

Algorithm Theoretical Baseline Document
Level 3 gridded specific humidity

Version 1.0
25 June 2025

NOAA STAR GNSS RO Science and Data Center (SDC)

1. Introduction

The purpose of this document is to describe the algorithms to generate the NOAA STAR gridded monthly mean specific humidity (Level 3 data product). Based on the NOAA STAR SDC Level 2 specific humidity profiles (Ho et al., 2022), Level 3 monthly mean data products are generated on a $10^\circ \times 10^\circ$ latitude-longitude grid for 21 pressure levels, spanning from 850 hPa to 100 hPa. Due to the varying spatial and temporal coverage among different RO missions, we must remove sampling errors for each mission to construct consistent climate data records. We used ERA-5 reanalysis as a reference to estimate and remove sampling error in the Level 3 monthly mean data products. We assessed the robustness of the sampling error removal method by comparing the results derived from three different reanalyses: ERA-5, MERRA-2, and JRA-55. The RO-based Level 3 monthly mean specific humidity is then validated by comparing it with those derived from ERA-5.

2. Input data needed for level 3 gridded data

2.1 STAR WetPrf version 1.0

The wet profiles (WetPrf) from the STAR GNSS RO 1D-var version 1.0 are used to construct the gridded specific humidity dataset covering the troposphere to lower stratosphere from 2002 to 2014. The WetPrf contains the profiles of the latitude and longitude of the RO perigee point, temperature, pressure, water vapor in terms of partial pressure, relative humidity, and specific humidity. We include profiles from seven RO missions: Metop-A, COSMIC-1, CHAMP, C/NOFS, GRACE, SAC-C, and TerraSAR-X. The temporal coverage of each mission is listed in Table 1.

Table 1 *RO Missions and Corresponding Temporal Coverage.*

RO mission	Temporal coverage
Metop-A	Oct 2007 to Nov 2021
COSMIC-1	Apr 2006 to Dec 2019
CHAMP	May 2001 to Oct 2008
C/NOFS	Mar 2010 to Dec 2011
GRACE	Feb 2007 to Nov 2017

SAC-C	Mar 2006 to Aug 2011
TerraSAR-X	Feb 2008 to Nov 2024

2.2 Ancillary Data

The ERA-5, JRA-55, and MERRA-2 reanalyses are used as references for sampling error estimation and correction in the STAR RO gridded specific humidity. STAR RO gridded specific humidity is further validated through comparisons with ERA-5.

2.2.1 ERA-5

ERA-5 reanalysis is the latest climate reanalysis produced by ECMWF (Hersbach et al., 2023). The atmospheric data have been available with multiple spatial and temporal resolutions since 1940 and continue to be extended forward. The global field of atmospheric temperature, sampled at 6-hour intervals (0000, 0600, 1200, 1800 UTC) on a regular $0.25^\circ \times 0.25^\circ$ grid and at 37 pressure levels, is used in this study. Spanning vertically from 1000 hpa to 1 hpa, the ERA-5 reanalysis temperature field can fully cover the upper-troposphere-to-lower-stratosphere (UTLS) region. The daily ERA-5 reanalysis data were used and downloaded from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview>.

2.2.2 MERRA-2

The second Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) is a NASA atmospheric reanalysis that began in 1980 (Global Modeling and Assimilation Office (GMAO), 2015). The MERRA-2 product, with 3-hour temporal resolution and on a regular $0.625^\circ \times 0.625^\circ$ grid with 42 pressure levels ranging from 1000 to 0.1 hPa, used in this study, is available online at <https://daac.gsfc.nasa.gov>.

2.2.3 JRA-55

The second-generation Japanese global atmospheric reanalysis project, known as JRA-55, was conducted by the Japan Meteorological Agency (Japan Meteorological Agency, 2013). It covers the period from 1958 to the present. The global atmospheric variables, sampled at 6-hour intervals (0000, 0600, 1200, 1800 UTC) on a regular $1.25^\circ \times 1.25^\circ$ latitude-longitude grid and 37 pressure levels spanning 1000-1 hPa, were used and downloaded from https://jra.kishou.go.jp/JRA-3Q/index_en.html.

3. Level 3 Algorithm

3.1 Quality control procedure

Before generating the gridded data, we first applied the essential quality control (QC) procedure to all RO-retrieved specific humidity profiles. Fig. 1 shows the monthly and yearly means and standard deviations (STDVs) of the differences between specific humidity from RO and ERA-5 forecasts. The STDV profiles are consistent across months/seasons. We use STDV as an index to evaluate the quality of RO retrievals. For instance, we define $\sigma - year$ as the yearly standard deviation for the year 2009 (Fig. 1). The RO-retrieved specific humidity at any pressure level that exceeds $7 \times \sigma_{year}$ at that level, it is excluded from the gridded data generation.

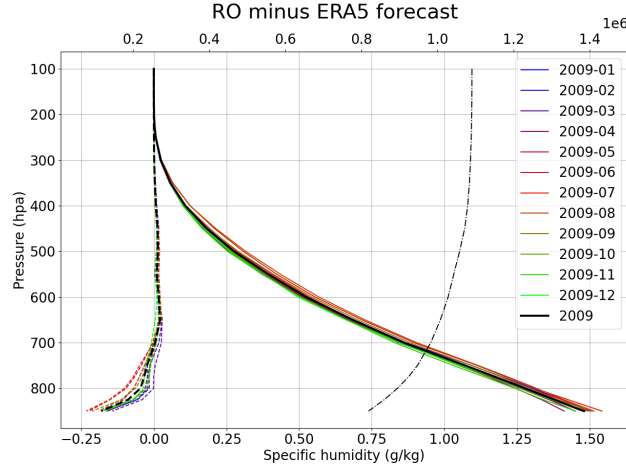


Figure 1 Monthly and yearly mean (dashed lines), STDV (solid lines) of O-B, and the number of valid data (dot-dashed line) in 2009.

3.2 Generating Gridded Monthly Mean Specific Humidity

The specific humidity profiles that pass the QC procedure are first interpolated to 21 pressure levels, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 650, 700, 750, 775, 800, and 850 hPa. They are then binned and area-weighted onto a $10^\circ \times 10^\circ$ latitude-longitude grid for each pressure level:

$$Gridded_{obs} = \frac{1}{\sum_{i=1}^n \cos \varphi_i} \cdot \sum_{i=1}^n Q_i \cos \varphi_i \quad (1)$$

where $Gridded_{obs}$ denotes the gridded data based on RO retrievals, n is the number of RO observations within each grid, Q_i is the i th RO retrieved specific humidity, and φ_i is the latitude. The final gridded data, combined from seven RO missions as mentioned above, span from January 2002 to December 2014.

3.3 Sampling Error Correction

The sampling error is a critical error source in $Gridded_{obs}$ that needs to be estimated and removed for reliable Gridded datasets. The sampling error can be calculated and limited by employing reanalyses through the following procedure (Zhou et al. 2025):

- (1) Interpolating the reanalysis model profiles to the times and locations of each RO profile and generating the sampled monthly mean ($Gridded_{Int}$) following the method described in Section 3.2. $Gridded_{Int}$ represents the gridded dataset with sampling error.
- (2) Calculating the model monthly mean ($Gridded_{full_grid}$) by binning and averaging the four-dimensional reanalysis model field with the weight of cosine latitude to the same grids:

$$Gridded_{full_grid} = \frac{1}{n_t n_\varphi n_\lambda} \cdot \frac{1}{\sum_{k=1}^{n_\varphi} \cos \varphi_k} \cdot \sum_{t=1}^{n_t} \sum_{k=1}^{n_\varphi} \sum_{l=1}^{n_\lambda} Q_{tkl} \cos \varphi_k \quad (2)$$

where Q_{tkl} and φ_k are the specific humidity and latitude at a model grid, respectively, and the summation loops over all n_t , n_φ , and n_λ (time-latitude-longitude) model grid points located within the grid box. $Gridded_{full_grid}$ represents the gridded data without any sampling error.

- (3) The sampling error (SE) in the grid is estimated as the difference between $Gridded_{Int}$ and $Gridded_{full_grid}$:

$$SE = Gridded_{Int} - Gridded_{full_grid} \quad (3)$$

- (4) The sampling error in $Gridded_{obs}$ is corrected by subtracting SE from $Gridded_{obs}$:

$$Gridded_{crr} = Gridded_{obs} - SE \quad (4)$$

Previous studies have shown that the uncertainty of reanalysis data under cloudy conditions, especially over oceans, remains very large (Lonitz and Geer 2017). Fig. 2 presents the time series of the monthly average of O-B and sampling errors estimated from three reanalyses from 60°N to 20°N at 500 hPa. Fig. 2a illustrates that, among the three reanalyses, the specific humidity derived from ERA-5 is the closest to that retrieved from RO. The O-B differences for the three reanalyses exhibit apparent seasonal variation with maxima in summer and minima in winter. This leads to

slight differences in sampling errors estimated by the three reanalyses, which also show seasonal variations (Fig. 2b)

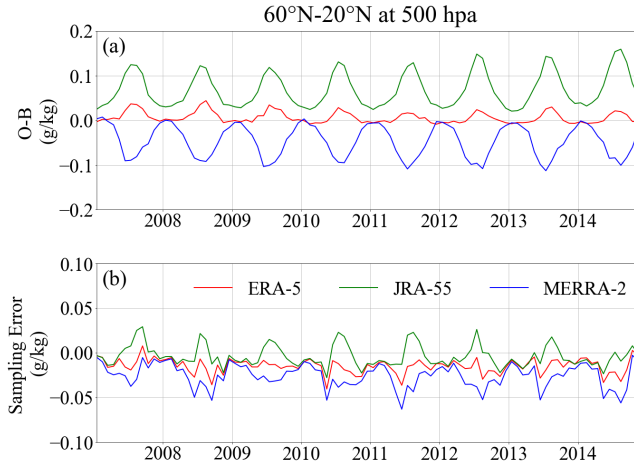


Figure 2 Time series of the monthly average of (a) $Gridded_{obs}$ relative to $Gridded_{Int}$, and (b) the sampling errors estimated from ERA-5 (red), JRA-55 (green), and MERRA-2 (blue) at 500 hPa between 60°N and 20°N.

To reduce the uncertainty caused by using different reanalyses in sampling error correction, the final $\overline{Gridded_{crr}}$ is calculated as the average of the $Gridded_{crr}$ with sampling error corrected by three reanalyses: ERA-5, JRA-55, and MERRA-2. Since MERRA-2 data are terrain-dependent and have missing values at lower altitudes, specific humidity is set to missing for grid points where MERRA-2 data are unavailable. Fig. 3 shows an example of the $\overline{Gridded_{crr}}$ at 500 hPa in January 2009 along with the corresponding STDV of the $Gridded_{crr}$ with sampling error corrected by three reanalyses.

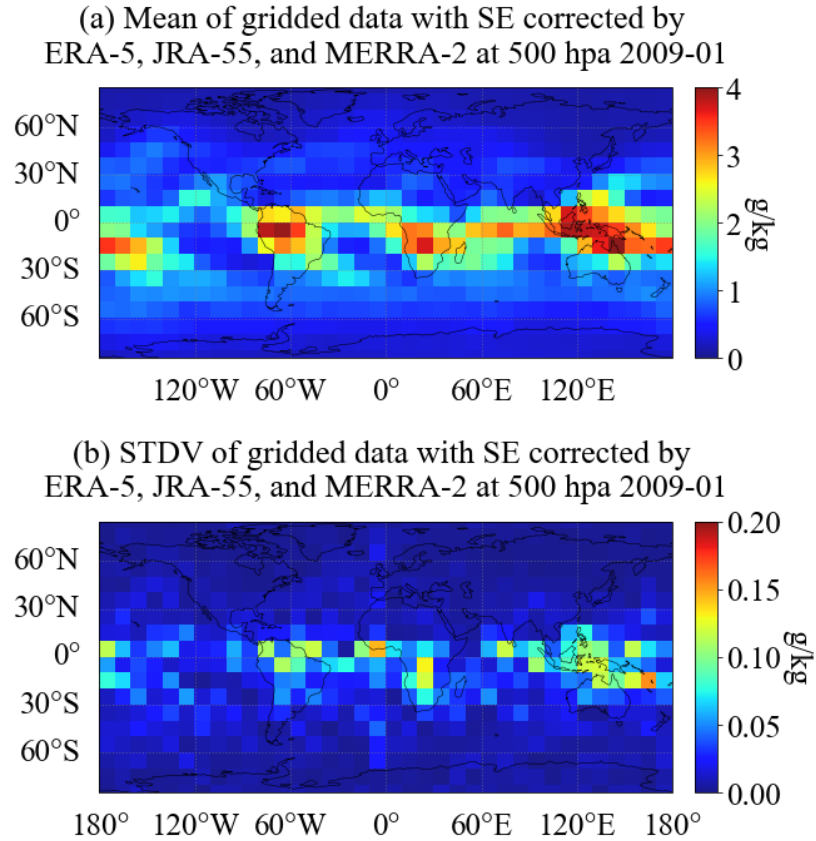


Figure 3 Mean (a) and STDV (b) of the Gridded data with sampling error corrected by ERA-5, JRA-55, and MERRA-2 at 500 hpa in January 2009

4. Data format

The output is archived in NetCDF format: q_mon_GNSS-RO-1-3_BE_gn_200201-201412_STAR.V1.0.nc. The data format is described in Tables 2 and 3.

Table 2 Data dimensions.

Dimension	Description
time	Month, expressed as days since 2000-01-01
plev	Pressure levels at 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 650, 700, 750, 775, 800, and 850 hpa. Unit: pa
lat	Latitude of grid center from 85°S to 85°N with 10° interval

lon	Longitude of grid center from 5° to 355° with 10° interval
-----	--

Table 3 Variable name, dimension, and description.

Variables	Dimension	Unit	Description
q_mean	(time, plev, lat, lon)	g/kg	Mean of the gridded data with sampling error corrected by three reanalyses, i.e. $\overline{Gridded_{crr}}$
q_stdv	(time, plev, lat, lon)	g/kg	STDV of the gridded data with sampling error corrected by three reanalyses
q_ro	(time, plev, lat, lon)	g/kg	RO gridded data before sampling error correction, i.e. $Gridded_{obs}$
q_era5	(time, plev, lat, lon)	g/kg	ERA-5 gridded data, i.e. $Gridded_{full_grid}$
q_jra	(time, plev, lat, lon)	g/kg	JRA-55 gridded data, i.e. $Gridded_{full_grid}$
q_merra	(time, plev, lat, lon)	g/kg	MERRA-2 gridded data, i.e. $Gridded_{full_grid}$
SE_era5	(time, plev, lat, lon)	g/kg	Sampling error estimated by ERA-5, i.e., SE
SE_jra	(time, plev, lat, lon)	g/kg	Sampling error estimated by JRA-55, i.e., SE
SE_merra	(time, plev, lat, lon)	g/kg	Sampling error estimated by MERRA-2, i.e., SE
N_sample	(time, plev, lat, lon)	—	Number of valid samples

5. References

Gorbunov, M. E., Radio-holographic analysis and validation of Microlab-1 radio occultation data in the lower troposphere (2002), J. Geophys. Res., 107, 10.1029/2001JD000 889.

- Gorbunov, M. E. and Lauritsen, K. B. (2004), Analysis of wave fields by Fourier integral operators and their application for radio occultations, *Radio Sci.*, 39, doi:10.1029/2003RS002 971.
- Ho, S.-P.; Peng, L.; Mears, C.; Anthes, R.A. Comparison of global observations and trends of total precipitable water derived from microwave radiometers and COSMIC radio occultation from 2006 to 2013. *Atmospheric Chem. Phys.* 2018a, 18, 259–274. <https://doi.org/10.5194/acp18-259-2018>. 13.
- Ho, S.-P.; Peng, L. Global water vapor estimates from measurements from active GPS RO sensors and passive infrared and microwave sounders. In *Green Chemistry Applications*; IntechOpen: London, UK, 2018b. <https://doi.org/10.5772/interchopen.79541>.
- Ho, S.-p., Kireev, S., Shao, X., Zhou, X., and Jing, X.: Processing and Validation of the STAR COSMIC-2 Temperature and Water Vapor Profiles in the Neutral Atmosphere, *Remote Sensing*, 14, 5588, <https://doi.org/10.3390/rs14215588>, 2022.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N. (2023): ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [Dataset], DOI: 10.24381/cds.bd0915c6 (Accessed on 06-08-2024).
- Global Modeling and Assimilation Office (GMAO) (2015), *inst3_3d_asm_Cp: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4*, Greenbelt, MD, USA: Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC) [Dataset], Accessed 06-08-2024 at doi: 10.5067/VJAFPLI1CSIV.
- Japan Meteorological Agency/Japan (2013), JRA-55: Japanese 55-year Reanalysis, Daily 3-Hourly and 6-Hourly Data, <https://doi.org/10.5065/D6HH6H41>, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated monthly) [Dataset]. Accessed 06-08-2024.
- Lonitz, K. and Geer, A.: Effect of assimilating microwave imager observations in the presence of a model bias in marine stratocumulus, EUMETSAT/ECMWF Fellowship Programme Research Reports, <https://www.ecmwf.int/node/17164> (last access: 7 November 2023), 2017

- ROM SAF ROPP (2021): EUMETSAT ROM SAF The Radio Occultation Processing Package (ROPP) Pre-processor Module User Guide, version 11.0, The ROM SAF Consortium, 2021, Ref:SAF/ROM/METO/UG/ROPP/004, [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://romsaf.eumetsat.int/romsaf_ropp_ug_pp.pdf](https://romsaf.eumetsat.int/romsaf_ropp_ug_pp.pdf).
- Shao, X., Ho, S.-P., Jing, X., Zhou, X., Chen, Y., Liu, T.-C., Zhang, B., and Dong, J.: Characterizing the tropospheric water vapor spatial variation and trend using 2007–2018 COSMIC radio occultation and ECMWF reanalysis data, *Atmos. Chem. Phys.*, 23, 14187–14218, <https://doi.org/10.5194/acp-23-14187-2023>, 2023.
- STAR-ROPP version 1.0 ATBD: Algorithm Theoretical Basis Document Inversion of Bending Angle and Refractivity profiles STAR ROPP Version 1.0 (Based on the ROPP Version 10.0) https://gpsmet.umd.edu/star_gnssro/img/ATBD_STAR_ROPP_final.pdf
- STAR 1D-Var version 3.2 ATBD: The NOAA/STAR GNSS RO 1D-VAR Retrieval Algorithm V.3.2 Algorithm Theoretical Basis Document Version 3.2 [STAR_GNSS_RO_1DVAR_3.2_ATBD.pdf](#)
- Teng, W.-H.; Huang, C.-Y.; Ho, S.-P.; Kuo, Y.-H.; Zhou, X.-J. Characteristics of global precipitable water in ENSO events revealed by COSMIC measurements. *J. Geophys. Res. Atmos.* 2013, 118, 8411–8425. <https://doi.org/10.1002/jgrd.50371>.
- Xue, Y.H.; Li, J.; Menzel, P.; Borbas, E.; Ho, S.-P.; Li, Z. Impact of Sampling Biases on the Global Trend of Total Precipitable Water Derived from the Latest 10-Year Data of COSMIC, SSMIS and HIRS Observations. *J. Geophys. Res. Atmos.* 2018, 124, 6966–6981.
- Zhou, J., Ho, S.-P., Zhou, X., Shao, X., Gu, G., Chen, Y., et al. (2025). Construction of temperature climate data records in the upper troposphere and lower stratosphere using multiple RO missions from September 2006 to July 2023 at NESDIS/ STAR. *Journal of Geophysical Research: Atmospheres*, 130, e2024JD041295. <https://doi.org/10.1029/2024JD041295>
- Vorob'ev, V. V., and T. G. Krasil'nikova (1993), Estimation of the accuracy of the atmospheric refractive index recovery from Doppler shift measurements at frequencies used in the NAV-STAR system. *Phys. Atmos. Ocean*, **29**, 602–609.