RESEARCH ARTICLE

ENHANCING VERTICAL RESOLUTION OF SATELLITE ATMOSPHERIC PROFILE DATA: A MACHINE LEARNING APPROACH.

Venugopal T and Venkatramana K
1. Department of Meteorology and Oceanography, Andhra University, Visakhapatnam, India.
2. The Energy and Resources Institute, New Delhi, India.

Manuscript Info

Abstract

We developed a statistical approach using the Artificial Neural Networks (ANN) to improve the vertical resolution of tropospheric relative humidity profiles (RH) from 20 pressure levels to 171 pressure levels. The model is based on an unconventional method in which we used the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) Global Positioning System Radio Occultation (GPS RO) data and the corresponding observed values of RH data. The model was developed using 3 years COSMIC daily data during 2007-2009 over the north Indian Ocean and produced high vertical resolution RH (171 pressure levels) output data from the coarse resolution inputs (20 pressure levels). We achieved the best performance in generating high vertical resolution data with a Pearson’s correlation coefficient (CC) of greater than 0.94 and scatter index (SI) of less than 0.1 throughout all pressure levels. Thus, the present approach is an efficient method to achieve the better vertical resolution of RH data from geostationary satellites.

Introduction:

The accurate high vertical, spatial and temporal resolution relative humidity (RH) profiles up to tropopause play an important role to understand the atmospheric stability and weather forecasting. High-resolution RH data also required to find the coupling between atmospheric parameters and climate change, as well as to validate the global models (Neerja et al., 2012). Different in situ profiling techniques like Radiosonde and Dropsonde have very high vertical resolutions but poor in spatial and temporal coverage, especially over global Oceans. Remote sensing data from polar-orbiting satellites and Global Positioning System Radio Occultation (GPS RO) observations also follows the same. While sounders on geostationary satellites provide very good spatial and temporal resolution in RH data, the vertical resolution is very poor with having only 20 pressure levels.

Thus, RH measurements in 3 dimensions (spatial, temporal and vertical) are not available as of today by any space or land-based technique. Neerja et al. (2012) have demonstrated an approach to full fill this gap by improving the vertical resolution of temperature profiles up to tropopause obtained from geostationary satellite measurements using an ANN technique to meet the required resolution in all three dimensions. Motivated by this study and results, here we extended and validated the approach with atmospheric RH profiles.
There are two basic types of atmospheric sounding: vertical sounding, in which the sounding instrument senses radiation coming from the atmosphere and the earth’s surface; and the limb sounding, in which only the limb of the atmosphere is sensed. Limb sounding has several advantages over vertical sounding (Gille and House, 1971; Kidder et al., 1995). GPS RO is a limb geometry active remote sensing technique that provides accurate all weather high vertical resolution profiles of measurements closely related to atmospheric variables, due to the insensitivity of the GPS signal wavelengths to clouds, aerosols and precipitation (Kursinski et al., 1997, 2000; Zhang et al., 2011; Terrestrial, Atmospheric and Oceanic Sciences (TAO), 2011). TAO also describes the Radio Occultation (RO) method; the application of RO to weather, climate, and ionosphere research. The report also documented the details on Taiwan’s Formosa Satellite Mission #3/Constellation Observing System for Meteorology, Ionosphere & Climate (FORMOSAT-3/COSMIC) in short F3C mission (Fig. 1).

The F3C is a joint Taiwan/US science mission for weather, climate, space weather, and geodetic research. The F3C mission was successfully launched on 14 April 2006 (15 April 2006, Taiwan time). Six identical microsatellites, each carrying an advanced GPS radio occultation (RO) receiver, a Tiny Ionospheric Photometer (TIP), and a Tri-Band Beacon (TBB) were deployed. The COSMIC payload science data are routinely downloaded every orbit via two National Oceanic and Atmospheric Administration (NOAA) Telemetry, Tracking, and Command Systems (TT&C) stations (in Alaska and Norway) and one National Aeronautics and Space Administration (NASA) station (in McMurdo, Antarctica); they are then transferred to the COSMIC data Analysis and Archival Center (CDAAC) at the University Corporation for Atmospheric Research (UCAR) in Boulder. CDAAC currently processes the COSMIC science data in near real-time: ninety percent of the RO profiles are delivered to operational weather centers within three hours of observation. CDAAC also reprocesses data in a more accurate post-processed mode (within six weeks of observation) for COSMIC as well as other missions. Formosat-3/COSMIC data are making a positive impact on operational global weather forecast models and particularly over regions void of data such as Oceans and Polar regions. Presently the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction (NCEP) and the UK Meteorological Office are using Formosat-3/COSMIC data operationally.

GPS RO generates the refractivity profiles which can be used to derive profiles of electron density in the ionosphere, the temperature in the stratosphere, and temperature and water vapor in the troposphere. The processing of GPS RO data usually assumes that wave propagation of the GPS signals in the Earth’s atmosphere can be well approximated by a single ray leading to the geometrical optics (GO) refractivity profiles (Poli et al., 2003). Thus, refractivity is derived from bending angles obtained by GO, assuming spherical symmetry via an Abel transform (Hajj et al., 2002). The air refractivity $N$ is a linear function of refractive index $n$ and can be related to atmospheric physical quantities via,

$$N = 10^6 (n - 1) = \frac{P}{T} (b1) + \frac{P_w}{T^2} (b2)$$

(Smith and Weintraub, 1953) (neglecting scattering and in a neutral atmosphere), where $P$ is the total atmospheric pressure (dry air and water vapor) in hPa, $T$ the temperature in K, $P_w$ the partial pressure in water vapor in hPa, $b1 = 77.6$ K hPa$^{-1}$, and $b2 = 3.73 \times 10^3$ K$^2$ hPa$^{-1}$. 

![Radio Occultation Technique and Representation of Tangent point (source: COSMIC)](image_url)
Since launch, the COSMIC mission has provided over three million GPS radio occultation atmospheric profiles to support science and operational applications. However, some spacecrafts have started to show degradation during this long run. Nevertheless, the data have already demonstrated their value for operational weather forecasting, hurricane forecasting, and investigations of the atmospheric boundary layer. COSMIC data has shown to be useful in improving the skill of weather prediction models. With the ability to penetrate deep into the lower troposphere using an advanced open loop tracking technique, the Formosat-3/COSMIC RO instruments have shown the capability to observe the structure of the tropical atmospheric boundary layer, providing valuable information on low-level atmospheric water vapor changes. COSMIC GPS RO data also have the potential to be of great benefit to climate studies due to their demonstrated high precision and global and diurnal sampling coverage.

There are many studies which emphasized the credibility of the GPS RO technique/data. Murphy et al. (2015) used airborne GPS RO refractivity profiles to observe the tropical storm environments. Poli et al. (2003) evaluated the Challenging Mini satellite Payload (CHAMP) RO refractivity using data assimilation office analyses and radiosondes. Tae et al. (2008) assimilated the GPS RO data from CHAMP and SAC-C missions over high southern latitudes with Mesoscale Model 5 (MM5) 4DVAR to assess the impact of the GPS RO data on analyses and short-range forecasts over the Antarctic. Jagadheesha et al. (2011) formulated the stability index based on atmospheric refractivity at ~500 hPa level and surface measurements of temperature, pressure, and humidity. There are several other studies which have used the GPS RO data to study the state of the atmosphere.

Further, GPS RO measurements are capable of providing climate records that are free from the constraints associated with any other space-borne and ground-based measurements. Foelsche et al. (2007) have studied the errors relevant for climatological investigation with GPS RO measurements from CHAMP and concluded that this data provides a valuable signature for climate monitoring. However, spatial and temporal coverage is the limitation of this technique. On the other hand, Indian Geostationary Satellite INSAT-3D under Indian Space Research Organization’s (ISRO) Indian National Satellite (INSAT) program provides continuous measurements of temperature and humidity data for the atmospheric studies. INSAT-3D sounder accuracy for near surface layer temperature is ~2K (RMSE) with a vertical resolution of 1-2 km (humidity accuracy of ~25% with a vertical resolution of ~2-3 km). The details of algorithm, payloads, and sensors used onboard (imager and sounders) and their channel frequencies are provided in technical documents of the ISRO and Indian Meteorological Department (2007 and 2010).

INSAT-3D have an 18-channel infrared sounder (plus a visible channel) along with a 6 channel imager. An algorithm is designed for retrieving vertical profiles of atmospheric temperature and moisture along with total column ozone content in the atmosphere from clear sky infrared radiances in different absorption bands observed through INSAT-3D. INSAT-3D sounder channels are similar to those in Geostationary Operational Environmental Satellite (GOES-12) system sounder. Hence, a present algorithm for INSAT-3D sounder is adapted from the operational high-resolution infra-red sounder (HIRS) and GOES algorithms developed by Cooperative Institute for Meteorological Satellite Studies (CIMSS), University of Wisconsin. Final accuracy of the retrieved profiles will largely depend upon the accuracy of the fast radiative transfer model used in the retrieval procedure. Therefore, a fast radiative transfer model needs to be properly validated and the forward model errors should be within instrument noise level.

The aim of the present study is to generate the high vertical resolution of RH profiles of INSAT-3D, using GPS RO data by applying statistical approach. The paper organized as follows. Section 2 describes the ANN approach and the algorithms used. In section 3, we have discussed the data and methodology used. Section 4 and 5 contains the results and discussions, respectively.

**ANN Models:**
An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way the biological nervous system works. The analysis can be used as a standalone application or as a complement to statistical analysis. ANN consists of an interconnected assembly of simple processing elements, known as nodes whose functionality is based on neurons (Pelliccia et al., 2010). ANN technique was also used earlier in meteorological (Badran et al., 1991; Butler et al., 1996), oceanographic and in satellite parameter retrievals (Krasnopolsky et al., 1995; Krasnopolsky and Schiller, 2003; Rumelhart., 1986). There are many advantages of the ANN technique compared to multiple regression (Borst et al., 1995). ANN analysis requires three sets of data, namely: training, verification, and validation. Training data set is used to train the model and verification sets to test
the model during the training process. Finally, ANN stores the trained model to predict the output using the input parameters that are helpful in validating the model. There are many ANN models which are in use for statistical analysis and prediction of different physical and mathematical problems. Among them, radial basis function, multi-layer perception, and the linear models are a few which are in broad usage.

A radial basis function network (RBF), has a hidden layer of radial units, each actually modelling a Gaussian response surface. Since these functions are nonlinear, it is not actually necessary to have more than one hidden layer to model any shape of function: sufficient radial units will always be enough to model any function. The remaining question is how to combine the hidden radial unit outputs into the network outputs. It turns out to be quite sufficient to use a linear combination of these outputs (i.e., a weighted sum of the Gaussians) to model any non-linear function. The standard RBF, therefore, has an output layer containing dot product units with identity activation function.

Multilayer Perceptron is perhaps the most popular network architecture in use today, due originally to Rumelhart and McClelland (1986). The units each perform a biased weighted sum of their inputs and pass this activation level through a transfer function to produce their output, and the units are arranged in a layered feedforward topology. The network thus has a simple interpretation as a form of the input-output model, with the weights and thresholds (biases) as the free parameters of the model. Such networks can model functions of almost arbitrary complexity, with the number of layers, and the number of units in each layer, determining the functional complexity. Important issues in Multilayer Perceptron (MLP) design include specification of the number of hidden layers and the number of units in these layers.

The number of input and output units are defined by the problem. The number of hidden units to use is far from clear. As good as a starting point as any is to use one hidden layer, with the number of units equal to half the sum of the number of input and output units. One major problem with the approach outlined above is that it doesn't actually minimize the error that we are really interested in - which is the expected error the network will make when new cases are submitted to it. In other words, the most desirable property of a network is its ability to generalize to new cases. In reality, the network is trained to minimize the error on the training set, and short of having a perfect and infinitely large training set, this is not the same thing as minimizing the error on the real error surface - the error surface of the underlying and unknown model. However, RBF networks have a number of advantages over MLPs. First, as previously stated, they can model any nonlinear function using a single hidden layer, which removes some design-decisions about numbers of layers. Second, the simple linear transformation in the output layer can be optimized fully using traditional linear modelling techniques, which are fast and do not suffer from problems such as local minima which plague MLP training techniques. RBF networks can, therefore, be trained extremely quickly (i.e., orders of magnitude faster than MLPs).

A neural network with no hidden layers, and an output with dot product synaptic function and identity activation function, actually implements a linear model. The weights correspond to the matrix and the thresholds to the bias vector. When the network is executed, it effectively multiplies the input by the weights matrix then adds the bias vector. The linear network provides a good benchmark against which to compare the performance of our neural networks. It is quite possible that a problem that is thought to be highly complex can actually be solved just as well by linear techniques as by neural networks. If you have only a small number of training cases, you are probably anyway not justified in using a more complex model. In the present work, we tried and tested all 3 models and found the linear model giving the least error and best result. Details presented in the sections below.

Data and Methodology:
In this approach, we used data over north Indian Ocean region and Indian subcontinent covering from 0° to 35°N latitude and 50°E to 100°E longitude. The spatial distribution of data collected from the RO technique from 2007 to 2009 is shown in figure 2.
There are 10,490 1-dimension variational (1-DVAR) temperature and humidity profiles available in which after series of quality and redundancy analysis we have picked 8878 profiles during the study period. We used temperature profiles along with the RH profiles to check the dependency of the RH on temperature variability. Theoretical analysis has shown that the temperature retrieved from GPS RO soundings is accurate to 0.5 K between 5-25 km altitude and to 0.2 K at tropopause (Kursinski et al., 1997), water vapor is to ~20-25% or better precision from near the surface to the upper troposphere in clear conditions (Gettelman et al., 2006). The accuracy of 1-DVAR temperature and humidity profiles from COSMIC, compared with radiosonde observations, is ~1 K, at lower levels and 0.5 K at upper levels (Anthes et al., 2000). However, if the temperature is known independently to within an accuracy of 2 K, estimates of water vapor pressure can be determined with an accuracy of about 0.5 mb. Original 1-DVAR temperature and humidity values are sampled at uneven intervals varying from 10 hPa to 0.1 hPa. For the ANN technique, we require temperature and humidity at fixed levels. Thus temperature and humidity values are interpolated at an interval of 5 hPa using simple linear interpolation method. Since RH profiles are having good correlation with the temperature profiles, we have run the model with and without including corresponding profiles of temperature with RH at each pressure level. The average error between interpolated temperature and humidity profiles and the actual 1-DVAR temperature and humidity profiles is within ~0.5 K and 20%, respectively.

The INSAT-3D Infra-red sounder provides temperature and humidity at 40 standard pressure levels from 1000 hPa to 0.1 hPa. Since the temperature and humidity data from GPS RO technique at most of the times are not available at the surface and because atmospheric processes are mainly confined to tropopause we used temperature and humidity from 950 hPa to 100 hPa. Thus, we have 171 temperature and 171 humidity (total 342) values with 5 hPa difference from 950 to 100 hPa. The INSAT-3D have 20 pressure levels from 950 hPa to 100 hPa (950, 920, 850, 750, 700, 670, 620, 570, 500, 475, 430, 400, 350, 300, 250, 200, 150, 135, 115 and 100 hPa). From the COSMIC interpolated data we selected temperature and humidity from INSAT-3D at 20 levels (from 950 to 100 hPa). The assumption involved in this simulation is that the sensor/algorithms of COSMIC and INSAT-3D are sensitive/accurate enough to record the same temperature and humidity at INSAT-3D pressure levels. Another point of concern is the non-comparability of the horizontal resolutions of COSMIC (200 to 300 km) and INSAT-3D (10 km). However, since COSMIC temperature and humidity measurements are well correlated with the point radiosonde observations (Anthes et al., 2000, Sun et al., 2010, Zhang et al., 2011, Foelsche et al., 2007), we presume that they will compare much better with INSAT-3D soundings with horizontal resolution of 10 km. By selection of COSMIC estimations, we miss a portion of about 500 m above the ground unlike radiosondes, but we compromise to miss this information in view of a large number of COSMIC observations. On any day the COSMIC has ~4000 occultations globally (over land and ocean) compared to only ~1000 radio soundings. Besides, radiosondes are not available over the ocean except during ship observations. One of the requirements of the ANN technique is to have a large number of data set
covering all the circumstances to get the best outcome. Thus we did not use the actual measurements from radiosondes because of the less number of observations spatially.

**Results and Discussions:-**
Statistical analysis has been carried out to estimate the errors involved in the ANN model. We run the model twice: (1) simulated RH profiles as input (2) simulated temperature and RH both as input. The Absolute Error Mean (AEM: average of absolute differences between estimated and observed values), absolute mean percentage error (AMPE: the percentage of the AEM to data mean), standard deviation (SD) errors in estimations (ESD), SD ratio (SDR: ratio of ESD to data SD), Root Mean Square Difference (RMSD) and scatter index (SI: ratio of RMSD to mean of in situ observations), for the training, verification and prediction data sets for two models are shown in table 1 and table 2. The range of GPS RO RH values varies from 0.1% to 100% with a mean value of 50.62% and SD of 28.59%. The AEM, AMPE, SDR, RMSD, and SI are 2.49, 5.32, 0.12, 5.53 and 0.1, respectively for validation data set of the model run with temperature and RH profiles. There is no significant difference observed between the two models. Results from the random selection process did not change significantly from the year wise selection. The correlation coefficient (CC) and SDR between the estimated and actual observations for training, testing and validation profiles at different levels are shown in figure 3(a) to 3(c), respectively.

![Figure 3](image-url)

**Figure 3:** Correlation coefficient (CC) and the standard deviation ratio (SDR) for (a) training, (b) testing, and (c) validation.

The SDR is shown in solid line at the left side and CC is shown in dashed line on the right side of the each figure. As mentioned earlier the ANN model input layer has 20 pressure levels and these layers are also repeated in the output layer of the model. As expected, the model predicted exactly the same value as output at those 20 pressure levels. As a result, the CC and SDR are perfectly one and zero, respectively at those pressure levels for all the three datasets. The CC is greater than 0.94 and the SDR is less than 0.35 throughout the pressure levels for training, testing and validation datasets. The statistics of the ANN model are shown in the below tables 1 and 2.
Table 1: Statistics of ANN model without temperature

<table>
<thead>
<tr>
<th>Case</th>
<th>No. of obs.</th>
<th>GPS RO RH (%)</th>
<th>Estimated Relative Humidity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
<td>Max</td>
</tr>
<tr>
<td>Train</td>
<td>3519</td>
<td>100</td>
<td>0.1</td>
</tr>
<tr>
<td>Test</td>
<td>2818</td>
<td>100</td>
<td>0.1</td>
</tr>
<tr>
<td>Validation</td>
<td>2541</td>
<td>100</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 2: Statistics of ANN model with temperature

<table>
<thead>
<tr>
<th>Case</th>
<th>No. of obs.</th>
<th>GPS RO RH (%)</th>
<th>Estimated Relative Humidity (%) with temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
<td>Max</td>
</tr>
<tr>
<td>Train</td>
<td>3519</td>
<td>100</td>
<td>0.1</td>
</tr>
<tr>
<td>Test</td>
<td>2818</td>
<td>100</td>
<td>0.1</td>
</tr>
<tr>
<td>Validation</td>
<td>2541</td>
<td>100</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Vertical variability of statistical parameters:
We also selected three random samples of estimated RH profiles from the validation dataset. The selected profiles are located over the Arabian Sea (16.17° N; 51.59° E), Indian mainland (18.95°; 76.48°E) and over the Bay of Bengal (10.98°N; 97.02°E) having tangent points, respectively. The error between 1-DVAR RH and ANN estimated RH of these profiles is shown in figures 4(a) to 4(c), respectively. The maximum error between estimated and 1-DVAR RH is around ±10%, which allows the temperature and RH accuracy within theoretical margin over the entire range at all three locations. This shows the perfect match between the estimated and actual profiles demonstrates the accuracy of the estimation.

Figure 4: The error between 1D-VAR and ANN estimated RH at (a) 16.17° N; 51.59° E. (b) 18.95°N; 76.48°E. (c) 10.98°N; 97.02°E
Summary and Conclusion:-
We have used a machine learning approach to improve the vertical resolution of tropospheric relative humidity profiles (RH) obtained from the geostationary satellite. We developed a statistical model to enhance the availability of data from 20 pressures levels to 171 pressure levels in the troposphere. The data used in the present study from the COSMIC mission was derived using the refractivity of GPS signal passing through the limb of the atmosphere. We used the tropospheric temperature and RH data from 950 hPa to 100 hPa with 5 hPa resolution with simple linear interpolation. Thus we have 171 interpolated pressure intervals in the selected profile. We then selected 20 pressure levels from COSMIC data collocated with INSAT-3D pressure levels. We presume that the INSAT-3D sensors/algorithms have same sensitivity as COSMIC. We compromise with the low horizontal resolution of COSMIC data and also neglecting the 500 m above the ground while choosing the COSMIC data. The model was developed using 3 years RH data during 2007-2009. We selected best algorithm among 3 algorithms those used to develop the ANN model. Each individual algorithm (RBF, MLP, and the Linear) was run twice, with and without including the corresponding temperature profiles along with the RH profiles. Thus, we have selected a linear model as the best one and we also observed the improvement in the model by using the temperature profiles along with the RH profiles. RH data from a year each was used to train, test and to validate the algorithm, respectively. The average RMSD between estimated and actual observations was 5.53 for the validation data set. The accuracy of our estimation was within the accuracy limits of 1-DVAR humidity profiles (more or less 10 % at lower levels and 15 to 20 % at upper levels). The estimated profiles coincide well with the actual profiles with a CC of greater than 0.94 and SI of less than 0.1 throughout the pressure levels. The model performed well at 3 randomly selected locations over the Indian mainland, Bay of Bengal and the Arabian Sea. Thus, the ANN approach is a better method to increase the vertical resolution of RH profiles obtained from geostationary satellites. However, the ultimate resolution and accuracy of the temperature and RH from INSAT-3D depends on the interpolation methods used and the radiative transfer algorithms used in retrieving the data from the sensors.

Acknowledgments:-
Authors wish to acknowledge the CDAAC for providing the COSMIC GPS RO data. We also thank M.M.Ali, Indian Institute of Tropical Meteorology (IITM), Pune, Neerja Sharma, Satellite Applications Center (SAC), ISRO, Ahmedabad and Prof. Vijayabhaskar Rao, Department of Physics, Sri Venkateswara University, Tirupati for their valuable guidance and support to carry out the work and to improve the quality of the work.

References:-