https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

A Deep Convolutional Neural Network for Retrieving Tropospheric Temperature and Moisture Profiles from Refractivity over Tropical Oceans: Framework Development and Characterization

Santosh Muralidharan*[1,2], Harris John[3], Deepak Mishra[4]

[1] Space Physics Laboratory, Vikram Sarabhai Space Centre, Indian Space Research Organisation, Thiruvananthapuram, 695022, Kerala, India

[2] Department of Physics, Faculty of Science, University of Kerala, Thiruvananthapuram, Kerala, India

[3] Vikram Sarabhai Space Centre, Indian Space Research Organisation, Thiruvananthapuram, 695022, Kerala, India

[4] Department of Avionics, Indian Institute of Space Science and Technology, Thiruvananthapuram 695547, Kerala, India https://orcid.org/0000-0001-6532-942X

Correspondence to: **Santosh Muralidharan,** *email:* santosh.spl.isro@gmail.com, *ORCID:* https://orcid.org/0000-0002-9131-0420

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Abstract

Retrieving profiles of temperature and water vapor from atmospheric refractivity over tropical oceans constitutes an inherently underdetermined problem. Conventional one-dimensional variational methods resolve this through numerical weather prediction (NWP) priors, potentially propagating model biases into retrieved profiles and limiting the independence desirable for climate monitoring applications. We present a deep learning retrieval that substitutes learned statistical constraints for model-dependent priors. A convolutional neural network, trained on approximately 20,800 high resolution radiosonde profiles —combining reference-grade GCOS Reference Upper-Air Network (GRUAN) measurements with quality-controlled operational GCOS Upper Air Network (GUAN) and field campaign data— predicts dry refractivity and partial pressure of dry air as intermediate targets; temperature, water vapor pressure, and relative humidity are derived analytically. The model achieves water vapor pressure root-mean-squared errors of ~0.5 hPa near the surface (decreasing with height) and relative humidity errors below 6 % from 100 m to 10 km, performance comparable in magnitude to the mean measurement uncertainty of state-of-the-art radiosondes. Temperature errors are ~1.5 K near the surface, improving to ~0.5 K in the 10–15 km range where dry refractivity dominates. Evaluation against more than 27,400 geographically independent radiosonde ascents across the tropical Pacific, Indian Ocean, and Atlantic—including observations from the 1992–93 TOGA-COARE field campaign, two decades predating the training period—demonstrates robust generalization within the tropical marine atmosphere. The framework accepts only refractivity vertical structure as input, with no dependence on geographic coordinates or NWP background states. This paper establishes the retrieval framework and characterizes performance under ideal input conditions; Part 2 addresses application to satellite observations.

Keywords

Water vapor retrieval, Temperature retrieval, Deep learning, Inception Net, GNSS RO

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

1 Introduction

Atmospheric water vapor is the dominant greenhouse gas and the principal amplifier of climate sensitivity, governing the magnitude of the global temperature response through positive feedback on longwave absorption, cloud formation, and the hydrological cycle (Held and Soden, 2000). Over the tropical oceans — where sea surface temperatures are highest and the thermodynamic coupling between ocean and atmosphere most intense — water vapor drives the large-scale atmospheric meridional and zonal overturning circulations, controls the vertical distribution of latent heat release, and amplifies cloud-radiative feedbacks that constitute the leading source of uncertainty in equilibrium climate sensitivity (Sherwood et al., 2020). Accurate vertical profiles of temperature and water vapor over these regions are therefore indispensable — for numerical weather prediction, for the detection and attribution of climate change, for validating the physical parameterizations, and for constraining the cloud feedbacks on which climate projections depend (Santer et al., 2008; Eyre et al., 2022; Allan et al., 2022). Yet the observational network that provides such profiles in situ — the global radiosonde network — is by its nature confined to discrete land and island locations, operationally resource-intensive, and essentially absent over the vast expanses of the open tropical ocean. The scientific and societal value of closing this observational gap through satellite-based atmospheric profiling is therefore high, and the methodological challenge of doing so at the vertical resolution and accuracy that physical applications demand remains open.

The operational landscape for satellite thermodynamic profiling relies on hyperspectral infrared and passive microwave sounders aboard polar-orbiting platforms. Prominent systems include the Infrared Atmospheric Sounding Interferometer (IASI) on MetOp, the Atmospheric Infrared Sounder (AIRS) on Aqua, and the Cross-track Infrared Sounder (CrIS) on Suomi-NPP, each paired with complementary microwave sounding units for simultaneous temperature and humidity retrievals (Nalli et al., 2013,2018; Trent et al., 2019, 2023). These instruments retrieve thermodynamic profiles through radiative transfer inversion of measured spectral radiances, an approach that has yielded scientifically valuable long-term datasets. Nevertheless, several

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

fundamental constraints limit their utility for the problem at hand. Retrievals are layer-averaged rather than continuous, with typical vertical resolution of 2 to 5 km — insufficient to resolve the sharp moisture gradients of the marine atmospheric boundary layer and the lower free troposphere where tropical water vapor is most abundant. Accuracy deteriorates in cloudy conditions, which prevail over much of the tropical ocean (Nalli et al., 2018). Fixed sun-synchronous satellite orbit geometries restrict temporal sampling to one or two overpasses per day per location, precluding capture of the full diurnal variability of tropical convection. These limitations are most consequential precisely at the altitudes and in the meteorological conditions that govern the tropical water vapor budget.

Global Navigation Satellite system (GNSS) Radio Occultation (RO) provides a physically distinct observational capability that directly addresses several of these shortcomings. By measuring the bending and excess phase delay of GNSS signals as they traverse the atmospheric limb, RO satellite constellations such as COSMIC/FORMOSAT-3 (Anthes, 2011; Ho et al., 2020a) and COSMIC-2/FORMOSAT-7 (Ho et al., 2020b; Schreiner et al., 2020) retrieve vertical profiles of atmospheric refractivity with vertical resolution of the order of few hundred metres, global geographic coverage, all-weather capability, and sampling at all local times — characteristics that passive sounders cannot simultaneously achieve. The self-calibrating nature of GNSS-RO measurements confers the long-term stability required for climate monitoring applications (Gleisner et al., 2022; Steiner et al., 2020; Anthes 2011; Ho et al., 2020a). The scale of this observational resource is already substantial: COSMIC-2 alone delivers thousands of refractivity profiles daily over tropical oceanic regions (Schreiner et al., 2020). These augment measurements from earlier missions including COSMIC-1 (National Oceanic and Atmospheric Administration and Taiwan's National Space Organization) (Anthes et al., 2008), Metop-A/B/C (European Organisation for the Exploitation of Meteorological Satellites and European Space Agency) (Luntama et al., 2008), and Fengyun-3 (China Meteorological Administration) (Bai et al., 2018), as well as Megha-Tropiques (Indian Space Research Organisation and French National Centre for Space Studies) (Karouche et al., 2012). Additionally, these observations are now complemented by small satellites and CubeSats deployed in commercial constellations comprising tens of satellites each—

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

collectively providing tens of thousands of vertical refractivity profiles per day globally—including Spire Global (Mo et al., 2024; Ho et al., 2023), GeoOptics (Chang et al., 2022), PlanetIQ (Zhran et al., 2024), and Tianmu-1 (Li et al., 2025). Yet all these satellite constellations only produce observational data where temperature and water vapor are fused into a single observable—atmospheric refractivity (N)—in the neutral atmosphere, expressed as:

$$N = 77.6890\frac{P_d}{T} + 71.2952\frac{P_w}{T} + 375463\frac{P_w}{T^2} \quad (1)$$

where P_d and P_w are the partial pressures of dry air and water vapor (hPa) respectively and T is absolute temperature (K) (Reuger, 2002). N is directly measurable by GNSS-RO; however, this single equation with three thermodynamic unknowns constitutes an inherently underdetermined system that admits no direct analytical solution.

The established operational response to this underdetermined problem is the one-dimensional variational (1D-Var) method, which resolves the system by incorporating a priori thermodynamic state estimates from numerical weather prediction (NWP) analyses or reanalysis products — principally NCEP GFS and ECMWF — as additional constraints on the inversion (Healy and Eyre, 2000; Von Engeln et al., 2003; Jagadheesha et al., 2009; Wee et al., 2022). The approach has been operationally refined over two decades and produces retrievals with quantifiable uncertainty metrics and physical internal consistency. One of its characteristic is that in altitude ranges where the observational constraint from refractivity on thermodynamic state is weakest — notably the moisture-laden lower troposphere — the retrieved solution necessarily draws more heavily on the NWP background and its associated error statistics. The extent to which this dependency affects the retrieved profiles depends on how accurately the background error covariances are specified and how well observation errors are characterized.

Wee et al. (2022) demonstrate that with carefully specified error covariances and a two-step variational approach, the UCAR 1D-Var system achieves remarkable background independence and long-term stability even in challenging moist conditions, producing retrievals that converge toward consistent solutions regardless of the

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

background source — a significant methodological advance for the GNSS-RO retrieval community. It is worth noting, however, that background independence and absolute accuracy address distinct aspects of retrieval quality. Convergence toward a consistent solution ensures that 1D-Var retrievals are not merely reproducing the prior state, yet a converged solution may still reflect the statistically optimal compromise between observation and prior rather than the true atmospheric condition — particularly in altitude regimes where refractivity's sensitivity to individual thermodynamic variables is intrinsically limited (e.g., where temperature and humidity effects on N partially compensate). Similarly, long-term stability does not by itself exclude the possibility of time-invariant structural biases shared between the retrieval and its background sources. These are not deficiencies unique to any particular 1D-Var implementation; they are inherent properties of any retrieval framework that incorporates model-derived prior information, however skillfully.

For applications where fully independent observational products are desirable — such as long-term climate monitoring, detection and attribution of trends, and independent validation of NWP systems — retrievals derived without any NWP input would offer a complementary capability that exploits the self-calibrating characteristics of GNSS-RO measurements most directly. This motivation — pursuing observational independence as a complement to established variational methods, rather than as a correction of their known limitations — drives the exploration of data-driven approach in the present work.

Deep learning has been proposed as a data-driven alternative for resolving the underdetermined refractivity retrieval problem, capable in principle of learning the non-linear mapping from refractivity to thermodynamic state without recourse to explicit NWP priors. Initial demonstrations established the concept in tropical and Arctic environments (Bonafoni et al., 2009, Bonafoni 2010; Pelliccia et al., 2011; Shyam et al., 2013), and subsequent work has extended these efforts to regional GNSS-RO applications (Hooda et al., 2023) and global scales (Lasota, 2021). A fundamental methodological limitation, however, is shared across these studies: training and evaluation are conducted against reanalysis products — primarily ECMWF ERA5 — as the definition of atmospheric truth. Models trained against reanalysis inherit its resolution constraints, its

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

systematic biases, and its model-state dependencies, meaning that the retrieval output is not independent of NWP in any meaningful sense. The use of the same reanalysis database for both training and testing additionally precludes independent validation. The scientific case that deep learning adds value over 1D-Var therefore remains unestablished, because both approaches share the same fundamental dependency on model background states.

This study proceeds from a different foundation. In the hierarchy of atmospheric observations, co-located and coincident in situ radiosonde measurements — subject only to their own measurement uncertainties — represent the most accurate available description of the true thermodynamic state of the atmosphere at any given location and time, superior to any NWP analysis or reanalysis product at the same spatial footprint. We develop and train a convolutional neural network (CNN) on approximately 20,800 high-resolution radiosonde ascents —anchored by reference-grade measurements from the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN)— providing traceable, uncertainty-characterized reference measurements—and selected quality-controlled stations from the GCOS Upper Air Network (GUAN) and dedicated tropical field campaigns, restricted to island locations within the tropical Pacific, Indian Ocean, and Atlantic Ocean regions to minimize land-contamination effects. The model is evaluated against more than 27,000 geographically independent radiosonde ascents from GUAN island stations located across the tropical Pacific, Indian Ocean, and Atlantic Ocean regions, supplemented by open-ocean ship-based observations from the Dynamics of Madden–Julian Oscillation (DYNAMO) 2011–12 field campaign over the Indian Ocean and from the Tropical Ocean Global Atmosphere – Coupled Ocean–Atmosphere Response Experiment (TOGA-COARE) 1992–93 campaign over the western Pacific — none of which contributed to model training. This evaluation dataset — the largest assembled for independent in situ assessment of any deep learning-based refractivity retrieval to date — encompasses observations from recent multi-year records extending back to the TOGA-COARE field campaign of 1992–93, representing a wide range of tropical atmospheric conditions, and observing periods. Yet even this dataset is composed of discrete point-location observations from island stations and ship tracks; the atmosphere over the vast majority of the open tropical ocean remains unobserved by radiosondes at

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

any given time, while GNSS-RO missions deliver profiles numbering in the tens of thousands daily over these same regions and growing (e.g., Schreiner et al., 2020). This asymmetry—dense satellite coverage versus sparse in situ stations—defines the central challenge: whether a model trained on geographically constrained radiosonde data can learn statistical thermodynamic relationships sufficiently general to perform across the broader tropical ocean domain.The CNN provides a data-driven approximation to the inherently underdetermined refractivity retrieval problem, substituting statistical constraints derived from measured atmospheric structures for the NWP background constraints employed in 1D-Var methods.

The CNN also implicitly employs a prior—the empirical joint distribution of (N, T, P_w) learned from training data—rather than eliminating prior constraints, which remain mathematically necessary for this underdetermined system. The distinction from 1D-Var lies in provenance: empirical-radiosonde versus NWP-model derived. The approach seeks to approximate a plausible solution through learned covariability between refractivity and thermodynamic state, linking the precision of ground-based measurements to the spatial reach of satellite refractivity profiling forming a bridge between the two bypassing use of NWP models.

The solution rests on two independently testable hypotheses. The first concerns accuracy: we hypothesize that a deep learning (DL) model trained on quality-controlled in situ data can retrieve thermodynamic profiles from refractivity with RMSE approaching radiosonde measurement uncertainty in altitude regimes where the retrieval is well-constrained, recognizing that the fundamental underdetermination precludes uniformly optimal performance across all variables and altitudes. The second hypothesis concerns the nature of what the model has learned: we hypothesize that the mapping captured by the DL model generalizes to atmospheric conditions over the open tropical ocean, despite training on spatially discrete point locations, and that this generalization reflects robust statistical associations between refractivity and thermodynamic structure rather than station-specific or instrument-specific artifacts. This hypothesis is contingent on the spatial stationarity of vertical thermodynamic state-refractivity relationships across the tropical marine troposphere.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Radiosonde observations, however carefully assembled, are confined to spatially discrete point locations that sample only a small fraction of the atmosphere over tropical oceanic regions typically at synoptic times. GNSS-RO, by contrast, provides profiles numbering in the thousands to tens of thousands daily over tropical oceanic regions—substantially denser spatial coverage than the global radiosonde network, though with irregular temporal sampling determined by orbital geometry. A retrieval model that does not generalize beyond its training locations has no practical utility, because the locations where it works are exactly those that already have radiosonde observations. The scientific and operational value of this approach materializes only if the learned statistical associations extend to the open oceanic atmosphere well beyond the training locations, and if these associations prove sufficiently robust to serve as practical approximations to the true thermodynamic structure. Evaluating both hypotheses rigorously, through independent in situ datasets drawn from island stations and ship tracks across the tropical Pacific, Indian Ocean, and Atlantic, constitutes the scientific core of this paper. The specific objectives of this study are: (i) to develop a customized CNN architecture for retrieving temperature and water vapor profiles from atmospheric refractivity, using dry refractivity and partial pressure of dry air as physically motivated intermediate prediction targets; (ii) to quantify retrieval accuracy of T, P_w, and relative humidity (RH) against geographically independent, multi-year, multi-location, multi-instrument in situ radiosonde datasets from tropical oceanic regions, with GRUAN radiosonde measurement uncertainty serving as the reference standard for interpreting retrieval skill. Diagnostic analyses of vertical resolution and sensitivity to controlled perturbations additionally assess the physical plausibility of the retrievals.

The Part 2 of this work addresses the application of this validated retrieval framework to GNSS-RO refractivity profiles from the COSMIC-2/FORMOSAT-7 mission.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

1.1 Scope of this study

This paper addresses framework development and characterization rather than operational validation. The experimental design employs radiosonde-derived refractivity as CNN input, with the same radiosonde measurements serving as reference truth. This closed-loop configuration isolates the CNN's learned thermodynamic mapping from confounding factors present in satellite observations — including measurement noise due to atmospheric multi-path, horizontal averaging, super-refraction etc. — enabling rigorous characterization of the retrieval framework itself. We explicitly acknowledge that this configuration does not constitute independent validation; rather, it establishes the accuracy achievable under ideal input conditions and provides the baseline against which satellite-based retrieval performance can be assessed. The evaluation dataset, comprising more than 27,000 radiosonde ascents from geographically independent stations across three ocean basins, tests the generalization of learned thermodynamic relationships beyond training locations — the central scientific question this framework must answer before satellite application is warranted.

2 Database and methodology

2.1 Theoretical framework of the deep learning model

The CNN receives as inputs, only vertical profiles of refractivity (N) and its wavelet covariance transform (WCT_N); scalar geographic or temporal variables are deliberately excluded. This constraint forces the model to learn generalizable physical relationships between refractivity vertical structure and thermodynamic state rather than memorizing location-specific climatologies, directly testing the hypothesis 2. Refractivity alone conflates temperature and moisture effects into a single observable; the wavelet covariance transform with dilation 150 m, partially disambiguates these contributions by detecting and amplifying sharp vertical transitions—inversions, moisture boundaries, and boundary-layer tops—whose characteristic scales are physically diagnostic of thermodynamic partitioning. Explicit provision of WCT_N as a distinct input channel is essential for retrieval fidelity, as the transform amplifies localized gradient information that the CNN architecture would otherwise extract inefficiently from raw profiles alone,

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

enabling accurate prediction without recourse to external geographic or temporal covariates (see Supplementary Section S6 for effect of WCT_N on model predictions). This architecture extends to the retrieval step the background-independence that characterizes the GNSS-RO measurement itself, ensuring that performance on independent open-ocean evaluation data (e.g., TOGA-COARE, DYNAMO) reflects generalization rather than interpolation from known geographic regimes.

Altitude information (100 m – 20 km above mean sea level, encompassing the tropical troposphere and lower stratosphere) is embedded at 10 m vertical resolution and standardized across all profiles owing to CNN architectural constraints (see Sect. 2.3 & Sect. 2.4, Sect.S1). The model outputs vertical profiles of dry refractivity (N_d) and partial pressure of dry air (P_d; referred to as dry pressure hereafter for simplicity). Refractivity, defined in Eq. (1), decomposes into dry and wet components:

$$N = N_d + N_w \tag{2}$$

where,

$$N_d = 77.6890 \frac{P_d}{T} \tag{3}$$

$$N_w = 71.2952 \frac{P_w}{T} + 375463 \frac{P_w}{T^2} \tag{4}$$

Total pressure comprises dry and wet contributions:

$$P = P_d + P_w \tag{5}$$

The linear dependence of dry variables N_d and P_d on N (Eq. 1) renders these quantities straightforward targets for CNN prediction. Conversely, temperature (T) and water vapor pressure (P_w) involve nonlinear, coupled terms (1/T and $1/T^2$), increasing model complexity and risking overfitting or underfitting for finite training datasets. To streamline the architecture, reduce computational burden, and enhance accuracy, N_d and P_d are designated as prediction targets, with P_w and T derived subsequently.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Radiosonde measurements of pressure (P), temperature (T), and relative humidity (RH) serve to compute reference profiles of N, N_d and P_d via Eqs. (1–5) for supervised training and validation. Saturated water vapor pressure ($P_{w\,sat}$) is calculated using the Hyland and Wexler formulation (Hyland and Wexler, 1983); P_w is then obtained from:

$$RH = \frac{P_w}{P_{w\,sat}} \times 100 \tag{6}$$

Following meteorological convention and GRUAN data processing standards (Dirksen et al., 2014), RH herein (and throughout the study) denotes relative humidity with respect to liquid water, applied even at temperatures substantially below 0 °C.

The wavelet covariance transform of N (WCT_N; Gamage and Hagelberg, 1993) identifies sharp transitions using the Haar wavelet:

$$h\left(\frac{z-b}{a}\right) = \begin{cases} 1: b - \frac{a}{2} \le z \le b \\ -1: \; b \le z \le b + \frac{a}{2} \\ 0: \quad elsewhere \end{cases} \tag{7}$$

where z denotes altitude, a is the dilation parameter, and b locates the wavelet centre. The localized convolution W_f is defined as:

$$W_f(a,b) = \int_{z_b}^{z_t} f(z) h\left(\frac{z-b}{h}\right) dz \tag{8}$$

with $f(z)$ representing the signal (N), and and Z_b and Z_t denoting profile boundaries (10 m and 20 km, respectively). A dilation a of 150 m is selected to resolve fine-scale vertical structures. Explicit inclusion of WCT_N is central to retrieval fidelity (see supplementary material Sect. S6).

Vertical profiles of N_d and P_d constitute the CNN model output; at no stage of inference are T, P_w, or RH provided to the model. Retrieved temperature is computed from N_d and P_d via Eq. (3); wet refractivity (N_w) follows from Eq. (2). These retrieved T and N_w

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

values yield P_w through Eq. (4), with $P_{w\,sat}$ calculated from T using the Hyland and Wexler formulation (Hyland and Wexler, 1983). Relative humidity is then determined via Eq. (6). The resulting retrieved profiles of T, P_w, and RH—derived from CNN-predicted N_d and P_d—are validated against independent radiosonde measurements. As reference and retrieved profiles originate from identical radiosonde ascents, the comparison represents perfectly collocated, coincident observations.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

2.2 Radiosonde data

Reference-grade and rigorously quality-controlled data are essential for training the neural network to prevent measurement biases and errors from propagating into model predictions. The radiosonde training database in this study draws from three complementary sources: certified data products from the GCOS Reference Upper-Air Network (GRUAN; details in Table 1), high-resolution observations from the GCOS Upper Air Network (GUAN), and quality-controlled data from the DYNAMO 2011–12 tropical field campaign. Details of all training stations are provided in Table 2.

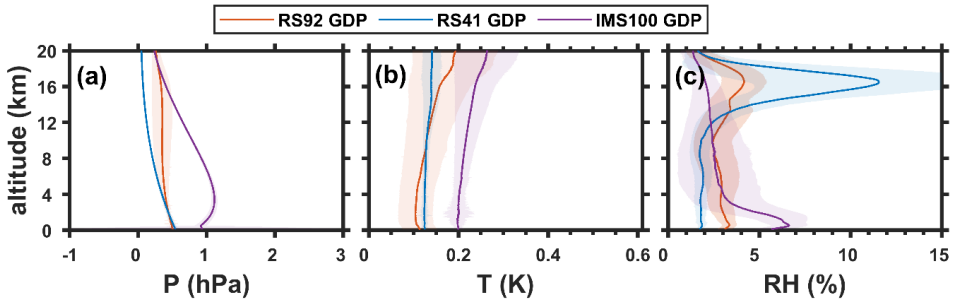
Table 1. The GCOS Radiosonde Upper-Air Network (GRUAN) Data Products (GDP) used in the study.

Certified GRUAN Data Product & Radiosonde Make	DOI	References
RS92-GDP.2 Vaisala RS92	https://doi.org/10.5676/GRUAN/RS92-GDP.2	Dirksen et al., 2014 Immler et al., 2010
RS41-GDP.1 Vaisala RS41	https://doi.org/10.5676/GRUAN/RS41-GDP.1	Von Rohden et al., 2022 Immler et al., 2010
iMS-100-GDP.2 Meisei iMS-100	https://doi.org/10.5676/GRUAN/IMS-100-GDP.2	Hoshino et al., 2022 Immler et al., 2010

GRUAN is a network of radiosonde sites established for measuring essential climate variables above the Earth's surface, providing high vertical-resolution atmospheric observations with a well-defined set of guidelines facilitating the quality and traceability required for climate monitoring (Seidel et al., 2009; Immler et al., 2010). GRUAN data products incorporate stringent corrections for all known measurement errors — including solar radiation heating, sensor time-lag, and humidity dry bias — with total measurement uncertainties fully characterized at each altitude for every ascent (Dirksen et al., 2014; Von Rohden et al., 2022; Hoshino et al., 2022). GRUAN data are routinely used to validate satellite missions including GNSS-RO constellations (Ladstädter et al., 2015; Nalli et al., 2018; Shao et al., 2021; Trent et al., 2019, 2023). The GRUAN data products used in the present study are the RS92-GDP.2, RS41-GDP.1, and iMS-100-GDP.2, whose details are provided in Table 1. Four GRUAN stations serve as

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

346 training sites: Singapore (1.34°N, 103.88°E), Manus (2.09°S, 146.87°E), and Nauru
347 (0.52°S, 166.93°E) — three near-equatorial western and central Pacific island stations —
348 and Minamitorishima (24.29°N, 153.98°E), an off-equatorial western Pacific island.



349

350 **Figure 1. Mean measurement uncertainties** for pressure **(a)**, temperature **(b)** and
351 relative humidity **(c)** from GRUAN data products RS92 GDP (3,327 profiles), RS41 GDP
352 (2,934 profiles) and iMS 100 GDP (2,782 profiles) from all data listed in Table 2. All
353 uncertainties are shown at coverage factor k = 1 (standard uncertainty); RS41-GDP
354 values have been divided by 2 from the published k = 2 specifications to enable direct
355 comparison. The shaded areas correspond to one standard deviation of measurement
356 uncertainty variability at each altitude.

357 Figure 1 presents an integrated view of measurement uncertainties for pressure,
358 temperature, and relative humidity from all GRUAN data used in this study (9,043
359 profiles). Following the standard uncertainty representation Y = y ± kU (where k denotes
360 the coverage factor, typically 1–3) (Sommer et al., 2023), all uncertainties are presented
361 at k = 1 (standard uncertainty) in Fig. 1. The values shown — approximately ±1 hPa for
362 pressure, ±0.2 K for temperature, and ±5–6 % for relative humidity, together represent
363 the lowest achievable measurement uncertainty for any multi-year, multi-station in situ
364 radiosonde dataset. Notably, these values indicate expanded uncertainties, with actual
365 magnitudes varying according to the coverage factor selected.

366 The uncertainty profiles reveal altitude-dependent and instrument-specific
367 characteristics. Mean relative humidity uncertainty for RS41 GDP increases sharply
368 above 12 km, reaching approximately 12-15 % at 16 km. At these moisture-depleted

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

altitudes, dry refractivity exceeds wet refractivity by more than 3 orders of magnitude (N_d/N_w > 5000), rendering such relative humidity uncertainties negligible in their contribution to total refractivity. Consequently, deep learning model temperature retrievals in this altitude range remain largely unaffected, as water vapor partial pressure approaches zero. Measurement uncertainty in pressure decreases with altitude, whereas uncertainties for temperature and relative humidity partially increase with height (Fig. 1). Additionally, iMS-100 GDP radiosondes lack a pressure sensor, deriving pressure recursively via the hydrostatic equation to match surface observations from ground-based barometers (Hoshino et al., 2022), potentially increasing uncertainty near the surface and accounting for the elevated pressure uncertainty from the surface to ~5 km in Fig. 1a.

Among all measurement systems for the thermodynamic state of the troposphere and lower stratosphere, radiosondes represent the best standard. The uncertainties presented in Figure 1 therefore constitute the practical limit of achievable measurement uncertainty for any multi-location, multi-year in situ dataset profiling tropical oceanic regions, attributable to stringent quality-control procedures and subsequent corrections implemented in GRUAN processing. Figure 1 thus provides the benchmark against which model retrieval accuracy is assessed in this study.

Three GUAN island stations are included in the training dataset, selected to provide representation in each ocean basin and to extend latitudinal coverage to off-equatorial tropical latitudes and extend longitudinal coverage to the western hemisphere: San Juan, Puerto Rico (18.43°N, 65.99°W) in the tropical Atlantic, Hilo, Hawaii (19.7°N, 155.05°W) in the central Pacific, and Willis Island (16.29°S, 149.00°E) in the Coral Sea of the south-western tropical Pacific. These stations form part of the Global Climate Observing System (GCOS) Upper Air Network (GUAN), which serves the needs of global climate applications under the auspices of the World Meteorological Organization (WMO) and other international bodies through the WMO Integrated Global Observing System (WIGOS) framework; they employ Graw and Vaisala radiosondes in accordance with WMO guidelines for network participation (https://library.wmo.int/idurl/4/48325). Together with the four near-equatorial GRUAN training stations, these off-equatorial GUAN stations broaden the model's

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

exposure to the range of atmospheric conditions encountered across the tropical ocean domain. High vertical resolution data for GUAN stations were obtained from the University of Wyoming Atmospheric Science Radiosonde Archive (https://weather.uwyo.edu/upperair/sounding.shtml). The locations of these stations are presented in Fig. 2.

High vertical resolution radiosonde data from the DYNAMO 2011–12 field campaign (Yoneyama et al., 2013) supplement the training database with two island stations from the Indian Ocean — Gan Island in the Maldives (0.70°S, 73.16°E) and Diego Garcia (7.32°S, 72.42°E) — providing near-equatorial Indian Ocean representation in the training data. Level 3 (L3) island station data from these two sites are used for training; the ship-based radiosonde data from the same campaign are reserved exclusively for model evaluation, as described below. All DYNAMO sites used the Vaisala RS92 SGP radiosonde. Details on the sounding system and quality control procedures are available in Ciesielski et al. (2014). Only data records with a quality flag of 'Good' were considered.

Since the focus of the present work is tropical oceans, only island-based stations are used for training. A small island can be considered a large ship permanently anchored in the ocean, capable of making atmospheric observations representative of the surrounding oceanic atmosphere provided its atmospheric boundary layer is accounted for; the typical mean daytime (nighttime) boundary layer height would be ~850 m (~150 m) (see M.Santosh (2022a) and references therein). The high native vertical resolution of the data (typically ≤10 m; Table 2) ensures that fine-scale vertical variations in temperature, water vapor pressure, relative humidity, and refractivity within the troposphere and lower stratosphere are captured. The mean and standard deviation of the native vertical resolution of raw data for the altitude range 0–20 km are provided in Table 2 for each station.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

425 **Table 2. Radiosonde data used for Training the Deep Learning model.**

Sr. No.	Group	Station / Location / Native Vertical Resolution	Observation Period	Ascents Used	Database Type
1	Near-Equatorial Stations	Singapore (1.34°N, 103.88°E) 5 ± 1.60 m	2016-04-13 to 2021-12-30	2934	RS41-GDP.1 Certified GRUAN Data Product (GDP)
2		Manus (2.09°S, 146.87°E) 3.97 ± 0.85 m	2012-01-01 to 2014-07-06	1899	RS92-GDP.2 Certified GRUAN Data Product (GDP)
3		Nauru (0.52°S, 166.93°E) 4.13 ± 1.08 m	2011-01-01 to 2013-08-26	1428	RS92-GDP.2 Certified GRUAN Data Product (GDP)
4		Gan, Maldives (0.70°S, 73.16°E) 9.07 ± 2.00 m	2011-09-22 to 2012-02-09	799	DYNAMO L3 data [NCAR/EOL]
5		Diego Garcia (7.32°S, 72.42°E) 4 ± 1.20 m	2011-09-30 to 2012-01-15	480	DYNAMO L3 Data [NCAR/EOL]
6	Off-Equatorial Stations	Minamitorishima (24.29°N, 153.98°E) 6.46 ± 0.90 m	2018-08-01 to 2023-12-31	2782	iMS-100-GDP.2 Certified GRUAN Data Product (GDP)
7		San Juan, Puerto Rico (18.43°N, 65.99°W) 5.08 ± 1.31 m WIGOS : 0-20000-0-78526	2018-01-01 to 2025-12-15	2670	GUAN high resolution data
8		Hilo, Hawaii (19.7°N, 155.05°W) 5.18 ± 1.29 m WIGOS: 0-20000-0-91285	2018-04-05 to 2024-01-27	3207	GUAN high resolution data
9		Willis Island, Australia (16.29°S, 149.00°E) 9.80 ± 1.95 m WIGOS: 0-20000-0-94299	2018-01-01 to 2025-12-15	4627	GUAN high resolution data

426 The trained model is evaluated using geographically independent radiosonde
427 datasets withheld entirely from model development. The term 'evaluation data' is used
428 throughout this study to refer to these datasets.

429 Radiosonde data from nine tropical island GUAN stations (Fig.2) are used for
430 evaluation (Table 3), providing multi-year independent assessments across the tropical
431 Pacific, Indian Ocean, and Atlantic Ocean regions. These stations — Guam (13.48°N,
432 144.79°E), Guadeloupe (16.26°N, 61.53°W), Grantley Adams Barbados (13.07°N,
433 59.5°W), Rochambeau French Guiana (4.82°N, 52.37°W), Lihue Kauai Hawaii (21.99°N,
434 159.3°W), PAGO American Samoa (14.3°S, 170.7°W), Atuona French Polynesia (9.8°S,

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

139.03°W), Tahiti (17.56°S, 149.6°W), and Cocos Island (12.1°S, 96°E) — are geographically independent of the training data and distributed across both hemispheres and all three tropical ocean basins. Multi-year records at each station ensure that the full range of seasonal, intra-seasonal, and interannual atmospheric variability at each location is represented in the evaluation. These data with high vertical resolution (typically < 10 m) were obtained from the University of Wyoming Atmospheric Science Radiosonde Archive.

Ship-based and island radiosonde observations from two dedicated tropical field campaigns constitute the second independent evaluation dataset (Table 4). From the DYNAMO 2011–12 campaign, four ship cruises operating over the open tropical Indian Ocean provide 1,001 ascents for evaluation. From TOGA-COARE conducted during 1992–93 over the western Pacific, data from four island sounding stations and six ship-based platforms are used, spanning locations over the western tropical Pacific. All TOGA-COARE stations used the Vaisala RS80 sonde; data quality checks and corrections are presented in Loehrer et al. (1996), with additional corrections applied for the known humidity measurement errors of the RS80 radiosonde (Wang et al., 2002). Only data records with a quality flag of 'Good' were considered. Nearly 50 % of TOGA-COARE profiles were excluded during pre-processing (Sect.2.3), mostly due to low balloon burst altitude. TOGA-COARE station data were collected at 10-second intervals, yielding a native vertical resolution of approximately 40 m — the coarsest of any dataset used in this study. The inclusion of TOGA-COARE data, collected over three decades before the training period and using a different radiosonde generation at coarser native resolution, provides a stringent test of the model's temporal robustness and adaptability across instrument generations strengthening confidence in the model's generalization capability over the tropical atmosphere.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

460 **Table 3. List of high vertical resolution in situ radiosonde data from GUAN used**
461 **for model evaluation.**

Sr. No.	Station Name / WIGOS Station ID	Location (Lat, Lon)	Native Vertical Resolution	Ascents Used	Data Period
1	**Guam, Mariana Islands, USA** 0-20000-0-91212	(13.48°N, 144.79°E)	5.10 ± 1.13 m	4325	1 Jan 2018 – 4 Apr 2024
2	**Guadeloupe Islands, French Territory** 0-20000-0-78897	(16.26°N, 61.53°W)	9.21 ± 1.96 m	2706	1 Jan 2018 – 27 Jan 2024
3	**Grantley Adams, Barbados** 0-52-130-78954	(13.07°N, 59.5°W)	4.83 ± 0.75 m (to Oct 2020) 24.1 ± 4.34 m (Nov 2020 onwards)	2196	14 Nov 2019 – 27 Jan 2024
4	**Rochambeau, French Guiana** 0-20000-0-81405	(4.82°N, 52.37°W)	9.26 ± 2.17 m	3738	2 Jan 2018 – 27 Jan 2024
5	**Lihue, Kauai, Hawaii** 0-20000-0-91165	(21.99°N, 159.3°W)	5.46 ± 1.63 m	2652	1 Jun 2021 – 28 Dec 2025
6	**PAGO, American Samoa** 0-20000-0-91765	(14.3°S, 170.7°W)	5.02 ± 1.08 m	3027	6 Apr 2018 – 15 Dec 2025
7	**Atuona, French Polynesia (Marquesas Islands)** 0-20000-0-91925	(9.8°S, 139.03°W)	10.24 ± 2.16 m	1372	14 Nov 2018 – 15 Dec 2025
8	**Tahiti, French Polynesia** 0-20000-0-91938	(17.56°S, 149.6°W)	9.62 ± 2.09 m	3106	14 Nov 2018 – 27 Jan 2024
9	**Cocos Island** 0-20000-0-96996	(12.1°S, 96°E)	10.21 ± 1.94 m	1397	17 Jan 2018 – 29 Dec 2023

462

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Table 4. Radiosonde data from ship cruises and island/coastal stations used for model evaluation (DYNAMO 2011–12 and TOGA-COARE 1992–93).

Sr. No.		Station / Ship Name	Location	Native Vertical Resolution	Ascents Used	Period of Observation
1	DYNAMO 2011–12	**RV *Baruna Jaya III***	(95°E, 7°S)	< 10 m	54	5–18 Dec 2011
2		**RV *Mirai***	(80.5°E, 0°)	< 10 m	490	25 Sep – 10 Dec 2011
3		**RV *Roger Revelle***	(80.5°E, 8°S)	< 10 m	443	29 Aug 2011 – 10 Feb 2012
4		**RV *Sagar Kanya***	(84°E, 0°)	< 10 m	14	25 Sep – 10 Oct 2011
5	TOGA-COARE 1992–93	**Kapingamarangi**	(154.8°E, 1.07°N)	40 ± 22 m	492	11 Jul 1992 – 29 Jun 1993
6		**Kavieng**	(150.8°E, 2.58°S)	46 ± 14 m	198	29 Oct 1992 – 1 Mar 1993
7		**Manus**	(147.43°E, 2°S)	46 ± 14 m	297	1 Jul 1992 – 30 Jun 1993
8		**Nauru**	(166.93°E, 0.5°S)	40 ± 15 m	233	2 Sep 1992 – 30 Jun 1993
9		**RV *Kexue***	(156°E, 4°S)	45 ± 16 m	55	9 Nov 1992 – 19 Feb 1993
10		**RV *Shiyan 3***	(158°E, 2.2°S)	45 ± 14 m	202	10 Nov 1992 – 18 Feb 1993
11		**RV *Xiangyanhong 5***	(155°E, 2.2°S)	42 ± 23 m	128	13 Nov 1992 – 18 Feb 1993
12		**RV *Moana Wave***	(156°E, 1.7°S)	44 ± 12 m	43	11 Nov 1992 – 13 Feb 1993
13		**RV *Hakuho Maru***	(156°E, 0°)	9 ± 7 m	129	1 Nov – 4 Dec 1992
14		**RV *Vickers***	(156°E, 2.1°S)	32 ± 4 m	103	28 Jan – 27 Feb 1993

The trained model is evaluated using independent radiosonde datasets withheld entirely from model development. Note that the term 'validation data' in this context and in this study, refers to the dataset used during neural network training as part of supervised learning. The *trained* model is further validated using an independent, larger in situ dataset (Tables 3,4) representing tropical oceanic regions, which is referred to as 'evaluation data' for clarity. For all datasets in Tables 2, 3 and 4, ascent counts were determined following pre-processing of the raw data (Section 2.3), and the native vertical resolutions indicated represent means ± standard deviations derived solely from the actual data used in this study. The locations of training and evaluation stations are shown

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

in Fig. 2. Manus and Nauru appear in both the training dataset (Table 2; GRUAN, 2011–2014) and the TOGA-COARE evaluation dataset (Table 4; 1992–93); the observing periods are separated by nearly two decades and use different radiosonde types, so these constitute temporally and instrumentally independent samples from the same locations.

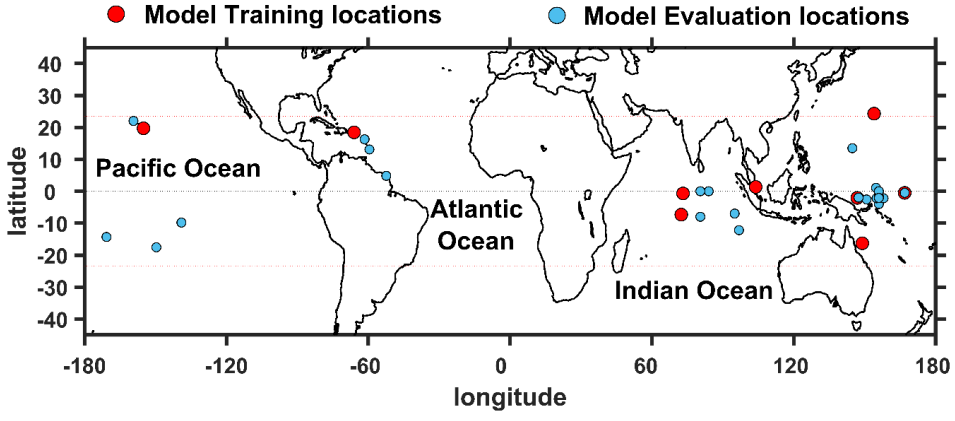


Figure 2. Locations of radiosonde ascent sites used for training and evaluation of the convolutional neural network (CNN) model are shown (details in Tables 2, 3, and 4).

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

2.3 Pre-processing of radiosonde data for model

Radiosonde observations from GRUAN, DYNAMO 2011–12, Wyoming GUAN, and TOGA-COARE 1992–93 were subjected to standardized quality control prior to model training and evaluation:

1. **Descent exclusion**: Descent-phase measurements were rejected. Ascent records were chronologically and vertically ordered, with duplicate altitude entries eliminated.
2. **Surface-layer WCT_N stabilization**: Observations below 100 m were truncated to mitigate edge-artifacts in wavelet covariance transforms. Notably, WCT_N at 100 m retains information from the excluded sub-100 m layer through the convolution operation, preserving input quality while accommodating station elevations lacking 10 m-level data. Valid model retrievals span 100 m to 20 km and so do model outputs.
3. **Anomaly excision**: Profiles exhibiting visually identified spurious excursions in pressure, temperature, or relative humidity were discarded.
4. **Vertical extent threshold**: Ascents failing to reach 20 km maximum altitude were rejected; for qualifying profiles, records above 20 km were excised, constraining both inputs and outputs to this altitude ceiling.
5. **Gap tolerance**: Profiles with cumulative data voids exceeding 1 km within 0–20 km, or surface-proximal gaps >500 m in pressure, temperature, or humidity, were excluded. Residual discontinuities were interpolated to a uniform 10 m grid—marginally coarser than native resolution for all datasets except TOGA-COARE. To accommodate the rapid tropospheric decrease in pressure, linear interpolation was performed on its logarithm, with the result subsequently exponentiated back to the original domain; temperature and relative humidity were interpolated linearly.
6. **Grid standardization**: All observations were resampled to 10 m vertical spacing. This represents sub-sampling for training and evaluation datasets (excluding TOGA-COARE) originally obtained at finer resolution.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

The 20 km altitude ceiling, gap-interpolation scheme and 10 m gridding derive from CNN architectural requirements: fixed-length input and output vectors of 2,000 elements.

7. **RH Sanity checks:** Unrealistically high positive (negative) were set to 100 (0) for training and evaluation data (model inputs and outputs). Such occurrences within the profiles was small (< 0.5 %) but this step ensures that effect of measurement errors was minimized.

8. **GRUAN specific uncertainty screening**: Ascents exceeding uncertainty thresholds—relative humidity >15 %, pressure >2 hPa, or temperature >1 K—for more than 50 levels within 0–20 km were rejected. Applied only to GRUAN profiles with complete uncertainty metadata for all parameters (*P, T, RH*).

Application of these criteria to the training data eliminated approximately 12 % of profiles through failure to meet the altitude threshold, with a further ~7 % excluded by ancillary constraints. Evaluation datasets (Tables 3,4) underwent identical processing. No smoothing was applied at any stage.

2.4 Convolutional Neural Network architecture

The neural network employs an InceptionNet-derived architecture (Szegedy et al., 2015), originally developed for hierarchical feature extraction in visual recognition tasks including image classification, object detection (Das et al., 2020; Dong et al., 2020; Pritt and Chern, 2017; Wang et al., 2019; Xia et al., 2017), and image retrieval (Bui et al., 2018). This framework was adapted for one-dimensional atmospheric profiling, restructuring the input layer to accept tensors of dimension 2 × 2,000 (vertical profiles and their wavelet transforms) rather than conventional 256 × 256 pixel arrays. The network generates vertical profiles of dry refractivity and dry pressure across a standardized 10 m grid spanning 0–20 km altitude. Separate network instances—architecturally identical but independently parameterized—predict N_d and P_d from refractivity inputs (Fig. 3).

Geometric altitude (referenced consistently to mean sea level) is implicit rather than explicit in both inputs and outputs, embedded through the fixed vertical discretization. Operations proceed within a normalized, dimensionally reduced latent space—equivalent to low-dimensional manifold embedding (Tenenbaum et al., 2000; Jia

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

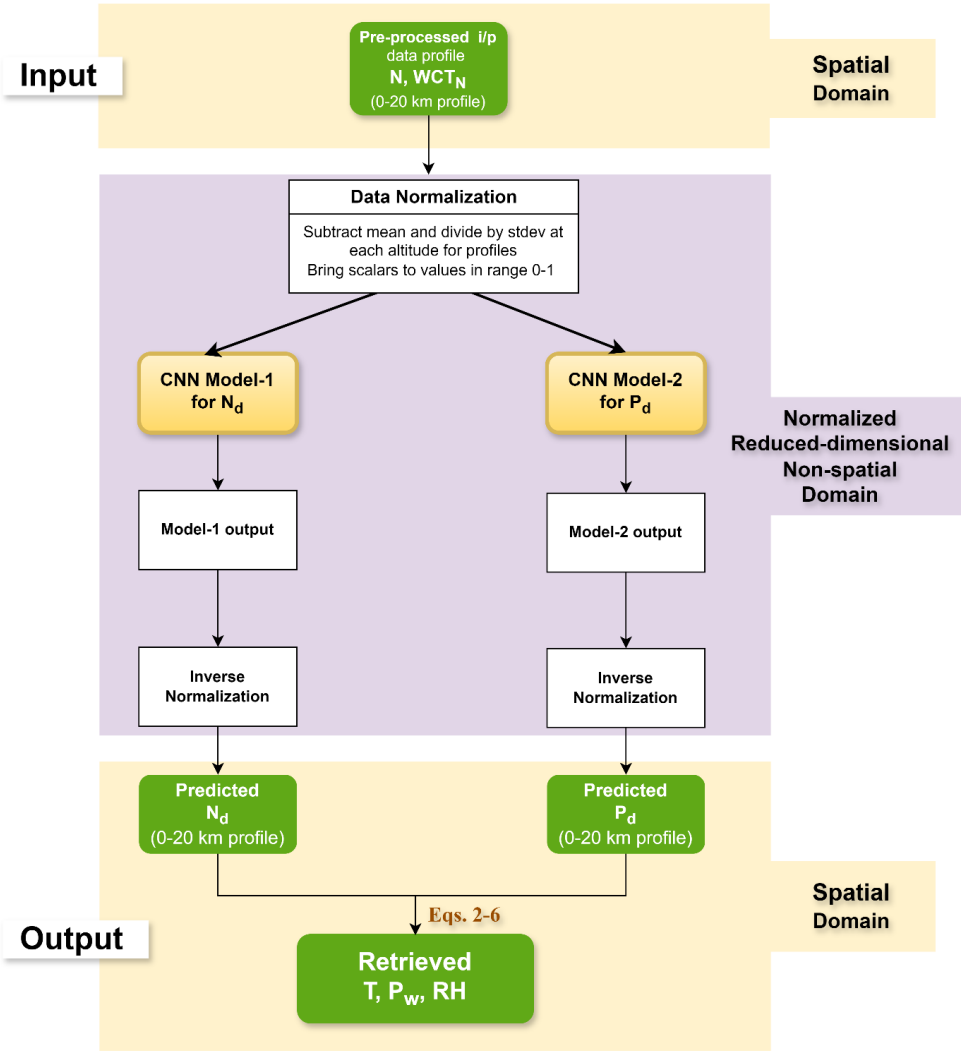
et al., 2022)—with altitude-relative normalization enforcing strict uniformity in vertical extent and sampling density. This architectural constraint necessitates fixed input and output vector cardinalities, as specified in criteria (4)–(6) of Sect. 2.3. Data below 100 m are excluded to avoid large gradients at the lower profile edge, thereby ensuring well-conditioned inputs from the outset; consequently, both model inputs and predictions span 100 m–20 km at 10 m intervals. Outputs are subsequently projected into the physical domain via these predefined altitude levels, maintained consistently across training, validation, evaluation, and operational datasets.

A detailed description of the CNN model architecture, including the composite loss function, training configuration, and all hyperparameter settings, is provided in Supplementary Section S1. The robustness of retrieval accuracy to variations in these hyperparameters — including latent space dimensionality, learning rate, batch size, and loss function weighting — was systematically evaluated across ten configurations using 5-fold cross-validation (Supplementary Section S2). Test-set retrieval errors for T, P_w, and RH proved largely insensitive to these variations (Supplementary Figs. S6–S8), confirming that retrieval accuracy is robust to the specific configuration selected and that marginal adjustments to model complexity or training dynamics do not yield meaningful improvements in retrieval accuracy for this architecture.

Terminological distinctions warrant clarification. "DL model" denotes the complete retrieval framework producing temperature, water vapor pressure, and relative humidity from refractivity. "CNN model" refers specifically to the neural network subset illustrated in Fig. 3 that generates N_d and P_d profiles through the modular components detailed in Sect. S1. The near-direct correspondence between these predicted quantities and the derived thermodynamic variables motivates interchangeable usage of both terms herein.

The 95 % confidence intervals for RMSE and mean bias at each altitude are computed from the sample statistics using standard parametric methods assuming independent profiles.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026



566

567 **Figure 3** Data flow for retrieval of thermodynamic variables *T, P_w* & *RH* from
568 refractivity in the deep learning (DL) model

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

3 Results

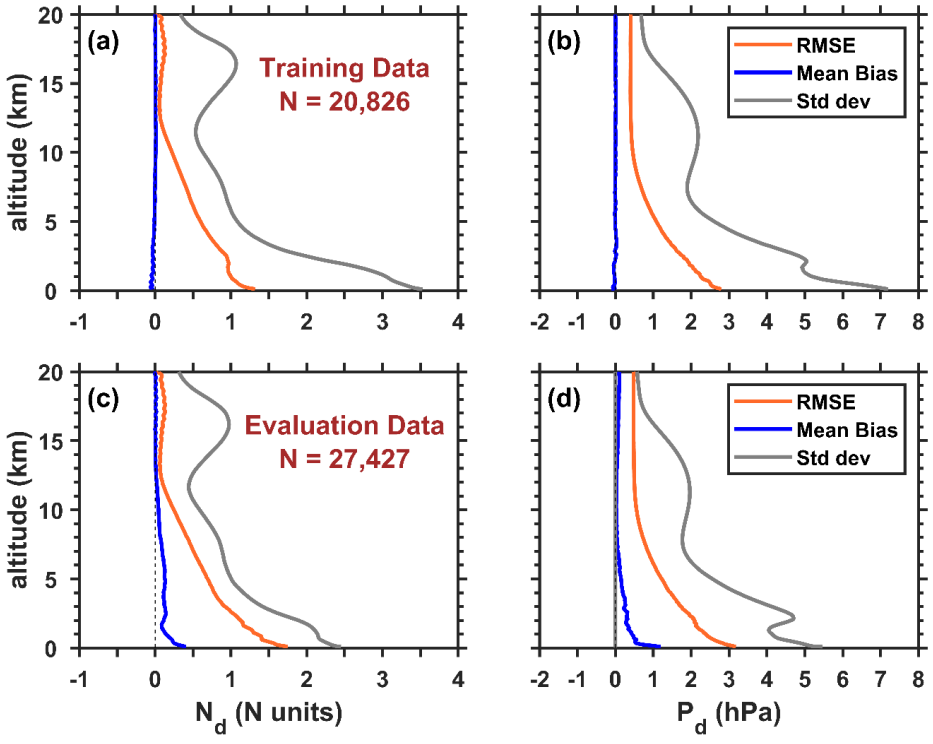
3.1 Retrieval of intermediate parameters: dry refractivity and dry pressure

The deep learning (DL) model predicts vertical profiles of dry refractivity (N_d) and dry pressure (P_d), from which temperature (T), relative humidity (RH), and water vapor partial pressure (P_w) are subsequently derived (Section 2.1). Retrieval accuracy for the thermodynamic parameters therefore depends fundamentally on the fidelity of the predicted N_d and P_d profiles. All results presented in this paper concern in situ radiosonde-based evaluation.

Figure 4 shows root-mean-squared error (RMSE), mean bias, and standard deviation for CNN-predicted N_d and P_d against radiosonde reference profiles, evaluated separately on training (N = 20,826) and independent evaluation (N = 27,427) datasets. The RMSE for training data is marginally lower than for evaluation data, and the mean bias is also slightly better (near-zero) for training data, both consequences of the iterative optimization process during model training. The close agreement in RMSE between the two datasets across all altitudes demonstrates robust generalization: the CNN has learned the underlying relationship between refractivity input and the dry-variable outputs rather than memorizing training samples.

For N_d, RMSE is approximately 1 N-unit near the surface, decreases with altitude, and remains below the standard deviation at all levels. A slight RMSE increase above 15 km coincides with greater natural variability in this region. For P_d, RMSE decreases from ~2 hPa near the surface and stabilizes at a small, non-zero value above 10 km. These trends confirm model accuracy across the full altitude range.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026



591

592 **Figure 4.** Root mean squared error (RMSE), mean bias, and standard deviation (Std dev)
593 for dry refractivity *(N_d)* **(a,c)** and dry pressure *(P_d)* **(b,d)** at 10 m vertical grid size across
594 the 100 m to 20 km altitude range. Standard deviation (grey lines) represents natural
595 parameter variability in the measured profiles, providing context for the magnitude of
596 retrieval errors. RMSE is calculated for CNN-predicted values on training data (**a,b**) and
597 evaluation data (**c,d**) relative to reference profiles derived from radiosonde
598 measurements. The number of radiosonde ascents used in each dataset is indicated.

599 ### 3.2 Temperature and water vapor retrievals from radiosonde-derived
600 refractivity

601 Retrieval skill is assessed against the two hypotheses stated in Section 1: retrieval
602 accuracy benchmarked against GRUAN measurement uncertainty (Fig.1), and
603 generalization evaluated through application to geographically independent data
604 (locations depicted in Fig. 2).

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Because water vapor content above 12 km is negligible ($N_w \ll N$), the moisture signal encoded in refractivity becomes too weak for reliable retrieval at these altitudes; accordingly, P_w and RH retrievals are presented up to 12 km throughout this study. However, as shown below, RH RMSE increases rapidly above 10 km; the comparison of retrieval accuracy against radiosonde measurement uncertainty is therefore restricted to below 10 km, where the moisture signal in refractivity remains robustly detectable. Expressing retrieval errors in terms of T and RH permits direct altitude-resolved comparison with GRUAN measurement uncertainties (Fig. 1). Since refractivity derives from the same radiosonde ascent used for evaluation, retrieved and reference profiles are inherently collocated and coincident. During model application, the model receives only the refractivity profile and its wavelet covariance transform (WCT_N; Section 2.4, Fig. 3); the thermodynamic reference profiles are withheld. No NWP prior or background state enters any retrieval presented in this paper; all model retrievals therefore constitute independent observational products.

Figure 5 summarizes RMSE, mean bias, and standard deviation for T, P_w, and RH across both datasets. Standard deviation (grey lines) represents natural parameter variability in the measured profiles. The comparison contextualizes RMSE magnitude: retrieval errors smaller than natural variability indicate the model preserves meaningful profile-to-profile distinctions rather than collapsing toward mean behaviour for each parameter. For temperature, RMSE in the training (evaluation) dataset is ~1 K (~1.5 K) near the surface, decreases to a minimum of ~0.5 K around 13 km, and increases again to ~1.6 K (~2.0 K) above 15 km (Fig. 5a,d). Benchmarked against the GRUAN temperature measurement uncertainty of ~0.2–0.4 K (k = 2; Fig. 1b), the model does not match this threshold in the lower troposphere — RMSE exceeds the radiosonde uncertainty by approximately a factor of 3–4 below 5 km. The 13–15 km range approaches but does not consistently reach the radiosonde benchmark. This progressive improvement — from ~1.5 K near the surface to ~0.5 K at 13 km — reflects the increasing dominance of dry refractivity over the moisture contribution to total refractivity at higher altitudes, reducing the number of effectively free thermodynamic unknowns and narrowing the solution space available to the model. Improving temperature retrieval accuracy, particularly in the lower troposphere, remains a clear direction for future work and is not

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

pursued further in the present study. The rise in RMSE above 15 km tracks the concurrent increase in N_d RMSE (Fig. 4) and reflects enhanced temperature variability near the cold-point tropopause driven by competing effects of deep convection and local subsidence. RMSE for *T* nonetheless remains below the standard deviation at all altitudes, confirming that the model carries meaningful temperature retrieval skill throughout the troposphere.

For moisture parameters, retrieval accuracy is markedly higher. For P_w, RMSE is ~0.5 hPa near the surface and decreases monotonically with altitude (Fig. 5b,e), primarily reflecting the natural decline in atmospheric moisture content. For *RH*, RMSE remains below 6 % from 100 m to ~8 km against *RH* standard deviation reaching 30 % (Fig. 5c,f). For context, mean GRUAN *RH* measurement uncertainty is approximately 5% at coverage factor k = 1 and ~10 % at k = 2 (Fig. 1c); individual profile uncertainty values vary around it. The model achieves *RH* RMSE below 6 % from 100 m to 10 km. While RMSE and measurement uncertainty are not statistically equivalent quantities — the former includes retrieval error while the latter characterizes measurement dispersion — their similar magnitudes suggest that CNN retrieval errors are comparable in scale to the irreducible uncertainty of the reference measurements themselves. This comparison provides practical context rather than formal equivalence. Above 10 km, *RH* RMSE increases sharply, consistent with the near-zero moisture content near the convective tropopause, and the accuracy comparison with radiosonde measurement uncertainty does not extend above this level. Mean bias is negligible for *T*, P_w, and *RH* in both datasets across nearly the full altitude range; the only notable exceptions are a small warm bias of ~0.25 K for *T* above 15 km and a small dry bias of ~2.5 % in *RH* in the evaluation data (Fig. 5d,f).

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

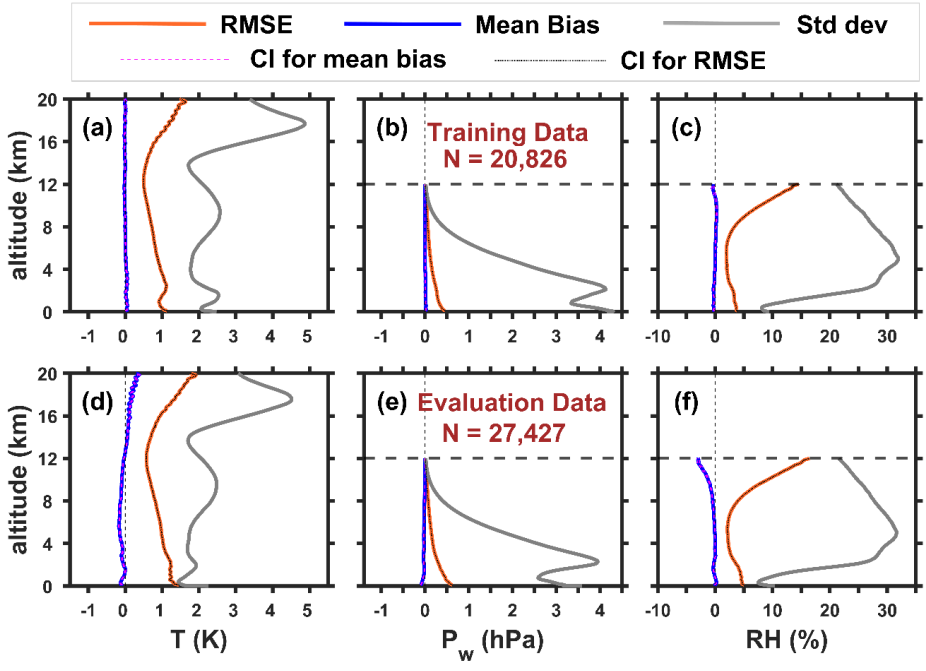


Figure 5. Vertical profiles of RMSE, mean bias (Retrieved – Reference), and standard deviation for temperature (T) **(a, d)**, water vapor pressure (P_w) **(b, e)**, and relative humidity (RH) **(c, f)** at 10 m vertical resolution from 100 m to 20 km. Standard deviation (grey lines) represents natural parameter variability in the measured profiles, providing context for the magnitude of retrieval errors. The 95 % confidence intervals for mean bias and RMSE are plotted as regions bounded by their upper and lower limits (as indicated in the legend); however, owing to the large number of profiles, the separation between the upper and lower bounds is narrower than the line widths used in the plots and is therefore not visually discernible. Panels **(a–c)**: training data; panels **(d–f)**: evaluation data. The number of radiosonde ascents used for training and evaluation is indicated (Tables 2-4). The dashed horizontal line at 12 km marks the upper limit for P_w and RH evaluation, above which atmospheric moisture content is too low for reliable retrieval.

The 95 % confidence intervals for both mean bias and RMSE (included in the figure as per legend) are extremely narrow at all altitudes — typically less than 0.01 K for temperature RMSE, less than 0.005 hPa for P_w RMSE, and less than 0.1 % for RH RMSE — confirming that the reported error metrics are stable estimates of the model's retrieval

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

performance and are not artefacts of the particular sample of profiles used. The narrow confidence intervals also establish that the small mean biases noted above (e.g., ~0.25 K warm bias in T above 15 km, ~2.5 % dry bias in RH) are statistically distinguishable from zero, representing genuine systematic tendencies in the retrieval rather than sampling fluctuations. Similarly, the RMSE differences observed between altitude ranges — for instance, the improvement in temperature RMSE from ~1.5 K near the surface to ~0.5 K at 13 km — are statistically robust, as the confidence intervals at these altitudes do not overlap. It should be noted, however, that these confidence intervals quantify the precision of the error statistics given the sample size; they do not account for systematic uncertainties arising from the retrieval methodology, the representativeness of the training data, or the measurement uncertainties in the reference radiosonde profiles themselves.

Figure 5d–f characterizes the model's overall retrieval capability across the tropical ocean domain. Station-by-station evaluations appear in Supplementary Figures S9–S11, Section S3; a collective summary follows.

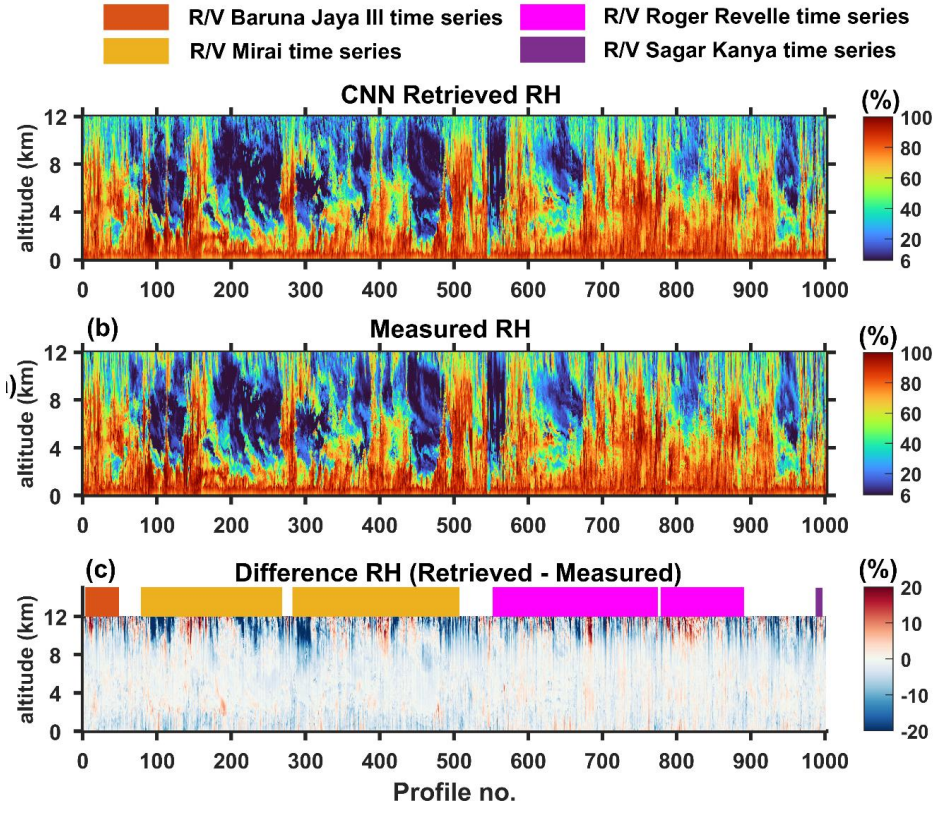
Across all individual stations and field campaigns, T RMSE is ~1.5 K near the surface, decreases to ~0.5 K in the 13–15 km range, and rises to ~2 K above 15 km. Mean bias is typically within ±0.3 K with no systematic altitude-dependent variability; a maximum negative bias of −1 K over Marquesas Island close to the surface, possibly reflects localized regional variability. For P_w, RMSE is ~0.5 hPa near the surface and decreases with altitude across all stations, with mean biases ranging from −0.5 to +0.5 hPa near the surface; although these magnitudes are small in absolute terms, at certain stations, bias magnitudes approach the RMSE, indicating that systematic offset contributes to the total error budget at those locations warranting further investigation. For *RH*, RMSE is below 8 % throughout 0–10 km — where the standard deviation reaches 30 % — and compares favourably with inter-sensor comparison errors reported for simultaneous radiosonde pairs (Sapucci et al., 2005; Xie et al., 2014). The 10–15 km altitude range consistently yields the best temperature retrieval (~0.5 K RMSE) and the highest *RH* retrieval error (~15 % RMSE 10-12 km). The 95 % confidence intervals for mean bias and RMSE for all three parameters at individual stations (Fig. 5d–f) remain narrow enough to confirm that the inter-station differences in retrieval accuracy

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

represent genuine regional variability rather than sampling noise. RMSE remains below the standard deviation for all three parameters at all altitudes, demonstrating model skill across the full range of atmospheric variability measured at multiple geographically independent locations spanning the tropical oceans.

3.2.1 Time-series evaluation of individual profiles: DYNAMO 2011–12, tropical Indian Ocean

Retrieval fidelity at the individual profile level over open ocean is assessed using ~1,000 ship-based radiosonde ascents from the DYNAMO 2011–12 field campaign over the tropical Indian Ocean (Table 4; Fig. 6), encompassing contrasting meteorological regimes and a broad range of local times. The dataset spans deep convective conditions (typically $RH > 80$ % at 10 km) through to subsidence-dominated episodes ($RH < 40$ % below 5 km), and includes diurnal sampling at ~3-hour intervals (~8 ascents per day).



https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Figure 6. CNN-retrieved relative humidity (*RH*) **(a)** and radiosonde-measured *RH* **(b)** at 10 m vertical resolution from 100 m to 20 km for 1,002 profiles, arranged chronologically by ship cruises from the DYNAMO 2011–12 campaign as per sequence listed in Table 4. As some timelines overlap, the x-axis is as per the 'Profile no.' The bottom panel **(c)** displays the difference between predicted and reference *RH* for each altitude and profile. Shaded sections at the top of panel **(c)** indicate time-series observations for cruises at specific locations (Table 4, Fig. 2). Other observations were taken during ship transits across the tropical Indian Ocean.

Within 100m – 10 km, the model retrieves *RH* with differences from measurements of less than 5 % under both deep convective and dry conditions (Fig. 6c), consistent with the GRUAN one-sigma measurement uncertainty. The marine atmospheric boundary layer — typically ~600 m deep over the tropical oceans (Johnson et al., 2001; Johnson and Ciesielski, 2017; Santosh, 2022b) — produces no anomalous retrieval errors relative to the overlying troposphere, confirming that the model captures water vapor variability within this dynamically complex layer. Subtle time-evolving structures in *RH* (e.g., profiles 100–500 in the RV *Mirai* time series; Fig. 6a,b) are faithfully reproduced. Above 10 km, differences increase without systematic sign, consistent with the diminishing moisture signal in refractivity. Low-bias retrievals for P_w (0–12 km) and T (0–15 km) across most individual profiles appear in Supplementary Figs. S12 and S13, confirming that the CNN recovers all three thermodynamic parameters simultaneously over the open ocean with the accuracy characterised in Section 3.2. It should be noted, however, a temporal overlap exists between the DYNAMO 2011–12 training and evaluation datasets (Tables 2 and 3); although the respective locations are spatially separated by hundreds of kilometres (Fig.2), shared large-scale synoptic conditions during this period cannot be excluded.

3.2.2 Fine-scale moisture sensitivity of the CNN model

The fine-scale moisture sensitivity of the model was probed through a controlled perturbation experiment. A +30 % *RH* delta spike was superimposed at 3000 m and 3020 m in a reference profile (profile no. 219 in Fig. 6; RV *Mirai*, 18 October 2011, 1200 UTC, 80.5°E, 0°), leaving *RH* at the intermediate 3010 m level unchanged. The

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

corresponding perturbed refractivity (N) and wavelet covariance transform (WCT_N) were supplied to the model.

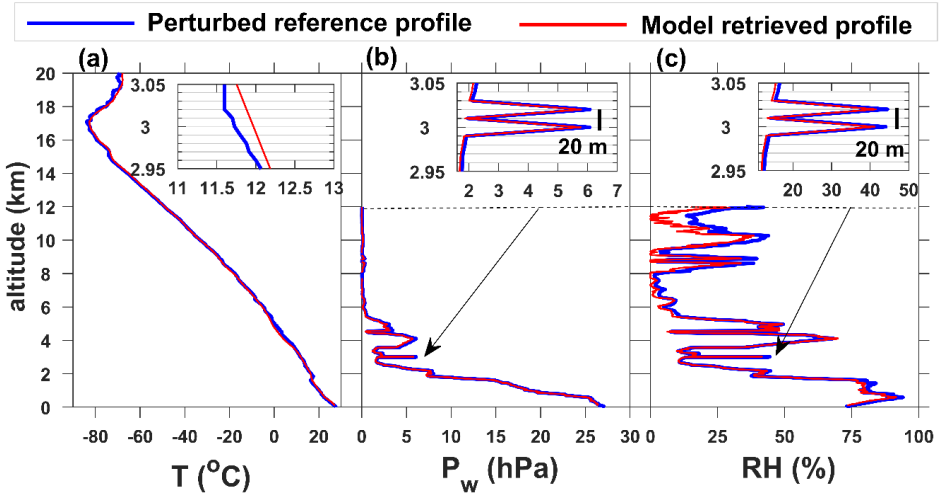


Figure 7. Reference and model-retrieved profiles for the radiosonde ascent aboard RV *Mirai* on *18th October 2011 at 1200 UTC* at location *(80.5°E,0)* over the Indian Ocean. The altitude range is 100 m to 20 km with a vertical grid spacing of 10 m for temperature (**a**), water vapor pressure (**b**), and relative humidity (**c**). Insets highlight regions of interest with consistent units for axes. Arrows indicate the altitude of *RH* delta perturbation in the original profile. The differences for the profiles (without perturbation) are given in supplementary Section.S5, Fig.S14 (d-f).

The model reproduced the delta perturbations in *RH* and P_w at 3000 m and 3020 m with only ~2 % underestimation, and left the retrieved value at 3010 m unchanged — matching the reference exactly (Fig. 7b,c). The retrieved *T* profile remained unaffected at the perturbed altitudes (Fig. 7a), demonstrating that the model correctly attributes the refractivity anomaly to moisture rather than temperature, despite the two being jointly encoded in *N*. These results demonstrate that the CNN preserves moisture gradients at the computational grid spacing (10 m) when provided with high-resolution input refractivity. This capability reflects the native resolution of the training data and the model architecture; fine-scale moisture sensitivity in satellite applications will be

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

fundamentally limited by the observation operator. For GNSS-RO, Fresnel zone averaging constrains achievable resolution to approximately 0.5–1.4 km regardless of retrieval grid spacing (Kursinski et al., 1997). The model's qualitatively different response to temperature perturbations (Supplementary Section S5) indicates that temperature retrievals do not achieve equivalent fine-scale fidelity.

Training data feature native vertical resolution finer than 10 metres (Table 1) and are re-gridded to 10 metre spacing, ensuring physically realistic vertical structures constrain the model. Thus, predictions are generated on a 10 metre native grid. However, this discretization reflects the model's computational architecture, not its resolving power. The model cannot retrieve 10 metre moisture features unless such information is present in the input refractivity—for example, GNSS radio occultation measurements inherently average over hundreds of metres, fundamentally limiting resolution regardless of output grid spacing.

3.2.3 Physical plausibility and learned atmospheric constraints

Supplementary Section S5 evaluates the response to single-level perturbations in input refractivity profiles. The architecture employs two independent CNNs—Model 0 predicting dry refractivity (N_d) and Model 1 predicting dry pressure (P_d)—trained on identical inputs of N and its wavelet covariance transform (Section 2.4), but with separate loss functions and no shared weights.

When unrealistic sharp temperature spikes (±2 °C) were introduced into input N profiles, neither model predicted corresponding anomalies in temperature; Model 0 and Model 1 independently output N_d and P_d values consistent with the background thermodynamic state (Fig. S15, Fig. S16, Section S5). Because the models were trained on the same quality-controlled dataset—where sharp temperature discontinuities are absent due to buoyancy-driven mixing—their independent predictions yield P_d/N_d ratios that fall within the moist adiabatic range. Consequently, the analytically derived temperature ($T = 77.6890\ P_d/N_d$) remains smooth, while the residual wet refractivity ($N_w = N - N_d$) absorbs the perturbation variance during post-processing. Similarly, realistic sharp moisture gradients introduced into N resulted in Model 0 predictions of N_d that,

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

when subtracted from the perturbed N, preserved the humidity structure in the derived P_w with minimal distortion to the derived T.

This response is asymmetric: the CNN faithfully recovers sharp moisture perturbations but smooths over sharp temperature perturbations of comparable magnitude (Fig. S15, Section S5). This asymmetry has a physical basis. In the lower troposphere, refractivity is far more sensitive to moisture than to temperature ($|\partial N/\partial Pw| \gg |\partial N/\partial T|$; Ao et al., 2003), and buoyancy constraints render sharp temperature discontinuities thermodynamically unsustainable while permitting sharp moisture boundaries (Betts, 1986; Sherwood et al., 2010).

The CNN does not encode these constraints explicitly. Instead, each model independently learns to approximate the conditional expectation of its target variable given the input refractivity — E[Nd|N] and E[Pd|N] — from the training data. Because the training observations reflect tropical tropospheric states where sharp moisture gradients are common but sharp temperature discontinuities are rare, the learned mappings inherit this statistical structure. The result, when the independent N_d and P_d predictions are combined analytically, is that derived temperature profiles are smooth while derived moisture profiles preserve fine-scale gradients. This behavior is consistent with the physical constraints governing tropical tropospheric variability, though it arises from the statistics of the training data rather than from explicit physical encoding. A practical consequence is that real sharp thermal features — such as trade wind inversions or boundary layer caps — may be systematically smoothed in CNN temperature retrievals, a limitation that users should consider when interpreting model output.

3.3 Spatial, temporal, and instrument generalizability and robustness of the DL model

The geographically independent evaluation stations span all three tropical ocean basins across both hemispheres: Guam (13.48°N, 144.79°E), Guadeloupe (16.26°N, 61.53°W), Barbados (13.07°N, 59.5°W), French Guiana (4.82°N, 52.37°W), Hawaii (21.99°N, 159.3°W), American Samoa (14.3°S, 170.7°W), French Polynesia (9.8°S, 139.03°W), Tahiti (17.56°S, 149.6°W), and Cocos Island (12.1°S, 96°E), supplemented by

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

ship-based observations from DYNAMO 2011–12 and TOGA-COARE 1992–93. The model achieves consistently low RMSE and mean bias across all these sites (Supplementary Figs. S9–S11, Section.S3), demonstrating accurate generalization well beyond the geographic footprint of the training data within the tropical ocean domain.

This consistent performance extends across multiple radiosonde types — Vaisala RS92, RS41, iMS-100, Graw, and RS80 instruments — none of which share a common calibration lineage or humidity sensor design. The model's retrieval accuracy does not degrade systematically with instrument type (Supplementary Figs. S9–S11, Section S3), a result that is consistent with the CNN having learned physical vertical relationships intrinsic to the tropical marine atmosphere rather than the statistical characteristics of any specific instrument, training station, or observing period.

The model shows no significant increase in T, P_w, or RH RMSE within the lowest 1 km over either islands or open ocean, confirming that the island-based training sites introduce no systematic boundary-layer biases into retrievals over open-ocean locations. This consistency suggests that the CNN has learned marine atmospheric boundary layer (MABL) thermodynamic relationships that are robust to the island versus open-ocean distinction, likely because island stations in the deep tropics sample a broad range of MABL states — from strongly coupled to decoupled conditions — that are representative of the wider tropical marine atmosphere.

Consistent RMSE and mean bias across this dataset (Figs. 4, 5), smoothly convergent learning curves (for selected hyper-parameter combination; Supplementary Fig. S5), and stable performance across model hyper-parameter sensitivity experiments (Supplementary Section S2) collectively rule out overfitting or underfitting in the model.

Temporal generalizability further reinforces this picture. Multi-year retrievals at GUAN stations and successful retrieval from the TOGA-COARE 1992–93 dataset — collected more than two decades before the training period and using the older Vaisala RS80 sonde — confirm that model skill persists well outside the training temporal window. For TOGA-COARE stations, T, P_w, and RH RMSE and mean bias values are consistent with those obtained at finer-resolution evaluation stations across other ocean

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

basins (Section 3.2), confirming that retrieval accuracy remains within the bounds established for the model even for inputs with coarser native vertical resolution (~40 m).

The consistency of retrieval performance across instruments, ocean basins, and observing periods spanning more than three decades provides the observational basis for the interpretive assessments presented in Section 4.

4 Discussion

Retrieving temperature and water vapor from atmospheric refractivity is inherently underdetermined: a single refractivity observation constrains three thermodynamic unknowns simultaneously, precluding direct analytical inversion. Rather than inverting the refractivity equation point-by-point at each altitude, the CNN reconstructs complete vertical output profiles simultaneously, encoding both local gradients and global profile structure through hierarchical convolutional feature extraction. This approach yields a robust approximation to the solution of the underdetermined problem, particularly for moisture retrieval. This is feasible because real atmospheric states occupy only a small subset of the full theoretical solution space: relative humidity rarely exceeds 100 %, tropical temperatures span a narrow range at each altitude level, and exponentially decreasing pressure with altitude imposes strong structural regularity on vertical profiles. The CNN learns these physical bounds implicitly from training data, using them to constrain what is otherwise an underdetermined system.

Central to the model's accuracy is a deliberate architectural choice: the CNN predicts dry refractivity (N_d) and dry pressure (P_d) rather than temperature and water vapor directly. This is mechanistically justified by the mathematical structure of the refractivity equation. N_d and P_d share linear relationships with total refractivity N, whereas T and P_w appear through nonlinear, coupled $1/T$ and $1/T^2$ terms that substantially increase the risk of overfitting for a regionally constrained training dataset. This is important in the context of the asymmetry between the number of ground-based radiosonde profiles (training data) and the available satellite refractivity profiles (application data). Predicting N_d and P_d and deriving T, P_w and RH analytically preserves full thermodynamic consistency while operating in the most tractable decomposition of

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

the refractivity equation — a distinction from prior deep learning retrieval studies that predict thermodynamic variables directly, and one that leads to the accuracy levels demonstrated here.

The accuracy of any supervised learning model is bounded by the quality of its training reference. Training on reanalysis fields or 1D-Var retrievals — as adopted in previous deep learning studies (Lasota, 2021; Shyam et al., 2013; Hooda et al., 2023) — propagates the biases of those background systems into the trained model, producing retrievals that are not independent of NWP priors. Training on in situ data anchors the CNN to in situ physical reality within known uncertainty bounds, making its retrievals genuinely independent of any NWP prior — the foundational premise on which the validation strategy of this study is constructed. When evaluated against geographically independent radiosonde observations, the CNN achieves *RH* RMSE typically of 6%, remaining below 10 % up to 10 km altitude, against a GRUAN *RH* measurement uncertainty of approximately 10 % (coverage factor $k = 2$; Fig. 1c). Temperature RMSE remains below 1.5 K from 100 m to 15 km, with optimal performance of ~0.5 K in the 10–15 km range where radiosonde uncertainty is ±0.4 K (for k=2). These results hold consistently across geographically independent datasets spanning three tropical ocean basins.

The results collectively address the two hypotheses stated at the outset. The first hypothesis is partially supported for moisture parameters (*RH* and P_w) below 10 km. The CNN RH RMSE of below 6 % is comparable to the mean GRUAN measurement uncertainty (~5 % at k=1 and ~10 % at k=2), with individual profile uncertainties varying substantially around this value (Fig. 1, shaded regions). This near-equivalence presents an interpretive challenge: observed RMSE reflects an inseparable combination of retrieval error and reference data uncertainty. Whether further algorithmic refinement could reduce retrieval errors—or whether such improvements would be masked by reference uncertainty—cannot be determined from aggregate statistics alone. But CNN retrieval errors are clearly of the same order as the best available in situ measurement precision. For temperature, the model demonstrates useful and consistent retrieval skill across the full troposphere, but RMSE does not reach the radiosonde measurement uncertainty threshold, particularly in the lower troposphere where it exceeds the

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

radiosonde standard by a factor of approximately 3–4 below 5 km. The objective was to benchmark against GRUAN uncertainty, not to meet it uniformly across all altitude regimes. The progressive improvement in temperature RMSE with altitude — from ~1.5 K near the surface to ~0.5 K at 13 km — reflects the decreasing contribution of moisture to total refractivity aloft and identifies a possible physical pathway for future retrieval improvement.

Improving temperature retrieval accuracy, particularly in the lower troposphere, remains an open challenge not addressed in the present study. The second hypothesis — that the model generalizes because it has learned the physical vertical structure of the tropical marine atmosphere, rather than the statistical signature of specific training stations, instrument types, or observing periods — is consistent with the cross-instrument and cross-basin performance documented in Section 3.2 and Supplementary Section S3, Fig.S9-S11.

Sensitivity analysis confirms a noteworthy behavioural property the model: the CNN largely ignores unphysical sharp temperature perturbations, attributing their refractivity signature to moisture instead, yet accurately recovers realistic humidity gradients (Sect.3.2.3, Supp. Sect.S5). This behaviour is explained by the statistical structure of the training data: in the tropical lower troposphere, high-frequency vertical variability in refractivity is overwhelmingly dominated by moisture gradients rather than temperature gradients. The model framework therefore learns the dominant statistical covariance of its training distribution, wherein a localised spike in refractivity is almost exclusively associated with a moisture anomaly. This is a statistical property of the training dataset rather than evidence of learned physical laws, and the model's response to unphysical temperature perturbations simply reflects the low prior probability of such structures in the observed training data. While this behaviour is desirable from a retrieval perspective, its scope is inherently bounded by the training distribution, and caution is warranted in attributing it to generalizable physical reasoning.

It is emphasized that the vertical resolution of 10 m for moisture retrieval applies to the forward-model retrieval chain (radiosonde → refractivity → retrieval) under ideal input conditions. Real GNSS-RO observations possess fundamentally limited vertical

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

resolution due to Fresnel zone effects (~0.5–1.4 km) and diffraction in moist conditions (Kursinski et al., 1997; Gorbunov, 2002). The 10m grid spacing represents the model's native capability, not the information content resolution achievable from actual RO missions.

A property of foundational importance emerges when the model architecture is examined carefully. The CNN receives a vertical refractivity profile as input; nowhere in its specification is the provenance of that profile defined. The model cannot distinguish whether the input was derived from a radiosonde, a GNSS-RO satellite, or any other refractivity-profiling instrument — its domain of applicability is bounded by the physical quality of the input profile, not by the observing platform. The CNN therefore learns the intrinsic relationship between refractivity vertical structure and thermodynamic state as represented in the training data. This explains why retrievals from TOGA-COARE 1992–93 data — a different radiosonde generation, coarser native resolution, two decades before the training period — are accurate: they are the expected consequence of a model that has learned a relationship that is not specific to any single instrument, dataset, or epoch. Consistent performance across evaluation sites hundreds of kilometres from the nearest training location further supports the spatial stationarity assumption — that vertical refractivity-thermodynamic relationships are sufficiently stable across the tropical marine atmosphere to generalize beyond specific training locations. The framework is, in principle, a general-purpose learned inversion operator for atmospheric refractivity, applicable to any platform that delivers physically accurate refractivity profiles over tropical oceans. Whether this platform-agnostic property holds when confronted with the distinct characteristics of satellite-based refractivity —including its larger horizontal footprint, different noise structure (e.g., residual ionospheric calibration errors, horizontal refractivity structure violating the spherical symmetry assumption, and critical refraction and multipath in the moist lower troposphere etc.), and global spatial coverage — remains to be seen.

The CNN remains a black-box operator: its learned mappings lack analytical transparency, though perturbation experiments (Supplementary Section S5) provide behavioural diagnostics. The model accepts input refractivity as given, without mechanisms to detect or correct for input data quality; any biases or noise present in

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

refractivity propagate directly into retrievals. Furthermore, the model's learned statistical constraints favour smooth temperature profiles, inherently limiting retrieval of sharp thermal discontinuities (Section 3.2.3; Supplementary Section S5). These are intrinsic properties of the current architecture rather than limitations of the present study design; addressing them would require fundamental methodological extensions such as input quality assessment layers.

4.1 Limitations

1. *Closed-loop evaluation.* The experimental design employs refractivity inputs computed from the same radiosonde measurements used as reference truth. This characterizes the CNN's learned mapping under ideal input conditions but cannot assess performance degradation from real measurement errors inherent to satellite observations such as residual ionospheric calibration errors, horizontal refractivity structure violating the spherical symmetry assumption, and critical refraction and multipath in the moist lower troposphere.
2. *Geographic and thermodynamic scope.* Training and evaluation data are confined to island and ship locations within the tropical marine atmosphere (~25°S–25°N). Performance in substantially different thermodynamic regimes — subtropical dry zones, mid-latitudes, polar regions — remains uncharacterized.
3. *DYNAMO temporal overlap.* One potential concern is that DYNAMO 2011–12 contributes to both training (island stations: Gan, Diego Garcia) and evaluation (ship cruises). Although these observations are spatially separated by hundreds of kilometers, they share the same campaign period and large-scale synoptic environment. Retrieval statistics for DYNAMO ship data are consistent with temporally independent evaluation stations (Supplementary Figures S9–S11, Section S3), indicating that temporal overlap with training data does not confer measurable advantage. The model's performance on DYNAMO ships reflects learned physical relationships rather than memorized synoptic patterns suggesting that temporal overlap with training data does not confer measurable advantage.
4. *Point estimates only.* The model provides point estimates without per-profile uncertainty quantification. Comprehensive uncertainty characterization is beyond the scope of the present work and will be addressed separately.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

5. *Retrieval-reference uncertainty ambiguity.* For moisture retrieval below 10 km, where aggregate RMSE approaches the mean GRUAN measurement uncertainty, definitive attribution is not possible — observed errors reflect an unknown mixture of retrieval inaccuracy and reference data uncertainty. Individual profile uncertainties vary substantially around the mean values shown in Figure 1, further complicating this separation. Since the CNN was trained on radiosonde data carrying these same uncertainties, training data noise may also impose an accuracy ceiling on moisture retrieval independent of algorithmic capability. This ambiguity does not affect temperature, where retrieval RMSE substantially exceeds measurement uncertainty throughout the troposphere.

5 Conclusions

This study demonstrates that a convolutional neural network trained exclusively on high-resolution, quality-controlled in situ radiosonde observations can retrieve tropospheric water vapor and temperature from atmospheric refractivity without recourse to NWP-derived prior states, instead employing statistical constraints learned from radiosonde measurement-based atmospheric structures. The value of the approach lies in the pedigree of the priors, not the elimination of prior assumptions which remain mathematically necessary for an inherently underdetermined system. The model employs both refractivity (N) and its wavelet covariance transform (WCT_N) vertical profiles as inputs to predict dry refractivity (N_d) and dry pressure (P_d) as physically motivated intermediate targets, from which temperature, water vapor pressure, and relative humidity are derived analytically. It was trained on 20,826 radiosonde ascents from tropical island stations and evaluated against 27,427 geographically independent ascents across the tropical Pacific, Indian, and Atlantic Oceans. The scope of this work is explicitly confined to tropical oceanic regions.

The principal findings are:

1. **Moisture retrieval approaches the radiosonde measurement benchmark.** Relative humidity RMSE is typically 6 % and remains below 10 % from 100 m to 10 km, approaching the GRUAN standard measurement uncertainty (~5 % at coverage

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

factor k = 1; ~10 % at k = 2)—with negligible mean bias; retrievals extend to 12 km but benchmark comparison is restricted to 10 km where the moisture signal remains robust. Water vapor pressure RMSE is ~0.5 hPa near the surface, decreasing monotonically with altitude.

2. **Temperature retrieval demonstrates skill to 20 km with altitude dependent accuracy and negligible mean bias;** achieved through analytical derivation from predicted dry refractivity (N_d) and dry pressure (P_d). RMSE is ~1.5 K near the surface exceeding the GRUAN measurement uncertainty by a factor of three to four (at k=2) below 5 km. RMSE improves to ~0.5 K at 13–15 km. The progressive improvement with altitude is consistent with the increasing dominance of dry refractivity fractional contribution to total refractivity and associated reduction in thermodynamic degrees of freedom.

3. **Generalization across evaluation stations:** Consistent retrieval accuracy across training (N = 20,826) and evaluation (N = 27,427) datasets — spanning geographically distant stations and five radiosonde types (Vaisala RS80, RS92, RS41; Meisei iMS-100; Graw) — is consistent with the CNN having learned generalizable thermodynamic relationships from refractivity structure rather than station-specific or instrument-specific artifacts. However, this generalization is demonstrated only for radiosonde-derived refractivity inputs; transfer to satellite observations requires separate assessment.

4. **Fine-scale moisture sensitivity under ideal input conditions.** The model operates on a 10 m computational grid and preserves fine-scale moisture structure in retrievals when provided with high-fidelity input refractivity vertical structure. This grid-scale sensitivity reflects the model's native capability under ideal input conditions; application to GNSS-RO observations will be limited by Fresnel-zone averaging (ranging from ~1.4 km in the middle atmosphere to ~0.5 km in the lower troposphere where strong refractivity gradients exist). The model does not resolve sharp temperature perturbations at comparable scales, as learned thermodynamic constraints enforce smoothness consistent with convective quasi-equilibrium (Section 3.2.3).

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

5. **The framework is designed for input-source independence.** Because only the refractivity profile enters the model, the architecture imposes no constraints on observation platform providing model input and serves as a general purpose inversion operator for atmospheric refractivity. However, satellite observations introduce measurement characteristics absent from training data, like horizontal averaging, specific noise structures, and potential systematic biases whose effects on retrieval accuracy are not characterized by the present analysis.

Key limitations include: the CNN is a data-driven approximation to an inherently underdetermined problem (Section 4); performance outside the tropical marine atmosphere regime and sensitivity to input refractivity errors remain uncharacterized.

Part 2 of this work applies this framework to GNSS radio occultation profiles from COSMIC-2/FORMOSAT-7, assessing the extension of these capabilities to satellite-based refractivity.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Code and Data availability

GRUAN certified data products used in study are available at https://doi.org/10.5676/GRUAN/RS92-GDP.2 (Sommer et al., 2012), https://doi.org/10.5676/GRUAN/RS41-GDP.1 (Sommer et al., 2022), and https://doi.org/10.5676/GRUAN/IMS-100-GDP.2 (Hoshino et al., 2022).

DYNAMO 2011–12 field campaign data are provided by NCAR/EOL and are available at https://doi.org/10.26023/5CX4-D564-0M12 (Gan Island; NSF NCAR Earth Observing Laboratory and Atmospheric Radiation Measurement (ARM) user facility, 2012), https://doi.org/10.5065/D6QF8R2Z (Diego Garcia; NSF NCAR Earth Observing Laboratory, 2012), https://doi.org/10.26023/GT5S-GSCP-3Q05 (RV *Baruna Jaya III*; NSF NCAR Earth Observing Laboratory et al., 2013), https://doi.org/10.26023/ZFQ2-TG96-TA0P, (RV *Mirai*; NSF NCAR Earth Observing Laboratory and Japan Agency for Marine-Earth Science and Technology, 2013), https://doi.org/10.5065/D6TH8K3Z (RV *Roger Revelle*; NSF NCAR Earth Observing Laboratory, 2014), and https://doi.org/10.26023/2JNV-7YQ5-610T (RV *Sagar Kanya*; NSF NCAR Earth Observing Laboratory et al., 2013)

TOGA-COARE 1992–93 data are provided by NCAR/EOL and are available at https://doi.org/10.5065/D6TQ5ZTV (Kapingamarangi; NSF NCAR Earth Observing Laboratory, 2006a), https://doi.org/10.5065/D62805XG (Kavieng; NSF NCAR Earth Observing Laboratory, 2006b), https://doi.org/10.5065/D66W98CD (Manus; NSF NCAR Earth Observing Laboratory, 2006c), https://doi.org/10.5065/D6K64GB2 (Nauru; NSF NCAR Earth Observing Laboratory, 2006d), https://doi.org/10.5065/D67S7M21 (RV *Kexue*; NSF NCAR Earth Observing Laboratory, 2006e), https://doi.org/10.5065/D6416VBM (RV *Shiyan 3*; NSF NCAR Earth Observing Laboratory, 2006g), https://doi.org/10.5065/D6GB22BS (RV *Xiangyanhong 5*; NSF NCAR Earth Observing Laboratory, 2006h), https://doi.org/10.5065/D6VH5M3J (RV *Moana Wave*; NSF NCAR Earth Observing Laboratory, 2006f), https://doi.org/10.5065/D6513WHF (RV *Hakuho Maru;* NSF NCAR Earth Observing Laboratory, 2009a), and https://doi.org/10.5065/D6319T5S (RV *Vickers;* NSF NCAR Earth Observing Laboratory, 2009b).

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

GUAN high-resolution data are available at University of Wyoming Atmospheric Science Radiosonde Archive https://weather.uwyo.edu/upperair/sounding.shtml.

All figures and results presented in this study are derived from the publicly available radiosonde datasets listed above, processed according to the methodology described in Sect. 2.3. The CNN model architecture and all training parameters are fully described in Supplementary Section S1; trained model weights are available from the corresponding author upon reasonable request.

Retrieved and reference profiles for all evaluation stations and field campaigns (Tables 3 and 4; N = 27,427) which constitutes processed dataset are additionally archived at Zenodo (Santosh et al., 2026) and will be made publicly accessible upon acceptance of this manuscript. Access during peer review is available to reviewers via the journal editor.

CRediT authorship contributions

SM: Conceptualization, Methodology, Software, Data Curation, Validation, Formal analysis, Investigation, Visualization, Writing-Original Draft, Writing - Review and Editing. **HJ:** Software. **DM:** Writing - Review and Editing.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors wish to thank the scientific and technical staff at the four GRUAN observatories (Singapore, Manus, Nauru and Minamitorishima) for dutifully performing reference radio soundings and maintenance of data according to the GRUAN standards. The authors thank the GCOS, initiated by the World Meteorological Organisation (WMO), for making this valuable data available. The authors thank the University of Wyoming for high resolution radiosonde data. The authors wish to thank the organizers and scientific and technical staff responsible for the radiosonde data from ship cruises and island

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

stations as part of the DYNAMO 2011-12 and the TOGA-COARE 1992-93 field campaigns. These data are provided by NCAR/EOL under the sponsorship of the National Science Foundation, https://data.eol.ucar.edu/. The authors thank the Indian Space Research Organisation (ISRO) for supporting the work.

During the preparation of this manuscript the authors used ChatGPT and similar AI in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Allan, R. P., Willett, K. M., John, V. O., and Trent, T.: Global Changes in Water Vapor 1979–2020, JGR Atmospheres, 127, e2022JD036728, https://doi.org/10.1029/2022JD036728, 2022.

Anthes, R. A.: Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather, Atmos. Meas. Tech., 4, 1077–1103, https://doi.org/10.5194/amt-4-1077-2011, 2011.

Anthes, R. A., Bernhardt, P. A., Chen, Y., Cucurull, L., Dymond, K. F., Ector, D., Healy, S. B., Ho, S.-P., Hunt, D. C., Kuo, Y.-H., Liu, H., Manning, K., McCormick, C., Meehan, T. K., Randel, W. J., Rocken, C., Schreiner, W. S., Sokolovskiy, S. V., Syndergaard, S., Thompson, D. C., Trenberth, K. E., Wee, T.-K., Yen, N. L., and Zeng, Z.: The COSMIC/FORMOSAT-3 Mission: Early Results, Bull. Amer. Meteor. Soc., 89, 313–334, https://doi.org/10.1175/BAMS-89-3-313, 2008.

Ao, C. O., Meehan, T. K., Hajj, G. A., Mannucci, A. J., and Beyerle, G.: Lower troposphere refractivity bias in GPS occultation retrievals, J. Geophys. Res., 108, 2002JD003216, https://doi.org/10.1029/2002JD003216, 2003.

Bai, W., Liu, C., Meng, X., Sun, Y., Kirchengast, G., Du, Q., Wang, X., Yang, G., Liao, M., Yang, Z., Zhao, D., Xia, J., Cai, Y., Liu, L., and Wang, D.: Evaluation of atmospheric profiles derived from single- and zero-difference excess phase processing of BeiDou radio occultation data from the FY-3C GNOS mission, Atmos. Meas. Tech., 11, 819–833, https://doi.org/10.5194/amt-11-819-2018, 2018.

Betts, A. K.: A new convective adjustment scheme. Part I: Observational and theoretical basis, Quart J Royal Meteoro Soc, 112, 677–691, https://doi.org/10.1002/qj.49711247307, 1986.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Bonafoni, S.: Atmospheric Profiling in the Inter-Tropical Ocean Area Based on Neural Network Approach Using GPS Radio Occultations, TOASCJ, 4, 202–209, https://doi.org/10.2174/1874282301004010202, 2010.

Bonafoni, S., Pelliccia, F., and Anniballe, R.: Comparison of Different Neural Network Approaches for the Tropospheric Profiling over the Inter-tropical lands Using GPS Radio Occultation Data, Algorithms, 2, 31–45, https://doi.org/10.3390/a2010031, 2009.

Bui, T., Ribeiro, L., Ponti, M., and Collomosse, J.: Sketching out the details: Sketch-based image retrieval using convolutional neural networks with multi-stage regression, Computers & Graphics, 71, 77–87, https://doi.org/10.1016/j.cag.2017.12.006, 2018.

Chang, H., Lee, J., Yoon, H., Morton, Y. J., and Saltman, A.: Performance assessment of radio occultation data from GeoOptics by comparing with COSMIC data, Earth Planets Space, 74, 108, https://doi.org/10.1186/s40623-022-01667-6, 2022.

Ciesielski, P. E., Yu, H., Johnson, R. H., Yoneyama, K., Katsumata, M., Long, C. N., Wang, J., Loehrer, S. M., Young, K., Williams, S. F., Brown, W., Braun, J., and Van Hove, T.: Quality-Controlled Upper-Air Sounding Dataset for DYNAMO/CINDY/AMIE: Development and Corrections, Journal of Atmospheric and Oceanic Technology, 31, 741–764, https://doi.org/10.1175/JTECH-D-13-00165.1, 2014.

Das, D., Santosh, K. C., and Pal, U.: Truncated inception net: COVID-19 outbreak screening using chest X-rays, Phys Eng Sci Med, 43, 915–925, https://doi.org/10.1007/s13246-020-00888-x, 2020.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463–4490, https://doi.org/10.5194/amt-7-4463-2014, 2014.

Dong, N., Zhao, L., Wu, C. H., and Chang, J. F.: Inception v3 based cervical cell classification combined with artificially extracted features, Applied Soft Computing, 93, 106311, https://doi.org/10.1016/j.asoc.2020.106311, 2020.

Eyre, J. R., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., and Pavelin, E. G.: Assimilation of satellite data in numerical weather prediction. Part II : Recent years, Quart J Royal Meteoro Soc, 148, 521–556, https://doi.org/10.1002/qj.4228, 2022.

Gamage, N. and Hagelberg, C.: Detection and Analysis of Microfronts and Associated Coherent Events Using Localized Transforms, J. Atmos. Sci., 50, 750–756, https://doi.org/10.1175/1520-0469(1993)050%3C0750:DAAOMA%3E2.0.CO;2, 1993.

Gleisner, H., Ringer, M. A., and Healy, S. B.: Monitoring global climate change using GNSS radio occultation, npj Clim Atmos Sci, 5, 6, https://doi.org/10.1038/s41612-022-00229-7, 2022.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Gorbunov, M. E.: Canonical transform method for processing radio occultation data in the lower troposphere, Radio Science, 37, https://doi.org/10.1029/2000RS002592, 2002.

Healy, S. B. and Eyre, J. R.: Retrieving temperature, water vapour and surface pressure information from refractive-index profiles derived by radio occultation: A simulation study, Quart J Royal Meteoro Soc, 126, 1661–1683, https://doi.org/10.1002/qj.49712656606, 2000.

Held, I. M. and Soden, B. J.: Water Vapor Feedback and Global Warming, Annu. Rev. Energy. Environ., 25, 441–475, https://doi.org/10.1146/annurev.energy.25.1.441, 2000.

Ho, S., Anthes, R. A., Ao, C. O., Healy, S., Horanyi, A., Hunt, D., Mannucci, A. J., Pedatella, N., Randel, W. J., Simmons, A., Steiner, A., Xie, F., Yue, X., and Zeng, Z.: The COSMIC/FORMOSAT-3 Radio Occultation Mission after 12 Years: Accomplishments, Remaining Challenges, and Potential Impacts of COSMIC-2, Bulletin of the American Meteorological Society, 101, E1107–E1136, https://doi.org/10.1175/BAMS-D-18-0290.1, 2020a.

Ho, S., Zhou, X., Shao, X., Chen, Y., Jing, X., and Miller, W.: Using the Commercial GNSS RO Spire Data in the Neutral Atmosphere for Climate and Weather Prediction Studies, Remote Sensing, 15, 4836, https://doi.org/10.3390/rs15194836, 2023.

Ho, S.-P., Zhou, X., Shao, X., Zhang, B., Adhikari, L., Kireev, S., He, Y., Yoe, J. G., Xia-Serafino, W., and Lynch, E.: Initial Assessment of the COSMIC-2/FORMOSAT-7 Neutral Atmosphere Data Quality in NESDIS/STAR Using In Situ and Satellite Data, Remote Sensing, 12, 4099, https://doi.org/10.3390/rs12244099, 2020b.

Hooda, S., Gupta, M., Singh, R., and Ojha, S. P.: Retrieval of Atmospheric Water Vapor Profiles From COSMIC-2 Radio Occultation Constellation Using Machine Learning, IEEE Trans. Geosci. Remote Sensing, 61, 1–7, https://doi.org/10.1109/TGRS.2023.3331187, 2023.

Hoshino, S., Sugidachi, T., Shimizu, K., Kobayashi, E., Fujiwara, M., and Iwabuchi, M.: Comparison of GRUAN data products for Meisei iMS-100 and Vaisala RS92 radiosondes at Tateno, Japan, Atmos. Meas. Tech., 15, 5917–5948, https://doi.org/10.5194/amt-15-5917-2022, 2022.

Hoshino, S.; Sugidachi, T.; Shimizu, K.; Kobayashi, E.; Fujiwara, M.; Iwabuchi, M. (2022): iMS-100 GRUAN Data Product Version 2 (IMS-100-GDP.2). Tateno Aerological Observatory (JMA) [data set], https://doi.org/10.5676/GRUAN/IMS-100-GDP.2

Hyland, R.W., Wexler, A., 1983. Formulations for the thermodynamic properties of the saturated phases of H_2O from 173.15 K to 473.15 K. *ASHRAE Trans.* 89(2A), 500–519.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, https://doi.org/10.5194/amt-3-1217-2010, 2010.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Jagadheesha, D., Simon, B., Pal, P.-K., Joshi, P. C., and Maheshwari, A.: A New Technique for Estimation of Lower-Tropospheric Temperature and Water Vapor Profiles from Radio Occultation Refractivity, Journal of Atmospheric and Oceanic Technology, 26, 1075–1089, https://doi.org/10.1175/2008JTECHA1145.1, 2009.

Jia, W., Sun, M., Lian, J., and Hou, S.: Feature dimensionality reduction: a review, Complex Intell. Syst., 8, 2663–2693, https://doi.org/10.1007/s40747-021-00637-x, 2022.

Johnson, R. H. and Ciesielski, P. E.: Multiscale Variability of the Atmospheric Boundary Layer during DYNAMO, Journal of the Atmospheric Sciences, 74, 4003–4021, https://doi.org/10.1175/JAS-D-17-0182.1, 2017.

Johnson, R. H., Ciesielski, P. E., and Cotturone, J. A.: Multiscale Variability of the Atmospheric Mixed Layer over the Western Pacific Warm Pool, J. Atmos. Sci., 58, 2729–2750, https://doi.org/10.1175/1520-0469(2001)058%3C2729:MVOTAM%3E2.0.CO;2, 2001.

Karouche, N., Goldstein, C., Rosak, A., Malassingne, C., and Raju, G.: MEGHA - TROPIQUES satellite mission: In flight performances results, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, 4684–4687, https://doi.org/10.1109/IGARSS.2012.6350420, 2012.

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, J. Geophys. Res., 102, 23429–23465, https://doi.org/10.1029/97JD01569, 1997.

Ladstädter, F., Steiner, A. K., Schwärz, M., and Kirchengast, G.: Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002 to 2013, Atmos. Meas. Tech., 8, 1819–1834, https://doi.org/10.5194/amt-8-1819-2015, 2015.

Lasota, E.: Comparison of different machine learning approaches for tropospheric profiling based on COSMIC-2 data, Earth Planets Space, 73, 221, https://doi.org/10.1186/s40623-021-01548-4, 2021.

Li, J., Lu, C., Zheng, Y., Ban, W., and Zhang, X.: Quality Assessment and Assimilation of Tianmu-1 GNSS Radio Occultation Refractivity Observations: A Preliminary Study, IEEE Trans. Geosci. Remote Sensing, 63, 1–17, https://doi.org/10.1109/TGRS.2025.3615559, 2025.

Loehrer, S. M., Edmands, T. A., and Moore, J. A.: TOGA COARE Upper-Air Sounding Data Archive: Development and Quality Control Procedures, Bull. Amer. Meteor. Soc., 77, 2651–2671, https://doi.org/10.1175/1520-0477(1996)077%3C2651:TCUASD%3E2.0.CO;2, 1996.

Luntama, J.-P., Kirchengast, G., Borsche, M., Foelsche, U., Steiner, A., Healy, S., Von Engeln, A., O'Clerigh, E., and Marquardt, C.: Prospects of the EPS GRAS Mission for Operational Atmospheric Applications, Bulletin of the American Meteorological Society, 89, 1863–1876, https://doi.org/10.1175/2008BAMS2399.1, 2008.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Mo, Z., Lou, Y., Zhang, W., Zhou, Y., Wu, P., and Zhang, Z.: Performance assessment of multi-source GNSS radio occultation from COSMIC-2, MetOp-B/C, FY-3D/E, Spire and PlanetiQ over China, Atmospheric Research, 311, 107704, https://doi.org/10.1016/j.atmosres.2024.107704, 2024.

Nalli, N. R., Barnet, C. D., Reale, A., Tobin, D., Gambacorta, A., Maddy, E. S., Joseph, E., Sun, B., Borg, L., Mollner, A. K., Morris, V. R., Liu, X., Divakarla, M., Minnett, P. J., Knuteson, R. O., King, T. S., and Wolf, W. W.: Validation of satellite sounder environmental data records: Application to the Cross-track Infrared Microwave Sounder Suite, JGR Atmospheres, 118, https://doi.org/10.1002/2013JD020436, 2013.

Nalli, N. R., Gambacorta, A., Liu, Q., Barnet, C. D., Tan, C., Iturbide-Sanchez, F., Reale, T., Sun, B., Wilson, M., Borg, L., and Morris, V. R.: Validation of Atmospheric Profile Retrievals From the SNPP NOAA-Unique Combined Atmospheric Processing System. Part 1: Temperature and Moisture, IEEE Trans. Geosci. Remote Sensing, 56, 180–190, https://doi.org/10.1109/TGRS.2017.2744558, 2018.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from Kapingamarangi. Version 1.0[data set], https://doi.org/10.5065/D6TQ5ZTV, 2006a.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from Kavieng. Version 1.0[data set], https://doi.org/10.5065/D62805XG, 2006b.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from Manus. Version 1.0[data set], https://doi.org/10.5065/D66W98CD, 2006c.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from Nauru. Version 1.0[data set], https://doi.org/10.5065/D6K64GB2, 2006d.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Kexue 1. Version 1.0[data set], https://doi.org/10.5065/D67S7M21, 2006e.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Moana Wave. Version 1.0[data set], https://doi.org/10.5065/D6VH5M3J, 2006f.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Shiyan 3. Version 1.0[data set], https://doi.org/10.5065/D6416VBM, 2006g.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Xiangyanghong 5. Version 1.0[data set], https://doi.org/10.5065/D6GB22BS, 2006h.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Hakuho-Maru. Version 1.0[data set], https://doi.org/10.5065/D6513WHF, 2009a.

NSF NCAR Earth Observing Laboratory: TOGA COARE soundings from R/V Vickers. Version 1.0[data set], https://doi.org/10.5065/D6319T5S, 2009b.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

NSF NCAR Earth Observing Laboratory: DYNAMO: Diego Garcia Radiosonde L3 Data. Version 1.0[data set], https://doi.org/10.5065/D6QF8R2Z, 2012.

NSF NCAR Earth Observing Laboratory: DYNAMO: R/V Roger Revelle Radiosonde L3.1 Data. Version 1.0[data set], https://doi.org/10.5065/D6TH8K3Z, 2014.

NSF NCAR Earth Observing Laboratory and Atmospheric Radiation Measurement (ARM) user facility: DYNAMO: Gan ARM AMF Radiosonde L3 Data. Version 1.0[data set], https://doi.org/10.26023/5CX4-D564-0M12, 2012.

NSF NCAR Earth Observing Laboratory and Japan Agency for Marine-Earth Science and Technology: DYNAMO: R/V Mirai Radiosonde L3.1 Data. Version 1.0[data set], https://doi.org/10.26023/ZFQ2-TG96-TA0P, 2013.

NSF NCAR Earth Observing Laboratory, Japan Agency for Marine-Earth Science and Technology, and Colorado State University: DYNAMO: R/V Baruna Jaya Radiosonde L3.1 Data. Version 1.0[data set], https://doi.org/10.26023/GT5S-GSCP-3Q05, 2013a.

NSF NCAR Earth Observing Laboratory, Japan Agency for Marine-Earth Science and Technology, and Colorado State University: DYNAMO: R/V Sagar Kanya Radiosonde L3.1 Data. Version 1.0[data set], https://doi.org/10.26023/2JNV-7YQ5-610T, 2013b.

Pelliccia, F., Pacifici, F., Bonafoni, S., Basili, P., Pierdicca, N., Ciotti, P., and Emery, W. J.: Neural Networks for Arctic Atmosphere Sounding From Radio Occultation Data, IEEE Trans. Geosci. Remote Sensing, 49, 4846–4855, https://doi.org/10.1109/TGRS.2011.2153859, 2011.

Pritt, M. and Chern, G.: Satellite Image Classification with Deep Learning, in: 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 1–7, https://doi.org/10.1109/AIPR.2017.8457969, 2017.

Reuger, J., 2002. Refractive indices of light, infra-red and radio waves in the atmosphere, Uniserv Report. *School of Surveying and Spatial Information Systems, University of New South Wales*, Sydney, Australia. S-68, pp. 1–52. https://www.unsw.edu.au/content/dam/pdfs/engineering/civil-environmental/historical-documents/unisurv-special-series/s68.pdf

Santer, B. D., Thorne, P. W., Haimberger, L., Taylor, K. E., Wigley, T. M. L., Lanzante, J. R., Solomon, S., Free, M., Gleckler, P. J., Jones, P. D., Karl, T. R., Klein, S. A., Mears, C., Nychka, D., Schmidt, G. A., Sherwood, S. C., and Wentz, F. J.: Consistency of modelled and observed temperature trends in the tropical troposphere, Intl Journal of Climatology, 28, 1703–1722, https://doi.org/10.1002/joc.1756, 2008.

Santosh, M.: Estimation of daytime planetary boundary layer height (PBLH) over the tropics and subtropics using COSMIC-2/FORMOSAT-7 GNSS–RO measurements, Atmospheric Research, 279, 106361, https://doi.org/10.1016/j.atmosres.2022.106361, 2022a.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Santosh, M.: Structure and development of the atmospheric boundary layer over a small island (Mahé Island, Seychelles) in the equatorial Indian Ocean, Meteorol Atmos Phys, 134, 91, https://doi.org/10.1007/s00703-022-00924-3, 2022b.

Santosh, M., John, H., and Mishra, D.: Retrieved and reference atmospheric profiles for CNN-based refractivity retrieval evaluation, Zenodo [data set], https://doi.org/10.5281/zenodo.19475727, 2026.

Sapucci, L. F., Machado, L. A. T., Da Silveira, R. B., Fisch, G., and Monico, J. F. G.: Analysis of Relative Humidity Sensors at the WMO Radiosonde Intercomparison Experiment in Brazil, Journal of Atmospheric and Oceanic Technology, 22, 664–678, https://doi.org/10.1175/JTECH1754.1, 2005.

Schreiner, W. S., Weiss, J. P., Anthes, R. A., Braun, J., Chu, V., Fong, J., Hunt, D., Kuo, Y. -H., Meehan, T., Serafino, W., Sjoberg, J., Sokolovskiy, S., Talaat, E., Wee, T. K., and Zeng, Z.: COSMIC-2 Radio Occultation Constellation: First Results, Geophysical Research Letters, 47, e2019GL086841, https://doi.org/10.1029/2019GL086841, 2020.

Seidel, D. J., Berger, F. H., Diamond, H. J., Dykema, J., Goodrich, D., Immler, F., Murray, W., Peterson, T., Sisterson, D., Sommer, M., Thorne, P., Vomel, H., and Wang, J.: Reference Upper-Air Observations for Climate: Rationale, Progress, and Plans, Bull. Amer. Meteor. Soc., 90, 361–369, https://doi.org/10.1175/2008BAMS2540.1, 2009.

Shao, X., Ho, S.-P., Zhang, B., Zhou, X., Kireev, S., Chen, Y., and Cao, C.: Comparison of COSMIC-2 radio occultation retrievals with RS41 and RS92 radiosonde humidity and temperature measurements, Terr. Atmos. Ocean. Sci., 32, https://doi.org/10.3319/TAO.2021.12.30.02, 2021.

Sherwood, S. C., Roca, R., Weckwerth, T. M., and Andronova, N. G.: Tropospheric water vapor, convection, and climate, Rev. Geophys., 48, RG2001, https://doi.org/10.1029/2009RG000301, 2010.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., Von Der Heydt, A. S., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence, Reviews of Geophysics, 58, e2019RG000678, https://doi.org/10.1029/2019RG000678, 2020.

Shyam, A., Gohil, B.S., Basu, S., 2013. Retrieval of water vapour profiles from radio occultation refractivity using artificial neural network. *Indian Journal of Radio & Space Physics (IJRSP)* 42, 411–419.

Sommer, M., Dirksen, R., Immler, F., 2023. Uncertainty representation for GRUAN data products. *GRUAN Technical Document* (GRUAN-TD-8), GRUAN Lead Centre, Lindenberg, Germany. https://www.gruan.org/documentation/gruan/td/gruan-td-8

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

Sommer, M.; Dirksen, R.; Immler, F. (2012): RS92 GRUAN Data Product Version 2 (RS92-GDP.2). GRUAN Lead Centre (DWD) [data set], https://doi.org/10.5676/GRUAN/RS92-GDP.2

Sommer, M.; von Rohden, C.; Simeonov, T.; Dirksen, R.; Fiedler-Krüger, M.; Friedrich, H.; Körner, S.; Naebert, T.; Oelsner, P.; Tietz, R. (2022): RS41 GRUAN Data Product Version 1 (RS41-GDP.1). GRUAN Lead Centre (DWD) [data set] , https://doi.org/10.5676/GRUAN/RS41-GDP.1

Steiner, A. K., Ladstädter, F., Ao, C. O., Gleisner, H., Ho, S.-P., Hunt, D., Schmidt, T., Foelsche, U., Kirchengast, G., Kuo, Y.-H., Lauritsen, K. B., Mannucci, A. J., Nielsen, J. K., Schreiner, W., Schwärz, M., Sokolovskiy, S., Syndergaard, S., and Wickert, J.: Consistency and structural uncertainty of multi-mission GPS radio occultation records, Atmos. Meas. Tech., 13, 2547–2575, https://doi.org/10.5194/amt-13-2547-2020, 2020.

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9, https://doi.org/10.1109/CVPR.2015.7298594, 2015.

Tenenbaum, J. B., Silva, V. D., and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 290, 2319–2323, https://doi.org/10.1126/science.290.5500.2319, 2000.

Trent, T., Schröder, M., and Remedios, J.: GEWEX Water Vapor Assessment: Validation of AIRS Tropospheric Humidity Profiles With Characterized Radiosonde Soundings, JGR Atmospheres, 124, 886–906, https://doi.org/10.1029/2018JD028930, 2019.

Trent, T., Siddans, R., Kerridge, B., Schröder, M., Scott, N. A., and Remedios, J.: Evaluation of tropospheric water vapour and temperature profiles retrieved from MetOp-A by the Infrared and Microwave Sounding scheme, Atmos. Meas. Tech., 16, 1503–1526, https://doi.org/10.5194/amt-16-1503-2023, 2023.

Von Engeln, A., Nedoluha, G., Kirchengast, G., and Bühler, S.: One-dimensional variational (1-D Var) retrieval of temperature, water vapor, and a reference pressure from radio occultation measurements: A sensitivity analysis, J. Geophys. Res., 108, 2002JD002908, https://doi.org/10.1029/2002JD002908, 2003.

Von Rohden, C., Sommer, M., Naebert, T., Motuz, V., and Dirksen, R. J.: Laboratory characterisation of the radiation temperature error of radiosondes and its application to the GRUAN data processing for the Vaisala RS41, Atmos. Meas. Tech., 15, 383–405, https://doi.org/10.5194/amt-15-383-2022, 2022.

Wang, C., Chen, D., Hao, L., Liu, X., Zeng, Y., Chen, J., and Zhang, G.: Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model, IEEE Access, 7, 146533–146541, https://doi.org/10.1109/ACCESS.2019.2946000, 2019.

https://doi.org/10.5194/egusphere-2026-2027
Preprint. Discussion started: 29 May 2026

1426 Wang, J., Cole, H. L., Carlson, D. J., Miller, E. R., Beierle, K., Paukkunen, A., and Laine, T. K.:
1427 Corrections of Humidity Measurement Errors from the Vaisala RS80 Radiosonde—
1428 Application to TOGA COARE Data, J. Atmos. Oceanic Technol., 19, 981–1002,
1429 https://doi.org/10.1175/1520-0426(2002)019%3C0981:COHMEF%3E2.0.CO;2, 2002.

1430 Wee, T.-K., Anthes, R. A., Hunt, D. C., Schreiner, W. S., and Kuo, Y.-H.: Atmospheric GNSS
1431 RO 1D-Var in Use at UCAR: Description and Validation, Remote Sensing, 14, 5614,
1432 https://doi.org/10.3390/rs14215614, 2022.

1433 Xiaoling Xia, Cui Xu, and Bing Nan: Inception-v3 for flower classification, in: 2017 2nd
1434 International Conference on Image, Vision and Computing (ICIVC), 783–787,
1435 https://doi.org/10.1109/ICIVC.2017.7984661, 2017.

1436 Xie, Q., Huang, K., Wang, D., Yang, L., Chen, J., Wu, Z., Li, D., and Liang, Z.: Intercomparison
1437 of GPS radiosonde soundings during the eastern tropical Indian Ocean experiment, Acta
1438 Oceanol. Sin., 33, 127–134, https://doi.org/10.1007/s13131-014-0422-9, 2014.

1439 Yoneyama, K., Zhang, C., and Long, C. N.: Tracking Pulses of the Madden–Julian Oscillation,
1440 Bull. Amer. Meteor. Soc., 94, 1871–1891, https://doi.org/10.1175/BAMS-D-12-00157.1,
1441 2013.

1442 Zhran, M., Mousa, A., Wang, Y., Hasher, F. F. B., and Jin, S.: Assessment of Commercial GNSS
1443 Radio Occultation Performance from PlanetiQ Mission, Remote Sensing, 16, 3339,
1444 https://doi.org/10.3390/rs16173339, 2024.