Data Release Notes for Airborne/Balloon-borne GNSS Radio Occultation Observations (Version IV)

Bing Cao, Jennifer S. Haase, Michael J. Murphy, Pawel Hordyniec

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA

July 10, 2025

Contents

1 Version Notes of this Document

- First version: 2019-04-19
- Second version: 2020-05-04
- Third version: 2020-09-30
- Fourth version: 2021-06-25
- Fourth version with clarifications: 2022-07-09
- Fifth version: 2023-04-19

2 Use of the Data

The Strateole-2 balloon-borne Radio Occultation data is freely and openly available to all users, with the requirement that the following acknowledgment is included in any use or publication of the data or derived products, and the following publications are cited:

"Development of the ROC instrument and data collection were supported by National Science Foundation award 1642650. The data were collected by Principal Investigator Dr. Jennifer S. Haase and her research team at Scripps Institution of Oceanography. Infrastructure for the balloon campaign was provided by NSF and the National Centre for Space Studies (CNES) and the National Centre for Scientific Research, Dynamic Meteorology Laboratory (LMD). CNES provided the balloons, flight control systems, flight control gondola 'Euros' and Flight Control Center. LMD provided the science payload gondola 'Zephyr' which carried the scientific instruments including ROC, as well as the science data transmission equipment, and provided the associated Mission Control Center (MCC) for monitoring and control of the instruments and the dataset collection."

Cao, B., J. S. Haase, M. J. Murphy, M. J. Alexander, M. Bramberger and A. Hertzog (2022), Equa- torial waves resolved by balloon-borne Global Navigation Satellite System radio occultation in the Strateole-2 campaign, Atmospheric Chemistry and Physics, 22, 15379–15402, doi:10.5194/acp-22-15379-2022

Haase, J. S., M. J. Alexander, A. Hertzog, L. Kalnajs, T. Deshler, S. M. Davis, R. Plougonven, P. Cocquerez, and S. Venel (2018), Around the world in 84 days - Strateole-2 investigates the tropical atmosphere with long duration superpressure balloons, EOS, 99.

The AR Recon airborne Radio Occultation data is freely and openly available to all users, with the requirement that the following acknowledgment is included in any use or publication of the

data or derived products, and the following publications are cited:

"AR Recon Airborne Radio Occultation instrumentation development and data collection by Principal Investigator Dr. Jennifer S. Haase and her research team at Scripps Institution of Oceanography were supported by National Aeronautics and Space Administration Grant NNX15AU19G, National Science Foundation award 1642650, and funding from the Center for Western Weather and Water Extremes Atmospheric Rivers Program and its sponsors, including the California Department of Water Resources and the Army Corps of Engineers. Aircraft flights were provided by the National Oceanic and Atmospheric Administration Aircraft Operations Center."

Haase, J. S., Murphy, M. J., Cao, B., Ralph, F. M., Zheng, M., and Delle Monache, L.: Multi-GNSS Airborne Radio Occultation Observations as a Complement to Dropsondes in Atmospheric River Reconnaissance, Journal of Geophysical Research: Atmospheres, 126, e2021JD034865, https://doi.org/10.1029/2021JD034865, 2021JD034865, 2021.

Ralph, F. M., F. Cannon, V. Tallapragada, C. A. Davis, J. D. Doyle, F. Pappenberger, A. Subramanian, A. M. Wilson, D. A. Lavers, C. A. Reynolds, J. S. Haase, L. Centurioni, B. Ingleby, J. J. Rutz, J. M. Cordeira, M. Zheng, C. Hecht, B. Kawzenuk, and L. Delle Monache (2020), West Coast Forecast Challenges and Development of Atmospheric River Reconnaissance, Bulletin of the American Meteorological Society, 101, E1357.

3 Acronym

- **GNSS** Global Navigation Satellite System
- **GPS** Global Positioning System
- **ARO/RO** Airborne Radio Occultation/Radio Occultation
- **GO** Geometric Optics
- **PM** Phase Matching
- **FSI** Full Spectrum Inversion
- **PPP** Precise Point Positioning
- **IMU** Inertial Measurement Unit
- **CDAAC** COSMIC Data Analysis and Archive Center
- **MSL** Mean Sea Level

- **EGM96** Earth Gravitational Model 1996
- **EGM2008** Earth Gravitational Model 2008
- **WGS84** World Geodetic System 1984
- **WMO** World Meteorological Organization

4 Version Control

Format Version (SSSS)
0001 = Murphy format for PREDICT dataset
0002 = Format version of data release from 2019-04-19
0003 = Format version of data release from 2020-05-05
0003 = (Same #) Format version of data release from 2023-04-19 with minor updates for some global attributes

Software version (VVVV)

00X1 = Brian Murphy, closed_loop, do_aro (provide date)
00X1 = Jeff Sussmen, closed_loop, do_aro

02X4 = Brian Murphy, open_loop, geometric optics (provide date)
02X6 = Brian Murphy, open_loop, phase matching (2013-09-01)
02X7 = Feiqin Xie, open_loop, full spectrum inversion (provide date)
Notes: the above legacy versions are not used in AR Recon and Strateole-2 data processing, and X is the not-assigned people identifier.

002Y (00 = close loop, 2 = Bing Cao, Y=1,2,3...7, version)
0021 = Bing Cao of 2019-04-15 closed loop, do_aro, doppler correction, geometric optics
0022 = Bing Cao of 2020-05-04, closed loop, do_aro, doppler correction, geometric optics
0023 = Bing Cao of 2020-09-30, closed loop, pppar, doppler correction, geometric optics
0024 = Bing Cao of 2021-06-25, closed loop, pppar, doppler correction, geometric optics
0025 = A test version that exists only for a short time, no data was processed by it.
0026 = Bing Cao of 2023-04-19, closed loop, pppar, doppler correction, geometric optics with updates and folder reorganization
Notes: The main change among those versions is the Doppler smoothing method.
Notes: The folder "code_30026", 30026 = 0003.0026 = format version 0003 and software version 0026. Starting from 0026, the processing folder structure was reorganized.
0027 = Bing Cao planned next version in Python.

003Z (00 = close loop, 3 = Kate Lord, Z=6, 7... version)
0036 = Kate Lord as of 2024-02-28, closed loop, pppar, phase matching, starting with version 0026 from Bing Cao.
0037 is Kate's next version and now has nothing to do with Bing's 0027. (Branch)

5 Data Releases

2023-04-19 AR Recon 2023 airborne RO data release for 2022.253 to 2023.032
processing version: V.0003.0026.
2021-06-25 Strateole2 balloon-borne RO data release for dates 2019.340-356, 2021.001, and partial dates 2021.009-013
processing version: V.0003.0024.
2020-09-30 Strateole2 balloon-borne RO data release for date 2019.346
processing version: V.0003.0023
2020-05-05 Strateole2 balloon-borne RO preliminary data release for date 2019.346
processing version: V.0002.0022
2019-04-15 AR Recon 2018 airborne data release for dates 2018.026 to 2018.027
processing version: V.0002.0021

6 Introduction

This release note is specifically for airborne and balloon-borne RO. In this context, "receivers" means GNSS receivers onboard either an aircraft or a balloon. The "satellites" are solely used in reference to GNSS satellites, not LEO satellites as is the case for other spaceborne RO.

This release note covers the basic information about how meteorological parameters such as pressure and temperature and refractivity profiles are retrieved from airborne and balloon-borne RO measurements. The data format of the final products in netCDF format is described in detail. The format follows conventions based on the standard CDAAC data format, but some changes are made to accommodate specific features of airborne/balloon-borne RO profiles.

6.1 Brief of Retrieval Procedures

The retrieval processes start from the high-accuracy aircraft/balloon trajectory determination. The precise positions are calculated using either the Applanix POS-Pac software (desired equipment used with or without IMU data) or the PANDA PPP-AR software developed at Wuhan University GNSS Research Center. Then, using the raw GNSS carrier phase observation data, the receiver trajectory, and the satellite orbit ephemerides, the excess phase/Doppler residuals are calculated using either our internal **do_aro** or **calres** methods. The radius and center of local curvature

are determined to approximate the ellipsoidal Earth surface so that the positions of satellites and receivers can be transformed into a locally spherically symmetric coordinate system. The excess Doppler observations are used to determine the bending angle in this shifted geometry using geometric optics or phase matching. Then, bending angles (positive and negative) are calculated, and a partial bending angle is obtained after some smoothing. The inverse-Abel transform is used to derive the refractivity. Finally, meteorological parameters such as dry pressure and dry temperature, with water vapor neglected, are derived using the hydrostatic equilibrium assumption and ideal gas law. An alternate way of retrieving pressure, temperature, and humidity is to use the 1D-var method with model inputs as the first guess. However, this has not yet been implemented.

Users should understand that dry temperature is not a good representation of actual temperature below 9 km altitude and by preference, should use the refractivity profiles for analysis. For data recovered from a tropical area, such as in the Strateole2 campaign, the dry temperature deviates from true temperature from ∼12 km below which moisture is not negligible.

6.2 Horizontal Location of Drifting Tangent Points

The refractivity profile is derived using the Inverse-Abel transform for each tangent point height and corresponding impact parameter. Because of the interpolation of impact parameter during this process, an individual value of refractivity retrieved at a certain tangent point altitude does not have an associated time of observation, nor does it have a horizontal location. However, the positions of the satellite and receiver, bending angle, and impact parameter are all a function of time. With the known satellite and receiver positions and a climatological refractivity profile (CIRA-Q), we run a ray-tracing forward model (ROSAP) to simulate the bending angle at each time, and in the process establish the relationship between time and the tangent point location and height. In this way, we can assign a tangent point height and location, and time to each retrieved value of refractivity in the level2 **atmPrf** file.

However, CDAAC explains that they do not mix time and height together as independent variables. "It was decided to leave these data (time information) in the atmPhs file as it is difficult to accurately and unambiguously assign a correspondence between time (the independent variable in the atmPhs files) and height(the independent variable in **atmPrf** files)." We point out a significant difference with the spaceborne RO in which it is possible to determine geometrically the horizontal location of the tangent point without establishing the time correspondence. Because we associate time with each tangent point can determine the relationship between spatial and temporal resolution. In this case, the resolution in altitude is set as 100m, and the corresponding temporal resolution is roughly 3–8 sec.

6.3 Time Marks for an Occultation

All the discussions in this section will be first described for the case of a typical SETTING occultation. Differences in the case of a RISING occultation will be described at the end. For a typical occultation, it is clear and straightforward to label an occultation in time using the ending time

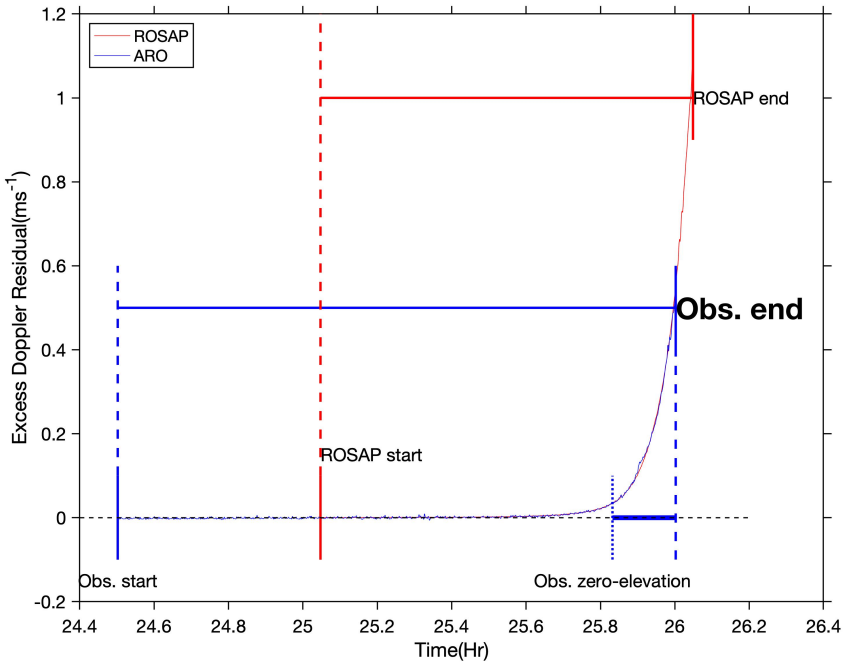

Figure 1: Time marks of a typical setting occultation with corresponding ROSAP forward simulation results.

(**Obs.end** in Figure 1) when a GNSS receiver loses track of the signal, for example, due to low SNR. This time mark is unique for an occultation and is defined in the atmPrf data files. In the current data processing algorithm, measurements beginning from 1 to 2 hours before the **Obs.end** are selected for processing and saved in a standard dop2alp_in format. This staring point shown as **Obs.start** in Figure 1 (**Obs.start** – **Obs.end** =1–2 hr). This **Obs.start** time mark has no physical meaning, however, it is used in the name of all intermediate files in the processing and saved as **old_fileStamp** in atmPrf files for the purpose of internal reference.

Another important time mark for an occultation is the time when the GNSS satellite reaches a zero-elevation angle with respect to the receiver. This time (**Obs. zero-elevation**) marks the transition from positive elevation angle bending to negative elevation angle bending. At this time, the tangent point is at the location of the aircraft's GNSS antenna and it is the starting of the slanted profile where the tangent point starts to drift away and downward from the GNSS antenna. It is determined by calculating the straight-line elevation angle between GNSS satellite and receiver, without considering the bending of the ray path. The time period between **Obs.zero-elevation** and **Obs.end** properly and unambiguously labels this occultation in the time domain. In the final atmPrf products, these two time points are labelled as **start_time(=Obs.zero-elevation)** and **end_time(=Obs.end)**. As stated in the previous section, the time is not an independent variable for atmPrf files but we interpolate the **[time, tangent_point_height]** relation determined in the forward ray tracing to specify the time at each of the tangent points. The interpolated time range

closely matches the [**start_time, end_time**], but not exactly because the time of the actual zero-elevation angle is slightly different from the straight line **Obs.zero-elevation** due to ray path bending.

An oblate correction is necessary prior to the spherically symmetric Abel-transform step in the refractivity retrieval. The global attribute **occpt_offset** is introduced to label the time offset from the start of the occultation **(Obs.start)** to the time point at which the oblate correction is determined. In our dataset, the oblate correction is always applied when the tangent point reaches its minimum height, which corresponds to the end of the occultation **end_time(=Obs.end)**. So this **occpt_offset** is calculated as **end_time** – **start_time**, which also corresponds to the duration of an occultation. For a RISING occultation, the **Obs.start** is the time the receiver starts to track of GNSS signal. **Obs.end** is obtained by extending 1–2 hr from the **Obs.start**. In this case, the occultation is label as **start_time(=Obs.start)** and **end_time(=Obs.zero-elevation)**. The global attribute **occpt_offset** is still defined as **end_time** – **starting_time**.

6.4 Smoothing of Bending Angle

For a setting occultation, the bending angle was calculated continuously from the time when the GPS satellite is still above the horizon to the time that the GPS satellite sets below the horizon and loses track. The corresponding segments of bending angle are called positive and negative bending angles, respectively. The positive bending angle and part of the negative angle just below the horizon, are generally small and associated with larger uncertainties. To reduce the propagation of the uncertainties into further retrieved refractivity at lower altitudes in the inverse Abel transform, the original bending angle is smoothed for the partition of positive bending and negative bending near the horizon. The noisy section of the bending angle profile is replaced with smoother values with noticeable outliers removed. Then the partial bending is calculated from this optimized/smoothed bending angle profile and the final refractivity profile is calculated. In the **atmPrf** files, the smoothed bending angle (positive and negative segments) and calculated partial bending angle are both included and named as **Bend_ang** and **Opt_bend_ang**, respectively. We keep the original variable names but give them new definitions. For a rising occultation, the positive and negative bending angles are the same except the time is reversed, aka, the negative bending is retrieved first, then the positive bending angle.

6.5 Occultation Point and Oblate Correction

One important characteristic of airborne/balloon-borne RO is the retrieved profiles are slant. The obliqueness of typical profiles is too large to be neglected, considering the limited altitude range. In the example shown in the Figure 2, the tangent point (blue dots in 2(a) and 2(b), and yellow line in 2(c)) of the airborne/balloon-borne RO profile is slant with a horizontal drift distance of 500–600 km, and corresponding altitude range is 0–14 km (for airborne) or 0–20 km (for balloon-borne). For such a slant profile, in order to give this occultation profile a geographical coordinate, the location of

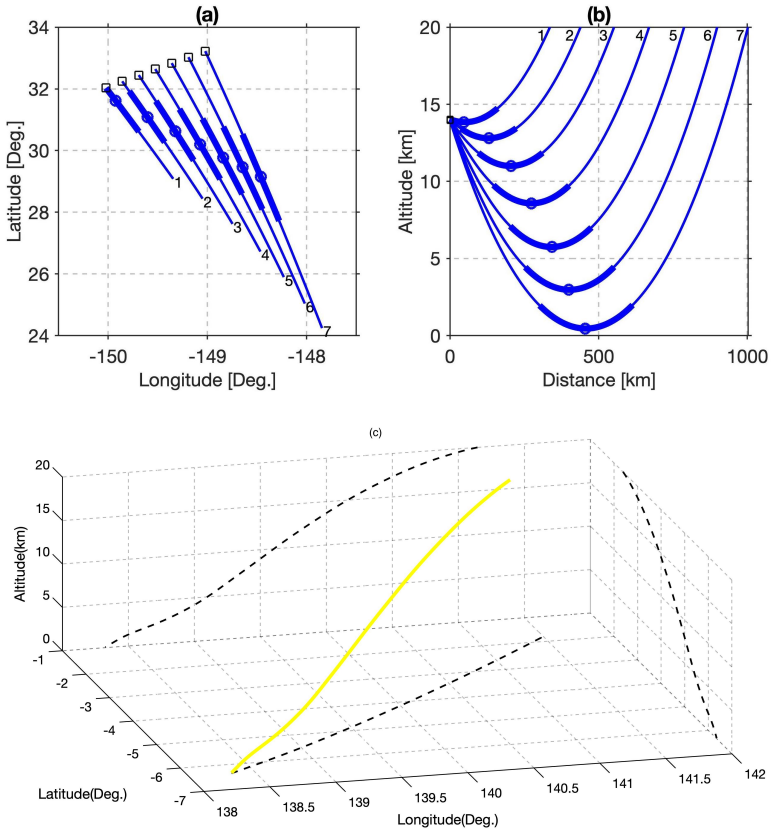

Figure 2: Drift of the tangent point of one airborne/balloon-borne RO profile. (a) the plan view and (b) altitude-distance cross-section of the signal ray path (blue line) and tangent point (blue dot) of each ray path. The square is the aircraft that is moving during the occulation lasting about 10–15 minutes. (c) the 3-D view of the tangent point drifting for a balloon-borne occultation.

the lowest tangent point (148.3° W, 29.2° N in 2(a), 138.3° W, 6.4° S in 2(c)) is selected, rather than any mean or median of the whole slanted profile. Please note the difference with space-borne RO where the point at which 500 meters of excess phase is accumulated, is selected as the 'occultation point'. We did not choose a similar definition because the conventional close-loop GNSS receivers were used so far, there is no guarantee of tracking of signal to accumulate excess phase to a certain value. The lowest points correspond to the largest accumulated excess phases.

The aforementioned 'oblate correction' is also applied at this lowest point. Because most profiles are truncated at certain altitudes, not always reaching the Earth's surface. Therefore, the location of the lowest tangent point projected to the Earth's surface that was described by the ellipsoid WGS84 was found. Then, the orientation of the lowest signal ray path (line 7 in 2(a)) is determined. A sphere tangential to the projection point on the Earth's surface, along with the ray path orientation is found, and the center and radius of the tangential sphere are determined. The position/velocity of aircraft/balloons and satellites are then shifted to the new origin of coordinates.

6.6 How to calculate geometric height from geopotential height?

Typical RO products, including the airborne/balloon-borne RO, are described in geometric heights above the ellipsoid or orthometric heights, above the Mean Sea Level (MSL). Observations such as radiosonde and dropsonde are often described in the geopotential heights or simply geopotential. The geopotential height in meters can be derived simply by dividing geopotential ($\mathrm{m^2\,s^{-2}}$) by World Meteorological Organization (WMO) standard gravity $g_0 = 9.80665\ \mathrm{m\,s^{-2}}$ prior to transformation to the ellipsoidal height. The conversion between geopotential and geometric altitudes can be expressed by the following formulas.

$$\begin{aligned} H_\phi(Z) &= \frac{r_\phi Z}{\frac{r_\phi g_\phi}{g_0} - Z} \\ Z_\phi(H) &= \frac{r_\phi H}{r_\phi + H} \cdot \frac{g_\phi}{g_0} \end{aligned} \tag{1}$$

where:

Z_ϕ = altitude in geopotential meters [in m],
H_ϕ = altitude in geometric meters [in m],
g_0 = WMO standard gravity acceleration 9.80665 [in $\mathrm{m\,s^{-2}}$],
g_ϕ = sea level acceleration of gravity at latitude ϕ [in $\mathrm{m\,s^{-2}}$],
r_ϕ = effective earth radius for latitude ϕ [in m],
ϕ = the latitude.

6.6.1 Snider/Muradyan's Formulas

The gravity acceleration at certain latitude ϕ can be described by an empirical formula:

$$g_\phi = 9.80616 \times \left[1 - 0.0026373 \times \cos(2\phi) + 0.0000059 \times \cos^2(2\phi)\right], \tag{2}$$

The geocentric radius vector to the surface of the WGS84 ellipsoid, which approximates the surface of the earth, is computed from the relationship:

$$r_\phi = \sqrt{\frac{a^4 \cos^2\phi + b^4 \sin^2\phi}{a^2 \cos^2\phi + b^2 \sin^2\phi}}, \tag{3}$$

where: a = equatorial radius = 6378.160 km, and b = polar radius = 6356.775 km. The Eq. 2 and 3 are taken from Paytsar Muradyan's thesis [Snider & Goldman, 1975; Muradyan, 2012, the radius a and b are also taken from]. These two formulas are is currently used for ARO-dropsonde and ARO-ERA5 comparison.

6.6.2 ROPP/Pawel's Formulas

r_ϕ is the effective Earth radius for the ellipsoid defined as

$$r_\phi = \frac{a}{1 - \frac{e^2}{2} + g_r - e^2 \sin^2 \phi}, \tag{4}$$

where $g_r = 0.003449787$ is a gravity ratio. The semi-major a and semi-minor axes b of the WGS84 ellipsoid are 6378.137 km and 6356.752 km. And the ellipsoid eccentricity $e^2 = 1 - b^2/a^2$. The acceleration due to gravity g for a given node's latitude ϕ is calculated on the reference ellipsoid according to Somigliana [1929]

$$g_\phi = g_e \left(\frac{1 + 0.0019 \sin^2 \phi}{\sqrt{1 - e^2 \sin^2 \phi}} \right), \tag{5}$$

where $g_e = 9.7803253359$ is equatorial gravity acceleration. This is taken from Appendix A in Hordyniec, Kapłon, Rohm, & Kryza [2018] and the ROPP documents [Lewis, 2007]. They are also implemented in the ROPP software package [ROM SAF, 2021, ropp11.0/ropp_utils/geodesy/].

6.6.3 Geiod Undulation

While geopotential strictly refers to the geoid and the geometric altitude is calculated relative to the ellipsoid, it is required to include the geoid undulation after conversion using e.g., the Earth Gravitational Model 1996 (EGM96) [Lemoine et al., 1998]. The geometric height relative to the ellipsoid h_e (referred to as ellipsoid height) can be converted from the orthometric height h_g (relative to the geoid) given a geoid undulation u

$$h_e = h_g + u, \tag{6}$$

where u is estimated by the EGM96 model at given locations.

6.6.4 Possible Differences and Recommendation

The difference in the gravity acceleration as defined in Eq. 2 and Eq. 5 is very small ($<10^{-4}\,\mathrm{m\,s^{-2}}$). There are also several other versions of the empirical formulas to calculate the gravity acceleration, the difference among them is subtle and negligible. The difference in the effective radius as defined in Eq. 3 and Eq. 4 (in the order of 20–50 km, or 0.3 %–0.8 % of the Earth radius) is related to the definition of the radius, either geocentric or geodetic. The differences in the final converted geometric altitudes using two types of definitions are up to 30 cm at 15 km, up to 20 cm at 10 km, up to 10 cm at 8 km, and up to 5 cm at 5 km. The differences are largest at the equator and pole, and smaller at mid-latitudes (about half of the aforementioned values). There is a difference of several percent in the geoid heights between the two versions of the Geoid models (EGM96 and EGM2008) in most places, except near the polar area where the differences are much larger. **We**

recommend using the same formulas as the ones used in ROPP.

6.7 How to calculate refractivity from dropsonde observations?

The atmospheric refractivity N for radio waves at GNSS frequencies (L band) in the neutral atmosphere is described by [Haase et al., 2014, 2021; Rüeger, 2002]

$$N = (n-1) \times 10^6 = 77.689\frac{p}{T} - 6.3938\frac{p_w}{T} + 3.75463 \times 10^5 \frac{p_w}{T^2}, \tag{7}$$

where n is the refractive index, p, p_w and T are atmospheric pressure [in hPa], water vapor pressure [in hPa] and temperature [in K], respectively. In the atmosphere higher than ∼10 km, assuming water vapor is negligible, the equation for refractivity can be simplified to

$$N = (n-1) \times 10^6 = 77.689\frac{p}{T_d}, \tag{8}$$

where T_d is commonly referred as dry temperature and deviates from true temperature if moisture is present. However, there is also a different form of the formula with only two terms. [**?**]

$$N = (n-1) \times 10^6 = 77.6\frac{P}{T} + 3.73 \times 10^5 \frac{p_w}{T^2}. \tag{9}$$

The difference between the refractivity defined as Eq. 7 and Eq. 9 is about +0.1 % (Eq. 7 - Eq. 9) nearly constant throughout the altitude range.

6.7.1 Specific Humidity

When airborne/balloon-borne RO refractivity profiles are compared with other observations such as dropsonde/radiosonde or model output such as the ERA5 reanalysis, the corresponding refractivity needs to be calculated from temperature, pressure, and humidity. In the definition of refractivity in Eq. 7, water vapor pressure p_w is not a direct measurement or model variable. In general, absolute humidity provides a better representation of water vapor content than relative humidity, especially when the water vapor concentration is low. We selected the following formula to calculate water vapor pressure from absolute (specific) humidity:

$$p_w = p\frac{r}{\epsilon + r}, \tag{10}$$

where r is the water vapor mixing ratio, calculated from specific humidity q [in kg/kg] using $r = q/(1-q)$, and ϵ (R_d/R_w) is the ratio of gas constants for dry air and water vapor, assumed to be 0.622. This formula provides a more accurate estimate of water vapor pressure than other formulas that estimate water vapor pressure from relative humidity based on empirical equations.

This looks similarly to ROPP calculations in Forward Model Module User Guide: section 4.8.2 Compute partial water vapour pressure. Should we check the consistency?.

6.7.2 ITU's Formula of Relative Humidity

However, observations such as dropsonde/radiosonde only have relative humidity available. The water vapor pressure needs to be calculated from relative humidity r_H. The following formulas can be used.

$$p_w = r_H \cdot p_w(sat) \tag{11}$$

The saturated water vapor pressure $p_w(sat)$ can be written as [ITU, 2019]

$$p_w(sat) = EF \cdot a \cdot \exp\left[\frac{\left(b - \frac{T_c}{d}\right) \cdot T_c}{T_c + c}\right]. \tag{12}$$

and the coefficient EF

$$EF = 1 + 10^{-4} \times \left[7.2 + p\left(0.00320 + 5.9 \times 10^{-7} T_c^2\right)\right] \tag{13}$$

where T_c= temperature in (°C), p=pressure (hPa), r_H= relative humidity (%), $p_w(sat)$ =saturated water vapor pressure (hPa) at temperature T_c, and the coefficients a, b, c and d are:

	water	
	-40°–50°	
a	6.1121	
b	18.678	
c	257.14	
d	234.5	

6.7.3 Muradyan's Formula of Relative Humidity

The following formulas are taken from Paytsar Muradyan's thesis [Muradyan, 2012, original reference could not be located]. The $p_w(sat)$ can also be calculated to account a region of sub-zero temperature, where the phases of water can be presented simultaneously.

$$p_w(sat) = a_1 \exp\left(a_2 \frac{T - T_{melt}}{T - a_3}\right), \tag{14}$$

where $T_{melt} = 273.16$ K, a_1=610.78 Pa, and the constant a_2 and a_3 linearly vary between the values in the $[T_{melt} - 15, T_{melt}]$ interval as follows:

$$a_2 = \begin{cases} 17.269\ K & for\ T \geq T_{melt} \\ 21.875\ K & for\ T \leq T_{melt} - 15 \end{cases} \tag{15}$$

$$a_3 = \begin{cases} 35.86\ K & for\ T \geq T_{melt} \\ 7.66\ K & for\ T \leq T_{melt} - 15 \end{cases} \tag{16}$$

The two sets of constants correspond to the water vapor saturation pressure over a surface of liquid water and ice respectively. The use of a combination in the 15-degree window below the freezing temperature accounts for the effect of three phases of water in that temperature interval. In our calculations, the a_2 and a_3 values have been interpolated for the $[T_{melt} - 15, T_{melt}]$ temperature region.

6.7.4 Recommendation

ERA5 products have both specific and relative humidity, specific humidity is selected to calculate the water vapor pressure and refractivity. For radiosonde/dropsonde products, relative humidity was used to calculate water vapor pressure and refractivity, based on Muradyan's formula. ITU's formula was used to calculate aircraft flight level refractivity, where there is not much water vapor. A comparison was made using ERA5 products that have both specific and relative humidity, it turns out that Muradyan's formula could estimate the saturated water vapor pressure closest to the one estimated by specific humidity. **We recommend using specific humidity to calculate water vapor pressure and refractivity if it is available. If not, choose Muradyan's formula to calculate water vapor pressure and refractivity from relative humidity.** Please see section 5. in Vedel, H [2001] for recommendations regarding humidity calculation from sonde data.

7 Data Format

For each retrieved airborne/balloon-borne Radio Occultation profile, final products are delivered in both netCDF and ASCII formats, with the content exactly the same in both files.

7.1 Directory Name Format

Strateole2 Directory Name Format:
strateole2/data/YYYY-MM-DD_CAID/data_release_YYYY-MM-DD/level2/atmPrf/YYYY.DOY

First **YYYY-MM-DD**	Start date of flight/campaign
CAID	Campaign ID or Balloon ID
Second **YYYY-MM-DD**	Date of data release
YYYY.DOY	Date of daily data subdirectory

Example: strateole2/data/2019-12-05_str2/data_release_2020-09-30/level2/atmPrf/2019.346

AR Recon. Mission Directory Name Format:
airborne_ro/data/YYYY.DOY_CAID/data_release_YYYY-MM-DD/level2/atmPrf/YYYY.DOY_FLID
Example: airborne_ro/2018.023_ar2018/data_release_2019-04-15/level2/atmPrf/2018.027_rf01

France ISPL Data Sever Directory Name Normat:
/DATA/CampaignID/InstrID/ProjID_CampaignID_BallonID_ConfigID_InstrID_YYYYDOY_vYYYYMMDD
Example: /DATA/C0/ROC/ST2_C0_06_STR1_ROC_2019346_v20210625
Notes: YYYYDOY is the flight date, vYYYYMMDD is the data release date.

7.2 File Name Format

Current File Name Template in use:
/atmPrf/YYYY.DDD_RFID/atmPrf_IIII.YYYY.DDD.HH.MM.GXX.SSSS.VVVV_nc

IIII	Receiver/Site ID (SEPT, R24T, N49T, C023, C125)
YYYY	Year
DDD	Day of year
RFID	Research Flight or IOP ID if any
HH	Hour
MM	Minute
GXX	Low elevation GNSS satellite PRN G=GPS, E=Galileo, R=GLONASS, C=Beidou
SSSS	Data format version
VVVV	Software algorithm version

Legacy File Name Template, not longer used:
/atmPrf/YYYY.DDD_RFID/atmPrf_IIII.YYYY.DDD.HH.MM.GXXS_GYYH.SSSS.VVVV.nc

IIII	Receiver/Site ID (SEPT, R24T, N49T, C023, C125)
YYYY	Year
DDD	Day of year
RFID	Research Flight or IOP ID if any
HH	Hour
MM	Minute
GXX	Low elevation GNSS satellite PRN G=GPS, E=Galileo, R=GLONASS, C=Beidou
S/R	S=Setting or R=Rising occultation
GYYH	High elevation GNSS satellite PRN, H=High.
SSSS	Data format version
VVVV	Software algorithm version

7.3 Directory Name and File Name Format for uploads to air.larc.nasa.gov

ARO data were collected for the West-coast & Heartland Hyperspectral Microwave Sensor Intensive Experiment (WHyMSIE). These data were collected on the NASA ER-2, and have been processed to yield atmPrf products. They are archived in the conventional ARO format as described in this document. They may be viewed at `https://agsweb.ucsd.edu/gnss-aro/whymsie2024/`. Observation error and QC related files for the WHyMSIE campaign are also available at that site. To participate in the communal archive, file names were changed to be compliant with ICARTT standards in ways that conflict with 7.1 and 7.2. For results uploaded to `https://www-air.larc.nasa.gov/cgi-bin/ArcView/whymsie#HAASE.JENNIFER/`, the formats described in this subsection supersede 7.1 and 7.2 where they conflict.

7.3.1 Directory Name Format

All WHyMSIE ARO product files are uploaded to `https://www-air.larc.nasa.gov/cgi-bin/ArcView/whymsie#HAASE.JENNIFER/` under the 'whymsie-aro' data ID. No provision exists on the site for folders to separate data by flight as is typical (e.g. 7.1), and ARO lab flight numbers do not necessarily match NASA "launch numbers".

7.3.2 File Name Format

ARO product files were renamed for compliance with International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) File Format Standards 2.2. As discussed in this document, ARO typically follow CDAAC conventions. Those wishing to use the CDAAC-compliant filenames may look to TBD. The ICARTT standard, as applied to ARO data, is described below.

Current File Name Template in use:

/dataID_locationID_YYYYMMDD[hh[mm[ss]]]_R#[_L#][_V#][_comments].extension

dataID	short description of measured parameter/species instrument, or model, always "whymsie-aro"
locationID	short description of site, station, platform, laboratory or institute, always "ER2"
YYYY	four-digit year, always "2024"
MM	two-digit month
DD	two-digit day
hh	optional two-digit hour
mm	optional two-digit minute
ss	optional two-digit second
R	revision number of data, currently only "0" exists
L	optional launch number, not used
V	optional volume number, not used
comments	optional additional information, not used
extension	file extension, always "nc"

In practice, this means that all ARO data files uploaded to air.larc.nasa.gov will have the format whymsie-aro_ER2_YYYYMMDDhhmmss_R0.nc

7.4 Global Attributes

WARNING: The following attributes marked in the box are unique for airborne/balloon-borne RO, not used by CDDAC.

1. **year, month, day, hour, minute, second**
   - **Description:** Date/time of the occultation, determined as the time for the lowest ray path.
   - **Date Type:** double(8)
   - **Units:** NaN
2. **GPS_week** – Removed, no longer in use
   - **Description:** GPS week of the occultation
   - **Date Type:** double(8)
   - **Units:** NaN
3. **start_time**
   - **Description:** Starting time for the occultation
   - **Date Type:** double(8)

- **Units:** GPS epoch in second

4. **stop_time**
   - **Description:** Ending time for the occultation
   - **Date Type:** double(8)
   - **Units:** GPS epoch in second
5. **lat**
   - **Description:** Latitude of perigee point at occultation point
   - **Date Type:** double(8)
   - **Units:** decimal degrees
   - **Valid Range:** -90, 90
6. **lon**
   - **Description:** Longitude of perigee point at occultation point
   - **Date Type:** double(8)
   - **Units:** decimal degrees
   - **Valid Range:** -180, 180
7. **curv**
   - **Description:** The X, Y and Z offset of the center of sphericity used in the inversion from the ECEF Earth center
   - **Date Type:** 3*double(8)
   - **Units:** km
   - **Valid Range:** -200, 200
8. **occpt_offset**
   - **Description:** The time offset from the start of the occultation to the lowest occultation point, at which **curv** is measured. It equals to the duration of the occultation.
   - **Date Type:** double(8)
   - **Units:** sec
   - **Valid Range:** -1000, 1000
9. **rgeoid**

- **Description:** The height of the geoid ABOVE the reference ellipsoid (WGS84) at the lat/lon position of the lowest occultation point. MSL altitude = Ellip altitude – geoid height. In the most recent release of the data, geoid of EGM2008 was used for AR Recon data (AR-21, AR-22). However, the older datasets used the geoid of EGM96. **The standard is EGM96, so we have to reformat all of our data to be in EGM96**
- **Date Type:** double(8)
- **Units:** km
- **Valid Range:** -150, 150

10. **rfict**

- **Description:** Local curvature radius of the reference ellipsoid for the occultation point
- **Date Type:** double(8)
- **Units:** km
- **Valid Range:** 6200, 6600

11. **azim**

- **Description:** Azimuth angle of the occultation plane at tangent point measured clockwise with respect to North at 0 degrees. The angle is measured between North and the direction of the raypath from aircraft/balloon antenna towards GNSS satellite.
- **Date Type:** double(8)
- **Units:** decimal degrees
- **Valid Range:** -180, 180

12. **fileStamp**

- **Description:** The complete ID for this occultation (IIII.YYYY.DDD.HH.MM.GXX).
- **Date Type:** char

13. **prefix, was called old_fileStamp**

- **Description:** The internal reference ID for this occultation (gXXs_XX-Z).
- **Date Type:** char
- **Note:** This is used in all the intermediate files during the data processing.

14. **bad**

- **Description:** Badness flag. 1 = Profile flunked quality control, 0 = Profile passed QC
- **Date Type:** int

15. **errstr**
    - **Description:** Error string. If bad = 1, contains the description of the problem found, else contains the null string
    - **Date Type:** string
16. **reference_sat_id**
    - **Description:** PRN of the satellite at high elevation used to do the single differencing
    - **Date Type:** int
17. **occulting_sat_id**
    - **Description:** PRN of the satellite at the occultation position
    - **Date Type:** int
18. **fiducial_id**
    - **Description:** Placeholder, not sure what it means
    - **Date Type:** int
19. **occ_id**
    - **Description:** Placeholder, not sure what it means, but put zero for now
    - **Date Type:** int
20. **usebits**
    - **Description:** Placeholder, not sure what it means, but put zero for now
    - **Date Type:** int
21. **center**
    - **Description:** Organization name of data processing
    - **Date Type:** string
22. **version**
    - **Description:** Data processing software and data format versions
    - **Date Type:** string
23. **mission**
    - **Description:** Campaign name of the experiment, either AR or TC, and the year.
    - **Date Type:** string

24. **tailnumber**
    - **Description:** Aircraft tail number
    - **Date Type:** string
25. **receiver**
    - **Description:** ARO receiver model number
    - **Date Type:** string
26. **antenna**
    - **Description:** ARO antenna model number
    - **Date Type:** string
27. **contact**
    - **Description:** Contact information regarding all issues
    - **Date Type:** string
28. **creation_time**
    - **Description:** Date and time this file was created
    - **Date Type:** string

7.5 Profile Variables

WARNING: The following variables marked in the box are unique for airborne/balloon-borne RO, are not used by CDAAC, or have different physical meanings. Please see notes above.

1. **Time**
    - **Description:** Corresponding time of each ray path
    - **Date Type:** float
    - **Units:** GPS second of week
    - **Valid Range:** 0, 999999
    - **Missing Value:** -999
    - **Note:** See previous section for the explaination of the time.
2. **MSL_alt**
    - **Description:** Mean sea level altitude of perigee point
    - **Date Type:** float

- **Units:** km
- **Valid Range:** 0, 9999
- **Missing Value:** -999

3. **Ellip_alt**
   - **Description:** Ellipsoidal altitude of perigee point above the WGS84 ellipsoid
   - **Date Type:** float
   - **Units:** km
   - **Valid Range:** 0, 9999
   - **Missing Value:** -999
4. **Impact_para**
   - **Description:** Impact parameter
   - **Date Type:** float
   - **Units:** km
   - **Valid Range:** 0, 9999
   - **Missing Value:** -999
5. **Impact_height**
   - **Description:** Impact height=Impact parameter−Radius of Curvature (rfict)
   - **Date Type:** float
   - **Units:** km
   - **Valid Range:** 0, 9999
   - **Missing Value:** -999
6. **Bend_ang**
   - **Description:** Smoothed bending angle, composed of positive and negative bending angles, to be distinguished with **Opt_end_ang**
   - **Date Type:** float
   - **Units:** rad
   - **Valid Range:** 0, π
   - **Missing Value:** -999
7. **Bend_ang_stdv**
   - **Description:** Uncertainties of Bend_ang, now empty

- **Date Type:** float
- **Units:** rad
- **Valid Range:** 0, π
- **Missing Value:** -999

8. **Opt_bend_ang**

   - **Description:** Partial bending angle, calculated from smoothed positive and negative bending angles, to be distinguished with **Bend_ang**
   - **Date Type:** float
   - **Units:** rad
   - **Valid Range:** 0, π
   - **Missing Value:** -999
   - **Notes:** Partial bending angle = negative elevation angle bending angle – positive elevation angle bending angle.

9. **Ref**

   - **Description:** Refractivity
   - **Date Type:** float
   - **Units:** N
   - **Valid Range:** 0, 9999
   - **Missing Value:** -999

10. **Ref_stdv**

    - **Description:** Uncertainties of refractivity, now empty
    - **Date Type:** float
    - **Units:** N
    - **Valid Range:** 0, 9999
    - **Missing Value:** -999

11. **Pres**

    - **Description:** Dry pressure
    - **Date Type:** float
    - **Units:** mb or hPa
    - **Valid Range:** 0, 1500
    - **Missing Value:** -999

12. **Temp**
    - **Description:** Dry temperature
    - **Date Type:** float
    - **Units:** C
    - **Valid Range:** 0, 9999
    - **Missing Value:** -999
13. **Lat**
    - **Description:** Latitude of perigee point
    - **Date Type:** float
    - **Units:** degree
    - **Valid Range:** -90, 90
    - **Missing Value:** -999
14. **Lon**
    - **Description:** Longitude of perigee point
    - **Date Type:** float
    - **Units:** degree
    - **Valid Range:** -180, 180
    - **Missing Value:** -999
15. **Azim**
    - **Description:** azimuth angle of the occultation plane at tangent point measured clockwise with respect to North at 0 degrees. The angle is measured between North and the direction of the raypath from aircraft antenna towards GNSS satellite.
    - **Date Type:** float
    - **Units:** degree
    - **Valid Range:** -180, 180
    - **Missing Value:** -999

8 Known Issues

1. Different methods might be used to calculate geodetic coordinates (latitude, longitude, altitude) from Cartesian coordinates (ECEF_xyz). Some corrections are needed to compensate for the possible difference. Matlab uses a function **ecef2geodetic** while the ROSAP model uses a simplified method to do the conversion.

2. Some variables, especially the coordinates, do not have consistent precision of the measurements.

3. Definition of the time, zero-elevation time as starting or ending time is not a good option but works sufficiently well.

4. In current airborne/balloon-borne RO data processing, the ionospheric correction was applied in the excess phase/Doppler estimation, so only the ionospheric-free bending angle was calculated.

5. The quality flag was not in use and the profiles that were detected with some issues were not processed and included in the released datasets.

9 Contacts

If any errors are noticed in the data, please contact Bing Cao (bic020@ucsd.edu) and Jennifer Haase (jhaase@ucsd.edu).

References

Haase, J. S., Murphy, B. J., Muradyan, P., Nievinski, F. G., Larson, K. M., Garrison, J. L., & Wang, K.-N. (2014). First results from an airborne GPS radio occultation system for atmospheric profiling. *Geophysical Research Letters*, *41*(5), 1759–1765. Retrieved from `http://dx.doi.org/10.1002/2013GL058681` (2013GL058681) doi: 10.1002/2013GL058681

Haase, J. S., Murphy, M. J., Cao, B., Ralph, F. M., Zheng, M., & Delle Monache, L. (2021). Multi-GNSS Airborne Radio Occultation Observations as a Complement to Dropsondes in Atmospheric River Reconnaissance. *Journal of Geophysical Research: Atmospheres*, *126*(21), e2021JD034865. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JD034865` (e2021JD034865 2021JD034865) doi: https://doi.org/10.1029/2021JD034865

Hordyniec, P., Kapłon, J., Rohm, W., & Kryza, M. (2018). Residuals of Tropospheric Delays from GNSS Data and Ray-Tracing as a Potential Indicator of Rain and Clouds. *Remote Sensing*, *10*(12). Retrieved from `https://www.mdpi.com/2072-4292/10/12/1917` doi: 10.3390/rs10121917

ITU, R. (2019, 8). *The radio refactive index: its formula and refractivity data* (Tech. Rep. No. P.453-14). Radiocommunication Sector of ITU.

Lemoine, F. G., Kenyon, S. C., Factor, J. K., Trimmer, R., Pavlis, N. K., Chinn, D. S., ... Olson, T. R. (1998). *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96* (Tech. Rep.). Greenbelt, Maryland, 20771 USA,: NASA Goddard Space Flight Center.

Lewis, H. (2007). *Geodesy Calculation in ROPP* (Tech. Rep. No. GRAS SAF Report 02). Met Office, UK.

Muradyan, P. (2012). *Profiling the atmosphere with the airborne GPS radio occultation technique using open-loop tracking* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN, U.S.A..

ROM SAF. (2021). *The Radio Occultation Processing Package (ROPP) Utilities Module User Guide* (Tech. Rep.). ROM SAF Consortium.

Rüeger, J. (2002). Refractive index formulae for electronic distance measurements with radio and millimetre waves. *Rep. Unisurv Rep*, *109*, 758-766.

Snider, D. E., & Goldman, A. (1975). *Refractive effects in remote sensing of the atmosphere with infrared transmission spectroscopy* (Tech. Rep. No. 1790). Ballstic Research Laboratories.

Somigliana, C. (1929). Teoria generale del campo gravitazionale dell'ellissoide di rotazione. *Memorie della Società Astronomica Italiana*, *4*, 425.

Vedel, H. (2001). *On calculation of zenith total delays from meteorological data and inter comparison with other data from nearby sites. For the COST ACTION 716 ztd calculation study group* (Tech. Rep.). Danish Meteorological Institute.