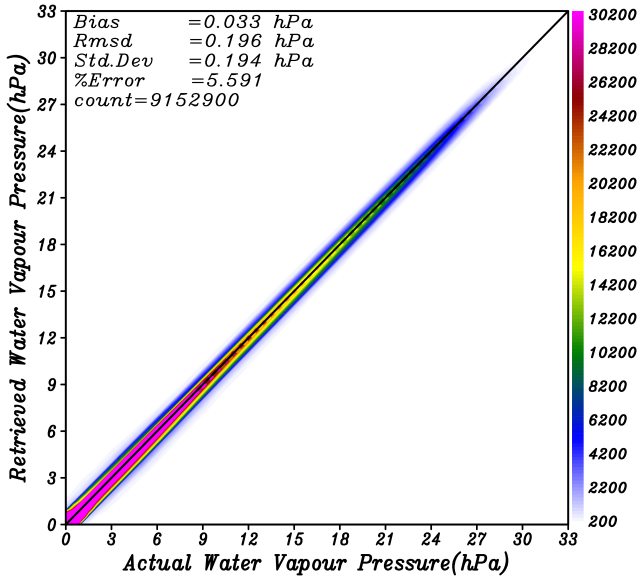
Deep Learning based COSMIC-2 Radio Occultation Atmospheric Profile Retrieval



Report submitted for completion of training
Satellite Meteorology and OceAnography Research and Training - **SMART**
Programme of Space Applications Centre

Submitted by
Soumil Hooda
(Registration Number RS00415)
B.E. EEE and M.Sc. Physics
BITS Pilani, Hyderabad

Under the guidance of
Dr. Satya Prakash Ojha
SCI/ENG-SF
PSD/BPSG/EPSA
Space Applications Centre (ISRO)

Space Applications Centre
Indian Space Research Organisation
Ahmedabad-380015, India
June 2022 - August 2022

Acknowledgement

I am extremely grateful to Space Applications Centre (ISRO), Ahmedabad for providing me with the opportunity to work on an amzing problem. I would like to thank Dr. S P Vyas (Head, Scientific Research and Training Division (SRTD)-RTCG/MISA) and the organisers of the Satellite Meteorlogy and OceAnography Research and Training programme. I would like to provide gratitude to Dr. Satya Prakash Ojha (PSD/BPSG/EPSA/SAC/ISRO) and Dr. Randhir Singh (ASD/AOSG/EPSA/SAC/ISRO) for the enriching learning experience working on a problem together. I am also indebt of the Research, Outreach and Training Coordination Group, Management and Information Systems Area and all staff members of Space Applications Center (ISRO), Ahmedabad for the continued support throughout. I would also like to thank Mr. Jai Gopal Singla (SIPG/SAC/ISRO) for providing with his computational resources. I would also like to thank Dr. Sarmistha Banik (HoD, Physics Department, BITS Pilani, Hyderabad), for enabling this opportunity for me.

Scientific Research and Training Division (SRTD)
Research, Outreach and Training Coordination Group (RTCG)
Management and Information Systems Area (MISA)

CERTIFICATE

This is to certify that **Mr. Soumil Hooda**, a student of B.E. (Electrical and Electronics) and M.Sc. (Physics) of BITS Pilani, Hyderabad, Telangana has completed a three month (1st June 2022 - 31st August 2022) project on "**Deep Learning based COSMIC-2 Radio Occultation Atmospheric Profile Retrieval**" under the supervision of Dr. Satya Prakash Ojha, Scientist-SF, PSD/BPSG/EPSA, Space Applications Centre (ISRO), Ahmedabad. The research work was carried out through the Scientific Research and Training Division (SRTD) of Space Applications Centre, Ahmedabad.

Abstract

Accurate and reliable data on tropospheric temperature and water vapour profiles are crucial for studies of weather and climate. Among the sensors used to assist tropospheric observations, the Global Navigation Satellite System (GNSS) Radio Occultation (RO) technology stands out because it provides precise and excellent meteorological profiles. RO, a form of active limb sounding used by the GNSS, involves GNSS satellites transmitting signals that are then picked up by a GNSS receiver on a low-earth orbiter (LEO) satellite as they pass through the Earth's atmosphere. The atmospheric refractivity profile is accurately produced by GNSS RO. Common techniques (like variational) to drive atmospheric profiles of temperature, water vapour pressure, and pressure require knowledge of the atmospheric condition beforehand (e.g. water vapour pressure, temperature and pressure). By training two Artificial Neural Network (ANN) models using simulated data, we hope to eradicate this dependency in this study. The wetPf2 dataset derived from COSMIC-2 (Formosa Satellite-7/Constellation Observing System for Meteorology, Ionosphere and Climate-2 mission) is used to extract the atmospheric temperature, pressure, and water vapour profiles. A more precise and recently published three-term refractivity formulation is employed to model the corresponding refractivity using the thermodynamic profiles from wetPf2. The inputs to our ANN model are the simulated refractivity, latitude, longitude, height, and month, and the target is the thermodynamic profiles. The model is trained and tested using data from the year 2020, and the results are encouraging. Temperature, pressure, and water vapour pressure have roots mean square errors (RMSE) that are, respectively, 1.28 K, 1.26 hPa, and 0.19 hPa when averaged vertically. Additionally, a completely independent data set obtained in 2021 is used to evaluate the model, and while the retrieval errors are slightly larger, they are still within accepted limits. The vertically averaged root mean square error (RMSE) for the independent data is 1.82 hPa for pressure, 1.64 K for temperature, and 0.24 hPa for water vapour pressure. The retrieval errors for temperature and pressure in this study are comparable to those achieved by the earlier studies, while the retrieval errors for water vapour are substantially lower in this study.

Contents

List of Figures

1 Introduction

Earth's atmosphere is highly nonlinear in nature, and continuous monitoring applications such as weather and climate modelling require reliable tropospheric thermodynamic profiling (i.e. temperature, moisture and pressure). Radiosonde observations are the most reliable measurements for this purpose. However, these observations are very sparse and not available over the oceans. Moreover, they have manual dependency too. Sounders onboard satellites remove this dependency and provide global observations at high spatial and temporal resolutions (*Xu et al.*, 2019).

Infrared (IR) sounders can provide thermodynamic profiles in clear-sky conditions, but they have limitations during cloudy and precipitating environments. Microwave (MW) sounders improve on the IR sounders, but they are also sensitive to heavy precipitation which limits their use in extreme weather events. On the contrary, the radio occultation (RO) measurements from the Global Navigation Satellite System (GNSS) have the advantage of monitoring Earth's atmosphere in all weather conditions with high vertical resolution. RO data is minimally affected by aerosols, clouds, or precipitation, does not require calibration and is free from instrument drift and satellite-to-satellite bias (*Anthes*, 2011; *Anthes et al.*, 2000, 2003, 2008; *Chen et al.*, 2021; *Hajj et al.*, 2004; *Kuo et al.*, 2004; *Rocken et al.*, 1997; *Wickert et al.*, 2004). These unique features make RO data ideal for weather and climate studies (*Scherllin-Pirscher et al.*, 2021; *Steiner et al.*, 2011). Due to the success of COSMIC-1, US agencies and Taiwan decided to move forward with a follow-on Global Navigation Satellite System Radio Occultation (GNSS-RO) mission called FORMOSAT-7/COSMIC-2. The six COSMIC-2 satellites launched successfully on June 25, 2019, into low inclination orbits. COSMIC-2 can provide 5000 RO profiles over the tropics and subtropics daily (*Xu et al.*, 2019). RO techniques have been successfully assimilated into numerical weather prediction models which greatly improved forecast quality (*Rennie*, 2010).

An RO occurs when a receiver in low-Earth orbit (LEO) views a GNSS satellite as it sets or rises behind the Earth's atmosphere. The measured signal phase and amplitude are analyzed to derive atmospheric refractivity using a limb-sounding geometry. The atmospheric state variables such as pressure, water vapour pressure and temperature (i.e. P, e and T) can be derived from the inversion of atmospheric refractivity observations. RO observations in weather and climate modelling are used in two manners. In the first approach, the refractivity or bending angle profiles are directly assimilated in NWP models. In the second approach, thermodynamic profiles are extracted from these refractivity profiles using optimum theory-based variational methods such as the 1Dvar approach. This approach requires *a priori* knowledge of the atmospheric state variables (P, e and T) which are obtained from either climatology or NWP model forecasts, the latter being the more widely used. There is redundancy and dependency in this method, as retrieval requires external data sources which introduces lag for real-time application. The use of *a priori* can be avoided by developing an empirical relationship between the atmospheric state variables and the RO refractivity observations using a collection of a large dataset (*Leroy et al.*, 2012).

Machine learning (ML) has seen a recent surge in earth-atmosphere applications due to its immense prowess. ML consists of algorithms that learn and establish relationships among the dependent and independent variables from a collection of datasets. Training an ML algorithm is a one-time process but is computationally expensive. As high-speed computers are readily available nowadays, the application of ML algorithms has become ubiquitous. Deep learning is a subfield of ML based on neural networks, which are highly flexible differentiable functions that can be fit to data to help learn factors of variation that explain the observed data. Neural networks consist of a hierarchy of layers that contain nodes performing weighted non-linear transformations of their inputs, through a series of hidden layers, to the desired output. Deep learning solves the central problem in representation learning by introducing representations expressed in terms of other, simpler representations. Deep learning allows the computer to build complex concepts out of simpler concepts (*Goodfellow et al.*, 2016).

This study aims to develop a deep learning-based algorithm to reduce continuous dependency on meteorological information from external sources while deriving thermodynamic profiles from RO refractivity observations. The structure of the document is as follows, section 2 discusses the GNSS RO-based technique, section 3 discusses the data and methods used, section 4 discusses the results and section 5 provides the conclusion.

2 GNSS RO Technique

The basic principle of the GNSS RO technique is the interaction between electromagnetic radiation and a medium i.e., in this case the terrestrial atmosphere using a limb-sounding geometry. As the signals propagate from the GNSS satellite to the receiver on the LEO satellite, they are delayed (due to bending as well as retardation) by the ionosphere and the neutral atmosphere (*Davis et al.*, 1985). The ionospheric delay is frequency dependent which can be cancelled using two different frequencies to obtain ionospheric free measurements. The tropospheric neutral atmosphere delay is directly proportional to the refractive index or refractivity and can be expressed as a function of atmospheric temperature, pressure and water vapour pressure.

Figure 1 schematically illustrates the geometry of the GNSS-LEO occultation of the Earth's atmosphere in the occultation plane. The overall effect of the atmosphere can be characterized by a total bending angle (α) as a function of the impact parameter (a). The impact parameter is defined, assuming spherical symmetry, as the perpendicular distance between the centre of the Earth (more precisely speaking, the centre of local curvature at the perigee of the occultation ray) and the ray asymptote at the GNSS or LEO satellite. The variations of α with a or r (radius to ray tangent point) depend primary on the vertical profile of the atmospheric refractive index.

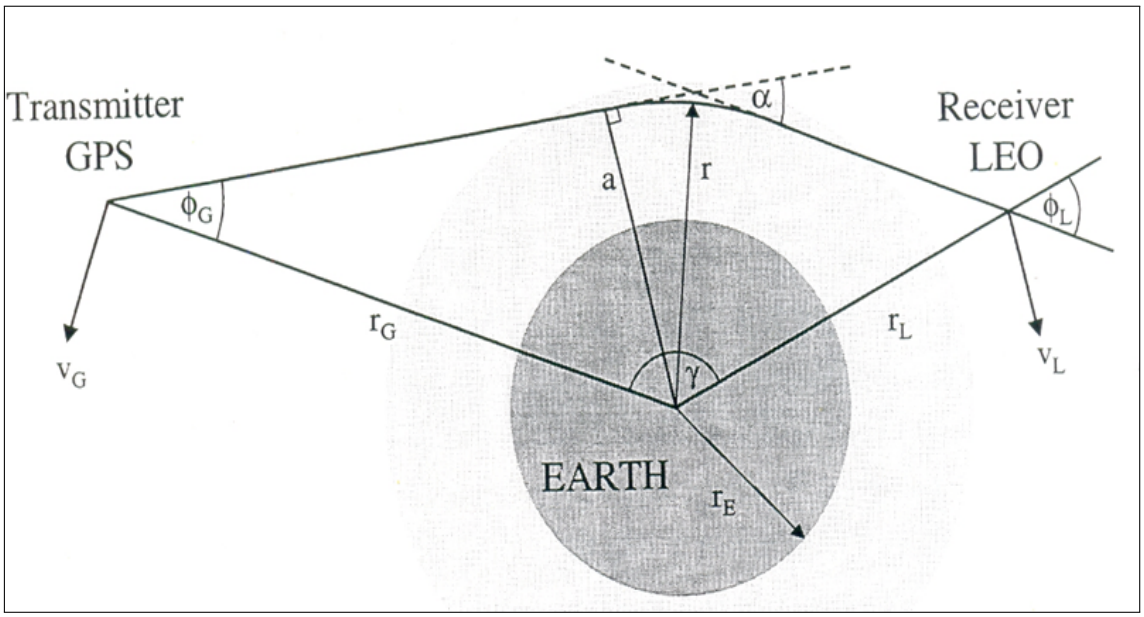


Figure 1: Conceptual sketch showing the geometry of the GNSS-LEO occultation of the Earth's atmosphere in the occultation plane. The GNSS satellite (radius vector r_G, velocity v_G) emit rays at the zenith angle ϕ_G, which are received by the LEO satellite (radius vector r_L, velocity v_L) at the zenith angle ϕ_L. The total bending angle is denoted by α and radius at the tangent point r.

Refractivity profile can be retrieved from measurements of $\alpha(a)$ during occultations assuming local spherical symmetry. In the case of the radio frequency domain, accurate angular measurements are difficult to achieve since they would require a very large dimension of the receiving antenna (*Gorbunov and Sokolovkiy*, 1993). Instead, it is possible to measure the phase (or Doppler frequency) shift of the signal with extremely large accuracy as a function of time. These measured phase paths include both the geometric paths from the transmitter to the receiver and the optical path due to atmospheric bending. By derivation of the signal phase with respect to time, the Doppler shifted frequency is calculated. The residual Doppler shift due to atmospheric bending can be combined with satellite position and velocity knowledge to estimate the bending angle and impact parameter. This gives a set of $\alpha(a)$ which has to be inverted into a vertical profile of refractivity close to the tangent point. Assuming that the set of tangent points are not significantly separated in the horizontal and that the atmosphere can be approximated locally as spherically symmetric, there is a direct solution to the inverse problem through an Abel transform.

2.1 Atmospheric bending and the inverse refractive problem

Phinney and Anderson (1968) deduced inversion relation for the determination of refractive index profiles from either the phase shift or the angular path. With the geometric optics assumption for the propagation of electromagnetic signals, the path of a ray through a region of varying refractive index is determined globally by Fermat's principle of least time. Thus, the trajectory of the electromagnetic ray connecting two points (P1, P2) can

be obtained through setting.

$$\int_{p1}^{p2} nds = Min \tag{1}$$

It is evident that it is not possible to retrieve the three dimensional variation of the refractive field in the vicinity of the ray perigee from the one dimensional measurements along the trajectory of the receiving satellite. A simplification of the problem has to be made in which it can be considered as the horizontal displacement of the transmitting ray perigee crossing the atmosphere to be much smaller than the characteristic horizontal scale of meteorological fields (*Gorbunov and Sokolovskiy*, 1993). This can make the approximation that the ray path is lying in a plane. Furthermore, it was shown by *Lusignan et al.* (1969) that the horizontal component of the refractive index gradient normal to the ray path may be ignored since it causes negligible small errors to angular and Doppler data. The variation of n along the limb path is dominated by the vertical gradient of the refractive index field. This gradient of n is directed radially by making the assumption that the local refractive index field is distributed spherically symmetrical in the vicinity of the ray perigee. For the ray changing its direction when passing through the refractive index field, basically Snell's law (*Born and Wolf*, 1980) may be applied

$$nsin\phi = const \tag{2}$$

where ϕ denotes the angle between the gradient of refraction and the ray path. If the ray path lies in one plane, then the change in direction can be expressed in polar coordinates. The differentiation of Snell's law $dn \sin\phi + n\cos\phi d\phi = 0$ leads to the equation for the incremental bending of the ray, which can be rearranged as:

$$d\alpha = d\phi_{snell} = -\tan\phi \frac{dn}{n} \tag{3}$$

The angle of incidence ϕ is related to the polar coordinate in that

$$\tan\phi = \frac{rd\theta}{dr} \tag{4}$$

Since Snell's law is defined for planner interfaces the total change in the incremental angle $d\phi$ has two contributions, the one described by Snell's law $(d\phi_{snell})$ plus an additional contribution due to the spherically symmetric interfaces, and thus the change of the incident angle with respect to the change of the polar angle $\frac{\partial\phi}{\partial\theta}\, d\theta$ (*Kursinski*, 1997). Using the geometrical relation $\theta + \phi - \alpha = {}^{\pi}/_{2}$ with its differential form

$$d\theta + d\phi - d\alpha = 0 \tag{5}$$

And substituting for $d\theta$ and $d\phi$ by using Eq. 3 and Eq. 4gives

$$\frac{dn}{n} + \frac{dr}{r} + \frac{d\phi}{\tan\phi} = 0 \tag{6}$$

Integration of Eq. 6 leads to the formula of Bouguers ()*Born and Wolf*, 1980) which represents the equation for the ray trajectory in a spherically symmetrical refractive field and replaces Snell's law for a plane stratified medium:

$$nr\sin\phi = cont = a \tag{7}$$

The quantity $r \sin\phi$ represents the perpendicular distance from the origin to the ray path tangent. As r goes to infinity, n goes to unity, so that the constant term must equal the impact parameter (a). At the point of closet approach, denoted by the ray tangent radius $r_o \sin\phi = 1$. So that a=n (r_0) r_0.

In order to derive the total refractive bending angle, the Bouguer equation is rearranged to get $\sin\phi = {}^{a}/_{nr}$ which substitute for tanϕ in Eq.6. Integration along the ray path then gives the integral equation

$$\alpha(a) = 2a \int_{r=r_0}^{r=\infty} \frac{1}{\sqrt{n^2 r^2 - a^2}} \frac{d\ln(n)}{dr} dr \tag{8}$$

This so called Abelian integral equation addresses the forward calculation of α*(a)* from a given *n(a).* The equation has to be inverted employing standard mathematical techniques, i.e., using the Abelian transformation to retrieve the refractive index as a function of the tangent

$$n(r_0) = \exp\left[\frac{1}{\pi} \int_{a=\infty}^{a=a_0} \frac{\alpha(a)}{\sqrt{a^2 - a_0^2}} da\right] \tag{9}$$

Here the impact parameter '*a*' corresponds to the radius at the tangent point $r_0 = n(r)/a_0$ for the ray of closest approach. This equation describes a set of rays in that *da* is being integrated from a_0 to infinity using the measurements of α(a). In practice a discrete approach with a numerical solution is performed. A description of the derivation of the Abelian integral equation and its solution with respect to n(r) leading to the Abelian integral transform pair was given by *Fjeldbo et al.* (1971).

2.2 Derivation of the bending angle from phase measurements

In the radio frequency domain, the effect of atmospheric bending can be measured as phase (ϕ) delays or excess phase ($\nabla\phi$) for the GNSS signal paths due to the slowing and bending of both the atmosphere and ionosphere. From these observables, computed excess phase paths (L) and Doppler shifts (dL/dt) can be combined with satellite position and velocity vectors measurements to determine the atmospheric bending angle. In the following, the standard methodology for the derivation of bending angles from phase measurements is described.

The excess phase paths (L_1 and L_2) for the dual frequency GNSS signals obey the following relation:

$$L_i = \Delta\phi_i \lambda_i = \int n(s_i)\, ds_i - R_{geom} \tag{10}$$

where R_{geom} denotes the geometrical straight line distance (vacuum path length) between the receiver and transmitter, and λ is the wavelength of the probing signal. The rays of the two GNSS frequencies f_1 and f_2 are influenced differently by the ionosphere because of its dispersive properties and travel along different paths. Since the main interest is refractive index of the neutral atmosphere, the contribution of the ionospheric bending in the excess phase path has to be removed by applying an ionospheric correction method.

For the standard ionospheric correction method, a linear combination of the excess phase paths of the two signals with harmonically related frequencies is performed:

$$L_c = \frac{f_1^2 L_1 - f_1^2 L_2}{f_1^2 - f_2^2} \tag{11}$$

This is a model correction of the ionospheric optical path length, removing contributions up to the first order. The excess phase path L_c now consists only of the delay due to the neutral atmospheric bending plus a small ionospheric residual, i.e., the higher order terms. The derivation of L with respect to time (dL_c/dt) then gives the atmospheric Doppler shift f_d on the carrier frequency f which can be related to kinematic properties such that

$$\frac{f_d}{f} c = \frac{dL_c}{dt} = v_G^c \sin\phi_G - v_L^c \sin\phi_L + v_G^c \cos\phi_G + v_L^c \cos\phi_L \tag{12}$$

The GNSS satellite emits electromagnetic waves at zenith angle ϕ_G and the signal is received at the LEO satellite at the zenith angle ϕ_L. With the assumption of local spherical symmetry and the use of Snell's law of the form $r_G \sin\phi_G = r_L \sin\phi_L$, where r_G and r_L are the geocentric distances of the satellites, n(r_G)=n(r_L)=1 is valid.

By expressing ϕ_G through Snell's law and combining it with Eq. 12 the ray zenith angle at the GNSS satellite and in analogues from the ray zenith angle at the LEO satellite is derived with the known position and velocities of the satellites. Finally the atmospheric bending angle and the ray parameters can be calculated with the knowledge of the angle γ from the satellite positions:

$$\alpha(a) = \phi_G + \phi_L + \gamma - \pi \tag{13}$$

$$a = r_G \sin\phi_G = r_L \sin\phi \tag{14}$$

2.3 Atmospheric property derivation from the refractive index

The total ionospheric and atmospheric refractivity (N) can be approximated for waves in the radio frequency domain (<10 GHz) by:

$$N = 77.6890 \frac{P}{T} - 6.3938 \frac{e}{T} + 3.75463 \times 10^5 \frac{e}{T^2} + 4.04 \times 10^7 \frac{n_e}{f^2} \tag{15}$$

where P denotes the pressure (mb), T is the atmospheric temperature in (K), e is water vapour partial pressure in (mb), n_e is the electron number density in (m^{-3}), and f denotes the transmitter frequency in (Hz). Three main sources, the dry neutral atmosphere, water vapour, and free electrons in the ionosphere contribute to the total refractivity in Eq. 15. Further, scattering from large rain drops is neglected in Eq. 15, as it is important only under very extreme conditions.

The first term in Eq. 15 is due to the dry part of the atmosphere. This dry term is caused by the polarizability of molecules in the atmosphere which means that the incident electric field induces an electric dipole in the molecule. The dry refractivity term is proportional to the molecular number density and dominates in altitudes from the surface up to about

50 km. The second and third terms, represent the wet part of the refractivity which is due to the large permanent dipole moment of water vapour. Throughout most of the troposphere, the dipole component of the refractivity is about 20 times larger than the non-dipole component (*Bevis et al.*, 1992). The moist terms have only a substantial impact on the magnitude of N in the lower troposphere, i.e., below 5 km, above altitudes of 7 to 10 km, the contribution to N from the water vapour terms is less than 2%.

The fourth term, the ionospheric term, is mainly due to free electrons in the ionosphere and becomes important above about 50 km. This term is proportional to the free electron number density and is an approximation of the Appleton-Hartee equation, which represents the full dispersion relation for ionised plasma. The dispersive nature of the ionosphere causes the ionospheric term to depend on the frequency. This term is estimated and removed to first order combination of the two frequency GNSS measurements to derive the neutral atmospheric refractivity. Removing the ionospheric term leaves the dry and the moist term which represents then the equation for neutral refractivity. This equation is valid for the radio frequency domain and provides an accuracy of approximation 0.5 % in N (*Smith and Weintraub*, 1953). *Thayer* (1974) improved the equation for the radio refractive index of air by taking into account the compressibility factor due to non-ideal gas behaviour.

It is clear from Eq. 15 that, after accounting for the ionospheric component, it contains three unknowns (P, T and e), which makes it challenging to retrieve thermodynamic profiles from known refractivity measurements. In the upper atmosphere where moisture content is low or negligible the hydrostatic equation and the equation of state can be used to retrieve the temperature and pressure from refractivity measurement by neglecting the moist terms in Eq. 15. The profiles are known as dry retrievals, and are very accurate in the upper troposphere and above. However, when the water vapour content is significant, which is the case in the warmer regions of the troposphere, particularly in tropical regions where the abundance is greatest, the separate contributions to refractivity by the dry and moist terms cannot be distinguished uniquely through radio occultation measurements. This introduces an ambiguity into the profiles of water vapour pressure, pressure and temperature. In this case, the use of a priori data from meteorological analyses is necessary and is a constraint to the retrieval problem.

3 Data and Methods

3.1 Data used

In the present study the wetPf2 level 2 product which is available at the COSMIC Data Analysis and Archive Center (CDAAC) website (https://data.cosmic.ucar.edu/gnss-ro/cosmic2/nrt/level2/) was used. The product contains meteorological profiles of temperature, pressure and water vapour pressure with a vertical resolution of 50 m from surface to 30 km height and a vertical resolution of 200 m from 30 km to 60 km height, derived from COSMIC-2 measured refractivity using the 1DVar approach with European Centre for Medium-Range Weather Forecasts (ECMWF) 6 hour forecasts as the prior (or first

guess) information. These thermodynamic profiles along with their respective spatial and temporal metadata were collected. These profiles were extracted for the Indian subcontinent (Latitude: 45°E to 110°E, Longitude: 0°N to 45°N) in the January 2020 - December 2021 time period. Note that the CDAAC profiles incorporate refractivity as well, computed using the 2-term refractivity formulation, and is compatible with these thermodynamic profiles (Equation 10 of Smith and Weintraub, 1953). However, recent literature showed that the 3-term refractivity formulation (*Singh et al.*, 2021) performs better than the 2-term refractivity formulation therefore we have recomputed refractivity from the CDAAC thermodynamic profiles using 3-term refractivity formulation. These refractivity profiles along with the spatial and temporal metadata such as latitude-longitude, height and month-day-hour formed the independent variables. Thermodynamic profiles, temperature, pressure and water vapour pressure were the dependent variables. Note that, in this study, we have limited our analysis from surface to 15 km height only because the water vapour is negligible above 15 km altitude. Thus above 15 km the dry retrievals can be accurately performed using COSMIC-2 observed refractivity. Further note that although the original data was accessible every 50 m in the vertical, it was decided to take the data at a 100 m vertical resolution to develop and test the model. Also note, that the study excludes the highly mountainuous regions (regions higher than 1 km) as the quality of data there is unknown.

In order to choose the most ideal number of independent parameters for the model development, we performed a correlation study between the dependent and independent variables. Figure 2 shows the correlation between these independent and dependent variables. The association between refractivity and moisture (e.g., water vapour pressure) is significant in the lower troposphere, whereas in the upper troposphere, the correlation between refractivity and pressure (P) is stronger than the correlation with water vapour pressure. The association between refractivity and temperature is much weaker than it is for pressure (P) and water vapour pressure (e). Overall, it can be observed that the association between refractivity and total pressure increases as height increases, whereas the correlation between refractivity and water vapour pressure drops. As would be expected, temperature, pressure, and water vapour pressure all negatively correlate with height, with pressure and height having a much stronger correlation.

Except for higher level temperature, which has a positive link with latitude, the majority of variables have negative correlations with latitude. Furthermore, the relationship between pressure and latitude is weaker in the lower troposphere and stronger in the upper troposphere. As one moves from the tropics to higher latitudes, both water vapour pressure and temperature decrease, so it makes sense that there is a negative correlation between latitude and water vapour pressure. The relationship between pressure and latitude, however, is not simple and depends on the dynamics and thermodynamics of various atmospheric processes in a very complex way. Also, longitude and the month have a positive correlation with water vapour pressure across the entire height. While less significant than the association between water vapour pressure and longitude and month, the correlation between temperature and longitude appears to be strongly positive in the middle troposphere and negative in the upper troposphere. The total pressure has a weak positive association with longitude and months, in contrast to water vapour pressure and tem-

perature, and is primarily observed in the upper troposphere. Day and hour correlations with all of the dependent variables are quite weak when compared to other independent variables. Therefore the independent variables day and hour were subsequently dropped during feature selection owing to their low correlation. Therefore, our model independent feature set after correlation analysis consisted of refractivity, latitude, longitude, height and month, which is unlike the feature set used by *Lasota* (2021), where hour was considered to be a useful feature and longitude was not while setting up the training procedure for retrieval of profiles at the global scale. We attribute this difference to the variability observed over the Indian subcontinent.

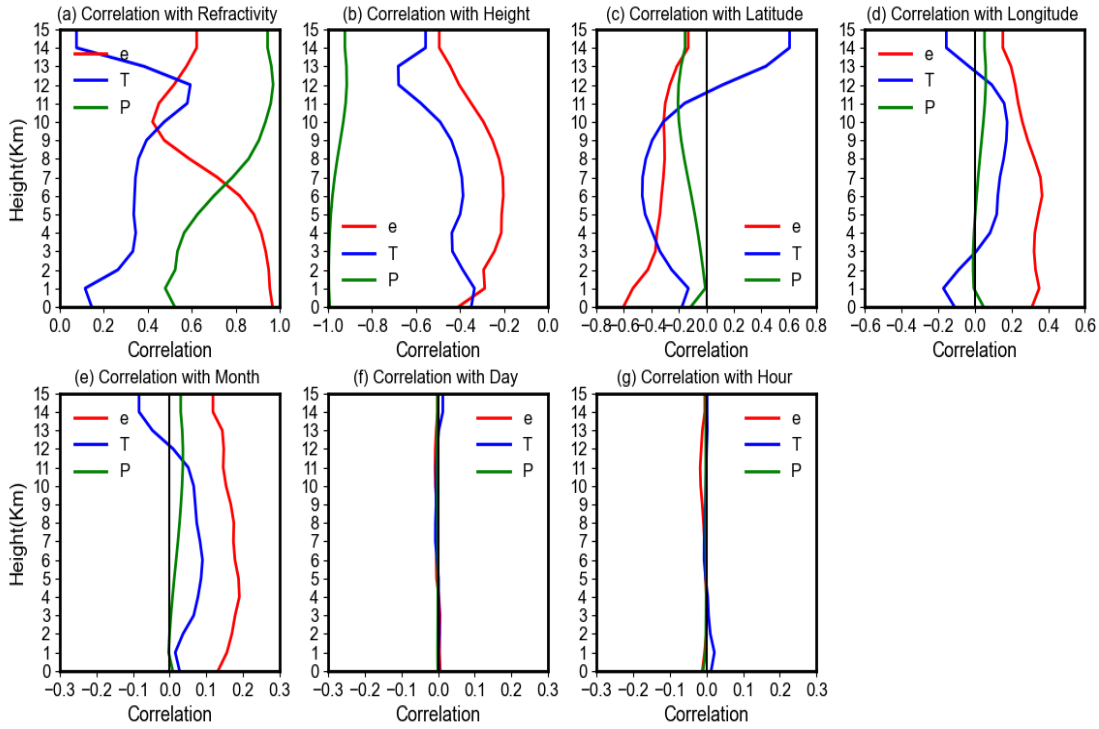


Figure 2: Correlation between dependent and independent variables using the data available within the chosen region (0°N-45°N, 45°E-110°E). The correlation is carried out for varous heights ranging from the surface to 15 km.

The model was trained using data from the year 2020. Thirty per cent of the data from the year 2020 was utilized for testing (Test20) and seventy per cent of the data from the year 2020 was used for training (Train20) the model. Further, along with the 2020 test data set, the entire 2021 data set (Test21) was employed as an entirely independent test to check the efficiency of the developed model. The climatological variability of the three targets has been plotted in Figure 3, 4 and 5. Figure 6 denotes the sample count for different heights used in the study, and Figure 7 shows the same spatially. Since COSMIC-2 is in a low-inclined orbit, as was already mentioned, the amount of data counts appears to be higher around the equator and decreases as we approach higher latitudes. Nevertheless, there are still enough data counts accessible from all the locations, with a noteworthy decrease in the number of counts over a few isolated areas in the domain's northern region.

The variability of the derived refractivity observations has been plotted in Figure 8. For the minimum, mean and maximum values, respectively, the water vapour pressure at the surface ranges between 0.154 hPa, 20.403 hPa and 34.656 hPa. The mean values range between 20.403 hPa at the surface and 0.002 hPa at the height of 15 km, with standard deviations of roughly 6.106 hPa at the surface and 0.0015 hPa at the height of 15 km. The minimum, mean and maximum values of the surface temperature are 252 K, 295 K and 322 K, respectively. The standard deviation spans from 2.6 K to 5 K, while the mean values vary from 295 K at the surface to roughly 207 K at 15 km. The surface pressures are 920 hPa, 995 hPa and 1036 hPa for the minimum, mean and maximum, respectively. The mean values range from 945 hPa at the surface to approximately 145 hPa at the height of 15 km, with a standard deviation that varies from 13 hPa at the surface to almost 2 hPa at the height of 15 km. Overall, it appears that both training and testing data sets have similar distributions of all the data. For the minimum, mean and maximum values, respectively, the refractivity at the surface ranges between 235, 335 and 400. The mean values range from 335 at the surface to around 55 at the height of 15 km, with standard deviations of about 28 at the surface and 3 at the height of 15 km.

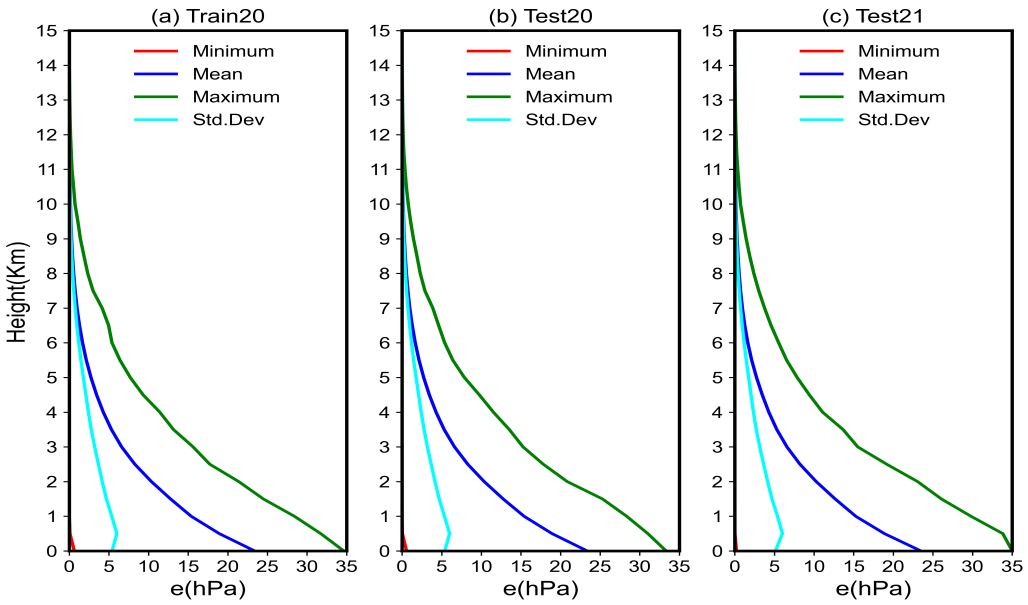


Figure 3: Statistical analysis of water vapour pressure data obtained from wetPf2. Leftmost panel indicates the training data from 2020, middle panel is the testing data from 2020, rightmost panel indicates the testing data from 2021.

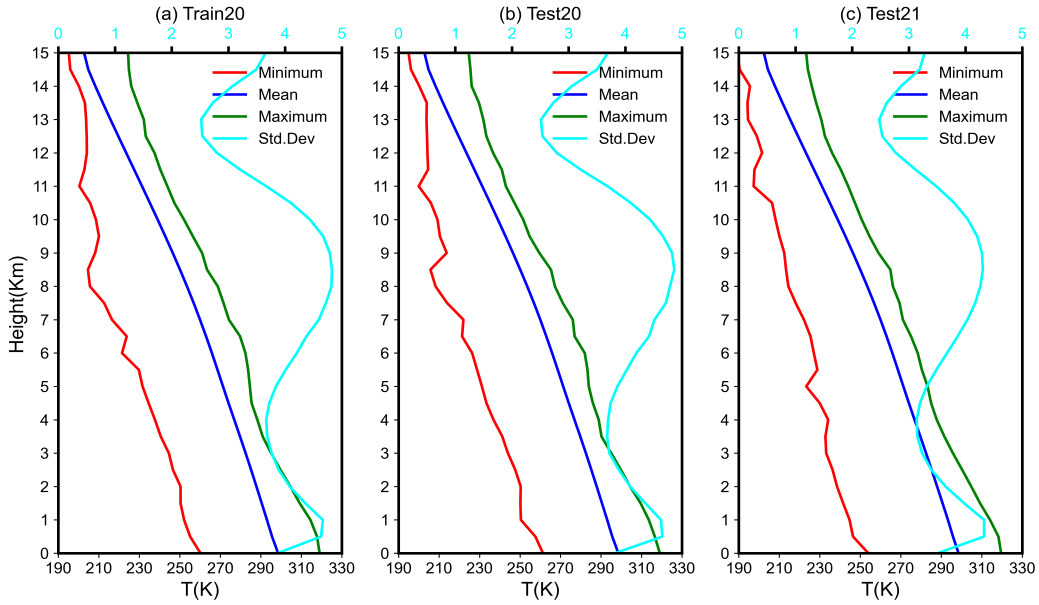


Figure 4: Statistical analysis of temperature data obtained from wetPf2. Leftmost panel indicates the training data from 2020, middle panel is the testing data from 2020, rightmost panel indicates the testing data from 2021. The top x axis is used to plot the standard deviation scale. The points accumulated over the entire domain are used to generate these statistics.

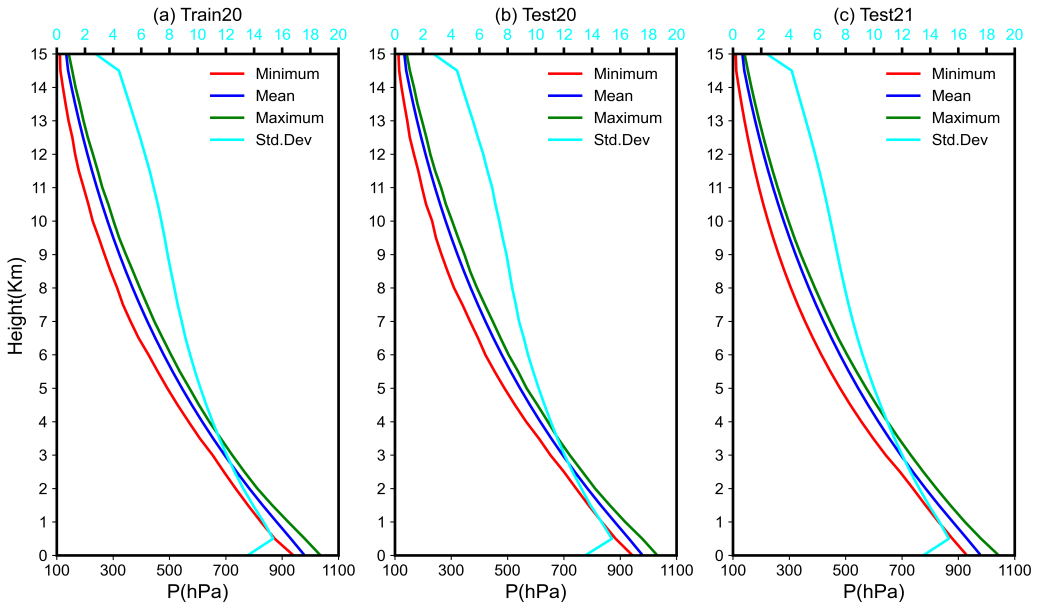


Figure 5: Statistical analysis of pressure data obtained from wetPf2. Leftmost panel indicates the training data from 2020, middle panel is the testing data from 2020, rightmost panel indicates the testing data from 2021. The top x axis is used to plot the standard deviation scale. The points accumulated over the entire domain are used to generate these statistics.

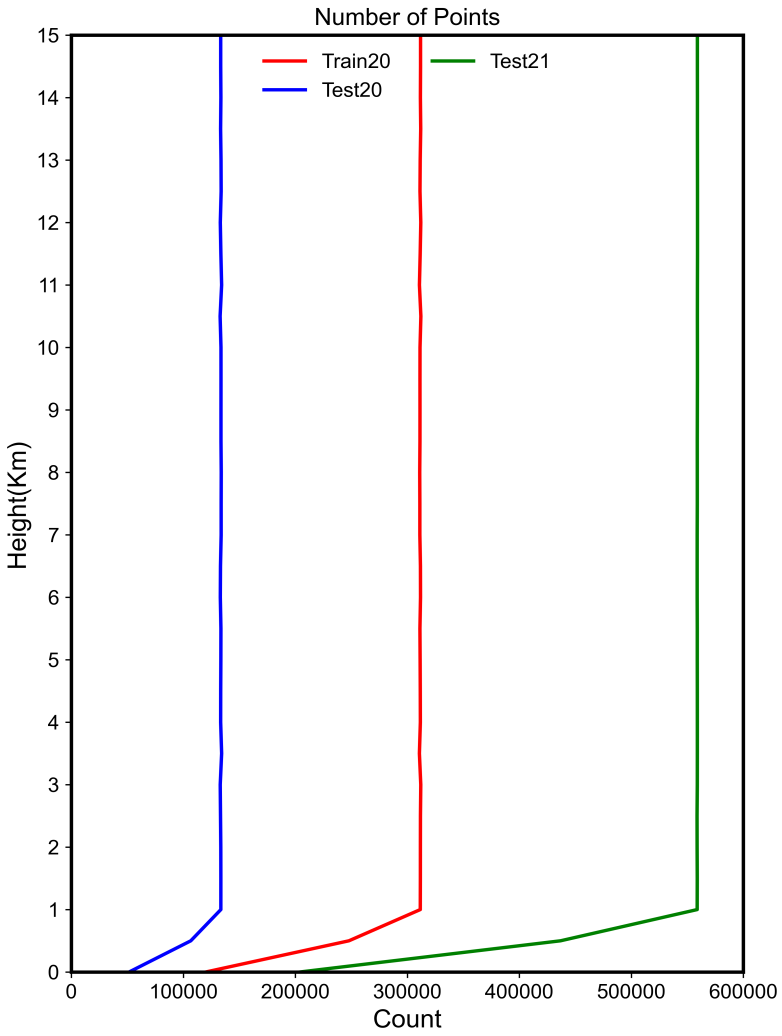


Figure 6: Total number of points in various height ranges employed for model training and testing. The points have been accumulated over the entire domain.

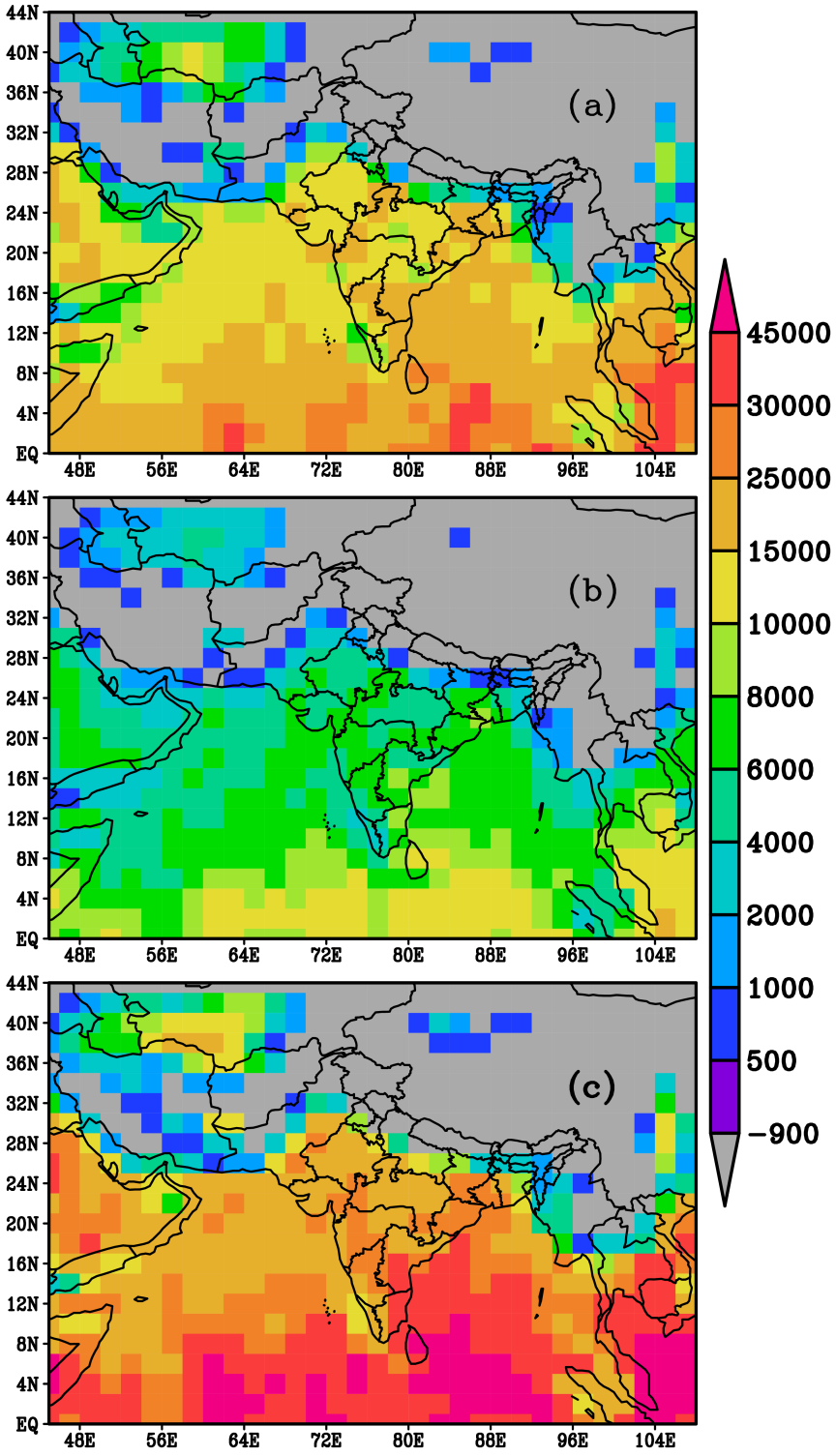


Figure 7: Spatial distribution of total number of points used for (a) training 2020 and (b) testing 2020 and (c) testing 2021 the model. The points have been accumulated over the entire height. The study's exclusion criteria eliminated areas with a color that is "gray", which are regions with surface elevation of more than 1000 m.

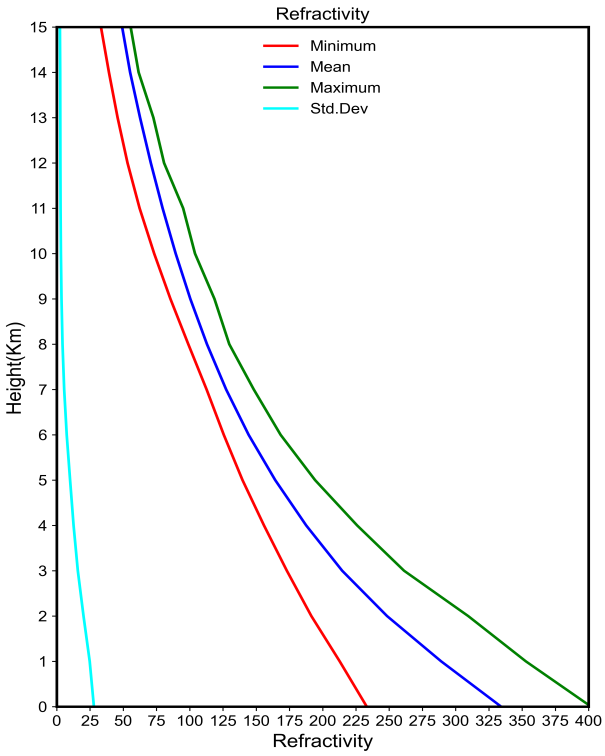


Figure 8: Statistics of the refractivity data derived using CDAAC thermodynamic profiles and the 3-term refractivity formulation. The data for years 2020 and 2021 over the selected domain are used to compile the statistics.

3.2 Methodology

For our study we focussed only on the ANN owing to its simplicity and effective results as shown by *Lasota* (2021). The ANN is an algorithm based on the works of *Hassoun* (1995) that are highly used in its feed-forward multilayer perceptron (MLP) configuration. An input, one or more hidden layers, and an output are the three types of layers that make up a typical model. Multiple basic processing units known as neurons or nodes connect these layers. The number of nodes in these layers and the number of hidden layers are arbitrary and are determined by the complexity and volume of available data. Each neuron in a given layer of an MLP network is connected to every neuron in the layer above it, and the strength of each connection is expressed by a numerical weight that is determined during training. The backpropagation algorithm is typically used in iterative learning processes. The neural network receives input data repeatedly, multiplies it by connection weights, adds it all up, and then sends the result to the following layer. Finally, based on the discrepancies between predicted and actual outputs, the model's error is estimated in the final layer. The calculated error is then fed back into the model and used to modify the connections weight, reducing the model's error and bringing the outputs closer to the desired results.

In the upper troposphere, where water vapour pressure values are close to zero, our investigation during training found that the water vapour pressure values occasionally fit toward the negative. Therefore, we employed a sigmoid activation function to manage

this circumstance. Water vapour pressure was normalised at input and re-normalised at the output with the maximum water vapour pressure value (34.6566 hPa) found in the training set, the same values were used to re-normalised for any independent datasets. To be able to use the sigmoid output activation function for water vapour pressure and linear output function for temperature-pressure, we build two different models that differ only in output activation as can be seen in Figure 9 and Figure 10.

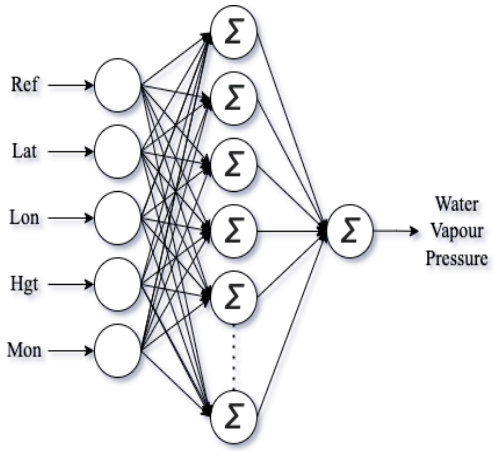


Figure 9: Model for retrieving water vapour pressure. Refractivity (Ref), height (Hgt), latitude (Lat), longitude (Lon) and month (Mon) are the input parameters, respectively. Note, that the inputs and outputs are for all 151 vertical levels (0 km to 15 km at a resolution of 100 m).

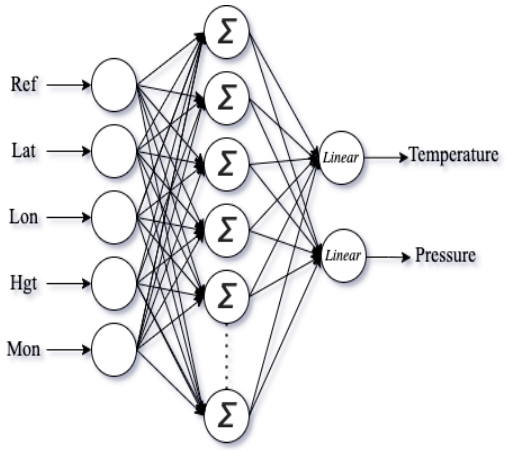


Figure 10: Model for retrieving pressure and temperature. Refractivity (Ref), height (Hgt), latitude (Lat), longitude (Lon) and month (Mon) are the input parameters, respectively. Note, that the inputs and outputs are for all 151 vertical levels (0 km to 15 km at a resolution of 100 m).

Differing to model parameters, which must be set beforehand before training, hyperparameters are configuration parameters that govern the learning process. Hyperparameters have a significant impact on how well a model performs, and choosing the best set of variables to include in a model is one of the most difficult tasks involved. Owing to parameters developed by *Lasota* (2021) we zero down and grid search around the following parameters,

- Hidden layers : 1, 2 and 3
- Hidden layer neurons: 50, 125, 250 and 500
- Dropout: 0 and 0.1
- Activation functions: rectified linear unit, leaky rectified linear unit, sigmoid and linear
- Epochs: 10, 50, 100, 500 and 1000
- Batch: 100, 500, 1000, 2000 and 10000

We use holdout validation as we want to emphasise on the independence of the testing data from across both years 2020 and 2021. The Adam optimisation algorithm first developed by *Kingma and Ba* (2014) was used to determine weights. We end up with optimal parameters at 1 hidden layer with 500 neurons, no dropout with sigmoid activation for the hidden layer at 500 batch size for 500 epoch, for both our models.

4 Results and Discussion

The scatter plot between the actual and retrieved water vapour pressure is shown in Figure 11 for the training and testing samples from 2020 and totally independent testing data from 2021. The root mean square error (RMSE) for the training (Train20), testing (Test20), and independent data set of the year 2021 (Test21) are in the range 0.196 hPa, 0.196 hPa and 0.24 hPa. The bias between the retrieved and actual water vapor pressure is very negligible. The retrieval error for the 2020 training and testing data sets is roughly 5.5%, whereas the error for the 2021 entirely independent testing data set is about 7%. As a result, the developed model displays excellent accuracy and negligible bias. In order to evaluate how retrieval errors vary vertically, Figure 12 shows the bias (such as mean differences), root mean square error, and standard deviation of differences between actual and retrieved water vapour pressure for training and testing data sets. For training and testing data for the year 2020, the lower troposphere exhibits a positive (i.e., model underestimate) bias of the order of 0.1 hPa. The RMSE is approximately 0.5 hPa close to the surface, and it decreases as height increases. As previously seen from the statistics of the data sets, the water vapour pressure decreases with height and, hence, error likewise decreases with height. Furthermore, because bias is low, the RMSE and standard deviation appear to be identical. When compared to the training and testing data from 2020, the model's retrieved water vapour pressure for the testing data from 2021 has a little bit greater bias, RMSE and standard deviations. The model overestimates the water vapour pressure by 0.2 hPa to 0.3 hPa (negative mean differences). But only by 0.1 hPa are the RMSE and standard deviation greater in 2021 than in 2020. Therefore, even with a completely independent data set, the overall model performs excellently.

For the training-testing split (2020) and independent (2021) datasets, Figure 13 displays the spatial distribution of the bias, RMSE, and standard deviation of the differences between actual and retrieved water vapour pressure. In the domain, bias ranges from 0 hPa to 0.02 hPa for the 2020 test and from -0.1 hPa to 0.2 hPa for the 2021 test. For the 2020 test and the 2021 test, respectively, the RMSE varies across the domain from 0.05 hPa to 0.45 hPa and from 0.12 hPa to 0.60 hPa. For the tests in 2020 and 2021, the standard deviation ranges from 0 hPa to 0.02 hPa and from -0.1 hPa to 0.2 hPa, respectively, across the domain. Furthermore, it is noteworthy to observe that RMSE and standard deviation increase as we move from lower to higher latitudes, with the equatorial region having the lowest errors. Our additional investigation (not provided) shows that the association between refractivity and water vapour pressure decreases as we move from the tropics, where the water vapour pressure is higher, to the higher latitudes, where the water vapour pressure is lower. As a result of the decreased association between refractivity and water

vapour pressure, larger errors are anticipated in the water vapour pressure retrievals over higher latitudes.

Overall, the vertically averaged errors in retrieving water vapour pressure are 0.196 hPa, 0.196 hPa, and 0.240 hPa for training (Train20), testing (Test20), and independent testing (Test21), respectively. The water vapour pressure retrieved errors in this study are significantly less than the error reported by *Lasota* (2021), which was 0.41 hPa. This could be for two reasons: first, this study only presents the retrieval accuracy for the Indian region, whereas *Lasota* (2021) reported accuracy for the global region (mainly global oceanic regions). Second, unlike *Lasota* (2021), who developed the model using COSMIC-2 refractivity and ERA5 thermodynamic profiles, we developed the model based on the simulated refractivity utilising wetPf2 data.

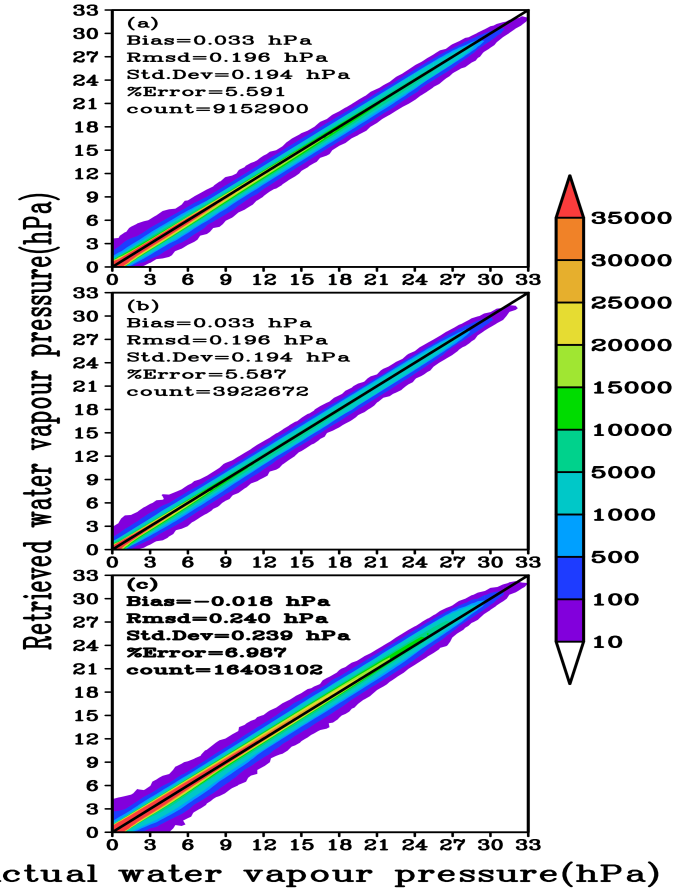


Figure 11: Scatter plot of actual versus retrieved water vapour pressure (a) for the training data set in 2020, (b) for testing data set in 2020, and (c) for completely independent data set fot the year 2021. Color bar represents the number of observations available for each water vapour pressure bin. This figure is based on data that has been accumulated over all vertical levels and the whole domain.

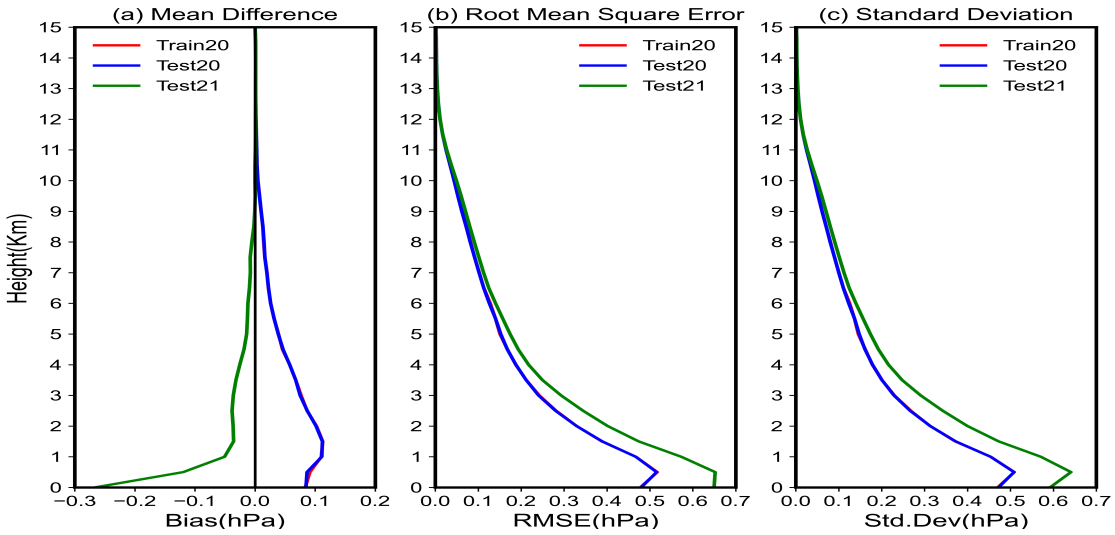


Figure 12: Vertical profiles (a) bias (i.e. mean differences), (b) root mean square error (RMSE), and (c) standard deviation (Std.Dev) of the difference between actual and model retrieved water vapour pressure, for training and testing data sets. The data acquired over the entire domain was used to create the figure.

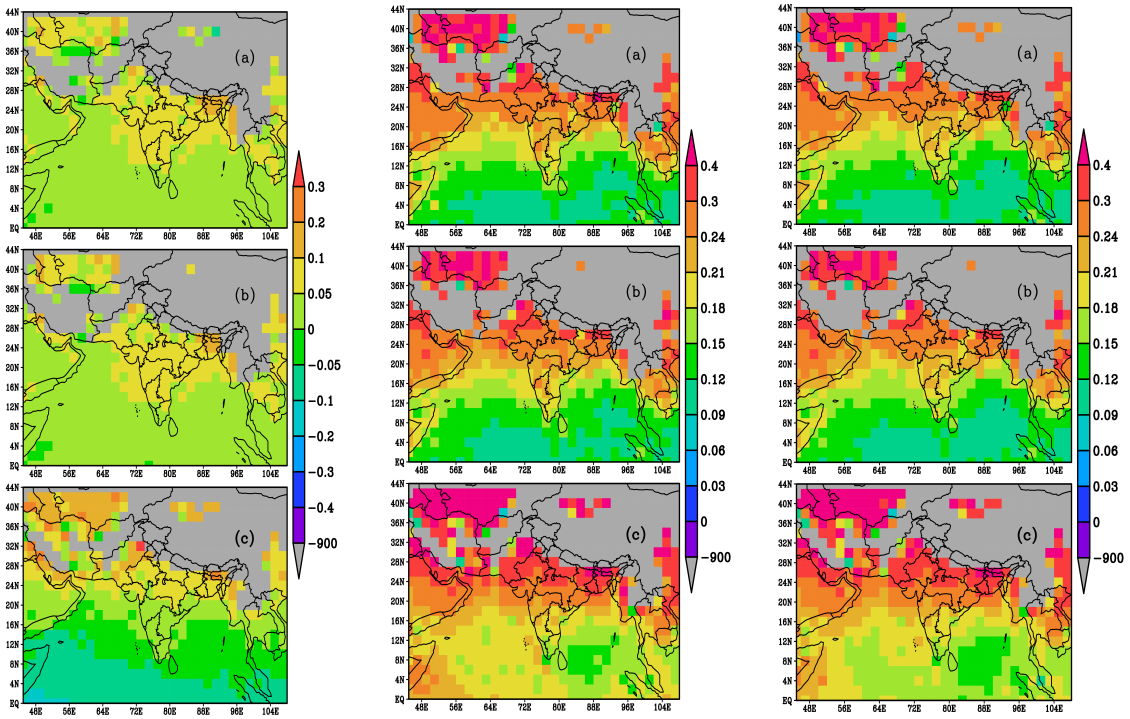


Figure 13: Spatial pattern of mean diffferences (i.e. bias), root mean square error (RMSE) and standard deviation (Std.Dev) of the differences between actual and model retrieved water vapour pressure. First column is for bias, second is for RMSE, and third is for standard deviation. Panels (a), (b) and (c) in each column represent the training data set for 2020, testing data set for 2020 and the testing data for 2021. The figure is produced using data collected across all heights.

The scatter plot between actual and retrieved temperature for the training (Train20) and testing (Test20 and Test21) datasets are displayed in Figure 14. For Train20, Test20, and Test21, the RMSE between actual and retrieved temperature is $1.3K,\ 1.3K, and\ 1.6K$, respectively. The bias (mean difference) between the actual and retrieved temperatures is roughly -0.2 K, but the bias for Train20 and Test20 is negligible. Overall, retrieval errors for training and testing (Test20 and Test21) datasets are comparable. The temperature retrieval errors are comparable to those published by *Lasota* (2021), who reported retrieval errors of the order of 1.4 K in training and around 1.7 K for the testing. The vertical profiles of the differences between actual and retrieved temperature for the training and testing datasets are shown in Figure 15 for bias, RMSE, and standard deviation. For the training and testing data sets from 2020, there is a very negligible bias between actual and retrieved temperatures over most levels; however, for testing data sets from 2021, the biases become more pronounced (as large as -0.8 K) as one move from the surface to the upper troposphere. The RMSE and standard deviation range from 1.2 K to 1.5 K and are almost constant with height for the training and testing datasets of 2020. The RMSE and standard deviation for the independent testing data set for 2021 range from 1.5 K to 1.8 K. These errors in the retrieved temperature profiles are less than the standard deviation in actual temperature, demonstrating the utility of the retrieved temperature information. Here it should be noted that the association between refractivity and temperature is much less than the correlation between refractivity and water vapour pressure, and so, a temperature retrieval error of less than 2 K is acceptable.

For the training-testing split (2020) and independent (2021) datasets, Figure 16 displays the spatial distribution of the bias, RMSE, and standard deviation of the differences between actual and retrieved temperature. In the domain, bias ranges from -0.1 K to 0.35 K for the 2020 test and from -0.8 K to 0.5 K for the 2021 test. For the 2020 test, RMSE ranges from 0.6 K to 3.0 K, and for the 2021 test, it ranges from 0.9 K to 3.6 K. For the tests in 2020 and 2021, the standard deviation ranges from 0.6 K to 3.0 K and from 0.9 K to 3.3 K, respectively. We notice similar patterns in the error, i.e., error rises with increasing latitudes, just like in the case of the water vapour pressure retrieval error.

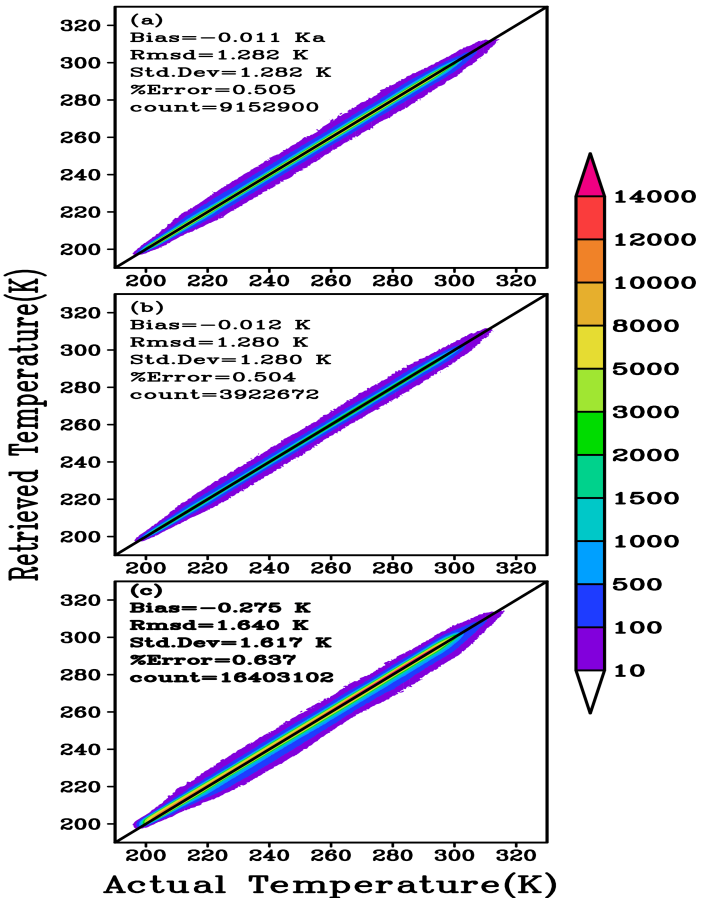


Figure 14: Scatter plot of actual versus retrieved temperature (a) for the training data set in 2020, (b) for testing data set in 2020, and (c) for completely independent data set fot the year 2021. Color bar represents the number of observations available for each temperature bin. This figure is based on data that has been accumulated over all vertical levels and the whole domain.

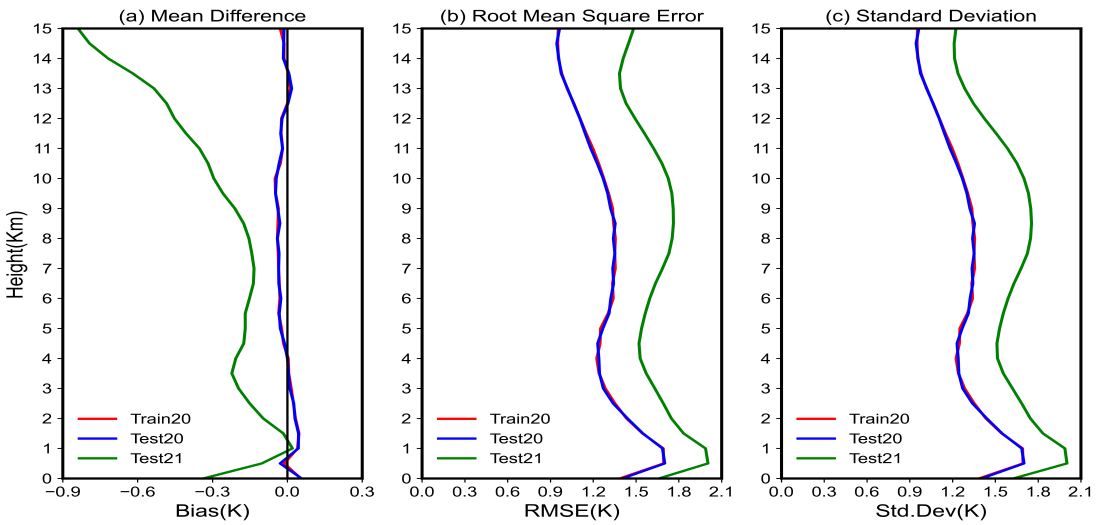


Figure 15: Vertical profiles (a) bias (i.e. mean differences), (b) root mean square error (RMSE), and (c) standard deviation (Std.Dev) of the difference between actual and model retrieved temperature, for training and testing data sets. The data acquired over the entire domain was used to create the figure.

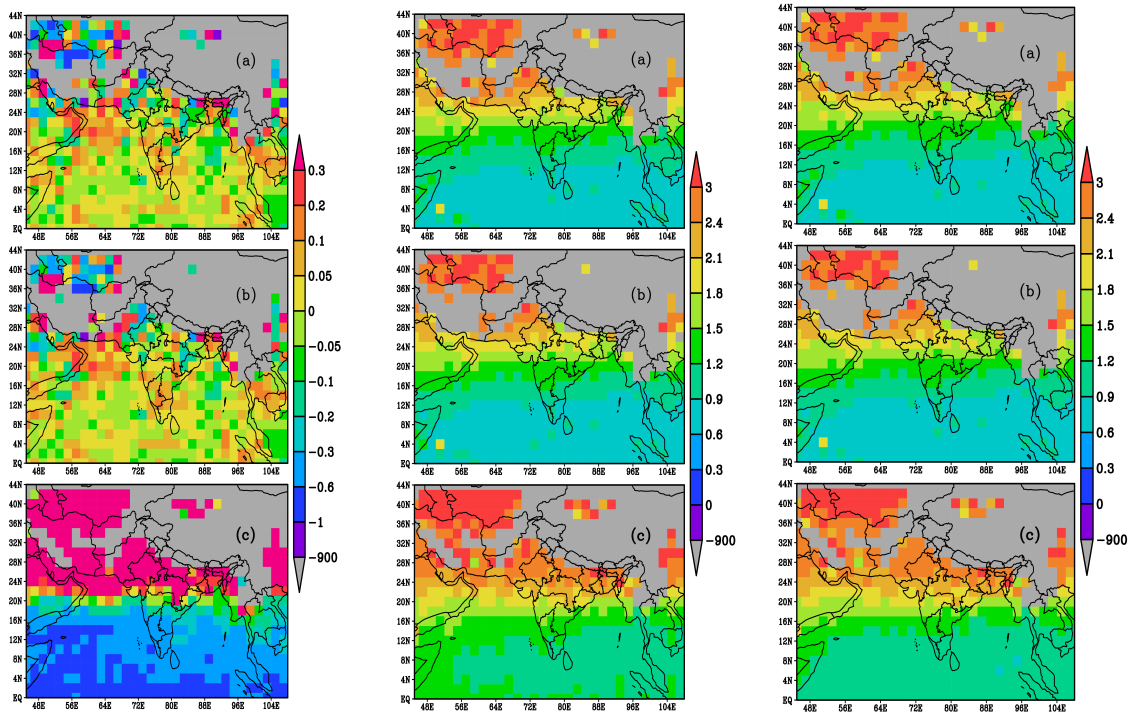


Figure 16: Spatial pattern of mean diffferences (i.e.bias), root mean square differences (RMSE) and standard deviation (Std.Dev) of the differences between actual and model retrieved temperature. First column is for bias, second is for RMSE, and third is for standard deviation. Panels (a), (b) and (c) in each column represent the training data set for 2020, testing data for 2020 and the testing data for 2021. The figure is produced using data collected across all heights.

The scatter plot between actual and retrieved pressure for the training-testing and independent datasets can be seen in Figure 17. For Train20, Test20, and Test21, the retrieved errors are 1.26 hPa, 1.26 hPa, and 1.82 hPa, respectively. The 2020 train and test datasets have very little bias, while the 2021 independent test datasets have a noticeable bias. The vertical profiles for bias, RMSE, and standard deviation are shown in Figure 18. The biases are minimal through the troposphere for both the training and testing datasets for the year 2020. However, with the 2021 independent sample, a sizable overestimation is seen in the retrieved profile, with a bias as high as -1.5 to -1.8 hPa. The RMSE varies from 1 hPa to 1.5 hPa for the training and testing data for the year 2020, and retrieval errors decrease as height increases. Except for near the surface, where errors are 2 hPa to 3 hPa, the RMSE for the independent dataset for the year 2021 ranges from 1.5 hPa to 1.8 hPa. Because pressure and refractivity are weakly correlated in the lower troposphere compared to the upper troposphere, there is a larger error in the retrieved pressure at lower levels. A retrieval error below 2 hPa to 2.5 hPa is tolerable since the surface pressure in the lower levels is highly variable, with a variability of about 10 hPa to 15 hPa. *Lasota* (2021) reported a similar pressure retrieval error.

For the training-testing split (2020) and independent (2021) datasets, Figure 19 displays the spatial distribution of the bias, RMSE, and standard deviation of the differences between actual and predicted pressure. For the 2020 test, the bias ranges from -0.2 hPa to 0.4 hPa, and for the 2021 test, it ranges from -1.2 hPa to 0.5 hPa. RMSE varies across the domain from 0.6 hPa to 3.3 hPa for the 2020 test and from 1.2 hPa to 3.5 hPa for the 2021 test. For the tests in 2020 and 2021, the standard deviation ranges from 0.6 hPa to 3.1 hPa and from 0.9 hPa to 3.4 hPa, respectively. Similar to the water vapour pressure and temperature retrieval errors, the error in retrieved pressure increases as one moves from lower latitudes to higher latitudes. This is most likely because there is less dependence of refractivity on the thermodynamic profiles in higher latitudes, as well as because the atmospheric conditions are more variable there than in the tropical region.

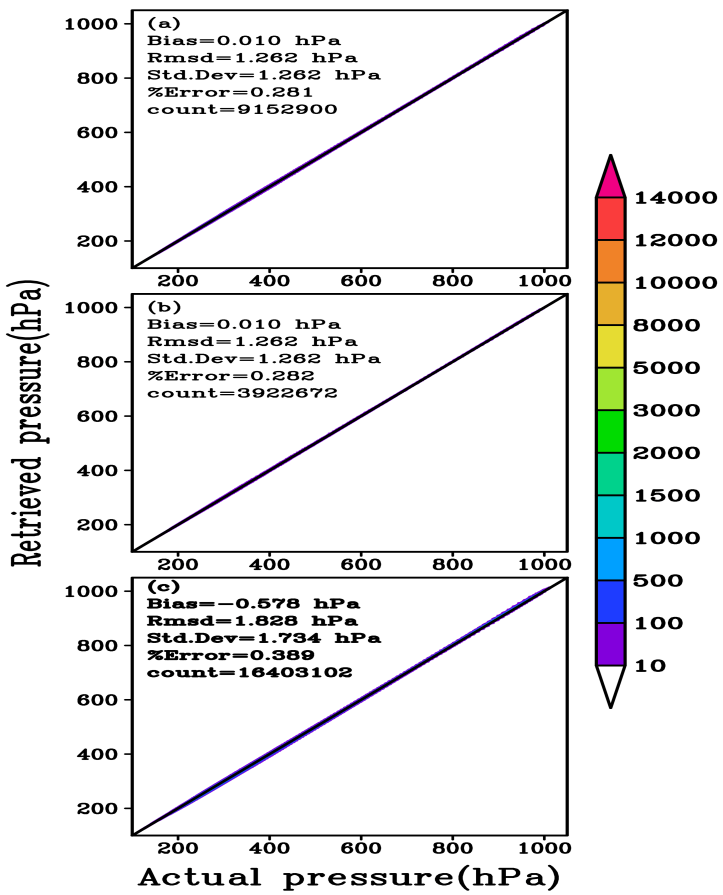


Figure 17: Scatter plot of actual versus retrieved pressure (a) for the training data set in 2020, (b) for testing data set in 2020, and (c) for completely independent data set fot the year 2021. Color bar represents the number of observations available for each pressure bin. This figure is based on data that has been accumulated over all vertical levels and the whole domain.

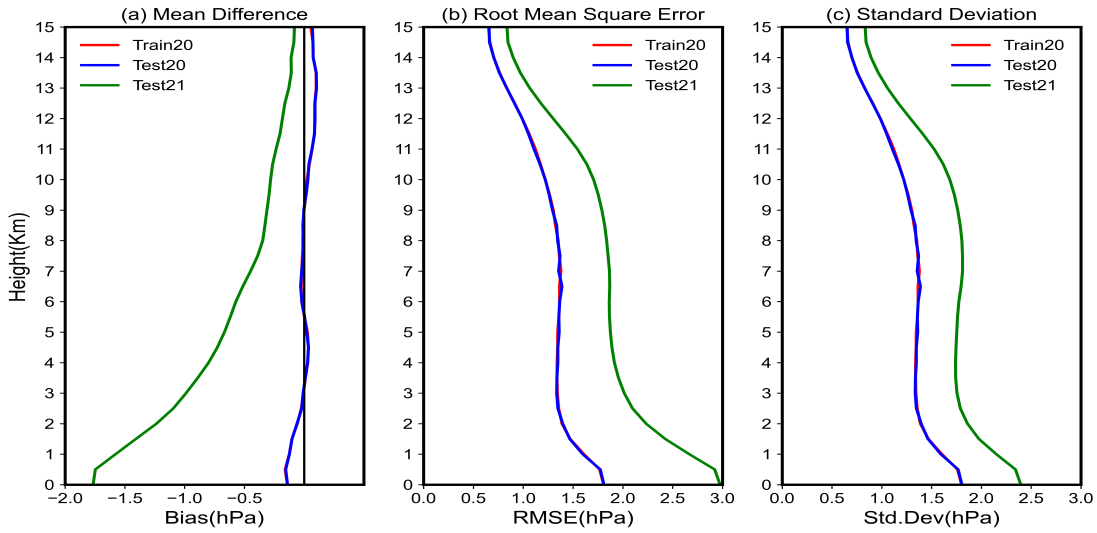


Figure 18: Vertical profiles (a) bias (i.e. mean differences), (b) root mean square error (RMSE), and (c) standard deviation (Std.Dev) of the difference between actual and model retrieved pressure, for training and testing data sets. The data acquired over the entire domain was used to create the figure.

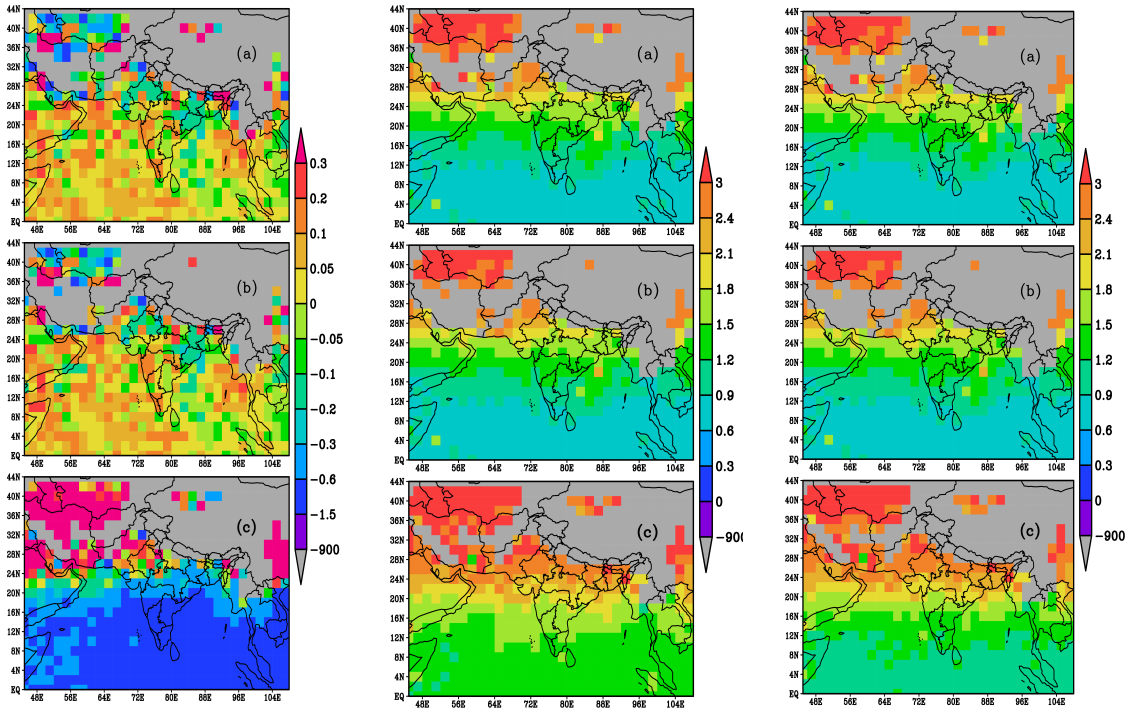


Figure 19: Spatial pattern of mean diffferences (i.e. bias), root mean square differences (RMSE) and standard deviation (Std.Dev) of the differences between actual and model retrieved pressure. First column is for bias, second is for RMSE, and third is for standard deviation. Panels (a), (b) and (c) in each column represent the training data set for 2020, testing data for 2020 and the testing data for 2021. The figure is produced using data collected across all heights.

5 Conclusion

In order to retrieve thermodynamic profiles from GNSS RO refractivity observations, we performed this work to develop a reliable system with less need on external data sources. In contrast to earlier studies, where refractivity and target profiles were taken from two different sources, we used the simulated data sets for developing the models. This may have avoided additional retrieval errors caused by errors made during the spatial and temporal collocation of input and output targets. As far as these added errors are concerned, the simulated observations have an advantage.

In order to retrieve tropospheric thermodynamic profiles from simulated refractivity profiles, two artificial neural network models were developed. Data from the two-year period of January 2020 to December 2021 was used in this process. While data from the year 2020 were utilized for developing and testing the proposed model, data from the year 2021 were used as a completely independent dataset to check the robustness of the model. On both, the testing dataset and the dataset that was completely independent, the developed model consistently delivered satisfactory results. When averaged vertically, the roots mean square errors (RMSE) for temperature, pressure, and water vapour pressure are 1.28 K, 1.26 hPa, and 0.19 hPa, respectively. With a completely independent data set acquired in 2021, the vertically averaged root mean square error (RMSE) for pressure is 1.82 hPa, for temperature it is 1.64 K, and for water vapour pressure it is 0.24 hPa. While the retrieval errors for water vapour pressure are significantly lower in this study, they are still comparable to those acheived by earlier studies for temperature and pressure.

Further analysis found that retrieval errors over the extra-tropical region are marginally higher than those over the tropical region. This is primarily because the association between refractivity and thermodynamic profiles is weaker over the extra-tropical region than it is in the tropical region. Furthermore, compared to the tropical region, the extra-tropical region has substantially higher degrees of thermodynamic profile variability.

By applying these models to the refractivity profiles provided by COSMIC-2, we hope to further the work. Additionally, in order to capture the variation in real refractivity profiles and makes our model more robust and generalized, we want to include some base altitude- and latitude-dependent bias to better represent the diversity in real refractivity profiles and to strengthen and extend our model. The variability information of the thermodynamic profiles over land toward northern latitudes will be better encapsulated by doing this. Finally, we advocate using such a model in an online learning system so that it can continue to train and update its weights as fresh data from various years is added.

References

[1] R. A. Phinney; D. L. Anderson. "On the radio occultation method for studying planetary atmospheres". In: *Journal of Geophysical Research* 73.5 (1968). DOI: 10.1029/JA073i005p01819.

[2] Richard A. Anthes. "Exploring Earth's atmosphere with radio occultation: Contributions to weather, climate and space weather". In: *Atmospheric Measurement Techniques* 4 (2011). DOI: 10.5194/Amt-4-1077-2011.

[3] G. Modrell; A. Morrison; J. Pomalaza; S.G. Ungar B. Lusignan. "Sensing the earth's atmosphere with occultation satellites". In: *Proceedings of the IEEE* 57.4 (1969). DOI: 10.1109/PROC.1969.7000.

[4] Diederik P. Kingma; Jimmy Ba: "Adam: A Method for Stochastic Optimization". In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2014). DOI: 10.48550/arXiv.1412.6980.

[5] J. L. Davis; T. A. Herring; I. I. Shapiro; A. E. E. Rogers; G. Elgered. "Geodesy by interferometry: Effects of atmospheric errors on estimates of baseline length". In: *Radio Science* 20.6 (1985). DOI: 10.1029/RS020i006p01593.

[6] E. R. Kursinski; G. A. Hajj; J. T. Schofield; R. P. Linfield; K. R. Hardy. "Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System". In: *Journal of Geophysical Research Atmospheres* 102 (1997). DOI: 10.1029/97JD01569.

[7] Mohamad H Hassoun. *Fundamentals of artificial neural networks*. MIT Press, 1995.

[8] Shu-Ya Chen; Chian-Yi Liu; Ching-Yuang Huang; Shen-Cha Hsu; Hsiu-Wen Li; Po-Hsiung Lin; Jia-Ping Cheng; Cheng-Yung Huang. "An Analysis Study of FORMOSAT-7/COSMIC-2 Radio Occultation Data in the Troposphere". In: *Remote Sensing* 13.4 (2021). DOI: 10.3390/rs13040717.

[9] Randhir Singh; Satya P. Ojha; Richard Anthes; Douglas Hunt. "Evaluation and Assimilation of the COSMIC-2 Radio Occultation Constellation Observed Atmospheric Refractivity in the WRF Data Assimilation System". In: *Journal of Geophysical Research Atmospheres* 126.18 (2021). DOI: 10.1029/2021JD034935.

[10] Aaron Courville Ian Goodfellow Yoshua Bengio. *Deep Learning*. MIT press, 2016.

[11] A. K. Steiner; B. C. Lackner; F. Ladstädter; B. Scherllin-Pirscher; U. Foelsche; G. Kirchengast. "GPS radio occultation for climate monitoring and change detection". In: *Radio Science* 46.6 (2011). DOI: 10.1029/2010RS004614.

[12] Richard A. Anthes; Christian Rocken; Ying-Hwa Kuo. "Applications of COSMIC to meteorology and climate". In: *Terrestrial, Atmospheric And Oceanic Sciences* 11 (2000). DOI: 10.3319/Tao.2000.11.1.115(Cosmic).

[13] T.-K. Wee; S. Sokolovskiy; C. Rocken; W. Schreiner; D. Hunt; R. A. Anthes Kuo; Y.-H. "Inversion and error estimation of GPS radio occultation data." In: *Journal of the Meteorological Society of Japan* 82 (2004). DOI: 10.2151/jmsj.2004.507.

[14] Elżbieta Lasota. "Comparison of diferent machine learning approaches for tropospheric profling based on COSMIC-2 data". In: *Earth, Planets and Space* 73 (2021). DOI: 10.1186/s40623-021-01548-4.

[15] J. Wickert; A. G. Pavelyev; Y. A. Liou; T. Schmidt; C. Reigber; K. Igarashi; A. A. Pavelyev; S. Matyugov. "Amplitude variations in GPS signals as a possible indicator of ionospheric structures". In: *Geophysical Research Letters* 31.24 (2004). DOI: 10.1029/2004GL020607.

[16] Max Born; Emil Wolf Oxford. "Principles of Optics Electromagnetic Theory of Propagation, Interference and Diffraction of Light". In: *Pergamon Press* (1980).

[17] Fjeldbo G.; Kliore A. J.; Eshleman V. R. "The Neutral Atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments". In: *Astronomical Journal* 76 (1971).

[18] M. P. Rennie. "The impact of GPS radio occultation assimilation at the Met Office". In: *Quarterly Journal of the Royal Meteorological Society* 136 (2010). DOI: 10.1002/qj.521.

[19] Gorbunov M. E.; S. V. Sokolovskiy. "Remote sensing of refractivity from space for global observations of atmospheric parameters". In: *Report 119 Max-Planck-Institute for Meteorology* (1993).

[20] Gordon D. Thayer. "An improved equation for the radio refractive index of air". In: *Radio Science* 9.10 (1974). DOI: 10.1029/RS009i010p00803.

[21] Stephen S. Leroy; Chi O. Ao; Olga Verkhoglyadova. "Mapping GPS Radio Occultation Data by Bayesian Interpolation". In: *Journal of Atmospheric and Oceanic Technology* 29.8 (2012). DOI: 10.1175/JTECH-D-11-00179.1.

[22] Michael Bevis; Steven Businger; Thomas A. Herring; Christian Rocken; Richard A. Anthes; Randolph H. Ware. "GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system". In: *Journal of Geophysical Research Atmospheres* 97 (1992). DOI: 10.1029/92JD01517.

[23] Ernest K. Smith; S. Weintraub. "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies". In: *Proceedings of the IRE* (1953).

[24] Lennart Bengtsson; Gary Robinson; Rick Anthes; Kazumasa Aonashi; Alan Dodson; Gunnar Elgered; Gerd Gendt; Robert Gurney; Mao Jietai; Cathryn Mitchell; Morrison Mlaki; Andreas Rhodin; Pierluigi Silvestrin; Randolph Ware; Robert Watson; Werner Wergen. "The use of GPS measurements for water vapor determination". In: *Bulletin of the American Meteorological Society* 84 (2003). DOI: 10.1175/Bams-84-9-1249.

[25] Chunming Wang; George Hajj; Xiaoqing Pi; I. Gary Rosen; Brian Wilson. "Development of the Global Assimilative Ionospheric Model". In: *Radio Science* 39.1 (2004). DOI: 10.1029/2002RS002854.

[26] Barbara Scherllin-Pirscher; Andrea K. Steiner; Richard A. Anthes; M. Joan Alexander; Simon P. Alexander; Riccardo Biondi; Thomas Birner; Joowan Kim; William J. Randel; Seok-Woo Son; Toshitaka Tsuda; Zhen Zeng. "Tropical Temperature Variability in the UTLS: New Insights from GPS Radio Occultation Observations". In: *Journal of Climate* 34 (2021). DOI: 10.1175/JCLI-D-20-0385.1.

[27] R. A Anthes; P. A Bernhardt; Y. Chen; L. Cucurull; K. F. Dymond; D. Ector; S. B. Healy; S.-P. Ho; D. C Hunt; Y.-H. Kuo; H. Liu; K. Manning; C. McCormick; T. K. Meehan; W J. Randel; C. Rocken; W S. Schreiner; S. V. Sokolovskiy; S. Syndergaard; D. C. Thompson; K. E. Trenberth; T.-K. Wee; N. L. Yen; Z Zeng. "The COSMIC/Formosat-3 mission: Early results". In: *Bulletin of the American Meteorological Society* 89 (2008). DOI: 10.1175/Bams-89-3-313.

[28] C. Rocken; R. Anthes; M. Exner; D. Hunt; S. Sokolovskiy; R. Ware; M. Gorbunov; W. Schreiner; D. Feng; B. Herman; Y.-H. Kuo; X. Zou. "Analysis and validation of GPS/MET data in the neutral atmosphere." In: *Journal of Geophysical Research* 102 (1997). DOI: 10.1029/97JD02400.

[29] Xu Xu; Xiaolei Zou. "COSMIC-2 RO Profile Ending at PBL Top with Strong Vertical Gradient of Refractivity". In: *Remote Sensing* 14.9 (2022). DOI: 10.3390/rs14092189.