Volume : 4 | Issue : 6 | June 2015 • ISSN No 2277 - 8179

Research Paper

Data Warehousing and Data Mining Applications for Recent Water Vapour Distribution Studies Over India

Engineering

KEYWORDS : Meteorology, Atmospheric water vapour, climate change, Data warehousing, Data mining, Decision supporting, Hidden patterns, Information

Venkata Sheshanna Kongara — Research Scholar, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur-515 003,Andhra Pradesh, India.

Dr. D.Punyasesudu — Professor, Department of Physics, Rayalaseema University, Kurnool, 518 003, Andhra Pradesh, India.

ABSTRACT

Meteorology is one of the most significant area of practice and research of the atmospheric concerns that focuses on weather circumstances and exploration, In that 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. In general radiosonde observations are the primary source of upper air water vapour data and are being used for estimation of different atmospheric parameters, on the other hand different statistical and scientific methods are using to process the meteorological datasets in water vapour studies and have many data level limitations, challenges in storing, retrieving, managing and exploration of these structured and un-structured data; which is very large in size, accordingly the approaches of data maintenance and conversion involves in time consuming, expensive, and complex to mitigate the accurate results. Currently there are extraordinary prospects in the Information Systems to analyze these data and explore. The data warehousing and data mining applications are the most emerging features which are endorsed that information to be accessed easily and efficiently. This paper presents various analytics of the recent water Vapour distribution Studies over India by using these applications, The most recent version of the Meteorological Radiosonde data from British Atmospheric Data Centre (BDAC) are used for these studies. As a part of this study, radiosonde observations over 34 Indian stations pertaining to a period of 17 years from 1997 to 2014 obtained from the British Atmospheric Data Centre(BADC) are integrated and analyzed from the different met parameters of air temperature, dew point temperature and pressure levels for water vapour concentrations estimation at different height (pressure levels). The 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 surface level water vapour distribution patterns and trends are presented in the form of Physiographic Divisions, annual and seasonal charts for better decision supporting systems. In fact it is an improved version of the detailed studies related to water vapour distribution over India.

INTRODUCTION

1.1. Importance of the 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 are using to process the water vapour datasets and measure the correlated innovations [11][12] [15][16][18].However due to the data limitations researchers are facing challenges in storing, retrieving, managing and exploration of these structured and un-structured data which is very large in size. Hence the approaches of data maintenance and conversion involves in time consuming, expensive, and complex to mitigate the accurate results; consequently most of the researchers choosing these datasets specific to the locations or regional levels with the limited time periods and limiting the scope of the study. 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 [7][8]. The "Data warehouse" is pretend by W. H. Inmon [3] 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 are include association Rules, clustering, classification, prediction, sequence mining, web mining, text mining and spatial data mining [9] [10]. This paper reveals various analytics of the recent water vapour distribution studies over India by using these applications. As a part of this study, radiosonde observations over 34 Indian stations pertaining to a period of 17 years from 1997 to 2014 obtained from the British Atmospheric Data Centre(BADC) are integrated and analyzed from the different met parameters of air temperature, dew point temperature and pressure levels for water vapour concentrations estimation at different height (pressure levels). The datasets assimilation and classification techniques are carried out by using data warehousing and data mining techniques and applied better data analytics for precise results. The observational results of surface level water vapour distribution patterns and trends are presented in the form of regional, physiographic divisions, annual and seasonal charts for better decision supporting systems. In fact it is an improved version of the detailed studies related to water vapour distribution over India.

1.2. Physiographic features of India

India is a country of great geographical extent with a diversity and rich cultural heritage. The physical features of India are not same across the country. They are varying from place to place with significant features which are highest mountains in the north, wide plains in the centre, desert region in the northwest, narrowing peninsula to the south, mountain ranges to the northeast and a lower range running along the west coast of the peninsula. The vast expanse of the Indian Ocean lies to the

Volume : 4 | Issue : 6 | June 2015 • ISSN No 2277 - 8179

south of the land mass with the Bay of Bengal and Arabian Sea pushing north engulfing the peninsula [4], Hence Physiographic features plays an important role on the climatic variations of the country. And also the magnitude of these climatic changes varies from different regions and point in time which are observed that different climatic seasons in a year commonly known as winter, summer, monsoon and post monsoon. The seasons are categorized from various months incorporated as winter [December to February], summer [March to May], Monsoon [June to September] and the post monsoon [October to November]. The physiographic divisions as shown in the map [5] figure A and the radiosonde station's water vapour seasonal trends with physiographic divisions mentioned in table I are analyses in this study. The observational results of surface level water vapour distribution trends are presented in the form of Physiographic divisions, annual and seasonal charts for better understanding and decisions supporting.

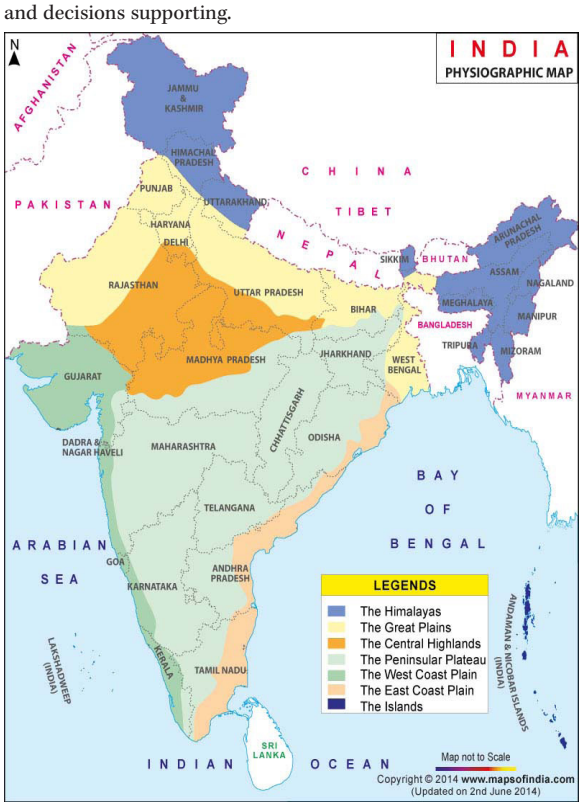


Figure A: The physiographic divisions by Maps of India

Following are the different Physiographic divisions over India are considered in the study

1.The Kashmir Region: These are highest mountain ranges are located in North Western Region of India. This region is uninhabited precipitation due to the Himalayan Mountains

2.The Assam Region: Covers different states in North Eastern Region of India, the climate is cold and rainfall most of the months

3.The Northern Plains: The Indo-Gangetic plains located in Northern and Eastern Region of India. The climate is very hot in summers and very cold in winters

4.The Central Highlands: These are composed of different Plateaus which are lies in North Central Region of India. The climate is with hot summers and mild winters

5.The Desert Region: These are lies in the North Western region of India, The climate is very hot, rainfall is inadequate and unpredictable

6.The Peninsular Plateau: The Peninsular plateau lie to the south of the northern plains of India with different weather

7.The Eastern Coastal Plains: These are lies along the east coast of the Southern Region of India and washed by Bay of Bengal.

8.The Western Coastal Plains: These are lies along the west coast of Southern Region of India and washed by Arabian sea. The rainfall is heavy in this region

9.The Island: The one in Bay of Bengal called as Andaman and Nicobar islands which is in South Eastern Region of India and Arabian Sea called as Lakshadweep islands which is in South Western Region of India

Station	State	Physiographic Divisions	Height (M)	Latitude degree (°)	Longitude degree (°)	Altitude (M)	Surface Pressure (Mb)
AGARTALA	Tripura	The Assam Region	15	23.88	91.25	15-25	990-1020
AHMADABAD	Gujarat	The Western Coastal Plain	55	23.07	72.63	50-70	1000-1014
AMINI_DIVI	Union Territory of Lakshadweep	The Island	3	11.12	72.73	10-150	990-1000
AURANGABAD_CHIKALTHAN_AE	Maharashtra	The Peninsular Plateau	582	19.85	75.4	575-585	940-960
BANGALORE	Karnataka	The Peninsular Plateau	920	12.97	77.58	900-925	990-1000
BHOPAL_BAIRAGARH	Madhya Pradesh	The Central Highlands	523	23.28	77.35	515-525	940-960
BHUBANESWAR	Orissa	The Eastern Coastal Plains	45	20.25	85.83	40-60	990-1016

BOMBAY__SANTACRUZ	Maharashtra	The Western Coastal Plain	8	19.12	72.85	05_20	990-1020
CALCUTTA_DUM_DUM	West Bengal	The Northern Plains	4	22.65	88.45	0-20	1000-1020
COCHIN_WILLINGDON	Kerala	The Western Coastal Plain	1	9.95	76.27	0-10	1000-1020
CWC_VISHAKHAPATNAM_ WALTAIR	Andhra Pradesh	The Eastern Coastal Plains	45	17.7	83.3	40-60	1000-1020
DIBRUGARH_MOHANBARI	Assam	The Assam Region	110	27.48	95.02	90-115	995-1000
GAUHATI	Assam	The Assam Region	47	26.1	91.58	45-55	995-1000
GOA_PANJIM	Goa	The Western Coastal Plain	58	15.48	73.82	55-65	995-1015
GORAKHPUR	Uttar Pradesh	The Northern Plains	76	26.75	83.37	70-85	985-1015
GWALIOR	Madhya Pradesh	The Central Highlands	205	26.23	78.25	180-220	980-1010
HYDERABAD_AIRPORT	Telangana	The Peninsular Plateau	530	17.45	78.47	480-540	940-970
JAGDALPUR	Chhattisgarh	The Peninsular Plateau	552	19.08	82.03	540-570	940-1000
JODHPUR	Rajasthan	The Desert Region	217	26.3	73.02	210-230	980-1000
KARAIKAL	Pondicherry	The Eastern Coastal Plains	15	10.92	79.83	0-20	1000-1020
LUCKNOW_AMAUSI	Uttar Pradesh	The Northern Plains	122	26.75	80.88	110-129	970-1010
M.O.RANCHI	Jharkhand	The Peninsular Plateau	652	23.32	85.32	600-700	920-970
MACHILIPATNAM	Andhra Pradesh	The Eastern Coastal Plains	14	16.2	81.15	0-20	995-1020
MADRAS_MINAMBAKKAM	Tamil Nadu	The Eastern Coastal Plains	10	13	80.18	0-20	995-1020
MANGALORE_PANAMBUR	Karnataka	The Western Coastal Plain	32	12.95	74.83	20-40	1000-1020
MINICOY	Kerala	The Western Coastal Plain	1	8.3	73.15	0-20	1000-1020
NAGPUR_SONEGAON	Maharashtra	The Peninsular Plateau	308	21.1	79.05	280-320	970-990
NEW_DELHI_SAFDARJUNG	Delhi	The Northern Plains	211	28.58	77.2	210-211	975-990
PATIALA	Punjab	The Northern Plains	249	30.33	76.47	240-260	950-1000
PATNA	Bihar	The Northern Plains	51	25.6	85.1	45-60	980-1015
PORT_BLAIR	Andaman Nicobar Islands	The Island	73	11.67	92.72	70-85	985-1020
SILIGURI	West Bengal	The Northern Plains	122	26.67	88.37	110-129	970-1010
SRINAGAR	Jammu &Kashmir	The Kashmir Region	1585	34.08	74.83	1580-1600	820-990
THIRUVANANTHAPURAM	Kerala	The Western Coastal Plain	60	8.48	76.95	55-65	990-1081

Table I : Describes the Radiosonde stations used in the study and Latitude and Longitude degrees of the locations. And also the Physiographic divisions of the India

The rest of the paper is organized as follows. The literature review and related work for water vapour studies are discussed in section II. The details of the 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 work-

bench, 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 [8] 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

Meghali A. Kalyankar et al [9] discussed on Data mining Process for weather data to study weather data using data mining technique like clustering technique. By using this technique one can acquire Weather data and can find the hidden patterns inside the large dataset so as to transfer the retrieved information into usable knowledge for classification and prediction of climatic condition.

Gaurav J. Sawale et al [10] discussed how to use a data mining technique to analyze the Metrological data like Weather data. A variety of data mining tools and techniques are available in the industry, but they have been used in a very limited way for meteorological data. And also explained how a neural network-based algorithm for predicting the atmosphere for a future time and a given location is presented.

Prakash et al [11] conducted a study of 11 years (1995-2005) Perceptible Water (PW) derived from radiosonde data of Santacruz airport (19.7 N, 72.51 E) Mumbai for both 0000 and 1200 UTC has been carried out. It has been seen that 0000 UTC perceptible Water vapour trend over the years (1995-2005) is quite smooth as compared to 1200 UTC. They concluded that Studies above indicates that the regional models after assimilating the above sounding data may improve the temperature and moisture field's initial fields and in turn overall forecasting.

Suresh Raju et al [12] described the estimation of daily Integrated Water Vapor (IWV) using ground-based GPS at Bangalore, India. The GPS data was analyzed using GAMIT-10.07 to derive total zenith delay (TZD). The IWV was derived from TZD after deducting hydrostatic delay component using empirical models established for Bangalore based on atmospheric parameters and ray-tracing techniques. The GPS derived IWV was compared with radiosonde observations.

R.M. Gairola et al [13] studied the south-west monsoon seasonal variations of the water vapor and rainfall over the Indian Subcontinent and oceans using microwave (MW) and near-Infrared (NIR) satellite measurements on monthly scales. The total precipitable water (TPW) derived from multi channel imaging data acquired with the Moderate Resolution Imaging Spectrometer (MODIS) on the Terra Spacecraft and rainfall data from merged infrared estimates calibrated against Tropical Rainfall Measuring Mission (TRMM) microwave data respectively are used. And also discussed in order to examine the variations and the complete flow patterns of the moisture from Indian Ocean, Arabian Sea and Bay of Bengal to Indian land masses and rainfall over India and adjoining areas, characterizing good and bad monsoon are revealed for all the three years.

P. Kishore et al [14] 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).

Sanjay Kumar et al [15] presents an analysis of GPS data recently deployed (as of May 2007) on the campus of Banaras Hindu University, Varanasi (latitude 25 150N, longitude 82 590E). Further, they have compared the variability of water vapor from Kanpur GPS, AERONET and MODIS water vapor data for the year 2007. The monthly variability of water vapor shows characteristic features and dynamics of water vapor between two closely spaced GPS stations. The limited GPS derived water vapor data analyzed simultaneously over Varanasi and Kanpur in the Indo-Gangetic region for the year 2007–2008.The column water vapor estimated from GPS data is found to be very sensitive with surface temperature, monsoon onset and rainfall.

III. Data processing & Methodology

3.1. Source of the data:

The global atmospheric radiosonde data has millions of soundings from worldwide upper air stations over the period 1996 to present. Each sounding is composed of records of pressure, geopotential height, temperature, dew-point temperature, windspeed and wind-direction at standard and significant pressure levels proposed by World meteorological Organization (WMO) from the surface to approximately 20-30 km. This data was extracted from the daily occurrences of atmospheric soundings compiled by British Atmospheric Data Centre (BDAC) and is made available for individuals and organizations for atmospheric studies[17].The radiosonde takes measurements at intervals of approximately 2 seconds. The high resolution data files contain all such data. The standard resolution data files contain measurements taken at particular levels of the atmosphere. we obtained different stations details, air pressure, air temperature, dew point temperature, wind speed, wind direction and various other parameters for the time period 1997-2014 over India as shown in table I

3.2. Water Vapour Data Integration:

As acknowledged in the Introduction of this paper data warehousing and data mining applications are the best suitable for atmospheric water vapour studies. The dimensional modeling is plays an important role in data warehouse design. We analyzed source data and its parameters, applied transformations and built a WV dimensional model. The dimension members which are including with stations details, air pressure, air temperature, dew point temperature, wind speed, wind direction and its parameters are processed with calculated measures 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 B;

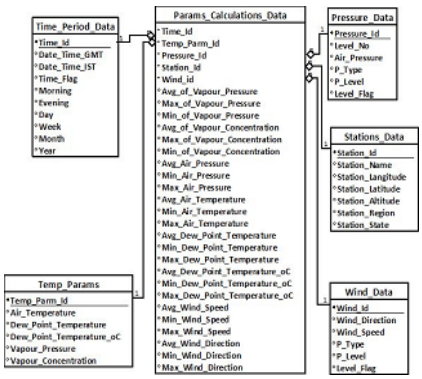


Figure B: Star Schema based WV dimensional model for water vapour data.

Based on the data availability we have selected the necessary stations for various geographical locations specified in Table I and data cleansing activity is conducted to focus on the target data. As per the accessibility all the datasets are collected in GMT and converted the dates from GMT time to India IST for 00:00 GMT as 5:30 AM IST and 12:00 GMT as 17:30 PM IST for all the datasets. The Pressure levels are considered at the surface levels while validating the altitude in Meters to identify Surface Pressure for various stations based on the station height. Extracted the rules and map them into different dimensions and measures to produce understandable and useful knowledge for analysis

3.3. Integrated Water Vapour Concentration:

H. Sarkar et al [18] conducted Comparative Study of Integrated Water Vapor (IWV) and of Attenuation of 94 GHz Signal from Radiometer and Radiosonde Observations during Monsoon Period over Kolkata, India. The Radiosonde data they have calculated the integrated water vapor content at different heights for the specific dates using regular empirical formula. However Khamphoui et al [19] 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) by using following equation. We are also using the identical equation for our studies for evaluation of the integrated water vapour as below.

The VAPOUR_CONCENTRATION (M) are calculated based on the Vapour Pressure (i.e.) → 1

Vapour Concentration (M) in grams/meter cube

$$(g/m^3) = 216.7 * \left[\frac{Vapour\ Pressure\ (E)}{Air\ Temperature\ (K)}\right]$$

The VAPOUR_PRESSURE (E) are calculated based on the Empirical formula (i.e.) → 2

Vapour Pressssure (E)=

$$6.11 * Exp\left[\frac{19.7*Dew\ Point\ Temperature\ oC}{Dew\ Point\ Temperature\ oC+273.15}\right]$$

The DEW_POINT_TEMPERAT JRE are converted from Kelvin to Degree Centigrade (i.e.) → 3

Dew point Temperature oC=
ew point Temperature (K)-273.15

The calculated water vapour concentration datasets are arranged in the dimensional tables of the data warehouse for our analyses, which may be reusable for the further studies. We have analyzed all the datasets with various analytics, the detailed results are presented in section IV.

3.4. Water Vapour Data Pre-Processing and Classification:

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 radiosonde source data also there are some missing values for various parameters for different stations and different dates of the years. The BDAC filled those missing data sets with 999999999 in their data, Hence we have handled those missing values filled with précised one by using cluster based missing values algorithm is consider to process the water vapour data for better results.

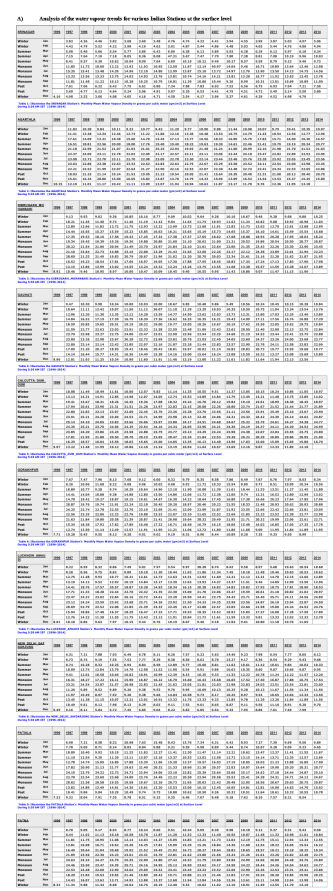
Classification is a supervised machine learning technique used to build a model, once model has been built and applied to unseen data for prediction of class label. Building the accurate and efficient classifiers for large databases is one of the essential tasks of data mining and machine learning research. 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. We conducted a comparative review for various methods of classification and measured the accuracy of the results. It show that using Neural Networks obtains the best result among the other methods therefore in the water vapour studies we have used improved version of neural networks algorithm to classify the water vapour data for better classification and prediction

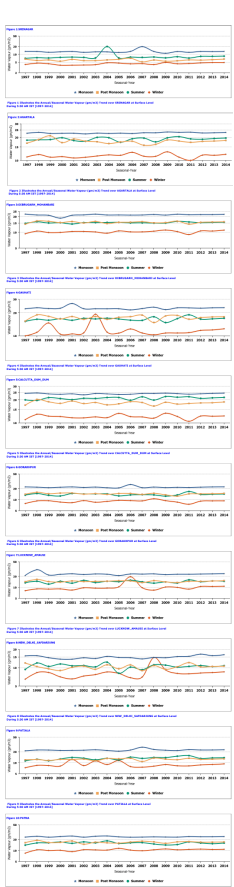
IV. Data Analysis and Results

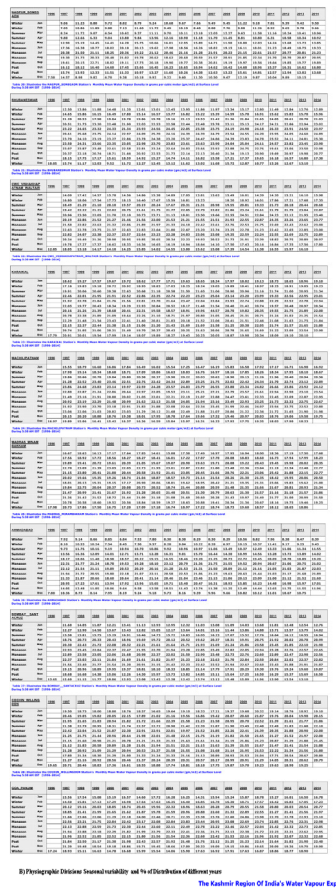
The water vapour data analytics provide the details on the seasonal mean water vapour distributions from different locations and physiographical regions have been generated through the improved version of the equations. The graphs Illustrates the various Station's monthly mean water vapour density trends in grams per cubic meter (gm/m3) at Surface Level during 5:30 AM IST for 1996-2014. The detailed results shown in the below sections A, B, C and D as follows

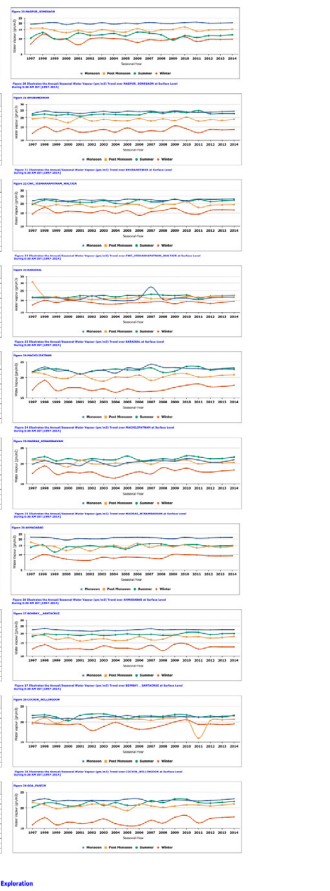
A) Analysis of the water vapour trends for various Indian Stations at the surface level
B) Physiographic Divisions Seasonal variability and % of Distribution of different years
C) Analysis of the water vapour trends from different seasons within various years with different physiographic divisions
D) Physiographic divisional water vapour Increase and Decrease trends for different years for various seasons

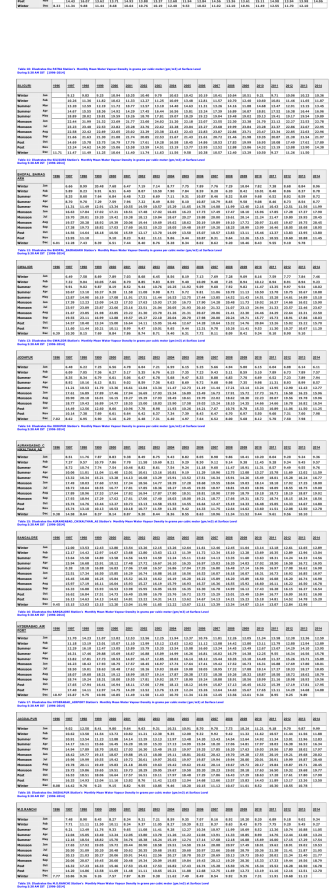
A) Analysis of the water vapour trends for various Indian Stations at the surface level

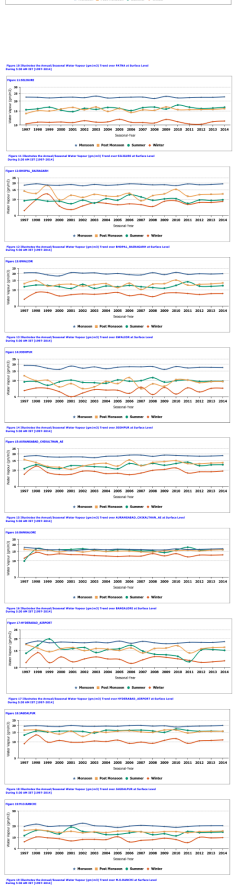


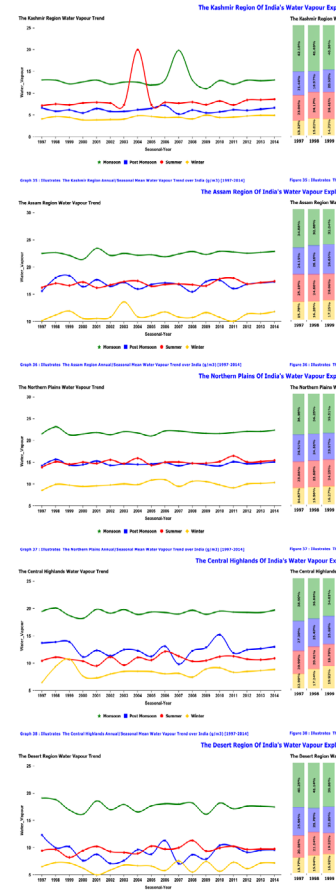


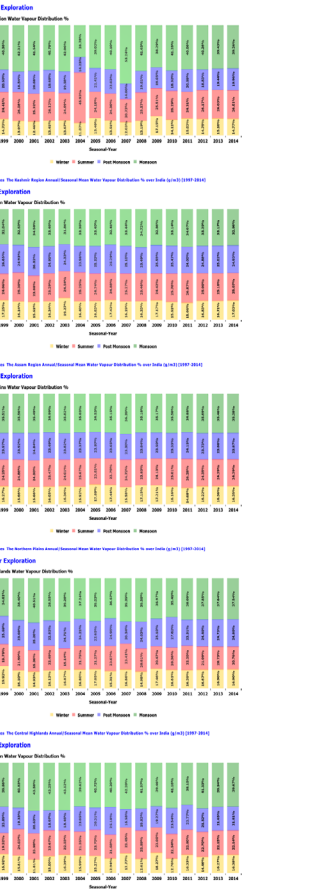














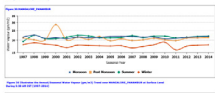


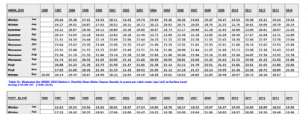


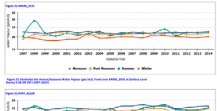


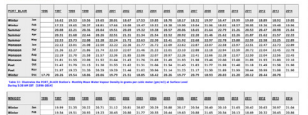
B) Physiographic Divisions Seasonal variability and % of Distribution of different years



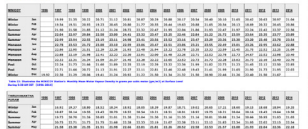
















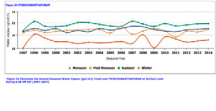
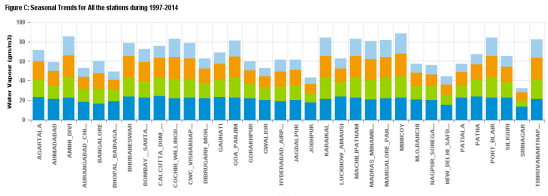
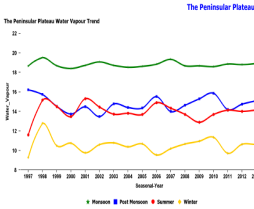
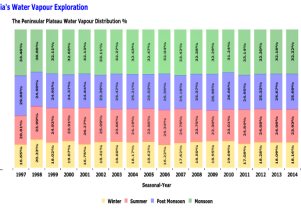


Figure C: Seasonal Trends for All the stations during 1997-2014



The Peninsular Plateau Of India's Water Vapour Exploration

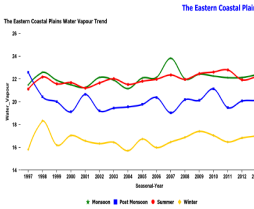


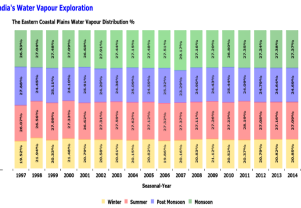


Graph 40 : Illustrates The Peninsular Plateau Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 40 : Illustrates The Peninsular Plateau Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Eastern Coastal Plains Of India's Water Vapour Exploration

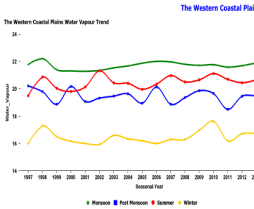


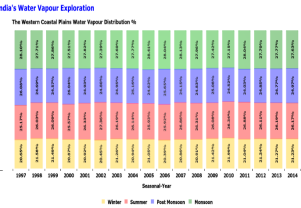


Graph 41 : Illustrates The Eastern Coastal Plains Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 41 : Illustrates The Eastern Coastal Plains Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Western Coastal Plains Of India's Water Vapour Exploration

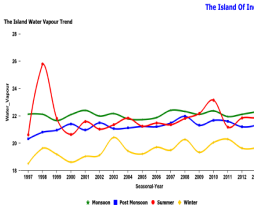


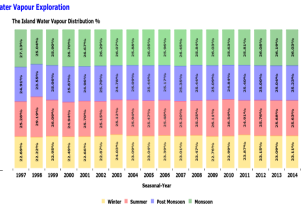


Graph 42 : Illustrates The Western Coastal Plains Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 42 : Illustrates The Western Coastal Plains Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Island Of India's Water Vapour Exploration





Graph 43 : Illustrates The Island Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 43 : Illustrates The Island Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Peninsular Plateau Of India's Water Vapour Exploration

Graph 40 : Illustrates The Peninsular Plateau Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 40 : Illustrates The Peninsular Plateau Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Eastern Coastal Plains Of India's Water Vapour Exploration

Graph 41 : Illustrates The Eastern Coastal Plains Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 41 : Illustrates The Eastern Coastal Plains Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

The Western Coastal Plains Of India's Water Vapour Exploration

Graph 42 : Illustrates The Western Coastal Plains Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

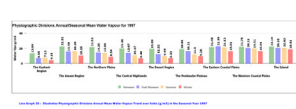
Figure 42 : Illustrates The Western Coastal Plains Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

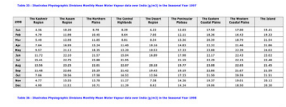
The Island Of India's Water Vapour Exploration

Graph 43 : Illustrates The Island Annual/Seasonal Mean Water Vapour Trend over India (g/m3) [1997-2014]

Figure 43 : Illustrates The Island Annual/Seasonal Mean Water Vapour Distribution % over India (g/m3) [1997-2014]

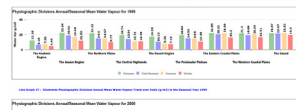








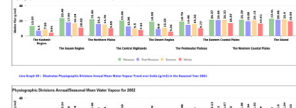


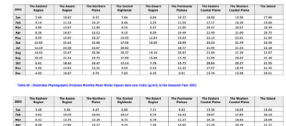




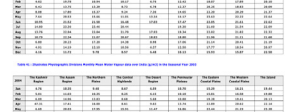




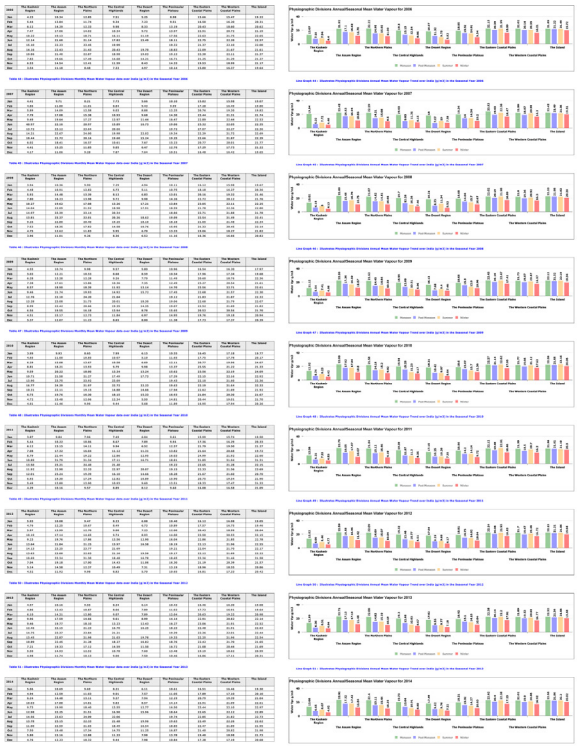








Volume : 4 | Issue : 6 | June 2015 • ISSN No 2277 - 8179



D) Physiographic divisional water vapour Increase and Decrease trends for different years for various seasons

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. In common there are different statistical and empirical methods are using to process the water vapour datasets and measure the correlated innovations. However data assessment and usage is critical and challenging, due to that most of the researchers limiting the scope of the comprehensive analysis in their studies. 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 datasets assimilation and classification techniques 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 water vapour studies.

Following are the accomplishments in the study.

- Massive collection of radiosonde water vapour data is managed about 17 Years since from 1997 to 2014 and there are 34 India stations are considered with multiple atmospheric parameters to assess the meaningful data.
- Most recent radiosonde observational data is considered and also improved version of the empirical equations are used to calculate the water vapour concentration for different pressure levels
- Build the water vapour dimensional model for ease of use and analysis, the evaluated data is maintained in the data warehouse for further studies
- Identified the gaps of missing data and accomplished missing data using improved clustering methods to evaluate and fill the gaps for precise results
- Data classification is carried out by using the improved version of neural networking algorithm from various Indian geographical locations and seasonal variations. The water vapour distributional trends are presented rich looking graphs for better understand and analysis
- In future work is anticipated for validation of the forecasting model based on the current trends with water vapour distribution by using neural networks for different physiographic divisions and seasonal variations in prospect trends.

ACKNOWLEDGEMENTS

We wish to thanks to British Atmospheric Data Centre (BADC) for providing the radiosonde data and the supporting documentation to understand the data and usage for our studies

REFERENCE

[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] Inmon. W, "Build the data warehouse", John Wiley and Sons, New York, 1996. | [4] http://www.yourarticlelibrary.com/geography/the-main-physiographic-divisions-of-india-geography/5461/ | [5] http://www.mapsofindia.com/maps/india/physiographic.htm | [6] Xiaoguang Tan, "Data Ware housing and its potential using in weather forecast", Institute of Urban Meteorology, CMA, Beijing, China | [7] 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 | [8] Folorunsho Olaiya, Adesesan Barnabas Adeyemo "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies",(MECS), 2012,page(s) 51-59 | [9] Meghali A. Kalyankar,Prof. S. J. Alaspurkar "Data Mining Technique to Analyse the Metrological Data", (IJARCSSE), 2013, Volume :3, Page(s) :114-118 | [10] Gaurav J. Sawale, Dr. Sunil R. Gupta, "Use of Artificial Neural Network in Data Mining For Weather Forecasting", (IJCSA), 2013, Volume:6, Page(s) 383-387 | [11] Satish Prakash, R.K. Giri and Adesh, "Precipitable Water Vapour at Santacruz Airport Mumbai from Radiosonde Measurements: A Study" International Journal of Physics and Mathematical Sciences, 2012 Vol. 2 , pp.120-130 | [12] Suresh Raju, Korak Saha, Bijoy V. Thampi and K. Parameswaran," Measurement Of Integrated Water Vapor Over Bangalore Using Ground Based GPS Data", Proceeding of the XXVIIIth URSI General Assembly in New Delhi 16 | [13] R.M. Gairola, and V.K. Agarwal, "Measurements of Water Vapour and Rainfall Over Indian Ocean and Sub-Continent with a TERRA-MODIS and TRMM"www.ursi.org/Proceedings/ProcGA05/pdf/F10P.6(01549).pdf | [14] 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 | [15] Sanjay Kumar, A.K. Singh, Anup K. Prasad and R.P. Singh, "Variability of GPS derived water vapor and comparison with MODIS data over the Indo-Gangetic plains", Physics & Chemistry of the Earth, doi:10.1016/j.pce.2010.03.040, 2010 | [16] Swastika Chakraborty and Animesh Maitra, "A Comparative Study of Cloud Liquid Water Content from Radiosonde Data at a Tropical Location", International Journal of Geosciences, 2012, 3, 44-49 | [17] Met Office (2006): Met Office Global Radiosonde Data. NCAS British Atmospheric Data Centre (BDAC), date of citation.http://catalogue.ceda.ac.uk/uuid/f2afaf808b61394b78bd342ff068c8cd | [18] H. Sarkar, S.K. Midya and S. Goswami," A Comparative Study of Integrated Water Vapor (IWV) and of Attenuation of 94 GHz Signal from Radiometer and Radiosonde Observations during Monsoon Period over Kolkata, India."The Pacific Journal of Science and Technology, Volume 13. Number 1. May 2012 (Spring) | [19] Khamphoui Southisombath, Toshio Wakabayashi and Yoshiaki Moriya, "The impact of water vapor in atmosphere on Ku band satellite DTV reception in a subtropical region", Proceedings of ISAP'04, Sendai, JAPAN, IEICE , 1381-1384 | [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