Data Warehousing and Data Mining Applications for studies on distribution of water vapour observed by COSMIC GPS Radio Occultations over Indian Regions (8° to 38°N and 66° to 98°E)

Venkata Sheshanna Kongara[1], Dr. D.Punyasesudu[2]

[1]*Research Scholar, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India.*

[2] *Professor, Department of Physics, Rayalaseema University, Kurnool, Andhra Pradesh, India.*

Abstract— **Atmospheric water vapour plays an important role in radio communications for both terrestrial and earth space communication systems as well as for global climate change studies. Hence the water vapour data and its information plays vital role for scientists and researchers to evaluate the hidden patterns, trends for analysis and forecasting. Global Positioning System Radio Occultation (GPS RO) is a new technique for retrieving the properties of the atmospheric conditions and the data retrieved from GPS RO contributes significant level of information for research and development. Formosa Satellite mission-3/Constellation Observing System for Meteorology, Ionosphere, and Climate (FORMOSAT-3/COSMIC) is a joint Taiwan-United States (U.S.) mission for monitoring global weather; and providing the various atmospheric parameters like vertical profiles of bending angles, refractivity, electron density, temperature, pressure, and water vapor in the atmosphere. The datasets are managed and produced by CDAAC (COSMIC Data Analysis and Archival Center) and made available in multidimensional data formats of NetCDF (Network Common Data Form) for further analysis. As part of the study considered the most recent RO water vapor datasets for various locations of Indian region (latitudes between 8° to 38°N, and longitudes 66° to 98°E). There are different statistical and scientific methods being used to process these datasets, however they have different data level limitations, challenges in retrieving, storing, managing and exploration of these structured and un-structured data; which are also very large in size, therefore the approaches of data assessment and conversion involves in time consuming, expensive, and complex to mitigate the accurate results. Presently there are extraordinary prospects in the Information Technology to analyze these data and explore in the better manner. The data warehousing and data mining applications are the most emerging features which are endorsed that information to be accessed easily and efficiently. Hence the GPS RO water vapour datasets assimilation and classification have been carried out by using data warehousing and data mining techniques and applied better data analytics for precise results. The observational results of RO water vapour distribution patterns and trends are presented in the form of different zones of the atmosphere like Troposphere, Middle and Low levels of Troposphere and Stratosphere about altitude 40 KM with different seasons for better decision supporting systems.**

Index Terms—Atmospheric Water Vapour, Radiosonde, Radio Communications, Climate Change, Hidden Patterns, Global Positioning System, Radio occultation, Data Warehousing, Data Mining, Decision Supporting Systems

I. INTRODUCTION

1.1. Importance of the GPS Radio Occultation water vapour and objectives of the present work

The Water vapor (WV) is the most significant greenhouse gas and plays an important role in radio communications for both terrestrial and earth space communication systems as well as for global climate change studies and weather forecasting. The Water Vapour and Oxygen needs to be considered in the effective estimation performance of earth space communication connections. In the clear air the Water Vapour and oxygen absorption causes frequency dependant signal attenuation, propagation delay, ray bending and medium noise. Therefore it is an essential area of practice and research about Atmospheric water vapour [1] [2]; consequently the water vapour datasets usage and the method of data computation are the key factors in evaluation of atmospheric water vapour trends and usage. There are different statistical and scientific methods being used to process the GPS Radio occultation datasets and measure the correlated innovations [3] [4] [5]. However due to the data limitations it is a challenging task in retrieving, storing, managing and exploration of these structured and un-structured data which are very large in size. Hence the approaches of data maintenance and conversion involves in time consuming, expensive, and complex to mitigate the accurate results; In recent years the data warehousing and data mining applications are the most emerging technologies with powerful data managing features, which are endorsed that information can be accessed easily and efficiently to build and deploying data driven analytics for better knowledge in assisting the right decision making activities [6] [7]. The "Data warehouse" is pretend by W. H. Inmon [8] in the book "Building the Data Warehouse" (1996). He gave the first definition of data warehouse as "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process". At the outset the data warehouse is a massive collection of storage area which serves as a centralized repository of all the data collected from various atmospheric parameters or processing entities from the large combination and managed systematically for meaningful information and analysis for effective decision supporting.

Data mining is defined as a technique to get useful, previously unknown information from databases and data warehouses. Data mining techniques include association Rules, clustering, classification, prediction, sequence mining, web mining, text mining and spatial data mining [9]. This paper reveals various analytics of the recent RO water vapour distribution studies over India by using these applications. As part of the study obtained the COSMIC GPS RO vertical profiles of bending angles, refractivity, electron density, temperature, pressure, and water vapor in the atmosphere of various locations of Indian region (latitudes between 8° to 38°N, and longitudes 66° to 98°E). The GPS RO water vapour datasets assimilation and classification has been carried out by using data warehousing and data mining techniques and applied better data analytics for precise results. The observational results of RO water vapour distribution patterns and trends are presented in the form of different zones of the atmosphere like Troposphere, Middle and Low levels of Troposphere and Stratosphere with altitudes from 0 to 40 KM vertical with different seasons for better decision supporting systems.

1.2. The structure of the Earth's Atmosphere

The vertical distribution of temperature, pressure, density and composition of atmosphere constitutes atmospheric structure. These quantities also vary with season and location in latitude and Longitude, as well as from night to day; however under the topic of atmospheric structure, the focus is on the average variations with height above sea level. The envelope of gas surrounding the earth changes from the ground up. There are four distinct horizontal layers have been identified based on thermal and convective characteristics (temperature changes), chemical composition, movement, and density [10]; consequently from the surface of the Earth upwards the layers are Troposphere, Stratosphere, Mesosphere and Thermosphere. In this paper we have presented the studies on water vapour distribution with different atmospheric parameters and its analytical observation of Troposphere, Middle and Low levels of Troposphere and Stratosphere over India.

1.3. FORMOSAT-3/COSMIC: Global Positioning System Radio Occultation (GPS RO)

Formosa Satellite mission-3/Constellation Observing System for Meteorology, Ionosphere, and Climate (FORMOSAT-3/COSMIC) is a joint Taiwan-United States (U.S.) mission for monitoring global weather and demonstrate the use of GPS RO data in operational weather prediction and climate analysis. This project is targeted to place six micro-satellites into six different orbits at 700~800 kilometer above the earth ground. These satellites orbit around the earth to form a low-earth-orbit constellation that will receive signals transmitted by the 24 US GPS satellites. The satellite observation covers the entire global atmosphere and ionosphere, providing over 2,500 global sounding data per day. These data distribute uniformly over the earth's atmosphere. The global climate information collection and analysis can be completed in three hours while the sounding data will be updated every 90 minutes [11].

The Global Positioning System Radio Occultation (GPS RO) technique is for retrieving the properties of the atmospheric conditions for numerical weather models and climate change studies [12]. It enables the measurements of the global atmospheric density structure from different meteorological condition. This is a quite new technique which involves a GPS receivers placed on low Earth Orbit (LEO) Satellite to sound the earth's atmosphere. This limb sounding technique works under all weather conditions due to the insensitivity of the GPS Signal wave length to scattering by clouds, aerosols and precipitation, with a vertical resolution about 1 km but a poor horizontal resolution about 200 km [13].Then, such GPS LEO System as shown in the figure B is exploited to obtain profiles of refractivity, temperature; pressure and humidity in the atmosphere at the global scale are extremely valuable for weather forecasting and research, climate monitoring and space weather forecasting.

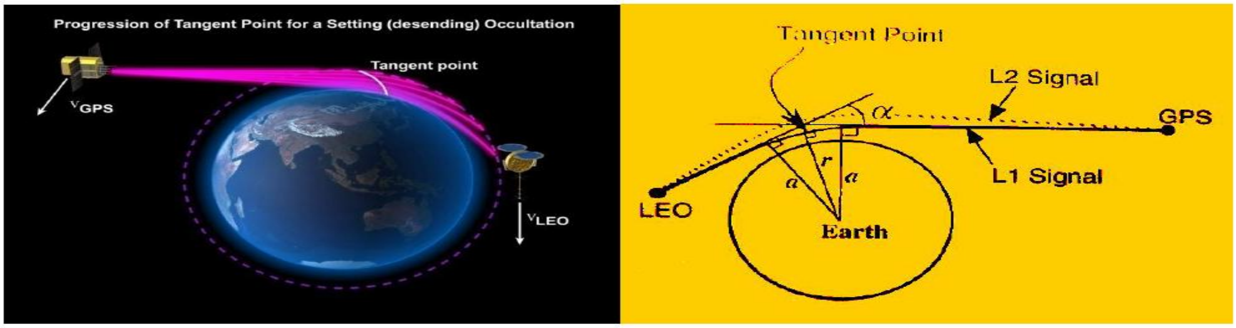


Figure 1: Radio Occultation technique and representation of the tangent point (Image Source: COSMIC)

As part of this study we have considered the most recent RO water vapor datasets for various locations of Indian region (latitudes between 8° to 38°N, and longitudes 66° to 98°E).

The rest of the paper is organized as follows. The literature review and related work for GPS RO water vapour studies are discussed in section II. The details of the GPS RO water vapour data evaluation, processing, integration and classification are described in Section III. The data analysis and results are presented in section IV and concluded the paper with accomplishment in the study and future work in Section V.

II. RELATED WORK

Xiaoguang Tan et al [6] expressed that data warehouse as a new kind of Artificial Intelligence (AI) system that combines database and meteorological graphics technology. It helps forecasters accumulate, manage and use their knowledge in operational forecast. It is a new generation of DSS. Obviously data warehouse will not become whole system of forecaster's workbench, because operational forecast mission is very complex. But it is a system to help forecasters accumulate, manage and use their knowledge

V. Sreenivasarao et al [7] discussed an overview of scientific data warehouse and OLAP technologies, with an emphasis on their data warehousing requirements, The methods that were used include the efficient computation of data cubes by integration of MOLAP and ROLAP techniques, the integration of data cube methods with dimension relevance analysis and data dispersion analysis for concept description and data cube based multi-level association, classification, prediction and clustering techniques

Folorunsho Olaiya et al [9] investigated the use of data mining techniques in forecasting maximum temperature, rainfall, evaporation and wind speed. This was carried out using Artificial Neural Network and Decision Tree algorithms and meteorological data. A data model for the meteorological data was developed and this was used to train the classifier algorithms. The performances of these algorithms were compared using standard performance metrics, and the algorithm which gave the best results used to generate classification rules for the mean weather variables. A predictive Neural Network model was also developed for the weather prediction program and the results compared with actual weather data for the predicted periods

Abhineet Shyam et al [14] Artificial neural network (ANN) technique have been used to derive water vapour pressure profiles in the troposphere from radio occultation data over India and adjoining region. Fully connected 3 layer networks with 1 hidden layer have been constructed and standard back propagation algorithm is used to train the network. While month, latitude and vertical profiles of refractivity/bending angle constitute the input vector. The water vapour partial pressure profiles forms the output vector. Only the moisture-laden summer monsoon months of June, July, Aug, Sept of 2010 have been consider for developing the retrieval algorithm. There are 2120 input and output pairs, out of which 1696 pairs from the training set and remaining pairs constitute the validation set. The retrieved profiles of water vapour pressure in the validation set are 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 the root mean square error (RMSE).The statistics show that the better performance of the algorithm and the retrieved data has found to be highly consistent.

Neerja Sharma et al [15] discussed about the importance of the GPS RO data that assimilating radio occultation refractivity in numerical weather predication models improved the weather forecasts significantly. The main focus of their study is explains that how the RO refractivity can be used as an indicator of atmospheric instability. Hence as part of their study the stability and moisture parameters are observed from COSMIC refractivity and derived temperature and humidity profiles over India and the surrounding region during May-August 2007.Those parameters have been analyzed with collected three hourly rainfall estimated from the Tropical Rainfall Measurement Mission (TRMM) and found that stability and moisture parameters can be associated with different degrees of probability of rainfall. The results are concluded with the importance of atmospheric refractivity and moisture parameters from the predication of rainfall occurrence.

P. Kishore et al [16] studied about global (50 S–50 N) distribution of water vapor and investigated using COSMIC GPS RO measurements for the period of September 2006–December 2009, As a first step towards building the COSMIC water vapor climatology, reasonable global maps of seasonal mean of specific humidity at different pressure levels. The detailed comparisons have been made between COSMIC and high resolution GPS radiosonde measurements across 13 tropical stations and model outputs (ERA-Interim, NCEP, and JRA-25 reanalyses datasets). In comparison within dependent techniques like radiosonde (Vaisala) and it is found that COSMIC GPS RO wet profiles are accurate up to 7–8km (assuming radiosonde as standard technique).

C.J. Johnny et al [17] discussed about the spatial and temporal distribution of water vapour in the upper troposphere and the lower stratosphere (UTLS) region over India including Arabian Sea and Bay of Bengal is presented using COSMIC/FORMOSAT3 radio occultation measurements. Water vapour plays a crucial role in many aspects of UTLS chemistry. The influence of Asian summer monsoon can be seen in the seasonal patterns of water vapour in the UTLS region. It is observed that water vapour in the lower stratosphere follows the seasonal cycle in upper tropospheric water vapour with a time lag of one month. The time scale of cross tropopause trasnport of air mass in the region is also discussed.

D. Narayana Rao et al [18] conducted studies on Validation of the COSMIC Radio Occultation Data over Gadanki(13.48 o N, 79.2 o E).As part of the study, Radiosonde (Vaisala RS-80H and RS-92 Type) specially launched for validation of OCSMIC RO data from Gadanki, a tropical site in India during July 2006 to March 2007. The atmospheric parameters (pressure, temperature, water vapour and horizontal winds) were determined. They have also used post-processed GPS RO data obtained by COSMIC Satellite from July 2006 to March 2007, which are processed by TAAC data center with temperature and water vapour profiles (1D- var retrieval) in the troposphere and lower stratosphere. And another set of Temperature data from co-located Nd: YAG Rayleigh Lidar also considered, which provides temperature information right from 30 to 80 km to validate the temperature profiles from 30 to 40 km. The results and discussions are presented in the form of i) Comparison of water vapour between COSMIC and Radiosonde (Wet Region), ii) Comparison of Temperature between COSMIC and Radiosonde in Upper Troposphere and Lower Stratosphere (UTLS) (Dry Region), and iii) Comparison of temperature between COSMIC and Rayleigh Lidar in Middle and Upper Stratosphere.

III. METHODOLOGY

3.1. Source of the data:

FORMOSAT-3/COSMIC uses a constellation of six remote sensing micro-satellites to collect atmospheric data to demonstrate their uses and applications in weather prediction, ionosphere, climate and gravity research. The mission sponsors encourage the use of these data in a wide variety of applications and research projects. The datasets are managed and produced by CDAAC (COSMIC Data Analysis and Archival Center). As part of the study we have downloaded the GPS Radio Occultations level 2 product profiles from the CDAAC website [19]. The datasets are made available in multidimensional data formats of NetCDF (Network Common Data Form) covering the period from 1996-2013 and the wet atmospheric profiles (wetprf) of refractivity, temperature and water vapor pressure with altitude range is 0-40 km at 100 meter vertical resolution, these profiles are based on 1D variation analysis using ECMWF low resolution analysis data. We have considered the recent radio occultation datasets for various positions of Indian region (latitudes between 8° to 38°N, and longitudes 66° to 98°E) from January to December 2013 with different mean Sea Level geometric height in Kilo meters and Pressure levels in Mb, Figure B shows the COSMIC data availability for Indian region latitude, 8° to 38° north and longitude, 66° to 98° east from 1 to 365 Days of 2014 and 1 to 150 Days of 2015.

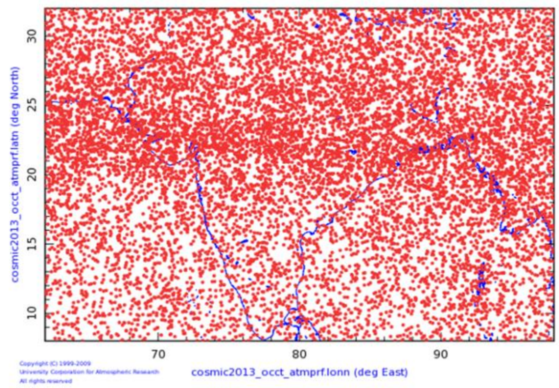


Figure 2: COSMIC data availability in the year 2014 & 2015 for Indian region (latitude 8° to 38° N and longitude 66° to 98° E)

3.2 Data pre-processing:

Data preprocessing is a fundamental stage of data analysis for better performance and good quality of results. In general missing data are unintended and uncontrolled by the researchers, but the overall result is that the observed data cannot be analyzed because of the incompleteness of the data sets. Missing Values and its problems are very common in the data cleaning process. Several methods have been proposed [20] [21] so as to process missing data in datasets and avoid problems caused by it. In the COSMIC GPS RO source data also there are some missing values for different parameters for different MSL Altitude, Pressure, Latitude and longitudes. The CDAAC filled those missing data sets with 999 in their data, Hence we have handled those missing values filled with précised one by using cluster based missing values algorithm to process the GPS RO water vapour data for better results.

Radio Occultation data evaluation and Integrated Water Vapour Concentration:

The GPS Radio Occultation (RO) and Radiosonde (RS) data assessment and computation of integrated water vapour have been carried by using different functions, B.-R. Wang et al [22] discussed on the atmospheric refractivity of radio occultation was calculated using function (1)

$$N = 77.6\frac{P}{T} + 3.73 \times 10^5 \frac{e}{T^2} \quad (1)$$

Where N Refractivity Index, T is temperature in Kelvin and P is total air pressure and e is the water vapour pressure hpa, respectively. Hence as part of the study we have assessed the obtained radio occultation data by using this function; and considered water vapour pressure and temperature parameters to compute the Integrated Water Vapour Concentration (M) for further analysis. Khamphoui et al [23] discussed on the impact of water vapor in atmosphere and clear sky attenuation is mainly due to the absorption caused by water vapour and oxygen molecules. It increases with the relative humidity as well as the temperature and alleged that the water vapour concentration is strong function temperature and humidity. In their studies they have used most revised version of the equation to calculate the water vapour concentration (M) as specified in function (2).

Water Vapour Concentration (M) in g/m^3 $= 216.7 \times \frac{e}{T}$ (2) Where e is water vapour pressure hPa and the T is temperature in Kelvin. Hence we are using this function our calculation and the calculated water vapour concentration datasets are arranged in the dimensional tables of the data warehouse for our analyses, which are may be reusable for the further studies. We have analyzed all the datasets with various analytics, the detailed results are presented in section IV.

3.3. GPS RO Water Vapour Data Integration and Classification:

As acknowledged in the Introduction of this paper data warehousing and data mining applications are the most appropriate for atmospheric water vapour studies. The dimensional modeling plays an important role in data warehouse design. Hence we analyzed GPS RO source data and its parameters, then applied transformational rules and built a star schema based GPS RO Water Vapour dimensional model., The dimension members which are including with Latitude, Longitude, Altitude, Temperature, Pressure, Vapor Pressure, Refractivity, Time Period and Location and its attributes are processed with calculated measures of min, mean and max values of the set of GPS RO parameters into star based dimensional model (i.e.) each entity parameters comes up with a context; which is a dimension table and qualifies a measurable number (Fact Table), each fact table is associated in centre of the schema surrounded by multiple dimension tables like a star as shown in Figure 3;

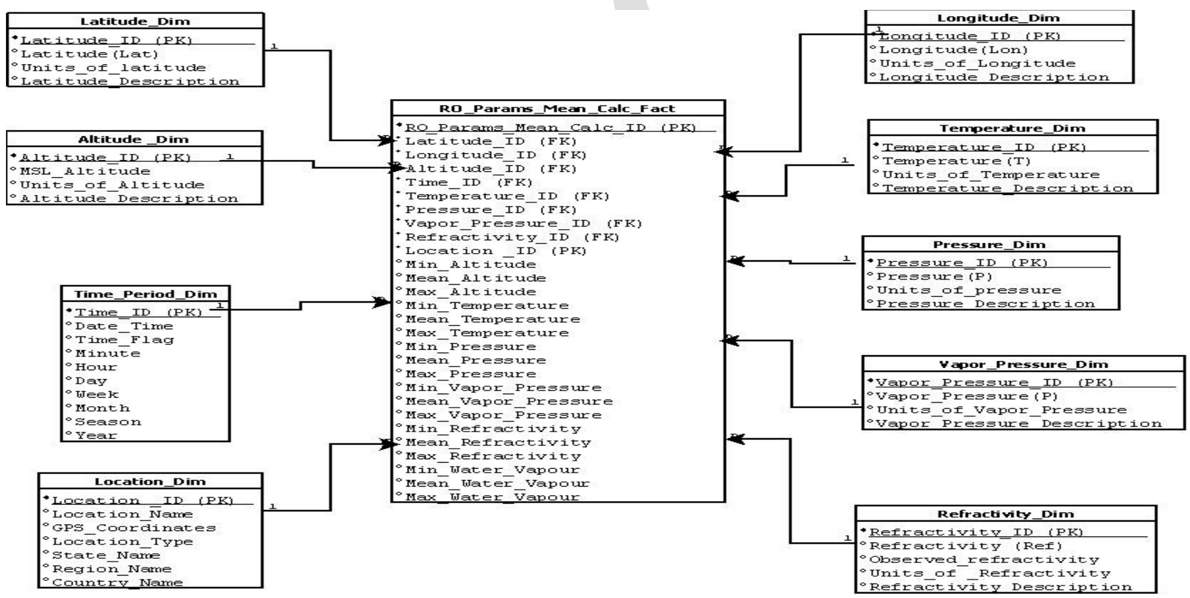


Figure 3: Data warehousing star schema based dimensional model for GPS RO water vapour data.

Based on the data availability we have selected the radio occultations data for Indian region latitude, 8° to 38° north and longitude, 66° to 98° east from 1 to 365 Days of 2014 and 1 to 150 Days of 2015. Extracted the rules and map them into different dimensions and measures to produce understandable and useful knowledge for analysis Table I: illustrates the list of attributes and the units of measurements are used for GPS Radio Occultation data computations. The seasons are categorized based on the different climatic conditions of India for different months that follows like winter (Dec, Jan, Feb), Summer(Mar, Apr, May), Monsoon(Jun, Jul, Aug, Sep) and Post-Monsoon(Oct, Nov)

Parameter_ name	Units_of_ measures	Attributes Description
Latitude(Lat)	degrees_N	latitude, 8° to 38° North
Longitude(Lon)	degrees_E	longitude, 66° to 98° East
MSL_Altitude	KM	Mean Sea Level geometric height in Kilo meters
Temperature(T)	Degree C	Celsius
Pressure(P)	millibars	pressure level
Vapor_Pressure(Vp)	millibars	H2O vapor pressure
Water Vapour	g/m3	Grams/Cubic Meter

Table I: Different parameters and unit of measures used for GPS Radio Occultation data computations.

Classification of the vertical structure of the atmosphere in accordance of temperature:

The vertical temperature profile is important parameter to understand the actual concentration of the atmospheric constituents and atmospheric pressure at a particular height from the sea level in the earth's atmosphere. The scientist divided the atmosphere into different layers according to major changes in temperature. The Figure D illustrates the classification of the layers of atmosphere based on the temperature changes with different altitudes and the pressure levels.

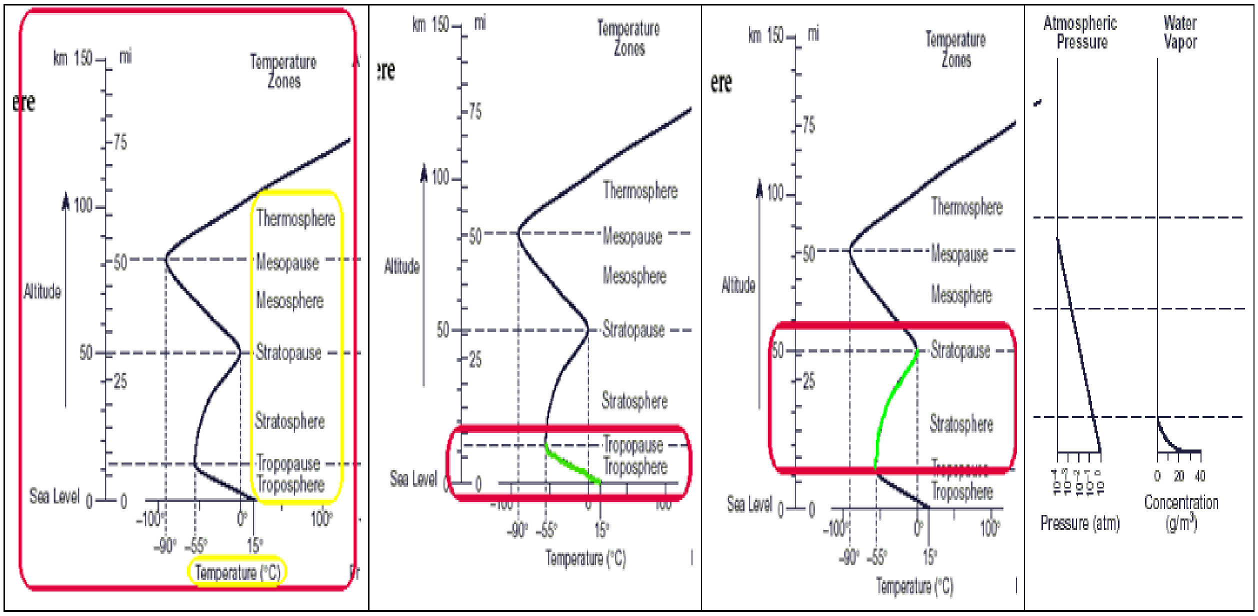


Figure 4: The Layers of the Atmosphere (Image from regentsprep.org)

Data mining based classification is a supervised machine learning technique used to build a model, once model has been built and applied to unseen data for prediction of class label. Building the accurate and efficient classifiers for large databases is one of the essential tasks of data mining and machine learning research. Building effective classification systems is one of the central tasks of data mining. Many different types of classification techniques are available that includes Decision Trees, Naive-Bayesian methods, Neural Networks, Logistic Regression. So we have conducted a comparative review on various methods of classification and measured the accuracy of the results. It shows that using Neural Networks based classifier obtains the best result among the other methods to classify the GPS RO water vapour data.

Following are the different classification zones being considered for our study with different Mean Sea Level geometric height in Kilo Meters.

MSL Altitude_ Classification	Atmospheric Zone	Description
0-10 KM	Tropospheric Zone	Temperature varies from ~10 to -60 deg C at the tropopause Air pressure and the density of the air also decrease with altitude (pressure ranges from 1000 to 200 millibars)
11-19 KM	UTLS Zone	The height of the bottom of the stratosphere varies with latitude and with the season. UTLS is Upper Troposphere and Lower Stratosphere.
20-40 KM	Stratospheric Zone	Temperature varies from ~0 deg to -70 deg C

Table II: Classification of different Atmospheric zones.

Graphical representation of various atmospheric parameters variations with Altitude about 40 KM:

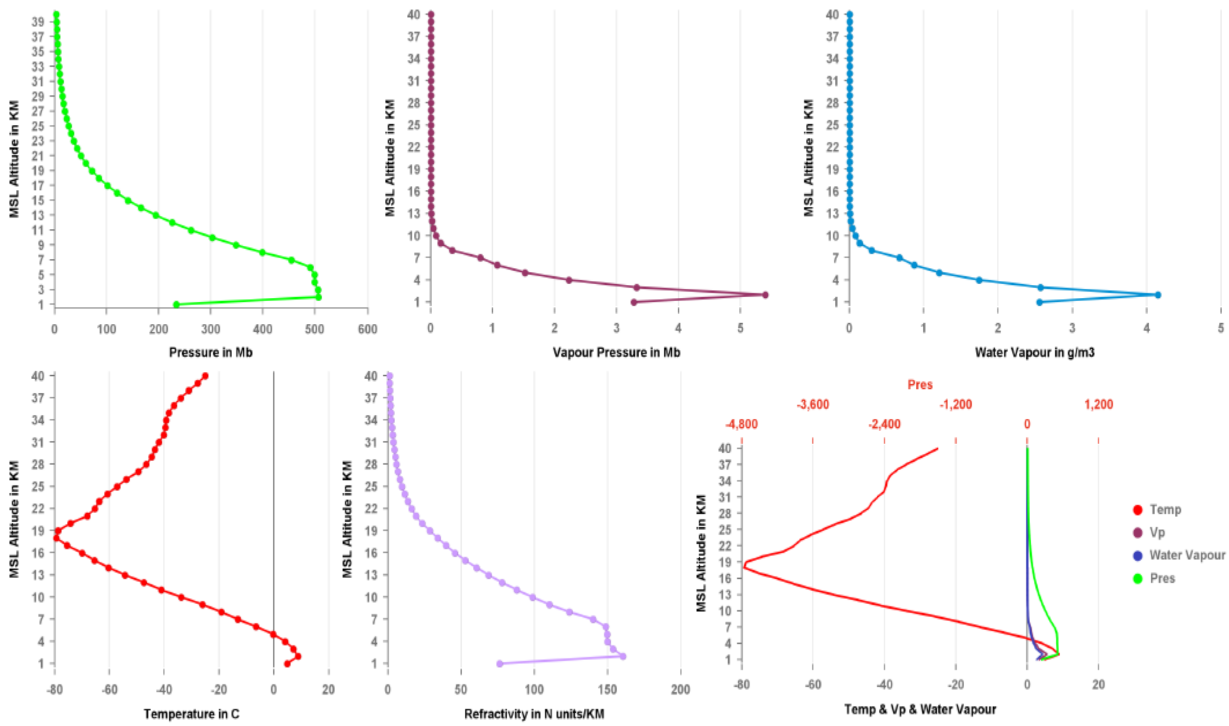


Figure 5: graphical representation of the different atmospheric parameter's changes with MSL Altitude about 0-40 KM in the atmosphere.

IV. DATA ANALYSIS AND RESULTS

The Concentration of water vapour, the third most abundant atmospheric gas, is highly variable, mainly because the maximum possible vapour pressure is a strong function of temperature. Once the air is saturated with water vapour, any reduction in temperature will cause condensation and No other constituent gas changes phase under normal atmospheric conditions. Water vapour is responsible for most of the weather people experience, condensed into clouds and precipitation. The concentration of water vapour in the atmosphere is temperature limited, and winds can easily transport vapour thousands of kilometers, this gas is highly variable in space and time. Its concentration is 0-4% by volume .Almost all water vapour in the atmosphere is confined to the troposphere, where clouds and storms occur. Low temperatures at the top of the troposphere (-50 to 70°C) assure that condensation will remove all but trace amounts of vapour before it can reach the stratosphere, the layers of atmosphere immediately above the troposphere. The principal source of water vapour is evaporation from the oceans, mostly in the tropics where the temperature is relatively high. Evaporation from lakes and soils, and transpiration from plants are other important source of vapour. Precipitation removes water vapour from the atmosphere. Water vapour is naturally occurring greenhouse gas, that is, it absorbs little incoming solar radiation (visible wavelengths), but absorbs significant outgoing long wave radiation (infrared wavelengths) [24].

In this study the GPS RO water vapour data analytics provide the details on the mean water vapour distribution patterns and trends are presented in the form of different zones of the atmosphere like Troposphere, Middle and Low levels of Troposphere and Stratosphere about altitude 40 KM with different seasonal distribution for better decision supporting systems.

The detailed results shown in the below sections A, B, C, D and E as follows

A). Seasonal Distribution of Water Vapour for 2014 and 2015 with different atmospheric zones

B). Seasonal % of Water Vapour Distribution for 2014 with different atmospheric zones

C). Seasonal % of Water Vapour Distribution for 2015 with different atmospheric zones

D). Seasonal Water Vapour Distribution % of Increase or Decrease Trends, altitude about 0-40 KM in the atmosphere

E). Seasonal Distribution of Water Vapour for 2014 and 2015 with different atmospheric zones and for various fixed pressure levels.

A). Seasonal Distribution of Water Vapour for 2014 and 2015 with different atmospheric zones:

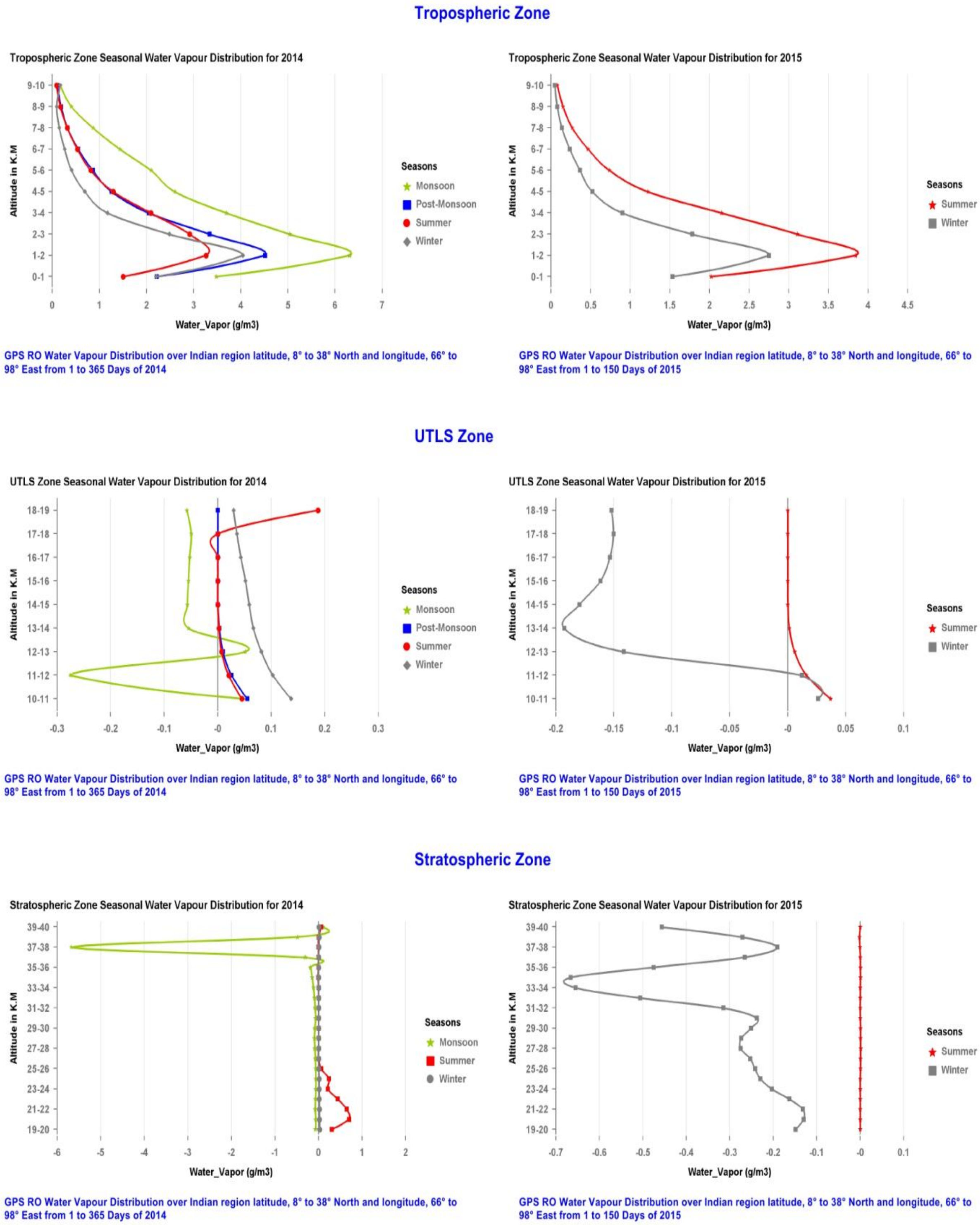
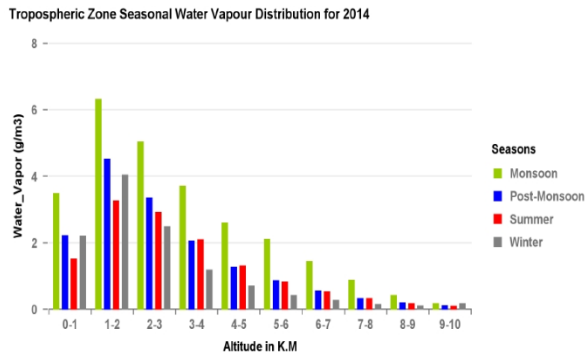


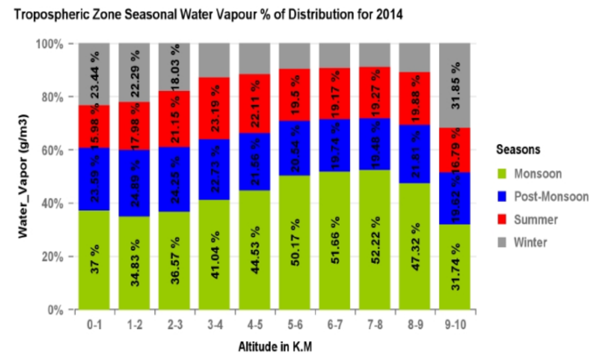
Figure 6: GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014 and 1 to 150 days of 2015 changes with MSL Altitude about 0-40 KM in the atmosphere.

B). Seasonal % of Water Vapour Distribution for 2014 with different atmospheric zones:

Tropospheric Zone

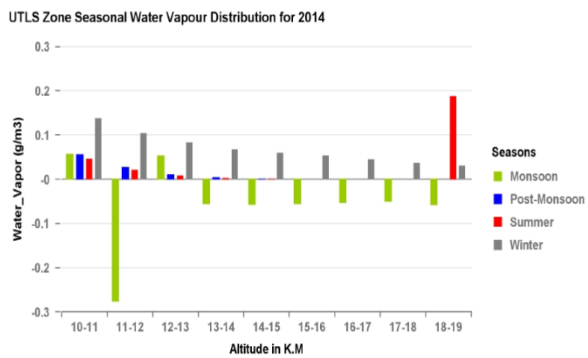


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

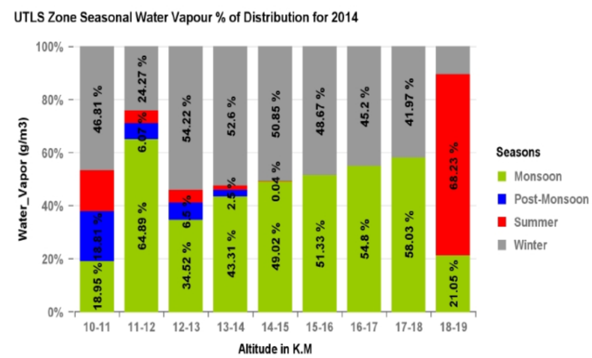


GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

UTLS Zone

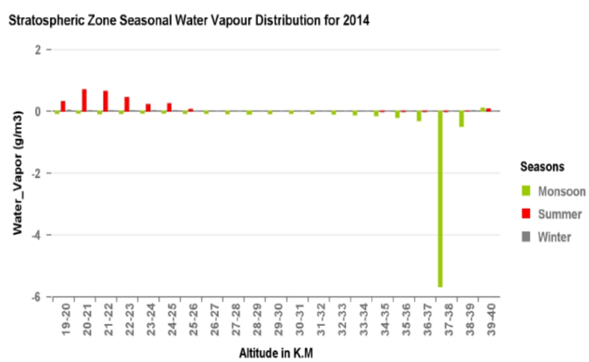


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

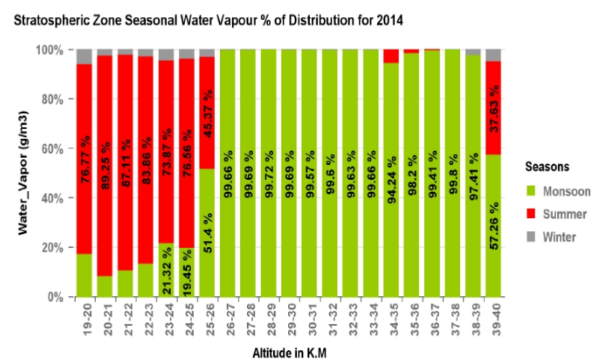


GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

Stratospheric Zone



GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

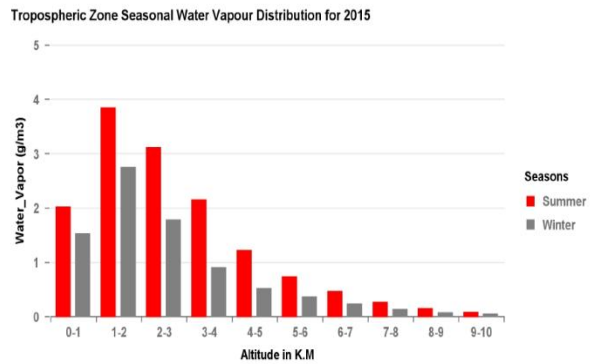


GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014

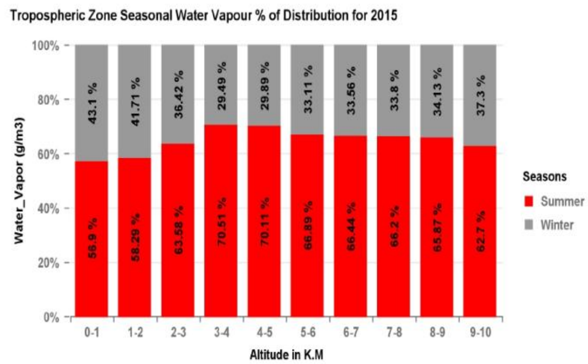
Figure 7: GPS RO Water Vapour Distribution % over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014 and changes with MSL Altitude about 0-40 KM in the atmosphere for different seasons

C). Seasonal % of Water Vapour Distribution for 2015 with different atmospheric zones:

Tropospheric Zone

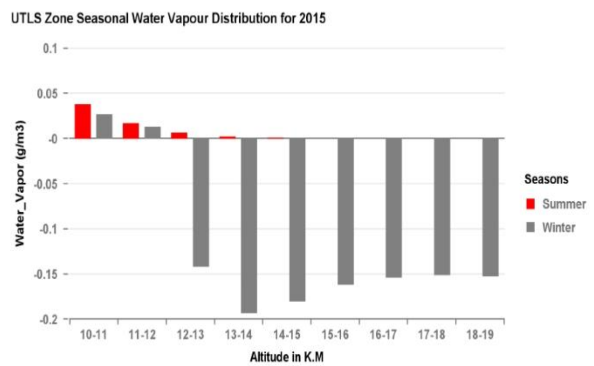


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015

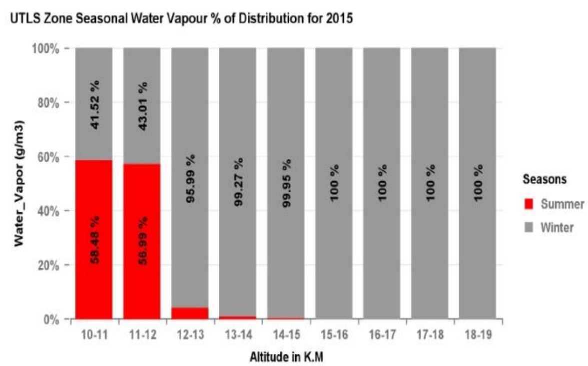


GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015

UTLS Zone

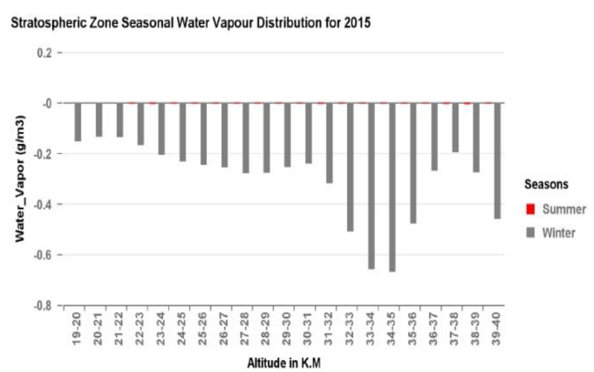


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015

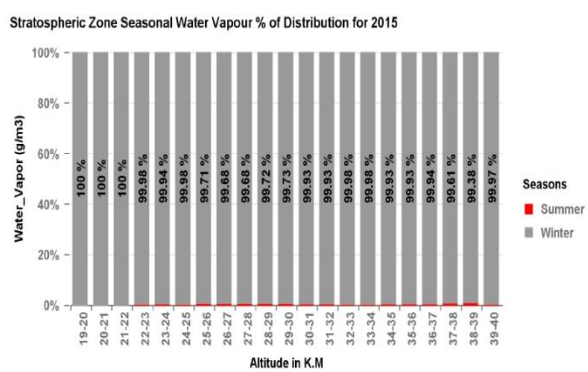


GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015

Stratospheric Zone



GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015



GPS RO Water Vapour % of Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015

Figure 8: GPS RO Water Vapour Distribution % over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 days of 2015 changes with MSL Altitude about 0-40 KM in the atmosphere for different seasons

D). Seasonal Water Vapour Distribution % of Increase or Decrease Trends, altitude about 0-40 KM in the Atmosphere:

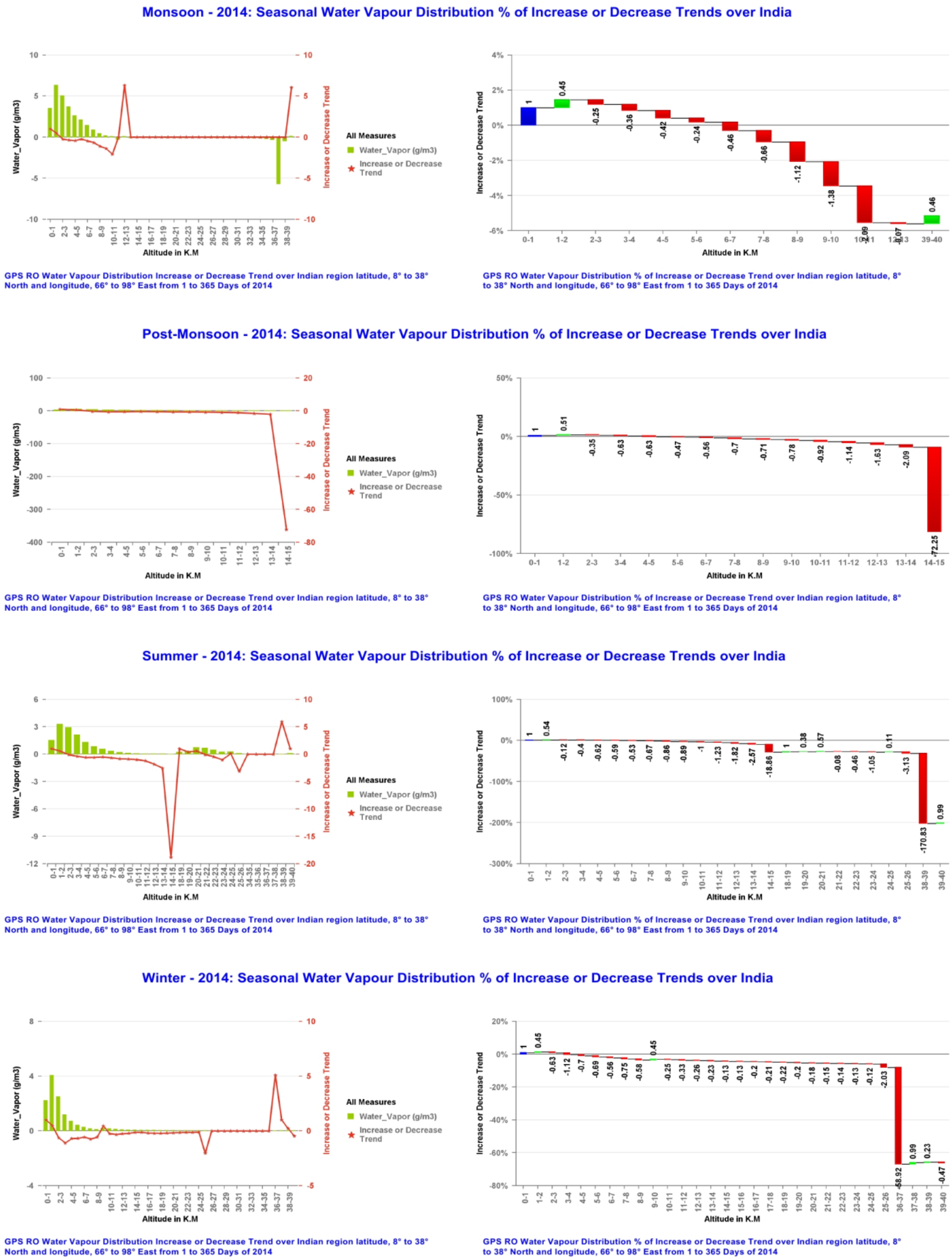
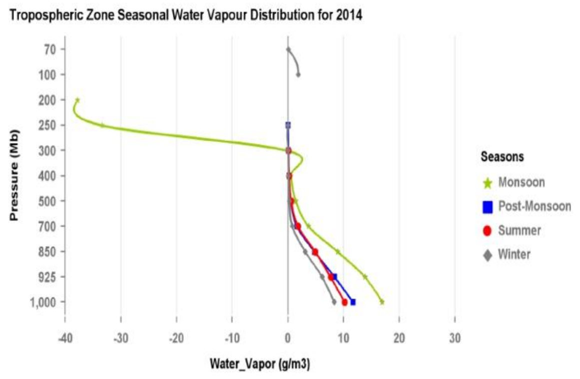


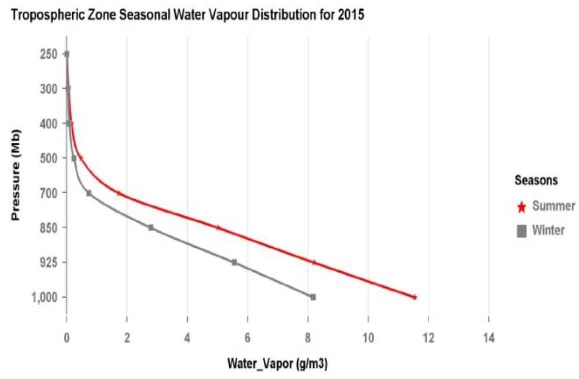
Figure 9: GPS RO Water Vapour Distribution % of Increase or Decrease Trend over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014, altitude about 0-40 KM in the atmosphere

E). Seasonal Distribution of Water Vapour for 2014 and 2015 with different atmospheric zones and for various fixed pressure levels:

Tropospheric Zone

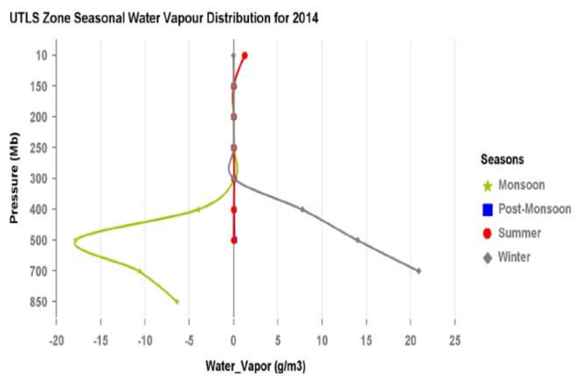


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014 for different Pressure Levels in Millibars

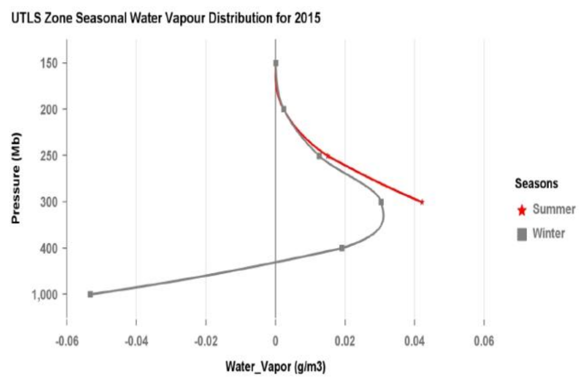


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015 for different Pressure Levels in Millibars

UTLS Zone

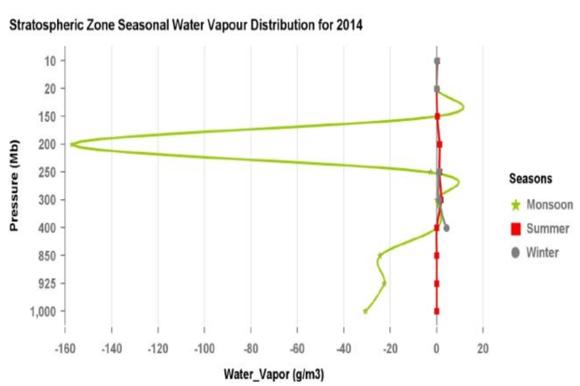


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014 for different Pressure Levels in Millibars

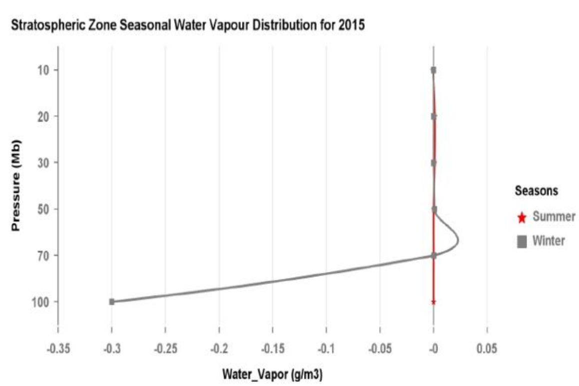


GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015 for different Pressure Levels in Millibars

Stratospheric Zone



GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 365 Days of 2014 for different Pressure Levels in Millibars



GPS RO Water Vapour Distribution over Indian region latitude, 8° to 38° North and longitude, 66° to 98° East from 1 to 150 Days of 2015 for different Pressure Levels in Millibars

Figure 10: Seasonal Distribution of Water Vapour for 2014 and 2015 with different atmospheric zones and for various fixed pressure levels

V. CONCLUSION AND FUTURE WORK

The water vapor is the most significant greenhouse gas and plays an important role in global climate change studies and weather forecasting. However atmospheric parameter's data assessment and usage is critical and challenging. The data warehousing and data mining applications are the most emerging technologies with powerful data managing features for better data processing and analysis for effective decision supporting systems. Hence as part of the study GPS RO datasets assimilation and classification are carried out by using by these applications and also applied better data analytics for precise results on recent water vapour distribution studies over India. The Pre-processed datasets are reusable for further GPS RO based water vapour studies.

In our earlier papers we have conducted the water vapour distribution studies by using radiosonde data obtained from the British Atmospheric Data Centre(BADC) with different met parameters of air temperature, dew point temperature and pressure levels for water vapour concentrations estimation at different height (pressure levels).The observational results of surface level water vapour distribution patterns and trends are presented, hence in future work we would like to conduct the Comparison studies between GPS Radio Occultation and Radiosonde Water Vapour Distribution over India. And also as part of the future studies we would like to incorporate the forecasting GPS RO water vapour distribution model based on the current trends by using neural networks for different layers of atmosphere with different seasonal variations in prospect trends.

ACKNOWLEDGMENT

We wish to thank COSMIC Data Analysis and Archive Center (CDAAC) for the availability of the occultation data and the supporting documentation to understand the data and usage for our water vapour distribution studies over India.

REFERENCES

[1] IM Held and BJ Soden, "Water vapor feedback and global warming 1", Annual Review of Energy and the Environment 25 (1), 441-475

[2] S.K. Sarkar, H.N. Dutta, M.V.S.N. Prasad and B.M. Reddy," TROPOSPHERIC WATER VAPOUR OVER INDIA", Antennas and Propagation, 1989. ICAP 89., Sixth International Conference on (Conf. Publ. No.301) 302 - 306 vol.2

[3] Bomin Sun, Anthony Reale, Dian J, Seidel and Douglas C.Hunt, "Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collection on comparison statistics" Journal of Geophysical Research, Vol.115, D23104, doi:10.1029/2010JD014457,2010

[4] Shu-peng Ho, Xinjia Zhou, Ying-Hwa Kuo, Douglas Hunt and Jun-Hong Wang "Global Evaluation of Radiosonde Water Vapor Systematic Biases using GPS Radio Occultation from COSMIC and ECMWF Analysis", Remote Sensing 210,2,1380-1330;doi:10.3390/rs2051380

[5] F.Ladstadter, A.K.Steiner, M.Schwarz and G.Kirchengast, "Climate intercomparision of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002 to 2013",Atmospheric Measurement Techniques, doi:10.5194/amt-8-1819-2015

[6] Vuda Sreenivasarao, Venata Subbareddy Pallamreddy, "Advanced Data Warehousing Techniques for Analysis, Interpretation and Decision Support for Scientific Data", (CCIS 198), 2011, Page(s): 162-174

[7] Folorunsho Olaiya, Adesesan Barnabas Adeyemo "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies",(MECS), 2012,page(s) 51-59

[8] Inmon. W, "Build the data warehouse", John Wiley and Sons, New York, 1996.

[9] Folorunsho Olaiya, Adesesan Barnabas Adeyemo "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies",(MECS), 2012,page(s) 51-59

[10] Ahrens, C. D. Meteorology Today: An Introduction to Weather, Climate, and the Environment. 9th Edition. Belmont, Calif., 2009.

[11] http://www.cosmic.ucar.edu/satStatus/index.html

[12] E. R. Kursinski, G. A. Hajj, J. T. Schofield, R. P. Linfield, and k. R. Hardy, "Observing Earth's atmosphere with radio occultation measurements using the global positioning system", Journal of Geophysical Research D,1997,vol.102,no.19,pp 23429-23465

[13] R.Biondi and T.Neubert, "Bending Angle and Temperature Climatologies from Global Positioning System Radio Occultations", Hindawi Publishing Corporation, Dataset Paper in Geosciences, Volume 2013, http://dx.doi.org/10.7167/2013/795749 pp 1-5

[14] Abhineet Shyam,B.S. Gohil and Sujit Basu,"Retrival of water vapour profiles from radio occultation refractivity using artificial neural network", Indian Journal of Radio & Space Physics Vol.42, Dec 2013 ,pp 411-419

[15] Neerja Sharma, D Jagadheesha, P.C. Joshi and P.K.Pal, "Atmospheric stability estimation using radio occultation data over India and surrounding region", Indian Journal of Radio & Space Physics Vol.38, Dec 2009 ,pp 317-385

[16] P. Kishore, M.VenkatRatnam , S.P.Namboothiri , IsabellaVelicogna , GhouseBasha , J.H.Jiang , K. Igarashi, S.V.B.Rao and V.Sivakumar, "Global (501S–501N) distribution of water vapor observed by COSMIC GPS RO Comparison with GPS radiosonde, NCEP, ERA-Interim, and JRA-25 reanalysis datasets", Journal of Atmospheric and Solar-Terrestrial Physics (2011) 1849–1860

[17] C.J. Johny, S.K.Sarkar and D.Punyasesudu, "Spatial and temporal variations of water vapour in upper troposphere and lower stratosphere over Indian region", Current Science, Vol.97, No.12, Dec 2009, pp 1735-1741

[18] D. Narayana Rao, M.VenkatRatnam, Sanjay Mehta, Debashis Nath, S.Ghouse Basha, V.V.M Jaganadha Rao, B.V.Krishna Murthy, T.Tsuda and Kenji Nakamura, "Validation of the COSMIC Radio Occultation Data over Gadanki(13.48 o N, 79.2 o E):A Tropical Region", Terr.Atmos.Ocean.Sci., Vol.20,No.1,59-70,February 2009,doi:10.3319/TAO.2008.01.23(F3C)

[19] University Corporation for Atmospheric Research (UCAR), UCAR-COSMIC Data Analysis and Archive Center (CDAAC), 2053,http://cdaaac-www.cosmic.ucar.edu/cdaac/.

[20] Ding Youqing, Yumei Fu, Zhu Fang, and Zan Xinwu," Comparison of Missing Data Filling Methods in Bridge Health Monitoring System", Proceedings. 12th IEEE International Conference on Cognitive Informatics & Cognitive Computing 978-1-4799-0783-0/13/2013 IEEE

[21] B. Mehala, P. Ranjit Jeba Thangaiah, and K. Vivekanandan, "Selecting Scalable Algorithms to Deal With Missing Values" International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009, 80-83

[22] B.-R. Wang, X.-Y.Lui, and J.-K. Wang, "Assessment of COSMIC radio occultation retrieval product using global radiosonde data", Atmospheric Measurement Techniques,6,1073-1083,2013, doi:10.5194/atm-6-1073-2013

[23] Khamphoui Southisombath, Toshio Wakabayashi and Yoshiaki Moriya, "The impact of water vapor in atmosphere on Ku band satellite DTV reception in a sub-tropical region", Proceedings of ISAP'04, Sendai, JAPAN, IEICE , 1381-1384

[24] Thomas W. Schlatter, "Atmospheric Composition and Vertical Structure", Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA, 2009, Environmental Impact and Manufacturing, ArticleID: eae31MS.