in χ. The spatial structure of background error covariance $\mathbf{B}_{s_{\chi i}}$ is modeled as

$$\mathbf{B}_{s,\chi i}^{1/2} = \mathbf{C}_v \mathbf{B}_{v,\chi i}^{1/2} \mathbf{B}_{h,\chi i}^{1/2} \tag{2.7.12}$$

Here, $\mathbf{C}_v$ denotes a vertical coordinate transformation introduced to limit the terrain effect of the vertical coordinate within the lower troposphere.

The vertical background error covariance matrix $\mathbf{B}_{v,\chi i}$ is assumed to be horizontally homogeneous over the domain. The square root of $\mathbf{B}_{v,\chi i}$ is given by

$$\mathbf{B}_{v,\chi i}^{1/2} = \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{V}^T \ , \tag{2.7.13}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{B}_{v,\chi i}$, and $\mathbf{V}$ is the orthogonal matrix ($\mathbf{V}^T \mathbf{V} = \mathbf{I}$) whose columns are the related eigenvectors.

The horizontal background error correlation $\mathbf{B}_{h,\chi i}$ is defined on the vertical level of the model. In the calculation of $\mathbf{B}_{h,\chi i}^{1/2}$, the recursive filter technique (Purser *et al.* 2003) is used in each of the x- and y-directions.

The background error statistics $\mathbf{B}_{v,\chi i}$ and $\mathbf{B}_{h,\chi i}$ are estimated using the NMC method (Parrish and Derber 1992). However, the error profiles in the lower levels are modified artificially to localize spatial correlations so that surface observations are assimilated more appropriately. The seasonal variation of background error statistics is not taken into account.

2.7.4 Observation Terms

2.7.4.1 Observation Data

Assimilated observation types and brief outlines of each data type are provided in Table 2.1.3.

2.7.4.2 Observation Error

The observation error covariance matrix is assumed to be diagonal, and cross-correlation between different observations is not considered as it is in MA.

2.7.4.3 Observation Operators

The observation operator for surface observations (i.e wind at a height of 10 m and temperature, relative humidity at a height of 1.5 m) is based on the surface diagnostic scheme (Beljaars and Holtslag 1991) in the LFM. In the tangent linear operator for surface observation, perturbations of transfer coefficients for surface fluxes are not considered in diagnostic equations.

The observation operator for brightness temperature is RTTOV10.2 (Saunders *et al.* 2012), but perturbations of surface elements and ozone are ignored as inputs to the RTTOV tangent-linear model.

The observation operator for soil moisture is provided using a linear regression equation with coefficients estimated from cumulative distribution function matching.

2.7.5 Parallelization

The data assimilation domain is two-dimensionally decomposed into blocks, each of which is assigned to an MPI process. The loops for the y-direction and certain fused horizontal loops are forked via OMP parallelization similar to that of the LFM (see Section 3.5.10). Observations are distributed to blocks according to their location, and innovations are calculated in each process. In addition, the recursive filter method is parallelized using decomposed blocks.

In calculation of minimization, the inner product of general vectors in the L-BFGS (Nocedal and Wright 2006) is determined for each block, and loops for observations and model variables are forked via OMP parallelization.

2.8 Snow Depth Analysis

2.8.1 Global Snow Depth Analysis

Global snow depth analysis is executed every day separately from global atmosphere analysis. Global snow depth data with a 1.0° latitude/longitude resolution are analyzed using SYNOP snow depth data on the same day. Analysis involves two-dimensional optimal interpolation (2D-OI). Analysis snow depth data are interpolated into GSM model grids (TL959) that are converted to snow water equivalents as an initial condition for the land-surface process (see Subsection 3.2.9) in the GSM.

The first guess is produced for 2D-OI as follows:

$$G = C + \frac{1}{2}A_C \tag{2.8.1}$$

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

Snow depth analysis excludes grids with evergreen broadleaves as the tree vegetation type and those with no climatological potential for snow cover as estimated from past SSM/I observation statistics. The snow

depths of these grids are set to 0 cm. Grids covering land ice are also excluded from such analysis. The snow depths of these grids are set to C, the climatological values.

2.8.2 Mesoscale Snow Depth Analysis

In the Meso Scale Model (MSM, Section 3.5), snow cover data are used to gauge the status of ground snow, with depths exceeding 5 cm iin individual grid squares being classified as snow-covered. Snow depth analysis data are produced via a two-dimensional OI (2D-OI) in the high-resolution snow depth analysis system. The first guesses for 2D-OI are set using an offline version of the land surface model (LSM) with the same domain and grid spacing as the MSM. The offline LSM and the 2D-OI are outlined below.

The offline LSM, which includes a multi-layer snowpack model, simulates typical snow processes such as accumulation, compaction and ablation. The atmospheric forcing data necessary to drive the LSM are air temperature and wind velocity at the lowest atmospheric model level and radiative fluxes toward the surface as predicted by the MSM. Radar/Raingauge Analyzed Precipitation data (see Subsection 4.4.1) are used as rain and snowfall inputs to the LSM.

The model first guesses and observations reported from SYNOP and AMeDAS stations are handed over to the 2D-OI system in the snow depth analysis system. The methodology of OI is based on Brasnett (1999), where the correlation coefficient, μ_{kl}, is given by

$$\mu_{kl} = \alpha(r_{kl})\beta(\Delta z_{kl}) \tag{2.8.2}$$

with the horizontal and vertical separation, r_{kl} and Δz_{kl} between points k and l. $\alpha(r_{kl})$ and $\beta(\Delta z_{kl})$ are the horizontal and vertical structure functions:

$$\alpha(r_{kl}) = \left(1 + \frac{r_{kl}}{L}\right)\exp\left(-\frac{r_{kl}}{L}\right), \tag{2.8.3}$$

$$\beta(\Delta z_{kl}) = \exp\left\{-\left(\frac{\Delta z_{kl}}{h}\right)^2\right\}, \tag{2.8.4}$$

where L and h are set to 25 km and 500 m, respectively. The standard deviations of observation and background errors are set at 4 and 3 cm, respectively.

2.9 Non-real-time Quality Control

2.9.1 GDPFS-RSMC Operational Activities

JMA is designated as a Regional Specialized Meteorological Center (RSMC) of the World Meteorological Organization (WMO) Global Data-processing and Forecast System (GDPFS), and is known in this role as RSMC Tokyo. In March 1991, WMO Commission for Basic Systems (CBS) 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 six-monthly report containing a consolidated list of stations suspected of reporting low-quality observation data on station level pressure, mean sea level pressure and geopotential height during the six-month periods ending June and December. This report is available on JMA's website[8].

RSMC Tokyo also 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 on JMA's website[9].

Data quality evaluation is based on differences between observations and first guess fields (three to nine-hour forecasts) from the global model. Standard procedures and formats for the exchange of monitoring results are given in the Manual on GDPFS (WMO-No.485).

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

2.9.2 WDQMS Operational Activities

The WMO Integrated Global Observing System (WIGOS) is a framework for all WMO observing systems and WMO contributions to co-sponsored observing systems in support of all WMO Programmes and activities. At WIGOS workshops on Quality Monitoring and Incident Management held in December 2014 and December 2015, plans were developed for a WIGOS Data Quality Monitoring System (WDQMS). The Task Team on WDQMS (TT-WDQMS) is working to develop the WDQMS under the Inter-Commission Coordination Group on WIGOS. Four NWP centers (European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP), JMA and Deutscher Wetterdienst (DWD)) contribute to the NWP Quality Monitoring Pilot Project on WDQMS, providing monitoring output in near-real time to the WMO Secretariat. Contributions began with surface pressure data, and now also include information on surface humidity, wind, temperature and upper-air soundings.

2.9.3 Blacklist Management

As mentioned in Section 2.3, low quality observational data can result in significant forecast degradation. The cause of low quality may be instrumental failure, which can continue for a long time. Such observation data should be excluded in the first step of QC, and a blacklist is kept to meet this need. Blacklist management is one of the most important activities in QC. The quality of all observations is evaluated based on differences between observations and first guess fields from the global model (three to nine-hour forecasts), the meso-scale model (zero to three-hour forecasts) and the local forecast model (one-hour forecasts). Providers of problematic observation data are added to the blacklist.

2.10 Climate Data Assimilation System

For climate system monitoring and seasonal prediction, it is essential to comprehend both current and past climate conditions accurately and consistently over a long period of time. To this end, a high-quality, spatio-temporally homogeneous long-term climate dataset based on reanalysis of past observations using a state-of-the-art NWP technique is maintained.

JMA conducted its second global atmospheric reanalysis (the Japanese 55-year Reanalysis, or JRA-55; Kobayashi *et al.* 2015) covering the period from 1958, when regular radiosonde observations began on a global basis. Production is continued to the present on a near-real-time basis as an operational climate data assimilation system, thereby supporting real-time climate monitoring and seasonal prediction. JRA-55 data are also used in various research and development activities at JMA and in wider communities of fundamental academic fields such as meteorology, climatology and oceanography, practical application fields such as agricultural meteorology and renewable energy, and other areas.

The data assimilation system used to produce JRA-55 data is described in Kobayashi *et al.* (2015). The system is based on the TL319 version of JMA's operational data assimilation system as of December 2009 (JMA 2007, 2013), and features numerous improvements made since Japan's first global atmospheric reanalysis (the Japanese 25-year Reanalysis, or JRA-25; Onogi *et al.* 2007). These include a revised longwave radiation scheme, a 4D-Var data assimilation system and a variational bias correction scheme for satellite radiances. The system also incorporates several newly available observational datasets produced as a result of ongoing efforts to improve the quality of past observation data, including homogenization of radiosonde temperature observations (Haimberger *et al.* 2008, 2012) and reprocessing of satellite data at major meteorological satellite centers (e.g. van de Berg *et al.* 2002; Oyama 2010).

These improvements make JRA-55 products considerably superior to JRA-25 products. Two major issues with JRA-25 (a cold bias in the lower stratosphere and a dry bias in the Amazon basin) are mitigated in JRA-55, and the temporal consistency of temperature analysis is considerably better than in previous reanalysis products.