Data Mining Techniques for Interpretation of Weather Forecast

Rattan K. Datta[1] and Kavita Pabreja[2]
[1]Mohyal Educational and Research Institute of Technology (MERIT), New Delhi
[2]Computer Science Department, Maharaja Surajmal Institute (GGSIPU)
Email: kavita_pabreja@rediffmail.com

ABSTRACT

In this paper we have selected the output products of European Center for Medium-range Weather Forecasting (ECMWF) and Weather Research and Forecasting (WRF) models with reference to occurrences of Cloudbursts and Tornado in India. The NWP model output products have been pre-processed using Multidimensional Data Model which considerably improved the storage and retrieval of required weather variables on a selected time and space scale. The important ingredients to tornado and cloudburst formation have been derived using primary output forecasts provided by the models. Clustering technique of data mining has been applied first on the synoptic scale data of well formed Low Pressure Systems (LPS) which form near head Bay of Bengal and move west, west north-west to north-west and produce rainfall in their wake. Same technique has been applied on sub-grid scale weather processes. Clusters of ensemble (forecast valid for a particular time based on different initial conditions) of derived weather variable viz. convergence at various atmospheric pressure levels for a real life case of tornado and four real life cases of cloudbursts have been generated and analyzed using k-means clustering technique. This approach resulted in locating patterns conducive to formation of cloudbursts four days in advance. The data availability was a limitation but promising advanced signals of formation of patterns have been shown in the study.

Keywords*: Data mining, Clustering, Artificial Neural Networks, Cloudburst, Rainfall, Tornado and Low Pressure System.*

1. Introduction

India is considerably sensitive to weather and climate. But timely weather forecasting can help to considerably minimise the adverse effect. Severe weather phenomena such as tornados, cloudbursts, thunderstorms and floods, annually cause significant loss of life and property, crop destruction and disruption of the transportation systems. Any mitigation of the effects of these weather events would be helpful.

Of the many approaches to weather forecasting, NWP is the most popular method. Datta (1992) illustrates that NWP models attempt to capture the dynamics of various atmospheric variables (like temperature, wind speed, etc.) and how physical processes (like convection, radiation, etc.) affect the future state of these variables. The prediction made hence, is required to be interpreted by the man-machine mix into weather forecasts that are understandable to the general public.

Data mining is a promising new area of research. This area combines methods and tools from the fields of databases, statistics and machine learning. Data mining is a technique for searching and explaining patterns in data, for helping to explain that data and make forecasting from it. Data mining techniques have been applied to many real-life applications and new applications continue to drive research in this area. Since the weather data are generally voluminous, they can be mined for occurrence of particular patterns that distinguish specific weather phenomena.

According to literature Survey, Intelligent systems have been applied generally for processes like decision making applications of evacuation of public in case of cyclone hit, probability of occurrence of rainfall / snow etc. It has been noted that Data Mining (DM) techniques have been used for prediction of snow/ no snow, to determine the relationship between the trajectories of Mesoscale Convective systems moving out of the plateau of Tibet and their environmental physical field values, to predict temperature and pressure. But Data Mining has not necessarily been used to derive actual weather phenomenon from NWP output products. Especially limited work has been done for interpretation of Sub-Grid scale related extreme weather like Tornado and Cloudburst. Keeping in view of these significant considerations, we have discussed six case studies that utilize DM techniques and have demonstrated promising results.

2. Case Study: Analysis of Movement of "Low Pressure Systems" over Indian Region

2.1 Datasets used

Two different datasets corresponding to synoptic scale weather system such as Low Pressure Systems (LPS) were examined in the years given below:

1. Years 1984, 1985, 1988 to 1994

2. Years 1995 to 2003.

The LPS data have been made available with courtesy of (Mooley and Shukla, 1987) and (Sikka, 2006) and this data set is for the monsoon months June to September for each year under consideration.

2.2 Technique Used

An open Source Data Mining tool, WEKA has been used that offers many techniques for mining of datasets. The k-means method of clustering is used to generate clusters corresponding to the day of formation of the system to the day it dies. The number of clusters required has been taken as equal to two. The selection of the LPS dataset has been done so that it contains only the rows corresponding to the first and last day of LPS.

For the other analysis to locate favored zones of LPS movement during months of June and July, here the important feature for clustering is date.

2.3 Findings

For visualization of clusters, the x-axis is chosen as longitude and y-axis as latitude. The area under consideration is from latitude 15.0ºN to 25.0ºN and longitude 66.5 ºE to 90.0 ºE so that focus remains on Indian region. Using data mining technique has resulted in locating a cluster of formation of LPS on 1st day and a cluster of disappearance on 4^{th} or 5th day of formation, on a spatio-temporal scale for the months of June-July and Aug-Sept separately, for two sets of 9years. These clusters corresponding to June-July for years (year 1984, 85, 88, 89, 90, 91, 92, 93, 94) and (year 1995 to 2003) are shown in Figures 1&2. Similarly clusters corresponding to Aug-Sept for years (year 1984, 85, 88, 89, 90, 91, 92, 93, 94) and (year 1995 to 2003) are shown in Figures 3&4.

There is a strong resemblance between the location of clusters of formation and disappearance in Figures 1&2, which are for Jun-July months, for the two sets of years. There is also a strong resemblance between the location of clusters of formation and disappearance in Figures 3&4 which is for Aug-Sept months, for the two sets of years. Hence the favored zones of formation and disappearance of LPS on a spatio-temporal scale, over Indian region during June to September could be identified.

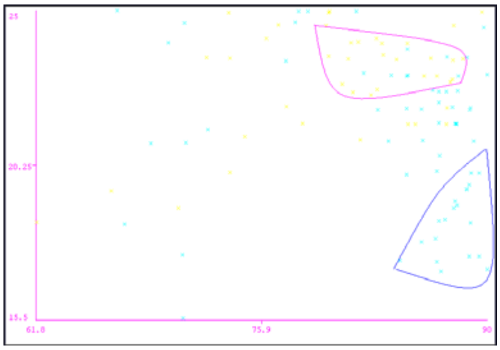

Figure1: Clusters of formation and disappearance of LPS during June-July 1984, 1985, 1988 to 1994. Blue data points represent formation of LPS and Yellow data points represent disappearance of LPS.

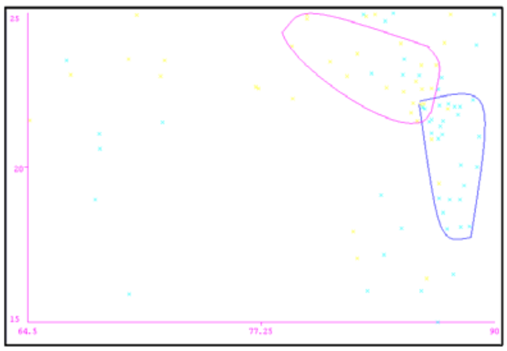

Figure2: Clusters of formation and disappearance of LPS during June-July 1995 to 2003. Blue data points represent formation of LPS and Yellow data points represent disappearance of LPS.

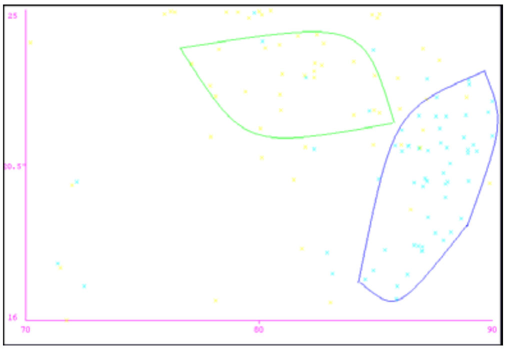

Figure3: Clusters of formation and disappearance of LPS during Aug-Sept 1984, 85, 88 to 94. Blue data points represent formation of LPS and Yellow data points represent disappearance of LPS.

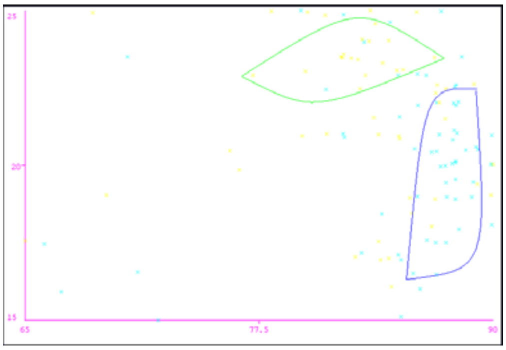

Figure4: Clusters of formation and disappearance of LPS during Aug-Sept 1995 to 2003. Blue data points represent formation of LPS and Yellow data points represent disappearance of LPS.

Clusters of LPS formation, movement, and dissipation during June, July 1984-94 are shown in Figure 5.

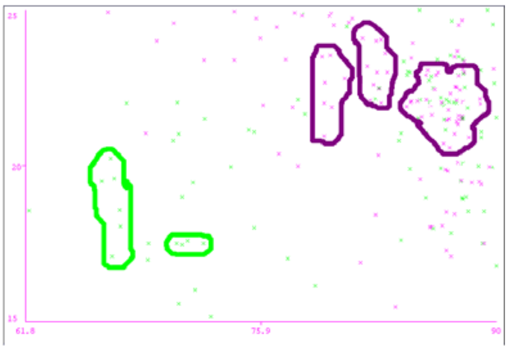

Figure5: Clusters of LPS formation, movement and dissipation during June, July 1984, 85, 88, 89, 90, 91, 92, 93, 94. Green represents LPS during June and Pink represents LPS during July.

3. Case Study: Artificial Neural Network for Rainfall forecasting

3.1 Datasets used

An effort has been made to make use of ANN for forecasting the rainfall based on previous year's rainfall for the months June to September. The daily rainfall dataset taken into consideration for the training of Neural Network is from longitude 70.5 °E to 90.0 °E and latitude 17.5°N to 37.0°N for the time period June to September for the years 1989 to 1992.

3.2 Application of ANN to forecast Rainfall

We have used a two layer MLP Back Propagation network. The input dataset comprises of day number (day 1 corresponds to June 1, day 2 to June 2 and so on till day number 122 that corresponds to September 30), latitude and longitude. The output data corresponds to rainfall in mm. A sample of dataset is shown in Table 1. From this table, columns 1 to 3 are used as input and column 4 is used as target.

Table 1. Sample of location-wise rainfall for year 1989

Day no.	Latitude	Longitude	Rainfall(mm)
1	35.5	76.5	6.4
1	35.5	77	6.5
1	35.5	77.5	1
1	35.5	78	7.3
1	35.5	78.5	2.8
1	35	76	7.7
1	35	76.5	7.4

(*Source*: as a result of pre-processing rf1989.grd provided by IMD)

Before training, the inputs and outputs have been scaled so that they fall in the range [-1,1].

Transfer function used in hidden layer of the back propagation network is tangent-sigmoid while pure linear transfer function is used in output layer.

3.3 Learning Algorithms used

ANN developed for prediction of rainfall is trained with different learning algorithms, learning rates, and number of neurons in its hidden layer.

3.4 Findings

Daily rainfall data for 122 days in a year i.e. months June to September were chosen for training and testing. Networks were trained with data of year 1989 and tested using rainfall data of the year 1990. The training has been done using three different training functions traincgf, trainrp and trainscg and the error comes out to be less than 0.005 in 5 epochs for training functions trainscg and traincgf. With trainrp function, it takes 35 iterations to train.

4. Case Study: A 5-Dimensional Data Model for Meteorological Datasets

In order to understand the impact of an event like cloudburst in a district/state or a river catchment area in a particular time period, we have developed a 5-D multidimensional cube (Pabreja 2010a and 2010b).

4.1 Datasets used

The dataset corresponding to daily rainfall data for each year for the period 1984–2003 has been used. The data is for the geographical region from longitude 66.5 ºE to 100.5 ºE and latitude 6.5 ºN to 38.5 ºN for each day of the year. The LPS datasets as discussed earlier has also been used. The datasets corresponding to river catchment area and district with state has been added.

4.2 Implementation of 5-Dimensional data cube

In order to analyze the rainfall across five different dimensions – Gridded location, Time, LPS formation, River Catchment area and District, a 5 dimensional data cube has been implemented. That is the rainfall during a particular time period of a year, in a particular district or districts, in a particular river catchment area or areas because of movement of LPS system (LPS data from Mooley and Shukla, 1987; Sikka, 2006) can be analyzed. For this purpose, the data has been restructured so as to adapt itself to Multidimensional OLAP environment. Five dimension tables - Location, Time, Low Pressure System, river catchment area and district, have been created. For these two dimensions – River catchment area and district, the tables are normalized each represented by two tables per dimension. Here, the primary keys are riverid for river table and districtid for district table. A 5-Dimensional cube for the data to be examined is shown in Figure 6.

Snowflake schema has been used to model the dimensions datasets and the fact dataset as shown in Figure 7, using Microsoft SQL Server Business Intelligence Development Studio. The schema contains a central fact table for Rainfall that contains foreign keys to each of the five dimensions' primary keys, along with measure for Rainfall. The primary keys are time_key as primary key for Time dimension, loc_key as primary key for Location dimension, (LPSno + Date of LPS) as primary key for Low Pressure System dimension, riverid for river table and districtid for district table.

This 5-D cube comprises of a total of 32 cuboids (2^5) and search of a particular cuboid has been represented in Figure 8 where each node corresponds to a unique cuboid. Statistically the search of a node would save a lot of time in comparison of flat 2-D data storage in Relational databases.

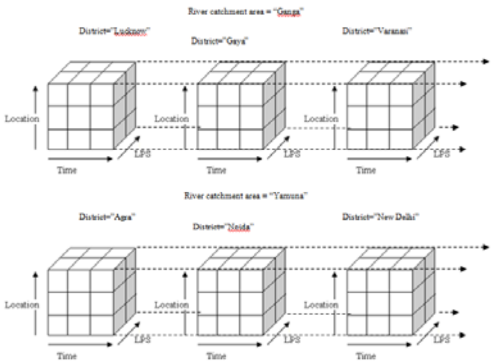

Figure 6: A 5-D data cube representation of rainfall data, according to dimensions time, gridded-location, Low Pressure system, river catchment area and district

4.3 Findings

This MDDM has been browsed to view rainfall value against LPS vs. time vs. location dimensions as shown in Figure 9. SQL server offers facility to apply any filter based on any condition on one/more attributes of dimensions of the cube. The cube can be browsed to view aggregate rainfall at all locations separately, month-wise. The cube has also been browsed to view aggregate rainfall falling in different river catchment areas against various districts, month-wise. These results serve as input to Data mining tools and depict efficient mode of pre-processing of meteorological datasets.

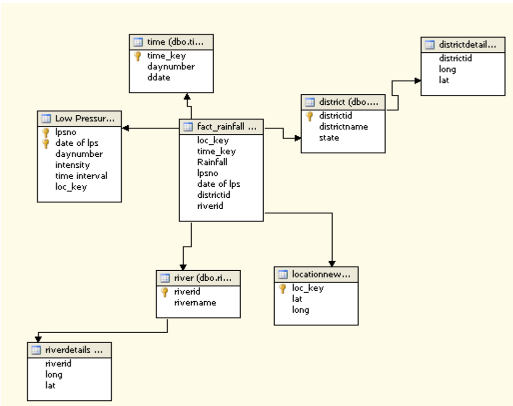

Figure 7: Snowflake Schema for Rainfall corresponding to Tables 4.5, 4.6, 4.7, 4.9, 4.10, 4.11, 4.12, 4.13.

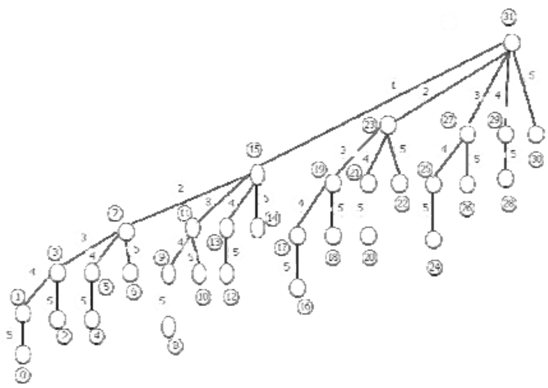

Figure 8: Search of a node corresponding to cuboid in a 5 dimensional data hypercube.

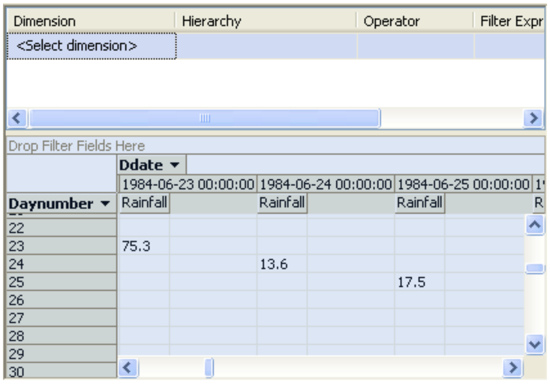

Figure 9: Analysis of Aggregate Rainfall against date and LPS.

6. Case Study: Data mining for Interpretation of sub-grid scale weather system –"Tornado" using NWP output - A Case Study of Tornado in Orissa on 31st March 2009

A tornado accompanied with wind speed of about 250 kmph, thunderstorm, rainfall and hailstorm affected Rajakanika block of Kendrapara district of Orissa in the afternoon of 31st March 2009 (Tornado over Orissa on 31st March 2009 :A preliminary report by IMD).

6.1 Datasets of ECMWF model under analysis

With the courtesy of IMD, ECMWF T-799 model outputs based on weather variables input on 0000GMT 29 March 2009, 0000GMT 30 March 2009 and 0000GMT 31 March 2009 for forecast valid for 0600 GMT 31 March 2009 and 1200GMT 31 March 2009 have been obtained.

The model output has been pre-processed so as to generate vorticity and divergence. For this case of tornado, the location was Rajakanika block of Kendrapara district of Orissa which is at 86° East longitude and 20° North latitude so the convergence and vorticity have been considered for a window of 2.5° around the sides i.e. the area 83.5° longitude to 88.5° longitude and 17.5 ° latitude to 22.5 ° latitude. There are no missing values in the datasets.

We have tried to pick up updraft of air mass that is primary cause of formation of tornado, represented by convergence in the dataset. Hence clustering technique has been applied for feature viz. convergence.

6.2 Visualization and interpretation of clusters of convergence

The derived value of convergence based on weather variables input as on 0000GMT 29 March 2009, 0000GMT 30 March 2009 and 0000GMT 31 March 2009 to ECMWF model for the forecast valid for 0600GMT on 31 March 2009 and 1200GMT on 31 March 2009 have been considered for analysis. The ensemble of convergence for 54hr forecast made based on 0000GMT 29 March 09 weather variables, 30hr forecast made based on 0000GMT 30 March 09 and 6hr forecast based on 0000GMT 31 March 09, valid for 0600GMT 31 March 09 is created. Similarly, the ensemble of convergence for 60hr forecast made based on 0000GMT 29 March 09 weather variables, 36hr forecast made based on 0000GMT 30 March 09 and 12hr forecast based on 0000GMT 31 March 09, valid for 1200GMT 31 March 09 is created.

The clusters of convergence (green and orange points) and divergence (black points) for forecast on 1200GMT 31 March 09 have been plotted as shown in Figure 10. The clusters indicate a very large vertical motion field based on every forecast from 29 March 09 onwards, up to atmospheric pressure level 700hPa. The presence of strong vertical motion field depicts that pattern conducive to formation of tornado is there and a broad interpretation can be made. From the 60hr forecast itself, the indication is provided which starts becoming clearer as we move forward to 36hr and 12hr forecast.

6.3 Datasets of WRF model under analysis

The output products of WRF model, initialized with inputs from NCEP analysis have been utilized. The window under analysis has been selected as 82º longitude to 90º longitude and 14º latitude to 26º latitude.

Now, for data mining, the z-wind component at atmospheric pressure levels of 850hPa, 750hPa, 700hPa and 550hPa has been selected as the convergence at these levels are good indicative of development of tornado.

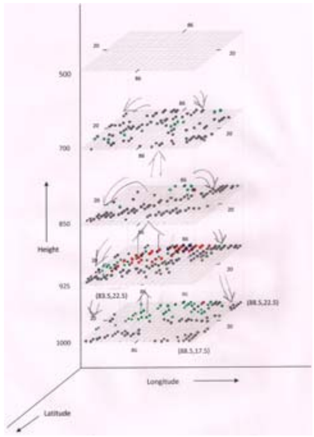

Figure 10: Forecast of convergence valid for 1200 GMT 31 March 09 (Location of tornado: 86°E, 20°N).

(green : -12X10^{-5} per second or -14X10^{-5} per second

red : -16X10^{-5} per second or -18X10^{-5} per second or -20X10^{-5} per second

purple: -22X10^{-5} per second or -24X10^{-5} per second

black: 10X10^{-5} per second or 8X10^{-5} per second or 6X10^{-5} per second or 4X10^{-5} per second)

6.4 Data mining of WRF output field

The two ensembles of vertical wind motion values made for forecasts valid for 0600GMT 31 March 09 and 1200GMT 31 March 09 have been analyzed (Pabreja, 2010c) using clustering technique of data mining. For each ensemble, the k-means method of clustering is used to generate clusters corresponding to the positive and negative values to vertical wind at every atmospheric pressure level under consideration. We have selected z-wind component as the attribute on the basis of which the clustering should be done. There are other research works (Singh, Ganju and Singh, 2005; Stojanova, Panov, and Koblar, 2006) that are based on clustering technique of data mining for prediction of weather events.

6.5 Visualization of clusters of z-wind component

The 3D view of clusters of ensemble of vertical wind component for forecast valid for 0600GMT 31 March 09 and 1200GMT

31March 2009 has been shown in Figure 11 and 12 respectively..

6.6 Interpretation of clusters of z-wind component

The forecast of 0600GMT 31 March09 indicates quite weak vertical motion field whereas the forecast of 1200GMT 31 March 09, which is more near the time of occurrence of tornado (at 1640 hrs IST the tornado hit the ground) shows presence of strong vertical motion field based on all forecasts made i.e. on 0000GMT 28 March 09, 0000GMT 29 March 09, 0000GMT 30 March 09 and 0000GMT 31 March 09. Thus this pattern of convergence is an indicative of early signal of tornado formation

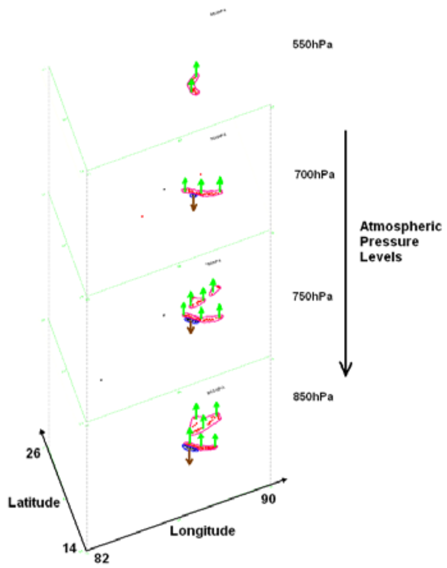

Figure 11: 3-dimensional visualization of forecast of z-wind (arrows represent vertical motion field), valid for 0600GMT 31 March 09 (Location of tornado: 86°E, 20°N).

Within the clusters -

Red points : less than -1m/s

Blue points : more than 1m/s

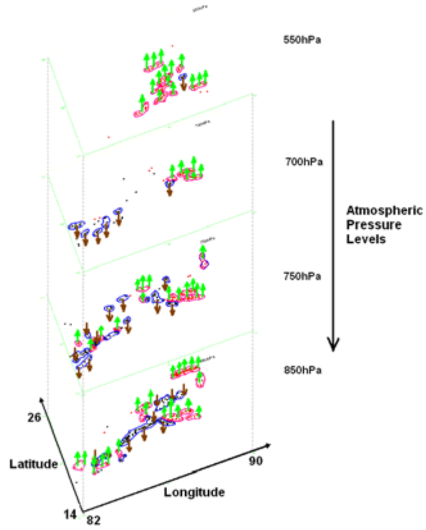

Figure 12: 3-dimensional visualization of forecast of z-wind (arrows represent vertical motion field), valid for 1200GMT 31March 09 (Location of tornado: 86°E, 20° N).

Within the clusters -

Red points : less than -1m/s

Blue points : more than 1m/s

7. Case Study: Data mining for Interpretation of sub-grid scale weather system –"Cloudburst" using NWP (ECMWF) outputs

Cloudburst is a very frequent weather phenomenon that takes place in hilly as well as coastal regions of India. Few of such cases have been identified and the ECMWF model outputs for 84 hour forecast, 60 hour forecast, 36 hour forecast, 12 hour forecast at various surface levels, valid for the following dates of cloudburst in India have been collected from IMD.

1. 29 July 2009, Dhaka, Bangladesh
2. 8 August 2009, Pittorgarh distt. Of Uttarakhand
3. 18 July, 2009, Chamoli distt. Of Uttarakhand
4. 7 August 2009, Shimla, Himachal Pradesh

7.1 Technique applied

The ensemble of eight forecasts, four before time of cloudburst and four after the time of cloudburst, at each of the selected atmospheric pressure levels has been mined. Out of the four case studies, we have shown one here.

Cloudburst case under consideration-Dhaka

Date: 29 July, 2009 (between 1:00am and 7:00am)
Location: Dhaka, Bangladesh (23.5°N and 90.25°E)
Area under consideration: 2.5° X 2.5° window surrounding the location of cloudburst i.e.21.0°N, 87.75°E to 26.0°N, 92.75°E

Forecasts used for creating two ensembles:

- Forecast made on 0000GMT 25 July 09, 0000GMT 26 July 09, 0000GMT 27 July 09 and 0000GMT 28 July 09 valid for 1800GMT 28 July 09.
- Forecast made on 0000GMT 25 July 09, 0000GMT 26 July 09, 0000GMT 27 July 09 and 0000GMT 28 July 09 valid for 0000GMT 29 July 09.

Data mining:

After pre-processing as mentioned above, the clusters using k-means clustering technique have been generated. Corresponding to ensemble created for the forecast made on 0000GMT 25 July 09, 0000GMT 26 July 09, 0000GMT 27 July 09 and 0000GMT 28 July 09 valid for 1800GMT 28 July 09, two clusters of convergence and divergence have been generated by the tool. These clusters have been generated for the different atmospheric pressure levels viz. 300hPa, 400hPa, 500hPa, 700hPa, 850hPa, 925hPa, and 1000hPa. The 3-dimensional visualization of the convergence and divergence at various atmospheric pressure levels for forecast on 1800GMT 28July 09 is also being plotted, in Figure 13.
Corresponding to ensemble created for the forecast made on 0000GMT 25 July 09, 0000GMT 26 July 09, 0000GMT 27 July 09 and 0000GMT 28 July 09, valid for 0000GMT 29 July 09, two clusters of convergence and divergence have been generated by the tool. The 3-dimensional visualization of the convergence and divergence at various atmospheric pressure levels for forecast valid for 0000GMT 29July 09 is also being plotted, in Figure 14.

7.2 Interpretation of visualization of clusters of convergence

There is a very strong vertical motion field up to atmospheric pressure level 700hPa in the clusters of ensemble of forecast, shown in Figure 13 and Figure 14. It is observed that there is an active region of convergence which is an early signal of formation of cloudburst.

Within the clusters – For Figure 14 also

Red points : $-8X10^{-5}$ per sec. or $-10X10^{-5}$ per sec. or $-12X10^{-5}$ per sec.

Green points: $-14X10^{-5}$ per sec. or $-16X10^{-5}$ per sec. or $-18X10^{-5}$ per sec.

Purple points : $-20X10^{-5}$ per sec. or $-22X10^{-5}$ per sec. or $-24X10^{-5}$ per sec. or $-26X10^{-5}$ per sec.

Black points : $8X10^{-5}$ per sec. to $18X10^{-5}$ per sec.

Blue points: $20X10^{-5}$ per sec. to $40X10^{-5}$ per sec.

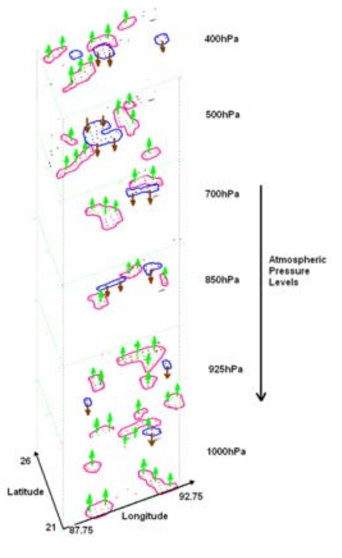

Figure 13: 3-dimensional visualization of forecast of convergence (arrows represent vertical motion field), valid for 1800GMT 28July 09 (Location of cloudburst: 23.5°N, 90.25°E - Dhaka).

For the cloudbursts cases of hilly regions *viz.* Pittorgarh on 8th Aug'09 , Chamoli on 18July'09 and Shimla on 7th Aug'09, it is important to focus on the upper atmosphere levels only as the levels at 1000hPa and 925hPa are virtual areas because of presence of the hills. So, we have to observe the levels

850hPa and above. In these three cases, because of the orography of the area which NWP model cannot integrate well while forecasting, the vertical wind motion field is not so well formed. Although there is indication of presence of vertical motion field, but it is not as clear as in case of coastal region's cloudburst.

8. Conclusions

The output products of NWP model were used to derive the required weather parameters that can be mined so as to generate patterns depicting early indication of sub-grid scale weather events viz. cloudburst and tornado. The meteorological datasets being multidimensional in nature so restructured and stored in Multidimensional Database System. These MDDM facilitated selection of required derived weather variables across a filtered time and space scale as per the time and location of sub-grid scale events.

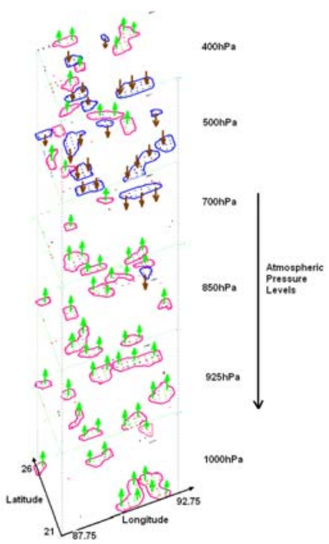

Figure 14: 3-dimensional visualization of forecast of convergence (arrows represent vertical motion field), valid for 0000GMT 29July 09 (Location of cloudburst: 23.5°N, 90.25°E - Dhaka).

The favored zones of formation and disappearance of LPS on a spatio-temporal scale, over Indian region during June to September for years 1984-2003 could be identified. Also the movement of LPS during June to July month during mentioned years could be located along with the favored zones of heavy rainfall. These convincing results of synoptic scale systems provided an encouragement to further look into sub-grid scale weather events viz. tornado and cloudbursts.

The analysis of patterns conducive to formation of sub-grid scale weather systems has been done. An area of 2.5° X 2.5° surrounding the location of the event has been analyzed. The clusters (WRF and ECMWF models for tornado) demonstrated strong vertical wind presence at an area surrounding the location of tornado. This verifies that the WRF model and ECMWF model have a capability of providing an early signal for formation of tornado. Hence it has been observed that the NWP model is able to provide indication of tornado and cloudburst formation 3-4 days in advance, as the forecast made 4days in advance is providing clear indication of strong convergence.

Artificial Neural Networks have been used for the purpose of forecasting Rainfall based on previous year's data. Networks were trained with data of previous year and tested using rainfall data of the following year. It is concluded that ANN has demonstrated promising results and is very suitable for solving the problem of rainfall forecasting but ANN can not be applied for interpretation of formation of sub-grid scale weather phenomenon.

The study demonstrates interesting, useful and fairly clear signals of formation of sub-grid scale weather phenomenon. The forecast of these systems is extremely important for the society. The weather events like tornadoes and cloudbursts are disastrous and threatening to life. This study has clearly brought out that Data Mining techniques when applied rigorously can help in providing advance information for forecast of sub-grid phenomenon.

References

Datta R.K., 1992: Computer as a Window to Deterministic Weather Forecasting. Proceedings of Eighteenth Session of Indian Science Congress Association (ISCA), Calcutta, India, 1992-93, 5-15

Mooley, D. A., Shukla J.,1987: Characteristics of the westward moving summer monsoon low pressure systems over the Indian region and their relationship with the monsoon rainfall, Department of Meteorology, University of Maryland, College Park, MD, USA, Centre for Ocean-Land-Atmosphere interactions, a report, 1987

Pabreja K., 2010a: Mapping of spatio-temporal relational databases onto a multidimensional data hypercube, Proceedings of Einblick – Research Paper Competition held during Confluence 2010 organized by Amity University in association with EMC data storage systems (India) Pvt. Ltd., Noida, UP, India, 2010, Jan. 22-23, 127-133.

Pabreja K., 2010b: Mapping of spatio-temporal weather relational databases onto a multidimensional data hypercube. In: Knowledge Management. (Lathar A.S., Saini A.K., Dhingra S., Eds.), Macmillan Publishers India Ltd., 2010, 253-265

Pabreja K., 2010c: Identification of weather system from Numerical Weather Prediction output data using Data Mining. Proceedings of the 4th International Conference on Advanced Computing & Communication Technologies sponsored by IEEE, Oct 30, 2010, 630-635.

Sikka D.R.,2006: A study on the Monsoon Low Pressure Systems over the Indian Region and their Relationship with Drought and Excess Monsoon Seasonal Rainfall,May 2006.

Singh D., Ganju A., Singh A., 2005: Weather prediction using nearest neighbour model. Current Science, April 2005, vol. 88, no. 8, 25: 1283-1289. Available from: http://www.ias.ac.in/currsci/apr252005/1283.pdf

Stojanova D, Panov P, Koblar A., 2006: Learning to predict Forest Fires with different data mining techniques, Proceedings of the Conference on Data mining and Data Warehousing (SiKDD 2006), Ljubljana, Slovenia, 2006. Available from: http://kt.ijs.si/Dunja/SiKDD2006/Papers/Stojanova.pdf

Weka 3- Data mining with open source machine learning software. (Weka 3, downloaded from Internet) Available from: http://www.cs.waikato.ac.nz/ml/weka/