Indian Journal of Radio & Space Physics
Vol 42, December 2013, pp 411-419

Retrieval of water vapour profiles from radio occultation refractivity using artificial neural network

Abhineet Shyam[$,*], B S Gohil & Sujit Basu

Geophysical Parameters Retrieval Division (GRD), Atmospheric & Oceanic Sciences Group (AOSG), Earth, Ocean, Atmosphere, Planetary Sciences & Applications Area (EPSA), Space Applications Centre (SAC), ISRO, Ahmedabad 380 058, India

[$]E-mail: abhineetshyam@gmail.com

Received 8 March 2013; revised 21 August 2013; accepted 3 September 2013

Artificial neural network (ANN) technique has been used to derive water vapour pressure profiles in the troposphere from radio occultation data over India and adjoining region. A fully connected three-layer network, with one hidden layer, has been constructed and standard back propagation algorithm has been used to train the network. While month, latitude and vertical profile of refractivity/bending angle constitute the input vector, the water vapour partial pressure profile forms the output vector. Only the moisture-laden summer monsoon months of June, July, August and September of 2010 have been considered for developing the retrieval algorithm. There are 2120 input and output pairs, out of which 1696 pairs form the training set while the remaining pairs constitute the validation set. The retrieved profiles of water vapour pressures in the validation set have been compared with the corresponding COSMIC operational products of water vapour pressure profiles. The effectiveness of the algorithm is apparent from this comparison and also from the vertical profiles of bias and root mean square error (RMSE). The statistics show better performance of the algorithm with refractivity as one of the inputs than with bending angle. The RMSE in water vapour retrieval from refractivity is within 1.5-2 hPa compared to markedly higher values of 6 hPa when derived from bending angle. Additionally, the algorithm has also been tested in an independent year 2009 and the performance of the refractivity based retrieval has been found to be highly consistent in the year 2010, with RMSE within 1.5 hPa.

Keywords: Artificial neural network, Radio occultation refractivity, Water vapour pressure profile, Back propagation algorithm

PACS Nos: *92.60.hv; 92.60.hf*; 92.60.Jq; 07.05.Mh; 95.10.Gi

1 Introduction

Monitoring of climate change inevitably requires numerical weather prediction models and knowledge of atmospheric temperature and water vapour profiles is an essential prerequisite of such models. Although the radiosonde networks are distributed over the globe, these are simply inadequate for global numerical weather prediction models. Moreover, the vast oceanic area cannot be monitored by such networks. Hence, remote sounding via satellites is the only viable alternative for global profiling of atmosphere, and particularly so over the oceans. Radio occultation (RO) plays a vital role in such type of sounding[1,2]. During such an occultation, the GPS receiver placed on a Low-Earth Orbiting (LEO) Satellite records the excess phase delay induced by the Earth's atmosphere, from which the atmospheric Doppler shift, bending angle and refractivity can be derived[3,4]. In the dry atmosphere, where the value of water vapour mixing ratio is generally less than a value of ~ 10^{-4}, pressure and temperature can be derived quite easily with hydrostatic equation and ideal gas law[2,5]. However, the atmosphere is seldom dry, and particularly so in the lower troposphere. Hence, an accurate knowledge of water vapour in such a wet atmosphere is crucial in making accurate numerical weather prediction.

As mentioned, the bending angle obtained from radio occultation is used to derive the vertical profile of refractivity from which atmospheric temperature profile can be derived in the case of dry atmosphere. However, the derivation of vertical profile of water vapour requires additional a priori information[6]. This is often termed as water vapour ambiguity.

There are three distinct approaches for deriving water vapour profiles from RO measurements, which are: (i) A direct method based on *a priori* knowledge of the temperature and pressure profiles[7]; (ii) An

iterative method using the combination of the equation of state, hydrostatic equation and prior knowledge of the temperature profile[6,8]; (iii) An optimal estimation inverse method based on the prior temperature and water vapour pressure profiles[9-11]. Recently, a variant of the approaches (i) and (ii) has been devised using modeled wet component of the total refractivity, and surface temperature and surface pressure as a priori information[12].

All these approaches suffer from the major drawback that a prior knowledge of temperature profile is absolutely essential for the derivation and hence, the results are pre-determined to a certain extent. To overcome this difficulty, the data adaptive approach of artificial neural network (ANN) has been advocated in more recent time[13]. The retrieval using ANN was, however, restricted to China and nearby regions and input-output pairs of bending angle and water vapour pressure was used for training the network[13]. In the present study, ANN technique is used for deriving water vapour pressure profiles over India and adjoining regions. Moreover, unlike the earlier study, both refractivity profiles and bending angle profiles are used separately as input to the network and studied the performance of the algorithms inter se. The former is subsequently being designated as RANN (i.e. refractivity based ANN) and the latter by BANN (bending angle based ANN).

2 Method

ANNs have been used in a wide variety of geophysical applications and a comprehensive review exists[14]. Nevertheless, the salient features are recapitulated here for the sake of self-consistency. ANN is a data-driven technique, also known as machine-learning. A generic ANN mapping between a set of input and output patterns can be symbolically written as:

$$Y = M(X), \quad X \in R^n, \text{ and } Y \in R^m \qquad \ldots(1)$$

where, X and Y are the input and output vectors belonging to n and m dimensional vector spaces, respectively. The mapping represented by M, or rather a very close approximation to it, is performed by the network consisting of simple interconnected elements known as neurons, the terminology being inspired by biological nervous systems. As in nature, the connections between the neurons largely determine the network function. The network is trained to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, neural networks are adjusted or trained, so that a particular input leads to a specific target output. Thereafter, the network is adjusted, based on comparison of the output and the target, until the network output matches the target, albeit within specified error limits. Typically, many such input/target pairs are needed to train a network. The training of the network can be affected using a variety of algorithms. But more often than not, the back propagation algorithm is preferred over others because of its relative simplicity. This algorithm has been used in the present study.

The back propagation algorithm computes the gradient of the output error of the multilayered network with respect to each of the connection weights[15] and adjusts the ANN weights (and biases) in the descending gradient direction. In other words, it tries to improve the performance of the neural network by reducing its error along its gradient. The error is expressed by the mean squared error (MSE) or the root-mean-square error (RMSE) given by:

$$\text{MSE, } e = \frac{\sum (y_i - o_i)^2}{n}$$

where, y_i, is the predicted target output; and o_i, the actual target output vector. Further, RMSE = $(\text{MSE})^{1/2}$. In each training step, the weights (w) are adjusted towards the direction of maximum decrease, scaled by some learning rate,

lambda (λ): $w_{new} = w_{old} + \lambda (\nabla e)$

where, $\nabla e = \left(\frac{\delta e}{\delta w_1}, \frac{\delta e}{\delta w_2}, \ldots\ldots, \frac{\delta e}{\delta w_n}\right)$ gives the components of the error gradient with respect to all the weights. There are two popular variants of the back propagation algorithm based on the minimization algorithm used, viz. Gradient Descent back propagation algorithm and Levenberg-Marquardt backpropagation algorithm. The latter has been used owing to faster and assured convergence.

A typical ANN architecture is shown in Fig. 1. This consists of an input layer, an output layer and one or more hidden layers. Whereas the numbers of input and output neurons are dictated by the problem at hand, the number of neurons in each hidden layer is largely a matter of trial and error. In the present case,

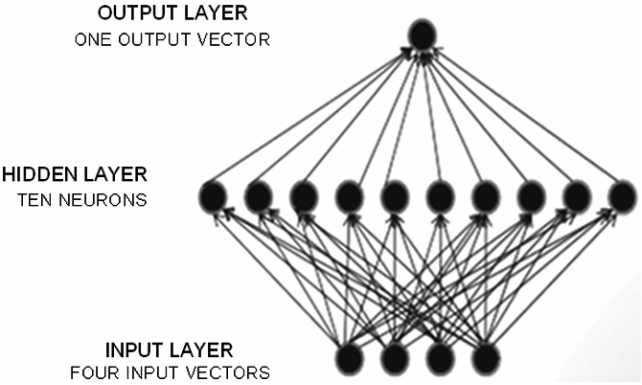

Fig. 1 — A typical ANN architecture based on feed-forward back propagation method

the number of input neurons is 46 [n in Eq. (1)], while the number of output neurons is 22 [m in Eq. (1)]. More specifically, the input consists of month, latitude, 22 altitude levels and refractivity values at these levels. The output is simply the water vapour pressure at these altitudes. As regards the number of hidden layers and the number of neurons in each of these layers, various network topologies have been tested by varying the number of hidden layers from one to three and the number of neurons from six to ten. Mean squared error for each of these trials were compared. The topology with smallest RMSE, in turn producing the best input output match, was selected. It is worth pointing out that many of the distinct topologies among the trials did have similar performance, and so the one with minimum number of hidden layers and number of neurons in that order has been selected. This topology, consisting of one input layer with 46 neurons, one hidden layer with 10 neurons and one output layer with 22 neurons, has been used in the paper.

3 Data

The area of the present study is the Indian landmass and its adjoining region (8°-38°N; 66°-98°E). The data consist of COSMIC occultation data[16] for the summer monsoon months of June, July, August, September for the years 2009 and 2010. The data are accessed from the website http://tacc.cwb.gov.tw/ for the above mentioned periods. The occultation data of interest in this study comprise of refractivity, bending angle and operational water vapour partial pressure products from COSMIC at various vertical levels up to 14 km. To begin with, the occultation data are preprocessed to select only those RO profiles which penetrate at least up to 2 km and are confined to the chosen geographical domain. The 2 km penetration criterion is a consideration following from the presence of significant N-bias in COSMIC refractivity data, especially in the tropics and subtropics, due to super-refraction effects characterizing the atmospheric boundary layer. A total of 2120 and 2808 occultations data have been obtained for the years 2010 and 2009, respectively on applying the two criteria. The data for 2010 have been subdivided into two sets. One is the training set and another is the validation set. The training set consists of 1696 points, which have been used for developing the algorithm. The remaining set has been used for validation. As far as the data for 2009 is concerned, the entire set has been used for validation. Further, bending angle, refractivity and water vapour profiles in the selected datasets are retained upto 14 km and interpolated at the following variable vertical resolutions which are: 200 m up to 1 km; 0.5 km in the range 1-5 km; and 1 km in the range 6-14 km. The particular choice of the assumed vertical resolutions is intended to capture the sharp gradients of water vapour, known to exist in the lower troposphere, while at the same time keeping the number of neurons at a reasonable level. The latter is significant from the point of view of the complexity of the network and consequently the computational time, both of which may be affected by the number of neurons at the input level. Beyond 14 km, the water vapour content in the atmosphere has been found to be negligible.

4 Results and Discussion

The locations of the occultation events used in the study are shown in Fig. 2. The dots are the profiles used for training while the profiles set aside for retrieval and validation are earmarked as '+' symbol. Quantitatively, the numbers are 1696 and 424, respectively. These numbers are same for the bending angle as well as the refractivity, since the latter is derived from the former. Following the usual practice, at the beginning of the ANN's training process, a desired value of the average error, is set as the convergent criterion of the forward computations. The root mean squared error of 0.1 hPa has been considered as the convergent criterion. As soon as the criterion is reached, the training is stopped. The performance of the training can be best judged from a scatter plot between the target water vapour pressure and the water vapour pressure obtained as a result of training. Ideally, the plot should be a linear line with

negligible bias and slope near unity. This plot is shown in Fig. 3(a) for RANN. Separately, the algorithm is also developed for bending angle as input and the training procedure is carried out similar to that of refractivity. Figure 3(b) shows the scatter plot between the target water vapour partial pressure and the one retrieved with BANN. It is apparent from the two plots that the training of the neurons is effective in replicating the target profiles exceedingly well from refractivity than from bending angle, especially in the lower troposphere. Relatively larger scatter of points for bending angle [Fig. 3(b)] may possibly be attributed to the complicated implicit relationship between atmospheric water vapour partial pressure and bending angle in contrast to the more direct physical relationship between refractivity and water vapour partial pressure[7].

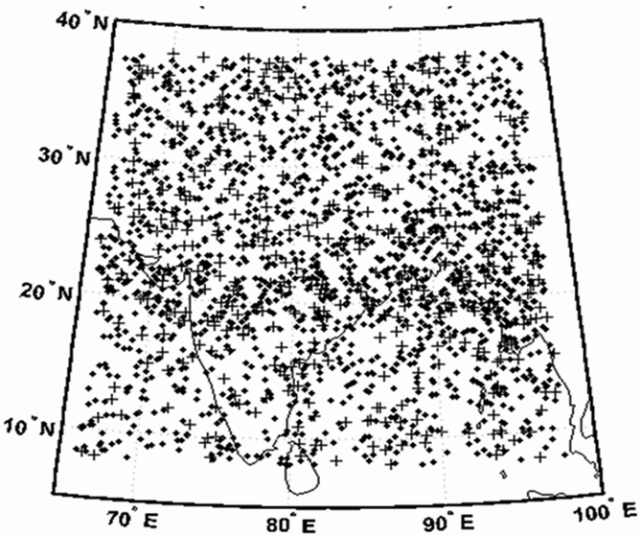

Fig. 2 — Geo-location of COSMIC occultation events over Indian region for the year 2010 (June-September) [Same geographical region is selected for the validation dataset (year 2009)]

Once the training is finished, the parameters of the trained network are used to derive water vapour pressure profiles from an independent dataset, earmarked for validation in the beginning, using the months, latitudes, altitude levels and RO refractivity profiles as inputs for RANN. Bending angle profiles replaces the refractivity profiles in case of BANN, keeping all other input parameters intact. In order to quantify the quality of the retrieved profiles, the results are compared with the corresponding water vapour profiles available from COSMIC dataset. These comparative water vapour profiles have been obtained using one-dimensional variational optimal estimation method (1DVAR), operational for COSMIC products. The bias and root mean square error (RMSE), in absolute term, of the retrieved water vapour profiles for RANN and BANN are shown in Figs 4(a) and 4(b) respectively. Figures 5(a) and (b) show the same in percent terms. The retrieval bias is better than 0.2 hPa below ~8 km, translating to a difference of less than 10% in the water vapour dominant region. Above 8 km, bias shows an increasing trend with a peak value of 40% at 10 km for RANN [Fig. 5(a)] and about 50% for BANN [Fig. 5(b)] and sporadic decrease at a few vertical levels. By and large, the nature of bias above 8 km is similar in both RANN and BANN as is evident from the Figs 4(a-b) and 5(a-b), although the latter is largely in excess of the former. At 8 km, the bias is

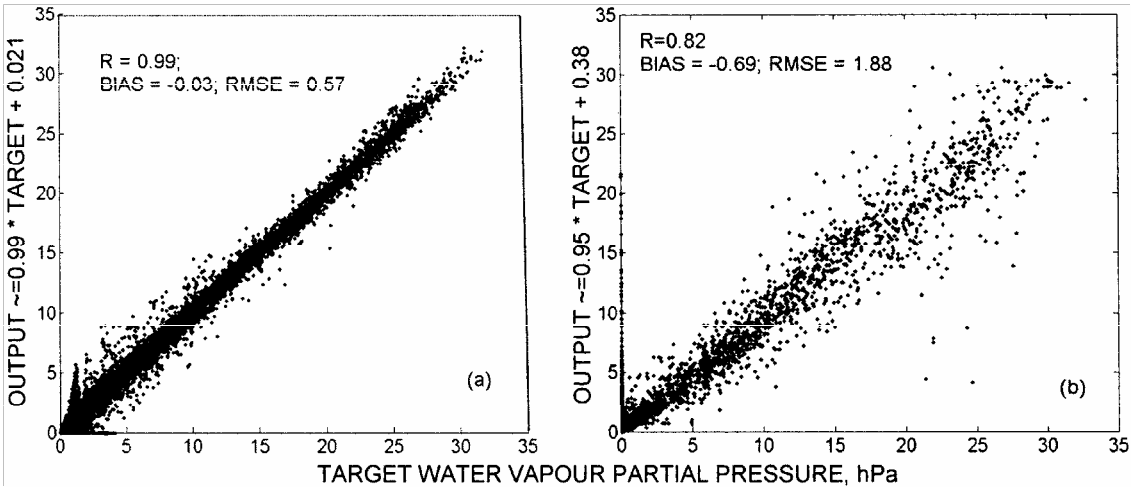

Fig. 3 — Linear relationship between the target water vapour pressures and the training results of ANN forward computations with COSMIC network input as: (a) refractivity and (b) bending angle

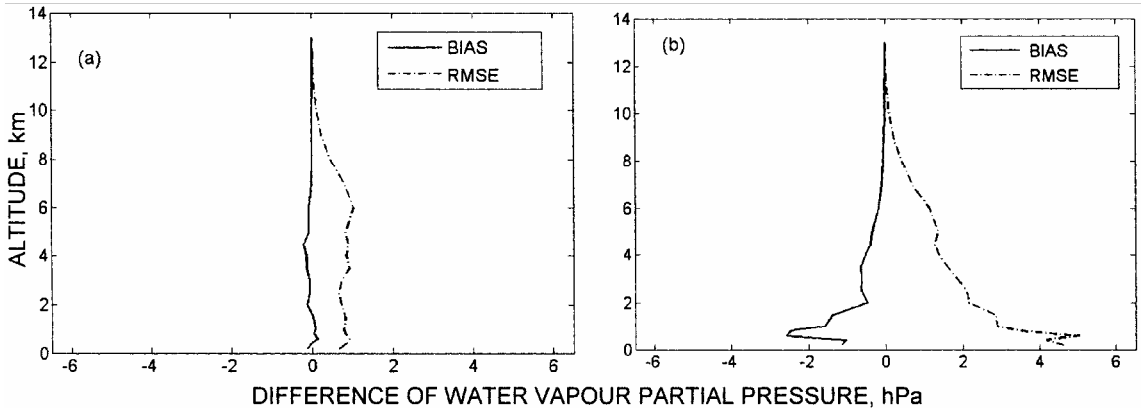

Fig. 4 — Vertical profiles of bias and root mean square error in water vapour partial pressure derived from ANN with: (a) refractivity and (b) bending angle, as input for the validation dataset of 2010, vis-a`-vis the COSMIC operational water vapour products from the same dataset

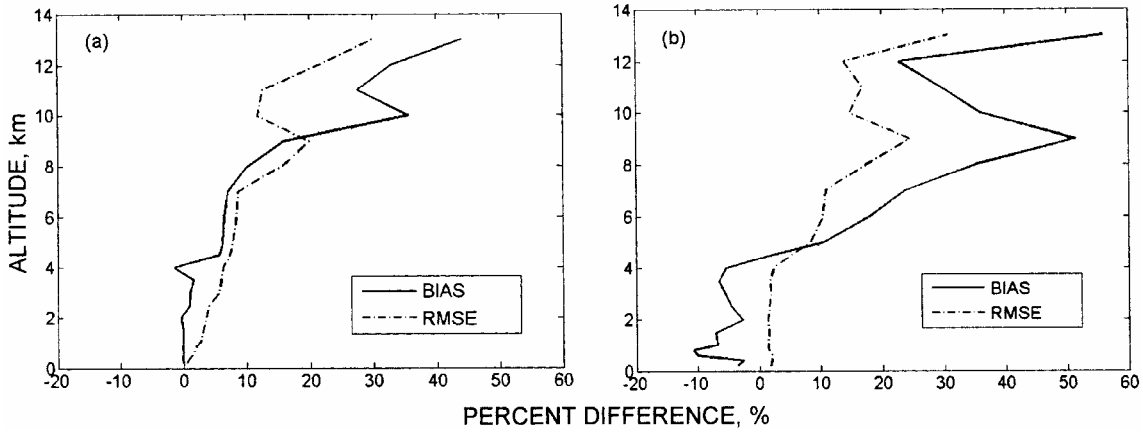

Fig. 5 — Vertical profiles of bias and root mean square error, in percent, in water vapour partial pressure derived from ANN with: (a) refractivity and (b) bending angle, as input from the validation dataset of 2010, vis-a`-vis the COSMIC operational water vapour products from the same dataset

within 15% for RANN, while it is 35% for BANN [Fig. 5(a & b)]. The differences are remarkable below 8 km with both bias and RMSE being better than 10-15% for RANN. Below 4 km, bias and RMSE are better than 0.2 hPa (0.5-1%) and 1 hPa (up to 5-6%) in case of RANN, as shown in Figs 4(a) and 5(a). The result for BANN is markedly distinct compared to that for RANN, particularly in the humid lower troposphere (i.e. below 6 km). Not only does the bias show large under estimation of up to 2 hPa at around 1 km, the RMSE has increased by more than 5 times the peak RMSE for RANN. In percent terms, the bias and RMSE are in the range -2.0 to -10% and 1-3%, respectively, for BANN as shown in Figs 4(b) and 5(b).

The comparative results for six retrieval examples are shown in Fig. 6(a-f) for RANN. The retrieval is of reasonably good accuracy for all the profiles and matches well with the COSMIC 1DVAR profiles of water vapour pressure. Some discrepancy is seen for the third and fourth example plots as shown in Fig. 6(c and d), respectively. For the third profile, this is evident in the altitude range of 4-10 km. Below 3 km, the retrieval performs quite well. In the fourth example, there are discrepancies in the altitudes between 6 and 10 km, although the discrepancy is not as high as in the case of third example.

Although the performance of the retrieval algorithm has been tested on an independent validation data set of the year 2010, specifically earmarked for this purpose. It is quite interesting to see whether the algorithms performs equally well for any other year. For this, the RANN algorithm is applied on the COSMIC RO data for the monsoon months of the year 2009 also. The retrieval statistics in terms of mean bias and RMSE, in absolute and percent values, are shown in Fig. 7(a and b) As can be clearly seen, Fig. 7(a-b) closely resemble

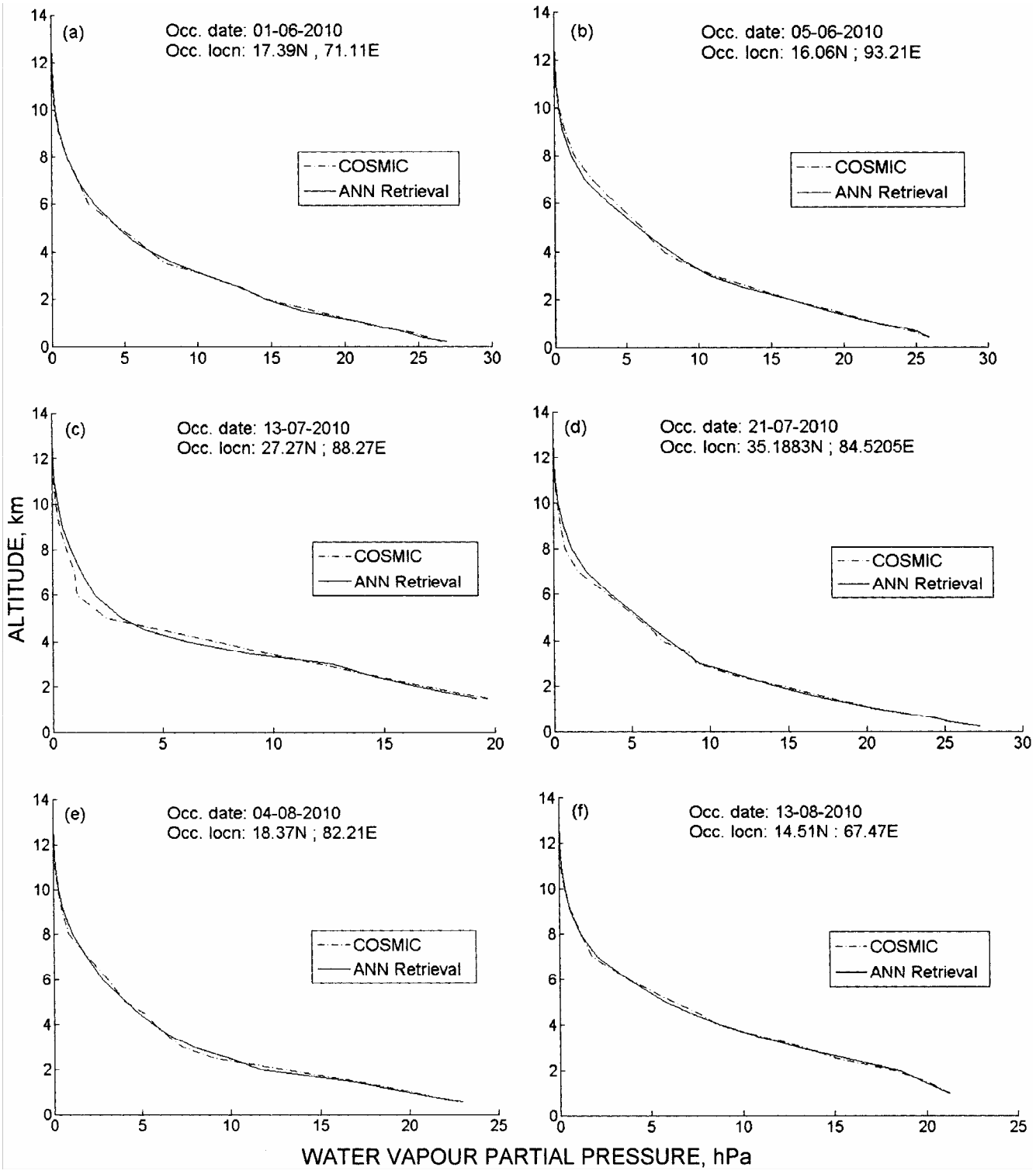

Fig. 6 (a-f) — Comparison of water vapour pressures retrieved for RANN with the COSMIC water vapour pressures for six examples from the validation set of 2010

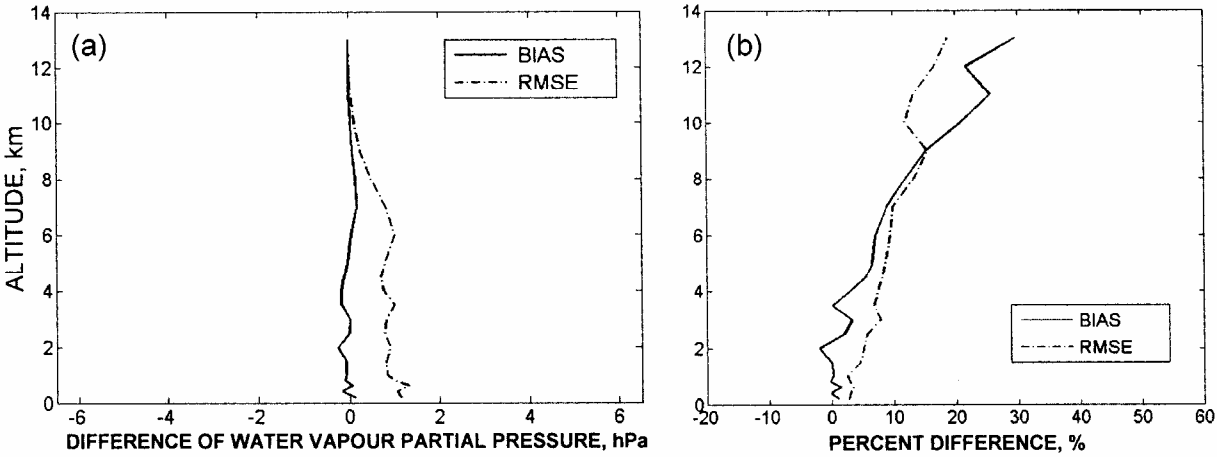

Fig. 7 — Vertical profiles of bias and root mean square error: (a) in absolute unit and (b) in percent, of water vapour partial pressure derived from ANN with refractivity as input from the validation dataset of 2009, vis-a`-vis the COSMIC operational water vapour products from the same dataset

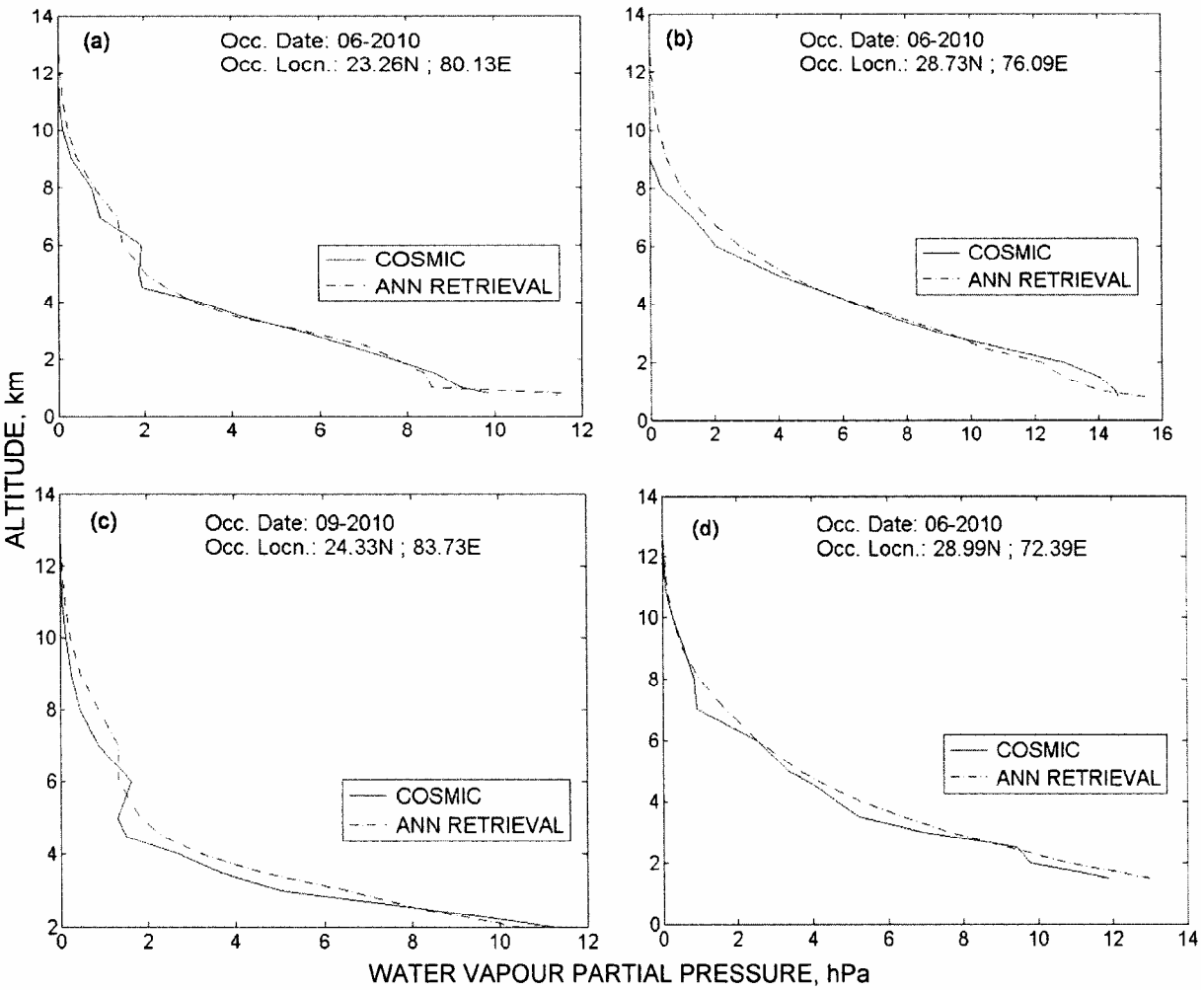

Fig. 8 (a-d) — Comparison of water vapour partial pressures retrieved for RANN with the COSMIC water vapour pressures for the less humid (< 8-10 hPa at 2-3 km) cases taken from the validation set of 2010

analogous to Figs. 4(a) and 5(a) barring very minor differences, especially below 2-4 km. However, these inconsistencies are rather due to significantly lesser number of occultation profiles available at depths of 2 km and below, than to the incapability of the ANN algorithm to retrieve water vapour profiles accurately. Unfortunately, the issue of lack of or meagre amount of data available at depths below 2 km or so persists with the radio occultation processed data. Notwithstanding such aberrations, the overall quality of retrieval can be seen from the consistency in the results above 2 km.

The lowest near-surface humidity values for the study domain and during the selected months were searched and zeroed-in on the profiles with humidity range of 8-10 hPa at around 2-3 km as relatively dry cases. Figure 8 (a-d) shows the comparison of derived water vapour partial pressure with corresponding COSMIC 1DVAR retrievals for the relatively less humid conditions. Figure 8(a and d) show discrepancy of ~1-1.5 hPa for 1-2 km, with observable deviations at altitudes in the range 4-8 km. Similarly, reasonable deviations are observed for the remaining two figures. In summary, all the Fig. 8(a-d) show over estimation by ANN at various vertical levels. The results indicate the existing ANN network limitations in deriving very accurate humidity values in typical dry cases, as has also been discussed in the statistical analysis of retrieval errors for the altitudes above 8 km, where the humidity has scant presence.

5 Conclusions

The technique of artificial neural network (ANN) has been used for the retrieval of water vapour pressure profiles in the atmosphere from radio occultation measurements of refractivity over the Indian landmass and adjoining region. The advantage of the technique is that the profile of water vapour pressure could be derived directly from radio occultation measurements without the need for using other auxiliary measurements, e.g. the temperature profiles from climate/NWP models. The result shows that the network is able to quickly learn the hidden relationship known to exist between the refractivity - water vapour pressure pairs than between bending angle - water vapour pressure pairs. The few typical retrieval examples shown in the study confirm this inference. This inference holds good for humidity abundant regions of atmosphere. However, for the drier regions of atmosphere, the retrieval accuracy, whether in respect of RANN or BANN, degrades albeit rather adequately for BANN than for RANN, thus indicating the limitation of the ANN-based algorithm in dry regions. Nevertheless, the consistency of the RMSE for the two years of data appears to suggest that the ANN technique has produced results which are quite promising. Apart from its use as a standalone retrieval technique, the technique can be used as a precursor to other operational iteration-based and optimal estimation methods by treating the retrievals as 'initial guess'. As a priori temperature profile information is also required in optimal method like the 1DVAR, it will be worthwhile to explore the retrieval of temperature from refractivity. It is believed to be plausible, since temperature, similar to water vapour pressure, is physically related to atmospheric refractivity.

In this study, the algorithm has been developed using satellite (COSMIC) data alone. It would be worthwhile to see if the results are equally good for water vapour pressure data derived from *in situ* source like radiosonde over the chosen region. However, due largely to lack of collocated good quality radiosonde data over Indian and adjoining region, satellite products alone have been considered for the study. Quality of the radiosonde apart, the quasi-random nature of radio occultation events make it difficult to find spatio-temporal collocation with radiosonde, known to be launched only at 00Z and 12Z (Z for Zulu time or equivalent time in UTC).

Acknowledgment

The authors are indebted to the Director, Space Applications Centre, for constant encouragement. They express their sincere gratitude to the Deputy Director, Earth, Ocean, Atmosphere, Planetary Sciences and Applications Area, and to the Group Director, Atmospheric and Oceanic Sciences Group, for their continuous support and motivation. They further acknowledge the TACC Centre at Taiwan for COSMIC data access and MOSDAC at Space Applications Centre, Ahmedabad, for data download.

References

1 Lu D, Fan Y & Jiyao X, Advances in studies of the middle and upper atmosphere and their coupling with the lower atmosphere, *Adv Atmos Sci* (*China*), 21 (2004) pp 361-368.

2 Rocken C, Anthes R, Exner M, Hunt D, Sokolovskiy S, Ware R, Gorbunov M, Schreiner W, Feng D, Herman B, Kuo Y & Zou X, Analysis and validation of GPS/MET data in the neutral atmosphere, *J Geophys Res (USA)*, 102 (1997) pp 29849-29866.

3 Kursinski E R**,** Hajj G A, Schofield J T, Linfield R P & Hardy K R, Observing Earth's atmosphere with radio occultation measurements using the Global Position System, *J Geophys Res (USA),* 102 (1997) pp 23429-23465.

4 Yunck T P, An overview of atmospheric radio occultation, *J Global Position Syst* (*Canada*), 1 (2002) pp 58-60.

5 Kursinski E R**,** Hajj G A, Leroy S S & Herman B, The GPS radio occultation technique, *Terr Atmos Ocean Sci (Taiwan)*, 11 (2000) pp 53-114.

6 Healy S B & Eyre J R, Retrieving temperature, water vapor and surface pressure information from refractive-index profiles derived by radio occultation: A simulation study, *Q J R Meteorol Soc* (*UK*), 126 (2000) pp 1661-1683.

7 Kursinski E R, Hajj G A, Hardy K R, Romans L J & Schofield J T, Observing tropospheric water vapor by radio occulation using the Global Positioning Sytem, *Geophys Res Lett* (*USA*), 22 (1995) pp 2365-2368.

8 O'Sullivan D B**,** Herman B M, Feng D, Flittner D E & Ward D M, Retrieval of water vapor profiles from GPS/MET radio occultations, *Bull Am Meteorol Soc (USA)*, 81 (2000) pp 1031-1040.

9 Engeln A V**,** Nedoluha G, Kirchengast G & Buler 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 (USA),* 108 (2003) pp 4337, doi: 10.1029/2002JD002908.

10 Palmer P**,** Barnett J, Eyre J & Healy S, A non-linear optimal estimation inverse method for radio occultation measurements of temperature, humidity, and surface pressure, *J Geophys Res (USA),* 105 (2000) pp 17513-17526.

11 Von Engeln A & Nedoluha G, Retrieval of temperature and water vapor profiles from radio occultation refractivity and bending angle measurements using an optimal estimation approach: a simulation study, *Atmos Chem Phys (Germany)*, 5 (2005) pp 1665-1677.

12 Jagadheesha D, Simon B, Pal P K, Joshi P C & Maheshwari A, A new technique for estimation of lower-tropospheric temperature and water vapor profiles from radio occultation refractivity, *J Atmos Ocean Technol (USA)*, 26 (6) (2009) pp 1075.

13 Xin W & Daren L, Retrieval of water vapor profiles with radio occultation measurements using an artificial neural network, *Adv Atmos Sci* (*China*), 22 (2005) pp 759-764.

14 Krasnopolsky V M, Neural network emulations for complex multidimensional geophysical mappings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling, *Rev Geophys* (*USA*), 45 (2007) pp RG3009.

15 Grossberg S, Nonlinear neural networks: Principles, machines and architectures, *Neural Netw (USA)*, 1 (1988), pp 17-61.

16 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 & Zeng Z, The COSMIC/FORMOSAT-3 mission early results, *Bull Am Meteorol Soc (USA),* 89 (2008) pp 313-333.