GNSS Radio Occultation Data in the AWS Open Data Registry

AWS Region: us-east-1

AWS Services: s3://gnss-ro-data/

Author: Stephen Leroy[1]

Reviewers: Erik Fanny, James Lynch, Amy McVey (AER); Chi O. Ao, Kuo-Nung Wang, Panagiotis Vergados (Jet Propulsion Laboratory, California Institute of Technology); Hailing Zhang, Doug Hunt (COSMIC Project Office, UCAR); Hans Gleisner, Kent Lauritsen, Stig Syndergaard (Danish Meteorological Institute); Michael Gorbunov, Vladimir Irisov (Spire Global)

Date: 4 November 2021

Radio occultation is a relatively new remote sensing technique of the Earth's atmosphere and so is generally unfamiliar to the broader Earth atmospheric science community. Interest in it is expected to grow rapidly in the coming years because of some of its extraordinary characteristics: it is the most easily and most strongly calibrated remote sensing technique and so is ideally suited to climate monitoring, and it obtains extraordinarily high vertical resolution and so is well suited to atmospheric process studies that remain unaddressed by all other remote sensing techniques. Despite these unique strengths, it poses its own set of unique challenges: that it is ultimately a limb sounding technique and so its horizontal resolution is poor; and that it cannot unambiguously distinguish the influence of water vapor and that of the "dry atmosphere" – nitrogen and oxygen – on the fundamental retrievable of radio occultation, the microwave index of refraction. A large body of work on the inversion of radio occultation data has been performed and published in the last three decades, but this document is not intended to be a review paper. Instead, it is intended to introduce an atmospheric scientist to the data that is provided by Earth radio occultation, justify the archival of lower-level radio occultation data and document an archival format to the professional radio occultation scientist, and provide some introduction on the nature of the inversion of radio occultation data for temperature, pressure, etc., to the radio occultation non-professional so that he/she gains some understanding of the quality and provenance of the atmospheric variables that are contained in the archived data sets. For the uninitiated atmospheric scientist, a few references are provided to assist in researching the various finer points that are brought up. The professional radio occultation scientist will already be familiar with these references and many more.

In any case, GNSS radio occultation[2] is now in the AWS Open Data Registry and thus accessible to the world, and there are many reasons why the world's citizens might be interested in these data. There have been approximately 30 publicly funded GNSS radio occultation receiver satellites flown as of year 2021 and a few dozen commercial satellites as well. Also, there are

---

[1] Contact: Stephen Leroy, Atmospheric and Environmental Research, 131 Hartwell Ave., Lexington, Massachusetts, USA, (+1) 781-761-2234, sleroy@aer.com.

[2] GNSS radio occultation refers to Earth atmospheric occultation that use the satellites of the Global Navigation Satellite Systems (GNSS) as transmitters of microwave signals and receivers that are flown in low-Earth orbit.

four organizations with semi-independent inversion algorithms that process various subsets of these satellite data. Just two of them are mostly independent and process the vast majority of the data: the COSMIC Project Office at the University Corporation for Atmospheric Research (UCAR), and the NASA Jet Propulsion Laboratory of the California Institute of Technology. Another organization with its own retrieval system is the Radio Occultation Meteorology Satellite Application Facility (ROM SAF) based at the Danish Meteorological Institute in Copenhagen. The ROM SAF has processed a subset of the data that has been processed by UCAR and JPL, but its reach is steadily expanding in time. The ROM SAF is funded by EUMETSAT.

One curious characteristic of radio occultation that is unique to it is that the trends and anomalies in atmospheric variables as retrieved from its measurements are the same no matter which retrieval system inverted the measurements. This is a consequence of its extraordinary calibration, with traceability to the international definition of the second as opposed to passive sounders, whose traceability is complicated by unstable spectral response functions, complicated antenna patterns, and less accurate physical standards. For this reason, and because the data volume of radio occultation is relatively small when compared to passive remote sensors, it is fruitful to archive and distribute the retrievals of atmospheric variables from radio occultation as produced by several of the world's major independent retrieval centers. That is done in the AWS Open Data archive of Earth radio occultation data.

In the remainder of this document, first a brief background to Earth radio occultation is given followed by presentations of the various levels of radio occultation data products that are made available in the AWS Open Data GNSS Radio Occultation archive. The rationale is given for the variables contained in each file type, and documentation of the data formats and descriptions of the variables are presented as well.

1. Background

Radio occultation (RO) is a remote sensing technique wherein a precisely timed microwave signal is transmitted outside an atmosphere, transects the atmosphere's limb, and is received outside the atmosphere. The atmosphere refractively bends the signal as it passes through the limb, and the bending is measured as a Doppler phase shift in the frequency of the received signal. That information can be analytically inverted to profile the microwave index of refraction of the atmosphere at extremely high vertical resolution. This technique has provided some of the best information available for temperature and pressure in planetary atmospheres and has proven itself in many applications of weather, atmospheric process, and climate studies.

Since 1995, the technique has been applied to the Earth's atmosphere using the signals of the Global Positioning System (GPS) originally and currently the signals of many of the Global Navigation Satellite Systems such as GPS, the Russian GLONASS, the European GALILEO, and the Chinese BeiDou. Two salient and unique qualities of RO data are that it is capable of extremely high vertical resolution by virtue of phase-tracking GNSS signals and that it is capable of extraordinary overall accuracy by virtue of calibration against atomic clocks. Trends and anomalies in retrieved variables have proven to be mostly independent of the retrieval algorithm. Demonstrating this requires inter-comparison of retrieved products from different retrieval centers. Three of the world's major RO retrieval centers are represented in the AWS OpenData data set: the CDAAC (ucar), the NASA Jet Propulsion Laboratory/Caltech (jpl), and the Radio

Occultation Meteorology Satellite Application Facility (romsaf). More contributors are under consideration.

GNSS RO undergoes processing which is radically different than that of most atmospheric sounders but can still be categorized by processing step: (1) uncalibrated data as provided by the satellite instrument with communication information stripped – level 1a; (2) calibrated data wherein clock biases of the transmitters and receivers are removed and precise orbits determined – level 1b; (3) high vertical-resolution profiles of RO bending angle and microwave refractivity – level 2a; and (4) retrievals of temperature, pressure, and specific humidity on a coarser vertical grid – level 2b. Note that the definition of the level of processing, levels 0 through 3, differ between various processing centers: the ROM SAF and EUMETSAT in general have a convention that differs from UCAR and JPL, which follow a NASA convention. The levels of processing for the data formats defined below most closely align with the NASA convention.

All files are NetCDF-4. When a variable is not provided by a contributing processing center, the contents of the file for that variable are set to _FillValue.

2. Level 1b: Calibrated GNSS RO data (calibratedPhase)

In GNSS RO, calibration is performed by removing as much bias drift in timing as possible and determining the positions and velocities of the transmitting and receiving satellites as best as possible. As opposed to most remote sensing techniques, GNSS RO is a timing measurement rather than a radiance measurement. More specifically, RO is a measurement in the rate of well-timed signals, which is manifested as a Doppler shift in the frequency of the signals emitted by the GNSS transmitters. As such, RO places strong requirements on drifts in timing biases, and those drifts are usually stated in the form of optical path drifts. For example, if a bias in timing is δt_b, the bias (δL_b) drift is typically communicated as

$$\frac{d}{dt}\delta L_b = c\frac{d}{dt}\delta t_b$$

in which c is the speed of light. Because Doppler shifting of the GNSS signals is the fundamental observable in RO, the uncertainties in the velocities of the transmitting and receiving satellites contribute to the overall uncertainty of the RO data. Consequently, both the bias drifts of the clocks on board the transmitting and receiving satellites must be removed in the process of calibrating RO data, but so also must the velocities of the transmitting and receiving satellite be determined with stringent accuracies as well. All told, the overall uncertainty for timing bias drifts (and velocity uncertainty) is required to be less than 0.4 mm s^{-1} for the purpose of climate monitoring[3]. Lower accuracy is acceptable for the purpose of numerical weather prediction. Thus, the variables that constitute a calibrated RO observation are the total optical path of GNSS signals from the time of transmission from a GNSS satellite to the time of reception on the receiving satellite with all clock biases removed and the positions and velocities of both of those satellites. How this is done has been documented in depth elsewhere[4].

---

[3] Kursinski et al., *J. Geophys. Res.,* **102**, 23429–23465, doi:10.1029/97JD01569, 1997.

[4] Hajj et al., *J. Atmos. Solar-Terr. Phys.,* **64**, 451–469, doi:10.1016/S1364-6826(01)00114-6, 2002.

GNSS RO measures the timing and Doppler shifts of multiple signals simultaneously for reasons of necessity and maximizing signal quality. The necessity arises because the free electrons in the ionosphere induce refraction as well as the neutral molecules of the atmosphere: profiling of atmospheric variables necessitates the removal of the influence of the ionosphere. The refraction induced by free electrons in the ionosphere is dispersive, meaning that the index of refraction changes as a function of frequency, as opposed to the refraction induced by the neutral molecules of the atmosphere. By tracking GNSS signals at two different carrier frequencies simultaneously, linear combinations can be formed that eliminate the influence of the free ionospheric electrons to first order. For GPS signals, these two carrier frequencies are those associated with L1 (1.57542 GHz) and L2 (1.22760 GHz). Other GNSS constellations offer a wider variety of carrier frequencies. A second reason for tracking multiple signals derives from complications associated with code and data modulations on the transmitted signals. The GNSS signals are not simply sinusoids – if they were, it would be impossible to distinguish the transmissions of one satellite from those of all of the others – and the binary modulations of those signals must be removed before they can be interpreted in RO processing. Some of the signals are weaker than others because their modulations are so rapid, because their transmitted power is less, or because the modulations are unknown in advance of reception. Weak signals can theoretically be ameliorated through the combination of multiple signals of a GNSS satellite broadcast at the same carrier frequency, thus the desirability of tracking several GNSS signals simultaneously during an RO observation. For example, the C/A and P code signals are commonly both tracked at L1 during a GPS RO observation[5].

Finally, tracking a GNSS RO signal involves a prior model of the optical path of the signal, primarily for profiling well into the Earth atmosphere's planetary boundary layer. The highest quality RO observations are those obtained when the transmitter and receiver appear to be setting into the Earth's limb from the vantage of the other because the receiver "locks on" to the GNSS signal when the transmitter is far above the atmospheric limb, when the signal is propagating through the vacuum of space. This advantage is absent when the satellites appear to be rising, thus making it far more difficult to obtain high quality data in the lower atmosphere, or any data at all! Also, diffraction, atmospheric multi-path, and super-refraction greatly complicate the RO signal structure because of large vertical gradients in the index of refraction in the lower atmosphere. Obtaining RO measurements in such an environment requires "open-loop" tracking, in which received signals are correlated against sine and cosine waves at frequencies modeled *a priori*[6]. Most importantly, removal of the binary modulations when the signal is weak and complex necessitates a model of the total optical path between the transmitting and receiving satellites as well as knowledge of the sequence of bits in the modulation as well. This type of model is known as a "range" model in GNSS RO retrieval science.

In some cases, a single range model does not suffice to maximize the quality of RO signal detection because of variability in the range model and especially when the range model produced local maxima in auto-correlation at non-zero lags. For this reason, the range model

---

[5] Detail on the structure of GPS signals and those of other GNSS can be found by simple internet searches.

[6] Sokolovskiy et al., *Geophys. Res. Lett.,* **33**, L12813, doi:10.1029/2006GL025955, 2006.

must be archived and an allowance must be made for multiple range models being employed simultaneously for a particular GNSS signal during an RO observation.

Taking the above into consideration, the information that must recorded for calibrated GNSS RO data so that retrieval systems can be applied to the data into the foreseen future are listed as essential variables in Table 1a. The table records the NetCDF variable names, their units, and their descriptions. Table 1b lists the metadata that describes the RO observation. The timing of an occultation is defined by the variables startTime and endTime. The flag navBitsRemoved indicates whether the bit stream associated with the navigation message on a signal has been removed in the excess phase. It is meaningless if the signal is a data-less "pilot", defined as a signal that is not modulated by a navigation message, but is set to True for those tones anyway. The optical path in excess of a vacuum path is given as time series for each signal in variable excessPath, and the signal-to-noise ratio corresponding to optical path for the same signals in variable snr. The RINEX 3 observation codes[7] that define those phase and amplitude are given as phaseCode and snrCode. A phase code always has "L" as the first character, and the signal-to-noise ratio code "S". The carrier frequency for each signal is given as carrierFrequency. The RO observations are given as functions of time, and they are excessPhase for the phase in excess of what would have been incurred in the absence of any refraction and snr for the signal-to-noise ratio. The positions of the transmitting and receiving satellites are also given as functions of time as positionGNSS and positionLEO. The "LEO" refers to low-Earth orbit, because that is where RO receiving satellites are flown. The positions of the LEO are exactly the positions of the phase center of the receiving antenna at the specified high-rate time in the file; however, the positions of the transmitter are given at the time the signal was *transmitted* by the GNSS satellite, corresponding to the signal received at the times given in the file. In order to compute the transmitter positions, it has been necessary to account for the light travel time between the transmitter and receiver and interpolate the transmitter backward in time by those amounts. If range and phase models were used in RO, then they are given as rangeModel and phaseModel. Recall that these models are essential for recording RO events in which the transmitting satellite appears to rise and for high-quality retrieval of atmospheric variables in the lower troposphere. Various other information is given in the metadata, including the identification of the satellites, the names of a reference satellite or a ground station if either was used in calibrating satellite clocks, and a list of references to papers in the form of digital object identifiers ("doi:…") that are helpful in understanding how the calibration of the RO event was executed. For example, if the metadata in the global attributes contains *year*=2009, *month*=3, *day*=4, *hour*=5, *minute*=21, mission='*cosmic1*', leo='*cosmic1c5*', occGnss='*G23*', then the occultation occurred on March 4, 2009, at 05:21 UTC, the receiver was FM5 of the COSMIC-1 mission, and the occulted transmitter was GPS PRN 23.

3. Level 2a: Refractivity and dry atmospheric retrievals (refractivityRetrieval)

Calibrated measurements of phase and amplitude and precise orbits can be inverted to obtain diffraction-corrected profiles of bending angle as a function of impact parameter, atmospheric microwave refractivity as a function of altitude, and "dry pressure" and "dry temperature" as

---

[7] The RINEX 3 observation codes are documented completely in the official documentation distributed through the International GNSS Service, https://igs.org/. For example, the codes and their explanations can be found in Tables 4 through 9 of the RINEX3.04 documentation.

functions of geopotential height. RO by itself does not offer enough information to profile temperature, pressure, and water vapor unambiguously, and so it is customary to generate "dry" pressure and temperature, wherein refractivity is interpreted as having no contribution from water vapor. No assumption need be made about water vapor when retrieving bending angle as a function of impact parameter and refractivity, dry pressure, and dry temperature as functions of geopotential height.

Excess phase measurements are first inverted to obtain bending angles as functions of impact parameter for at least two GNSS signals recorded for the RO sounding. Those two must be associated with different carrier frequencies. Typically, this step in the inversion involves a physical optics retrieval in the form of Fourier Integral Operators that disentangles the effects of atmospheric multi-path and diffraction that is applied to at least two GNSS signals. Several implementations of FIOs exist to do this, including backpropagation, canonical transform types 1 and 2, full spectrum inversion, and phase matching. Documentation[8] on these techniques is plentiful. The resulting bending angle profiles are very noisy in the upper atmosphere, and so several techniques have been implemented to merge these bending angle profiles with climatological bending angle profiles in a process referred to as "statistical optimization"[9]. Also, the bending angle profiles as obtained at two different carrier frequencies are combined to yield bending angle profiles that eliminate the influence of the ionosphere to first order. The traditional method of forming this ionosphere-corrected bending angle leaves behind a residual ionospheric influence that comes about because of horizontal gradients in the ionosphere and ray-path separation of the two signals away from the ray tangent points. Approximate methods are currently being implemented to eliminate this second-order ionospheric influence that are commonly referred to as "kappa-correction" techniques. In any case, an ionosphere-corrected bending angle profile is the result.

The independent coordinate for profiles of bending angle is impact parameter, the asymptotic miss distance of a geometric ray-path if projected as a straight line from its actual ray-path near the transmitting GNSS satellites with respect to the center of the Earth. (If the atmosphere is locally spherically symmetric, the same values result when projected from the ray-path near the receiving satellite.) Because the Earth is not a sphere, and the curvature of the surface depends not only on latitude but also on the orientation of the ray in azimuth, the local center of curvature varies from the center of mass of the Earth. The center of curvature is computed based on the geometry of the occultation alone and is singly defined for an entire occultation event.

The next step in inversion is an application of an Abel transform that converts profiles of bending angle vs. impact parameter to microwave refractivity vs mean sea-level altitude. Refractivity N is related to the index of refraction n through $N = (n - 1) \times 10^6$. The value of refractivity ranges from 320 to 380 in the surface air and about 30 near the tropopause. The mean sea-level altitude is determined as an altitude above an estimate of mean sea level, which is expressed as a best-fit ellipsoid that approximates the Earth's shape plus an undulation that

---

[8] For an overview and discussion of the various physical optics approaches, see Gorbunov et al., *Atmos. Meas. Tech.,* **14**, 853–867, 2021, doi:10.5194/amt-14-853-2021.

[9] For a description of statistical optimization, see Gorbunov, *Radio Sci.,* **37**, 1084, doi:10.1029/2000RS002370, 2002.

accounts for the effects of local perturbations to the gravity field. The best-fit ellipsoid is completely described by the equatorial and polar radii of the ellipsoid, and that ellipsoid is completely defined by two parameters: equatorialRadius and polarRadius. The vertical coordinate is altitude, which is the distance above the mean sea level geoid. Because the occultation rays' tangent points drift horizontally in the course of an occultation sounding, every altitude is associated with a unique longitude and latitude, and so those coordinates can be profiled in altitude as well as the atmospheric variables.

Without additional information on atmospheric variables, water vapor and temperature cannot be retrieved unambiguously just from RO observations. For that reason, a different set of variables are retrieved: dry pressure and dry temperature. These quantities are the pressure and temperature that are retrieved when the contribution of water vapor to refractivity, the equation of state, and the hydrostatic equation is ignored[10]. Simply, the refractivity is related to dry pressure p_{dry} and dry temperature T_{dry} through

$$N = k_1 \, p_{\mathrm{dry}} / T_{\mathrm{dry}}$$

in which k_1 is 0.776 *N*-units K Pa^{-1}. These variables are profiled in the vertical, and the natural independent coordinate is geopotential height[11]. Geopotential height itself is the geopotential energy associated with a point in Earth-centered Earth-fixed (ECEF) coordinates divided by the gravitational acceleration g, which is commonly (but not necessarily) taken to be the WMO standard value of 9.80665 J kg^{-1} m^{-1}.

Finally, super-refraction in RO occurs when ray bending is stronger than the curvature of the Earth in distinct atmospheric layers. This commonly occurs in the planetary boundary layer in the Subtropics. The consequence for RO retrievals is substantial: retrievals of refractivity are biased negative by several per cent[12]. It is extremely difficult to detect the presence of super-refraction in RO observations alone, and it is impossible to remove the bias in retrieved refractivity using the RO observations alone.

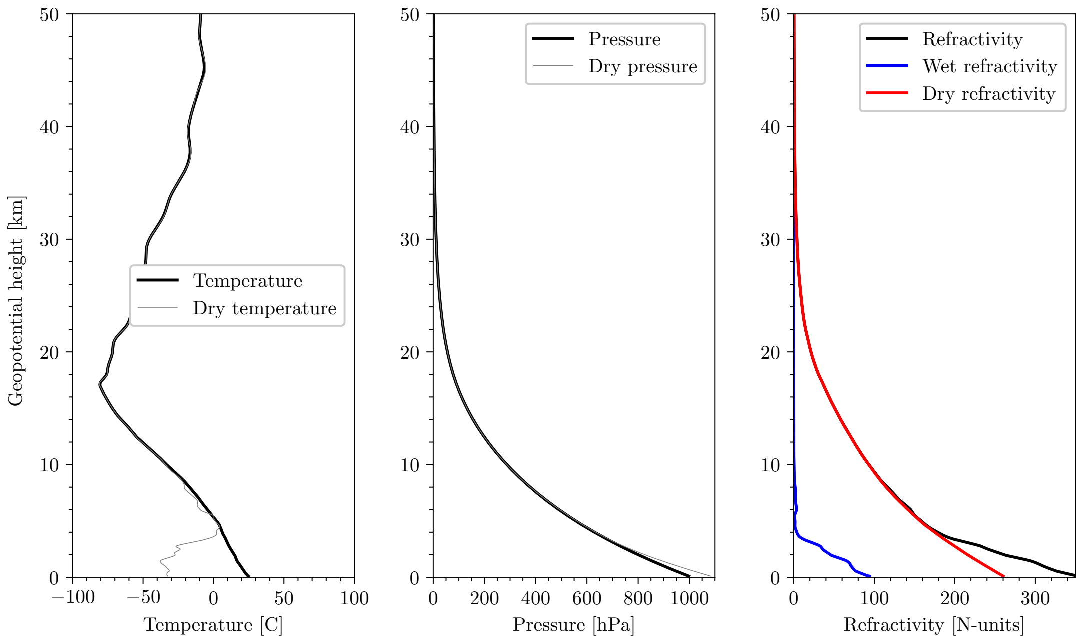
The retrievals of bending angle, refractivity, dry pressure and dry temperature are logged in the refractivityRetrieval format output file. It is NetCDF version 4. The variables and their dimensions are specified in Table 2a. The global attributes are specified in Table 2b. The reference time (refTime), longitude (refLongitude), and latitude (refLatitude) are the nominal time and position of an RO sounding. Because an RO sounding does not occur in a single instant in time because the duration of the event can be in excess of 200 s, the reference time is non-uniquely defined: different processing centers can produce reference times and positions that differ by ~200 s and a few degrees in longitude and latitude. The equatorial and polar radii (equatorialRadius and polarRadius) of the Earth are sufficient to define the shape of the ellipsoid that best fits the mean sea-level geoid of the Earth while undulation is the deviation of the mean sea-level geoid from that ellipsoid. A positive undulation means that mean sea level lies above

---

[10] Danzer et al., *Atmos. Meas. Tech.,* **7**, 2883–2896, doi:10.5194/amt-7-2883-2014, 2014.

[11] Scherllin-Pirscher et al., *J. Geophys. Res.,* **122**, 1595–1616, doi:10.1002/2016JD025902, 2017.

[12] For a full and detailed exposition of super-refraction and its impact on retrieval, see Sokolovskiy, *Radio Sci.,* **38**, 1058, doi:10.1029/2002RS002728, 2003.



Profiles of dry temperature and temperature, dry pressure and pressure, refractivity and its wet and dry components for occultation cosmic2e5_G09_201911012312. This profile is characteristic of RO sounding in the Tropics.

the reference ellipsoid. The local center of curvature of the RO sounding is given in Earth-centered inertial Cartesian coordinates as centerOfCurvature. It is never very far from the center of mass of the Earth. The first step in the post-calibration retrieval chain is the computation of bending angle as a function of impact parameter for at least two signals given as bendingAngle and impactParameter, respectively. Removal of ionospheric influence and blending with a climatology in the upper atmosphere, typically beginning near the stratopause, is given as combinedBendingAngle. The next step in retrieval is the inversion for microwave refractivity as a function of altitude using an Abelian transform, given as refractivity and altitude. Because RO soundings slide in position as they progress, the longitude and latitude corresponding to each latitude are also given. The independent coordinate of retrieval for an atmospheric scientist is geopotential, which is geopotential energy per unit mass and is related to height – rather than altitude – through a multiplicative constant, typically 9.80665 J/kg/m. “Dry” atmospheric profiles of pressure and temperature are given as dryPressure and dryTemperature. The global attributes contain lists of reference regarding the removal of ionospheric influence in the processing (ionospheric_references), the fusion of a model or climatology in the upper atmosphere (optimization_references), and any other references of relevance to the retrieval algorithm. All references are given as digital object identifiers (“doi:…”).

4. Level 2b: Atmospheric retrievals of temperature, pressure, and water vapor (atmosphericRetrieval)

The contributions of water vapor and temperature to refractivity cannot be inferred from RO data alone; however, they can be if auxiliary information on temperature, pressure, or water vapor is supplied. Most often such auxiliary information is taken from the forecast of a numerical weather prediction system. It can also come from a climatological prior. The most sophisticated approaches are optimal estimation in the form of 1DVAR, but simpler approaches are also possible, wherein temperature forecasts are used to estimate water vapor in the lower troposphere and humidity forecasts are used to estimate temperature in the upper troposphere and above. The independent coordinate for these retrievals is geopotential. Because the prior on temperature and water vapor is taken from models that are much more poorly resolved in the vertical than RO observations obtain, it is not unusual for full atmospheric retrievals (temperature, pressure, and water vapor) to be obtained on much coarser vertical resolution grids than the dry retrievals.

The retrievals of temperature, pressure, and water vapor are logged in atmosphericRetrieval. Table 3a describes the variables in the NetCDF file, and Table 3b describes the global attributes. The variable naming is nearly the same as for the dry retrieval format but with the addition of actual pressure and temperature and the partial pressure of water vapor.

5. Naming conventions

The path to the files presented above in the AWS S3 bucket for GNSS RO is

contributed/**center**/**mission**/**filetype**/**yyyy**/**mm**/**dd**/**filetype_mission_center_version_occid**.**suffix**

The bold characters are mnemonics, which are defined in Table 4. The occultation ID **occid** is defined as **leo-ttt-yyyymmddhhnn**. The GNSS transmitters (**ttt**) are defined as three-character strings, the first character defining the GNSS constellation ("G" for GPS, "R" for GLONASS, "E" for Galileo, "C" for BeiDou) and the second and third characters are two-digit zero-padded integer that defines the PRN (or slot number for GLONASS) of the satellite. This is the RINEX standard naming convention for GNSS satellites. The listing of the receiver (**leo**) satellites is given in the second column of Table 5. Three independent GNSS RO processing centers (**center**) have contributed RO data thus far: the COSMIC Project Office at UCAR, the NASA Jet Propulsion Laboratory, and the Radio Occultation Meteorology Satellite Application Facility (ROM SAF). The University of Graz is also considering contributing to this archive.

6. Quality

RO, like any other remote sensing technique, is subject to measurement error that can corrupt the retrieval process in ways foreseen and unforeseen. Until now there has been no uniform standard for flagging the quality of RO retrieval: different processing centers have implemented different approaches to judging the quality of the retrieval of atmospheric variables from each RO sounding. In the case of one processing center, no quality indicator is provided at all. Thus, the inclusion of a quality indicator is considered optional in this repository of RO data. Some centers provide a quality indicator for each level in retrieval, sometimes as a bit-wise integer, sometimes as a score as a floating point scalar. When a quality variable is provided, a global attribute "quality_reference" will provide a pointer to documentation on the quality control indicator contained in each file.

7. The DynamoDB database

AWS offers database service called a DynamoDB, which is a key-value document database utility. Its hallmark feature is search times of a few milliseconds at any scale of the database. It is ideal for cataloging RO soundings. It is implemented for GNSS RO data in the AWS Open Data

Registry. The AWS Open Data Registry does not provide DynamoDB as a service. Instead, a comma-separated value file is provided in the registry as catalogue.csv that can be converted into a DynamoDB within the user's own AWS computing environment.

In this implementation of a DynamoDB database, every entry is a unique RO sounding. Identification of a sounding is governed first by a *partition key* and second by a *sort key*. The search algorithm employed by the DynamoDB service operates on the sort key for each partition, and so the number of RO soundings in each partition must be manageable. In the RO database, the partition key is defined as **leo-ttt** and the sort key as **yyyy-mm-dd-hh-nn**; see the mnemonics in Table 4. The transmitter, the receiver, and the time of the occultation to within a few minutes is sufficient to uniquely identify an RO sounding. The information provided in the database contains information on the geolocation of the sounding and pointers to all of the data files relevant to that occultation, including calibratedPhase, refractivityRetrieval, atmosphericRetrieval files as contributed by all participating processing centers. In effect, the database has done all of the work in matching up occultations as processed by the different processing centers.

TABLE 1A

NetCDF variables for level 1b excess phase format calibratedPhase. The dimensions are (1) time, the number of high rate times recorded in the RO observation; (2) signal, the number of GNSS signals recorded simultaneously in the RO observation; (3) obscode, used in 3-character RINEX 3 code observation nomenclature, and (4) xyz, three elements that correspond to *x, y, z* Cartesian coordinates.

Variable	**Units**	**Description**
double startTime	GPS seconds[13]	Time of first data point at the receiver
double endTime	GPS seconds	Time of last data point at the receiver
byte navBitsRemoved(signal)		This logical variable declares whether the bit-stream of a navigation message was removed in the generation of excess phase. If non-zero, then the navigation message bits have been removed; if zero, then the navigation message bits have not been removed. If the signal is a data-less "pilot" tone, meaning there is no navigation message on the signal, the variable is meaningless but is set to True nonetheless.
char snrCode(signal,obscode)		RINEX 3 observation code[14] for the signal-to-noise ratio observation
char phaseCode(signal,obscode)		RINEX 3 observation code for the phase observation
double carrierFrequency(signal)	Hz	Carrier frequency of the tracked GNSS signal
double time(time)	GPS seconds	Observation times, the independent coordinate
double snr(time,signal)	V/V (1 Hz)	Signal-to-noise ratio
double excessPhase(time,signal)	m	Excess phase, or phase in excess of the vacuum optical path
double rangeModel(time,signal)	m	The model for pseudo-range used in open-loop tracking less transmitter and receiver clock biases
double phaseModel(time,signal)	m	The model for phase used in open-loop tracking
double positionLEO(time,xyz)	m	Receiver (low-Earth orbiter – LEO) satellite position in Cartesian Earth-centered fixed (ECF) coordinates[15]
double positionGNSS(time,xyz)	m	GNSS transmitter position in Cartesian Earth-centered fixed (ECF) coordinates at the time of transmission of the signal[16]

[13] GPS seconds is the number of seconds that have elapsed since 00:00 UTC on 6 January 1980 without an accounting for leap seconds. "00:00 UTC on 6 January 1980" corresponds exactly to "00:00 GPS on 6 January 1980". As of June, 2021, 18 leap seconds have elapsed, and thus GPS time leads UTC time by 18 seconds. Data on leap seconds can be downloaded from here.

[14] RINEX 3 observation codes are documented in section 5.1 of the RINEX 3.04 documentation.

[15] There are many definitions of Earth-centered fixed coordinate systems that differ by several meters at the surface of the Earth. Extreme precision in the positions of the satellites is unimportant in RO, as it is far more important that the *velocities* are precise to ~ 0.4 mm s^{-1}.

[16] The positions of the GNSS transmitter are given at the time of transmission of the signal that was received by the receiver at time "time". Recording the GNSS transmitter position at the time of signal transmission is the coordinate that directly enters into the RO retrieval process.

TABLE 1B

Global attributes in the calibrated excess phase file format calibratedPhase.

Global attribute	Description
char file_type	"GNSS-RO-in-AWS-Open-Data-calibratedPhase"
int year	Year of RO sounding
int month	Month of RO sounding [1-12]
int day	Day of month of RO sounding [1-31]
int hour	Hour of RO sounding (UTC) [0-23]
int minute	Minute of RO sounding [0-59]
float second	Seconds of RO sounding [0-59.99]
int doy	Day of year of RO sounding [1-366]
char mission	Name of the GNSS RO mission, defined in Table 5
char leo	Name of the receiver low-Earth-orbiting satellite, defined in Table 5
char occGnss	Three-character definition of the occulted GNSS satellite
char refGnss	Three-character definition of the GNSS satellite used in calibration if single-differencing or double-differencing is used in receiver clock calibration; empty string if zero-differencing is used
char refStation	Four-character definition of the ground station used in calibration if double-differencing is used in receiver clock calibration; empty otherwise
char processing_center	The contributing RO processing center [ucar, jpl, romsaf]
char processing_center_version	A string that identifies the version the retrieval system used to generate the file, as defined by the contributing processing center
char processing_center_path	The full path of the original file as contributed by the RO processing center to AER
char data_use_license	Website for the data use license for this file
char references	A list of digital object identifiers of papers that detail the retrieval method

TABLE 2A

NetCDF variables for level 2a bending angle and dry pressure and temperature retrieval format refractivityRetrieval. The dimensions are (1) xyz, corresponding to *x, y, z* Cartesian coordinates; (2) signal, the number of GNSS signals recorded simultaneously in the RO observation; (4) impact, the number of elements in the bending angle/impact parameter retrieved profiles; and (5) level, the number of vertical levels in the atmospheric retrieval.

Variable	Units	Description
double refTime	GPS seconds	The reference time of the occultation.
float refLongitude	degrees east	The reference longitude of the occultation
float refLatitude	degrees north	The reference latitude of the occultation
double equatorialRadius	m	The equatorial radius that describes the ellipsoid that approximates the Earth's mean sea level
double polarRadius	m	The polar radius that describes the ellipsoid that approximates the Earth's mean sea level
double undulation	m	The geoid undulation at the location of the RO sounding; add this quantity to the mean sea level ellipsoid described by equatorialRadius and polarRadius to determine the position of the mean sea level geoid
double centerOfCurvature(xyz)	m	The reference center of curvature for the occultation; reference_frame = ECEF
double impactParameter(impact)	m	Impact parameter is the independent coordinate of retrievals of bending angle
double carrierFrequency(signal)	Hz	Carrier frequency of the tracked GNSS signal
double rawBendingAngle(impact,signal)	radians	The bending angle for each signal, unoptimized, not fused with a model, not corrected for the ionospheric influence, positive for downward bending
double bendingAngle(impact)	radians	The unoptimized bending angle calibrated to eliminate ionospheric influence, positive for downward bending
double optimizedBendingAngle(impact)	radians	The bending angle calibrated to eliminate ionospheric influence and fused with a model by a statistical method, positive for downward bending
float altitude(level)	m	Altitude above the mean sea-level geoid
float longitude(level)	degrees east	Longitude of the occultation tangent point
float latitude(level)	degrees north	Latitude of the occultation tangent point
float orientation(level)	degrees	The direction of the occultation ray, transmitter to receiver, at the occultation tangent point, measured eastward from north
double geopotential(level)	J/kg	geopotential energy per unit mass at the occultation tangent point; divide by the WMO standard constant for gravity (J/kg/m) to obtain geopotential height (gpm)
double refractivity(level)	*N*-units	microwave refractivity at the occultation tangent point
double dryPressure(level)	Pa	Dry pressure at the occultation tangent point; it is the pressure retrieved when ignoring the contribution of water vapor to microwave refractivity, the equation of state, and the hydrostatic equation; see doi:10.5194/amt-7-2883-2014
double superRefractionAltitude	m	The altitude above the mean sea-level geoid of the highest super-refracting layer. If super-refraction is not analyzed, leave as fill values; if no super-refraction is found, set to -1000.0.

TABLE 2B

Global attributes in the dry retrieval file format refractivityRetrieval.

Global attribute	Description
char file_type	"GNSS-RO-in-AWS-Open-Data-refractivityRetrieval"
int year	Year of RO sounding
int month	Month of RO sounding [1-12]
int day	Day of month of RO sounding [1-31]
int hour	Hour of RO sounding (UTC) [0-23]
int minute	Minute of RO sounding [0-59]
float second	Seconds of RO sounding [0-59.99]
int doy	Day of year of RO sounding [1-366]
char mission	Name of the GNSS RO mission, defined in Table 5
char leo	Name of the receiver low-Earth-orbiting satellite, defined in Table 5
char occGnss	Three-character definition of the occulted GNSS satellite
char processing_center	The contributing RO processing center [ucar, jpl, romsaf]
char processing_center_version	A string that identifies the version the retrieval system used to generate the file, as defined by the contributing processing center
char processing_center_path	The full path of the original file as contributed by the RO processing center to AER
char data_use_license	Website for the data use license for this file
char optimization_references	A list of digital object identifiers of all papers that detail the statistical method for fusing data and model/climatology in the upper atmosphere; leave empty if no such fusion is employed
char ionospheric_references	A list of digital object identifiers of all papers that detail the methods implemented for removing ionospheric influence from bending angles; leave empty if no ionospheric correction is employed
char references	A list of digital object identifiers of all other papers that detail the retrieval method

TABLE 3A

NetCDF variables for level 2b retrievals of temperature, pressure, and water vapor in the atmosphericRetrieval format. The only dimension level, an index for the vertical structure of the retrieved and analyzed profile.

Variable	Units	Description
double refTime	GPS seconds	The reference time of the occultation
float refLongitude	degrees east	The reference longitude of the occultation
float refLatitude	degrees north	The reference latitude of the occultation
float altitude(level)	m	altitude above the mean sea-level geoid
float geopotential(level)	J/kg	Geopotential energy per unit mass at the occultation tangent point; divide by a standard constant for gravity to obtain geopotential height (gpm), typically the WMO standard (9.80665 J/kg/m)
float orientation(level)	degrees	The direction of the occultation ray, transmitter to receiver, at the occultation tangent point, measured eastward from north
float refractivity(level)	*N*-units	Observed microwave refractivity at the occultation tangent point
float pressure(level)	Pa	Atmospheric pressure retrieved by statistical methods using a prior
float temperature(level)	K	Atmospheric temperature retrieved by statistical methods using a prior
float waterVaporPressure(level)	Pa	Partial pressure of water vapor retrieved by statistical methods using a prior
double superRefractionAltitude	m	The altitude above the mean sea-level geoid of the highest super-refracting layer. If super-refraction is not analyzed, leave as fill values; if no super-refraction is found, set to -1000.0.

TABLE 3B

Global attributes in the dry retrieval file format atmosphericRetrieval.

Global attribute	Description
char file_type	"GNSS-RO-in-AWS-Open-Data-atmosphericRetrieval"
int year	Year of RO sounding
int month	Month of RO sounding [1-12]
int day	Day of month of RO sounding [1-31]
int hour	Hour of RO sounding (UTC) [0-23]
int minute	Minute of RO sounding [0-59]
float second	Seconds of RO sounding [0-59.99]
int doy	Day of year of RO sounding [1-366]
char mission	Name of the GNSS RO mission, defined in Table 5
char leo	Name of the receiver low-Earth-orbiting satellite, defined in Table 5
char occGnss	Three-character definition of the occulted GNSS satellite
char prior	The atmospheric prior that was used to resolve the wet-dry ambiguity; for example, "ERA5 forecasts", "NCEP GFS forecasts", "MERRA-2", etc.
char processing_center	The contributing RO processing center [ucar, jpl, romsaf]
char processing_center_version	A string that identifies the version the retrieval system used to generate the file, as defined by the contributing processing center
char processing_center_path	The full path of the original file as contributed by the RO processing center to AER
char data_use_license	Website for the data use license for this file
char references	A list of digital object identifiers of all other papers that detail the retrieval method

TABLE 4

Mnemonics used in the definitions of paths and occultation identifiers.

Mnemonic	Description	Examples
center	The RO retrieval center that contributed the data	"ucar", "jpl", "romsaf"
mission	The RO mission	See Table 2
filetype	The file type	"calibratedPhase", "refractivityRetrieval", "atmosphericRetrieval"
yyyy	The year of the RO sounding	1995 … 2020, etc.
mm	The zero-padded month of the RO sounding	01, 02, … 12
dd	The zero-padded day of the month of the RO sounding	01, 02, … 31
hh	The zero-padded hour of the day	00, 01, … 23
nn	The zero-padded minute	00, 01, … 59
version	A string defining the processing version	(defined by contributing center, without underscores)
occid	The occultation ID as registered in the AWS OpenData program	See definition in the text
suffix	A suffix defining the file format; "nc" for NetCDF	"nc" for calibratedPhase, refractivityRetrieval, atmosphericRetrieval

TABLE 5

Definition of the names of GNSS RO missions and the names of the receiver satellites.

Mission	Receivers/LEOs	Long name
gpsmet	gpsmet, gpsmetas	GPS/MET FORMOSAT-3
grace	gracea, graceb	Gravity Recovery and Climate Experiment (GRACE)
sacc	sacc	Satellite de Aplicaciones Cientifico-C (SAC-C)
champ	champ	Challenging Mini-satellite Payload (CHAMP)
cosmic1	cosmic1c1, cosmic1c2, … cosmic1c6	Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC)
tsx	tsx	Terra Synthetic Aperture Radar - X (TerraSAR-X)
tdx	tdx	TerraSAR add-on for Digital Elevation Measurement (TanDEM-X)
cnofs	cnofs	Communications/Navigation Outage Forecasting System (C/NOFS)
metop	metopa, metopb, metopc	Metop-A, Metop-B, Metop-C
kompsat5	kompsat5	Korean Multi-Purpose Satellite 5 (KompSat 5)
paz	paz	Radio Occultations and Heavy Precipitation with PAZ (ROHP-PAZ)
cosmic2	cosmic2e1, cosmic2e2, … cosmic2e6	Constellation Observing System for Meteorology, Ionosphere and Climate 2 (COSMIC-2)

TABLE 6

List of the contributing, independent RO processing centers.

Processing center	**“center”**
UCAR COSMIC Project Office	ucar
ROM SAF	romsaf
Jet Propulsion Laboratory, California Institute of Technology	jpl