

# The NOAA/STAR GNSS RO 1D-VAR Retrieval Algorithm V.3.2 Algorithm Theoretical Basis Document

NOAA/STAR GNSS RO Team  
2023-11-17

## *Prefix*

Compared with the Algorithm Theoretical Basis Document (ATBD) for the STAR RO 1D-VAR algorithm V.1.0, Section 6 is added in this document to describe features of version 3.2. The rest of the document is almost identical to ATBD for V.1.0.

## **1. Introduction**

The active global navigation satellite system (GNSS) radio occultation (RO) remote sensing technology can detect the vertical distribution of density variation. In the neutral atmosphere, the refractivity is a function of atmospheric pressure, temperature, and moisture [1-4]. With an optimal inversion algorithm and a priori atmospheric thermal information, the refractivity profiles can be inverted into temperature and water vapor profiles [5,6]. Numerous studies [7-27] have demonstrated the RO-derived water vapor profiles in the neutral atmosphere are complemented with those from satellite infrared (IR) and microwave (MW) sounders and provide water vapor information within and below clouds.

This ATBD aims to describe the one-dimension variational inversion algorithm (1D-VAR) to derive neutral atmospheric temperature and moisture profiles [28] developed by the NOAA Center for Satellite Applications and Research (STAR) GNSS RO data processing and science center (<https://www.star.nesdis.noaa.gov/smcd/gnssro/RO/index.php>). The input refractivity for the STAR 1D-VAR is from RO atmPrf files. Currently, we use RO atmPrf refractivity profiles from the UCAR COSMIC Data Analysis and Archive Center (UCAR/CDAAC) at <https://cdaac-www.cosmic.ucar.edu/cdaac/index.html>). In the future, we could use RO atmPrf files generated by the STAR GNSS RO team. The CDAAC implementation procedures for converting the COSMIC-2 raw measurements to bending angle and refractivity profiles are detailed at [https://cdaac-www.cosmic.ucar.edu/cdaac/doc/documents/Sokolovskiy\\_newroam.pdf](https://cdaac-www.cosmic.ucar.edu/cdaac/doc/documents/Sokolovskiy_newroam.pdf) and are not further described in this document. Description of the STAR 1D-VAR inversion approaches, pre-defined background covariance matrix and error covariance matrix, and initial validation are detailed in [28] and are also summarized below.

## **2. Dry and Wet Retrieval**

In a neutral atmosphere, the refractivity ( $N$ ) profile is a function of pressure ( $P$ ), temperature ( $T$ ), and the partial pressure of water vapor ( $P_W$ ) [2]:

$$N = 77.6 \frac{P}{T} + 3.73 \cdot 10^{-5} \frac{P_W}{T^2} \quad (1)$$

The units for  $P$ ,  $T$ , and  $P_W$  are mbar, Kelvin, and mbar, respectively. Refractivity is in N-units:  $N(z) = 10^6(n(z) - 1)$ , where  $n(z)$  is an index of refraction.

The refractivity information is contributed by temperature and moisture in the troposphere, where moisture is the dominant contributor (below 6-8 km altitude, depending on latitude and season of observation). In the upper troposphere and above, where the moisture is negligible, refractivity information mainly comes from the temperature.

Final pressure, temperature, and water vapor profiles are combinations of profiles obtained with “dry” and “wet” retrievals. Dry retrieval is applied for the upper portion of profiles where the contribution of the water vapor (second term in Equation (1)) into refractivity is negligible and can be omitted. Here, we use the dry hydrostatic equation and Equation (1) (with the second term removed) to solve the dry temperature and dry pressure (two equations to solve for two unknowns – a well-posed problem). The derived dry pressure and temperature shall be very close to the actual ones.

We use “wet” retrieval in the troposphere because the water vapor cannot be negligible. However, since we only have one observable (N) and two unknown, temperature and water vapor, the inversion becomes an ill-posed problem. We use the maximum likelihood method (the optimal estimator) introduced by [29] to retrieve temperature and water vapor from refractivity at specific altitudes. For a given RO refractivity value (the “observation”) at a particular height  $z$  (i.e.,  $Y_{OBS} = N(z)$ ), the optimal estimation equation is

$$X_{j+1} = X_0 + (K_j^T E^{-1} K_j + B^{-1})^{-1} \times K_j^T E^{-1} \{ (Y_{OBS} - Y(X_j)) + K_j (X_j - X_0) \}, \quad (2)$$

where  $X_0 = (T_0, P_{W_0})$  is the first guess,  $K = \left( \frac{\partial N}{\partial T}, \frac{\partial N}{\partial P_W} \right)$  is a Jacobian vector,  $B$  is the a priori background state covariance matrix, and  $E$  represents the combined instrument noise and the forward model-error covariance matrix.

Note that the retrieval vector  $X_j$  includes only temperature and water vapor, and the optimal estimator runs for each altitude separately, level by level, from the prescribed top of atmosphere (TOA) altitude to the bottom of the profile. The pressure profile requires integration from the TOA to the current altitude. The procedure to obtain a pressure profile is described separately in Section 4.

The index  $j$  in Equation 2 corresponds to the iteration number. The iteration procedures continue until the residual difference between the observed input refractivity and forward simulated refractivity (computed from Equation (1) using the retrieved temperature and water vapor as inputs) satisfies the condition

$$\left| \frac{N_{OBS} - N_j}{N_{OBS}} \right| < 0.001 \text{ (i.e., 0.1\%)}$$

A parameter  $z_{switch}$  is used to define the altitude where the switching from “dry” (for  $z > z_{switch}$ ) to “wet” (for  $z < z_{switch}$ ) retrieval happens. In Version 1.0,  $z_{switch} = 40 \text{ km}$ . The following sections describe each step of the algorithm in detail.

### 3. Detailed Algorithm Description

#### 3.1 The Algorithm Initialization

The algorithm initialization includes the following steps (also see [28]):

- Loading of the pre-computed background covariance matrix and error covariance matrix (used in optimal estimator);
- Loading of the First Guess (used in optimal estimator);
- Loading of the ocean/land mask ( $0.25^\circ$  latitude/longitude resolution; used only for transferring to the output).

#### 3.2 Pre-Computed Background Covariance Matrix and Error Covariance Matrix

The background covariance matrix  $B$  and the error covariance matrix  $E$  in Equation (2) must first be defined to run the optimal estimator. We used the NOAA Global Forecasting System 6 Hours Forecast (GFS-6hF) from 2018 to estimate the background covariance matrix. To take the seasonal and latitude variability of the Earth’s atmosphere into account, we evaluated  $B$  and  $E$  for seven latitude zones (two polar zones  $90^\circ\text{N}$  to  $60^\circ\text{N}$  and  $60^\circ\text{S}$  to  $90^\circ\text{S}$ ; four mid-latitude zones  $60^\circ\text{N}$  to  $45^\circ\text{N}$ ,  $45^\circ\text{N}$  to  $20^\circ\text{N}$ ,  $20^\circ\text{S}$  to  $45^\circ\text{S}$ , and  $45^\circ\text{S}$  to  $60^\circ\text{S}$ ; and one tropical zone  $20^\circ\text{N}$  to  $20^\circ\text{S}$ ) for each month of the year.

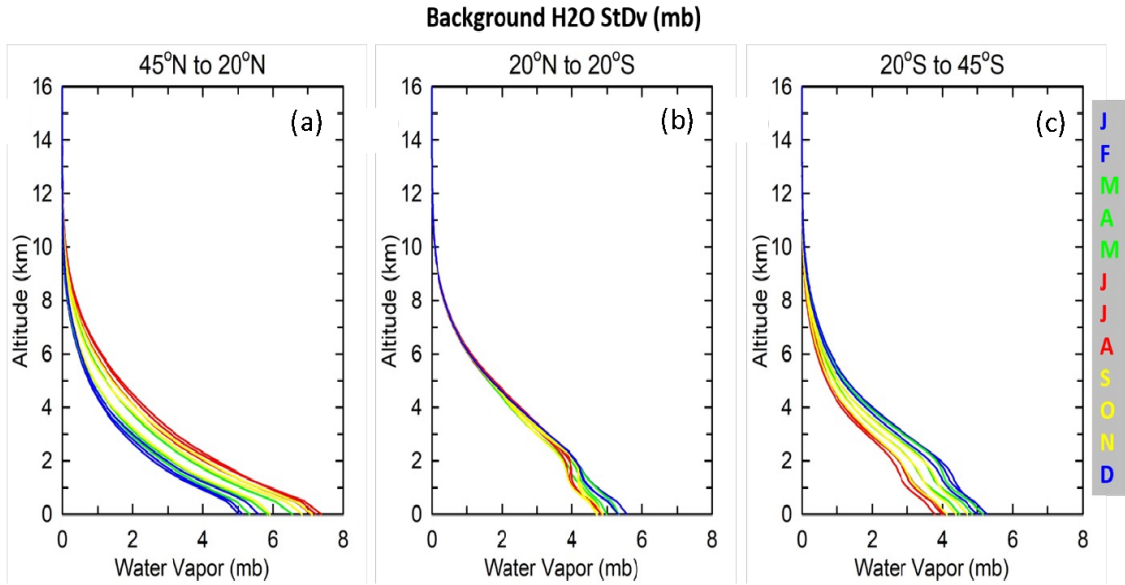
We have 84 ( $7 \times 12$ ) estimated background states from monthly mean profiles and standard deviation for atmospheric temperature and water vapor. Figures 1 and 2 present monthly standard deviations for water vapor partial pressure and temperature, which are used to define matrix  $B$ . In the covariance matrix  $B$ , the diagonal elements are equal to the square of the corresponding standard deviation, while the off-diagonal elements are all zero. Averaging has been performed for each month and in seven latitude zones separately. Only three latitude zones in the  $45^\circ\text{N}$  to  $45^\circ\text{S}$  range are shown since this is where COSMIC-2 measurements are located. Figures 1 and 2 show the atmospheric state's latitudinal and seasonal dependence, with the seasonal variability most vital in mid-latitudes and weakest in the tropics.

We applied the same year-long set of GFS background states to derive the set of simulated refractivity  $N$  by using a RO forward model (Equation 1). The current algorithm uses the error covariance matrix  $E$  to stabilize the optimal estimator (see Equation 2). Similar to the covariance matrix  $B$ , the error covariance matrix  $E$  has zero off-diagonal elements, and diagonal elements are proportional to the variance of  $N$  obtained for each month and each latitude zone:

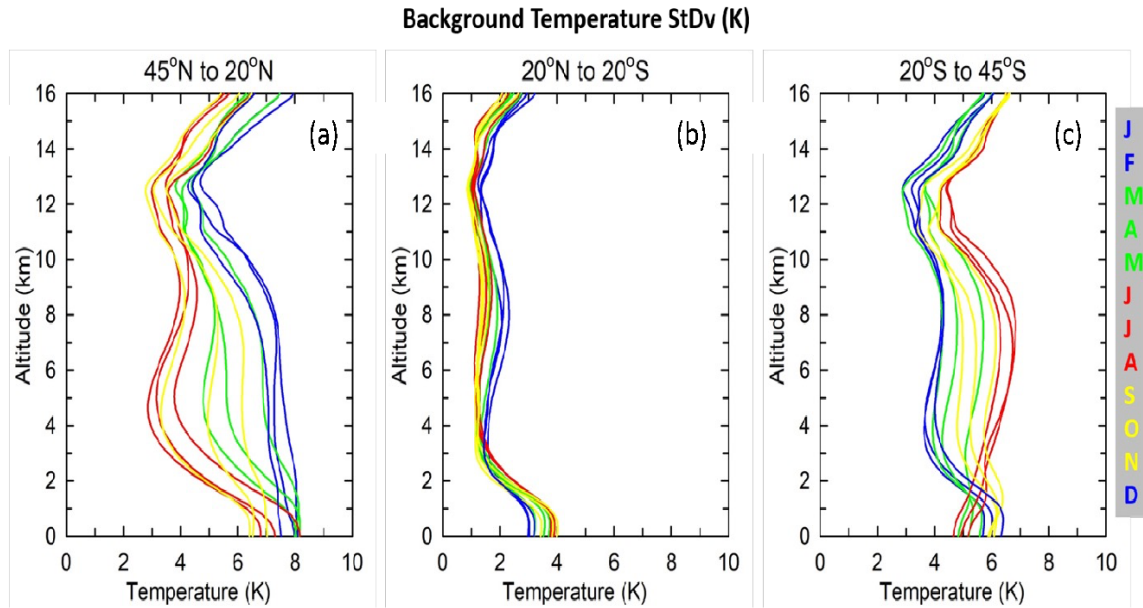
$$E_{ij} = \{\gamma_{Noise} \times \sigma(N_i)\}^2, \quad i = j$$

$$E_{ij} = 0, \quad i \neq j$$

The  $\gamma_{Noise}$  is a scaling factor, and we define  $\gamma_{Noise}=0.1$  to stabilize the matrix inversion and optimally use refractivity information in the retrieval. Our sensitivity study indicates that small  $\gamma_{Noise}$  leads to the instability of inversion, while high  $\gamma_{Noise}$  results in retrievals being close to the first guess and losing information obtained from measurements. This ensures the retrievals fit more to the observed refractivity than the first guess. Note that the error covariance described here does not reflect the combined error from the forward model and receiver but is designed to weight information obtained from the observation and first guess optimally.

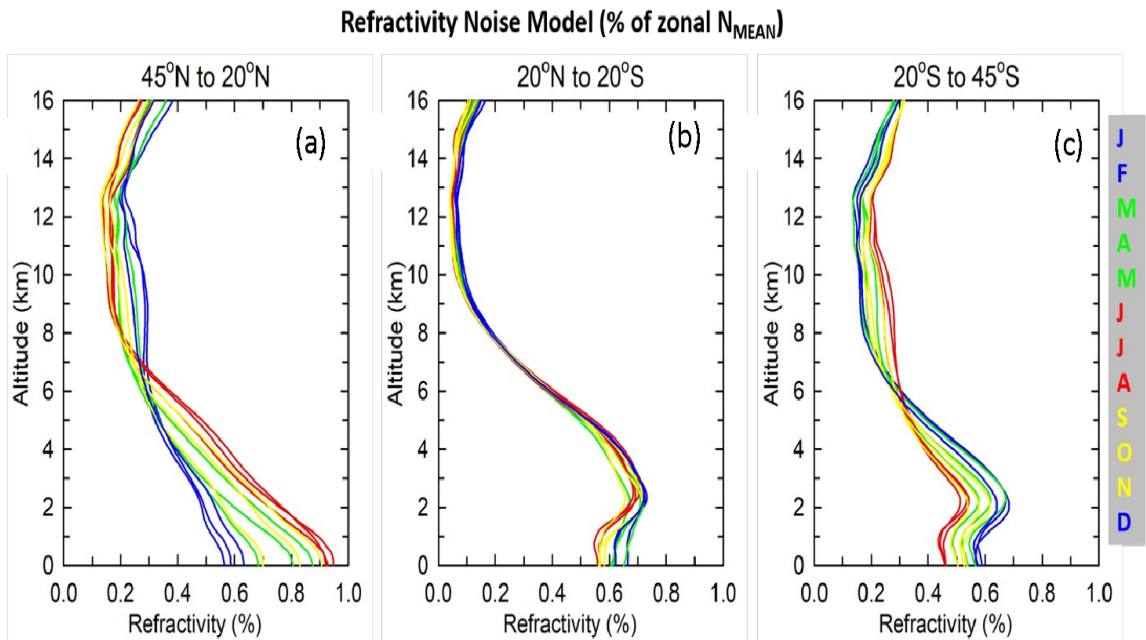


**Figure 1.** Background standard deviation of water vapor used in the optimal estimation: (a) northern mid-latitudes 45°N to 20°N, (b) tropics 20°N to 20°S, (c) southern mid-latitudes 20°S to 45°S. Each panel presents 12 months where blue lines are for northern winter (December, January, and February), green lines are for northern spring (March, April, and May), red lines are for northern summer (June, July, and August), and yellow lines are for northern autumn (September, October, and November).



**Figure 2.** The same as Figure 1 but for background temperature standard deviation.

Figure 3 shows the error covariance model used in the optimal estimations from 45°N to 45°S range. The estimated monthly errors are below 1% relative to the expected observed refractivity in all latitude zones below 16 km altitude.



**Figure 3.** The error covariance model (in % to zonal monthly mean refractivity) used in the optimal estimation: (a) northern mid-latitudes 45°N to 20°N, (b) tropics 20°N to 20°S, (c) southern mid-latitudes 20°S to 45°S. Each panel presents 12 months where blue lines are for northern winter (December, January, and February), green lines are for northern spring (March, April, and May), red lines are for northern summer (June, July, and August), and yellow lines are for northern autumn (September, October, and November).

### 3.3 The First Guess

In Version 1.0, the four-dimensional (latitude, longitude, fixed pressure levels, and time) global fields of the atmospheric temperature and water vapor are downloaded from the NOAA National Centers for Environment Information GFS website <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>. We use the NOAA Global Forecast (GFS) system data six hours forecast interpolated onto the RO locations and time as the a priori (first guess) atmospheric state (vector  $X_0$  in Equation (2)) to initialize the optimal estimation retrievals. We use GFS  $1^\circ \times 1^\circ$  latitude/longitude grid data. For a given day of observation, we have forecasts for 0:00, 6:00, 12:00, 18:00, and 24:00 UTC. The GFS fixed pressure levels grid varies depending on the time of interest (from 26 levels for the year of 2005 to 41 levels starting from the year of 2021) and includes pressure ranges from 1000 mbar to 1 mbar (in 2005) or 0.01 mbar (in 2021).

For a given RO profile, we first find the closest GFS profile location based on the nominal latitude/longitude from the RO profile, and two GFS forecast profiles are picked based on RO observation time, one before and one after the observation time. Then, we averaged these two profiles with weights proportional to the time difference between forecast time and the actual RO observation. Weighted profiles are lastly interpolated to the altitude grid of the input RO observation.

Starting from the NOAA/STAR GNSS RO 1D-VAR Version 3.2 package, we also support the ECMWF ERA5 six-hour forecast as the first guess. More details on using ERA5 can be found in Section 6.1 of this document.

### **3.4 The Main Input and Output**

The STAR GNSS RO 1D-VAR package is designed to run daily as a post-processing mode. The daily limitation is caused by the need to load the first guess, which takes a large volume of the computer's memory. Currently, we use UCAR atmPrf files (NetCDF format) to obtain input refractivity profiles. Full-length profiles of latitude and longitude and some other variables from atmPrf files are loaded for transferring to the main output file only (see Appendix A). Besides  $N(z)$ , the STAR GNSS RO 1D-VAR also uses impact height and altitude, the Earth's local curvature radius, nominal coordinates (latitude and longitude), and time of the RO observation from UCAR atmPrf files.

Input UCAR atmPrf profiles usually begin from  $\sim 60$  km of altitude to the surface with  $\sim 20$  m vertical resolution. The bottom altitude depends on the penetration depth achieved in the particular RO event and varies on average from 3 km to 0 km (the latter is less than 10% of the total number of profiles). The STAR GNSS RO 1D-VAR retrieval is performed on the input atmPrf altitude grid, i.e., with  $\sim 20$  m vertical resolution. When the particular profile is finished, the output profile is thinned to the fixed altitude grid similar to UCAR wetPf2, which is from 0 km (if available) to 20 km with a vertical resolution of 50 m and from 20 km to 60 km with a vertical resolution of 100 m. Thinning is performed as a linear interpolation by altitude.

The main output of the STAR GNSS RO 1D-VAR package is one file per RO event in NetCDF format, close to the one used in UCAR RO wetPrf and wetPf2 files. Appendix A provides a detailed description of the content and file name convention for the STAR wetPrf files.

## 4. Pressure Integration

### 4.1 Dry Pressure and Dry Temperature

Starting from the observed refractivity profile  $N(z)$ , the first step is to obtain a dry pressure profile  $P_{dry}(z)$ . Dry profiles are needed because actual pressure  $P(z)$  and temperature  $T(z)$  are very close and are equal to dry pressure and dry temperature, respectively, for altitudes  $z > z_{switch}$ .

The hydrostatic equation in the differential form to get  $P_{dry}(z)$  is as follows:

$$\frac{d \ln P_{dry}(z)}{dz} = \frac{-g(z)N(z)}{RkP_{dry}(z)} \quad (3)$$

Here  $z$  is altitude;  $g(z)$  – acceleration of gravity, computed as a function of geographic latitude and altitude above the reference ellipsoid derived from the Somigliana equation;  $N(z)$  – refractivity;  $R=287.05 \text{ J kg}^{-1} \text{ K}^{-1}$  – the dry air constant; and  $k=77.6 \text{ N-unit K hPa}^{-1}$  – the ideal gas refractivity constant.

The 4<sup>th</sup> order Runge-Kutta integration of Equation (3) is used to ensure accurate  $P_{dry}(z)$  values. The initial top value of dry pressure  $P_{dry}(z_{top})$  is needed to start the integration. One way to obtain  $P_{dry}(z_{top})$  (not used in Version 1.0 of the STAR GNSS RO 1D-VAR) is to calculate it as

$$P_{dry}(z_{top}) = \frac{-g(z_{top})N(z_{top})}{Rk\left(\frac{d \ln N}{dz}\right)}$$

which assumes the boundary condition that  $\frac{dT_{dry}}{dz} = 0$  at  $z_{top}$ . However, the vertical gradient of refractivity at the very top of the profile, computed from the actual input data, is unstable and cannot be estimated accurately. For this reason, we currently use the  $P_{dry}(z_{top})$  value, taken from the input atmPrf dry pressure profile.

Having a dry pressure profile, the dry temperature profile is derived from the simple relation

$$T_{dry}(z) = \frac{kP_{dry}(z)}{N(z)}$$

Above  $z_{switch}$ , retrieved pressure and temperature are equal to dry pressure and dry temperature, respectively, while water vapor partial pressure has an infinitesimally constant value ( $10^{-5}$  mbar).

### 4.2 Wet Pressure Retrieval

Below  $z_{switch} = 16 \text{ km}$  (Versions 1.0), where water vapor contribution into observed signal becomes noticeable, and increases while altitude goes down, the hydrostatic equation of state in moist air is used:

$$\frac{d \ln P(z)}{dz} = -\frac{g(z)}{RT_v(z)} \quad (4)$$

The equations for virtual temperature  $T_v(z)$  and conversion of water vapor partial pressure  $P_W(z)$  to specific humidity  $q(z)$  are [30]:

$$T_v(z) = T(z) \cdot (1 + 0.608q(z)) \quad (5)$$

$$q(z) = 0.622 \frac{P_W(z)}{P(z) - 0.378P_W(z)} \quad (6)$$

Equations (4-6) are well-known fundamental equations of atmospheric physics [30]. Here 0.622 is the ratio of water molar mass to air molar mass, and other numerical constants follow unit conversion from partial pressure-to-mixing ratio-to-specific humidity. Equation (4) is integrated using the same 4th-order Runge-Kutta method.

The problem with retrieving the atmospheric state from RO observations is that to get actual pressure  $P$  on altitude  $z$ , we need to know the temperature and water vapor at this level, but to know them, we need to know pressure first, as follows from Equations (4-6). To escape from this loop, the following approach has been implemented:

$$P(z) = P_{dry}(z) \text{ for } z \geq z_{switch}$$

$$T(z) = T_{dry}(z) \text{ for } z \geq z_{switch}$$

Starting from  $z_{switch}$  and below, the first guess pressure is calculated as

$$P_{FG}(z_i) = P_{rtr}(z_{i-1}) + \frac{g(z)P_{rtr}(z_{i-1})}{RT_{rtr}(z_{i-1})} \cdot |z_i - z_{i-1}|$$

Here, subscript ‘‘FG’’ means First Guess, and subscript ‘‘rtr’’ is used for values retrieved on a previous altitude level. Using  $P_{FG}$ , then  $T_{rtr}(z_i)$  and  $P_{W_{rtr}}(z_i)$  are obtained from the optimal estimation technique (described above). Having  $T_{rtr}(z_i)$  and  $P_{W_{rtr}}(z_i)$ , Equations (4-6) are used again to obtain  $P_{rtr1}(z_i)$ . Then optimal estimation runs again to update  $T_{rtr}(z_i)$  and  $P_{W_{rtr}}(z_i)$  using  $P_{rtr1}(z_i)$ . Finally,  $P_{rtr2}(z_i)$  is calculated using updated  $T_{rtr}(z_i)$  and  $P_{W_{rtr}}(z_i)$ .

As tests have shown, two runs of the optimal estimator are enough to obtain the stable value of wet pressure  $P$ . The check is performed by comparing the following values:

$$\max \left| \frac{P_{FG} - P_{rtr1}}{P_{rtr1}} \right| \sim 0.03 \%$$



$$\max \left| \frac{P_{rtr1} - P_{rtr2}}{P_{rtr2}} \right| < 0.005 \%$$

The maximum pressure difference values (0.03% and 0.005%) are taken over the whole altitude range for one month of COSMIC-2 processing (~100,000 profiles). It demonstrates that the final pressure is only a minimal adjustment from the  $P_{FG}$  value. It is also worth emphasizing that the wet pressure profile depends on retrieved values of temperature and water vapor below  $z_{switch}$ .

## 5. Quality Control

The STAR GNSS RO 1D-Var algorithm has thorough quality control (QC) of input data, processing, and output results. The total output files with retrieved atmospheric quantities can be less than the complete input atmPrf files. The reasons why the RO event (i.e., input atmPrf file) can be rejected are listed below:

1. The first guess files are not found.
2. The input atmPrf file has flag = “bad”.
3. Error in dry or wet pressure integration. It may occur if i)  $z_i - z_{i-1} \geq 100 \text{ m}$  while it is expected that  $z_i - z_{i-1} < 0$  (level numbering goes from TOA to the surface), and/or ii) pressure at the particular level is negative.
4. Interpolation error. The error may occur if i) the x-value is beyond the x-range but extrapolation is not allowed, ii) the x or y value is negative but logarithmic interpolation is required. The interpolation is widely used over the whole processing, for example, for the first guess, background covariance matrices, and output profiles thinning.
5. The number of successfully retrieved levels is too small (currently, less than half of the input refractivity profile length).
6. we also mark the side viewing angle between 60 degrees and 120 degrees as “bad” since they are of high uncertainty.
7. L1 Signal-Noise-Ratio less than 300 V/V as “bad” also because they are of high uncertainty.

In all other cases, the corresponding output wetPrf file will be produced. It has three QC variables:

1. Scalar integer variable **Overall\_retrieval\_quality** with a value range from zero to five. Zero value means that retrieval of the complete profile is successful. Values from 1 to 5 indicate that when thinning to the output altitude grid is performed, the altitude gap is more than 0.5, 1.0, 1.5, 2.0, and 2.5 km, respectively. The altitude gap is caused by too many consecutive levels where the retrieval has failed, i.e., N-residual convergence was not achieved.
2. Character one-symbol-length variable **bad** can be “0” or “1”. Zero means successful retrieval; bad=“1” when Overall\_retrieval\_quality>0.
3. Integer vector **QC\_lev** that has the same length as output atmospheric profiles and may be equal to 0 or 1 for each level. It specifies retrieval quality for each level on the output altitude grid. Zero means that retrieved quantities for this level were obtained with

interpolation over an altitude gap wider than 0.5 km, or interpolation has failed (i.e., retrievals are probably bad).  $QC\_lev(i)=1$  means that thinning was successful for this level.

It is worth reminding that actual retrievals are performed at atmPrf altitude grid (i.e., with high vertical resolution  $\sim 20$  m), while output retrieved profiles have a fixed altitude grid from 0 km to 60 km with a resolution of 50 m below 20 km and 100 m above 20 km. An internal QC vector, similar to QC\_lev but for atmPrf levels, controls the retrieval quality over the whole processing. This internal “high resolution” quality vector determines all output QC variables.

## **6. Main differences between the NOAA/STAR GNSS RO 1D-VAR Retrieval Algorithms version 1.0 and versions 3.2**

The NOAA/STAR RO team has released the version 3.2 of the RO 1D-Var algorithm in 2023. Basic physical concepts, described in Sections 2-4 of this document, remain the same in V.3.2 as in V.1.0. This section focuses on what is different in the versions.

### **6.1 The first guess**

Version 3.2 uses ECMWF ERA5 six-hour forecast as the default first guess. The option to use GFS is still kept in the package for users to choose. We use ERA5 six-hour forecast global datasets at  $0.25^\circ \times 0.25^\circ$  latitude/longitude grid in three moments, including 0:00, 12:00, and 24:00 UTC. ERA5 fixed pressure levels grid has 37 levels ranging from 1000 mbar to 1 mbar. The interpolation of ERA5 profiles to the actual RO observation (by location and time) is performed similarly to the GFS.

### **6.2 Switching from dry to wet retrieval**

Beginning from version 3.0, we use static switching from dry to wet retrieval instead of dynamic one, described in Section 4.2. It occurs at the predefined altitude  $z_{switch}$ . The input parameter  $z_{switch}$  is equal to 25 km for v.3.2 and 40 km for v.3.0 and v.3.1.

### **6.3 Output quality control**

Output quality control (QC) is based on the size of the altitude gap where the optimal estimator has failed (see Section 5). However, internal changes in the procedure affect output atmospheric profiles. Before version 3.1, not-retrieved levels (atmPrf grid) were marked as bad, and only pressure integration was continued through in these levels using first guess values for temperature and water vapor. The validation of versions 3.0 and 3.1, both configured with  $z_{switch} = 40$  km, has shown that for certain missions (specifically COSMIC-1 and SPIRE), this method leads to a significant issue with approximately 10% of a total number of problematic profiles. These profiles have substantial data gaps, predominantly within the 30-40 km altitude range. However, the COSMIC-2 mission does not have such a problem. Unsuccessful retrievals in this region can be attributed to the challenges posed by low signal-to-noise ratios and low sensitivity of refractivity to atmospheric temperature and water vapor.

Beginning with version 3.2, the dry/wet switching altitude has been adjusted to 25 km to address these challenges. Any failed levels are also filled with first-guess temperature and water vapor values. This adjustment ensures that all levels are adequately populated before the thinning process. The thinning process has been improved, which involves projecting atmospheric profiles from the atmPrf grid to the output wetPrf grid. One particular enhancement includes a sliding averaging with  $\pm 40 m$  window to effectively smooth output profiles and incorporate more information from the finer retrieval altitude grid into the coarser output altitude grid, surpassing the capabilities of simple interpolation.

Furthermore, an additional quality check involves identifying and flagging profiles with significant deviation from the first guess values. The threshold values are set at a sufficiently broad range, exceeding 50 K for temperature and/or 50hPa for water vapor partial pressure. The primary objective of this quality control measure is to exclude these obvious outliers from the dataset. Validation conducted over one month of data from COSMIC-1, COSMIC-2, and SPIRE missions has confirmed that, on average, such outliers occur less than once per day.

#### 6.4 Other features of the STAR RO 1D-VAR Retrieval Algorithm version 3.2

The distinguishing feature of the STAR RO 1D-VAR algorithm version 3.0 is its capability to detect and handle Super-Refraction (SR) cases. It's important to note that this algorithm aspect is currently under development, and a comprehensive description is presented in a separate ATBD. Version 3.0 of the software introduces a range of enhancements beyond the extensive updates in the SR processing. These improvements also include internal modifications to the input/output system, designed to optimize computer memory utilization and reduce the computational time required for stable operation of the RO 1D-Var algorithm in various computing environments. These changes enhance the efficiency and reliability of the software across different computing setups.

Table 1 lists the key features and notable enhancements introduced in different STAR RO 1D-Var retrieval algorithm versions.

**Table 1. Summary of key features and enhancements in STAR RO 1D-Var versions.**

Version	Main Feature and Enhancements	Comments
<b>v.1.0</b> Mar 2021	<ol style="list-style-type: none"> <li>1. Input: refractivity from atmPrf file</li> <li>2. First guess: GFS 6 hours forecast</li> <li>3. Composition of DRY and WET retrievals</li> <li>4. Dry retrieval above 16 km</li> <li>5. Dynamic switch from DRY to WET retrieval: starting from 16 km down, check <math>\Delta P =  P_{wet} - P_{dry} </math>: dry retrieval is continued until <math>\Delta P &lt; 0.001 hPa</math>, otherwise switch to wet retrieval.</li> </ol>	<u>Original version</u> The estimated value of dynamic switching altitude is $11.5 \pm 1$ km (stat over a couple of months of COSMIC-2 processing)
<b>v.2.0</b> Mar 2022	<ol style="list-style-type: none"> <li>1. Input: bending angle (BA) from atmPrf file</li> <li>2. The Abel inversion routine to obtain a refractivity profile is added</li> </ol>	<u>Note:</u> SR detection during Abel transform works if BA is computed with the STAR RO Forward Model. BA from atmPrf

	<p>3. Abel integration procedure includes the option to detect SR height location in cases when not-monotonic dependence <math>n(p)</math> is obtained</p> <p>4. SR refractivity correction is introduced as a user's option</p>	files always provide monotonic dependence $n(p)$
<b>v.2.1</b> Dec 2022	1. Input: get back to the refractivity profile from the atmPrf file	<u>Reason:</u> the top portion of atmPrf BA cannot be used for BA extrapolation above atmPrf top altitude, which is needed for proper Abel integration. External sources for BA modeling above 60 km are necessary.
<b>v.3.0</b> Mar 2023	<p>1. Input: refractivity from atmPrf file</p> <p>2. The option to use ERA5 6HF as the first guess is added</p> <p>3. Static switch from dry to wet retrieval, <math>z_{switch} = 40 \text{ km}</math></p> <p>4. Numerous updates in the SR procedure:</p> <ul style="list-style-type: none"> <li>• Dichotomy approach to evaluate top SR layer altitude</li> <li>• “Cut the Bubble” technique is introduced</li> <li>• Additional criteria to qualify RO events as “potential SR” and “confirmed SR” are added</li> </ul>	
<b>v.3.1</b> Jul 2023	<p>1. The function to output STAR wetPrf files is moved to the separate task</p> <p>2. “Cut the Bubble” technique is updated (SR correction)</p>	<u>Reason:</u> to avoid fatal crashes of the package at some servers due to memory insufficiency
<b>v.3.2*</b> Aug 2023	<p>1. Adjusted dry/wet switching altitude to 25 km</p> <p>2. Quality control checks for temperature and water vapor</p> <p>3. Filling failed levels with first guess values for temperature and water vapor</p> <p>4. Enhanced thinning procedure with sliding averaging</p>	<u>Reason:</u> high (=40 km) switching altitude results in fewer successful retrievals for some RO missions due to instability in the 30-40 km range.

\*We only released data generated by Version 1.0 and Version 3.2. The other versions between V.1.0 and V.3.2 are for user information. Note that the released 1D-Var-V3.2 data are processed with STAR Version 3.2 1D-Var package, which includes all the features implemented in the previous versions and uses ERA5 6HF as the first guess, but without SR correction. The approach to identifying SR and correcting the water vapor profiles for the SR layers is under development and is not detailed in this ATBD.

To use the STAR 1D-VAR data, please cite

Ho, S.-P.; Kireev, S.; Shao, X.; Zhou, X.; Jing, X. Processing and validation of the STAR COSMIC-2 temperature and water vapor profiles in the neutral atmosphere. *Remote Sens.* **2022**, *14*, 5588. <https://doi.org/10.3390/rs14215588>.

## 7. References

1. Ho, S.-P.; Goldberg, M.; Kuo, Y.-H.; Zou, C.-Z.; Shiau, W. Calibration of Temperature in the Lower Stratosphere from Microwave Measurements Using COSMIC Radio Occultation Data: Preliminary Results. *Terr. Atmos. Ocean. Sci.* **2009**, *20*, 87.

2. Bean, B.R.; Dutton, E.J. *Radio Meteorology. National Bureau of Standards Monogr., 92*; U.S. Government Printing Office: Washington, DC, USA, 1966; p. 435.
3. Ho, S.-P.; Zhou, X.; Kuo, Y.-H.; Hunt, D.; Wang, J.-H. Global Evaluation of Radiosonde Water Vapor Systematic Biases using GPS Radio Occultation from COSMIC and ECMWF Analysis. *Remote Sens.* **2010**, *2*, 1320–1330.
4. Ho, S.-P.; Kuo, Y.-H.; Schreiner, W.; Zhou, X. Using SI-traceable global positioning system radio occultation measurements for climate monitoring In "State of the Climate in 2009". *Bull. Am. Meteorol. Soc.* **2010**, *91*, S36–S37.
5. Healy, S.; Eyre, J. Retrieving temperature, water vapor, and surface pressure information from refractivity-index profiles derived by radio occultation: A simulation study. *Q. J. Royal Meteorol. Soc.* **2000**, *126*, 1661–1683.
6. Li, Y.; Kirchengast, G.; Scherllin-Pirscher, B.; Schwaerz, M.; Nielsen, J.K.; Wee, T.-K.; Ho, S.-P.; Yuan, Y.-B. A new algorithm for the retrieval of atmospheric profiles from GNSS radio occultation data in moist air and cross-evaluation among processing centers. *Remote Sens.* **2019**, *11*, 2729.
7. Ho, S.-P.; Smith, W.L.; Huang, H.L. The Retrieval of Atmospheric Temperature and Water Vapor Profile using Combined Satellite and Ground Based Infrared Spectral Radiance Measurements. *Appl. Opt.* **2002**, *41*, 4057–4069.
8. Ho, S.P.; Kuo, Y.H.; Sokolovskiy, S. Improvement of the temperature and moisture retrievals in the lower troposphere using AIRS and GPS radio occultation measurements. *J. Atmos. Ocean. Technol.* **2007**, *24*, 1726–1739. <https://doi.org/10.1175/JTECH2071.1>.
9. Ho, S.P.; Yue, X.; Zeng, Z.; Ao, C.O.; Huang, C.Y.; Kursinski, E.R.; Kuo, Y.H. Applications of COSMIC radio occultation data from the troposphere to ionosphere and potential impacts of COSMIC-2 data. *Bull. Am. Meteorol. Soc.* **2014**, *95*, ES18–ES22. <https://doi.org/10.1175/bams-d-13-00035.1>.
10. Ho, S.-P.; Peng, L.; Anthes, R.A.; Kuo, Y.-H.; Lin, H.-C. Marine Boundary Layer Heights and Their Longitudinal, Diurnal, and Interseasonal Variability in the Southeastern Pacific Using COSMIC, CALIOP, and Radiosonde Data. *J. Clim.* **2015**, *28*, 2856–2872. <https://doi.org/10.1175/jcli-d-14-00238.1>.
11. Ho, S.-P.; Peng, L.; Vömel, H. Characterization of the long-term radiosonde temperature biases in the upper troposphere and lower stratosphere using COSMIC and Metop-A/GRAS data from 2006 to 2014. *Atmospheric Chem. Phys.* **2017**, *17*, 4493–4511, <https://doi.org/10.5194/acp-17-4493-2017>.
12. 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.* **2018**, *18*, 259–274. <https://doi.org/10.5194/acp-18-259-2018>.
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, 2018. <https://doi.org/10.5772/interchopen.79541>.
14. Ho, S.-P.; Anthes, R.A.; Ao, C.O.; Healy, S.; Horanyi, A.; Hunt, D.; Mannucci, A.J.; Pedatella, N.; Randel, W.J.; Simmons, A. The COSMIC/FORMOSAT-3 Radio Occultation Mission after 12 Years: Accomplishments, Remaining Challenges, and Potential Impacts of COSMIC-2. *Bull. Amer. Meteor. Soc.* **2020**, *101*, E1107–E1136. <https://doi.org/10.1175/BAMS-D-18-0290.1>.

15. Ho, S.-P.; Zhou, X.; Shao, X.; Zhang, B.; Adhikari, L.; Kireev, S.; He, Y.; Yoe, J.; Xia-Serafino, W.; Lynch, E. Initial Assessment of the COSMIC-2/FORMOSAT-7 Neutral Atmosphere Data Quality in NESDIS/STAR Using In Situ and Satellite Data. *Remote Sens.* **2020**, *12*, 4099. <https://doi.org/10.3390/rs12244099>.
16. Huang, C.; Teng, W.; Ho, S.; Kuo, Y. Global variation of COSMIC precipitable water over land: Comparisons with ground-based GPS measurements and NCEP reanalyses. *Geophys. Res. Lett.* **2013**, *40*, 5327–5331. <https://doi.org/10.1002/grl.50885>.
17. 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>.
18. Biondi, R.; Randel, W.J.; Ho, S.-P.; Neubert, T.; Syndergaard, S. Thermal structure of intense convective clouds derived from GPS radio occultations. *Atmospheric Chem. Phys.* **2012**, *12*, 5309–5318. <https://doi.org/10.5194/acp-12-5309-2012>.
19. Biondi, R.; Ho, S.-P.; Randel, W.; Syndergaard, S.; Neubert, T. Tropical cyclone cloud-top height and vertical temperature structure detection using GPS radio occultation measurements. *J. Geophys. Res. Atmos.* **2013**, *118*, 5247–5259. <https://doi.org/10.1002/jgrd.50448>.
20. 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.
21. Zeng, Z.; Ho, S.-P.; Sokolovskiy, S. The Structure and Evolution of Madden-Julian Oscillation from FORMOSAT-3/COSMIC Radio Occultation Data. *J. Geophys. Res.* **2012**, *117*, D22108. <https://doi.org/10.1029/2012JD017685>.
22. Schröder, M.; Lockhoff, M.; Shi, L.; August, T.; Bennartz, R.; Brogniez, H.; Calbet, X.; Fell, F.; Forsythe, J.; Gambacorta, A.; et al. The GEWEX water vapor assessment: Overview and introduction to results and recommendations. *Remote Sens.* **2018**, *11*, 251. <https://doi.org/10.3390/rs11030251>.
23. Mears, C.; Ho, S.P.; Wang, J.; Huelsing, H.; Peng, L. Total Column Water Vapor [In “States of the Climate in 2018“]. *Bull. Amer. Meteor. Soc.* **2019**, *98*, S24–S25. <https://doi.org/10.1175/2019BAMSStateoftheClimate.1>.
24. Mears, C.; J. Wang, J.; Ho, S.P.; Zhang, L.; Zhou, X. Total Column Water Vapor [In “States of the Climate in 2020“]. *Bul. Amer. Meteor. Sci.* **2021**, *in press*.
25. Rieckh, T.; Anthes, R.; Randel, W.; Ho, S.-P.; Foelsche, U. Tropospheric dry layers in the tropical western Pacific: comparisons of GPS radio occultation with multiple data sets. *Atmospheric Meas. Tech.* **2017**, *10*, 1093–1110. <https://doi.org/10.5194/amt-10-1093-2017>.
26. Rieckh, T.; Anthes, R.; Randel, W.; Ho, S.-P.; Foelsche, U. Evaluating tropospheric humidity from GPS radio occultation, radiosonde, and AIRS from high-resolution time series. *Atmospheric Meas. Tech.* **2018**, *11*, 3091–3109. <https://doi.org/10.5194/amt-11-3091-2018>.
27. Liu, C.-Y.; Li, J.; Ho, S.-P.; Liu, G.-R.; Lin, T.-H.; Young, C.-C. Retrieval of Atmospheric Thermodynamic State From Synergistic Use of Radio Occultation and Hyperspectral Infrared Radiances Observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 744–756. <https://doi.org/10.1109/jstars.2015.2444274>.
28. Ho, S.-P.; Kireev, S.; Shao, X.; Zhou, X.; Jing, X. Processing and validation of the STAR COSMIC-2 temperature and water vapor profiles in the neutral atmosphere. *Remote Sens.* **2022**, *14*, 5588. <https://doi.org/10.3390/rs14215588>.

29. Rodgers, C.D. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.* **1976**, *14*, 609–624.
30. Salby, M.L.: Fundamentals of Atmospheric Physics. *Academic Press*, San Diego, **1996**.

## Appendix A

# The STAR GNSS Radio Occultation 1D-VAR Version 1.0

## Data Product User Guide

NOAA/STAR GNSS RO Team

2023-01-31

### 1. Introduction

The STAR Global Navigation Satellite System (GNSS) Radio Occultation (RO) one-dimension variational inversion algorithm (1D-VAR) software package is developed to retrieve atmospheric pressure, temperature, and water vapor from the refractivity obtained in RO observations. Input refractivity is taken from RO atmPrf files. RO atmPrf file includes RO bending angles, impact parameters, tangent point location, quality information, as well as refractivity. Currently, we use RO atmPrf generated by University Corporation for Atmospheric Research (UCAR) operational data processing center and distributed by COSMIC Data Analysis and Archive Center (CDAAC). In the future, we could use RO atmPrf files generated by the STAR GNSS RO team. The main output is collected in NetCDF files (one for each RO measurement). To be consistent with the name conversion from other RO processing centers, the output file name has a “wetPrf\_” prefix and “\_nc” suffix. The following sections provides more details on file name convention and its content.

### 2. File Name Convention

Output file name after the STAR GNSS RO-1D-VAR processing is generated using the following template:

**FileName** = “wetPrf\_*FileStamp*\_STAR.Vn.n\_nc”

Red color marks unchangeable parts of the FileName: “wetPrf” is a type of output, “STAR” is a processing center (NOAA/NESDIS/STAR), “Vn.n” is the software version number, and “\_nc” is for the file extension (NetCDF format).

**FileStamp** is a string with 23 symbols in the form “XXXX.yyyy.doy.hh.mm.gns”. Here “XXXX” is a mission identifier, which depends on the missions (see Table A.1 for different GNSS receiver satellite missions). Values of “yyyy”, “doy”, “hh”, and “mm” are year, Julian day of the year, hour, and minute of the RO observation, respectively. The “gns” is the GNSS satellite identifier, with the first letter as “G” for GPS, “R” for GLONASS, “E” for Galileo, or “C” for BeiDou, and the rest two are two-digit GNSS satellite number.

Table A.1. GNSS receiver satellite missions

<b>XXXX</b>	<b>Meaning</b>
C0nn	COSMIC-1, nn: 01-06, FM1 to FM6
C2En	COSMIC-2, n: 1-6, FM 1 to FM 6
KOM5	KOMPSAT5
MTPn	MetOp, n: A, B, or C
PAZn	PAZ, n: 1
Snnn	Spire, nnn: three-digit satellite ID
GOnn	GeoOptics, nn: two-digit satellite ID

### 3. Output File Content



### 3.1 Dimensions

STAR GNSS RO-1D-VAR output file has only one dimension variable, named MSL\_alt, which is the length (number of levels) for the output profiles.

### 3.2 Attributes

STAR RO-1D-VAR output file attribute variables are presented in Table A.2:

Table A.2. STAR RO-1D-VAR output NetCDF file attribute variables:

Name	Meaning
fileStamp	File Stamp: explained in Section 1 of this User Guide
year	Year of the RO observation
month	Month of the RO observation
day	Day of the month of the RO observation
hour	Hour of the RO observation
minute	Minutes of the RO observation
second	Seconds of the RO observation
DOY	Julian day of year of the RO observation
date	Date of the RO observation in format “yyyy-mm-dd hh:mm:ss.ssss”
atmPrf	Name of the input atmPrf file
fgsUsed	Description of the atmospheric model used as the First Guess in RO-1D-VAR; can be “ERA5” or “GFS”
lat	Nominal latitude (degrees North, -90:+90)
lon	Nominal longitude (degrees East, -180:+180)
landmask	“0”=Ocean; ‘1’=Land
Overall_retrieval_quality	Quality flag of the STAR RO-1D-VAR processing (“0” means “good”, otherwise “bad”)
H_switch	The altitude (unit=km) where switching from dry retrieval to wet retrieval happens
SR_check	v.3.0 and later: 0/1=do not/do SR processing
SR_Flag	v.3.0 and later: 0/1=case is not/is qualified as a “potential SR case”
SR_Flag_Corr	v.3.0 and later: 0/1=case is not/is qualified as a “confirmed SR case”
SR_h1	v.3.0 and later: bottom altitude of the SR Shadow layer, units=[km]
SR_h2	v.3.0 and later: top altitude of the SR Shadow layer, units=[km]
SR_h3	v.3.0 and later: top altitude of the SR layer, units=[km]
atmPrf_stdv	Copy of the attribute “stdv” from input atmPrf file
atmPrf_snr1avg	Copy of the attribute “snr1avg” from input atmPrf file
atmPrf_snr2avg	Copy of the attribute “snr2avg” from input atmPrf file
atmPrf_irs	Copy of the attribute “irs” from input atmPrf file
atmPrf_balmax	Copy of the attribute “balmax” from input atmPrf file
atmPrf_zbalmax	Copy of the attribute “zbalmax” from input atmPrf file
atmPrf_freq1	Copy of the attribute “freq1” from input atmPrf file

atmPrf_freq2	Copy of the attribute “freq2” from input atmPrf file
atmPrf_bad	Copy of the attribute “bad” (quality flag) from input atmPrf file
bad	Quality flag of the STAR RO-1D-VAR processing (“0” means “good”, otherwise “bad”)
version	STAR RO-1D-VAR software version number
center	Name of the processing center
NCProperties	NetCDF software version used to make file

### 3.3 Output Profiles

The collection of output profiles after STAR GNSS RO-1D-VAR processing is presented in Table A.3. All profiles are one-dimensional arrays of length MSL-alt.

Table A.3. STAR GNSS RO-1D-VAR output NetCDF file profiles:

Name	Type	Units	Range	Meaning
MSL_alt	float	km	0.0 to 60.0	Mean sea level altitude of perigee point
QC_lev	integer	n/a	0 or 1	Retrieval quality flag by level: 0=bad, 1=good
lat	float	degrees	-90 to +90	Latitude of perigee point
lon	float	degrees	-180 to +180	Longitude of perigee point
Temp	float	Celsius	-200 to 100	Retrieved temperature
Pres	float	mbar	0 to 1200	Retrieved pressure
Vp	float	mbar	0 to 100	Retrieved water vapor partial pressure
sph	float	g/kg	0 to 100	Computed specific humidity
rh	float	%	0 to 100	Computed relative humidity
ref	float	N-units	0 to 500	STAR thinning observed or corrected refractivity
temp_dry	float	Celsius	-200 to +100	Retrieved dry temperature
pres_dry	float	mbar	0 to 1200	Retrieved dry pressure
Temp_1gs	float	Celsius	-200 to +100	Used First Guess temperature
Vp_1gs	float	mbar	0 to 100	Used First Guess Water vapor partial pressure