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 hybrid four-dimensional variational method using the Local Ensemble Transform Kalman Filter is adopted in Global Analysis, while the four-dimensional variational method is adopted in Meso-scale Analysis. To reduce the computational cost, the analysis increment is calculated using a coarser-resolution inner model in each 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 hybrid 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 5.2), snow depth fields (see Section 2.8) and soil moisture fields (see Section 2.9) are also analyzed every day.

The Japanese Reanalysis for Three Quarters of a Century project covers the period from September 1947 to the present day on a near-real-time basis (see Section 2.11).

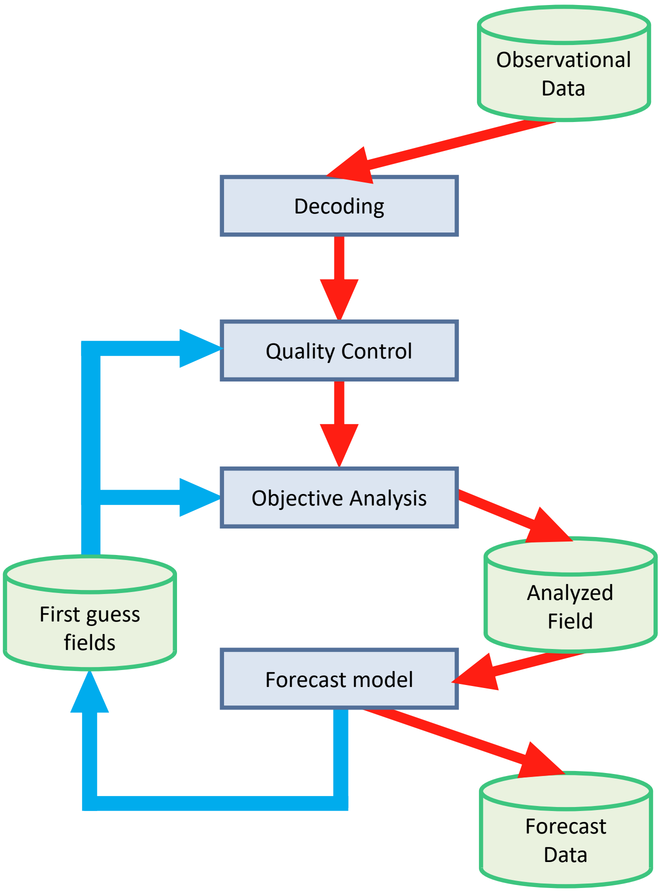

Figure 2.1.1: Major functional components and data flow in the JMA data assimilation system.

Table 2.1.1: Specifications of 4D-Var in Global Analysis (GA)

Analysis time	00, 06, 12 and 18 UTC
Analysis scheme	Incremental hybrid 4D-Var using LETKF
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 TQ959, Reduced Gaussian grid, roughly equivalent to 0.125° (13 km) [2880 (tropic) – 64 (polar)] × 1440
(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	128 forecast model levels up to 0.01 hPa + surface
Analysis variables	Wind, surface pressure, specific humidity and temperature
Observation (as of 31 March 2023)	SYNOP, METAR, SHIP, BUOY, TEMP, PILOT, Wind Profiler, AIREP, AMDAR, Typhoon Bogus; atmospheric motion vectors (AMVs) from Himawari-9, GOES-16, Meteosat-9 and Metop-[B, C]; polar AMVs from Aqua/MODIS, NOAA/AVHRR, Metop/AVHRR, Suomi-NPP/VIIRS and NOAA-20/VIIRS; LEO-GEO AMVs; ocean surface wind from Metop-[B, C]/ASCAT; radiances from NOAA-[15, 18, 19]/AMSU-A, Metop-[B, C]/AMSU-A, DMSP-F[17, 18]/SSMIS, Suomi-NPP/ATMS, NOAA-20/ATMS, GCOM-W/AMSR2, GPM-core/GMI, Metop-[B, C]/IASI, Suomi-NPP/CrIS, NOAA-20/CrIS, FY-3C/MWHS2, NOAA-19/MHS and Metop-[B, C]/MHS; clear sky radiances from the water vapor channels (WV-CSRs) of Himawari-9, GOES-16 and Meteosat-9; GNSS RO bending angle data from Metop-B/GRAS and TerraSAR-X/IGOR; zenith total delay data from ground-based GNSS
Assimilation window	6 hours

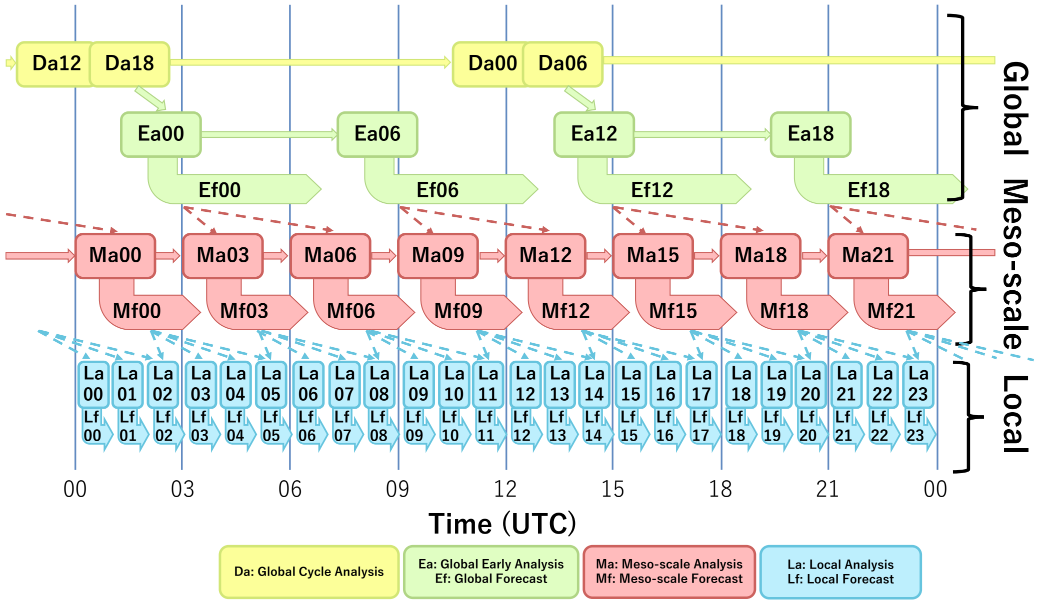

Figure 2.1.2: Main flow of JMA data assimilation systems. The first-guess and boundary conditions for Local Analysis are obtained from the latest MSM output.

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 ASUCA
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) 96 levels up to 37.5 km (Inner step) 48 levels up to 37.5 km
Analysis variables	Wind, potential temperature, surface pressure, pseudo-relative humidity, soil temperature and soil volumetric water content
Observations (as of 31 March 2023)	SYNOP, SHIP, BUOY, TEMP, PILOT, Wind Profiler, Weather Doppler radar (radial velocity, reflectivity), AIREP, AMDAR, Typhoon Bogus; AMVs from Himawari-9; ocean surface wind from Metop-[B, C]/ASCAT; radiances from NOAA-[15, 18, 19]/AMSU-A, NOAA-19/MHS, Metop-[B, C]/AMSU-A, Metop-[B, C]/MHS, DMSP-F[17, 18]/SSMIS, Suomi-NPP/ATMS, NOAA-20/ATMS, FY-3C/MWHS2, GCOM-W/AMSR2, GPM-core/GMI, Metop-[B, C]/IASI, Suomi-NPP/CrIS and NOAA-20/CrIS; clear sky radiances from the water vapor channels (WV-CSRs) of Himawari-9; 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-B/GRAS, TerraSAR-X/IGOR and TanDEM-X/IGOR; Total Precipitable Water Vapor from ground-based and shipborne 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 hybrid 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 21.8 km
Analysis variables	Wind, potential temperature, surface pressure, pseudo-relative humidity, skin temperature, ground temperature and soil moisture
Observations (as of 31 March 2023)	SYNOP, SHIP, BUOY, AMeDAS, TEMP, PILOT, Wind Profiler, Weather Doppler radar (radial velocity, reflectivity), AIREP, AMDAR; AMVs from Himawari-9; ocean surface wind from Metop-[B, C]/ASCAT; radiances from NOAA-[15, 18, 19]/AMSU-A, NOAA-19/MHS, Metop-[B, C]/AMSU-A, Metop-[B, C]/MHS, DMSP-F[17, 18]/SSMIS, Suomi-NPP/ATMS, NOAA-20/ATMS, FY-3C/MWHS-2, GCOM-W/AMSR2, GPM-core/GMI, Metop-[B, C]/IASI and NOAA-20/CrIS; clear sky radiances from the water vapor channels (WV-CSRs) of Himawari-9; soil moisture from GCOM-W/AMSR2 and Metop-B/ASCAT; Total Precipitable Water Vapor from ground-based and shipborne GNSS

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 31 March, 2023. 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 27,000 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 to observe near surface temperature and wind at about 930 stations in Japan at 10 minute intervals. JMA is proceeding with a plan for an upgrade of AMeDAS stations to attach humidity sensors up to FY 2026. Humidity observations started at 274 AMeDAS stations by FY 2022.

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 8,000 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 14,000 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 590 reports are produced at these times respectively.

2.2.2.7 PILOT

PILOT is a numerical code used for reporting upper-level wind observations performed by rawins or pilot balloons. Around 230, 210 and 70 reports are produced at 00, 12 and 06/18 UTC, respectively.

2.2.2.8 Aircraft

Aircraft observations are reported via Aircraft Report (AIREP) and Aircraft Meteorological Data Relay (AMDAR). While vertical profile data can be obtained at the vicinity of airports, only flight level data can be collected along the other airways. 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, Around 85,000 reports are produced every six hours covering areas around the world. Aircraft data availability is improving from the reduced volume caused by the COVID-19 pandemic that began around March 2020.

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, Sd: snow depth. 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, Sd	GA, MA, LA
AMeDAS	Land surface automated observation network in Japan	u, v, T, Rh	$MA(Rh), LA$
METAR	Routine weather reports from aerodromes	P, T, Rh	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 orbiting satellites	T_B	GA, MA, LA
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	$GA(ZTD)$, $MA, LA(Pwv)$
Shipborne GNSS	ZTD and PWV data estimated from shipborne GNSS receivers	Pwv	MA, LA
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, LA
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 microwave imager radiances 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 speed and direction 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 and water vapor patterns in successive satellite images. AMVs from geostationary (GEO) satellites (Meteosat-9, GOES-16 and Himawari-9), low-earth orbit (LEO) satellites (Aqua, NOAA and Metop), LEOGEO-AMVs and Dual-Metop 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. Dual-Metop AMVs are derived using tandem Metop-B and -C imagery, and have global coverage. AMVs from GEO satellites cover the area from 60°N to 60°S, and those from LEO satellites cover polar regions (i.e., latitudes higher than 60°).

2.2.2.11 Scatterometers

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

2.2.2.12 MW Sounders

Microwave (MW) sounder radiance data from AMSU-A (Advanced Microwave Sounding Unit-A) on NOAA-15, -18, -19, Metop-B and -C, MHS (Microwave Humidity Sounder) units on NOAA- 19, Metop-B and -C, ATMS (Advanced Technology Microwave Sounder) units on Suomi-NPP (the Suomi National Polar-orbiting Partnership) and NOAA-20, and MWHS-2 (Microwave Humidity Sounder 2) on FY-3C are used. AMSU-A is a sounder primarily monitoring atmospheric temperature profiles, while MHS and MWHS-2 are sounders primarily monitoring middle-to-upper tropospheric humidity profiles, and ATMS has temperature and humidity sounding channels. Related radiance data in meso-scale and local analysis are assimilated under clear-sky conditions (i.e., with no effect from cloud/rain). In global analysis, MHS, ATMS (humidity channels) and MWHS-2 radiances are assimilated under all-sky conditions, while other MW sounder radiances are assimilated under clear-sky conditions.

2.2.2.13 MW Imagers

Radiances from the following MW imagers are used; AMSR2 (Advanced Microwave Scanning Radiometer 2) on GCOM-W (Global Change Observation Mission 1st - Water), SSMIS (Special Sensor Microwave Imager Sounder) on DMSP (Defense Meteorological Satellite Program) -F17, -F18 and GMI (GPM (Global Precipitation Measurement mission) Microwave Imager) on GPM. Their radiances are sensitive to water vapor amounts in the lower troposphere. They are assimilated under all-sky conditions in global analysis, and assimilated under clear-sky conditions in meso-scale and local analysis. Precipitation amounts estimated from radiances using the MSC method (Takeuchi and Kurino 1997) are also used in meso-scale analysis. Radiances from AMSR2, SSMIS are used for snow-cover product as well (see Subsection 2.8.1).

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 GEO satellites (Meteosat-9, GOES-16 and Himawari-9) are used. CSR data of water vapor channels/bands which are sensitive to water vapor amounts in the upper and middle troposphere are used.

2.2.2.15 Hyperspectral IR Sounders

Radiances under clear-sky conditions from hyperspectral IR sounders are used. The data adopted are from IASI on Metop and CrIS on Suomi-NPP and NOAA-20.

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 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 TerraSAR-X and TanDEM-X, GRAS (GNSS Receiver for Atmospheric Sounding) onboard Metop-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 (Ishikawa 2010). GEONET is a ground-based GNSS receiver network operated by the Geospatial Information Authority in Japan using around 1,300 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 Shipborne GNSS

Shipborne GNSS data are provided every 10 minutes from two JMA observation vessels, four Japan Coast Guard vessels and ten merchant vessels (as of July 2023). Shipborne GNSS-derived precipitable water vapor (PWV) data (Shoji *et al.* 2017) are used in meso-scale and local analysis. One JMA vessel mission is to conduct targeted observations toward improved forecast accuracy for heavy rainfall in NWP during Japan's rainy season.

2.2.2.19 Radar Reflectivity

By March 2023, 10 of JMA's 20 C-band radars had been upgraded to dual-polarization types. With this system, three-dimensional reflectivity data are produced every 5 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 and used in meso-scale and local analysis. 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.

2.2.2.20 Radial Velocity of Doppler Radar

By March 2023, 10 of JMA's 20 Doppler Weather Radars (DWRs) and 8 of its 9 Doppler Radars for Airport Weather (DRAWs) had been upgraded to C-band dual-polarization types. With this system, three-dimensional radial velocity data are produced every 5 minutes within a radius of 150 km for DWRs (range resolution: 250 m) and 120 km for DRAWs (150 m). The azimuth resolution is 0.703° for both.

2.2.2.21 Analysis of Precipitation(R/A)

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.

2.2.2.22 Soil Moisture Content

Soil moisture content data estimated from microwave imager (AMSR2/GCOM-W) radiances and microwave scatterometer (ASCAT/Metop-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 measure prognostic variables in NWP such as pressure, temperature, wind and humidity. The QC system for these observations consists of internal and external QCs.

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 2.10). 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. Correction constants are determined from one-month statistics for the previous month. Kalman Filter is utilized to correct aircraft temperature biases in global analysis. Bias correction values are updated at each analysis time (00, 06, 12, 18 UTC) and calculated independent of analysis using a covariance of the first-guess departure separately for each aircraft, each vertical level and each flight phase (ascent, cruising and descent). In mesoscale analysis, aircraft temperature biases are corrected for each aircraft and vertical level using bias correction values statistically calculated using the first-guess departure from global analysis for the previous month.

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.

2.3.2 AMVs

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. Thinning is then performed based on a distance of 200 km. Around Japan region from 120°E to 150°E and from 20°N to 45°N, super-observation (SPOB) data are derived from Himawari-9 AMVs to arrange at 100 km intervals. (Yamashita 2016). Climatological checking (see Subsection 2.3.1.1) and external QC (see Subsection 2.3.1.2) are then performed. The details of QC and detailed settings for AMVs are given on the NWP SAF AMV monitoring page[1].

2.3.3 Scatterometers

Level 2 ocean surface wind data are used in global, meso-scale and local analysis, with initial rejection of low-quality data for land/sea ice areas. The most likely wind directions are then identified from inherent ambiguity scatterometer data with application of NWP nudging and median filtering. To avoid undesirable rejection of correct wind data in subsequent gross error checking (see Subsection 2.3.1.2) for areas in and around severe weather systems such as cyclones and fronts where wind direction/speed vary sharply, quality control (Group-QC) is applied. In this step, spatial consistency among wind vectors is checked in relation to smooth transition in wind direction/speed, and valid data are excluded from rejection (see the NWP SAF scatterometer monitoring page for details[2].

2.3.4 Satellite Radiance

Satellite radiance data are used in global, meso-scale and local analysis as a form of brightness temperature. The RTTOV-13.0 fast radiative transfer model (Saunders *et al.* 2020) is employed for radiance assimilation[3]. 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 2.10). 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, channel selection and gross error checking (see Subsection 2.3.1.2). Data passing this QC

---

[1] `https://nwp-saf.eumetsat.int/site/amv-usage-in-the-jma-nwp-model/`

[2] `https://nwp-saf.eumetsat.int/site/monitoring/winds-quality-evaluation/scatterometer-mon/use-in-nwp/jmamodel/`

[3] Coefficients of RTTOV-10.2 are still used.

Table 2.3.1: AMV blacklist summary. 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 other than Himawari-9; LEOGEO AMV: AMVs from a combination of low earth orbiting and geostationary satellites.; Dual-Metop AMV: AMVs from a combination of Metop-B and C.

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, and between 400 hPa and 825 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
Himawari-9 (IR/VIS)	below 700 hPa over land
Meteosat-9 (IR/VIS)	below 700 hPa over land
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
Dual-Metop AMV	between 50°N and 50°S, and at poleward of 88°N or 88°S
Dual-Metop AMV	above 300 hPa, and below 700 hPa over land

Table 2.3.2: Summary of microwave sounder channel sets used for each condition. In meso-scale and local analysis, AMSU-A Ch. are used up to 11 and 8, respectively.

	AMSU-A	MHS	ATMS
Clear sky ocean	Ch. 4–14	Ch. 3–5	Ch. 6–9
Clear sky land/coast/sea-ice	Ch. 6–14	Ch. 3–5	Ch. 7–9
Cloudy ocean	Ch. 7–14	Ch. 3–5	Ch. 8–9
Rainy ocean	Ch. 9–14	n/a	n/a

are thinned again to reduce the observation error correlation, and the thinned data are output for use in data assimilation systems. 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. The satellite radiance data used in the analyses are from MW sounders, MW imagers, CSR and hyperspectral IR sounders. The specific procedures for each data type are described in the following subsections.

2.3.4.1 MW sounders

For radiances assimilated under clear-sky conditions (not affected by cloud/rain), the sets of channels used are defined in advance according to individual surface and atmospheric conditions at the observation point (Table 2.3.2). In global analysis, microwave humidity sounders' radiances (183 GHz) are assimilated under all-sky conditions by assigning observation errors depending on cloud amount (see Subsection 2.5.7.2).

2.3.4.2 MW imagers

Vertically polarized-channel radiances are assimilated over ice-free ocean areas. Radiances are assimilated under all-sky conditions in global analysis (see Subsection 2.5.7.2), and assimilated under clear-sky conditions in meso-scale and local analysis. 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 (or temporally thinned to the hourly slots) 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-13.0, 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, meso-scale and local analysis. Data are horizontally thinned to divisions of 200 km for global analysis and 45 km for meso-scale and local analysis. Cloud screening with estimation of cloud top height are applied in quality control. Usage is based on CO2 slicing (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 higher than 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 from first guess are regarded as suspect. If the absolute difference of departure D of suspect data from averaged D of vicinity 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. The criterion C_P, C_R and C_S referred in Subsection 2.3.1.2 are set to 5 mm, 8 mm and 5 mm respectively for GNSS-PWV. 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.

2.3.7 Shipborne-GNSS

Quality control and data thinning for shipborne GNSS-PWV data are almost the same as for ground-based GNSS-PWV data (which have been used in mesoscale analysis since 2009 as per Subsections 2.2.2.17 and 2.3.6), except for the use of Dynamic QC (see Subsection 2.3.1.2) in gross error checking for shipborne GNSS-PWV. Quality parameters, the number of satellites and analysis times (reported only by shipborne GNSS) are also checked.

2.3.8 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 2.3.10). 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.

2.3.9 Radial Velocity of Doppler Radar

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 (< 10 km) 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.

2.3.10 R/A

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

2.3.11 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 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.

2.3.12 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) format, 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 format is 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 generated bogus 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 hybrid four-dimensional variational (4D-Var) data assimilation method using a Local Ensemble Transform Kalman Filter (LETKF, Hunt *et al.* 2007) 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. An incremental method (Courtier *et al.* 1994) is adopted to improve computational efficiency using outer-loop iterations, with increments evaluated at a lower (inner) resolution (TL319L128: grid roughly equivalent to 0.5625° (55 km) and up to 0.01 hPa). The increment is then interpolated and used to update the model trajectory at the original resolution (TQ959L128: grid roughly equivalent to 0.125° (13 km)), and the updated trajectory is used to refine the cost function for subsequent inner-loop iterations. 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), TL319L100 in March 2014. The hybrid 4D-Var data assimilation system using the LETKF with two outer-loop iterations was operated in December 2019 (Kadowaki *et al.* 2020). In March 2021, the number of ensemble members in LETKF was increased from 50 to 100, weighting for ensemble background error covariances was increased from 0.15 to 0.5 (Yokota *et al.* 2021) and the inner resolution was upgraded to TL319L128.

2.5.2 Incremental 4D-Var Formulation and Procedural Description

In 4D-Var, 3–9-hour forecasts from the GSM are used as a first guess (background), and all observations passed Quality Control (QC) from within three hours of analysis time (6-hour) are organized in hourly time slots as defined in Figure 2.5.1. 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 at a low-resolution, and determined by minimization of the cost function in the inner-loop iterations. The low-resolution analysis increment is interpolated to the high-resolution analysis increment. Adding this analysis increment to the guess field produces high-resolution analysis.

In the outer-loop iterations, procedures mentioned above are iterated as in Figure 2.5.2 :

1. The GSM forecast from “latest” high-resolution analysis is performed to update the guess field (or model trajectory).
2. QC processes of observations are reperformed.
3. The cost function is updated.
4. The low-resolution analysis increment is updated by minimization of the cost function in the inner-loop iterations and used to update the high-resolution analysis.

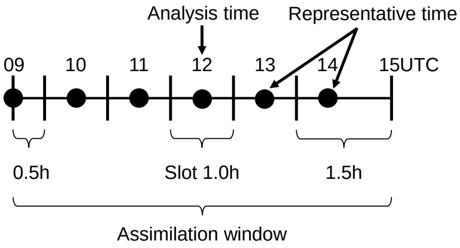

Figure 2.5.1: Schematic diagram of time slots for the analysis time 12 UTC. The black circles indicate the representative time of each time slot.

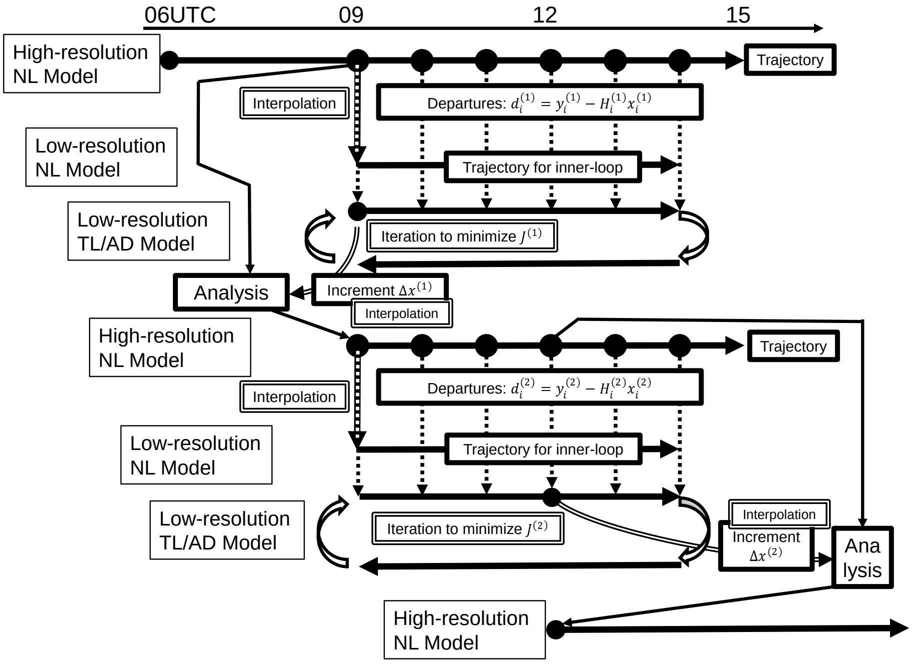

Figure 2.5.2: Flow of 4D-Var operation for the 12 UTC analysis time

In the $j(\geqq 1)$-th outer-loop, the cost function $J^{(j)}$ to determine the j-th low-resolution analysis increment update $\Delta x_i^{(j)}$ is defined by Eq. (2.5.1).

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

where the subscript $i(\geqq 1)$ is the index of time slot and n is the number of time slots. $\Delta x_0^{(j)}$ ($\Delta x_1^{(j)}$) is the low-resolution analysis increment at the initial time before (after) initialization, $\Delta x_{i\geqq 2}^{(j)}$ is the analysis increment evolved according to the tangent linear (TL) model to time of slot i. $\mathbf{M}_0^{(j)}$ is the TL operator of the nonlinear normal-mode initialization operator (Machenhauer 1977) and $\mathbf{M}_{i\geqq 1}^{(j)}$ is the TL model of the low-resolution nonlinear (NL) forecast model $M_i^{(j)}$ as detailed in Subsection 2.5.3 for integration from time of slot i to that of slot $i+1$. $\mathbf{R}_i^{(j)}$ 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 2.5.5 and Subsection 2.5.7. $\mathbf{H}_i^{(j)}$ is the TL operator of the observation operator $H_i^{(j)}$. The innovation vector is given for each assimilation slot by $d_i^{(j)} = y_i^{(j)} - H_i x_i^{(j)}$, where $x_i^{(j)}$ is the j-th guess field evolved using the high-resolution NL model, and $y_i^{(j)}$ is the observation data passed QC at time of slot i. $J_C^{(j)}$ is a penalty term used to suppress gravity waves described in Subsection 2.5.4.

To minimize the cost function $J^{(j)}$, 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.

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

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

$$p_0 = \mathbf{M}_0^{(j)\mathrm{T}} p_1 + \mathbf{B}^{-1}\Big(\sum_{l=1}^{j}\Delta x_0^{(l)}\Big) \tag{2.5.4}$$

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

where p_i is a auxiliary variable, $\mathbf{M}_i^{(j)\mathrm{T}}$ represents the adjoint (AD) model of the TL model $\mathbf{M}_i^{(j)}$, and $\mathbf{H}_i^{(j)\mathrm{T}}$ is the AD operator of $\mathbf{H}_i^{(j)}$. 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.

2.5.3 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. In particular, 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.

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.

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 ≈ 10 km) based on the inner NL model which is almost same as in GSM1705 (JMA 2018). 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.

2.5.4 Penalty Term

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

$$J_C^{(j)} = \frac{1}{2}\alpha \left(\left| \mathrm{N}_G \sum_{l=1}^{j} \Delta x_0^{(l)} \right|^2 + \sum_{i=2}^{\mathrm{n}} \left| \mathrm{N}_G \sum_{l=1}^{j} \Delta x_i^{(l)} \right|^2 \right) \tag{2.5.7}$$

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, it is also effective in stabilizing calculation.

2.5.5 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. In the 4D-Var, the evolution of the initial background error covariances over the length of the assimilation window is taken into account, although the initial background error covariances are climatological and do not represent the day-to-day weather situation. One way to take into account the error of the day is to use ensemble forecast as part of background error covariances. In JMA, the hybrid method composed of the LETKF and the 4D-Var is employed as the operational system. Figure 2.5.3 shows a schematic diagram of the hybrid 4D-Var using the LETKF.

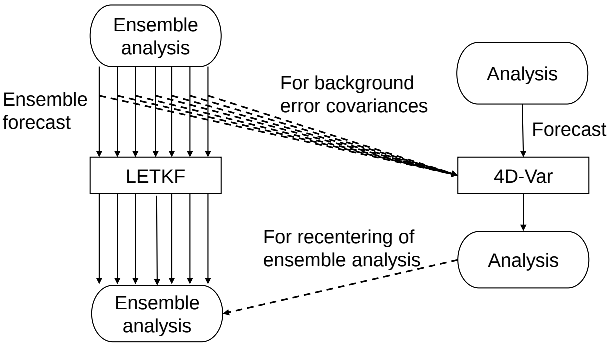

Figure 2.5.3: A schematic diagram of the hybrid 4D-Var using the LETKF.

Using the extended control variable method of Lorenc (2003), the background term in the hybrid 4D-Var is defined by Eq. (2.5.8).

$$\begin{aligned} \frac{1}{2}\Delta x^{\mathrm{T}}\mathbf{B}^{-1}\Delta x &= \frac{1}{2}\Delta x_{cl}^{\mathrm{T}}\mathbf{B}_{cl}^{-1}\Delta x_{cl} + \frac{1}{2}\Delta x_{en}^{\mathrm{T}}\mathbf{B}_{en}^{-1}\Delta x_{en} \\ \Delta x &= \beta_{cl}\Delta x_{cl} + \beta_{en}\Delta x_{en} \end{aligned} \tag{2.5.8}$$

where the subscript i and j in Eq. (2.5.1) are omitted here for simplicity. Δx is the low-resolution analysis increment. The subscript "*cl*" and "*en*" mean "climatological" and "ensemble". β_{cl} and β_{en} are the weight of "climatological" and "ensemble" background error covariances for making hybrid covariances. In JMA, both ${\beta_{cl}}^2$ and ${\beta_{en}}^2$ are 0.5 under 50 hPa, and they are approaching to 1 and 0 above that.

Two kinds of covariance inflations are applied for the ensemble background error covariances. One is an adaptive multiplicative covariance inflation used in the LETKF described in Subsection 2.5.6. The other is an additional covariance inflation to make vertical profiles of horizontal global mean of standard deviation from the ensemble covariances, consistent with those from the climatological background error covariances. A covariance localization is also applied for the ensemble background error covariances. The localization function is given as the Gaussian function of the distance between the analysis grid points. The localization scale l for which the localization function is $1/\sqrt{e}$ is set to 800 km for wind vectors, temperature and surface pressure and set to 400 km for specific humidity in the horizontal domain. The scale l is set to a 0.8 scale height in the vertical domain for all variables. A detailed description of the climatological background error covariances is provided in Subsection 2.5.5.1 - Subsection 2.5.5.5. A detailed description of the LETKF is provided in Subsection 2.5.6.

2.5.5.1 Climatological Background Error Covariance

For the climatological background error covariances, 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 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

$$\Delta\eta_U \equiv \Delta\eta - P\Delta\phi_B \qquad (\Delta\phi_B = \Delta\phi_B(\Delta\zeta)) \tag{2.5.9}$$

$$\begin{pmatrix}\Delta T_U \\ \Delta p_{sU}\end{pmatrix} \equiv \begin{pmatrix}\Delta T \\ \Delta p_s\end{pmatrix} - Q\Delta\phi_B - R\Delta\eta_U \tag{2.5.10}$$

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 2.5.5.2. 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.

2.5.5.2 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.

$$\Delta\tilde{\phi}_B = L\Delta\zeta = UWV^{\mathrm{T}}\Delta\zeta \tag{2.5.11}$$

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

$$\begin{aligned} \Delta\tilde{\phi}_{B n}^{\,m} &= c_n^m \Delta\zeta_{n-1}^m + c_{n+1}^m \Delta\zeta_{n+1}^m \qquad ((n,\, m) \neq (0,\, 0),\ n = m,\ m+1, \ldots,\ N), \\ c_n^m &= -\frac{2\Omega a^2}{n^2}\sqrt{\frac{n^2-m^2}{4n^2-1}}, \qquad c_{N+1}^m = 0, \qquad \Delta\tilde{\phi}_{B0}^{\,0} = 0 \end{aligned} \tag{2.5.12}$$

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 $\Delta\Phi$ [4] (as derived from temperature and surface pressure) and balance mass variable is calculated as follows

$$D_j = \frac{\left\langle \left(U^{\mathrm{T}}\Delta\Phi\right)_j^m \left(U^{\mathrm{T}}\Delta\tilde{\phi}_B\right)_j^m \right\rangle}{\left\langle \left[\left(U^{\mathrm{T}}\Delta\tilde{\phi}_B\right)_j^m\right]^2 \right\rangle} \tag{2.5.13}$$

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:

$$\Delta\phi_B = UDU^{\mathrm{T}}\Delta\tilde{\phi}_B = UDWV^{\mathrm{T}}\Delta\zeta = \tilde{L}\Delta\zeta \tag{2.5.14}$$

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.

---

[4] 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 calculation of $\Delta\phi_k$, $\bar{T}_k$ and $\bar{p}$ are also used and some approximation is done.

2.5.5.3 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:

$$P_n = \left\langle \Delta\eta_n^m \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle \left\langle \Delta\phi_{B_n}^{\ m} \left(\Delta\phi_{B_n}^{\ m}\right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.15}$$

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

$$R_n = \left\langle \left[\begin{pmatrix} \Delta T_n^m \\ \Delta p_{sn}^{\ m} \end{pmatrix} - Q_n \Delta\phi_{B_n}^{\ m} \right] \left(\Delta\eta_{U_n}^{\ m}\right)^{\mathrm{T}} \right\rangle \left\langle \Delta\eta_{U_n}^{\ m} \left(\Delta\eta_{U_n}^{\ m}\right)^{\mathrm{T}} \right\rangle^{-1} \tag{2.5.17}$$

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

2.5.5.4 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 $(\Delta T_U,\ \Delta P_{sU})$.

$$B_{\zeta n} = \left\langle \Delta\zeta_n^m \overline{\Delta\zeta_n^m}^{\mathrm{T}} \right\rangle, \qquad B_{\eta_U n} = \left\langle \Delta\eta_{U_n}^{\ m} \overline{\Delta\eta_{U_n}^{\ m}}^{\mathrm{T}} \right\rangle \tag{2.5.18}$$

$$B_{\begin{pmatrix} T_U \\ p_{sU} \end{pmatrix}_n} = \left\langle \begin{pmatrix} \Delta T_{U_n}^{\ m} \\ \Delta p_{sU_n}^{\ m} \end{pmatrix} \overline{\begin{pmatrix} \Delta T_{U_n}^{\ m} \\ \Delta p_{sU_n}^{\ m} \end{pmatrix}}^{\mathrm{T}} \right\rangle, \qquad B_{\ln q\, n} = \left\langle \Delta \ln q_n^m \overline{\Delta \ln q_n^m}^{\mathrm{T}} \right\rangle \tag{2.5.19}$$

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.

2.5.5.5 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:

$$J_n^{(x)} = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta x_n^m}\right)^{\mathrm{T}} B_n^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta x_n^m}\right)^{\mathrm{T}} \left(L_n L_n^{\mathrm{T}}\right)^{-1} \Delta x_n^m = \sum_{m=-n}^{n} \frac{1}{2} \left(\overline{\Delta y_n^m}\right)^{\mathrm{T}} \Delta y_n^m \tag{2.5.20}$$

$$\Delta y_n^m \equiv L_n^{-1} \Delta x_n^m \tag{2.5.21}$$

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.

2.5.6 LETKF

The specifications of the LETKF approach used in the GA to provide flow-dependent background error covariances are listed in Table 2.5.1.

Observation datasets assimilated in the LETKF are the same as those of GA (Table 2.1.1) except for those of hyperspectral sounders (AIRS, IASI and CrIS).

Table 2.5.1: Specifications of the LETKF used in GA

Data cut-off time	same as in the 4D-Var
First guess	3–9-hour forecast of its own
Horizontal resolution	Same as for the Inner step of the 4D-Var
Vertical resolution (model top)	Same as for the 4D-Var
Ensemble size	100 members
Analysis variables	Same as in the 4D-Var
Observation	Same as in the 4D-Var except for AIRS, IASI and CrIS
Assimilation window	6 hours (± 3 hours of analysis time)
Covariance inflation	Adaptive multiplicative covariance inflation
Initialization	Horizontal divergence adjustment based on the analysis of surface pressure tendency (Hamrud *et al.* 2015)
Localization	Gaussian function. The localization scale for which the localization function is $1/\sqrt{e}$ is set to 400 km in the horizontal domain (300 km for humidity-sensitive observations), a 0.6 scale height in the vertical domain (1.2 for surface pressure and ground-based GNSS zenith total-delay observations) and three hours in the temporal domain. For satellite radiance observations, the maximum of the square of the weighting function divided by its peak value and the Gaussian function with a $0.6\sqrt{2}$ scale height centered at the peak of the weighting function is used as the vertical localization function.
Model ensemble method	Stochastic physics scheme
Other characteristics	A total of 100 analyses are re-centered so that their ensemble mean is consistent with the hybrid 4D-Var.

Observation localization is applied in the LETKF. Observation errors are multiplied by the inverse of the localization function to give less weight to data collected farther from the analysis grid point. The localization function is given as the Gaussian function of the distance between the analysis grid point and the observation location. The localization scale l for which the localization function is $1/\sqrt{e}$ is set to 400 km in the horizontal domain (300 km for humidity-sensitive observations), a 0.6 scale height in the vertical domain (1.2 for surface pressure and ground-based GNSS zenith total-delay observations) and 3 hours in the temporal domain. The tail of the localization function is set to 0 farther than $2\sqrt{10/3}l$. For satellite radiance observations, the maximum of a square of the weighting function divided by its peak value and the Gaussian function with a length scale of $0.6\sqrt{2}$ scale height centered at the peak of the weighting function is used as the vertical localization function.

Multiplicative covariance inflation (Anderson 2001) is applied to first-guess (6-hour forecast from the previous analysis) ensembles. Inflation coefficients are estimated for each analysis grid point so that the following relation (Desroziers *et al.* 2005) is observed using locally assimilated observations:

$$tr\left[\mathbf{d}_{A-B}\mathbf{d^T}_{O-B}\right] = \alpha tr\left[\mathbf{HBH^T}\right] \tag{2.5.22}$$

where $\mathbf{d}_{A-B}$, $\mathbf{d}_{O-B}$, $\mathbf{H}$ and $\mathbf{B}$ represent the analysis increment projected onto the observation space, innovations, observation operator and background error covariance, respectively. $tr\,[]$ represents the trace of the matrix, and α represents the inflation coefficient. Based on (2.5.22), the raw inflation coefficient is estimated on analysis grid point j as

$$\alpha_{j,raw} = \frac{\sum_{m=1}^{N_{obs}} \rho_{m,j}\left(d_{A-B,m}d_{O-B,m}/\sigma^2_{o,m}\right)}{\sum_{m=1}^{N_{obs}}\left[\rho^2_{m,j}\left(\mathbf{H}_m\mathbf{X}_b\right)\left(\mathbf{H}_m\mathbf{X}_b\right)^{\mathbf{T}}/\left(K-1\right)/\sigma^2_{o,m}\right]} \tag{2.5.23}$$

where N_{obs} is the number of observations assimilated locally, K is the ensemble size, $\rho_{m,j}$ is the localization function of the mth observation and $\mathbf{H}_m\mathbf{X}_b$ is the first-guess perturbation projected onto the mth observation. The subscript *raw* represents a raw estimate. $d_{A-B,m}$ and $d_{O-B,m}$ are the observational increment and innovation of the mth observation, where $d_{A-B,m}$ is computed with a transformation matrix derived from LETKF analysis on grid j. Temporal smoothing is applied to the estimated inflation coefficient as

$$\alpha_{i,j} = \frac{\alpha_{i-1,j}\sigma_{o,j}^2 + \alpha_{i,j,raw}\sigma_b^2}{\sigma_{o,j}^2 + \sigma_b^2} \tag{2.5.24}$$

where the subscript i represents the value at the ith analysis step, and $\sigma_{o,j}^2$ and σ_b^2 are error variances of the estimated and prior coefficients and are set as

$$\sigma_{o,j}^2 = 1/\sum_{m=1}^{N_{obs}} \rho_{m,j} \tag{2.5.25}$$

$$\sigma_b^2 = 0.005 \tag{2.5.26}$$

Finally, the estimated coefficients are relaxed to the default values as the deviation from these defaults reaches $1/e$ at 10 days. The defaults are set to 1.21 from the surface to 45 hPa and linearly reduced with the logarithm of pressure to 1.0 at 0.85 hPa. Inflation to specific humidity is reduced by 30% of the estimated inflation coefficients below 230 hPa and linearly decreases with the logarithm of pressure to about 15% at 45 hPa.

Initialization based on analysis of surface pressure tendency (Hamrud *et al.* 2015) is applied after the LETKF analysis update. The surface pressure tendency of the first-guess ensembles $(\partial p_s/\partial t)_{guess}$ is diagnosed from the continuity equation and the hydrostatic balance as

$$\left(\frac{\partial p_s}{\partial t}\right)_{guess} = -\int_0^1 \nabla \cdot \left(\boldsymbol{v}_{guess}\frac{\partial p_{guess}}{\partial \eta}\right) d\eta \tag{2.5.27}$$

where p_s, $\boldsymbol{v}$, p and η are surface pressure, horizontal wind, pressure and the model vertical coordinates (1 at the bottom and 0 at the top), respectively. Surface pressure tendency is analyzed by adding the above variable to the first-guess state variables of the LETKF. The difference in surface pressure tendency diagnosed from horizontal wind and surface pressure analysis and that observed from LETKF analysis is distributed to each model layer so that the value is proportional to the global mean of the analysis spread of horizontal wind. Horizontal wind is adjusted using

$$\Delta\left[\nabla \cdot (\boldsymbol{v}_k dp_k)\right] = w_k \left[\left(\frac{\partial p_s}{\partial t}\right)_{diag} - \left(\frac{\partial p_s}{\partial t}\right)_{anl}\right] \tag{2.5.28}$$

where k is an index of the vertical model layer, w_k is the weight on the kth model layer and dp_k is the difference in half-level pressure adjacent to the kth full-level model layer. The horizontal wind increment is derived from (2.5.28) assuming $\Delta(dp_k) = 0$ and $\Delta(\nabla \times (\boldsymbol{v}_k dp_k)) = 0$.

The analysis ensemble is recentered so that the ensemble mean is consistent with global analysis.

The stochastic physics scheme (Subsection 3.3.4) is also applied to the forecast ensemble of the LETKF.

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. Surface pressure data at the station height or sea level pressure data reported from surface observation station are assimilated after conversion to the pressure at model surface height. While scatterometer data are reported as winds at 10 m above sea level, assimilated as wind at the lowest model 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-13.0 (Saunders *et al.* 2020). 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 2.5.2 and Table 2.5.3. For ocean surface wind data from scatterometers, such errors are defined as those with values of 4 or 6 m/s. For GNSS-RO 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.

For satellite radiance data of MW imagers and sounders assimilated under all-sky conditions, observation errors are given by an observation error model. Observation errors are assigned as a function of the symmetric (average of observation and first guess (FG)) cloud amount to handle the non-Gaussian distribution of observation errors. The symmetric cloud amount is based on Geer and Bauer (2011) using horizontal and vertical polarization difference at 37GHz for MW imagers, and Geer *et al.* (2014) using difference of scattering effect between 90GHz and 150GHz channels for MW sounders. The parameters of the observation error model are derived based on FG departure statistics as listed in Table 2.5.4. Observation error model for 183GHz channels is separately prepared for each surface types (ocean, land and sea ice). Only parameters for ocean are shown.

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-13.0 is used as the observation operator for satellite radiance data assimilation. ROPP8 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.

$$J(\Delta z_0) = \frac{1}{2}\Delta x_0^{\mathrm{T}}\mathbf{B}^{-1}\Delta x_0 + \frac{1}{2}\Delta\beta^{\mathrm{T}}\mathbf{B}_\beta{}^{-1}\Delta\beta + \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{H}_i\Delta x_i + \sum_{j=0}^{m}\Delta\beta_j p_{i,j} - d_i\right)^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\mathbf{H}_i\Delta x_i + \sum_{j=0}^{m}\Delta\beta_j p_{i,j} - d_i\right) + J_C \tag{2.5.29}$$

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.

Table 2.5.2: Observation error tables used in operational global analysis for (a) conventional observation, (b) AMV, (c) AMSU-A, (d) ATMS and (e) 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) 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	Metop-B	Metop-C	NOAA-15	NOAA-18	NOAA-19
4	x	0.45	0.45	0.45	0.45
5	x	0.3	0.3	x	0.3
6	0.3	0.3	x	0.3	0.3
7	0.3	0.3	0.3	0.3	0.3
8	0.3	0.3	0.3	x	x
9	0.3	0.3	0.3	x	0.3
10	0.3	0.3	0.45	0.3	0.3
11	0.45	0.45	x	0.45	0.45
12	0.64	0.64	0.64	0.64	0.64
13	1.02	1.02	0.85	1.02	1.02
14	2.83	2.83	x	2.63	2.63

(d) ATMS T_B (K)

Satellite Channel	Suomi-NPP	NOAA-20
6	0.3	0.3
7	0.3	0.3
8	0.3	0.3
9	0.3	0.3

(e) CSR T_B (K)

Satellite band (μm)	GOES-16	Satellite band (μm)	Meteosat-9	Satellite band (μm)	Meteosat-11	Satellite band (μm)	Himawari-9
8 (6.15)	1.5	5 (6.25)	1.5	5 (6.25)	1.5	8 (6.24)	1.5
9 (7.00)	1.5					9 (6.94)	1.5
10 (7.40)	1.5	6 (7.35)	1.5	6 (7.35)	1.5	10 (7.35)	1.5

Table 2.5.3: Observation error tables used in operational global analysis for hyperspectral IR sounders. (a) IASI and (b) CrIS.

(a) IASI T_B (K)					
Ch.	Metop-B	Metop-C	Ch.	Metop-B	Metop-C
38	1.26	1.08	249	0.69	0.75
49	1.20	1.08	252	0.72	0.75
51	1.23	1.05	256	0.72	0.72
55	1.20	1.05	262	0.69	0.72
57	1.23	1.02	265	0.78	0.72
61	1.20	0.96	267	0.69	0.63
63	1.14	0.99	269	0.72	0.78
70	1.11	0.96	271	0.66	0.69
85	1.11	0.93	273	0.63	0.66
87	1.08	0.99	275	0.72	0.75
109	0.96	0.84	278	0.72	0.72
116	0.87	0.87	282	0.72	0.72
122	0.87	0.87	284	0.72	0.69
128	0.87	0.84	288	0.69	0.72
135	0.78	0.81	292	0.66	0.66
141	0.78	0.87	294	0.75	0.78
146	0.72	0.75	296	0.78	0.81
148	0.75	0.78	306	0.90	0.87
154	0.75	0.78	308	0.69	0.72
159	0.72	0.72	316	0.69	0.72
161	0.72	0.75	320	0.72	0.72
167	0.72	0.75	323	0.69	0.66
173	0.72	0.81	327	0.72	0.75
180	0.72	0.84	333	0.72	0.75
185	0.75	0.81	347	0.72	0.75
187	0.72	0.81	350	0.69	0.72
193	0.81	0.90	354	0.72	0.69
205	0.90	1.02	356	0.69	0.69
212	0.87	0.78	360	0.69	0.66
217	0.81	0.87	666	1.62	1.62
219	0.87	0.81	1110	1.65	1.65
224	0.78	0.84	2951	3.51	3.51
226	0.81	0.84	2993	4.32	4.32
230	0.75	0.78	3207	4.32	4.32
232	0.81	0.84	5130	3.03	3.03
236	0.69	0.69	5315	1.74	1.74
239	0.75	0.81	5397	3.09	3.09
243	0.72	0.84	5492	3.51	3.51
246	0.72	0.78			

(b) CrIS T_B (K)		
Ch.	Suomi-NPP	NOAA-20
27	0.81	0.75
31	1.35	1.32
37	0.54	0.51
51	0.39	0.39
69	0.45	0.36
73	0.45	0.36
75	0.51	0.42
79	0.42	0.36
80	0.45	0.39
81	0.39	0.36
83	0.39	0.36
85	0.39	0.36
87	0.39	0.39
88	0.39	0.39
93	0.42	0.42
96	0.39	0.39
102	0.48	0.48
106	0.45	0.45
116	0.54	0.54
123	0.63	0.63
124	0.54	0.54
125	0.57	0.60
132	0.54	0.54
136	0.51	0.54
138	0.54	0.57
142	0.51	0.54
148	0.48	0.48
224	-	1.65
248	-	1.56
264	-	1.86
279	-	1.50
718	-	2.28
978	-	3.81
988	-	3.33
1003	-	3.93
1032	-	3.75

Table 2.5.4: Observation error (S_{clr}, S_{cld}) and threshold of symmetric cloud amount for identifying clear-sky and cloudy samples (C_{clr}, C_{cld}) for assimilated channel under all-sky conditions.

Satellite/Sensor	Ch.	Frequency(GHz) and polarization	S_{clr}	S_{cld}	C_{clr}	C_{cld}
GCOM-W/AMSR2	7	19V	4.8	43.8	0.0	0.6
	9	23V	6.9	22.8	0.0	0.5
	11	37V	6.0	34.2	0.0	0.5
DMSP-F17/SSMIS	9	183 ± 6.6	9.4	82.2	0.0	40.0
	10	183 ± 3.0	8.0	56.1	0.0	40.0
	11	183 ± 1.0	7.9	29.2	0.0	40.0
DMSP-F17,18/SSMIS	13	19V	5.4	40.8	0.0	0.6
	14	23V	7.2	19.8	0.0	0.6
	16	37V	6.9	35.1	0.0	0.5
GPM-core/GMI	3	19V	5.1	47.1	0.0	0.6
	5	23V	6.9	26.1	0.0	0.5
	6	37V	6.6	36.6	0.0	0.5
	12	183 ± 3	6.8	75.8	0.0	40.0
	13	183 ± 7	7.1	105.0	0.0	30.0
NOAA-19,Metop-B,C/MHS	3	183 ± 1	6.1	44.0	0.0	45.0
	4	183 ± 3	5.4	71.4	0.0	45.0
	5	190V	4.9	104.1	0.0	45.0
FY-3C/MWHS-2	11	183 ± 1.0	6.9	50.1	0.0	50.0
	12	183 ± 1.8	6.6	71.0	0.0	50.0
	13	183 ± 3.0	6.5	91.6	0.0	50.0
	14	183 ± 4.5	6.9	110.0	0.0	50.0
	15	183 ± 7.0	6.6	128.2	0.0	50.0
Suomi-NPP,NOAA-20/ATMS	18	183 ± 7.0	8.2	117.0	0.0	50.0
	19	183 ± 4.5	8.1	99.9	0.0	50.0
	20	183 ± 3.0	7.7	81.1	0.0	50.0
	21	183 ± 1.8	7.3	43.6	0.0	50.0
	22	183 ± 1.0	7.4	43.6	0.0	50.0

The second term on the right hand side of Eq. (2.5.29) 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 optimal 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 Data Assimilation (JNoVA; Honda *et al.* 2005)) in April 2009. 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). A further upgrade of the MSM forecast model in February 2017 replaced 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. In March 2020, a 4D-Var system based on ASUCA (ASUCA-Var; Ikuta *et al.* 2021) was introduced (Ikuta *et al.* 2020). This development enabled MA to produce initial conditions more consistent with the upgraded MSM forecast model. In March 2022, the model top was upgraded from 21,801 m to 37,500 m, with the number of vertical layers increased from 76 to 96 in the outer model and from 38 to 48 in the inner model.

Various observational data are used to improve the accuracy of prediction for meso-scale weather events, including information from weather radars, satellite observations and ground-based GNSS (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 ASUCA-Var system and produces initial conditions for MSM forecasts at 00, 03, 06, 09, 12, 15, 18 and 21 UTC. The data assimilation window is set as three hours, with the end of the window corresponding to the analysis time. Figure 2.6.1 shows the MA process:

1. With the initial condition produced in the previous MA, the high-resolution (5 km) forecast model is run within the three-hour data assimilation window to obtain the first guess.

2. Quality-control for observation (Section 2.3) and calculation of related deviations from the first guess are conducted as preparation for input to the next step.

3. 4D-Var analysis is performed to determine the optimal model state by assimilating observations in a low-resolution (15 km) space.

4. The low-resolution (15 km) analysis increment is added to the high-resolution (5 km) first guess via interpolation to determine the initial condition for the next step.

5. Based on this initial condition, the high-resolution (5 km) forecast model is run within the data assimilation window to determine 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 that in which ASUCA-Var in the low-resolution space is executed is called the inner step. ASUCA is used as the forecast model for both steps. The analysis domain is shown in Figure 2.6.2 with a topographic map (5-km resolution) as used by MA. The lateral boundary conditions are from Global Spectral Model (GSM, Section 3.2) forecasts, while the initial conditions of the first guess are taken from the previous MA (the three-hour forecast in the last outer step). Thus, MA frames the cycle analysis nested into the GSM.

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 with the observation time rounded to the nearest hour (as represented by the four stars in Figure 2.6.1). Accordingly, data observed from 4 hours before to 0.5 hours after analysis are assimilated in the inner step.

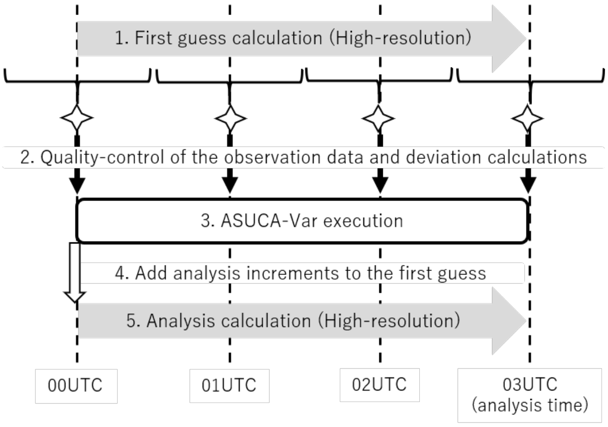

Figure 2.6.1: MA procedure (03 UTC analysis)

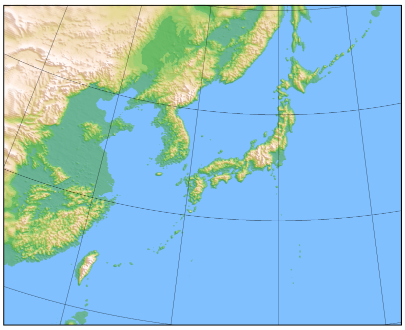

Figure 2.6.2: MA domain and topography

As described previously, ASUCA-Var in the inner step involves a data assimilation system based on a four-dimensional variational (4D-Var) method as detailed in Subsection 2.6.3. This approach is based on maximum likelihood estimation, with optimal values (i.e. analysis fields) determined by minimizing the cost function (Subsection 2.6.3.1 for details). This minimization requires iterative calculation of the cost function and its gradient (50 times at maximum), which takes considerable computational resources. To reduce this burden, operational ASUCA-Var 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 minimization. The basic field in this approach is updated with the low-resolution nonlinear model twice (20 and 35 steps). Operational formulation of minimization with this approach is detailed in Subsection 2.6.3.1. The model used in minimization for ASUCA-Var is called the inner model, with the specifications described in Subsection 2.6.3.3. The horizontal grid spacing is 5 km (817 × 661 grid points) with 96 vertical layers for atmosphere in the outer steps, and 15 km (273 × 221 grid points) with 48 vertical layers in the inner step.

2.6.3 Basic Formulation

2.6.3.1 Cost Function

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

$$\begin{aligned} J(\boldsymbol{x}_0) &= J_b + J_o + J_{vbc} + J_p \\ &= \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathbf{B}_0^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \sum_{t=0}^{N} \frac{1}{2}\Big(H_t(\boldsymbol{x}_t) - \boldsymbol{y}_t + P(\boldsymbol{b})\Big)^T \mathbf{R}_t^{-1}\Big(H_t(\boldsymbol{x}_t) - \boldsymbol{y}_t + P(\boldsymbol{b})\Big) \\ &\quad + J_{vbc} + J_p, \end{aligned} \qquad (2.6.1)$$

where the superscript T represents transpose.

The first and second terms of Eq. (2.6.1) are background and observation terms, representing deviations from the first guess and observation, 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 the 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 observation parameters and interpolation from model grid points to observation points. $P(\boldsymbol{b})$ is observation bias. The error covariance matrixes $\mathbf{B}_0$ and $\mathbf{R}_t$ specify the error profiles (uncertainty and error correlation) of $\boldsymbol{x}_0^b$ and $\boldsymbol{y}_t$, respectively (Subsection 2.6.3.2 and Subsection 2.6.4.2). Observation terms are modified via the introduction of variational quality control (Subsection 2.6.4.5).

The third term of Eq. (2.6.1) J_{vbc} is for variational bias correction (VarBC; Dee 2004) to estimate observation bias. It is given as

$$J_{vbc} = \frac{1}{2}\left(\boldsymbol{b} - \boldsymbol{b}^b\right)^T \mathbf{B}_{\mathrm{vbc}}^{-1}\left(\boldsymbol{b} - \boldsymbol{b}^b\right), \tag{2.6.3}$$

where $\boldsymbol{b}$ is a control variable for bias correction, $\boldsymbol{b}^b$ is the first guess of $\boldsymbol{b}$, and $\mathbf{B}_{\mathrm{vbc}}$ is the background error covariance matrix for VarBC, which is used to correct for satellite brightness temperature. VarBC background error is defined from Cameron and Bell (2018) as

$$B_{\mathrm{vbc}} = \frac{\sigma_o^2}{N_b}, \tag{2.6.4}$$

$$N_b = \max(m_{\mathrm{avg}}, m_{\mathrm{min}})\left(\frac{1}{2^{1/n} - 1}\right), \tag{2.6.5}$$

where σ_o is the observation error standard deviation, m_{avg} is the average number of observations including a particular channel assimilated over the last three days, m_{min} is the lower limit of observation numbers (500), and n is the bias halving time (set as 8).

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

$$J_p = \frac{1}{2}\left(\boldsymbol{x}_{N/2} - \overline{\boldsymbol{x}}_{N/2}\right)^T \mathbf{B}_{\mathrm{df}}^{-1}\left(\boldsymbol{x}_{N/2} - \overline{\boldsymbol{x}}_{N/2}\right) \tag{2.6.6}$$

where $\boldsymbol{x}_{N/2}$ is the model state analysis increment at the center of the data assimilation window ($t = N/2$), $\overline{\boldsymbol{x}}_{N/2}$ is the digitally filtered analysis increment at $t = N/2$, and $\mathbf{B}_{\mathrm{df}}$ is the background error covariance matrix for digital filtering. Based on Wee and Kuo (2004), $\mathbf{B}_{\mathrm{df}}$ is given as

$$\mathbf{B}_{\mathrm{df}}^{-1} = \frac{\lambda}{\mathrm{diag}(\mathbf{B}_0)}, \tag{2.6.7}$$

where λ is the factor 1.0×10^{-6}. 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

$$\overline{\boldsymbol{x}}_{N/2} = \sum_{k=0}^{N} h_{N/2-k} W_k \boldsymbol{x}_k \, , \tag{2.6.8}$$

where

$$h_k = \frac{\sin k\theta_c}{k\pi} \, , \tag{2.6.9}$$

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).

---

[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 operational MA.

In MA, an incremental approach is adopted to reduce computational cost. Optimization is performed using the inner model (Subsection 2.6.3.3) to determine an analysis increment for the low-resolution model space (inner step). The cost function is expanded around the basic field, which can be expressed as

$$\begin{aligned} J(\delta \boldsymbol{x}_0) = {} & \frac{1}{2}\delta \boldsymbol{x}_0{}^T \mathbf{B}_0^{-1} \delta \boldsymbol{x}_0 \\ & + \sum_{t=0}^{N} \frac{1}{2}\left(\mathbf{H_t}\mathbf{M_t}\delta \boldsymbol{x}_0 + \mathbf{P}\delta \boldsymbol{b} - \mathbf{d_t}\right)^T \mathbf{R}_t^{-1}\left(\mathbf{H_t}\mathbf{M_t}\delta \boldsymbol{x}_0 + \mathbf{P}\delta \boldsymbol{b} - \mathbf{d_t}\right) \\ & + J_{vbc} + J_p, \end{aligned} \tag{2.6.10}$$

where

$$\delta \boldsymbol{x}_0 = \boldsymbol{x}_0 - \boldsymbol{x}_0^b, \tag{2.6.11}$$

$$\delta \boldsymbol{b} = \boldsymbol{b} - \boldsymbol{b}_0^b, \tag{2.6.12}$$

$$\boldsymbol{d}_t = \boldsymbol{y}_t - H_t(M_t(\boldsymbol{x}_0^b)) - P(\boldsymbol{b}^b), \tag{2.6.13}$$

and $\mathbf{H_t}$, $\mathbf{M_t}$, and $\mathbf{P}$ are tangent-linear operators of H_t, M_t, and P, respectively.

The effect of the non-linear process was insufficiently incorporated in analysis values with linear optimization alone. To address this problem, a basic-field update (Trémolet 2008) was introduced to incorporate non-linear effects into optimization. Details of the update in ASUCA-Var are provided in *Section 2.1* of Ikuta *et al.* (2021). The field is updated twice during optimization.

The final analysis $\boldsymbol{x}_N$ is determined from forecasting with the high-resolution model over the data assimilation window (outer step).

$$\boldsymbol{x}_N = M_N(\boldsymbol{x}_0)\ . \tag{2.6.14}$$

2.6.3.2 Background Error Covariance

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

A group of parameters are defined as analysis variables, with background errors considered uncorrelated. The variables used in MA are

- u: x component of horizontal wind
- v: y component of horizontal wind
- $\left(T_g, p_s, \theta\right)$: underground and skin temperature, surface pressure and potential temperature
- $\left(W_g, \mu_p\right)$: soil moisture and pseudo-relative humidity ($\mu_p = q_v/q_{sat}^b$, where q_v is the mixing ratio of water vapour and q_{sat}^b is the saturated water vapour of the first guess).

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

$\delta \boldsymbol{x}_0$ is transformed from uncorrelated control variables χ as

$$\delta \boldsymbol{x}_0 = \boldsymbol{x}_0 - \boldsymbol{x}_0^b = \mathbf{B}^{1/2}\chi, \tag{2.6.15}$$

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

$$\mathbf{B}^{1/2} = \mathbf{K}_p \begin{pmatrix} \mathbf{B}_u^{1/2} & 0 & 0 & 0 \\ 0 & \mathbf{B}_v^{1/2} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{T_g,p_s,\theta}^{1/2} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{W_g,\mu_p}^{1/2} \end{pmatrix}. \tag{2.6.16}$$

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

$$\mathbf{B}_{\chi_i}^{1/2} = \mathbf{C}_v \mathbf{B}_{h,\chi_i}^{1/2} \mathbf{B}_{v,\chi_i}^{1/2}. \tag{2.6.17}$$

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

The square root of the vertical background error covariance matrix $\mathbf{B}_{v,\chi_i}$ is given by

$$\mathbf{B}_{v,\chi_i}^{1/2} = \mathrm{diag}\left(\mathbf{B}_{v,\chi_i}\right)^{1/2} \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{V}^T , \tag{2.6.18}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\left[\mathrm{diag}\left(\mathbf{B}_{v,\chi_i}\right)^{-1/2} \mathbf{B}_{v,\chi_i} \mathrm{diag}\left(\mathbf{B}_{v,\chi_i}\right)^{-1/2}\right]$, 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). Vertical background error covariance is calculated separately for sea and land grids (except u and v) and classified by local time.

Variational optimization is performed with respect to χ. The transform (2.6.15), called preconditioning, simplifies the background term of the cost function J (see Eq. (2.6.1) and Eq. (2.6.10); for simplicity, the present discussion does not deal with J_{vbc} and J_p). The cost function and its gradient after the transform are given as

$$J(\chi) = \frac{1}{2}\chi^T\chi + \sum_{t=0}^{N} \frac{1}{2}\left(\mathbf{H_t M_t B}^{1/2}\chi - \boldsymbol{d}_t\right)^T \mathbf{R}_t^{-1}\left(\mathbf{H_t M_t B}^{1/2}\chi - \boldsymbol{d}_t\right), \tag{2.6.19}$$

$$\nabla_\chi J = \chi + \sum_{t=0}^{N} \mathbf{B}^{T/2}\mathbf{M}_t^T\mathbf{H}_t^T\mathbf{R}_t^{-1}\left(\mathbf{H}_t\mathbf{M}_t\mathbf{B}^{1/2}\chi - \boldsymbol{d}_t\right), \tag{2.6.20}$$

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

2.6.3.3 Inner Model

In ASUCA 4D-Var, a high-resolution with the same configuration as the MSM (5 km horizontal grid spacing, 96 vertical layers), is used in the outer step. The inner step is executed using the nonlinear model (NLM), tangent-linear model (TLM) and adjoint model (ADM), with a lower resolution (15 km horizontal grid spacing, 48 vertical layers) to reduce computational cost. The specifications of the inner and outer models are listed in Table 2.6.1.

The NLM is essentially the same as the outer model except for the resolution and the configuration of convective parameterization. The dynamics of the TLM are fully tangent-linearized, although some physical processes are simplified or not implemented to avoid errors associated with non-linearity.

In regard to cloud microphysics, saturation adjustment is linearized but other processes are not. For convective parameterization, the NLM adopts the modified Kain-Fritsch scheme (Kain and Fritsch 1990) but the TLM does not. For the boundary layer scheme, Mellor-Yamada-Nakanishi-Niino level 3 (Nakanishi and Niino 2006) is employed, and is linearized for the TLM except for diffusion coefficients and partial condensation. The surface flux scheme based on Beljaars and Holtslag (1991) is linearized for the TLM except for bulk coefficients. The TLM Radiation scheme is simplified based on Mahfouf (1999).

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.

Table 2.6.1: Specifications of the outer and inner models employed in MA

	Outer model	Inner NLM	Inner TLM/ADM
Resolution	5 km, 96 layers	15 km, 48 layers	15 km, 48 layers
Cloud microphysics	6-class 3-ice bulk scheme, Subsection 3.5.4	6-class 3-ice bulk scheme, Subsection 3.5.4	Only saturation adjustment process tangent-linearized
Convection	Modified Kain-Fritsch	Modified Kain-Fritsch	None
Boundary layer	Mellor-Yamada-Nakanishi-Niino level-3	Mellor-Yamada-Nakanishi-Niino level-3	Tangent-linearized except for diffusion coefficients and partial condensation
Surface flux	Beljaars and Holtslag	Beljaars and Holtslag	Tangent-linearized except for bulk coefficients
Radiation	Subsection 3.5.6	Subsection 3.5.6	Mahfouf (1999)

2.6.4.2 Observation Error

The observation error covariance matrix **R** in Eq. (2.6.1) is assumed to be diagonal. 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 (Table 2.5.2(c) - (j)). 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. Observation error is calculated using 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 Radar/Raingauge Analyzed Precipitation (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 pre-analysis.

2.6.4.3 Observation Operator

The RTTOV-13.0 fast radiative transfer model is used as the observation operator for satellite radiance data assimilation, while ROPP8 for the assimilation of refractivity data from GNSS-RO. These 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 in Eq. (2.6.1), a Gaussian probability density function (PDF) for observation errors is assumed. However, as the PDF for precipitation-amount data is not Gaussian, the following observation term is used for one-hour precipitation-amounts (Koizumi *et al.* 2005):

$$J_o{}^{PREC}(x) = \sum_{j\ (\text{where } r_j^o \geq 0.5)}^{n} \frac{\left(H_j(x) - r_j^o\right)^2}{2\sigma_o(r_j^o)^2}. \tag{2.6.21}$$

Here, $H_j(x)$ is an observation operator used to convert the state variable x to one-hour cumulative precipitation values at the j-th grid point, r_j^o is precipitation observed at the grid point, and n is the number of grid points in

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

the inner model[6]. $\sigma_o(r_j^o)$ is the observation error standard deviation, defined as

$$\sigma_o(r_j^o) \equiv \begin{cases} C_{sat}\max(r_{min}, r_j^o) & \left(H_j(x) \le r_j^o\right) \\ C_{sat}C_a\max(r_{min}, r_j^o) & \left(H_j(x) > r_j^o\right) \end{cases}, \quad \begin{cases} C_a = 3,\ C_{sat} = 1 & \text{for R/A} \\ C_a = 5,\ C_{sat} = 2 & \text{for satellite retrievals} \end{cases}, \quad r_{min} \equiv 1\text{mm/h}. \tag{2.6.22}$$

Here, 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 values 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 observation rather than from one-hour cumulative observation.

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 observation error PDF is assumed to be a summation of a Gaussian function and a positive constant value within a certain range. This constant represents 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:

$$j_o^{VarQC} = -\log\left(\frac{\gamma + \exp(-j_o)}{\gamma + 1}\right), \quad \gamma \equiv \frac{A\sqrt{2\pi}}{(1-A)\,2d} \tag{2.6.23}$$

$$\nabla j_o^{VarQC} = W^{VarQC}\nabla j_o, \quad W^{VarQC} \equiv 1 - \frac{\gamma}{\gamma + \exp(-j_o)}. \tag{2.6.24}$$

---

[6] If one-hour cumulative precipitation P_{1h} is larger than 1 mm/h, $\left(3P_{1h}^{1/3} - 2\right)$ is used for $H_j(x)$ and r_j^o.

Here, A is the prior probability of rough error (e.g., 0.05 for SYNOP), d is the maximum standard deviation below which rough error is possible (e.g., 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. (2.6.24) 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 ($W^{VarQC} << 1$). Observation values satisfying $W^{VarQC} < 0.25$ are regarded for rejection in VarQC.

2.7 Local Analysis

2.7.1 Introduction

Local Analysis (LA) produces initial conditions for the Local Forecast Model (LFM) (Section 3.7) 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). In March 2022, the analysis cycle was updated to hybrid 3D-Var utilizing ensemble perturbation of Meso-Scale Ensemble Prediction System (MEPS) data (Section 3.6) (Yokota *et al.* 2022).

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 contributing to better forecasting of high-impact phenomena (see Table 2.1.3). LA was also adopted in January 2017 to make extensive use of satellite observations , 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 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 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. In 3D-Var, the weighted average of flow-dependent and climatological background error covariances (hybrid covariance) is used (hybrid 3D-Var), and the former is created from ensemble perturbations of MEPS.

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 hybrid 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 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 (Section 3.5), which also provides lateral and upper boundary conditions for LF1 throughout the three-hour data assimilation period. Sea surface temperature is given by HIMSST (see Subsection 5.2.2). The analyzed field in the last 3D-Var ((d) in Figure 2.7.2) is used as the initial condition of the LFM. Further LA specifications are provided in Table 2.1.3.

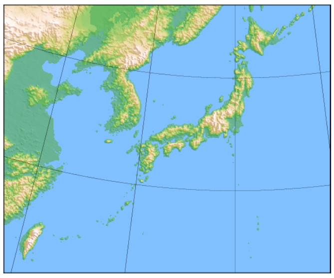

Figure 2.7.1: domains of LA and LFM.

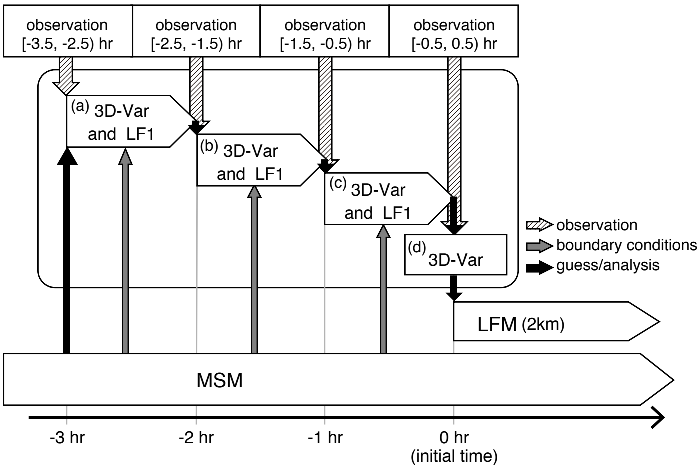

Figure 2.7.2: Schematic representation of LA analysis.

2.7.3 Basic Formulation

2.7.3.1 Cost Function

LA involves four time slots for the assimilation of observations using hybrid 3D-Var, in which the extended control variable method (Lorenc 2003) is used to take into account flow-dependent error covariance. The related cost function in the i-th time slot is defined as

$$
\begin{aligned}
J(\delta\boldsymbol{x}_i) =& \frac{1}{2}\delta\boldsymbol{x}_{i,cl}^T\mathbf{B}_{cl}^{-1}\delta\boldsymbol{x}_{i,cl} + \frac{1}{2}\delta\boldsymbol{x}_{i,en}^T\mathbf{B}_{en}^{-1}\delta\boldsymbol{x}_{i,en} && (2.7.1)\\
&+\frac{1}{2}\left(\mathbf{H}\delta\boldsymbol{x}_i - \boldsymbol{d}_i + \mathbf{P}\delta\boldsymbol{b}_i\right)^T\mathbf{R}^{-1}\left(\mathbf{H}\delta\boldsymbol{x}_i - \boldsymbol{d}_i + \mathbf{P}\delta\boldsymbol{b}_i\right) && (2.7.2)\\
&+\frac{1}{2}\delta\boldsymbol{b}_i^T\mathbf{S}_i^{-1}\delta\boldsymbol{b}_i, && (2.7.3)
\end{aligned}
$$

$$
\delta\boldsymbol{x}_i = \beta_{cl}\delta\boldsymbol{x}_{i,cl} + \beta_{en}\delta\boldsymbol{x}_{i,en}. \tag{2.7.4}
$$

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

$$
\boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}\left(\boldsymbol{x}_i^b\right) - \mathcal{P}\left(\boldsymbol{b}_i^b\right), \tag{2.7.5}
$$

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. $\mathcal{H}$ 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. The subscripts "*cl*" and "*en*" mean "climatological" and "ensemble", respectively. β_{cl} and β_{en} are the weights for the hybrid covariances, set as $(\beta_{cl}^2, \beta_{en}^2) = (0.5, 0.5)$.

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 hybrid 3D-Var and one-hour forecasting. The one-hour forecast operator $\mathcal{M}_{i,i+1}$ is configured specifically for LA. 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 LA cycle is conducted as outlined below.

1. $\mathbf{B}_{en}$ valid at $t-3\text{h}, t-2\text{h}, t-1\text{h}, t$ are created from MEPS ensemble perturbations.

2. $\boldsymbol{x}_{i=1}^b$ valid at $t - 3\text{h}$ 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 LA.

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

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

   $$\boldsymbol{x}_{i+1}^b = \mathcal{M}_{i,i+1}\left(\boldsymbol{x}_i^b + \delta \boldsymbol{x}_i\right), \tag{2.7.6}$$

   and the observation bias is given by

   $$\boldsymbol{b}_{i+1}^b = \boldsymbol{b}_i^b + \delta \boldsymbol{b}_i. \tag{2.7.7}$$

5. Steps 3 and 4 are repeated three times.

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

   $$\boldsymbol{x}^a = \boldsymbol{x}_4^b + \delta \boldsymbol{x}_4, \tag{2.7.8}$$

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

   $$\boldsymbol{b}^a = \boldsymbol{b}_4^b + \delta \boldsymbol{b}_4. \tag{2.7.9}$$

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 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 at land grids, as that at 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:

$$\delta \boldsymbol{x} = \mathbf{B}^{1/2}\chi = \begin{bmatrix} \beta_{cl}\mathbf{B}_{cl}^{1/2} & \beta_{en}\mathbf{B}_{en}^{1/2} \end{bmatrix} \begin{bmatrix} \chi_{cl} \\ \chi_{en} \end{bmatrix}, \tag{2.7.10}$$

$\mathbf{B}^{1/2}$ is the square root of $\mathbf{B}$.

The climatological background error covariance $\mathbf{B}_{cl}$ is modeled as described in Subsection 2.6.3.2, and statistics 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.

The ensemble background error covariance $\mathbf{B}_{en}$ is created from 100 ensemble perturbations using 5 lagged MEPS forecasts with 20 members, and spatial localizations with Gaussian functions (scales of $1/\sqrt{e}$ are set as 100 km horizontally and 0.5 km vertically) to reduce sampling errors. $\mathbf{B}_{en}$ is inflated by multiplication with a factor, which is the ratio of $\mathbf{B}_{cl}$ and the horizontal mean of $\mathbf{B}_{en}$ for potential temperature at 5.5 km above ground level, meaning that error variance is comparable to the magnitude of $\mathbf{B}_{cl}$.

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 RTTOV-13.0 (Saunders *et al.* 2020), 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 six hours with data at the same resolution as Global Analysis with cycle runs and early runs (TQ959: grid approx. 0.125°(13 km); Table 2.1.1) . Analysis incorporates forecast snow depth fields, satellite snow-cover data and SYNOP snow depth data with two-dimensional optimal interpolation (2D-OI).

The satellite snow-cover product is retrieved in the 0.25° × 0.25° latitude/longitude grid from microwave imager radiances[7] with the same scheme as used for JRA-55 (Kobayashi *et al.* 2015). In the algorithm, regions where the observation satisfies the following criteria are considered to be covered by snow[8]:

$$T_b\,(19GHz, \mathrm{V}) < 265K$$

$$T_b\,(19GHz, \mathrm{H}) - T_b\,(37GHz, \mathrm{H}) > \mathrm{threshold}\,(\mathrm{H})$$

$$T_b\,(19GHz, \mathrm{V}) - T_b\,(37GHz, \mathrm{V}) > \mathrm{threshold}\,(\mathrm{V})$$

where T_b is the brightness temperature, H and V are the horizontal and vertical channels, respectively. Thresholds (H, V) should be 0 K for ideal snow but in practice need to be adjusted depending on the snow and surface condition. Accordingly, those values are determined for each region, vegetation, and month. For grid with no satellite radiance for not more than 3 days, past snow-cover product is used. For grid with no satellite radiance more than 3 days, missing value is used.

The first guess snow depth fields used in 2D-OI are estimated mainly from the satellite snow-cover product (SSCP) and the forecast snow depth as below.

1. The SSCP is interpolated into gridding for the GSM.
2. The SSCP is modified under certain conditions. Where SSCP is zero or missing, the climatological probability of snow (CPS) exceeds 80 % and snow is forecast, the CPS value is applied. Where SSCP is otherwise missing, the CPS value or zero is applied depending on the forecast land surface temperature.
3. If there is snow in both the forecast and the satellite, the forecast snow depth is used as the first guess.
4. If there is no snow in satellite data and the forecast snow depth exceeds 5 cm, the first guess is set to the forecast snow depth multiplied by $\exp(-0.1)$.
5. If there is no snow in satellite data and the forecast snow depth is equal to or less than 5 cm, the first guess is set to 0 cm.

This first guess and observations reported from SYNOP stations are handed over to the 2D-OI system for the snow depth analysis[9]. The spatial correlation coefficient for the background error, μ_{kl}, is based on Brasnett (1999) and given by

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

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:

$$\alpha(r_{kl}) = \left(1 + \frac{r_{kl}}{L}\right)\exp\left(-\frac{r_{kl}}{L}\right),$$
$$\beta(\Delta z_{kl}) = \exp\left\{-\left(\frac{\Delta z_{kl}}{h}\right)^2\right\},$$

where L and h are set to 100 km and 800 m, respectively. The standard deviation ratio of observation error to background error is set as 0.75. Grids on ocean are skipped in snow depth analysis. Grids where there is no

---

[7] SSMIS (DMSP F-17, F-18) and AMSR2 (GCOM-W1) are used.

[8] There are some typographical errors in Kobayashi *et al.* (2015).

[9] SYNOP reports from Japan are not used in 2D-OI. Analysis for grid points over the country is described later in this section.

snow in both the forecast and the satellite are also skipped and the analysis are set to 0 cm. Grids where the forecast ground surface temperature is above the freezing point and there is very few snow only in the forecast or the satellite are also skipped and the analysis are set to 0 cm. Grids covering land ice are also skipped and the analysis are set to the climatological values.

Grid point values for analysis over Japan are replaced by inverse distance-weighted interpolation of the latest observations reported by SYNOP and AMeDAS stations.

Snow depths are converted to snow water equivalents as an initial condition for land-surface processing (see Subsection 3.2.10) in the GSM.

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 in 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

$$\mu_{kl} = \alpha(r_{kl})\beta(\Delta z_{kl}) \tag{2.8.1}$$

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:

$$\alpha(r_{kl}) = \left(1 + \frac{r_{kl}}{L}\right)\exp\left(-\frac{r_{kl}}{L}\right), \tag{2.8.2}$$

$$\beta(\Delta z_{kl}) = \exp\left\{-\left(\frac{\Delta z_{kl}}{h}\right)^2\right\}, \tag{2.8.3}$$

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 Soil Moisture Analysis

Global soil moisture analysis is performed as part of Global Analysis (Table 2.1.1) every six hours. Soil moisture fields for the first three soil layers are analyzed via a simplified Extended Kalman Filter (EKF) (Hess 2001, de Rosnay *et al.* 2013) using first-guess forecasts of soil moisture and analyzed screen-level temperature and relative humidity.

In this EKF, the analyzed state vector $\mathbf{x_a}$ at time t_i is calculated for each grid point as

$$\mathbf{x_a}(t_i) = \mathbf{x_b}(t_i) + \mathbf{K_i}\left[\mathbf{y_o}(t_i) - H_i(\mathbf{x_b})\right] \tag{2.9.1}$$

where $\mathbf{x_b}$, $\mathbf{y_o}$ and H_i are the first-guess state vector, observation state vector and nonlinear observation operator, respectively. In global soil moisture analysis, values (i.e., degrees of saturation) for the first three soil levels (from the surface to a depth of 0.19 m) correspond to $\mathbf{x_a}$ and $\mathbf{x_b}$. Analyzed screen-level temperature and relative humidity are treated as $\mathbf{y_o}$. The Kalman gain matrix $\mathbf{K_i}$ is calculated as

$$\mathbf{K_i} = \left[\mathbf{B^{-1}} + \mathbf{H_i}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H_i}\right]^{-1}\mathbf{H_i}^{\mathbf{T}}\mathbf{R^{-1}} \tag{2.9.2}$$

where **B** and **R** are background-error and observation-error covariance matrices, respectively. Following de Rosnay *et al.* (2013), these are static and diagonal matrices composed of error variances for simplification. The terms are based on soil moisture standard deviation $\sigma_b = 0.02\,\mathrm{m}^3/\mathrm{m}^3$, and screen-level parameter standard deviations of $\sigma_T = 1\,\mathrm{K}$ for temperature and $\sigma_{RH} = 4\,\%$ for relative humidity. $\mathbf{H_i}$ is a linearized observation operator approximated from finite differences between perturbed and unperturbed forecasts based on a low-resolution (TL319L128) version of GSM2103 (Table 3.1.1).

The analyzed screen-level parameters are derived from two-dimensional optimal interpolation (2D-OI) assimilating SYNOP observations. The spatial correlation coefficient for background error, μ_{kl}, is based on Drusch and Viterbo (2007) as

$$\mu_{kl} = \alpha(r_{kl})\beta(\Delta z_{kl}) \tag{2.9.3}$$

with 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:

$$\alpha(r_{kl}) = \exp\left\{-\frac{1}{2}\left(\frac{r_{kl}}{L}\right)^2\right\}, \tag{2.9.4}$$

$$\beta(\Delta z_{kl}) = \exp\left\{-\left(\frac{\Delta z_{kl}}{h}\right)^2\right\}, \tag{2.9.5}$$

where L and h are 200 km and 800 m, respectively. The standard deviations of background and observation errors are 1.5 and 2 K for temperature and 5 and 10 % for relative humidity. In Eq. (2.9.1), the analysis increments of screen-level parameters are used as background departures $\mathbf{y_o}(t_i) - H_i(\mathbf{x_b})$.

2.10 Non-real-time Quality Control

2.10.1 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[10].

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[11].

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).

2.10.2 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

---

[10] `http://qc.kishou.go.jp/clsf.html`

[11] `http://qc.kishou.go.jp/mmr.html`

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.

2.10.3 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.

2.11 Climate Data Assimilation System

2.11.1 Introduction

Climate system monitoring and seasonal prediction require accurate and consistent comprehension based on high-quality, spatio-temporally homogeneous long-term data. To this end, JMA has to date developed the Japanese 25-year Reanalysis (JRA-25; Onogi *et al.* 2007) dataset in conjunction with the Central Research Institute of Electric Power Industry, the Japanese 55-year Reanalysis (JRA-55; Kobayashi *et al.* 2015) dataset and the Japanese Reanalysis for Three Quarters of a Century (JRA-3Q; Kosaka *et al.* 2024) dataset. At present, JRA-3Q data are used for real-time climate monitoring, seasonal prediction and various R&D activities at JMA and in wider communities (e.g., fundamental academic fields such as meteorology, climatology and oceanography) and for practical application in fields such as agricultural meteorology and renewable energy.

2.11.2 JRA-55

JMA's second reanalysis (JRA-55, completed in 2013) covers the period from 1958, when regular global radiosonde observation began. Production is continued to the present on a near-real-time basis for operational climate data assimilation as described in Kobayashi *et al.* (2015). The system is based on the TL319 version of the JMA global NWP system as of December 2009 (JMA 2007, 2013), and features numerous improvements made since JRA-25 (Table 2.11.1). 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 new observational datasets produced from 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 from major meteorological satellite centers (e.g., van de Berg *et al.* 2002; Oyama 2010).

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

2.11.3 JRA-3Q

The latest Japanese reanalysis (JRA-3Q, completed in 2022) covers the period from September 1947, when Typhoon Kathleen caused extreme flooding in Japan. Production is continued to the present on a near-real-time basis for operational climate data assimilation. The system is based on the TL479 version of the JMA global NWP system as of December 2018 (JMA 2019), and incorporates developments in the operational NWP system, boundary conditions and forcing fields achieved at JMA since JRA-55 (Table 2.11.1). In JMA's GSM, for example, biases in radiation budgets, surface sensible/latent heat fluxes and precipitation are significantly reduced due to extensive improvement of parameterizations for physical processes. Boundary conditions over the ocean applied from June 1985 onward have also been replaced with satellite-based high-resolution (0.25-degree) sea surface temperature (SST) data (MGDSST; Kurihara *et al.* 2006, Subsection 5.2.1) to improve the representation of atmospheric processes around western boundary currents and associated SST fronts. JRA-3Q also incorporates rescued historical observations to temporally extend analysis to around 10 years earlier than JRA-55.

The significant JRA-55 issue of global energy imbalance with excess upward net energy flux at the top of the atmosphere and at the surface is significantly reduced in JRA-3Q. The former's trend toward artificial weakening of TCs is also resolved via the generation of TC bogus based on JMA's operational system.

Table 2.11.1: Data assimilation in JRA-55 and JRA-3Q

	JRA-55	JRA-3Q
Analysis period	1958 – 2023/early 2024	September 1947 onward
Base system	JMA operational system as of December 2009 (JMA 2007, 2013)	JMA operational system as of December 2018 (JMA 2019)
Horizontal resolution	TL319 (– 55 km)	TL479 (– 40 km)
Vertical levels	Surface and 60 levels up to 0.1 hPa (Iwamura and Kitagawa 2008; Nakagawa 2009)	Surface and 100 levels up to 0.01 hPa (Kawai *et al.* 2013)
Analysis scheme	4D-Var with T106 inner resolution	4D-Var with TL319 resolution (Kadowaki and Yoshimoto 2012)
Radiosonde temperature bias correction	Until 2006: RAOBCORE V1.4 (Haimberger *et al.* 2008) From 2007 onward: RAOBCORE V1.5 (Haimberger *et al.* 2012) • Based on comparison with ERA	RISE (RICH with solar elevation dependence) v1.7.2 (Haimberger *et al.* 2012) • Based on comparison with surrounding stations • Seasonally dependent (from 1979 onward)
Radiative transfer model for satellite radiances	RTTOV-9.3 (Saunders 2008)	RTTOV-10.2 (Saunders *et al.* 2012) • Improved accuracy • Inclusion of GHGs variations
Land surface analysis	Offline SiB	Cycle of land surface forecasts from modeling
SST and sea ice	COBE-SST (1 degree) (Ishii *et al.* 2005, Subsection 5.2.3)	Until May 1985: COBE-SST2 (1 degree) (Hirahara *et al.* 2014, Subsection 5.2.3) From June 1985 onward: MGDSST (0.25 degrees) (Kurihara *et al.* 2006, Subsection 5.2.1)
Ozone	Until 1978: Climatology From 1979 onward: MRI-CCM1 (T42L68) (Shibata *et al.* 2005)	MRI-CCM2.1 (TL159L64) (Deushi and Shibata 2011) • Produced with the new model for the whole period