GNSS RADIO OCCULTATION IN THE AWS CLOUD: BACKGROUND

S. Leroy (1), A. McVey (1), S. Leidner (1), J. Martin (1), H. Zhang (2), D. Hunt (2), S. Sokolovoskiy (2), C. Ao (3), K.-N. Wang (3), M. Oyola-Merced (3), H. Gleisner (4), S. Syndergaard (4), and K. Lauritsen (4)

(1) Verisk Atmospheric and Environmental Research; (2) University Corporation for Atmospheric Research; (3) Jet Propulsion Laboratory, California Institute of Technology; (4) Danish Meteorological Institute

The amount of GNSS radio occultation (RO) data that has accumulated since the proof-of-concept mission GPSMET/FORMOSAT-3 has hit approximately 20 million soundings. At least three fully functional and independent RO retrieval centers have processed these data; they have processed the data with algorithms that have been nonhomogeneous in time; the formats of the output are not only different amongst the retrieval centers but they are non-homogeneous even for a single center; the existing data formats are incompletely documented; no agreed upon standard for identifying a unique RO sounding exists; and different retrieval centers process different subsets of the RO data that have been recorded. The current paradigm of analyzing the data by downloading them onto local servers is proving increasingly cumbersome because of the amount of data, inadequate download interfaces and subsetting capabilities, and limits on bandwidth. We set out to address these problems by hosting as much RO data as are available and processed by as many independent retrieval centers as wish to participate. Current contributors are UCAR, JPL, and the ROM SAF. The data are hosted in the Amazon Web Services Open Data Registry, available for simple browsing, and the data formats are documented in Github. The data are free to browse and download.

We have defined new, AWS-native data formats into which we have translated all contributions of RO data: calibratedPhase contains excess phase, amplitude, and orbit data; refractivityRetrieval contains bending angle, refractivity, and dry atmosphere retrieved data; and atmosphericRetrieval contains pressure, temperature, water vapor, and height retrieved data. Translating required extended study of the details of the various generations of data formats, investigating details of RO retrieval systems, and documenting RO receiver tracking algorithms. Unique identifiers were assigned to RO soundings, and soundings and their metadata recorded in a database to force RO sounding matchups across retrieval centers. The resulting database can be used to select RO soundings depending on transmitter, mission, geolocation, time, sounding geometry, and contributing retrieval center. The result is an archive that enables efficient handling of the data especially when working in an AWS computing environment.

We will discuss details of the archive, problems solved, some simple science applications, and its capabilities in this presentation.