https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Cluster Analysis of Vertical Polarimetric Radio Occultation Profiles and Corresponding Liquid and Ice Water Paths From GPM Microwave Data

Jonas Katona[1,2], Manuel de la Torre Juárez[2], Terence L. Kubar[2,3], F. Joseph Turk[2], Kuo-Nung Wang[2], and Ramon Padullés[4]

[1]Yale University, Applied Mathematics Program
[2]NASA Jet Propulsion Laboratory, California Institute of Technology
[3]University of California, Los Angeles, Joint Institute for Regional Earth Systems Science and Engineering
[4]Institut de Ciencies de l'Espai, Consejo Superior de Investigaciones Cientificas

Correspondence: Jonas Katona (jonas.katona@yale.edu) and Manuel de la Torre Juárez (mtj@jpl.nasa.gov)

Abstract. The polarimetric phase difference between the horizontal and vertical components of GNSS radio signals is correlated with the presence of ice and precipitation in the propagation path of those signals. This study evaluates the ability of k-means clustering to find relationships among polarimetric phase difference, refractivity, liquid water path (LWP), ice water path (IWP), and water vapor pressure using over two years of data matched between the Global Precipitation Measurement (GPM) mission and Radio Occultations through Heavy Precipitation demonstration mission onboard the Spanish Paz spacecraft (ROHP-PAZ). A cluster hierarchy is introduced across these variables. A potential refractivity model for polytropic atmospheres is introduced to ascertain how different types of vertical thermodynamic profiles that can occur during different precipitation scenarios are related to changes in the polytropic index and thereby vertical heat transfer rates. The clustering analyses uncover a relationship between the amplitude and shape of deviations from the potential refractivity model and water vapor pressure and confirm the expected positive correlation between polarimetric phase difference and both LWP and IWP. For certain values, the coefficients of the potential refractivity model indicate when a profile has little to no moisture, and the study reveals a similar relationship between the clustering for these coefficients and different water vapor pressure profiles. The study also confirms the relationship between the integrated polarimetric phase difference and water vapor pressure columns, known as the "precipitation pickup," globally ($\rho_s = 0.971$ after averaging) and over different latitudinal ranges ($> 50°$, $\geq 20°$, and $< 20°$, with different ρ_s for each).

1 Introduction

GNSS (Global Navigational Satellite System) refers to the collection of satellites orbiting Earth that periodically send cirularly polarized radio signals indicating their positions globally. As these satellites occult from a low Earth-orbiting satellite with a GNSS receiver, the radio signal they receive has been refracted and bent by the atmosphere. The bending angle is caused by the atmospheric refractivity gradient in the region through which the signal traveled. The degree of bending can be calculated using the geometry between the emitting satellite and a receiver, and the shift in frequency of the signal between when it is

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

emitted and received. Hence, these GNSS radio occultations (RO) can provide us with the refractivity, N, which is related to pressure (p, in hPa), temperature (T, in K), and water vapor pressure (e, in hPa) as follows (e.g., Smith and Weintraub, 1953; Kliore et al., 1974) for an atmospheric air composition with approximately 78 percent nitrogen and 21 percent oxygen containing water:

$$N = \frac{k_1 p}{T} + \frac{k_2 e}{T^2}, \tag{1}$$

where $k_1 = 77.6$ and $k_2 = 3.73 \times 10^5$.

Quantities derived from RO have demonstrated high accuracy and resolution in space (e.g., Kursinski et al., 1997; Huang et al., 2010; Son et al., 2017). RO temperatures derived from refractivity have been shown to be of similar quantitative accuracy as temperatures directly measured by radiosondes, which are mostly limited to land (e.g., Nishida et al., 2000; Randel et al., 2003; Schmidt et al., 2004; Kim and Son, 2012).

One of the most powerful applications of RO has been in understanding climatic variability and trends—including intraseasonal-to-interannual atmospheric modes of variability such as the quasi-biennial oscillation (QBO), Madden–Julian oscillation (MJO), and El Niño–Southern Oscillation (ENSO)—as they relate to atmospheric structure over the tropics (Scherllin-Pirscher et al., 2021), especially in the upper-troposphere–lower-stratosphere (UTLS) region (Schmidt et al., 2004; Lackner et al., 2011).

RO observations have also been used to uncover and measure the upper-level thermal structures of deep convection in tropical storms both alongside and without precipitation radar data (Biondi et al., 2012; Xian and Fu, 2015; Scherllin-Pirscher et al., 2021).

Eq. (1) shows that using RO refractivity data to retrieve thermodynamic variables such as temperature, pressure, and water vapor remains underconstrained. Water vapor information is extracted from refractivity by assuming that the temperature profiles of a given weather analysis, ECMWF or NCEP, are correct at the location of each RO profile, even in cases where the RO and model refractivity may differ (e.g., Kursinski et al., 1997; Kuo et al., 2001). An inaccurate refractivity profile from the analysis will lead to erroneous water vapor retrievals. Because RO has a more valuable contribution to model improvement precisely in the profiles where the weather analysis and RO differ, the relationship between water vapor and refractivity has a higher error bar particularly in the most useful profiles. Moreover, GNSS RO measurements are sensitive to variations in temperature and water vapor within clouds (Kuo et al., 2001; Huang et al., 2010), but require other observables to confirm the presence of clouds and understand their structure.

Polarimetric radio occultation (PRO) provides a way to expand the applications of standard RO. PRO uses circularly polarized GNSS radio signals to measure atmospheric anisotropies like precipitating droplets and ice crystals, as these induce a phase difference between the horizontal (H) and vertical (V) components of the GNSS radio signal. In particular, the polarimetric phase difference, $\Delta\Phi$, between H and V is related to the amount of precipitation or ice in the atmosphere (Tomás et al., 2018; Cardellach et al., 2019; Wang et al., 2022; Padullés et al., 2023) using $\Delta\Phi$ and has promising applications in weather model assimilation (Hotta et al., 2023), climate monitoring (Cardellach et al., 2019; Gleisner et al., 2022), and atmospheric research (Turk et al., 2021; Padullés et al., 2023). Datasets from GNSS-PRO contain data on refractivity and $\Delta\Phi$, both as functions of

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

height. Unlike infrared instruments, PRO gives data even inside clouds with a higher vertical resolution than microwave (e.g. Turk et al., 2019).

Statistical correlations as a function of height between CloudSat integrated water content (or water path) along the radio occultation ray path and $\Delta\Phi$ are strong. There are models for how a given thermodynamic state of the atmosphere will affect a propagating RO signal and cause a $\Delta\Phi$ (e.g., Padullés et al., 2023, and references therein). However, a precise formula is missing for how a measured $\Delta\Phi$ relates to thermodynamic atmospheric states. Part of the challenge is that a given $\Delta\Phi$ at a specific height may be caused by both ice or precipitation. Therefore, we explore if different vertical distributions of precipitation- or moisture-related variables—$\Delta\Phi$, liquid water path (LWP), ice water path (IWP), and water vapor pressure—are interrelated.

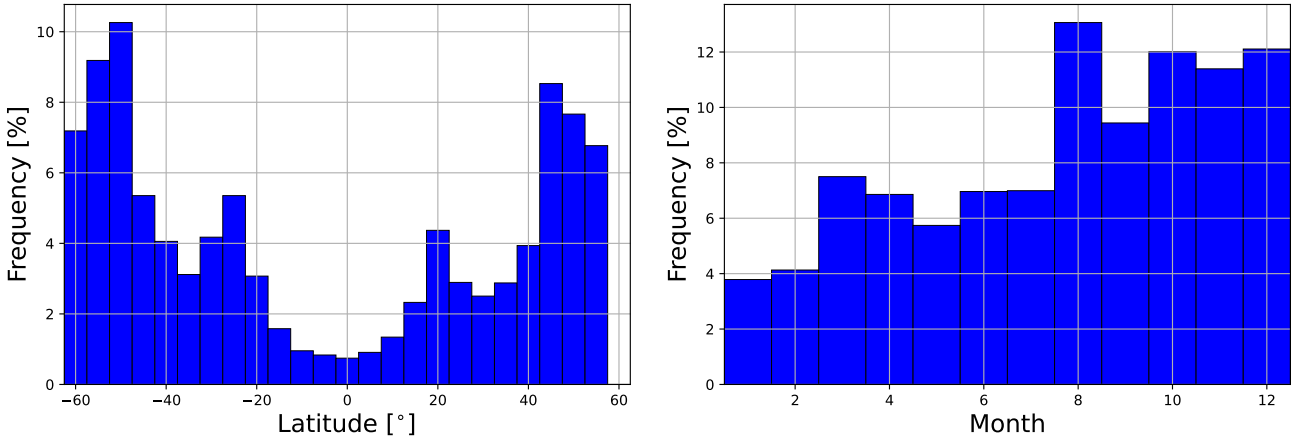
This study looks at how the vertical shape of $\Delta\Phi$ along the RO ray correlates with that of other thermodynamic variables such as refractivity, water vapor pressure, liquid water path (LWP), and ice water path (IWP) along the ray as functions of height at given latitudes and longitudes. A k-means cluster analysis is performed to see how cluster centroids relate to physical phenomena across different variables, the variables being the aforementioned ones and a physically interpretable model for potential refractivity similar to the one introduced in de la Torre Juárez et al. (2018). This analysis also looks at how the vertical integral of $\Delta\Phi$ relates to total column water vapor, and how this confirms results and observations from prior studies. We explore if vertical profiles of $\Delta\Phi$ and refractivity can help to distinguish possible thermodynamic states and even the contributions from ice vs. liquid water precipitation. Through new statistical and graphical analyses below, it is hoped to help understand and quantify these relationships.

To this end, in Section 2, we describe the dataset; in Section 3, we outline how we classify different thermodynamic states from refractivity profiles alone and provide an overview of how we apply k-means clustering to different variables, from PRO-derived refractivity and $\Delta\Phi$, to model-inferred water vapor, water path, and ice path; in Section 4, we use our cluster to search for a classification of disparate vertical structures and cross-correlate interpretations of clusters for different variables; and in Section 5, we summarize the aims and results of our study.

2 Data

The two datasets analyzed and used to train the data classification and the model (3) are Level 2 Global Precipitation Measurement (GPM) data from the NASA Goddard Space Flight Center and Level 2 Radio Occultations and Heavy Precipitation data from the PAZ satellite (ROHP-PAZ) (Cardellach et al., 2019). From the former, we retrieve the pressure (hPa), water vapor pressure (hPa), Liquid Water Path (LWP, kg/m^2), and Ice Water Path (IWP, kg/m^2), while the latter gives refractivity (N-units, measured) and $\Delta\Phi$ (mm), all as functions of height at different latitudes, longitudes, and times. Temperature (K) in turn is derived from Eq. (1). For more details on how the aforementioned variables are retrieved from Level 0 and Level 1 datasets, we refer the reader to Turk et al. (2021) and the references therein for the GPM dataset and Cardellach et al. (2019) for the ROHP-PAZ dataset.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) Histogram of latitudes

(b) Histogram of months

Figure 1. Sampling distribution in the dataset across different a) latitudes and b) times.

The GPM and ROHP-PAZ profiles are matched across different latitudes, longitudes, and times whenever they coincided within a given spatiotemporal range (e.g. Turk et al., 2021). We have

- 2362 coincidences from July 26th, 2018 to December 31st, 2018;
- 2943 coincidences from March 1st, 2019 to December 31st, 2019; and
- 1401 coincidences from January 1st, 2020 to August 22nd, 2020;

which yields a total of 6706 coincidences from July 26th, 2018 to August 22nd, 2020.

Most of the aforementioned coincidences lie poleward of 40° N or S, as shown in Figure 1(a), enabling good statistics in those regions. There was a low number of coincidences in the tropics (within 15° of the equator) which constrains our analysis in low-latitude regions. Furthermore, as Figure 1(b) shows, there is also a slightly higher number of coincidences in the last four months of the year vs. the first eight, but this poses less of a problem as we do not assess seasonality.

Turk et al. (2021) computed LWP and IWP by integrating the condensed water content (kg/m^3)—estimated via emissivity principal components passive microwave precipitation profiling (Turk et al., 2018; Utsumi et al., 2020)—along each RO ray path in the ROHP dataset coinciding with GPM data. Integrating the condensed water content along the ray paths ensures that their values are related to $\Delta\Phi$, which is also computed by integrating along each ray path. By checking when the retrieved temperature is above or below 273 K, we partition this integrated water content into LWP and IWP, respectively. Finally, as with $\Delta\Phi$, the values of the LWP and IWP at a given latitude, longitude, and height are given according to where the lowest level tangent point for the given ray path lies.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

To compute the total column water and ice paths from the aforementioned data for each profile, the water and ice paths are integrated, respectively, from 1 km to 10 km only if a profile has data at 1 and 10 km.[1] For computing the vertical integral of $\Delta\Phi$, since the error associated with this variable in the ROHP-PAZ dataset is roughly ± 2 mm at each height, $\Delta\Phi$ is integrated from 2.5 km to 10 km after rounding $\Delta\Phi$ to the nearest multiple of 2 mm if there exist data from 2 to 2.5 km and at 10 km. The latter condition ensures that the endpoints of the integral are correct and we exclude faulty retrievals which tend to deteriorate near the bottom of the profiles before the data become corrupted or missing. Finally, for computing the total column water vapor, the water vapor pressure is integrated from 2.5 km and 10 km, excluding profiles which feature no data at 2.5 km or 10 km, negative water vapor pressure values, or water vapor pressure values above 300 hPa—these situations are unphysical and likely result from model and/or retrieval errors. The number of profiles where these conditions were not met are as follows:

- For total column water vapor: 33 profiles (0.49% of all profiles in the dataset);
- For total column water path: 1 profile (0.01%);
- For total column ice path: 6 profile (0.09%);
- For total column water+ice path: 6 profiles (0.09%); and
- For the vertical integral of $\Delta\Phi$: 923 profiles (13.76%).

For all cases, the integration is implemented in Python using the composite trapezoidal rule (Atkinson, 1988).

3 Methods

The PRO observables are $\Delta\Phi$ and refractivity as functions of height, latitude, and longitude. Hence, this study explores how far one can get with PRO observables while remaining as independent from externally derived weather analyses as possible. To this end, we develop a model for potential refractivity as a function of height assuming a constant lapse rate (which can be non-adiabatic), hydrostatic balance, and a constant water vapor mixing ratio.

3.1 Potential refractivity in a polytropic atmosphere

A first classification criterion organizes profiles based on the differences between observed refractivity profiles and those expected for polytropic atmospheres in which air can expand and compress with adiabatic and non-adiabatic heat transfer.

If an air parcel moving vertically through the atmosphere follows a polytropic process—as occurs in a *polytropic atmosphere*—and the ideal gas law holds, then p/ρ^m and therefore $p^{1-m}T^m$ are constant, where m is the polytropic index of the atmosphere. Hence, we define $K := p^{1-m}(z)\,T^m(z) = p(z_0)^{1-m}T(z_0)^m$ for some reference height z_0.

[1] While requiring path data down to 1 km may seem too stringent, requiring this only ends up excluding six profiles at most for both water and ice path, or under 0.1% of all the profiles in the dataset. Thus, it is not too stringent unless these six profiles happen to be rather extreme cases. The analysis at hand aims at finding general trends and associations rather than specific, individual cases.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

In hydrostatic balance, we have $\frac{\partial p}{\partial z} = -\rho g$, and polytropy also implies that $\frac{\partial p}{\partial z} = \frac{\partial (K\rho^m)}{\partial z} = mK\rho^{m-1}\frac{\partial \rho}{\partial z}$. Balancing these two equations necessitates that $-\rho g = mK\rho^{m-1}\frac{\partial \rho}{\partial z}$, and after multiplying both sides by $\frac{m-1}{mK\rho}$, one gets

$$g\frac{1-m}{mK} = \frac{\partial \rho^{m-2}}{\partial z}.$$

At constant $m \neq 0, 1$, the solution is

$$\rho^{m-1}(z) = \rho(z_0) + g\frac{1-m}{m}(z - z_0). \tag{2}$$

At $m = 0$, the pressure is constant and cannot satisfy hydrostatic equilibrium unless $\rho = 0$, while at $m = 1$, the density decays exponentially, typical of an isothermal atmosphere. When $m = \gamma$, where $\gamma = 5/3$ is the adiabatic index, the change of temperature incurred by air parcels moving vertically in this atmosphere follows an adiabatic process—an *adiabatic atmosphere*.

Using Eq. (2) for the vertical profile of an ideal gas, where $\frac{p(z)}{\rho(z)} = RT(z)$, and by polytropy again, $p(z) = K\rho^m(z) = K\rho(z)\rho^{m-1}(z)$ implies that

$$RT(z) = \frac{p(z)}{\rho(z)} = K\rho^{m-1}(z) = K\left[\rho(z_0) + g\frac{1-m}{m}(z - z_0)\right].$$

This shows that an ideal gas atmosphere in hydrostatic equilibrium and with constant polytropic index $m \neq 0$ with height has a linear temperature profile $T(z) = \widehat{T}(z_0) - \widehat{\Gamma}(z - z_0)$, where $\widehat{T}(z_0) = \frac{K\rho(z_0)}{R}$ and $\widehat{\Gamma} = -Kg\frac{m-1}{mR}$. When $m = 1$, the solution holds with $\widehat{\Gamma} = 0$ and a constant temperature with height. At constant m and R, $\widehat{\Gamma} = -\frac{\partial T}{\partial z}$, and hence, the lapse rate is constant.

When including water vapor processes, one can characterize the temperature profiles in a polytropic atmosphere as 1) a completely dry atmosphere with (virtually) no water, 2) an *unsaturated* moist atmosphere (i.e., containing non-negligible water but not saturated). Additionally, one can approximate temperature via a linear relationship with height for 3) a *saturated* moist atmospheric layer where the expansion and contraction of air is reversible or 4) an atmosphere in which water that condenses in an air parcel is instantaneously removed via precipitation—a pseudoadiabatic atmosphere (e.g., Emanuel, 1994). The lapse rate, Γ, is called a dry adiabatic lapse rate in the first case, a moist-unsaturated adiabatic lapse rate in the second, a reversible moist-adiabatic lapse rate in the third, and a pseudoadiabatic lapse rate in the fourth. The temperature profile is precisely linear with height for only the first case and close to linear in the others.

Each of the four thermodynamic cases above would be represented by a different conservation law (Emanuel, 1994): dry adiabatic (for 1), moist adiabatic (for 2 and 3), pseudoequivalent potential temperatures (for 4), and, by analogy, via a different type of potential refractivity profile. These conserved quantities can be used to define different types of potential refractivity, $\widehat{N}$, based on fitting data to physical laws describing adiabatic and pseudoadiabatic processes (e.g. de la Torre Juárez et al., 2018).

$\widehat{N}$ is derived for an atmosphere with the following properties: 1) Eq. (1); 2) the ideal gas law; 3) a linear temperature profile with height representative of a polytropic atmosphere; 4) a constant specific humidity representative of a subsaturated atmosphere; and 5) in hydrostatic equilibrium. Deviations between the measured refractivity N and the fit to the model $\widehat{N}$ signal the presence of changes in mixing ratio, precipitation, or non-equilibrium physics (e.g., gravity waves or turbulence).

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

From the above assumptions, we derive in Appendix A the model for $\widehat{N}$:

$$\widehat{N}(z) = \frac{N(z_0)}{[1-c_1(z-z_0)]^2} \times \{(1-c_2)[1-c_1(z-z_0)]^{c_0} + c_2\}, \tag{3}$$

where $c_0 = \frac{g}{R\widehat{\Gamma}} + 1 = \frac{2m-1}{K(m-1)}$, $c_1 = \frac{\widehat{\Gamma}}{\widehat{T}_0} = \frac{g(m-1)}{m\rho(z_0)}$, and $c_2 = \frac{k_2\widehat{e}_0}{N(z_0)\widehat{T}_0^2}$ are coefficients which must be fit to a given refractivity profile and provide information about the polytropic index. $\widehat{T}_0 := T(z_0)$ & $\widehat{e}_0 := e(z_0)$ are the temperature and water vapor pressure, respectively, at reference height z_0. In particular, for $m = 1$, $\widehat{N}$ has an exponential relationship with z (e.g. Bean and Dutton, 1966).

We initialize $z_0 = 2.5$ km, since this height is in the domain for every profile we are considering and should be part of the range of heights where we would expect the model assumptions to hold approximately. The fit coefficients $\mathbf{c} = (c_0, c_1, c_2)$ are defined in terms of the following physical parameters: the acceleration due to gravity on Earth $g = 9.81\ \mathrm{m \cdot s^{-2}}$, specific gas constant of dry air $R = 287.05\ \mathrm{J \cdot kg^{-1} \cdot K^{-1}}$, mean tropospheric lapse rate $\widehat{\Gamma}$ (in $\mathrm{K \cdot km^{-1}}$), and constants k_1 and k_2 defined in Section 1.

The lapse rate can change with height across moist and dry sections, e.g., in the transition between the boundary layer and the free atmosphere (von Engeln et al., 2005; Ao et al., 2012), in the transition from the mid-troposphere to the tropical tropopause layer (TTL) (Fueglistaler et al., 2009), or when clouds are present in a real atmosphere (e.g., Peng et al., 2006; Mascio et al., 2021). Based on these three examples, the fits for $\widehat{N}$ are made in an altitude range that is likely to have a constant lapse rate under the five assumed properties from the last subsection: 2.5 km to 200 m below the estimated tropopause.

For this study, the tropopause was estimated by finding where $\left|\frac{\partial T}{\partial z}\right|$ is minimized for all heights above 5 km and where the temperature is within 10 K of the minimum temperature below 25 km, i.e., within 10 K of the temperature at the cold-point tropopause. Second-order central differences were used to estimate $\frac{\partial T}{\partial z}$ across all of the heights for each given profile (Atkinson, 1988). We use the GPM temperature to estimate T to establish a rigorous criterion across all of the profiles and ensure that the fit $\widehat{N}$ is consistently being used where it would be expected to hold, especially for accurate clustering in $N - \widehat{N}$.

More details on the numerical fitting of Eq. (3) are given in Appendix B.

3.2 Time series k-means clustering

Across the profiles in the merged dataset described above, we apply k-means clustering with $k = 8$ clusters for each of the following variables:

- RO measured variables: $\Delta\Phi$ (2.5 to 10 km), $N - \widehat{N}$ (2.5 to 8 km), the three fit coefficients for $\widehat{N}$ (the vector $\mathbf{c}$), and
- Variables from ancillary data: RO+model-derived water vapor pressure (2.5 to 10 km) and GPM+RO ray path computed liquid water path (LWP, 1 to 10 km), ice water path (IWP, 1 to 10 km), and total (liquid+ice) water path (TWP, 1 to 10 km).

In all cases aside from the clustering for coefficients (in which case we use standard k-means clustering with the standard Euclidean distance), a variation of naive k-means clustering called time series k-means with dynamic time warping (DTW)

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

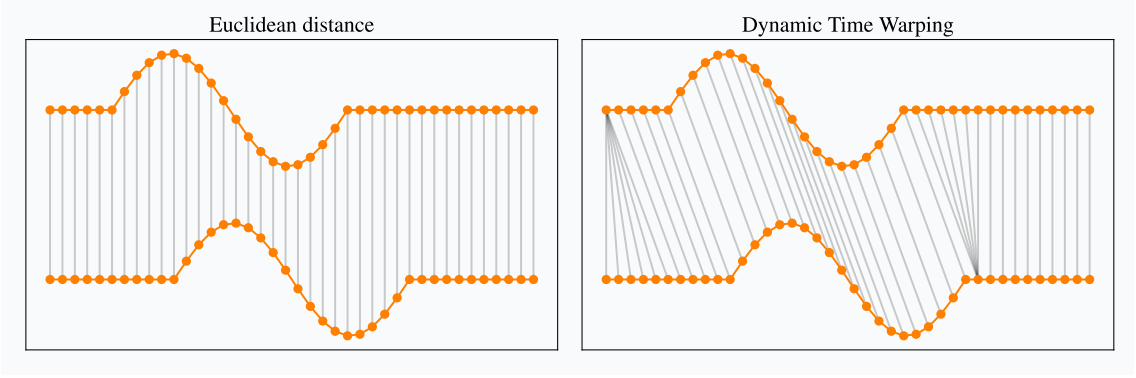


Figure 2. A visual comparison (Tavenard, 2021) showing the difference between Euclidean distance and the DTW measure. Time series are shifted vertically in the visualization, but assume that the y-axis values match. Thus, between the two time series shown, the Euclidean distance would be nonzero but the DTW measure would be zero.

(Izakian et al., 2015) is applied. As with naive k-means, the dataset is partitioned into k clusters, but instead of measuring the distances between profiles using the Euclidean distance, DTW is used.

The numerical procedure for running k-means clustering on the aforementioned variables is described in Appendix C.

We introduce quality-control criteria for each of the variables informed by how k-means clustering detected outliers and other physical considerations, e.g., $\Delta\Phi$ profiles when the retrieval for $\Delta\Phi$ cut off above 2 km or e clusters featuring profiles with nonphysically high water vapor pressure values (> 300 hPa). Including these faulty profiles affect the accuracy when we compare the shapes of the $\Delta\Phi$ profiles for clustering and compute the integral of $\Delta\Phi$. In particular, faulty retrievals tend to deteriorate near the bottom of the profiles before the data become corrupted or missing. The percent of profiles excluded ranged from 0.01% (for LWP clustering) to 13.76% (for $\Delta\Phi$ clustering). See Appendix D for more details on the precise quality-control criteria used for each clustering variable.

3.3 Dynamic time warping

DTW is a technique originating in time series analysis that measures the similarity between two signals which are functions of time (or some analogous variable—in this case, height) by finding an optimal alignment between the two signals by "warping" the sample points of each signal such that the measurements in each signal are matched to their nearest point(s) in the other signal as measured by the Euclidean norm, regardless of the times at which each point was measured (Müller, 2007). We still assume that the start and end points match in each case, that the ordering of measurements (with time) within each profile stay the same, and that each point in one signal is matched to at least one point in the other. This ensures the following:

1. For cases of missing or uneven data points within a given profile, we can still compare the rough shape of this profile with others, and

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

2. For translations in sampling (e.g., when two measurements are out of phase or when recorded heights are imprecise), DTW can make up for this by shifting the heights at which measurements are taken when comparing two profiles.

See Tavenard (2021) or Müller (2007) for more details on how DTW is calculated.

Figure 2 features an intuitive visualization of how DTW works when comparing time series. The featured example is taken from Tavenard (2021) and shows two signals consisting of horizontal lines combined with one period of a sinusoid. Note how DTW matches the patterns and overall shape of each time series, which intuitively should result in a more sound similarity assessment than when using the Euclidean distance, since the latter matches timestamps (or heights for this study) regardless of when they were sampled.

4 Results and analysis

Clustering provides an initial classification for the types of atmospheric profiles that can occur across the dataset by looking at the centroids in different clustering variables as in Fig. 5. The clustering centroids represent the general magnitude and shape of the profiles belonging to each cluster.

A second step in the analysis uses frequency histograms of different cluster groups—Tables 2, 3, and 4—to summarize the relationships between clusters in $N - \widehat{N}$ with water vapor pressure, the $\widehat{N}$ coefficients $\mathbf{c}$ with water vapor pressure, and $\Delta\Phi$ against the path variables (LWP, IWP, and TWP). These tables gauge the ability of $N - \widehat{N}$ to predict different distributions of vertical water vapor pressure and $\Delta\Phi$ to predict different types of water path profiles across the vertical profiles in the dataset.

4.1 Total column $\Delta\Phi$ and total column water vapor

Table 1. Pearson's correlation coefficient (r), Spearman's rank correlation coefficient (ρ), and Kendall's rank correlation coefficient (τ) on all pairs of the accumulated $\Delta\Phi$ vs. total column water vapor for the raw dataset (Table 1(a)) and for the moving averages (Table 1(b)) across varying latitudinal ranges. Each correlation coefficient has a p-value below 10^{-9}, indicating a high statistical significance for all coefficients.

Lat. range: → Cor. coef.: ↓	All	$> 50°$	$\geq 20°$ and $\leq 50°$	$< 20°$
Pearson's r_p	0.332	0.315	0.349	0.375
Spearman's ρ_s	0.216	0.223	0.206	0.287
Kendall's τ_k	0.147	0.151	0.139	0.194

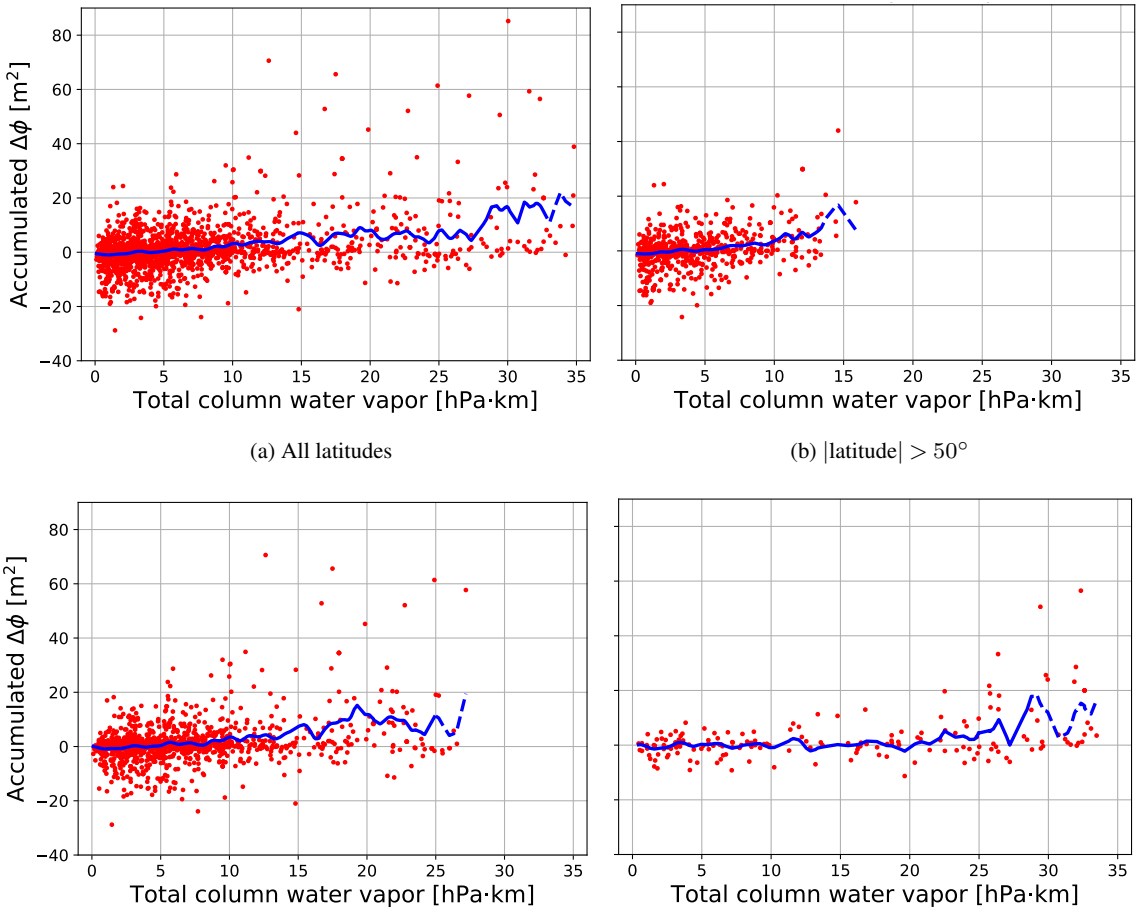
(a) Correlation tests on the raw dataset

Lat. range: → Cor. coef.: ↓	All	$> 50°$	$\geq 20°$ and $\leq 50°$	$< 20°$
Pearson's r_p	0.940	0.901	0.921	0.708
Spearman's ρ_s	0.971	0.964	0.947	0.683
Kendall's τ_k	0.864	0.847	0.803	0.508

(b) Correlation tests on the moving averages

Bretherton et al. (2004) showed a relationship between precipitation and total column water vapor over the tropics. Later studies (Muller et al., 2009; Holloway and Neelin, 2010; Emmenegger et al., 2022) demonstrate a positive relationship between precipitation and total column water vapor in the tropics, where under a certain total column water vapor value, precipitation is generally negligible in a given profile, and above a "pickup" threshold, precipitation *may* become non-negligible and increases exponentially. To test the validity of $\Delta\Phi$ as a proxy for precipitation (Padullés & Turk, private communication) and the ade-

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) All latitudes

(b) |latitude| $> 50°$

(c) |latitude| $\geq 20°$ and |latitude| $\leq 50°$

(d) |latitude| $< 20°$

Figure 3. Moving averages (blue) of accumulated $\Delta\Phi$ vs. accumulated water vapor pressure over scatter plots (red) across all latitudes and in different latitudinal ranges. These moving averages were done using the `filter1d` tool Generic Mapping Tools (`gmt`) Version 6.3 (Wessel et al., 2019). Averaging was done with a Gaussian filter of width 2 hPa · km (option `-Fg2`) excluding outputs where the input data has a gap exceeding 0.2 (option `-L0.2`) and including ends of the time series in the output (option `-E`). Dashed portions of the moving averages correspond to where each bin had less than 34 data points.

quacy of our dataset, we look for the precipitation pickup pattern (Holloway and Neelin, 2010) in the relationship between the total column water vapor and the total column of the PRO observable $\Delta\Phi$.

235 Fig. 3(d) shows insufficient data in the tropics to replicate their results with much fidelity ($r_p = 0.708$ for latitudes $< 20°$) compared to the higher latitudes shown in Fig. 3(b), as shown by the dashed line in Fig. 3(d).[2] Nonetheless, the positive correlation between accumulated $\Delta\Phi$ and total column water vapor was found across *all* latitudes ($r_p = 0.940$) and three latitudinal ranges; see Table 1(b). There is also an apparent total column water vapor threshold after which $\Delta\Phi$, the PRO

[2] 34 counts per bin was chosen as a consistent threshold for all plots in Fig. 3 to show where the density of data falls below a given reference value.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

signature of precipitation, starts increasing at a faster rate for latitudes poleward of $50°$; c.f. Bretherton et al. (2004); Muller et al. (2009) for the tropics. Due to sparse data, this threshold is difficult to identify in the tropics, as the dashed lines in Fig. 3(d) suggest, but still weakly present.

The strength of the correlation between accumulated $\Delta\Phi$ and total column water vapor also depends on which data the correlation analyses ran. The correlation coefficients in Table 1(a) indicate a low positive correlation between accumulated $\Delta\Phi$ and total column water vapor in the raw dataset. After applying the Gaussian filter with results in Fig. 3 and running correlation analyses on the filtered data, we find a high positive correlation between the same two quantities in Table 1(b). This suggests a relationship between the total column $\Delta\Phi$ and water vapor pressure.

Figures 3(b) and 3(d) show how the rough threshold past which the total column water vapor induces precipitation, i.e., the critical level at which the CWV pickup starts, is around half of what it is in the tropics (roughly 12-13 vs. 25-26 hPa $\cdot$ km).

4.2 $N - \widehat{N}$ and water vapor pressure

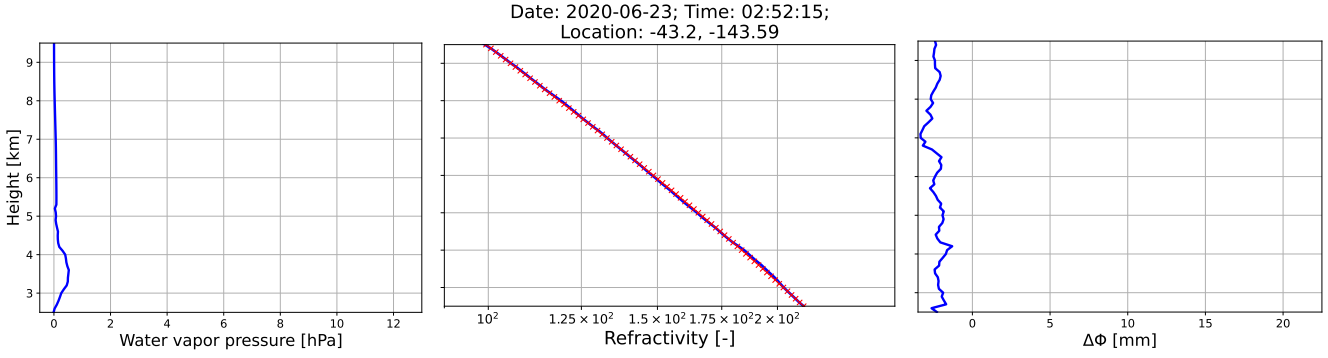
We represent the deviations of N from a profile with the properties outlined in Section 3.1 by looking at overlaid graphs of N and $\widehat{N}$ as functions of height and by plotting $N - \widehat{N}$ as a function of height. Fig. 4(b) shows that differences in N from $\widehat{N}$ tend to correspond with altitudinal excursions from a near-exponential water vapor pressure as expected for a constant c_2 in Eq. (B). Table 2 verifies this by measuring the frequency with which different $N - \widehat{N}$ clusters agree with specific e clusters; their centroids are shown in Figs. 5(a) and 5(f), respectively. For example, Cluster 1 for $N - \widehat{N}$ is the most flat and occurs most frequently, correlates most strongly, to Clusters 3 and 7 for e, the latter of which correspond to profiles with little to no moisture. Conversely, Cluster 6 for $N - \widehat{N}$ correlates well with the highest-moisture profiles in Clusters 1 and 5 for e and contains almost none of the low or no moisture profiles (Clusters 3, 7, 2, and 8 for e).

The $N - \widehat{N}$ centroids in Fig. 5(a) tend to largely deviate from non-zero in the negative direction—and particularly for $N - \widehat{N}$ clusters which correlate with higher moisture, e.g., $N - \widehat{N}$ Cluster 6—which suggests that $N < \widehat{N}$ within a profile correlates with the presence of moisture. This is because a higher relative humidity generally induces a higher refractivity (Friehe et al., 1975; Takamura et al., 1984, also see Eq. (1)), and hence, since $\widehat{N}$ is fit to regions of a profile both with and without moisture, the background measured refractivity N (i.e., in regions without moisture) would be below $\widehat{N}$.

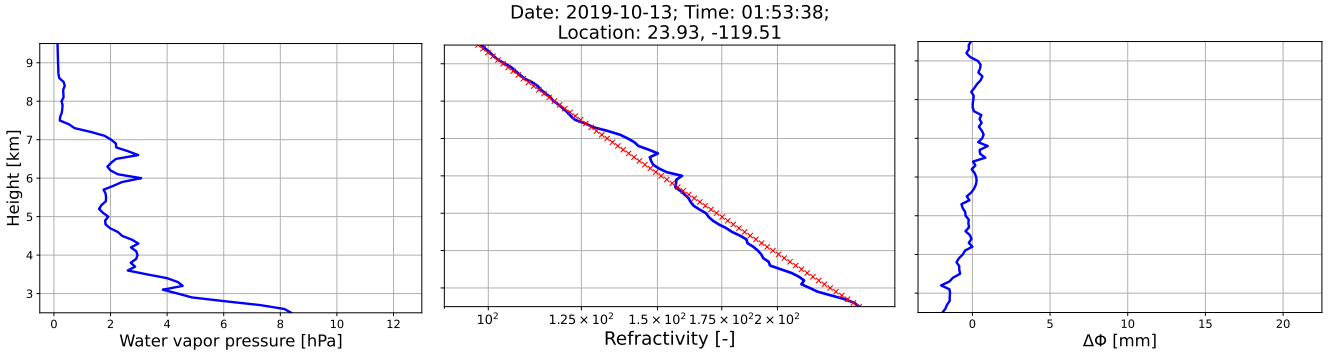
On the other hand, as shown in Table 2, Cluster 3 for $N - \widehat{N}$ features larger values of $\left|N - \widehat{N}\right|$ than Cluster 6 for $N - \widehat{N}$ yet does not correlate with profiles that have a higher water vapor pressure (i.e., Clusters 1 and 5 for e). The examples in Fig. 4 also demonstrate this; in particular, Fig. 4(c) features a profile with a notably higher value of e than the one in Fig. 4(b) yet exhibits smaller values of $\left|N - \widehat{N}\right|$ overall. This suggests that the actual *magnitude* of deviations of N from $\widehat{N}$ does not necessarily correspond with the magnitude of water vapor pressure. Nonetheless, the clustering indicates a weak inverse relationship between the $N - \widehat{N}$ and e—the upper-left and bottom-right corners of Table 2 consist mostly of red values while the bottom-left and upper-right corners consist mostly of green ones.

The aforementioned observation raises two possible hypotheses for why the relationship between the magnitudes of $N - \widehat{N}$ and e are not more direct. Firstly, it is possible that the relationship between e and $N - \widehat{N}$ is between the derivatives of one or both. Furthermore, $\widehat{N}$ is fit across most of the troposphere down to 2.5 km. Hence, $\widehat{N}$ is most effectively sensitive to

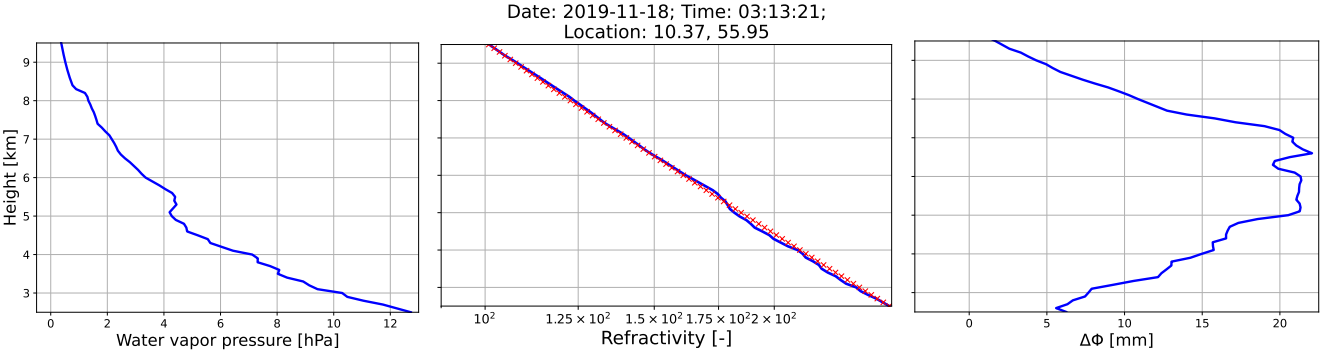
https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) Low moisture and no apparent precipitation



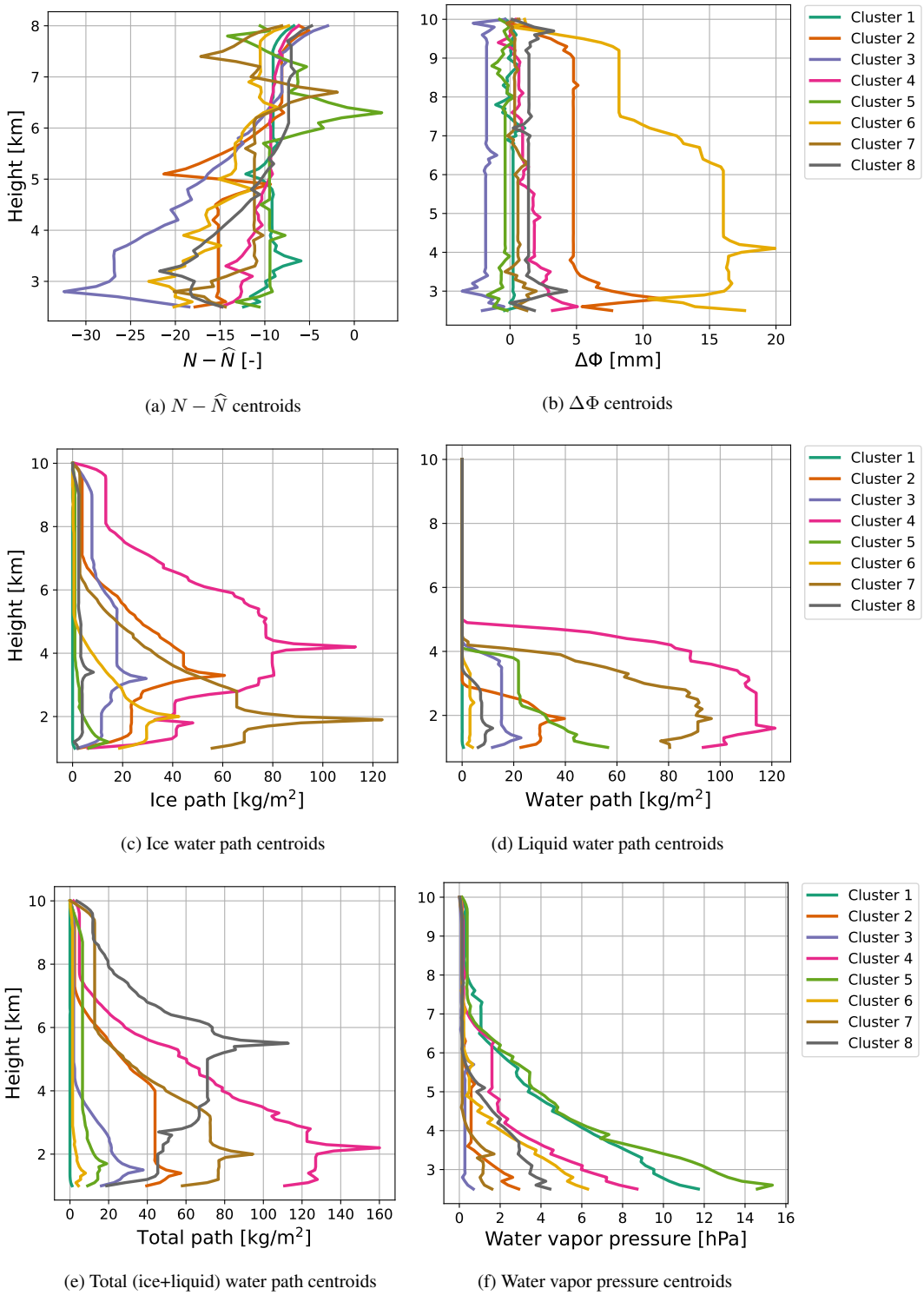
(b) Some moisture but no apparent precipitation



(c) High moisture and precipitation

Figure 4. Three examples of thermodynamic profiles at various times and locations with different moisture and precipitation contents. For each, we show the height on the y-axes and the following on the x-axes: e (left); N in blue and $\widehat{N}$ in red (center); and $\Delta\Phi$ (right).

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) $N - \widehat{N}$ centroids

(b) $\Delta\Phi$ centroids

(c) Ice water path centroids

(d) Liquid water path centroids

(e) Total (ice+liquid) water path centroids

(f) Water vapor pressure centroids

Figure 5. Cluster analysis centroids computed by applying time series k-means clustering across all variables of interest (excluding the $\widehat{N}$ coefficients $\mathbf{c}$).

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Table 2. Percent of profiles in each e cluster (column) for each $N - \widehat{N}$ cluster (row). Cluster numbers are ordered from smallest (most negative/zero) to largest (most positive) value by comparing their corresponding centroids in Fig. 5. Bold corresponds to percents higher than 1.5 times the STD weighted by the percentage of $N - \widehat{N}$ corresponding to each case (6.66%, 5.34%, 7.06%, etc.) greater or less than the mean percentage in the given row. Since there were profiles which were excluded from our cluster analyses for certain variables, the weighted averages for each column or row will not always add up as expected from the law of total probability. The percents in the topmost row and leftmost column percents were computed out of the total number of profiles that passed the clustering requirements for the clustering variable shown. Green and red correspond to the maximum and minimum percents for each row, respectively.

$N - \widehat{N}$: most negative … most positive

e: driest … wettest

$N-\widehat{N}$: →		6.66%	5.34%	7.06%	16.97%	3.75%	38.69%	18.12%	3.41%
e: ↓		3	2	6	8	7	4	1	5
21.35%	3	0.00%	0.29%	0.00%	5.09%	0.00%	30.86%	41.19%	0.90%
20.53%	7	**0.46%**	**1.73%**	**0.66%**	14.45%	2.88%	27.63%	31.66%	23.98%
19.63%	2	8.56%	10.12%	**0.66%**	**35.45%**	8.23%	23.05%	14.81%	23.53%
12.87%	8	14.35%	**24.57%**	**2.18%**	21.91%	12.76%	11.12%	7.15%	**28.51%**
10.89%	6	**30.09%**	**32.66%**	7.42%	15.91%	**29.22%**	4.90%	3.91%	14.93%
7.27%	4	**30.56%**	20.52%	19.43%	6.09%	**24.28%**	1.59%	0.60%	7.69%
6.16%	1	14.35%	8.96%	**55.46%**	0.82%	18.52%	0.36%	0.09%	0.00%
1.29%	5	1.16%	0.87%	**14.19%**	0.09%	3.70%	0.00%	0.00%	0.45%

concentrated moisture anomalies within narrow bands of the troposphere. The sensitivity to the derivatives with respect to height implies that there could be cases where a profile is moist yet the model $\widehat{N}$ is still a close fit for the observed N, e.g., when a moist-unsaturated adiabatic lapse rate (Emanuel, 1994) holds across most of the profile. For these cases, $N - \widehat{N}$ could be close to zero even when the water vapor pressure is non-negligible. As an example, the centroid for $N - \widehat{N}$ Cluster 7 is relatively flat (Fig. 5(a)), but Table 2 shows that e Clusters 4 and 6, both moderately high moisture cases (Fig. 5(f)), are the most commonly represented e clusters in $N - \widehat{N}$ Cluster 7.

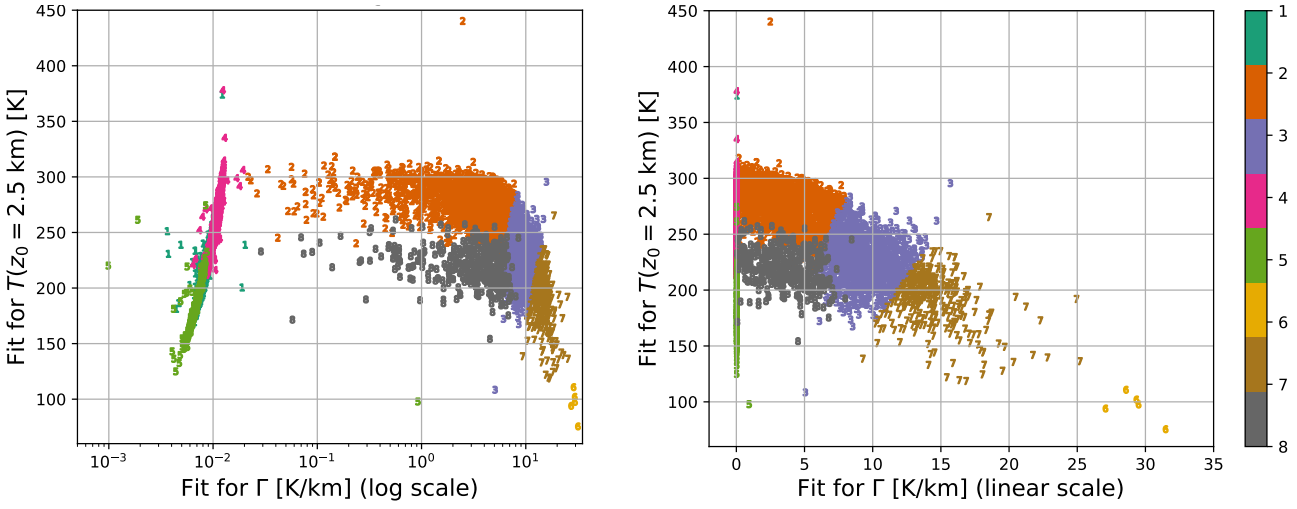
Fig. 4 shows three examples where $N - \widehat{N}$ does not correlate strongly with $\Delta\Phi$. Instead, these cases demonstrate the ability of the deviation from potential refractivity $N - \widehat{N}$ to predict moisture distributions. For instance, the profile in Fig. 4(b) exhibits no precipitation but the water vapor profile would indicate clouds from around 7.5 km down to somewhere near 5.5 km (see e.g. method in Peng et al., 2006).

4.3 $\widehat{N}$ model coefficients and cluster groups

The $\widehat{N}$ coefficients $\mathbf{c} = (c_0, c_1, c_2)$ tend to only exhibit two degrees of freedom across the profiles in the dataset. Fig. 6 shows how projecting the $\mathbf{c}$ clusters onto the $\left(\widehat{\Gamma}, \widehat{T}_0\right)$-plane leads to a clear partitioning across different $\mathbf{c}$ clusters. This suggests that the dominant clusters for $\widehat{e}$ (and therefore e) in the dataset are related to changes in $\widehat{\Gamma}$ and $\widehat{T}_0$. Note that changes in $\widehat{\Gamma}$ and $\widehat{T}_0$ are related to changes in the polytropic index m and therewith the underlying thermodynamics.

The clustering across $\mathbf{c}$ was generally able to partition the physical and nonphysical fits. Clusters 2, 3, 7, and 8 for $\mathbf{c}$ feature physical values of $\widehat{T}_0$ and $\widehat{\Gamma}$ while the other clusters feature nonphysically extreme values of $\widehat{T}_0$ (mainly Cluster 6), Γ (Clusters

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) All latitudes (log scale)

(b) All latitudes (linear scale)

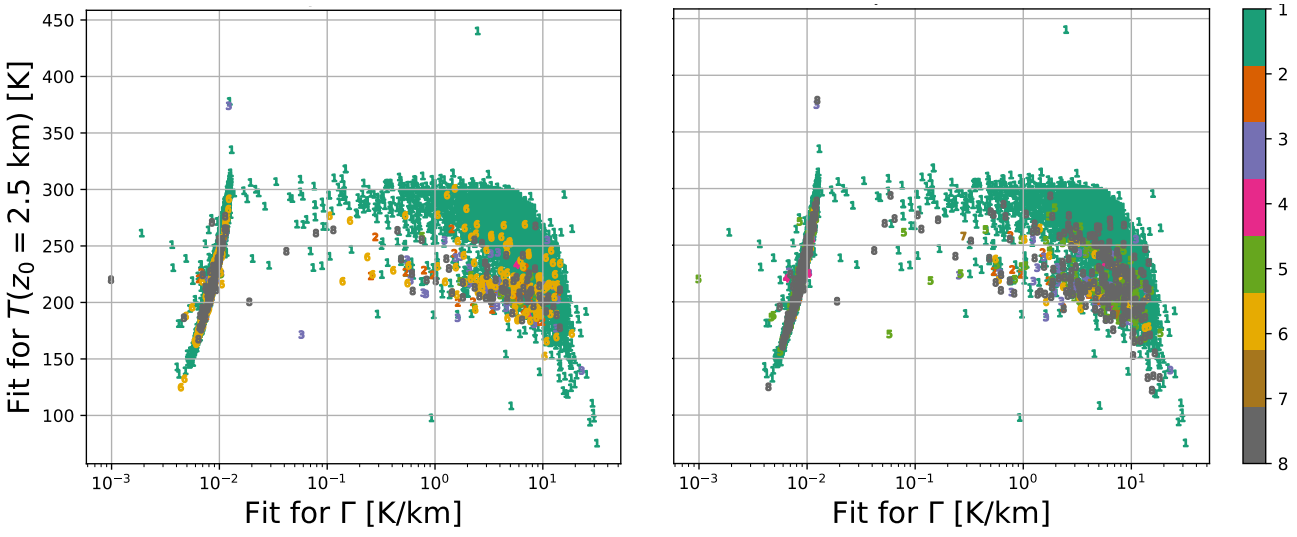
Figure 6. Scatter plots of the best-fit values of $\widehat{T}_0$ vs. $\widehat{\Gamma}$ across all latitudes in the dataset using a) logarithmic scaling and b) linear in $\widehat{\Gamma}$ to make the separation in $\widehat{\Gamma}$ more apparent for small values of $\widehat{\Gamma}$. The colors and symbols correspond to the associated $\widehat{N}$ coefficient (**c**) clusters for each point, as indicated by the colorbars on the right-hand sides.

1 and 4) or both (Cluster 5). Such nonphysical fits indicate where the assumed physics is not reflective of the actual physics in those profiles. Sometimes, we observed that faulty retrievals fell within these clusters with unphysical profiles, suggesting (and perhaps identifying) retrieval issues rather than physical phenomena. Although, since isothermal atmospheres ($\Gamma \approx 0$) are a subset of polytropic atmospheres, Clusters 1 and 4 for **c** could identify some of those as well and thereby still represent physically meaningful coefficients.

Fig. 6 shows a moderately negative linear correlation between $\widehat{T}_0$ and $\widehat{\Gamma}$ for the fits which feature physically realistic values of $\widehat{\Gamma}$. Between $\widehat{T}_0$ and $\widehat{\Gamma}$ across all latitudes for $\widehat{\Gamma} > 0.1$, we have a Pearson correlation coefficient of -0.697, a Spearman rank correlation coefficient of -0.676, and a Kendall rank correlation coefficient of -0.497. Each correlation coefficient has a p-value below machine epsilon (i.e., at least below 2.22×10^{-16}), thereby showing the statistical significance of this negative correlation. This correlation reflects that the moist adiabatic lapse rate has a negative relationship with temperature for profiles with sufficient moisture. Since the moist adiabatic lapse rate approaches the dry adiabatic lapse rate for temperatures roughly below 230 K, a higher lapse rate can be observed for colder profiles.

Fig. 7 shows how **c** relates to the path variables as a function of height. The fit values of $\widehat{\Gamma}$ and $\widehat{T}_0$ generally do not correlate with path clusters. However, when $\widehat{\Gamma} > 10^{-1}$ K/km and $\widehat{T}_0 > 280K$ for a given profile, that profile has little to no precipitation, as shown by the near-uniformity of Cluster 1 (turquoise) for either LWP or IWP in that region, as indicated by Fig. 7(a) and 7(b), respectively. That is, **c** is not too informative in confirming the presence of ice or precipitation, but it can rule out the presence of moisture, and thereby ice and precipitation. Similar yet weaker relationships between **c** and particular precipitation

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.



(a) Liquid water path clusters

(b) Ice water path clusters

Figure 7. LWP and IWP clusters over $\widehat{T}_0$ vs. $\widehat{\Gamma}$ across all profiles and latitudes.

Table 3. Percent of profiles in each e cluster (column) for each **c** cluster (row). e cluster numbers are ordered roughly from smallest to largest value by comparing their corresponding e centroids in Fig. 5(f) while the **c** cluster numbers are merely listed in numerically increasing order (arbitrarily). Bolding and coloring is as in Table 2.

e: driest … wettest

c:	→	21.73%	24.13%	16.19%	19.89%	7.60%	0.07%	4.82%	5.57%
e:	↓	1	2	3	4	5	6	7	8
21.35%	3	3.37%	43.54%	34.65%	11.55%	2.75%	40.00%	34.67%	3.75%
20.53%	7	9.34%	29.13%	29.49%	18.98%	8.06%	**0.00%**	**35.91%**	8.85%
19.63%	2	19.78%	16.57%	15.21%	**27.23%**	22.59%	**0.00%**	15.79%	16.09%
12.87%	8	15.73%	6.12%	9.31%	16.05%	**22.40%**	**40.00%**	6.50%	**21.18%**
10.89%	6	17.17%	2.66%	6.54%	11.18%	**27.11%**	20.00%	1.86%	18.50%
7.27%	4	**16.21%**	1.24%	3.13%	6.30%	12.57%	0.00%	1.55%	11.26%
6.16%	1	15.11%	0.37%	0.46%	6.98%	3.34%	0.00%	0.62%	**18.23%**
1.29%	5	**3.09%**	0.19%	0.65%	1.58%	0.20%	0.00%	0.62%	1.88%

regimes can also be seen across other ranges of $\widehat{\Gamma}$ and $\widehat{T}_0$ in Fig. 7, e.g., $\widehat{\Gamma} > 10$ K/km tends to also correlate with low or no moisture cases.

As the aforementioned relationship between $N - \widehat{N}$ and e suggests, **c** also exhibits an apparent relationship with e. Fig. 6 suggests that Cluster 2 for **c** tends to contain profiles where $\widehat{e}$ is near-zero. This tends to correspond to cases when e is too low for there to be precipitation; as seen in Fig. 3, for low e, there cannot be precipitation or ice formation. The relationship between **c** and e may be analyzed more precisely by looking at Table 3, which demonstrates the predictive power in using **c** clusters to predict representative water vapor pressure profiles, i.e., the centroids for e clusters shown in Fig. 5(f).

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Table 4. Percent of profiles in each cluster for the column variable listed—liquid water path (LWP), ice water path (IWP), and liquid+ice water path (TWP), respectively— for each $\Delta\Phi$ cluster indicated by the row. Cluster numbers are ordered from smallest (most negative/zero) to largest (most positive) value by comparing their corresponding centroids in Fig. 5. Bolding and coloring is as in Table 2.

$\Delta\Phi$: most negative/zero ... most positive

LWP: driest ... wettest

$\Delta\Phi$: →		5.20%	25.32%	34.62%	23.45%	3.87%	5.74%	1.56%	0.24%
LWP: ↓		3	5	1	7	8	4	2	6
89.14%	1	96.35%	95.70%	95.25%	91.22%	66.07%	**54.52%**	**17.78%**	**7.14%**
5.34%	6	3.32%	3.28%	3.15%	6.05%	11.61%	**15.96%**	6.67%	0.00%
2.85%	8	0.33%	0.75%	1.25%	1.99%	**12.05%**	**15.96%**	7.78%	**28.57%**
1.51%	3	0.00%	0.20%	0.15%	0.44%	8.04%	**11.14%**	**22.22%**	7.14%
0.81%	2	0.00%	0.00%	0.20%	0.22%	2.23%	2.11%	**27.78%**	**28.57%**
0.28%	5	0.00%	0.00%	0.00%	0.07%	0.00%	0.30%	**14.44%**	**14.29%**
0.04%	7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**2.22%**	**7.14%**
0.03%	4	0.00%	0.00%	0.00%	0.00%	**0.00%**	0.00%	**1.11%**	**7.14%**

$\Delta\Phi$: most negative/zero ... most positive

IWP: driest ... wettest

$\Delta\Phi$: →		5.20%	25.32%	34.62%	23.45%	3.87%	5.74%	1.56%	0.24%
IWP: ↓		3	5	1	7	8	4	2	6
85.78%	1	94.68%	94.26%	93.91%	85.03%	63.39%	**29.82%**	**3.33%**	**0.00%**
7.31%	8	2.99%	4.78%	4.80%	9.88%	**18.75%**	**19.58%**	3.33%	7.14%
3.75%	5	1.99%	0.75%	0.85%	4.06%	8.93%	**32.23%**	10.00%	14.29%
1.72%	3	0.33%	0.14%	0.35%	0.88%	7.14%	**13.25%**	**21.11%**	7.14%
0.75%	6	0.00%	0.00%	0.05%	0.07%	0.00%	4.22%	**31.11%**	**14.29%**
0.49%	2	0.00%	0.00%	0.05%	0.00%	1.34%	0.90%	**22.22%**	**28.57%**
0.13%	4	0.00%	0.00%	0.00%	0.07%	0.45%	0.00%	**7.78%**	0.00%
0.07%	7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.11%	**28.57%**

$\Delta\Phi$: most negative/zero ... most positive

TWP: driest ... wettest

$\Delta\Phi$: →		5.20%	25.32%	34.62%	23.45%	3.87%	5.74%	1.56%	0.24%
TWP: ↓		3	5	1	7	8	4	2	6
84.55%	1	93.02%	92.96%	93.26%	84.14%	60.71%	**28.92%**	**3.33%**	**0.00%**
8.46%	6	4.98%	5.81%	5.09%	12.09%	15.63%	**24.40%**	3.33%	0.00%
3.99%	5	1.99%	1.09%	1.15%	2.88%	17.41%	**29.82%**	12.22%	14.29%
1.79%	3	0.00%	0.07%	0.45%	0.66%	5.36%	**14.46%**	**25.56%**	7.14%
0.82%	2	0.00%	0.00%	0.05%	0.15%	0.45%	2.41%	**35.56%**	**42.86%**
0.22%	7	0.00%	0.00%	0.00%	0.00%	0.45%	0.00%	**12.22%**	**14.29%**
0.09%	8	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	**4.44%**	**7.14%**
0.07%	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**3.33%**	**14.29%**

4.4 $\Delta\Phi$ and both liquid & ice water path

Table 4 supports the correlation of PRO $\Delta\Phi$ with precipitation in a given profile. Clusters with large $\Delta\Phi$ tend to correlate with those of large LWP or IWP, and inversely, those with small $\Delta\Phi$ also relate to profiles with little to no LWP and IWP.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

This should already be expected, as prior studies (e.g., Cardellach et al., 2019; Wang et al., 2022; Padullés et al., 2023) already indicate relationships between $\Delta\Phi$ and both water precipitation and ice.

Despite how Clusters 2 and 6 for $\Delta\Phi$ feature large values of $\Delta\Phi$ (> 4 mm) quite deep into the atmosphere—up to around 9 km according to their respective centroids—Table 4 shows that ice precipitation is not necessarily deep for those cases. In particular, Clusters 2 and 6 for $\Delta\Phi$ both correlate well with Clusters 2 and 6 for IWP, but the centroids for the latter two drop to zero near 7 and 5 km, respectively. This could be because $\Delta\Phi$ across different heights need not correspond one-to-one with the LWP nor the IWP at those heights, and also because LWP and IWP do not necessarily signal precipitation right at the time they are measured.

Even though the height of a particular onset or peak in $\Delta\Phi$ might not correlate with onsets or peaks, respectively, in the path cluster centroids, the shapes of the $\Delta\Phi$ and total path cluster centroids appear to correlate in both precipitating and non-precipitation cases, as demonstrated in Table 4. This consistency in shape but not in height is a property of the DTW measure used for the clustering. Hence, the lack of height correlations in our clusters does not contradict the model predictions of Padullés et al. (2023) since their model directly matches features in $\Delta\Phi$ and precipitation as a function of height.

5 Conclusion

In summary, k-means clustering has been used to identify different types of correlations between the vertical distributions of precipitation- and moisture-related variables. Our work shows the application and physical interpretability of using a potential refractivity fit, $\widehat{N}$, when there is a linear temperature profile with height, as expected in a polytropic atmosphere. Deviations from $\widehat{N}$ relate to the presence of water vapor pressure anomalies at given latitudes, longitudes, and times (Section 4.2). In particular, Table 2 demonstrates a visibly strong yet non-monotonic relationship between the shapes and amplitudes of $N - \widehat{N}$ vs. e. For instance, the moderately negative Cluster 6 for $N - \widehat{N}$ corresponds well with very moist profiles, yet the more negative Clusters $N - \widehat{N}$ correspond to only moderately moist profiles. Inversely, the mostly flat Cluster 1 for $N - \widehat{N}$ corresponds to profiles with little to no moisture (Clusters 3 and 7 for e) yet the most positive Cluster 5 for $N - \widehat{N}$ corresponds to profiles with low to moderate moisture. This can be explained by how the deviation of N from $\widehat{N}$ will be muted if $\widehat{N}$ has been fit to a profile which is moist overall, and thereby $\left|N - \widehat{N}\right|$ will be largest when the moisture is large and relatively localized (e.g., in the presence of clouds).

$\widehat{N}$ coefficient (**c**) clusters can flag physical vs. nonphysical values of observed and derived variables (Section 4.3, Figure 6). As shown in Figure 6, Clusters 5-7 for **c** generally correspond to temperature values which are far too low, indicating either a problem with the data from the retrievals or a profile which does not satisfy the physical assumptions made in deriving $\widehat{N}$ (see Section 3.1). Inversely, the values of **c** for a given profile can identify when a profile has no moisture or precipitation with very high accuracy—as shown in Figure 7, profiles with $\widehat{\Gamma} > 10^{-1}$ K/km and $\widehat{T}_0 > 280K$ have little to no precipitation. Related correlations between different **c** and e clusters are also shown in Table 3, where we see that different clusters for **c** correspond to profiles with low, medium, and high water vapor pressure throughout.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Similarly, vertical distributions of $\Delta\Phi$ are found to correlate to specific vertical profiles of liquid and ice precipitation. In particular, the amplitude and shape of $\Delta\Phi$ centroids correlate with the amplitudes and shapes of LWP and IWP centroids, respectively (Section 4.4, Table 4). This correlation persists across low and high levels of LWP, IWP, and both combined, thereby demonstrating a strong one-to-one relationship between $\Delta\Phi$ and water path.

In conclusion, the clustering centroids (i.e., "representative" profiles) correlate with the general magnitude of a variable for a given profile and also the general shape of that variable as a function of height. The latter is especially evident for variables which correlate with water content: $\Delta\Phi$ and the path variables. As a demonstration of how the centroids capture the magnitude of profiles in their associated clusters, consider the ice water path (IWP) clusters shown in Figure 4(c): Clusters 4 and 7 for IWP both correspond to higher-than-average ice content in their respective profiles, and a similar comparison can be drawn between Clusters 2 and 6 for IWP. Relatedly, as a demonstration of how the centroids capture the shape, consider the liquid water path (LWP) clusters shown in Figure 4(d): Clusters 2 and 5 for LWP both correspond to non-negligible water precipitation, but Cluster 5 features profiles with deeper precipitation than those in Cluster 2. Thus, clustering in the manner introduced in this study confirms its value as a tool for quality control of profiles and automate the classification of—and condense information on—physical phenomena found across large datasets, thereby avoiding the need to inspect and compare profiles individually.

Appendix A: Derivation of $\widehat{N}$

Combining the equation for hydrostatic equilibrium and the ideal gas law, we have

$$p(z) = p(z_0)\exp\left(-\frac{g}{R}\int_{z_0}^{z}\frac{ds}{T(s)}\right) \tag{A1}$$

where $g = 9.8$ g/m^2 is the acceleration due to gravity on Earth and $R = 287$ J·kg^{-1}·K^{-1} is the specific gas constant for dry air. In a polytropic atmosphere, $T(z) = T(z_0) - \Gamma(z - z_0)$ for a lapse rate Γ to be determined by a fit to the data together with $T(z_0)$. The integral in Eq. (A1) for this temperature profile can be computed as

$$-\frac{g}{R}\int_{z_0}^{z}\frac{ds}{T(s)} = \frac{g}{R\Gamma}\log\left[1 - \frac{\Gamma(z - z_0)}{T(z_0)}\right]$$

which in turn implies that (e.g., Dutton, 1976)

$$p(z) = p(z_0)\left[1 - \frac{\Gamma}{T(z_0)}(z - z_0)\right]^{\frac{g}{R\Gamma}} \tag{A2}$$

Substituting Eq. (A2) and $T(z) = T(z_0) - \Gamma(z - z_0)$ into Eq. (1) and putting a hat on N since $\widehat{N}$ is the idealized model, we have

$$[l]\widehat{N}(z) = \frac{k_1 p(z_0)}{T(z_0)\left[1 - \frac{\Gamma(z-z_0)}{T(z_0)}\right]^2} \times \left\{\left[1 - \frac{\Gamma(z-z_0)}{T(z_0)}\right]^{\frac{g}{R\Gamma}+1} + \frac{k_2 e}{k_1 p(z_0) T(z_0)}\right\}. \tag{A3}$$

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

While $p(z_0)$ might not be available directly in a typical PRO profile (which only contains refractivity and $\Delta\Phi$), there will be data for $N(z_0)$, and hence, we solve for $p(z_0)$ in terms of $T(z_0)$ and $N(z_0)$ to constrain the number of fitting parameters. Rewriting Eq. (1) at $z = z_0$, we have

$$k_1 p(z_0) = T(z_0)\left[N(z_0) - \frac{k_2 e}{T(z_0)^2}\right] \tag{A4}$$

Substituting Eq. (A4) into Eq. (A3) for a constant, representative, e, and rewriting in terms of the fit coefficients c_0, c_1, and c_2, leads to Eq. (3).

Appendix B: Numerical fitting procedure for $\widehat{N}$

Once Eq. (3) has been fit to a given profile, we can use c_0 to solve for $\widehat{\Gamma}$, then use this and c_1 to solve for $\widehat{T}_0$, and finally, use c_2 and $\widehat{T}_0$ to solve for the representative value of $\widehat{e}$. To do this fitting routine in practice, since $k_2, \widehat{e}, N \geq 0$, we impose the constraint $c_2 \geq 0$ and use the curve-fitting utility `optimize.least_squares` from the SciPy package (version 1.7.3) in Python 3.7.4 with the initial conditions $c_0 = 4.5$, $c_1 = 0.01$ m^{-1}, and $c_2 = 0$ to fit Eq. (3) to each N profile for all cases. For reasons which are generally internal to the default `optimize.least_squares` algorithm, the nonlinear fitting procedure either did not always converge within the preset maximum number of iterations, 10000, with prescribed error tolerances `ftol=xtol=` 10^{-12}, or the profile in question was missing too much data for the model coefficients to be uniquely determinable—this only occurred in 5 profiles out of the 6706 in the dataset, or 0.07%. The latter could have either occurred because there were not enough data overall or because there were no refractivity data at $z_0 = 2.5$ km.

Appendix C: Clustering algorithm for k-means

For the numerical implementation of time series k-means clustering, we use version 0.6.2 of the Python package `tslearn`, which provides machine learning tools for the analysis of time series data and builds on the `scikit-learn`, `scipy`, and `numpy` libraries (Tavenard et al., 2020). To run time series k-means clustering for all variables in the dataset, we use `tslearn.clustering.TimeSeriesKMeans` with $k = 8$ clusters, the DTW metric, a maximum of 30 iterations of the algorithm, and we fix the random state to 0 to ensure that the cluster labels stay consistent upon each run.

There are ways to estimate the most "statistically meaningful" number of clusters for a given dataset, even when not using the Euclidean metric—e.g., the average silhouette method (Rousseeuw, 1987) or the gap statistic method (Tibshirani et al., 2001)—which could give different numbers of clusters for each variable. However, to keep a consistent number of clusters for each variable, and to give some semblance of the same hierarchy in magnitude across clustering in each variable, this study uses the same number of clusters for all variables and defers to using a number which is possibly too large rather than too small.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Appendix D: Quality-control criteria for clustering

Profiles are excluded from each cluster according to the quality-control criteria listed below.

- $\Delta\Phi$ (923 profiles excluded, or 13.76%): Files are excluded by the same criteria used for the vertical integral of $\Delta\Phi$.
- c_0, c_1, and c_2 (5 profiles excluded, or 0.07%): The fit for $\widehat{N}$ must converge, i.e., the algorithm for computing the best-fit coefficients c_0, c_1, and c_2 must converge, which means that there must be refractivity data at 2.5 km, there must be enough refractivity data between 2.5 km and the estimated lapse-rate tropopause (the latter of which was explained earlier), and the fit must converge within 10000 iterations for tolerance conditions `ftol=xtol=` 10^{-12}.
- $N - \widehat{N}$ (223 profiles excluded, or 3.33%): Along with the same criteria related to $\widehat{N}$ used for the coefficient clusters, cases where the tropopause is below 8.2 km are skipped and three files from clustering for $N - \widehat{N}$ are taken out manually and excluded. These three files contained unphysically large values of N ($N > 600$) and likely indicate an issue with retrieving the refractivity for the RO dataset.
- Water vapor pressure (33 profiles excluded, or 0.49%): Files are excluded by the same criteria used for the total column water vapor. It should be noted that the three files with unphysically large values of N that were manually excluded from clustering for $N - \widehat{N}$ also had unrealistically large water vapor pressure values (> 300 hPa), and hence, they were also excluded from clustering for water vapor pressure.
- LWP (1 profile excluded, or 0.01%): Files without LWP data from 1 to 10 km are excluded.
- IWP (6 profiles excluded, or 0.09%): Files without IWP data from 1 to 10 km are excluded.
- TWP (6 profiles excluded, or 0.09%): Files without LWP or IWP data from 1 to 10 km are excluded.

Code and data availability. The datasets associated with this study have been uploaded to the Jet Propulsion Laboratory's GENESIS (Global Environmental & Earth Science Information System) site: https://genesis.jpl.nasa.gov/ftp/paz_pol/.

Author contribution. Conceptualization: JK, MTJ, KNW; Data curation: JT, KNW, RP; Formal analysis: JK, MTJ, TK; Funding acquisition: MTJ; Investigation: All; Methodology: JK, MTJ, TK, KNW; Project administration: MTJ; Resources: MTJ, KNW; Software: All; Supervision: MTJ, TK, JT; Validation: All; Visualisation: JK, MTJ, TK; Writing – original draft preparation: JK, MTJ, TK; Writing – review & editing: JK, MTJ, TK, JT

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under the JPL Visiting Student Research Program with support from NASA's NH19ZDA001N-GNSS program, under a contract with the

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

National Aeronautics and Space Administration (80NM0018D0004), and with a stipend and teaching fellowship from the Yale Graduate School of Arts and Sciences. The authors would like to thank Joe Turk for collecting and preparing the GPM dataset, Kuo-Nung Wang and Ramon Padullés for preparing and managing the ROHP-PAZ dataset, and various technical support staff at the Jet Propulsion Laboratory for their tireless help with data, equipment, and account access. We would also like to thank Chi O. Ao for helping to manage and prepare the data and software resources used in this study.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

References

Ao, C. O., Waliser, D. E., Chan, S. K., Li, J.-L., Tian, B., Xie, F., and Mannucci, A. J.: Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017598, 2012.

Atkinson, K. E.: An Introduction to Numerical Analysis, Wiley, New York, 1988.

Bean, B. and Dutton, E.: Radio Meteorology, no. 92 in (National Bureau of Standards), U.S.Govt.Print.Off., https://api.semanticscholar.org/CorpusID:124549052, 1966.

Biondi, R., Randel, W. J., Ho, S.-P., Neubert, T., and Syndergaard, S.: Thermal structure of intense convective clouds derived from GPS radio occultations, Atmospheric Chemistry and Physics, 12, 5309–5318, https://doi.org/10.5194/acp-12-5309-2012, 2012.

Bretherton, C. S., Peters, M. E., and Back, L. E.: Relationships between Water Vapor Path and Precipitation over the Tropical Oceans, Journal of Climate, 17, 1517–1528, https://doi.org/10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2, 2004.

Cardellach, E., Oliveras, S., Rius, A., Tomás, S., Ao, C. O., Franklin, G. W., Iijima, B. A., Kuang, D., Meehan, T. K., Padullés, R., de la Torre Juárez, M., Turk, F. J., Hunt, D. C., Schreiner, W. S., Sokolovskiy, S. V., Hove, T. V., Weiss, J. P., Yoon, Y., Zeng, Z., Clapp, J., Xia-Serafino, W., and Cerezo, F.: Sensing Heavy Precipitation With GNSS Polarimetric Radio Occultations, Geophysical Research Letters, 46, 1024–1031, https://doi.org/10.1029/2018gl080412, 2019.

de la Torre Juárez, M., Padullés, R., Turk, F. J., and Cardellach, E.: Signatures of Heavy Precipitation on the Thermodynamics of Clouds Seen From Satellite: Changes Observed in Temperature Lapse Rates and Missed by Weather Analyses, Journal of Geophysical Research: Atmospheres, 123, 13,033–13,045, https://doi.org/10.1029/2017JD028170, 2018.

Dutton, J.: The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion, McGraw-Hill, https://books.google.com/books?id=9CxRAAAAMAAJ, 1976.

Emanuel, K.: Atmospheric Convection, Oxford University Press, https://books.google.com/books?id=VdaBBHEGAcMC, 1994.

Emmenegger, T., Kuo, Y.-H., Xie, S., Zhang, C., Tao, C., and Neelin, J. D.: Evaluating Tropical Precipitation Relations in CMIP6 Models with ARM Data, Journal of Climate, 35, 6343 – 6360, https://doi.org/10.1175/JCLI-D-21-0386.1, 2022.

Friehe, C. A., Rue, J. C. L., Champagne, F. H., Gibson, C. H., and Dreyer, G. F.: Effects of temperature and humidity fluctuations on the optical refractive index in the marine boundary layer, Journal of the Optical Society of America, 65, 1502–1511, https://doi.org/10.1364/JOSA.65.001502, 1975.

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W.: Tropical tropopause layer, Reviews of Geophysics, 47, https://doi.org/10.1029/2008RG000267, 2009.

Gleisner, H., Ringer, M. A., and Healy, S. B.: Monitoring global climate change using GNSS radio occultation, npj Climate and Atmospheric Science, 5, https://doi.org/10.1038/s41612-022-00229-7, 2022.

Holloway, C. E. and Neelin, J. D.: Temporal Relations of Column Water Vapor and Tropical Precipitation, Journal of the Atmospheric Sciences, 67, 1091 – 1105, https://doi.org/10.1175/2009JAS3284.1, 2010.

Hotta, D., Lonitz, K., and Healy, S.: Forward operator for polarimetric radio occultation measurements, Atmospheric Measurement Techniques, https://doi.org/10.5194/amt-2023-132, 2023.

Huang, Y., Leroy, S. S., and Anderson, J. G.: Determining Longwave Forcing and Feedback Using Infrared Spectra and GNSS Radio Occultation, Journal of Climate, 23, 6027–6035, https://doi.org/10.1175/2010jcli3588.1, 2010.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Izakian, H., Pedrycz, W., and Jamal, I.: Fuzzy clustering of time series data using dynamic time warping distance, Engineering Applications of Artificial Intelligence, 39, 235–244, https://doi.org/10.1016/j.engappai.2014.12.015, 2015.

Kim, J. and Son, S.-W.: Tropical Cold-Point Tropopause: Climatology, Seasonal Cycle, and Intraseasonal Variability Derived from COSMIC GPS Radio Occultation Measurements, Journal of Climate, 25, 5343–5360, https://doi.org/10.1175/jcli-d-11-00554.1, 2012.

Kliore, A., Cain, D. L., Fjeldbo, G., Seidel, B. L., and Rasool, S. I.: Preliminary Results on the Atmospheres of Io and Jupiter from the Pioneer 10 S-Band Occultation Experiment, Science, 183, 323–324, https://doi.org/10.1126/science.183.4122.323, 1974.

Kuo, Y.-H., Sokolovskiy, S., Anthes, R., and Vandenberghe, F.: Assimilation of GPS Radio Occultation Data for Numerical Weather Prediction, Terrestrial, Atmospheric and Oceanic Sciences, 11, 157–, https://doi.org/10.3319/TAO.2000.11.1.157(COSMIC), 2001.

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R.: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, Journal of Geophysical Research: Atmospheres, 102, 23 429–23 465, https://doi.org/10.1029/97JD01569, 1997.

Lackner, B. C., Steiner, A. K., Hegerl, G. C., and Kirchengast, G.: Atmospheric Climate Change Detection by Radio Occultation Data Using a Fingerprinting Method, Journal of Climate, 24, 5275–5291, https://doi.org/10.1175/2011jcli3966.1, 2011.

Mascio, J., Leroy, S. S., d'Entremont, R. P., Connor, T., and Kursinski, E. R.: Using Radio Occultation to Detect Clouds in the Middle and Upper Troposphere, Journal of Atmospheric and Oceanic Technology, 38, 1847–1858, https://doi.org/10.1175/JTECH-D-21-0022.1, 2021.

Muller, C. J., Back, L. E., O'Gorman, P. A., and Emanuel, K. A.: A model for the relationship between tropical precipitation and column water vapor, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL039667, 2009.

Müller, M.: Dynamic Time Warping, in: Information Retrieval for Music and Motion, pp. 69–84, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-74048-3_4, 2007.

Nishida, M., Shimizu, A., Tsuda, T., Rocken, C., and Ware, R. H.: Seasonal and Longitudinal Variations in the Tropical Tropopause Observed with the GPS Occultation Technique (GPS/MET), Journal of the Meteorological Society of Japan. Ser. II, 78, 691–700, https://doi.org/10.2151/jmsj1965.78.6_691, 2000.

Padullés, R., Cardellach, E., and Turk, F. J.: On the global relationship between polarimetric radio occultation differential phase shift and ice water content, Atmospheric Chemistry and Physics, 23, 2199–2214, https://doi.org/10.5194/acp-23-2199-2023, 2023.

Peng, G., de la Torre-Juárez, M., Farley, R., and Wessel, J.: Impacts of upper tropospheric clouds on GPS radio refractivity, in: 2006 IEEE Aerospace Conference, pp. 6 pp.–, https://doi.org/10.1109/AERO.2006.1655899, 2006.

Randel, W. J., Wu, F., and Ríos, W. R.: Thermal variability of the tropical tropopause region derived from GPS/MET observations, J. Geophys. Res., 108, 4024, https://doi.org/10.1029/2002JD002595, 2003.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Scherllin-Pirscher, B., Steiner, A. K., Anthes, R. A., Alexander, M. J., Alexander, S. P., Biondi, R., Birner, T., Kim, J., Randel, W. J., Son, S.-W., Tsuda, T., and Zeng, Z.: Tropical Temperature Variability in the UTLS: New Insights from GPS Radio Occultation Observations, Journal of Climate, 34, 2813–2838, https://doi.org/10.1175/jcli-d-20-0385.1, 2021.

Schmidt, T., Wickert, J., Beyerle, G., and Reigber, C.: Tropical tropopause parameters derived from GPS radio occultation measurements with CHAMP, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2004jd004566, 2004.

Smith, E. K. and Weintraub, S.: The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies, Proceedings of the IRE, 41, 1035–1037, https://doi.org/10.1109/JRPROC.1953.274297, 1953.

https://doi.org/10.5194/egusphere-2024-1278
Preprint. Discussion started: 12 August 2024
© Author(s) 2024. CC BY 4.0 License.

Son, S.-W., Lim, Y., Yoo, C., Hendon, H. H., and Kim, J.: Stratospheric Control of the Madden–Julian Oscillation, Journal of Climate, 30, 1909–1922, https://doi.org/10.1175/jcli-d-16-0620.1, 2017.

Takamura, T., Tanaka, M., and Nakajima, T.: Effects of Atmospheric Humidity on the Refractive Index and the Size Distribution of Aerosols as Estimated from Light Scattering Measurements, Journal of the Meteorological Society of Japan. Ser. II, 62, 573–582, https://doi.org/10.2151/jmsj1965.62.3_573, 1984.

Tavenard, R.: An introduction to Dynamic Time Warping, https://rtavenar.github.io/blog/dtw.html, 2021.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., and Woods, E.: Tslearn, A Machine Learning Toolkit for Time Series Data, Journal of Machine Learning Research, 21, 1–6, http://jmlr.org/papers/v21/20-091.html, 2020.

Tibshirani, R., Walther, G., and Hastie, T.: Estimating the Number of Clusters in a Data Set Via the Gap Statistic, Journal of the Royal Statistical Society Series B: Statistical Methodology, 63, 411–423, https://doi.org/10.1111/1467-9868.00293, 2001.

Tomás, S., Padullés, R., and Cardellach, E.: Separability of Systematic Effects in Polarimetric GNSS Radio Occultations for Precipitation Sensing, IEEE Transactions on Geoscience and Remote Sensing, 56, 4633–4649, https://doi.org/10.1109/tgrs.2018.2831600, 2018.

Turk, F. J., Haddad, Z. S., Kirstetter, P.-E., You, Y., and Ringerud, S. E.: An observationally based method for stratifying a priori passive microwave observations in a Bayesian-based precipitation retrieval framework, Quart. J. Roy. Meteor. Soc., 144, 145–164, https://doi.org/10.1002/qj.3203, 2018.

Turk, F. J., Padullés, R., Ao, C. O., Juárez, M. d. l. T., Wang, K.-N., Franklin, G. W., Lowe, S. T., Hristova-Veleva, S. M., Fetzer, E. J., Cardellach, E., Kuo, Y.-H., and Neelin, J. D.: Benefits of a Closely-Spaced Satellite Constellation of Atmospheric Polarimetric Radio Occultation Measurements, Remote Sensing, 11, https://doi.org/10.3390/rs11202399, 2019.

Turk, F. J., Padullés, R., Cardellach, E., Ao, C. O., Wang, K.-N., Morabito, D. D., de la Torre Juárez, M., Oyola, M., Hristova-Veleva, S., and Neelin, J. D.: Interpretation of the Precipitation Structure Contained in Polarimetric Radio Occultation Profiles Using Passive Microwave Satellite Observations, Journal of Atmospheric and Oceanic Technology, 38, 1727 – 1745, https://doi.org/10.1175/JTECH-D-21-0044.1, 2021.

Utsumi, N., Turk, F. J., Haddad, Z. S., Kirstetter, P.-E., and Kim, H.: Evaluation of precipitation vertical profiles estimated by GPM-era satellite-based passive microwave retrievals, J. Hydrometeor., 22, 95–112, https://doi.org/10.1175/JHM-D-20-0160.1, 2020.

von Engeln, A., Teixeira, J., Wickert, J., and Buehler, S. A.: Using CHAMP radio occultation data to determine the top altitude of the Planetary Boundary Layer, Geophysical Research Letters, 32, https://doi.org/10.1029/2004GL022168, 2005.

Wang, K.-N., Ao, C. O., Padullés, R., Turk, F. J., de la Torre Juárez, M., and Cardellach, E.: The Effects of Heavy Precipitation on Polarimetric Radio Occultation (PRO) Bending Angle Observations, Journal of Atmospheric and Oceanic Technology, 39, 149–161, https://doi.org/10.1175/jtech-d-21-0032.1, 2022.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., and Tian, D.: The Generic Mapping Tools Version 6, Geochemistry, Geophysics, Geosystems, 20, 5556–5564, https://doi.org/10.1029/2019gc008515, 2019.

Xian, T. and Fu, Y.: Characteristics of tropopause-penetrating convection determined by TRMM and COSMIC GPS radio occultation measurements, Journal of Geophysical Research: Atmospheres, 120, 7006–7024, https://doi.org/10.1002/2014jd022633, 2015.