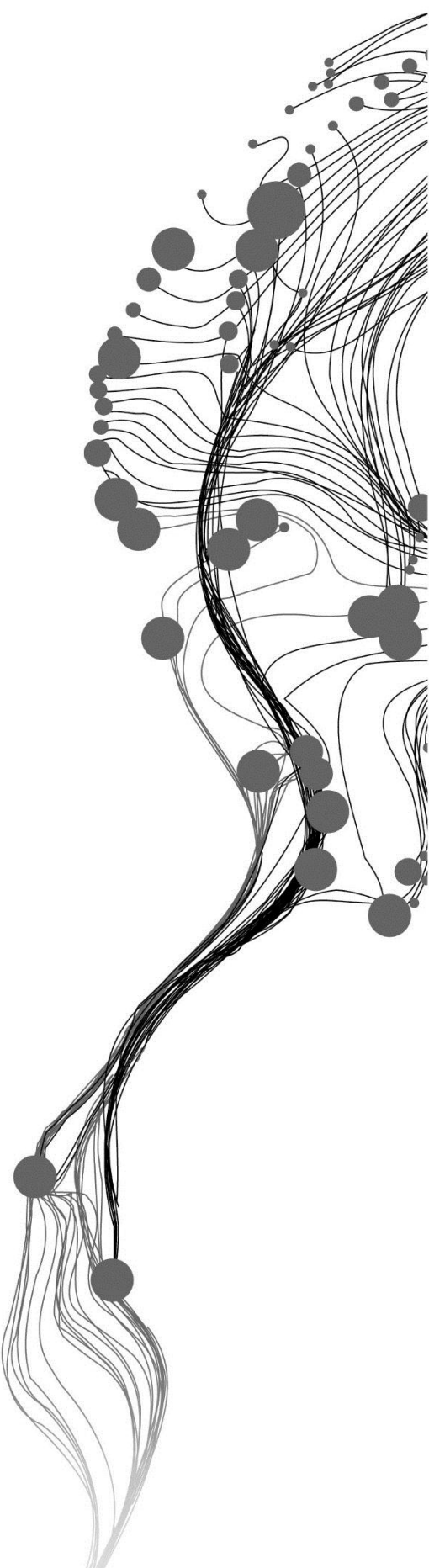


DESIGNING A SPATIAL OLAP BASED SYSTEM TO EXPLORE THE WEATHER BALLOON DATA

PRANAY KUMAR
MARCH, 2015

ITC SUPERVISOR
Dr. Raul Zurita Mila

IIRS SUPERVISORS
Dr. Sameer Saran
Dr. Yogesh Kant



DESIGNING A SPATIAL OLAP BASED SYSTEM TO EXPLORE THE WEATHER BALLOON DATA

PRANAY KUMAR

Enschede, The Netherlands, [March, 2015]

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the
requirements for the degree of Master of Science in Geo-information Science
and Earth Observation.

Specialization: Geoinformatics

THESIS ASSESSMENT BOARD:

Chairperson : Prof. Dr. M.J. Kraak
External Examiner: Prof. Dr. P.K.Garg, IIT, Roorkee
IIRS Supervisor : Dr. Sameer Saran
IIRS Supervisor : Dr. Yogesh Kant
ITC Supervisor : Dr. Raul Zurita Milla

OBSERVERS:

ITC Observer : Dr. Nicholas Hamm
IIRS Observer : Dr. S. K. Srivastav



UNIVERSITY OF TWENTE.

ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION



INDIAN INSTITUTE OF REMOTE SENSING

Indian Space Research Organisation

Department of Space, Government of India

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

With the recent advances in field of big data analysis, spatial data warehousing (SDW) is the need of the hour. SDW provides a structured approach for manipulating, analyzing and visualizing huge volumes of data. The architecture has been designed in such a way that it takes care of the data starting from data integration to pattern evaluation and knowledge discovery. Spatial online analytical processing (SOLAP) is a recent standard developed to describe analytical capabilities for understanding and visualizing the multidimensional data using the multi-dimensional expressions (MDX). It also provides a hierarchical structure to the data by representing it into multidimensional cubes. The cube structure not only provides a deeper understanding of the spatial properties of the data but also a platform to arrange the data into a conceptual schema. This is done using the extensible markup language for analysis (XMLA).

The primary objective is towards designing a Spatial OLAP based system to visualize the trends for inter-seasonal and inter-annual patterns of the atmospheric temperature at significant pressure levels for the Indian subcontinent using the characteristic features of SOLAP and SDW. This has been achieved by extract, transform, load (ETL), drill down and drill up operations. The regions have been classified on the basis of the latitudinal extent of the Indian subcontinent. The daily radiosonde data of 27 stations distributed over the regions have been used for the analysis. The regions include peninsular, mid-latitude and sub-humid tropical regions. The study has depicted that the temperature inversion at the lower levels were stable during the pre-monsoon season over both peninsular and mid-latitude regions. Variations in temperature have been observed during the post-monsoon seasons for both the regions. For the upper layers, stable variations were observed during the post-monsoon season over the peninsular and mid-latitude regions. High variations were observed over these regions during the pre-monsoon seasons due to the influence of the western disturbances. These analyses have been successfully visualized on the spatial visualization tool in the form of maps and graphs.

Keywords: Spatial data warehouse, spatial online analytical processing, multidimensional expressions, multidimensional cubes, analytical queries, extensible markup language for analysis, radiosonde data, extract, transform, load, drill down and drill up, temperature inversion.

ACKNOWLEDGEMENTS

As they say, “all well that ends well”. My eighteen months exciting journey at IIRS has finally come to an end. Foremost, I would like to express my thankfulness to ITC and IIRS joint education programme for providing me global exposure and an opportunity to enhance my technical skills in the field of Geoinformatics.

I would like to express my deepest gratitude to Dr. Sameer Saran, my IIRS supervisor for his continuous motivation, guidance, support and knowledge sharing in the field of spatial data warehousing. It is he, who had given such an innovative and interesting research topic and had been a true mentor at every step. I would also like to thanks Dr. Yogesh Kant my second supervisor, for his valuable suggestions and feedbacks on the meteorology part of the research.

I am grateful to Dr. Raul Zurita Mila, my ITC supervisor, for his valuable guidance and suggestions at every stage of this research work.

My sincere thanks to Dr. Y.V.N. Krishna Murthy, the Director, IIRS, for enlightening us with his expertise in the field of remote sensing. I am grateful to Mr. P.L.N Raju, Group Head, RS and GID, IIRS for providing his valuable feedback at each step of the research. Special thanks to all of the faculty and staff at IIRS for their cooperation. I would also like to thank Dr. S.K. Srivastava, Head, GID, IIRS and Dr. Nicholas Hamm, ITC for their valuable suggestions and support throughout the course.

I thank all my batch mates for being there in both good and bad times. I had a great time with you all and I’m going to cherish every moment that we had enjoyed together during the awesome “Euro trip”.

Lastly, I owe this achievement to my family for being the greatest source of inspiration and motivation.

TABLE OF CONTENTS

List of figures	v
List of tables	vi
1. INTRODUCTION.....	1
1.1. Background.....	1
1.2. Motivation and problem statement.....	1
1.3. Research Identification.....	3
1.3.1. Research Objective.....	3
1.3.2. Sub-objectives and Research Questions.....	3
1.4. Innovation aimed at.....	3
1.5. Thesis structure	3
2. LITERATURE REVIEW.....	5
2.1. Related work in SOLAP and SDW.....	5
2.2. GeoMondrian	7
2.3. WorkBench	8
3. DATA SETS, STUDY AREA AND TOOLS.....	9
3.1. Data.....	9
3.1.1. Format description of data files.....	9
3.1.2. Header format.....	10
3.1.3. Data format	10
3.2. Study Area.....	11
3.3. Tools used.....	13
3.3.1. Hardware tools.....	13
3.3.2. Software tools.....	13
4. DESIGN AND IMPLEMENTATION.....	14
4.1. High level architecture	14
4.2. Database creation (ETL).....	15
4.2.1. Extract (E)	15
4.2.2. Transform (T)	15
4.2.3. Load (L).....	16
4.3. Server setup.....	16
4.3.1. Java runtime environment (JRE)	17
4.3.2. GeoMondrian (SOLAP) SERVER	18
4.3.3. Connecting the servers with the database.....	18
4.4. Multidimensional cube creation.....	19
5. RESULTS AND DISCUSSIONS	21
5.1. Multidimensional cube construction and pattern evaluation.....	21
6. CONCLUSION AND RECOMMENDATIONS.....	35
6.1. Conclusion	35
6.2. Answers of research questions.....	35
6.3. Recommendations	36

List of references37

LIST OF FIGURES

Figure 1.1: Components of Spatial Data Warehouse.....	2
Figure 2.1: SOLAP (Integration of GIS and OLAP).....	5
Figure 2.2: The Taxonomy of Spatial and Spatiotemporal data warehouse.....	6
Figure 2.3: Star Schema Representation.....	6
Figure 2.4: MDX Query structure.....	7
Figure 3.1: Locations of 1500 stations of IGRA	9
Figure 3.2: Study Area - India.....	11
Figure 4.1: SDW Architecture	14
Figure 4.2: Conceptual Model of the dataset using Star Schema.....	15
Figure 4.3: Scripts used for the creation of the radiosonde database.....	16
Figure 4.4: System and Environment Variables.....	17
Figure 4.5: Startup of Tomcat Server.....	17
Figure 4.6: Tomcat Web Application Manager.....	18
Figure 4.7: GeoMondrian Server homepage	18
Figure 4.8: Successful connectivity of SOLAP server with database	19
Figure 4.9: Levels of hierarchy used for performing SOLAP operations	20
Figure 5.1: Meteorological cube construction using MDX editor and Workbench	21
Figure 5.2: MDX query for Pre-Monsoon season daily data at 850 hPa.....	22
Figure 5.3: MDX query for post-monsoon season daily data at 850 hPa	22
Figure 5.4: MDX query for average temperature for Pre-Monsoon season at 850 hPa	23
Figure 5.5: Seasonal variations in temperature with respect to region.....	24
Figure 5.6: MDX query for pre-monsoon.....	25
Figure 5.7: Seasonal variations in temperature at 850 hPa in the peninsular region during both the seasons.	26
Figure 5.8: Seasonal variations in temperature at 850 hPa in the mid latitude region during both the seasons.....	27
Figure 5.9: Seasonal variations in temperature at 700 hPa in the peninsular region during both the seasons.	28
Figure 5.10: Seasonal variations in temperature at 700 hPa in the mid latitude region during both the seasons.....	29
Figure 5.11: Seasonal variations in temperature at 500 hPa in the peninsular region during both the seasons.....	30
Figure 5.12: Seasonal variations in temperature at 500 hPa in the mid-latitude region during both the seasons.....	31
Figure 5.13: Geo-MDX query	32
Figure 5.14: Inter-Seasonal Variation in temperature at different pressure levels for the Indian region.	33

LIST OF TABLES

Table 3.1: Header record format of the IGRA data file	10
Table 3.2: Data format of the IGRA data file	10
Table 3.3: Climatological regions and stations of the Indian subcontinent.....	12
Table 3.4: Details of the hardware used	13
Table 3.5: Details of the software used.....	13
Table 5.1: Temperature Inversion	34

1. INTRODUCTION

1.1. Background

The weather balloon (radiosonde) data is obtained from radiosonde, where instruments are tethered to balloons that are launched from the ground, ascend through the troposphere into the stratosphere, and transmit back to a receiving station on the ground. This data investigates how the meteorological variables changes with altitude with respect to four calendar based seasons which are generally recognized as winter (December to February), pre-monsoon (March to May), summer (June to August) and post-monsoon (September to November) [1]. The Indian subcontinent, compared to other countries in the world is known for the diversities in weather conditions from region to region due to its vast latitudinal extent from 8° – 34° North. Some of the important weather related events affecting it are heat waves, cold waves, landslides, prolonged fog, snowfall, intense rain events, floods, droughts, and cyclonic storms. In addition to this there is also issue of large inter-seasonal variations in weather [2]. A considerate amount of variation is observed in the atmospheric temperature with respect to the significant pressure levels at different seasons and regions. For the Indian subcontinent, major variations in the atmospheric temperature are observed at 850, 700 and 500 hPa atmospheric pressure levels. These levels correspond to different geopotential height ranging from 0 – 1.5, 1.5 – 3 and 3 – 5 kilometres respectively.

Based on the latitudinal extent, the Indian subcontinent is categorized into three climatological regions- peninsular (0° - 20°) N, mid-latitude (20° - 30°) N and humid sub-tropical (30° – 45°) N. These regions consist of 64 stations distributed along the latitudinal extent. All these regions are influenced by different external factors resulting in the variations in the inter-seasonal patterns. The Peninsular region is influenced by the climatology of sea. The Mid-Latitude region is always under the influence of western disturbances [3]. The Humid Sub-Tropical region is influenced by the Himalayan region. Thus it becomes essential to analyze the variations in the atmospheric temperature over the regions to understand the changes inter-seasonally in order to categorize those as scientific evidences.

A traditional database management system is not capable enough to provide analytical capabilities to analyze and visualize this kind of multidimensional data; hence SOLAP has been developed to fully exploit the powerful concepts brought by the multidimensional database structure to add spatial extensions that provide highly interactive map visualization and data exploration. The SOLAP approach supports the iterative nature of the analytical process because it allows the user to explore and navigate across the different themes (dimensions) at different levels of detail and to rapidly visualize the facts or data at the intersections of these dimensions, whatever their level of aggregation [4].

1.2. Motivation and problem statement

Recent developments in large storage solutions have led to the rise of spatial data warehousing (SDW) and spatial online analytical processing (SOLAP) to store, manage and analyze large volumes of data. Currently, a single system that combines all the necessary features of SDW and SOLAP does not exist in the meteorological domain. Hence, it is anticipated that our system will allow for more comprehensive and thorough analysis of the multidimensional radiosonde data, leading to clear visualization of trends in inter-

seasonal patterns of the atmospheric temperature over the different climatological regions of the Indian subcontinent. This study aims at designing a system that can handle huge volume of data, perform SOLAP operations, and displays graphically the trends in inter-seasonal patterns of the atmospheric temperature with respect to regions.

SDW is a huge repository of data that provides means of carrying out spatial analysis together with agile and flexible multidimensional analytical queries performed by SOLAP [5]. SOLAP is defined as a visual platform especially built to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach [6]. SDW follows the extract, transform and load (ETL) process. These are the data pre- processing steps for the creation of database. The data is integrated from the distributed source files, cleaned for noise and invalid records and loaded into the database. Figure 1.1- shows the various components of SDW.

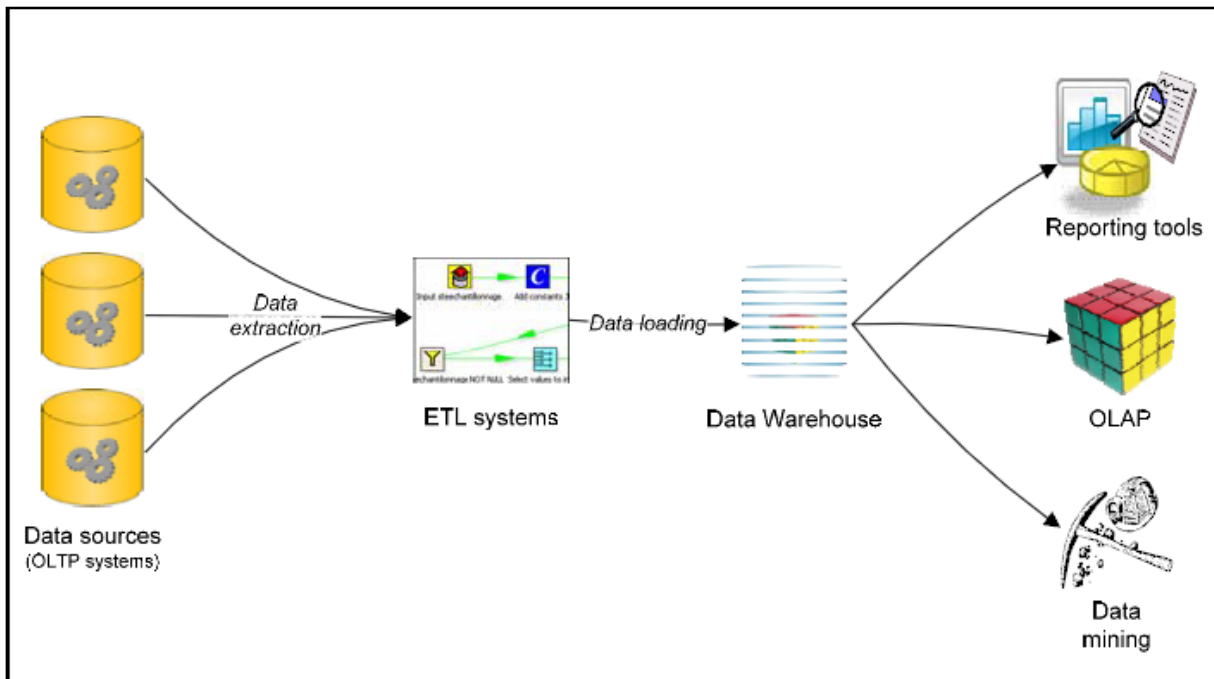


Figure 1.1: Components of Spatial Data Warehouse

The proposed SOLAP system architecture will consist of following main components: Apache Tomcat server, a system-based web applications enhancement to support deployment across the variety of environments, PostGIS spatial database management system, Geomondrian, an open source SOLAP server compatible with PostGIS. It will provide spatial extensions to the Multidimensional expressions (MDX) query language, and allows embedding of spatial analysis capabilities into analytical queries. Workbench, will allow creating SOLAP logical models using extensible markup language (XMLA) and Spatialytics a java based visualization tool will publish the result on its graphic user interface. These components will interact with each other to visualize and explore spatial information to perform SOLAP operations on the radiosonde data. It will also provide the technology to execute SOLAP queries with the help of MDX editor and, processes the data accordingly using SOLAP operations like roll up and drill down to manipulate the view of the data. The roll-up operation leads to more general view of the data. It is used to increase the granularity of a view for a specific set of feature dimensions, whereas a drill-down operation increases the granularity of a dimension and thus specializes the analysis of a measure [7].

1.3. Research Identification

1.3.1. Research Objective

This research aims to design a SOLAP based system to organize radiosonde data by applying the characteristic features of SDW and visualize the trends in inter-seasonal patterns with the altitudinal variations of the atmospheric temperature for the Indian subcontinent over a period of 20 years.

1.3.2. Sub-objectives and Research Questions

1. To identify suitable methods to perform the extract, transform and load (ETL) operations for the creation of a database with the radiosonde data.
 - What are the different data pre-processing techniques to be used for the database creation?
2. To setup the servers and develop the graphical user interface for the visualization of the trends inter-seasonal pattern of atmospheric temperature.
 - Which is the most suitable and server compatible GUI for the visualization of the results?
3. To create the multidimensional cubes by using the SOLAP queries.
 - What is the criterion for constructing the multidimensional cube using the SOLAP queries?
4. To perform the SOLAP operations, roll up and drill down on the multidimensional cubes to visualize the trends in inter-seasonal patterns of atmospheric temperature for the Indian subcontinent.
 - What is the trend in inter-seasonal patterns in the altitudinal variation of atmospheric temperature for Indian subcontinent over the years?

1.4. Innovation aimed at

As per literature review, there is no existing environment for the visualization of multidimensional radiosonde data on a SOLAP server. It is for the first time that GeoMondrian (SOLAP server) and Workbench have been used for research purpose. The innovation in this research is to perform inter-seasonal pattern analysis on multidimensional radiosonde data with the help of SOLAP technology for the Indian subcontinent. Based on the above research objectives and questions, the proposed SOLAP design will allow the user to create multidimensional cubes of the meteorological parameters and perform various SOLAP operations to visualize the trends in inter seasonal patterns of atmospheric temperature.

1.5. Thesis structure

The research work is organized as follows:

Chapter 1: Introduction, this chapter presents general overview about the research work. It describes the topic, motivation, problem statement, research objectives, and research questions.

Chapter 2: Literature Review, this chapter deals with the background of the study. It briefly explains important components of the research: GeoMondrian server, apache tomcat server, multi-dimensional

expressions (MDX), extensible markup language for analysis (XMLA), Work Bench, multidimensional cubes, schema, and Spatialytics (Discover).

Chapter 3: Data set, study area and tools, this chapter describes in detail the structure of the data, the study area and the software and hardware tools required.

Chapter 4: Design and Implementation, this chapter details about the high level architecture of the functionality and detailed implementation of each component of the architecture.

Chapter 5: Results and Discussions, this chapter discusses the results obtained at each implementation step, its discussion and analysis.

Chapter 6: Conclusion and Recommendations, this section provides the answer of the research questions highlights recommendations for further study.

2. LITERATURE REVIEW

In this chapter, we review the existing research on SDW, SOLAP and associated implementation strategies, leading to the list of essential requirements for designing a SOLAP system for radiosonde data.

2.1. Related work in SOLAP and SDW

To describe the trajectory of hurricanes, a SOLAP framework has been developed by using the weather balloon and flight telemetry system [8]. This framework has also used the SOLAP queries to construct weather events data cube for the analysis of hurricane trajectory with respect to wind speed and wind direction. Construction of the multidimensional cubes is also been shown in this research.

SOLAP concepts has also been used to support the spatio-temporal exploration of data to present the geovisualization, interactivity, and animation features in the SDW [4]. SOLAP provides platform for handling historical voluminous data. It provides storage, manipulation and retrieval of analytical queries, using MDX. The multi-dimensional data can also be represented in the form of multidimensional cubes and it supports various analytical operations as well. It also provides the time based analytics with visual interpretations. It supports all the ETL operations. It provides the user an interactive design to explore the multidimensional spatio-temporal data. This research describes the properties and characteristics of the SOLAP to support rapid and easy spatio temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram, as shown in the Figure 2.1

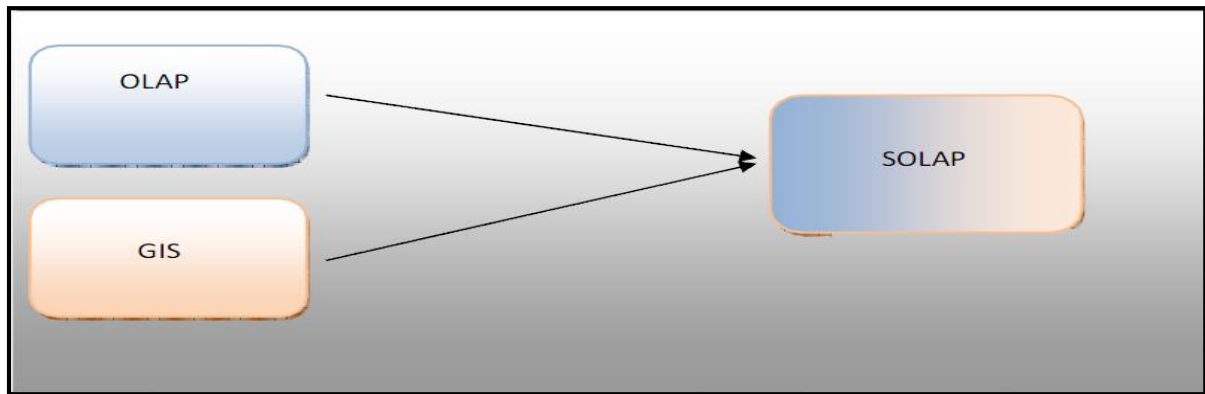


Figure 2.1: SOLAP (Integration of GIS and OLAP)

SDW aims at effective and efficient querying of spatial data. Spatial databases are suited for answering regular transactional queries where there is not a lot of historical component or aggregation. They are based on the concepts of Data warehouses (DW) and additionally provide support to store, index, and aggregate and analyze spatial data [9]. A data cube is a lattice of cuboids which represents hierarchies. The data cube may have cells which are pre computed for efficient query processing. Cube is a multidimensional, hierarchical data structure for analysis purposes. Dimension is the analysis axis in a cube. It can be thematic, temporal or spatial. Measure is a variable in a cube, representing the measured phenomenon. Hierarchy is the set of levels forming a roll-up path for members in a dimension. There can be more than one hierarchy in a dimension. These concepts are extended to spatial data in a spatial data warehouse. Integration of spatial data and data cubes make many interesting spatial aggregation queries possible. The Figure 2.2- shows the taxonomy of the Spatial and Spatiotemporal Data warehouse.

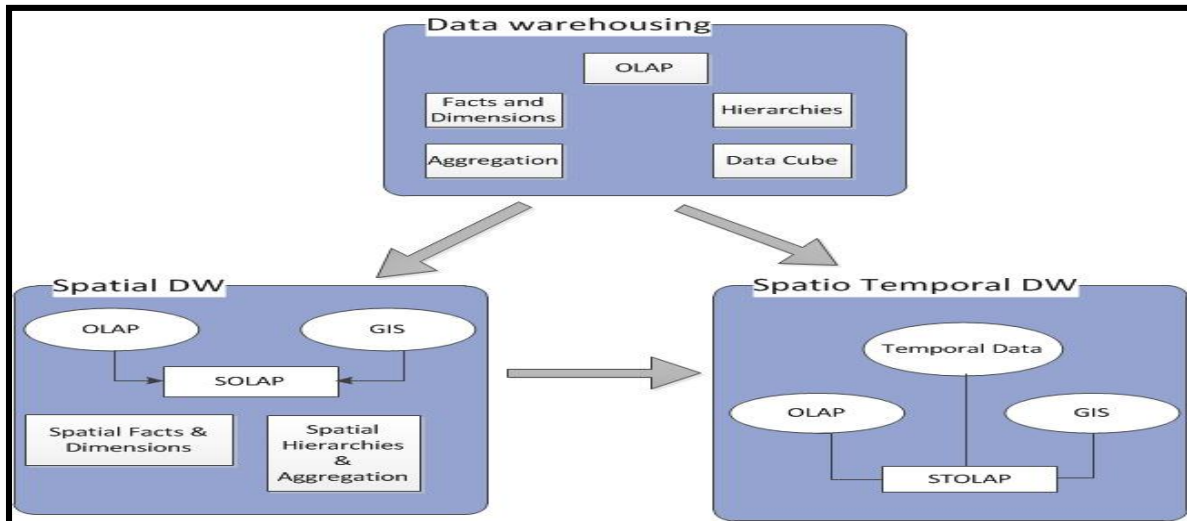


Figure 2.2: The Taxonomy of Spatial and Spatiotemporal data warehouse

The relevance of SOLAP has been widely recognized in several application domains like marketing, agriculture, environmental risk management. Designing multidimensional cubes for the agriculture data using OLAP [10] reflect the characteristic applications of SDW. This framework describes the various components of SDW. Highlights of this research are the comparisons between the star and snowflake schemas for developing the database. Based on the comparisons and our requirements, Star schemas has been selected as it characteristically consists of fact tables linked to associated dimension tables via primary/ foreign key relationships [11]. Each dimension in a star schema is represented with only one-dimension table. The dimension table contains the set of attributes. Star schema provide faster performance as the fact table is directly linked with the dimension tables hence, lesser number of joins are involved in the query processing. A representation of star schema is shown in the Figure 2.3.

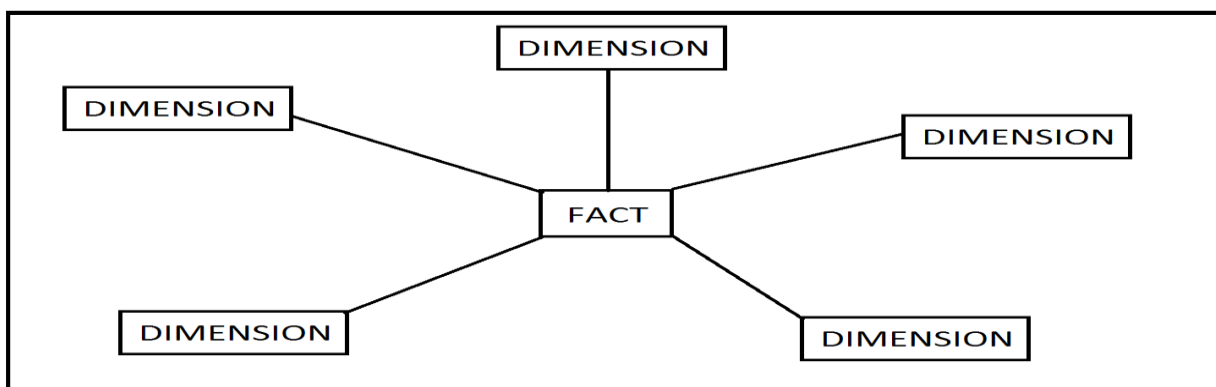


Figure 2.3: Star Schema Representation

It has also summarised the various SOLAP operations like, roll-up for summarization of data for the higher level of a hierarchy i.e. from lower level to upper level. Roll down (drill down, or drill through) for navigation among levels of data ranging from higher level (up) to lower level or detailed data (down). Slicing, allowing the selection of all data satisfying a condition along a particular dimension. Pivoting for changing the dimensional orientation of the cube, e.g. swapping the rows and columns, or moving one of the row dimensions into the column dimension, etc.

There are various frameworks constructed for trajectory data to investigate the behaviour of objects inside a given area as well as the movements of objects between areas in the same neighbourhood [12]. This data warehouse model is sufficiently flexible to deal with different types of mobile objects. Traditional spatial data warehouses do not support spatial vagueness. The VSCube conceptual model constitutes an expressive method to represent vague spatial data and spatial aggregation/predicates in SOLAP queries [13].

There are several location- based data warehouses for modelling and querying of uncertain data in SOLAP to support location-based services. The model allows for using probability distributions in facts/dimensions and algebra defines how to aggregate uncertain data over uncertain hierarchies [14]. To overcome the limit of OLAP analysis, location intelligence tools and techniques are integrated with geographical dimension into BI platforms, to enhance their capability of discovering patterns with specific reference to their territorial distribution [15].

The Spatial OLAP Visualization and Analysis Tool (SOVAT) is a new type of research application for community health assessments [16]. SOVAT integrates into one system many of the necessary characteristics needed to make comprehensive community health decisions. To process spatial OLAP query in spatial data warehouse, an OLAP-favoured search method which can utilize the pre-aggregation result in spatial data warehouse to improve the performance of spatial OLAP queries is developed [17].

2.2. GeoMondrian

GeoMondrian is an Open Source SOLAP java based Server, a spatially-enabled version of Pentaho Analysis Services (Mondrian) [18]. It is the first implementation of a true SOLAP server. It provides a consistent integration of spatial objects into the OLAP data cube structure, instead of fetching them from a separate spatial database, web service or GIS file. To make a simple analogy, GeoMondrian brings to the Mondrian OLAP server what PostGIS brings to the PostgreSQL database management system. It implements a native geometry data type and provides spatial extensions to the MDX query language, allowing embedding spatial analysis capabilities into analytical queries. It supports PostGIS based DW's.

Multidimensional Expressions (MDX) is the query language for retrieving multidimensional data. MDX is based on the XML for Analysis (XMLA) specification, with specific extensions for SQL Server Analysis Services. MDX utilizes expressions composed of identifiers, values, statements, functions, and operators that Analysis Services can evaluate to retrieve an object. Provides a specialized syntax for querying and manipulating the multidimensional data stored in OLAP cubes [19].

At the first look, MDX query looks like a simple SQL query but the difference lies in the syntax and the data model which it follows. A typical MDX query syntax is shown in the Figure 2.4.

```
SELECT
{ AXES OF MULTIDIMENSIONAL RESULT SET } ON COLUMNS,
{ AXES OF MULTIDIMENSIONAL RESULT SET } ON
ROWS
FROM ( NAME OF THE SOURCE CUBE )
WHERE ( [ SLICER ( FILTER ) ] )
```

Figure 2.4: MDX Query structure

- **SELECT** - Defines the axes on the multidimensional result set (COLUMNS, ROWS) MDX allows up to 128 axes on a query.

- FROM - Name of the source cube.
- WHERE - Slicer (filter) axis. It is the limiting factor.

2.3. WorkBench

The GeoMondrian Workbench allows visually creating and testing SOLAP cube schemas [20]. It provides the features of: Schema Explorer: An editor integrated with the underlying data source for validation. JDBC Explorer: Browses the connected database's structure. XMLA editor: for building a schema consisting of cubes, hierarchies and members and then mapping of this model to a visualization tool. XML for Analysis is a standard for accessing OLAP servers via SOAP (Simple Object Access Protocol) [21]. It is written as simple XML tags. It is also used for publishing the results of the workbench on the server.

Discover is a cartographic component of Spatialytics which enables navigation of SOLAP data cubes [22]. It is a java based web mapping client for the visualization of SOLAP data cubes. It aims to be integrated into existing dashboard frameworks in order to produce interactive geo-analytical dashboards. The Server relies on an embedded GeoMondrian instance API. The JavaScript client connects to server, using XML request calls to query a cube (with an MDX query). The response is a multidimensional data set. Styling rules are applied for the rendering of the output of the MDX and XMLA.

3. DATA SETS, STUDY AREA AND TOOLS

This chapter is divided into three sections; the first section describes the data and its complete structure. The second section describes the study area in detail. The third section describes the list of hardware and software tools required to successfully carry out the research activities.

3.1. Data

The data is compiled from the existing site <http://www.ncdc.noaa.gov/igra> of National Climatic Data Centre (NCDC). The Integrated Global Radiosonde Archive (IGRA) consists of radiosonde and pilot balloon observations at over 1500 globally distributed stations as shown in the figure 3.1, with varying periods of record, many of which extend from the 1960s to present [23].

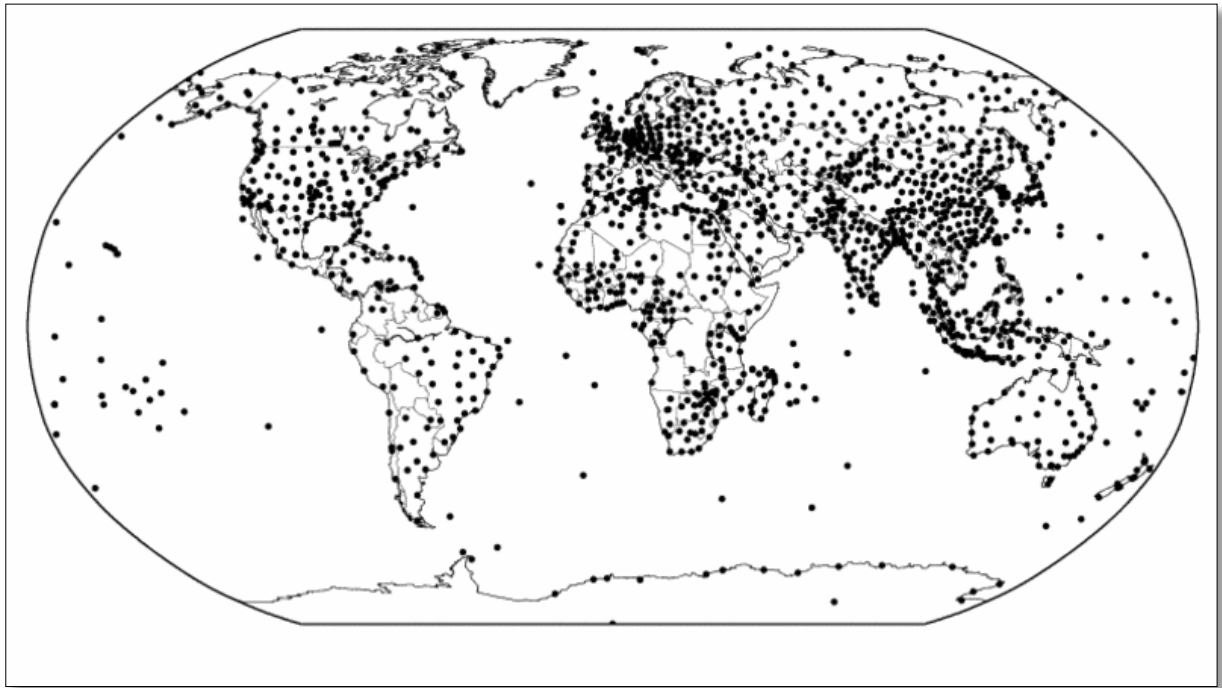


Figure 3.1: Locations of 1500 stations of IGRA

Observations include vertical profiles of pressure, temperature, geopotential height, dewpoint depression, wind direction and wind speed at standard, surface and tropopause, and significant levels. Radiosonde daily data at 00UTC for the period 1992–2012 has been used to compute the variations in inter-seasonal pattern of the atmospheric temperature over the Indian subcontinent for pre and post monsoon seasons.

3.1.1. Format description of data files

- Each file per station is given #####.dat
is the station number, and station number is obtained from igra-stations.txt file.
- Each sounding consists of one header record and n data records, where n (given in the header record) is the number of levels in the sounding.

- For all variables in the data record except the level type indicators, -8888 indicates that the original value has been removed by one of the IGRA quality assurance procedures, while -9999 means that the data value was never present.
- For each pressure, geopotential height, and temperature value, a one-character quality assurance flag indicates whether the corresponding value was checked by procedures based on climatological means and standard deviations. Possible flag values are:
 - i. Blank = no climatological check applied due to an insufficient number of data values for computing the relevant statistics,
 - ii. A = value falls within "tier-1" climatological limits based on all days of the year and all times of day at the station, and
 - iii. B = value passes checks based on both the tier-1 climatology and a "tier-2" climatology specific to the time of year and time of day of the data value.

3.1.2. Header format

Each data file is divided into two parts, the header and data formats. Each data file has all the 6 meteorological parameters available along with their respective flag values. The header format of the data file is shown in the Table 3.1. A sample header data looks like- **(#0100419930413129999 71)**

Table 3.1: Header record format of the IGRA data file

Variable Name	Columns Description
Header Record Indicator	1- 1 # character
Station Number	2- 6 WMO station number
Year	7- 10
Month	11- 12
Day	13- 14
13- 14	15- 16 00-23 UTC
Release Time	17- 20 0000-2359 UTC, 9999 = missing
Number of levels	21- 24 number of subsequent data records

3.1.3. Data format

The Data format of the data file of IGRA is shown in the Table 3.2. A sample data format looks like- **(21102900B 8 -193B 29 220 10)**

Table 3.2: Data format of the IGRA data file

Variable Name	Columns Description
Major Level Type	1- 1 1 = standard pressure level 2 = significant thermodynamic level 3 = additional wind level
Minor Level Type	2- 2 1 = surface, 2 = tropopause, 0 = other
Pressure	3- 8 units of Pa (mb * 100)
Pressure Flag	9- 9 A, B, or blank
Geopotential Height	10- 14 units of meters

Geopotential Height Flag	15- 15 A, B, or blank
Temperature	16- 20 units of degrees C * 10
Temperature Flag	21- 21 A, B, or blank
Dew point Depression	22- 26 units of degrees C * 10
Wind Direction	27- 31 units of degrees (0-360, inclusive)
Wind Speed	32- 36 units of (m/s)*10

3.2. Study Area

The study area used for this research is the Indian subcontinent. The spatial extent of India is 4.5°N to 38.5°N in latitudes and from 65.5°E to 98.5°E in longitudes. The time period of the study is from 1992 to 2012 i.e. two decades. Figure 3.2 shows the study area.

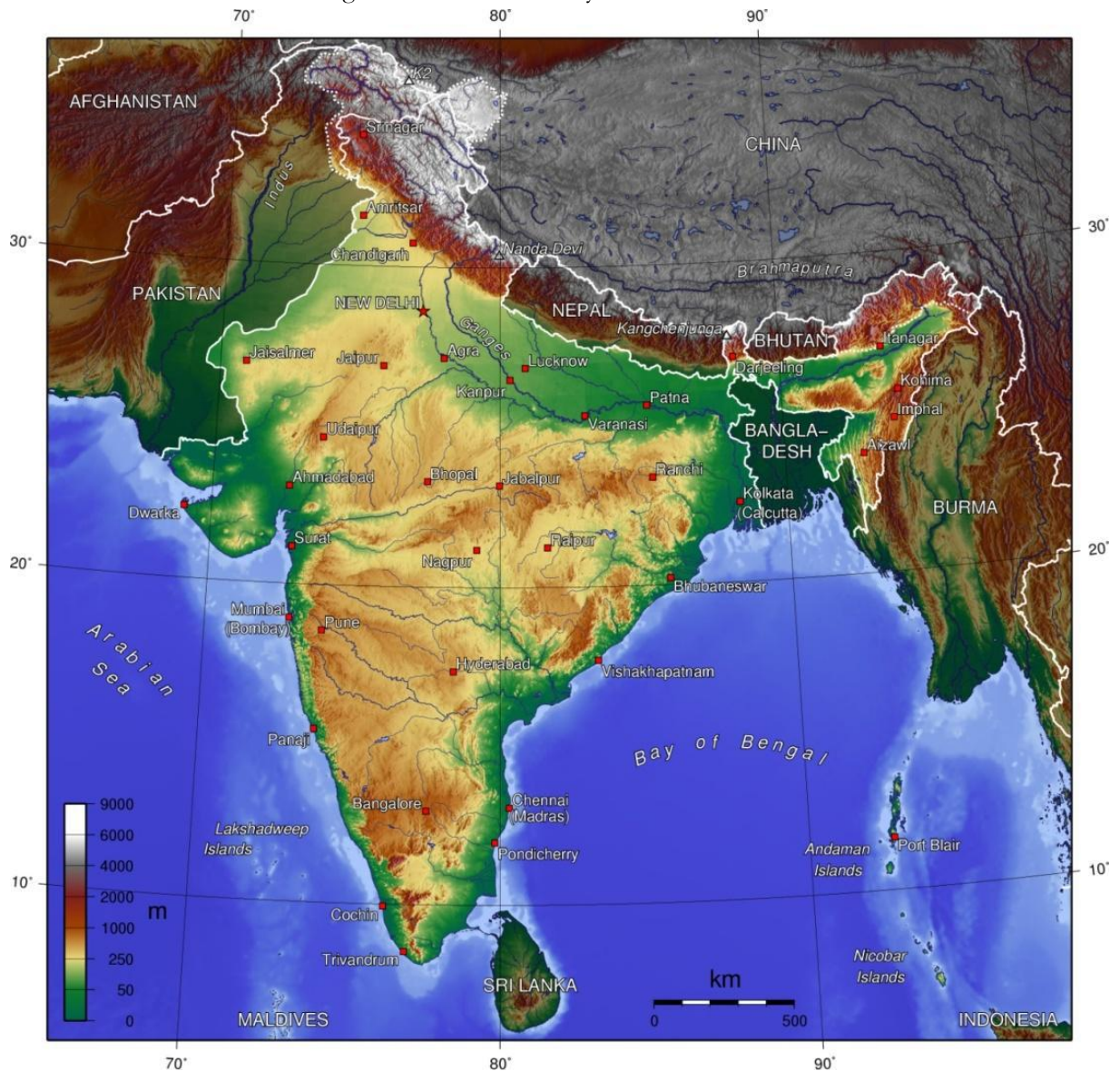


Figure 3.2: Study Area - India

Based on the longitudinal extent, the Indian subcontinent is divided into 3 climatological regions which covers all the 64 stations. Out of these 64 stations, only for 27 stations the daily data from 1992 to 2012 was available. The climatological region and station distribution is shown in the table 3.3.

Table 3.3: Climatological regions and stations of the Indian subcontinent

REGION	STATION_NAME	LATITUDE	LONGITUDE
Humid Sub-Tropical (30° – 45°)N	SRINAGAR	34.08	74.83
	PATIALA	30.33	76.47
MID_LATITUDE (20°-30°) N	NEW DELHI	28.58	77.2
	DIBRUGARH	27.48	95.02
	SILIGURI	26.67	88.37
	LUCKNOW	26.75	80.88
	GORAKHPUR	26.75	83.37
	JODHPUR	26.3	73.02
	GWALIOR	26.23	78.25
	GUWAHATI	26.1	91.58
	PATNA	25.6	85.1
	AGARTALA	23.88	91.25
	RANCHI	23.32	85.32
	BHOPAL	23.28	77.35
	AHMEDABAD	23.07	72.63
	KOLKATA	22.65	88.45
	RAIPUR	21.23	81.65
	BHUBANESHWAR	20.25	85.83
PENINSULAR (0°-20°) N	MUMBAI	19.12	72.85
	HYDERABAD	17.45	78.47
	PANJIM	15.48	73.82
	CHENNAI	13	80.18
	BENGALURU	12.97	77.58
	PORT BLAIR	11.67	92.72
	COCHIN	9.93	76.23
	THIRUVANANTHAPURAM	8.48	76.95
	MINICOY	8.3	73

The peninsular region consists of 9 stations, the mid latitude region consists of 16 stations and the humid sub-tropical region consists of only two stations respectively. The stations provide the full latitudinal coverage for our analysis. The inter-seasonal pattern analysis is based on the seasonal cycle which consists of the four seasons, namely:

- Winter - December-January- February (DJF)
- Pre Monsoon- March-April-May- (MAM)
- Summer - June-July-August-(JJA)
- Post Monsoon - September-October-November- (SON)

Out of these four, we have considered only the pre and post monsoon seasons. During these seasons all the regions in India experience dry atmosphere hence less atmospheric variability is observed in these seasons.

3.3. Tools used

3.3.1. Hardware tools

The hardware used for this research has been listed in table 3.4

Table 3.4: Details of the hardware used

Processor	Intel Core i5-2450M 2.40GHz
Ram	4 GB
Graphics card	AMD Radeon HD 7650M GPU
Video Ram	2 GB

3.3.2. Software tools

The software/packages used for this research have been listed in Table 3.5.

Table 3.5: Details of the software used

S.N.	Software/Packages	Purpose
1	Geomondrian 1.0	To generate support and generate MDX queries.
2	Apache tomcat 7.0.55	To host the SOLAP server
3	PostGIS 2.0	To be used as radiosonde database.
4	Work Bench 1.0	To generate the XMLA
5	Spatialytics 1.0	To Visualize the results

4. DESIGN AND IMPLEMENTATION

This chapter describes the high level architecture SDW implemented in our work, followed by detailed explanation of implementation of each component.

4.1. High level architecture

The high level architecture of SDW and its various components are shown in the Figure 4.1

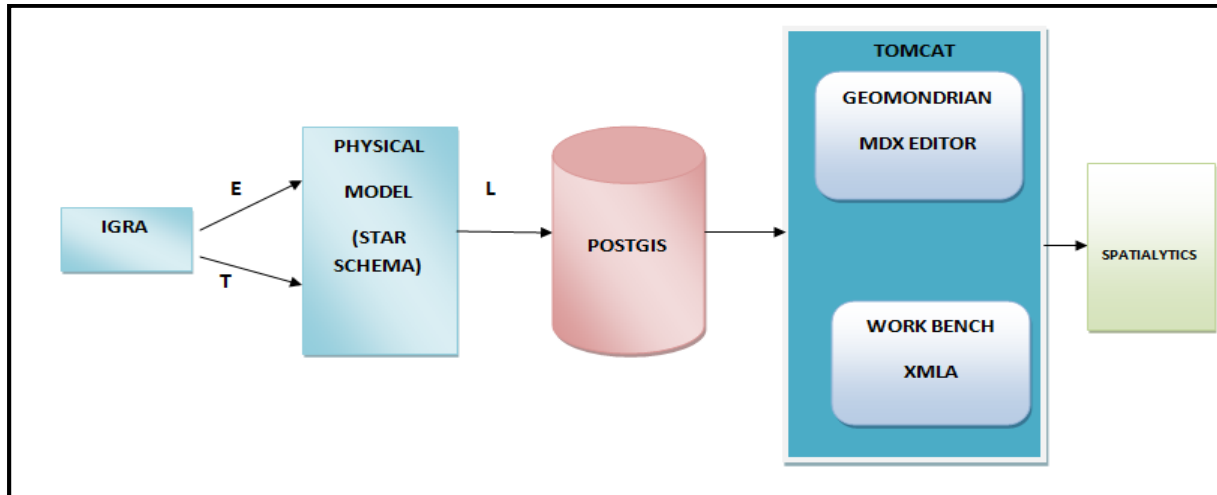


Figure 4.1: SDW Architecture

IGRA- The first component of the SDW architecture is the source of the data i.e. IGRA. The ASCII format daily data for the 64 stations of the Indian subcontinent is selected and downloaded. They are arranged with respect to regions.

Star Schema- The downloaded data undergoes the ETL operations. A logical model of the data is prepared using the star schema and then converted to a physical model by loading the data into the database.

PostGIS database- This is one of the key components of the SDW architecture. The Physical modelling of the data is done by converting the logical model into a physical model in the form of tables. Tables are created keeping in mind the structure of the star schema.

Apache Tomcat Server- This is another key component of the SDW architecture. It is a web container mounted on Java Runtime Environment (JRE). It supports user- as well as system-based web applications enhancement for the deployment across the variety of environments. It acts as base server for the SOLAP server.

GeoMondrian Server- It is the first implementation of a true SOLAP servers mounted on apache tomcat. It provides a consistent integration of spatial objects into the SOLAP data cube structure. It is compatible and connected to the PostGIS database and provides a platform for performing the SOLAP operations by using the MDX queries. It comes with MDX editor where all the query processing is done. It supports both spatial and non spatial MDX queries. MDX queries are a refined version of simple SQL (structured query language). It is used for querying multidimensional data and databases. It provides analytical capabilities to SQL.

WorkBench- The GeoMondrian Workbench allows visually creating and testing SOLAP cube schemas using XMLA. It is primarily used for building a schema consisting of cubes, hierarchies and members and then mapping of this model to a visualization tool.

Spatialytics (Discover) - It is a java based web mapping client for the visualization of SOLAP data cubes. It aims to be integrated into existing dashboard frameworks in order to produce interactive geo-analytical dashboards.

The proposed methodology adopted for this research is based on the above explained SDW architecture. The methodology is divided into 3 phases; Extract, transform and load (ETL) i.e. the database creation, the server setup, and finally the evaluation of results using the visualization tool. The methodology has been designed to meet the objectives and answer the respective research questions.

4.2. Database creation (ETL)

The database creation of the radiosonde data is divided into 3 steps.

4.2.1. Extract (E)

Daily radiosonde data from IGRA at 00 UTC for 64 stations well distributed over Indian subcontinent for the period 1971–2012 were extracted using online extraction technique. In this technique the data files are directly downloaded online from the NCDC data server of IGRA.

4.2.2. Transform (T)

The extracted data is transformed into a conceptual model using the schema creation of the dataset. The Star Schema is used for creating the conceptual model. Figure 4.2- shows the star schema implementation.

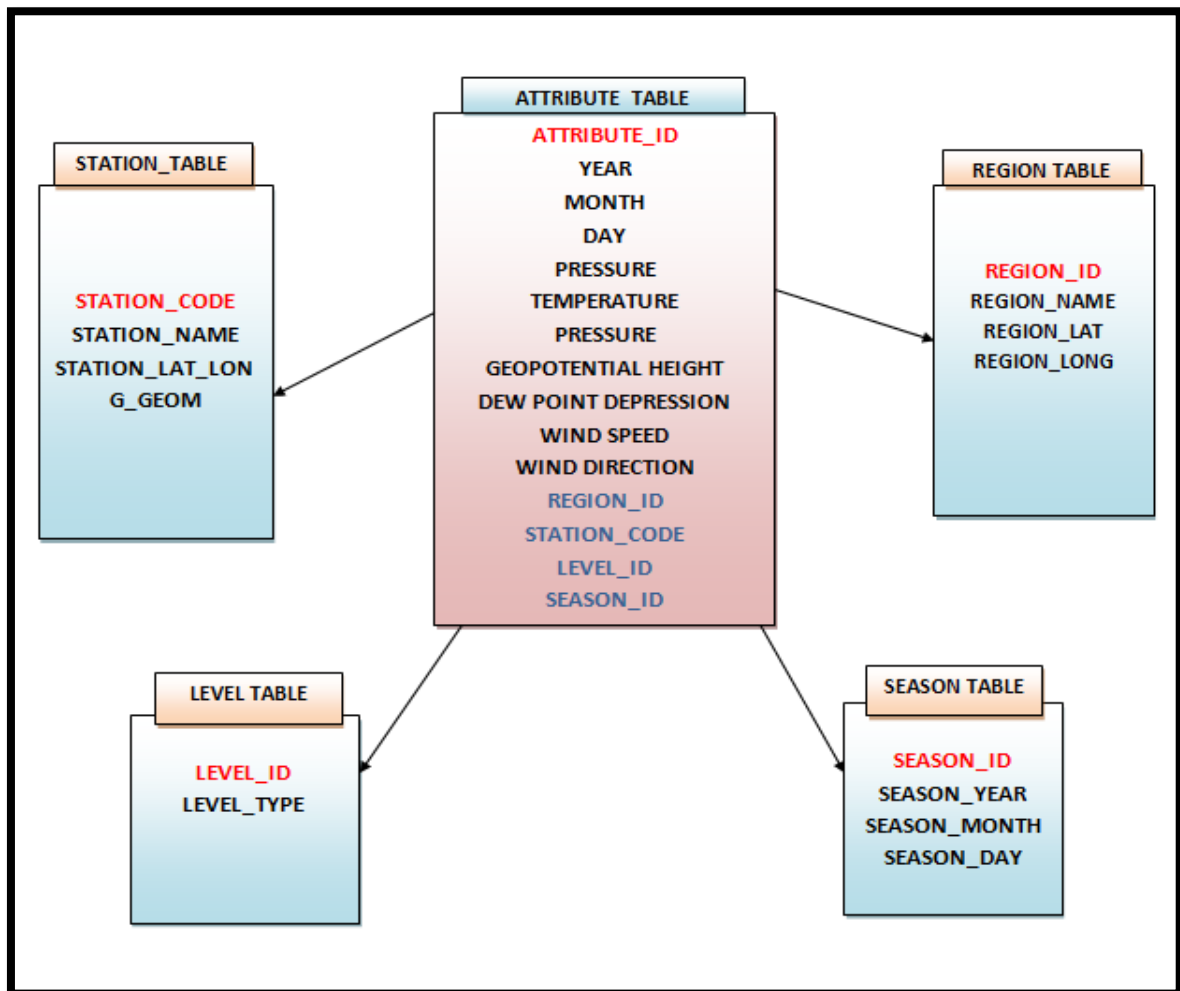


Figure 4.2: Conceptual Model of the dataset using Star Schema

The star Schema implementation consists of the base table i.e. the fact table and its corresponding dimension tables. The Attribute table is the fact table which is used for the creation of the

multidimensional cubes. The table consists of the year column which indicates the year for which the data has been captured. The month and day column highlights the corresponding month and day at which the data were captured. Temperature, pressure, geopotential height, dew point depression, wind speed and wind direction columns indicate their respective values corresponding to the year month and day. The remaining tables were the dimension tables and were connected to the fact table by the primary and foreign key values. Each primary key of the dimension tables has been used as the foreign key of the fact table. As the dimension tables were linked directly to the fact table, the number of joins required for querying is reduced and the execution is optimized. The station table and the region table were the spatial tables with point geometric columns. The attributes highlighted in red represent the primary key attributes and the one in blue represent the foreign key attributes.

4.2.3. Load (L)

The next process was to load the data into the database by converting the conceptual model in to a physical model by using SQL scripts. The scripts includes the creation of the table, insertion of records, integrity constraints and data manipulation operations. The name of the database has been kept as radiosonde_database. Figure 4.3- shows the scripts used for the creation database in PostGIS.

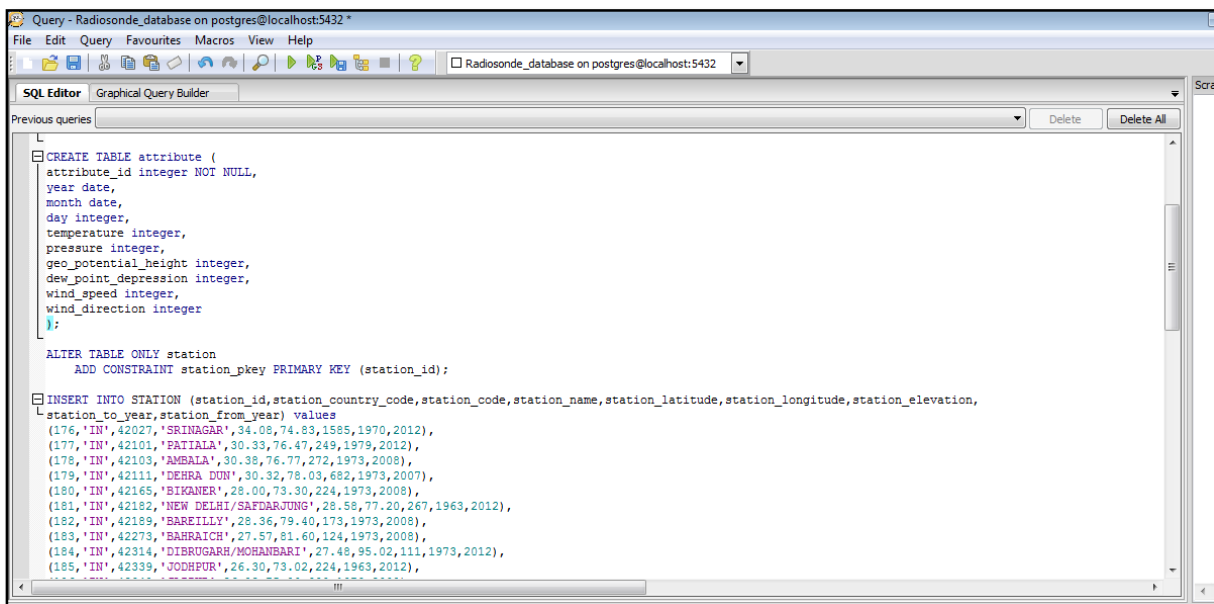


Figure 4.3: Scripts used for the creation of the radiosonde database.

The data loaded into the database had lot of garbage values i.e. -9999. The cleaning of the data was done to replace the garbage values with “zero”. For few stations there were data available only for selected years. In order to maintain consistency in the data, two decades data were selected from 1992 to 2012. As a result of this we were left out with only 27 stations data distributed across the three regions.

4.3. Server setup

This section describes the installation and configuration of the servers in detail. These servers had played an important role in setting up the environment for the performing the SOLAP operations.

4.3.1. Java runtime environment (JRE)

As both the servers used in this study are java based, therefore before installing the servers, the JRE has to be configured. The environment variables have to be set to the java folder. Figure 4.4- shows the configuring of the environment and system variables.

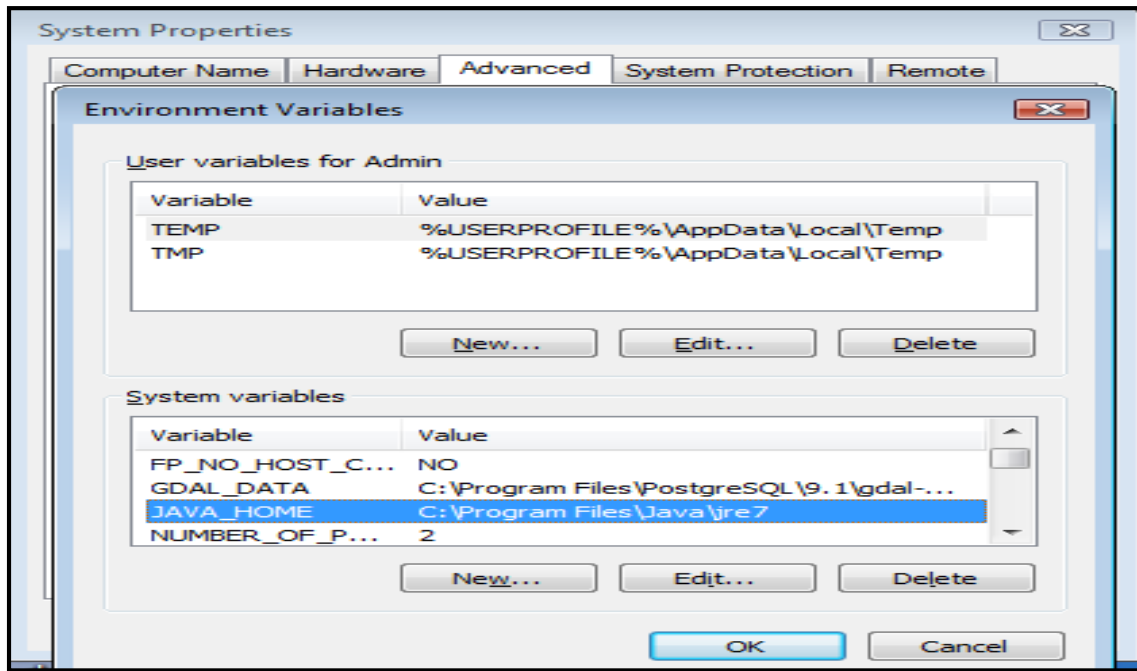


Figure 4.4: System and Environment Variables

Apache Tomcat Server version 7.0 is downloaded and installed and port number is set for local host. Server is configured and started by opening local host in browser with defined port number 8080. Figure 4.5- shows the successful starting of the tomcat server.

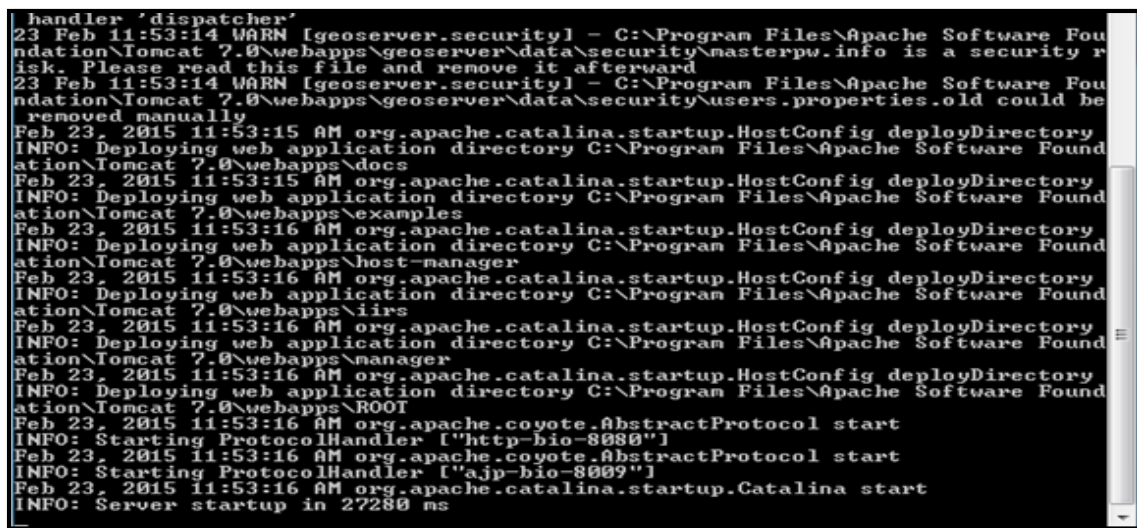


Figure 4.5: Startup of Tomcat Server

Tomcat web application manager is configured by making the changes in the (server.xml) file. Figure 4.6- shows the console of the web application manager.

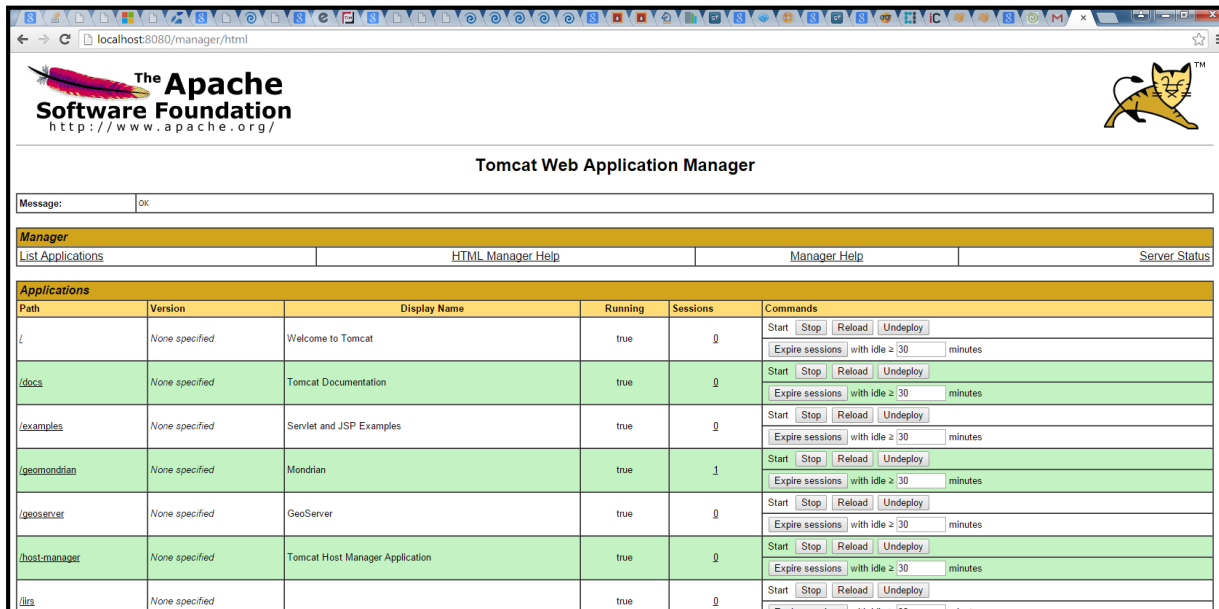


Figure 4.6: Tomcat Web Application Manager

4.3.2. GeoMondrian (SOLAP) SERVER

After the successful installation and configuration of the apache tomcat server, the GeoMondrian server is installed. To install the SOLAP server, GeoMondrian .war file is deployed on the tomcat server. This creates an instance of the server on tomcat. The Figure 4.7- shows the successful installation of the GeoMondrian server.

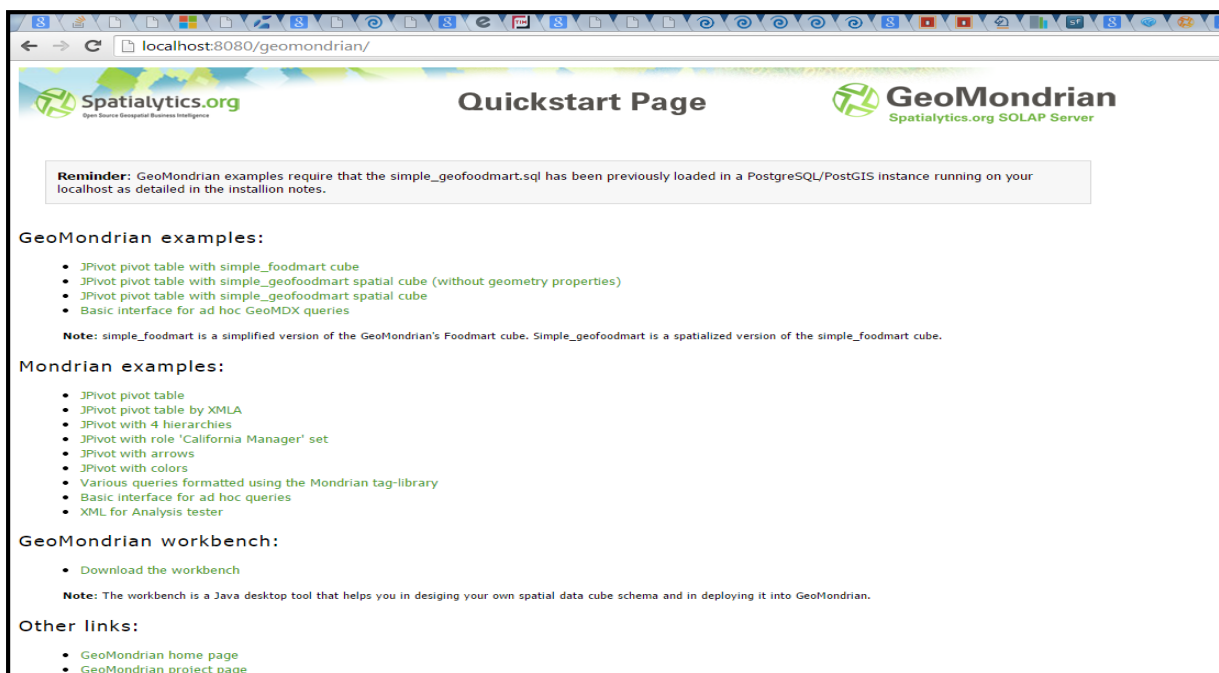


Figure 4.7: GeoMondrian Server homepage

4.3.3. Connecting the servers with the database

Java database connectivity (JDBC) drivers are used for establishing the connection of PostGIS with the Geomondrian. In the web.xml (configuration file) of tomcat, connection string is added along with the url

and the username and password. The default port for PostGIS is 5433 which is set at the time of installation. The connection is tested and saved. Figure 4.8 shows the successful connection of server with the database.

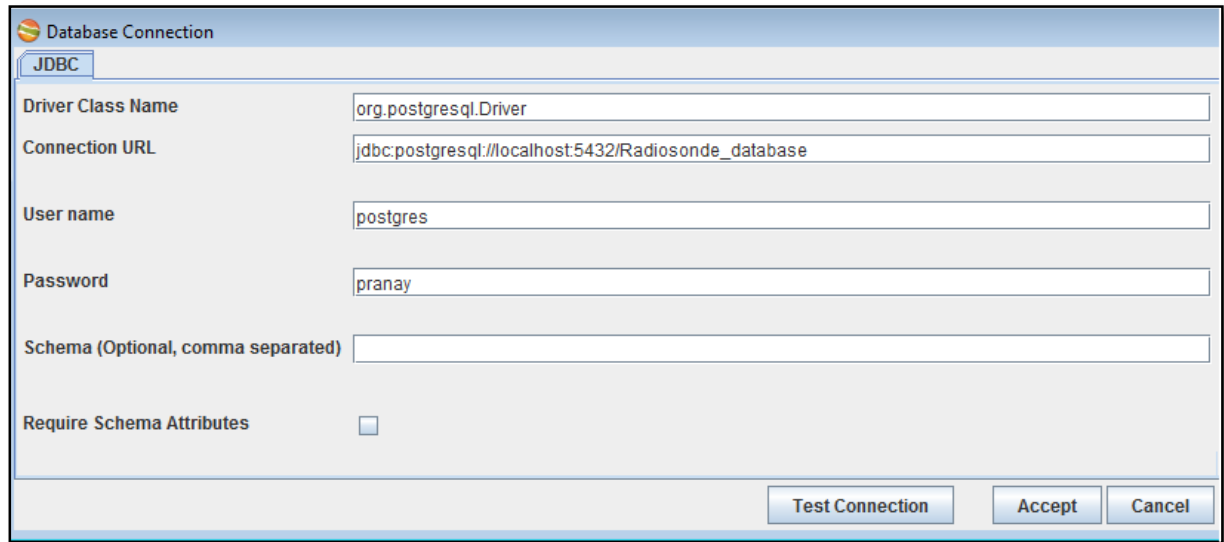


Figure 4.8: Successful connectivity of SOLAP server with database

4.4. Multidimensional cube creation

The biggest advantage of SOLAP is that it highlights the dimensionality of the data. The cube structure provides the analytical capabilities for visualizing complex hierarchical structure of the data, like the radiosonde data. In order to achieve the hierarchical structure, SOLAP operation namely drill down and drill up were performed using MDX and XMLA. It helps in analyzing the hierarchy structure of the data from daily level to seasonal level and station level to region levels. It also provides a better understanding for analysing the inter-seasonal patterns of atmospheric temperature at each level. Figure 4.9 shows the levels of hierarchy for performing the SOLAP operations on the multidimensional cube

- Daily data for one station for one year with all the parameters were selected and drill up operation was performed to get the data for all the years for all the stations.
- Average values were selected for the atmospheric temperature for based on the seasons and atmospheric pressure levels.
- Atmospheric temperature variance values are calculated for all the stations for the years using drill up operation.
- Specific date for which the temperature values for all the years for a particular station using drill down operation.
- Atmospheric temperature values for all the stations for the specific dates and significant pressure levels for 20 years based on region.

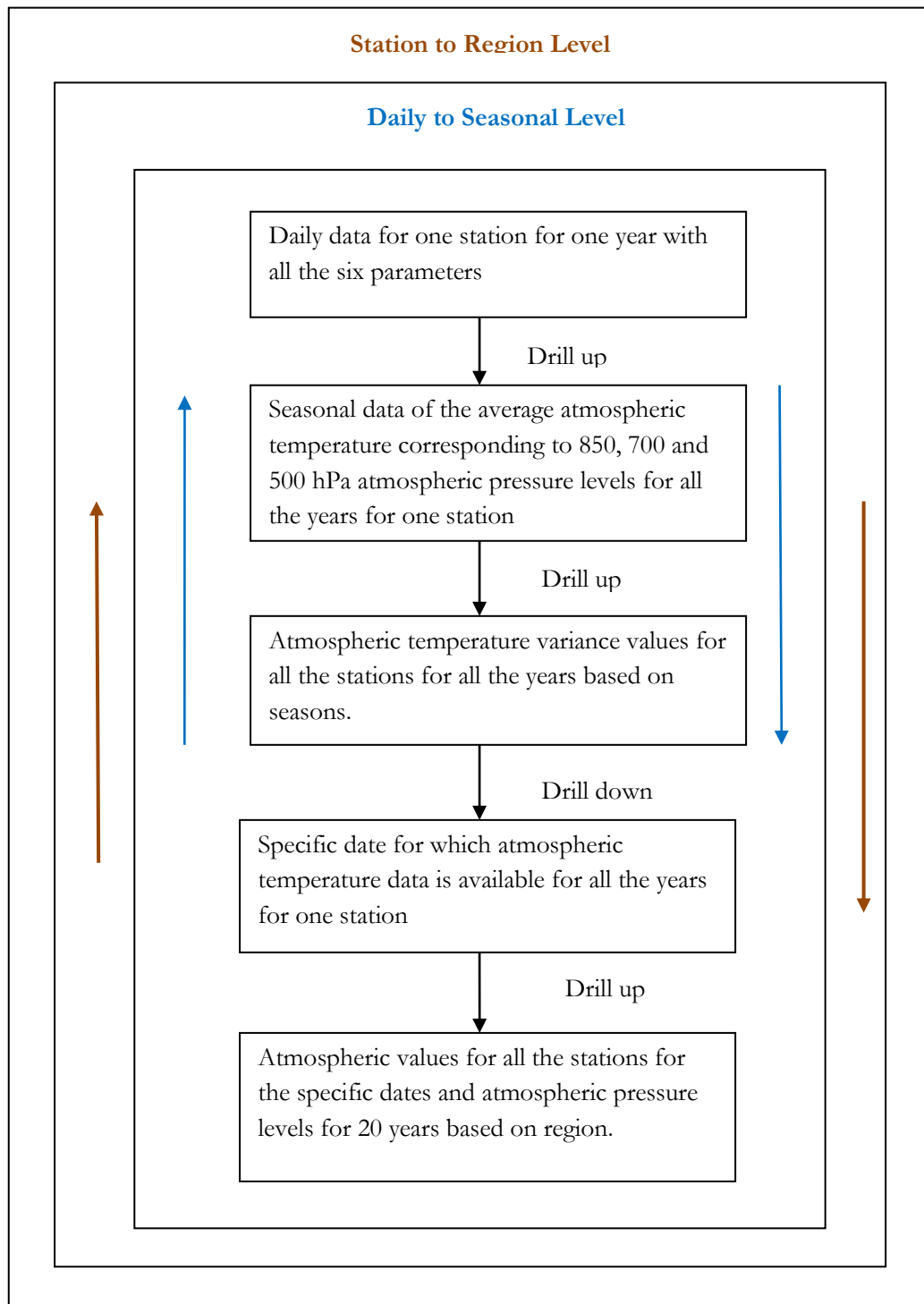


Figure 4.9- Levels of hierarchy used for performing SOLAP operations

5. RESULTS AND DISCUSSIONS

This chapter presents and discusses the results obtained at each implementation step.

5.1. Multidimensional cube construction and Pattern Evaluation

The Apache Tomcat and GeoMondrian (SOLAP) servers are successfully installed and configured. The database connection is made successfully with the servers. The SOLAP operations had provided a hierarchical structure to the ASCII format data. The daily radiosonde data has been analysed at different levels for a better understanding of the variations at seasonal, station and regional levels. The figure 5.1 shows the creation of meteorological cube using the MDX editor and XMLA. The MDX editor is used for writing the MDX queries and for performing the drill down and drill up operations. This section lists out the MDX queries and their results used for analyzing and visualizing the trends in inter-seasonal patterns over the Indian region for the two seasons.

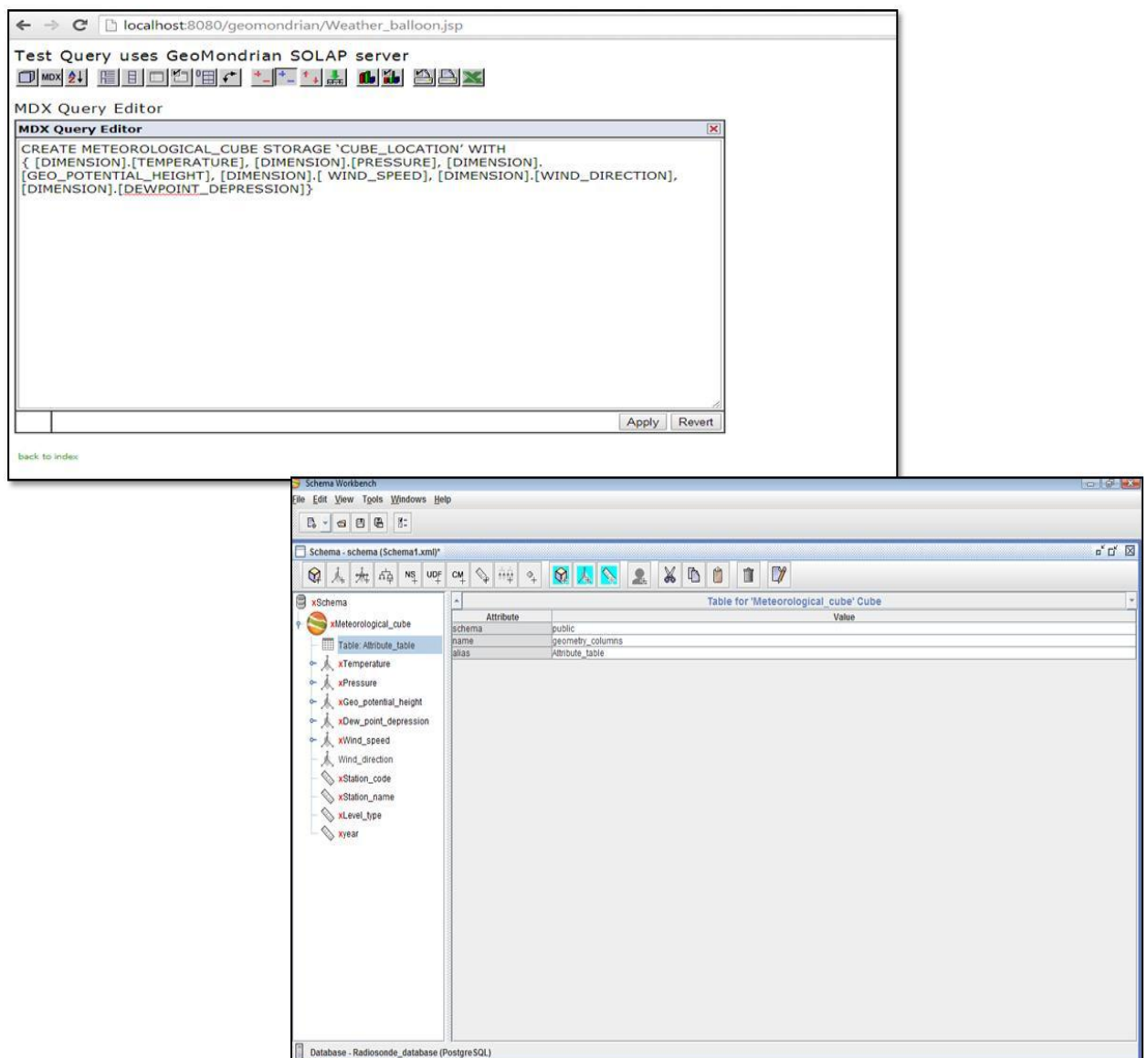


Figure 5.1: Meteorological cube construction using MDX editor and Workbench

The Figure 5.2 shows the MDX query which gives the daily data for pre-monsoon season of all the meteorological parameters for one station for one year based on 850 hPa pressure and 1500m geopotential height.

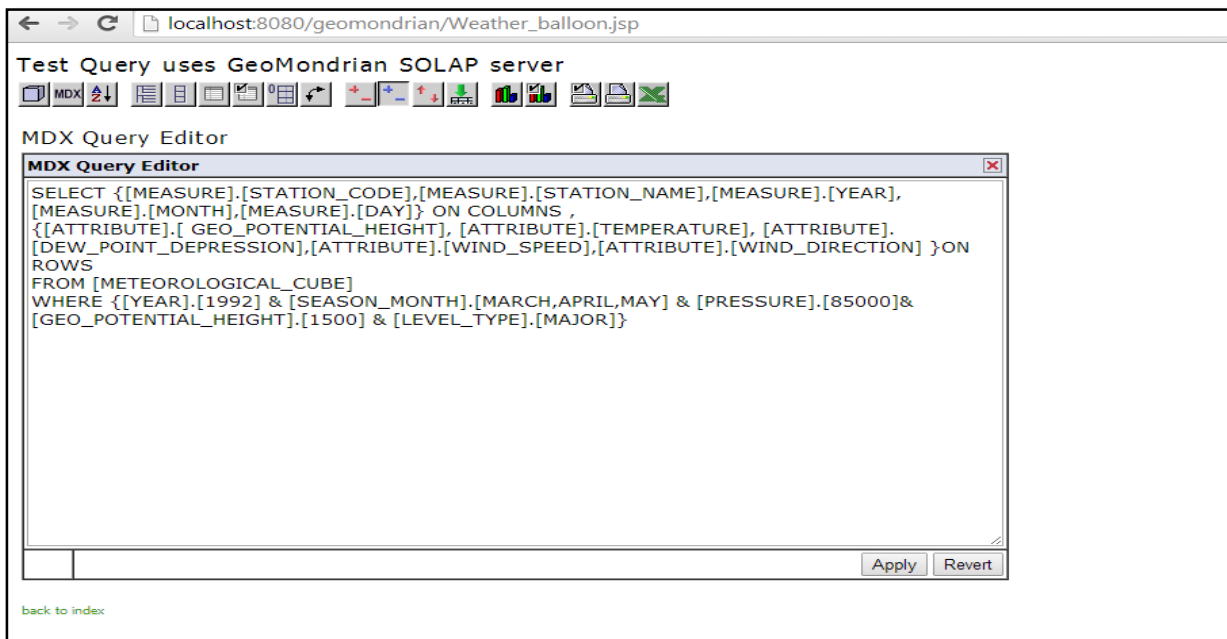


Figure 5.2: MDX query for Pre-Monsoon season daily data at 850 hPa

Similarly, the daily data for pre-monsoon season of all the meteorological parameters for one station for one year based on 700 hPa and 500 hPa pressure and 3000m and 5000m geopotential height is done respectively. This is referred to as drill up operation (i.e. from one station to all the stations).

The figure 5.3 shows the mdx query which gives the daily data for post-monsoon season of all the meteorological parameters for one station for the year 1992, based on 850 hPa pressure and 1500m geopotential height.

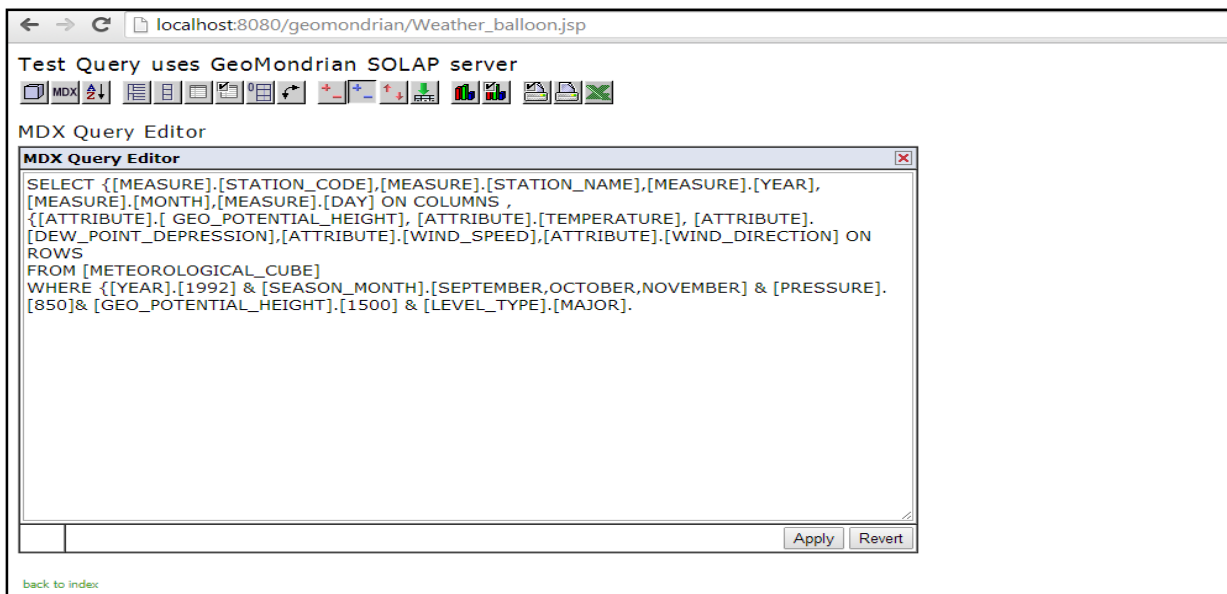


Figure 5.3: MDX query for post-monsoon season daily data at 850 hPa

Likewise, we have the daily continuous data of two decades, for all 27 stations at different significant pressure levels for the two seasons. Next set of queries used for finding the average temperature values for the pre and post monsoon seasons at different pressure levels.

The Figure 5.4 shows the query which gives the average atmospheric temperature values for pre-monsoon season for one the year 1992, based on 850 hPa pressure and 1500m geopotential height.

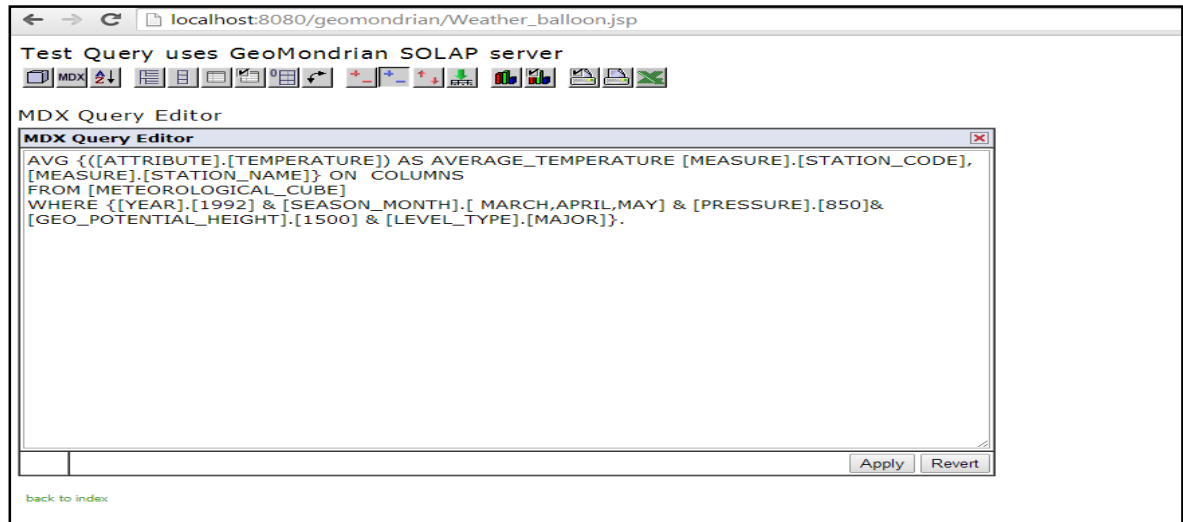


Figure 5.4: MDX query for average temperature for Pre-Monsoon season at 850 hPa

Similarly, we have the average temperature data of two decades, for all 27 stations at different significant pressure levels for the two seasons. By using the average temperature values, temperature variance has been calculated for all the 27 stations for two decades at different significant pressure levels for the two seasons. These variance values are used for analyzing the trend in inter-seasonal patterns of the Indian region. The Figure 5.5 shows the temperature variations in both pre and post monsoon seasons at significant pressure levels with respect to the regions, and the following inferences can be made on the basis of it.

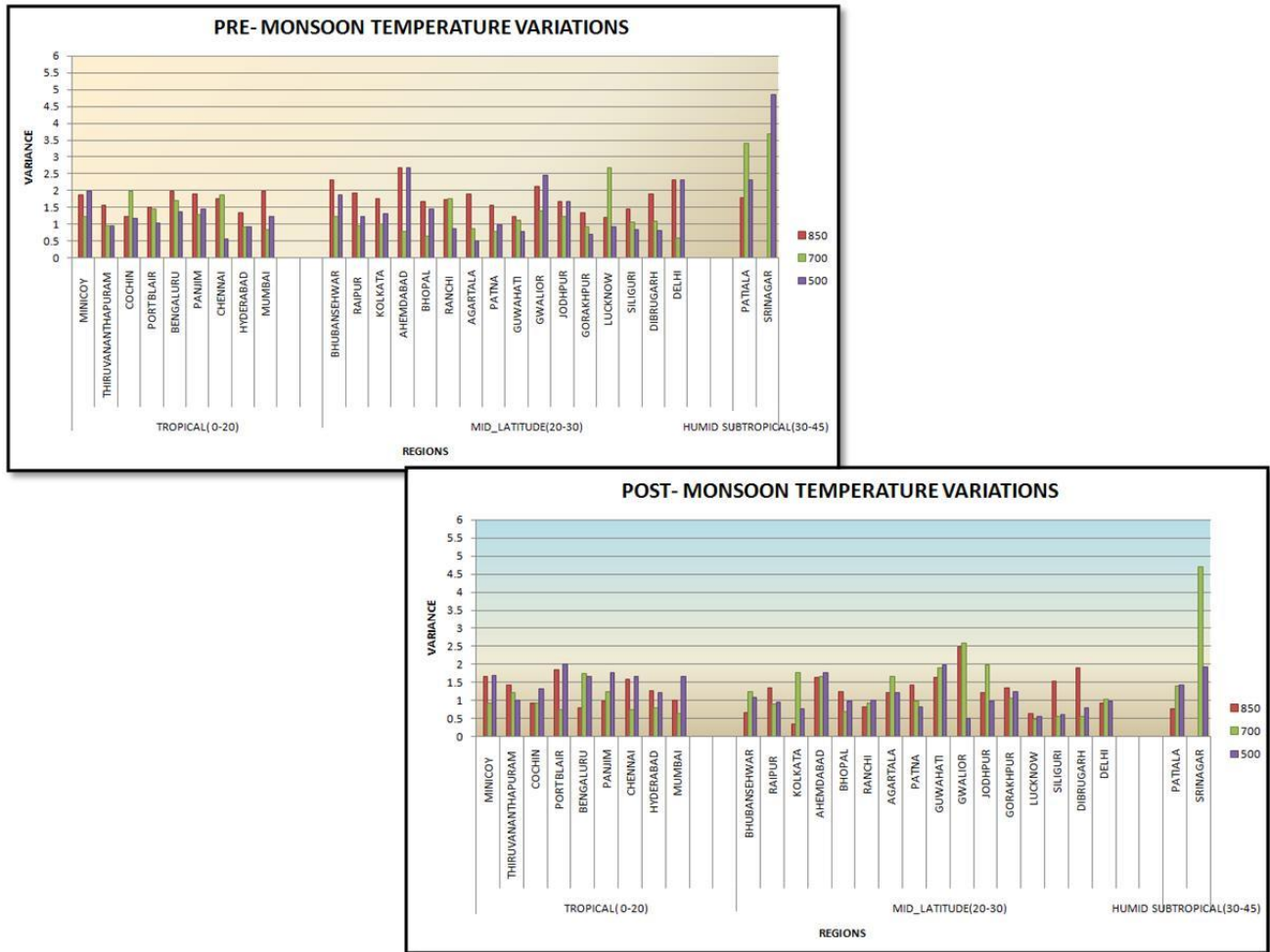


Figure 5.5: Seasonal variations in temperature with respect to region.

Peninsular Region: The peninsular region is more stable at all the pressure levels as compared to the mid-latitude and sub-tropical regions. Variability in the stations is seen clearly in the region. It is less influenced by external variations and is mainly governed by the land and ocean phenomena in both the seasons.

Mid-Latitude: The higher mid-latitude stations are influenced by the western disturbances (air masses coming from the western region) in both the seasons, hence comparative higher variability is seen at lower pressure level for both the seasons.

Humid Sub-Tropical: Highest variability is observed in this region as it is directly influenced by the Himalayan region which controls the atmospheric variability of the region. Maximum variability is seen in the lower pressure levels, i.e. 500 hPa and 700 hPa in both the seasons.

The Figure 5.6 shows the MDX query which gives the significant date and month for all the years for which we have temperature value for one station for the pre-monsoon season at 850 hPa and 1500m geopotential height.

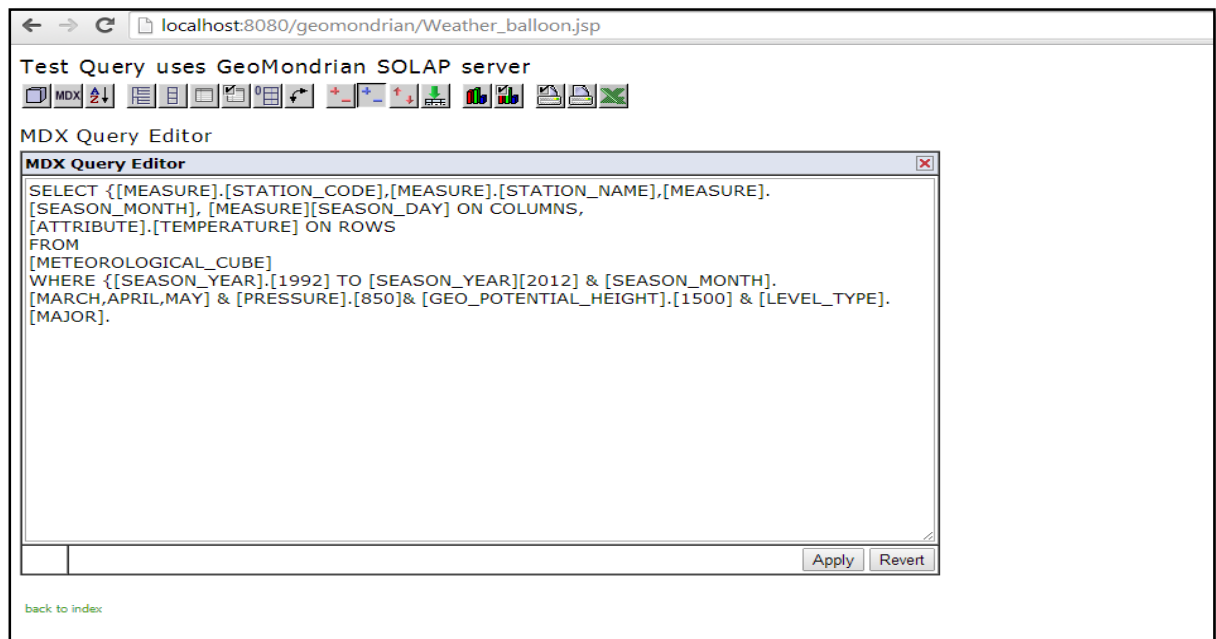


Figure 5.6: MDX query for pre-monsoon

Similarly, we have the significant date and month for all the years for which we have temperature value for all the 27 stations at different significant pressure levels for the two seasons. The result of these queries is used for analyzing the variation in temperature at the station level; this is referred to as drill down operation (i.e. from region to station level).

To analyze inter-seasonal atmospheric temperature variability at defined atmospheric pressure levels, we have considered the relationship amongst stations (i.e. locations) in a particular region. Figure 5.7 and 5.8 shows the graph between annual temperatures and station (i.e. date at which the temperature value was recorded for the particular location) for the both peninsular region and mid-latitude regions at 850 hPa for both the seasons. For the peninsular region the temperature at 850 hPa ranges within 18⁰ to 23⁰ C which depict stability in both the seasons. Wherein the mid-latitude region it is falling in the range of 18⁰ to 24⁰ C during pre-monsoon and 20⁰ to 25⁰ C during post-monsoon respectively.

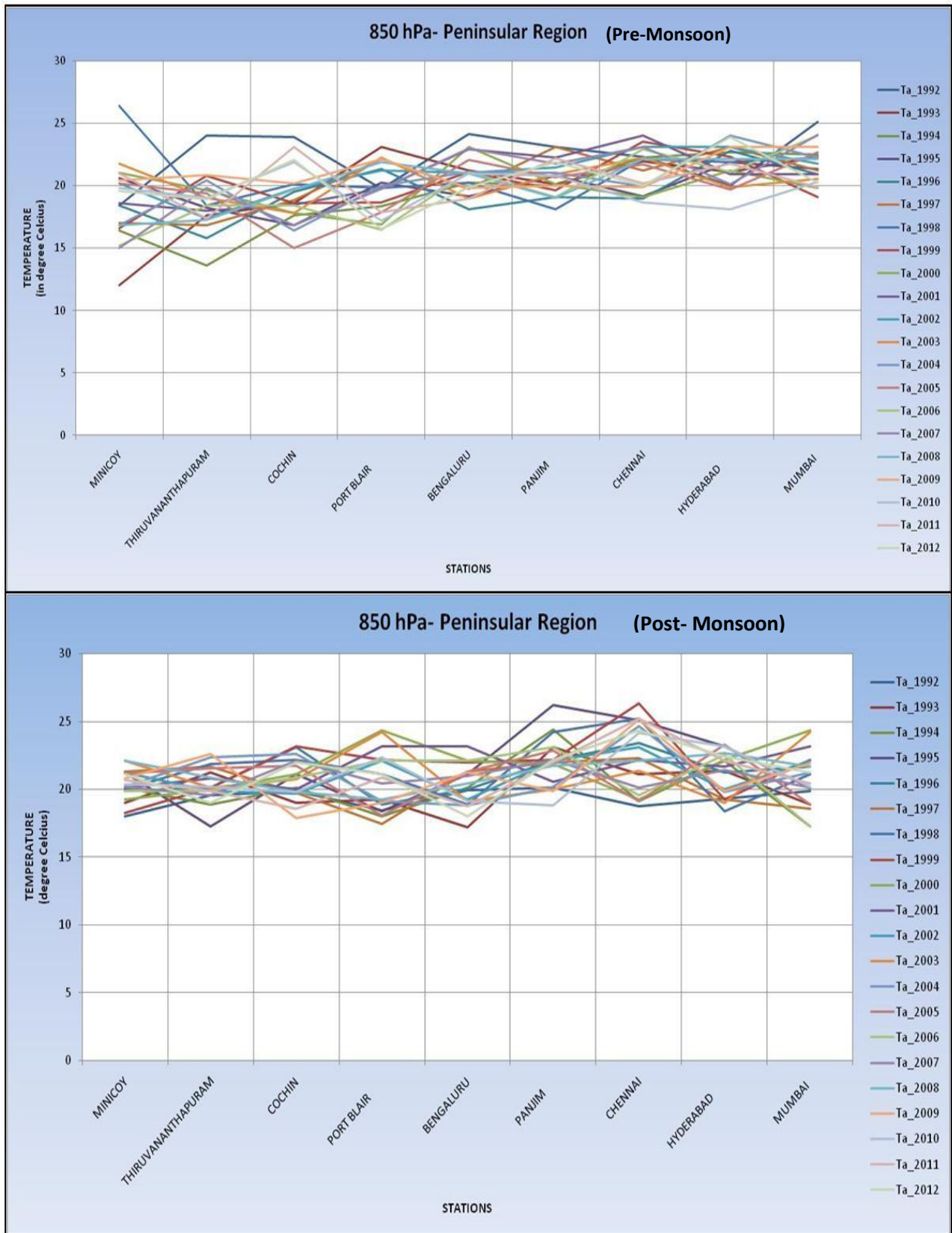


Figure 5.7: Seasonal variations in temperature at 850 hPa in the peninsular region during both the seasons.

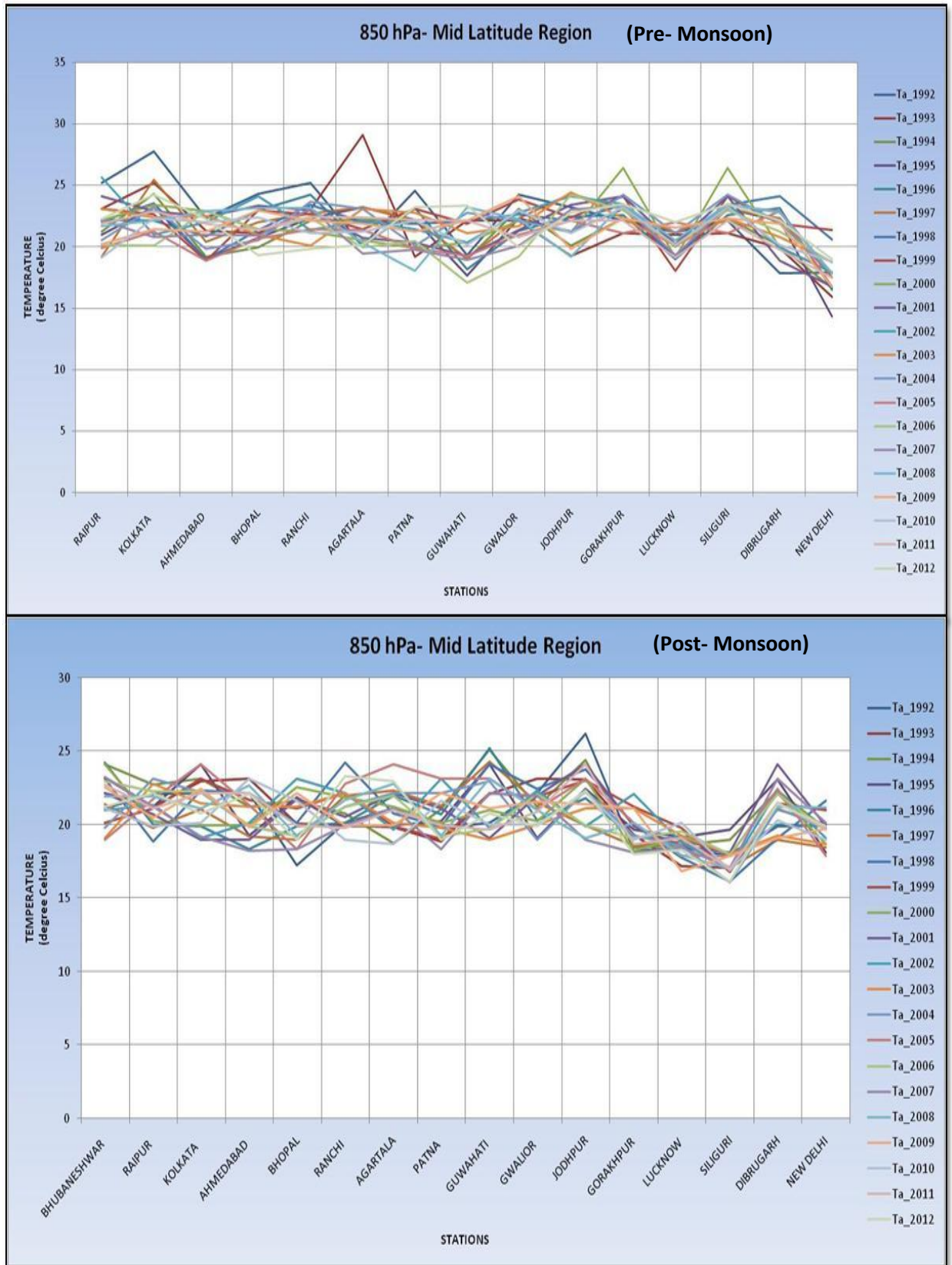


Figure 5.8: Seasonal variations in temperature at 850 hPa in the mid latitude region during both the seasons.

For the peninsular region the temperature at 700 hPa ranges from 8° to 12° C which is stable during both the seasons. Whereas in the mid-latitude region found to be in the range of 7° to 11° C during pre-monsoon and 7° to 14° C during post-monsoon respectively as shown in Figure 5.9 and 5.10

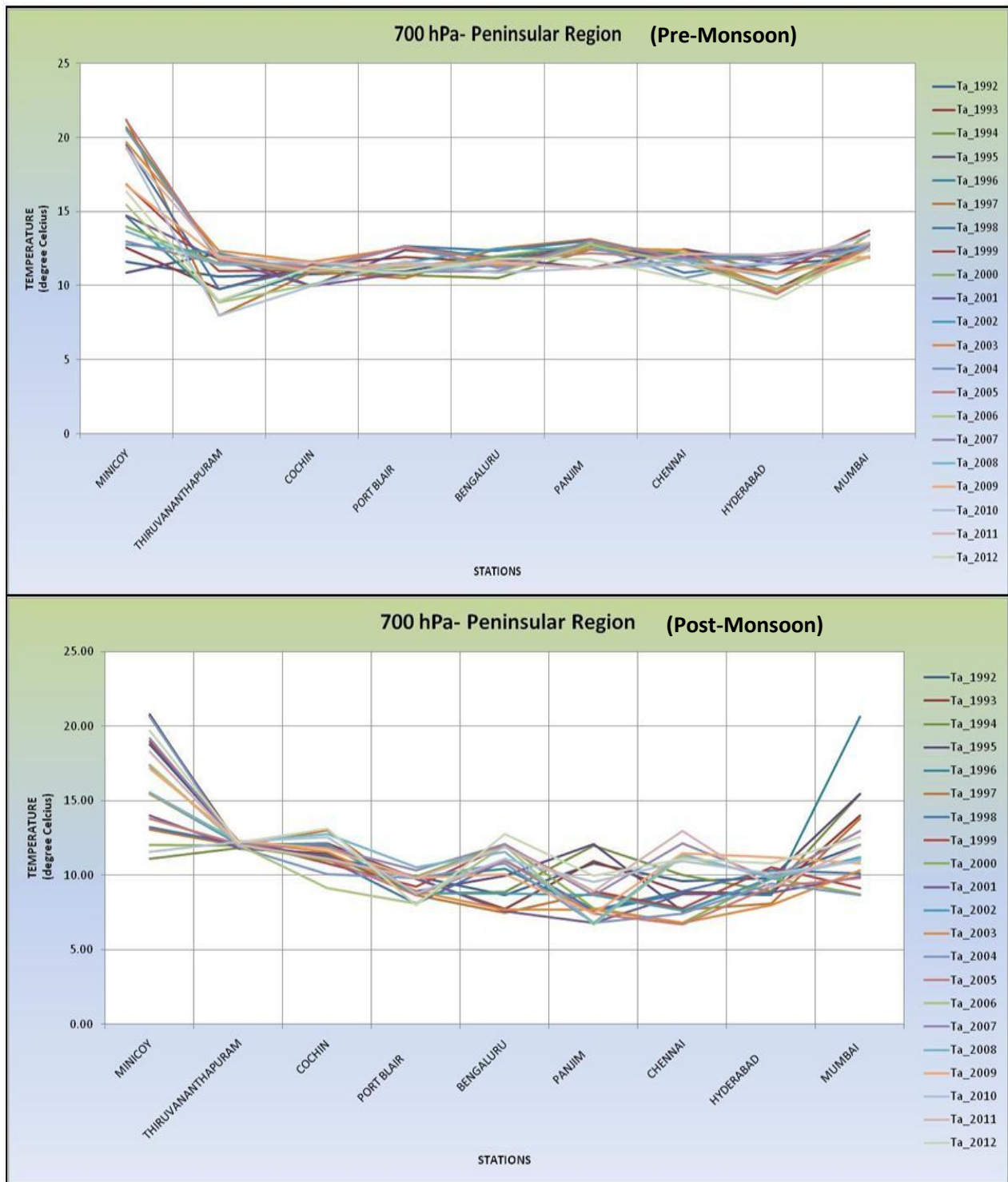


Figure 5.9: Seasonal variations in temperature at 700 hPa in the peninsular region during both the seasons.

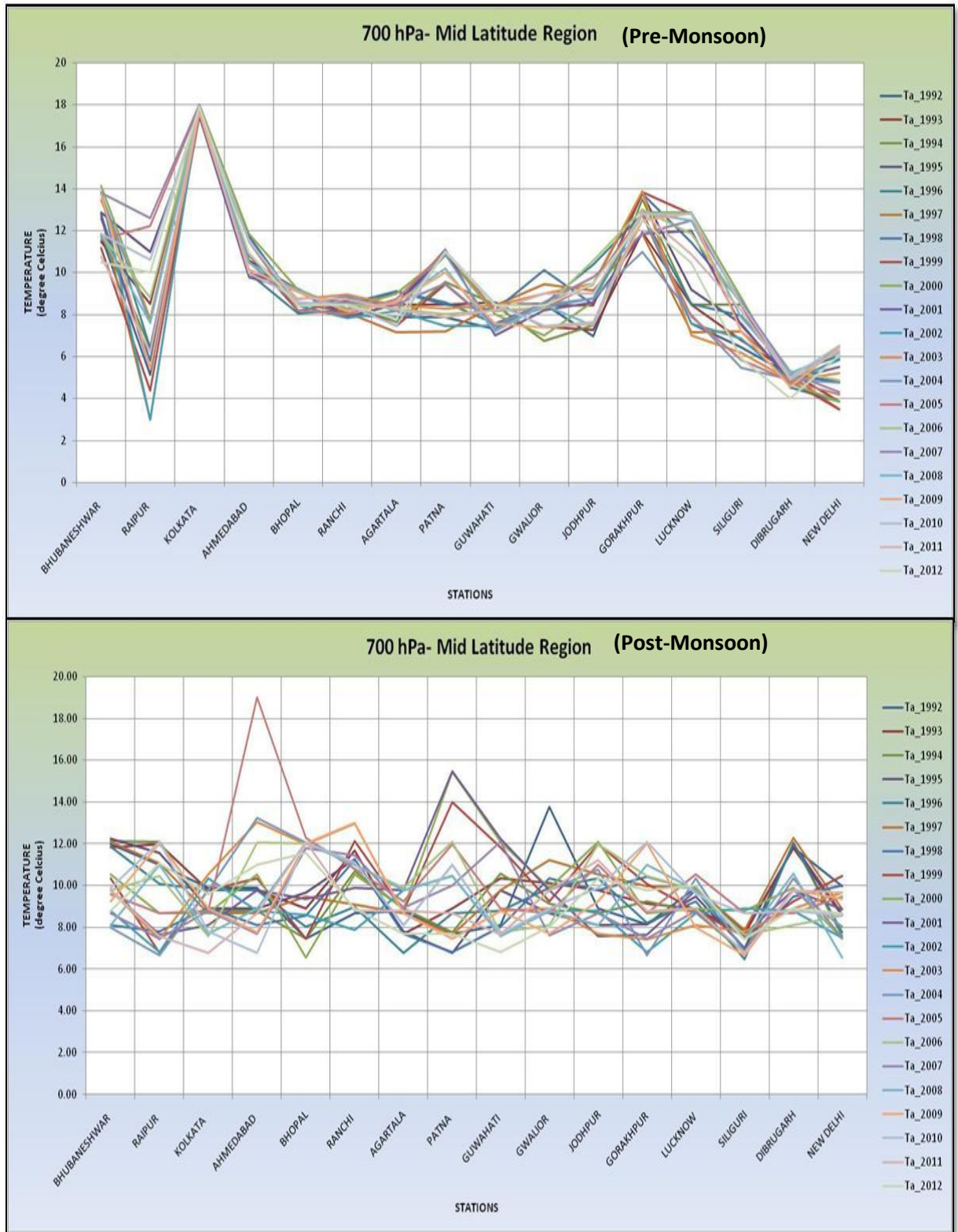


Figure 5.10: Seasonal variations in temperature at 700 hPa in the mid latitude region during both the seasons

For the peninsular region the temperature at 500 hPa ranges within -8° to -4° C during pre- monsoon and in the range of -2° to -7.5° C during post monsoon season, which is highly unstable. In the mid-latitude region it is in the range of -14° to -8° C during pre-monsoon and -9° to -4° C during post-monsoon respectively as shown in Figure 5.11 and 5.12

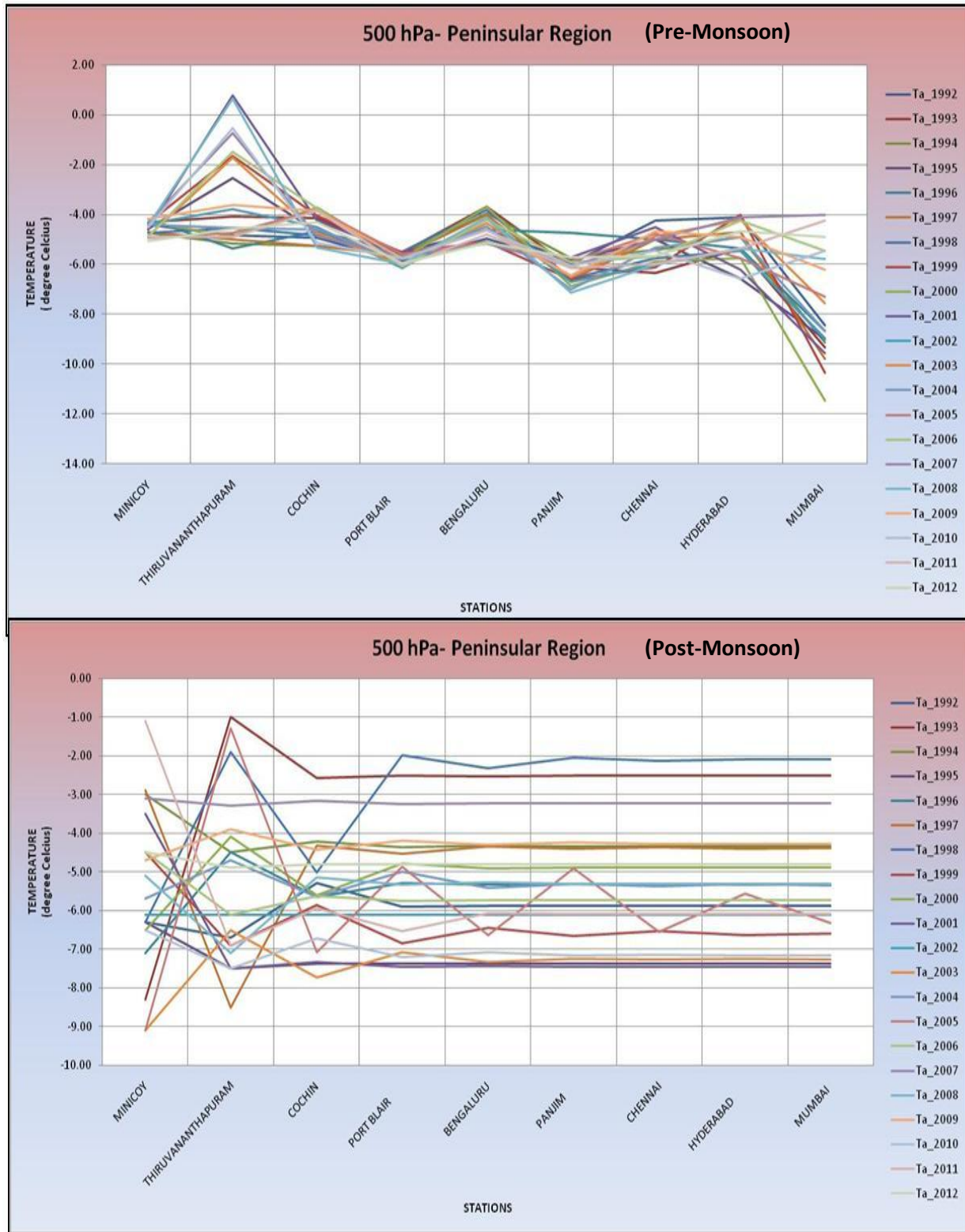


Figure 5.11: Seasonal variations in temperature at 500 hPa in the peninsular region during both the seasons.

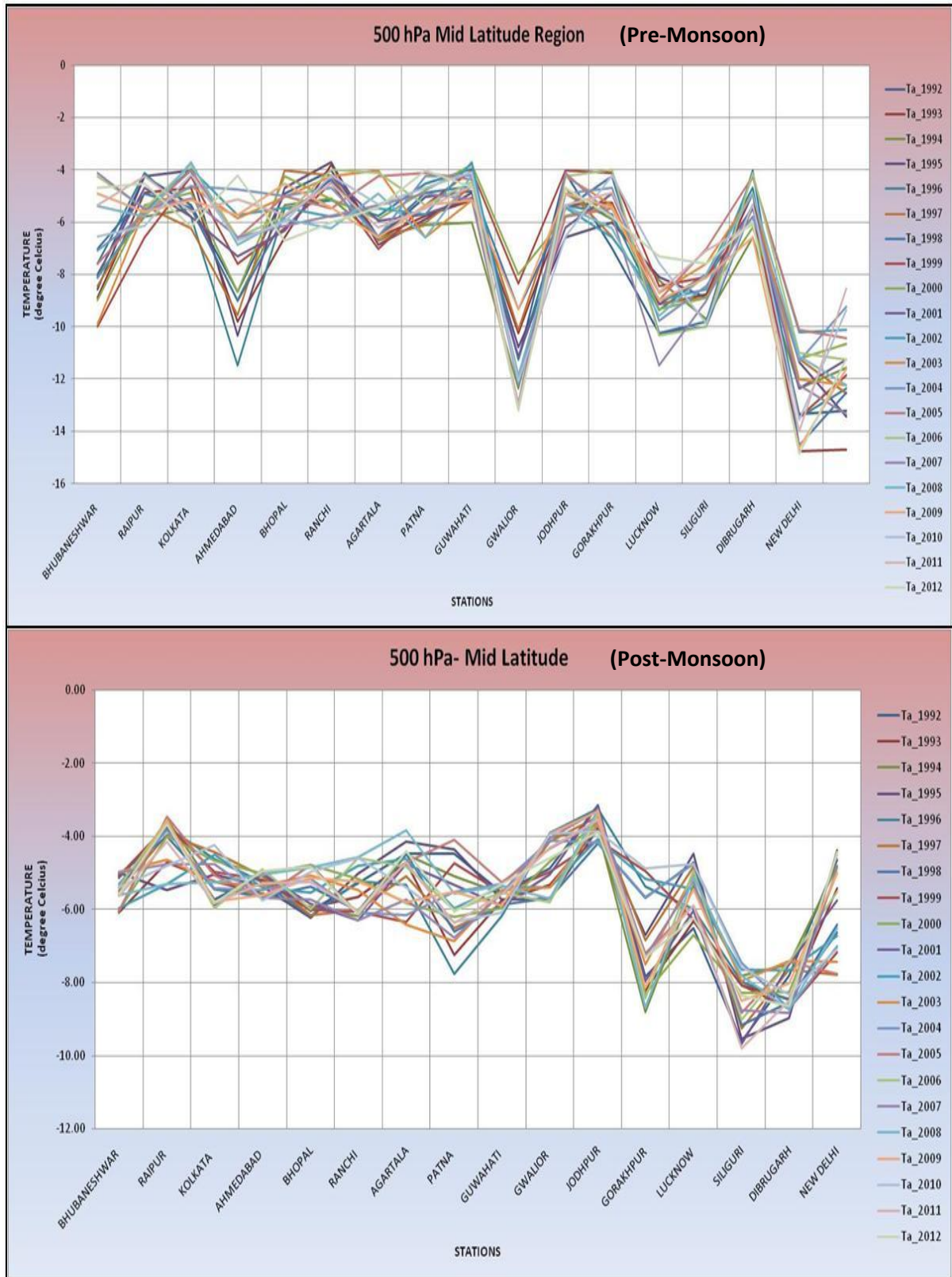


Figure 5.12: Seasonal variations in temperature at 500 hPa in the mid-latitude region during both the seasons

The Figure 5.13 shows the Geo-MDX query which gives the temperature of stations for the pre-monsoon season on the basis of their respective latitude and longitude at 850 hPa.

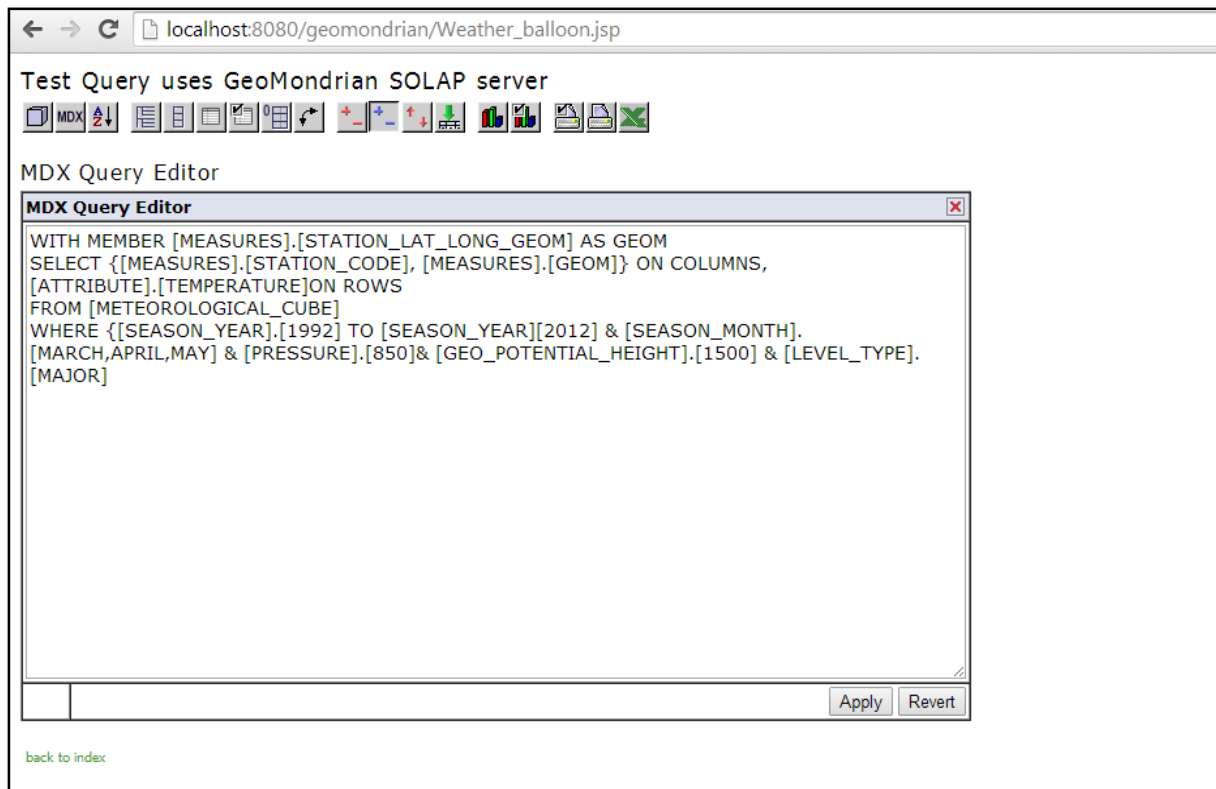


Figure 5.13: Geo-MDX query

Similarly, we have the temperature data of two decades, for all 27 stations at different significant pressure levels for the two seasons.

The above query is used for spatially representing the result on Spatialytics for the visualization and analysis of the trends in inter-seasonal patterns of the atmospheric temperature with respect to the significant pressure levels.

The inter-seasonal spatial variation in temperature over the Indian subcontinent at 850, 700 and 500 hPa for both the seasons are shown in the Figure 5.12

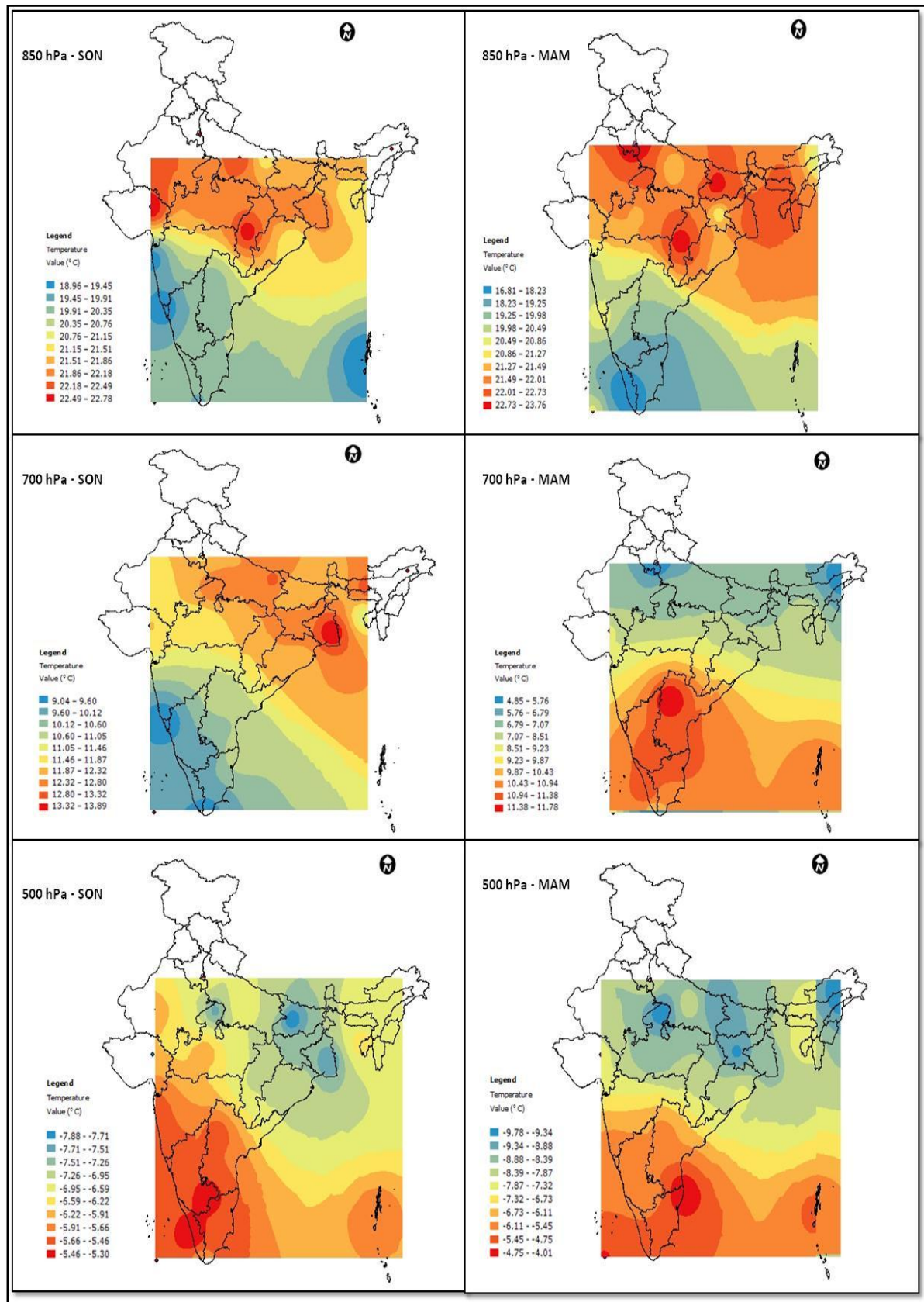


Figure 5.14: Inter-Seasonal Variation in temperature at different pressure levels for the Indian region.

To understand the stability of temperature at different atmospheric layers (850, 700 and 500 hPa), the temperature inversion is analysed. Table 5.1 shows the temperature inversion table.

Table 5.1: Temperature Inversion

Pressure Levels (hPa)	Temperature (° C) (Peninsular Region)		Temperature (° C) (Mid latitude)	
	Pre Monsoon	Post Monsoon	Pre monsoon	Post Monsoon
850	20	20.5	20.5	22.5
700	10	11	9.5	8.5
500	-6	-4.5	-3.0	-6.5
Temperature Inversion (ΔT)				
$\Delta T_{850-700}$	10	9.5	11.0	14.0
$\Delta T_{700-500}$	16	15.5	12.5	15.0

It is seen that between 850 to 700 hPa layers, the temperature inversion is nearly stable during pre-monsoon season over the Peninsular and mid-latitude region while variations are observed over these two regions during post-monsoon. While analysing the 700 to 500 hPa layers, it is evident that variation is stable during post-monsoon season over Peninsular and mid-latitude region, while high variation is observed over these regions during pre-monsoon. It has been established from studies [24] that during pre-monsoon season, the mid-latitude regions are under the influence of west air masses coming from far distance. These air masses mostly effect in the upper layers (700 to 500 hPa) of the atmosphere and hence change the inversion properties of the temperature which are clearly depicted from the results.

6. CONCLUSSION AND RECOMMENDATIONS

6.1. Conclusion

This research has successfully proposed and implemented the functionality to visualize as well as analyze the trends for inter-annual and inter-seasonal patterns of the complex radiosonde data by utilizing the standard features of SDW and SOLAP. These have provided a conceptual and logical representation of the data. Analysis has been performed at different levels of hierarchy (daily to seasonal and station to regional levels) using the multidimensional cube structure. The variation in the inter-seasonal patterns of the atmospheric temperature at different atmospheric pressure levels has been analysed successfully. The dimensionality of the multidimensional data has been analysed using the drill down and drill up operation. The study has successfully depicted that the temperature inversion at the lower levels were stable during the pre-monsoon season over both peninsular and mid-latitude regions. Variations in temperature have been observed during the post-monsoon seasons for both the regions. For the upper layers, stable variations were observed during the post-monsoon season over the peninsular and mid- latitude regions. High variations were observed over these reasons during the pre-monsoon seasons due to the influence of the western disturbances. These analyses have been successfully visualized on the spatial visualization tool in the form of maps and graphs. The development of such functionality can be helpful for the analysts for trend and pattern evaluation of big data.

6.2. Answers of research questions

1) What are the different data pre-processing techniques to be used for the database creation?

The three successful techniques which were used for the creation of the radiosonde database were extract, transform and load (ETL). The ASCII data from the IGRA was extracted using the online extraction technique. The extracted data set was transformed into a conceptual schema using the star schema. And finally the conceptual schema was converted into a physical schema by loading the data into the database using the structured query language (SQL) scripts.

In order to maintain consistency in the data, cleaning of the data was also done by replacing the garbage values (-9999) with "0".

2) Which is the most suitable and server compatible GUI for the visualization of the results?

This research involves the application of two servers, Apache Tomcat and GeoMondrian (SOLAP) server. Both these servers are written in java script. To provide ease of development and the compatibility of the servers, we have selected an existing SOLAP tool (Spatialytics) for the visualization purpose. This visualization tool is also written in java script and supports MDX and XMLA. It is compatible with PostGIS database as well. It has been configured as per the servers and has been used for visualizing the results on maps and graphs.

3) What is the criterion for constructing the multidimensional cube using the SOLAP queries?

The multidimensional cubes are constructed by two ways:

- 1- By the MDX query executed on the GeoMondrian server: The result of the MDX query is published on the SOLAP tool for visualization. The dimensions of the cube include the six meteorological parameters i.e. the values of temperature, pressure, geo potential height, wind speed, wind direction and dew point depression for the years 1992 to 2012. The measure of the cube includes the altitude and the two seasons i.e. Pre- Monsoon (spring) and Post- Monsoon (autumn). Drill down and drill up operations were applied on the multidimensional cubes for the hierarchical level analysis.

2- By the XMLA generated by Workbench Schema: The output of the Workbench is an XML for analysis which contains the hierarchical structure of the cube along with its dimensions, measures and tables used. This hierarchical structure of the multidimensional cube is created by using the drag and drop functionalities of the Workbench using the same criteria as used by the MDX. The resultant XMLA was published on the SOLAP tool for the analysis and visualization.

4) What is the trend in the inter-seasonal patterns in the altitudinal variation of the meteorological variables for the Indian climatological regions over the years?

We could see a reasonable constant and uniform pattern of atmospheric temperature at 850 hPa in both pre and post monsoon seasons for almost all the Indian region, except the humid sub-tropical region. Low to high variability in the atmospheric temperature is observed at 700 hPa and 500 hPa levels for both tropical and mid-latitude regions. The peninsular region is more stable as compared to the other regions as there are no external disturbances in that region. Higher mid- latitude ($>15^\circ$ latitude) are influenced by the western disturbances (air masses coming from the western region) in both the seasons, as a result of which high variability is observed in this region. For the humid sub-tropical region highest variability is observed as it is directly influenced by the Himalayan region.

6.3. Recommendations

Considering the results obtained from the research work, following aspects need more attention for future:

1. Snow flake Schema: Current research involves the star schema for the building the conceptual model of the dataset. The recommendation is to develop a conceptual model using the snow flake schema for analyzing the contrast in the analysis and variations in the results in terms of performance and optimizations. In the snowflake schema, the dimension tables are connected to each other as well. Hence more number of joins are required for the querying the data.

2. Global coverage: Utilizing the capabilities to explore the data of the remaining stations across the globe. It will provide a global coverage and more generalized trends in the inter-seasonal patterns of the radiosonde data.

3. Other Big Data Domains: Apart from meteorological domain, other domains dealing with huge volumes of data where the level of hierarchy is present in the multidimensional data can be used for trend and pattern analysis using this SOLAP based system. This system provides detailed analysis For example; the oceanographic study also involves multidimensional parameters and enormous data sets. So this SOLAP based system can be used for creating multidimensional cubes for the parameters and performing the standard SOLAP operations for the trend analysis.

REFERENCES

- [1] R. Madden, “Seasonal variations of the 40-50 day oscillation in the tropics,” *J. Atmos. Sci.*, 1986.
- [2] S. Dash and J. Hunt, “Variability of climate change in India,” *Curr. Sci.*, vol. Vol. 93, no. No.6, pp. 782–788, 2007.
- [3] S. Rana, Y. Kant, and V. K. Dadhwal, “Diurnal and Seasonal Variation of Spectral Properties of Aerosols over,” vol. 9, no. 1, pp. 32–49, 2009.
- [4] S. Rivest, Y. Bédard, M.-J. Proulx, M. Nadeau, F. Hubert, and J. Pastor, “SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data,” *ISPRS J. Photogramm. Remote Sens.*, vol. 60, no. 1, pp. 17–33, Dec. 2005.
- [5] R. G. Bianchi, G. Y. Hatano, and T. L. L. Siqueira, “On the performance and use of spatial OLAP tools,” in *2013 XXXIX Latin American Computing Conference (CLEI)*, 2013, pp. 1–12.
- [6] S. Bimonte, P. Wehrle, A. Tchounikine, and M. Miquel, “GeWOlap : A Web Based Spatial OLAP Proposal,” pp. 1596–1605, 2006.
- [7] O. Baltzer, “Computational methods for spatial OLAP,” no. April, 2011.
- [8] G. Viswanathan and M. Schneider, “On the requirements for user-centric spatial data warehousing and SOLAP,” *Database Syst. Advanced ...*, vol. 6637, pp. 144–155, 2011.
- [9] L. Gómez, B. Kuijpers, B. Moelans, and A. Vaisman, “A Survey of Spatio-Temporal Data Warehousing,” *Int. J. Data Warehous. Min.*, vol. 5, pp. 28–55, 2009.
- [10] K. K. Chaturvedi, A. Rai, V. K. Dubey, and P. K. Malhotra, “On-line Analytical Processing in Agriculture using Multidimensional Cubes,” vol. 62, no. 1, pp. 56–64, 2008.
- [11] Group Kimball, “Kimball Dimensional Modeling Techniques,” pp. 1–24, 2013.
- [12] L. Leonardi, S. Orlando, A. Raffaetà, A. Roncato, C. Silvestri, G. Andrienko, and N. Andrienko, “A general framework for trajectory data warehousing and visual OLAP,” *Geoinformatica*, vol. 18, no. 2, pp. 273–312, May 2013.
- [13] T. L. L. Siqueira, C. D. D. A. Ciferri, V. C. Times, and R. R. Ciferri, “Modeling vague spatial data warehouses using the VSCube conceptual model,” *Geoinformatica*, vol. 18, no. 2, pp. 313–356, Jan. 2014.
- [14] I. Timko, C. Dyreson, and T. B. Pedersen, *A probabilistic data model and algebra for location-based data warehouses and their implementation*, vol. 18, no. 2. 2013, pp. 357–403.
- [15] M. Golfarelli, M. Mantovani, F. Ravaldi, and S. Rizzi, “Lily: A Geo-Enhanced Library for Location Intelligence,” pp. 72–83, 2013.
- [16] M. Scotch and B. Parmanto, “SOVAT: Spatial OLAP Visualization and Analysis Tool,” vol. 00, no. C, pp. 1–7, 2005.

- [17] F. Rao, L. Zhang, X. L. Yu, Y. Li, and Y. Chen, “Spatial Hierarchy and OLAP-Favored Search in Spatial Data Warehouse,” pp. 48–55, 2003.
- [18] “Geomondrian Documentation centre.”[Online]. Available: <http://www.spatialytics.org/projects/geomondrian/> [Accessed: 10-Dec-2014].”
- [19] M. Kempa, “Multidimensional Expressions (MDX),” *Datenbank-Spektrum*, vol. 11. pp. 123–126, 2011.
- [20] J. Corporation, “OLAP Workbench Guide Table of Contents,” vol. 1, pp. 1–24, 2011.
- [21] “Extensible Markup language for analysis.”[Online].Available: http://en.wikipedia.org/wiki/XML_for_Analysis / [Accessed: 10-Dec-2014].
- [22] “Discover- spatialytics.” [Online]. Available: <http://www.spatialytics.com/apps/discover/> [Accessed: 12-Dec-2014].
- [23] N. Climatic and N. Carolina, “Overview of the Integrated Global Radiosonde Archive,” pp. 53–68, 2006.
- [24] Y. Kant, A. Singh, D. Mitra, D. Singh, P. Srikanth, A. S. Madhusudanacharyulu, and Y. N. V. K. Murthy, “Optical and Radiative Properties of Aerosols over Two Locations in the North-West Part of India during Premonsoon Season,” vol. 2015, 2015.