International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-9, July 2019

Forecasting Future Atmospheric Events using Machine Learning

Kaushik K. Rana, Ketan Sarvakar, Anamika Mittal

Abstract*:** ***Over the year's thunderstorms have been one of the major causes of death and one of the most catastrophic natural calamities in the country and the most challenging part is the prediction of thunderstorm beforehand, because of the random nature of our atmosphere. Through this research paper the attempt was made to do analyze the atmospheric stability indices (atmospheric instability causes thunderstorms) using INSAT-3D sounder data to predict thunderstorms by predicting their values using machine learning approach. By setting the indices possess a threshold value and also there is predictability in the data which can be used to predict their future values. The tephigram is used by meteorologist, scientist, weather observer, pilots to solve atmospheric temperature and humidity problems using simple graphical techniques. We can avoid extensive calculation for the mathematical relationships to generate diagram to predict the events. Meteorologists use the thermodynamic diagram daily to forecast cloud height and atmospheric stability and using the tephigram is use to generate and integrate LIVE events to show easier for the users to view the thermodynamic diagram instantly.

Index Terms: Atmospheric events, Determinism Weather, Machine learning, Predictability, Weather prediction.

I. INTRODUCTION

The ability to predict future atmospheric events, especially severe atmospheric phenomena, is goal shared by vast number of people, from atmospheric scientist to casual weather observers. Countless times throughout the history, severe weather has destroyed or damaged what humans have attempted to derive their existence from, including structures, agricultural products and human lives. The destructive capability of severe weather, especially the loss of human lives, added to the belief that any type of advance warning is desirable[1][6].

Sounder data provides the atmospheric sounding which determines the vertical and horizontal distribution of temperature, humidity, pressure, wind, and other geophysical parameters. This vertical sounding can be of use to meteorologist for weather forecasting and storm predictions and is depicted using computer generated maps. Tephigram is one of a number of thermodynamic diagrams which aids in the interpretation and providing graphical representation of vertical profile of the temperature and humidity in the atmosphere. The principle axes of a tephigram is temperature (T) and potential temperature (θ), hence tephi (Temperature – phi) diagram.[2],[7], [8]

Stability indices are designed to measure the ease with which an air parcel will rise through the atmosphere, using parcel-to-environment temperature difference from a few mandatory levels from atmospheric sounding. It has two purpose to serve:

1. To obtain a number to provide some measure of overall stability of the atmosphere
2. To evaluate the potential for severe weather/predict thunderstorms to occur.

Since atmospheric instability is extremely varying term only one index is not enough to predict convictive activity and hence three indices namely, Lifted Index, K Index, totals total Index are employed for the study [9][8].

For prediction of these indices there is requirement for developing a model which gives more accurate results. Here where machine learning comes to the picture. Machine learning means to make the machine learn from the past data and generate future outcomes. Machine learning that provides computers the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. Machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analysis [10].

II. OUR PROPOSED MODEL AND ARCHITECTURE

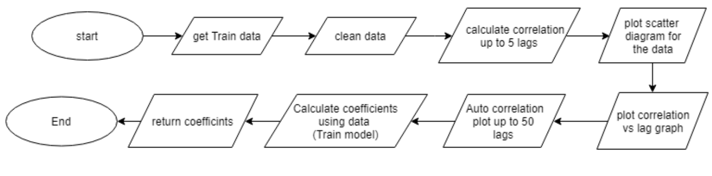


Figure 2.1 Flow Chart To Train The Model Using Test Data

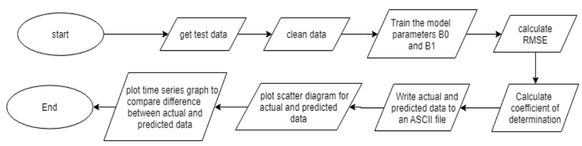


Figure 2.2 Flow Chart To Predict Using Auto-Regressive Model

Revised Manuscript Received on July 09, 2019

Kaushik K. Rana, Assistant Professor in VGEC Chandkheda, Gujarat, India.

Ketan Sarvakar, Assistance Professor, Information Technology department at U. V. Patel College of Engieering, Ganpat University, Kherva, Mehsana, and Gujarat, India.

Anamika Mittal, Assistance Professor, Information Technology department at Government Engineering College, Gandhinagar, and Gujarat, India.

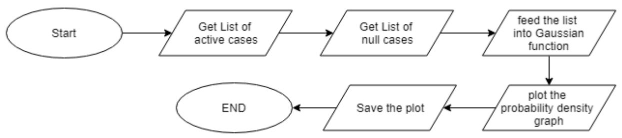


Figure 2.3 Flow chart to plot probability density function

A. Final Stage Methodology for Analyzing and Examine Indices

We take sounding which were taken in the month of April between the years 2016 – 2018 were examined. Also s soundings taken when a severe thunderstorm originated within 100 km of the location of the sounding and previous 6 hours of the time of the sounding were also examined to understand the development of the thunderstorm.

Since we didn't have an In-Situ (on site) data, daily rainfall data (HEM_DLY) from INSAT-3D was studied with a general threshold of 15mm/day to take into account cases of occurrence of thunderstorm, because during pre-monsoon season rainfall is usually accompanied by thunderstorm. Time series graphs of those cases, generated using INSAT-3D's HEM data were studied to get the date, time and peak values to generate a data base of active cases. Under this classification (of active and null cases) 70 cases of thunderstorm were studied. A total of 1134 null sounding remained.

Most of the indices analyzed describe the stability of the atmosphere, as opposed to shear or moisture. They can be determined by mathematical formulae or by plotting on a skew-T log-p diagram. Indices analyzed includes lifted index, K index, total totals. What follows is a list of the indices studied and a description of how each one was calculated. General cutoffs that forecasters use when predicting thunderstorms and their severity are also presented. Lifted Index (LI) is a common measure of atmospheric instability. Its value is obtained by comparing the temperature of parcel with temperature of environment at 500mb K Index Takes into account of the vertical distribution of moisture and temperature and Total totals Index Takes into account of static stability between 850 to 500 mb and low-level moisture at 850 mb. We show indexes along with attributes in Table 2.1, 2.2 and 2.3.

Table 2.1 cutoffs for lifted Index

Lifted Index	Thunderstorm Potential
> +2	No convective activity
0 to +2	Showers probable, isolated thunderstorms possible
-2 to 0	Thunderstorms probable
-4 to –2	Severe thunderstorms possible
< -4	Severe thunderstorms probable, tornados possible

Table 2.2 cutoffs for K Index

K value	Air mass T-Storm Probability
<15	0%
15-20	<20%
21-25	20-40%
26-30	40-60%
31-35	60-80%
36-40	80-90%
>40	>90%

Table 2.3 cutoffs for Total Totals Index

TT	T-Storm Potential
44-45	Isolated to few moderate
	scattered moderate, a few heavy
48-49	scattered moderate, a few heavy and isolated severe
50-51	scattered heavy, a few severe; isolated tornados
52-55	scattered to numerous heavy, few to scattered severe, a few tornados
>55	numerous heavy, scattered showers, scattered tornadoes

III. PREDICTING THUNDERSTORM

A. Autocorrelation of the Data

Autocorrelation (serial correlation), is the correlation of data with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. The autocorrelation function can be used for the following two purposes: To detect non-randomness in data. Here usually only the first (lag 1) autocorrelation is that of interest. To identify appropriate time series model if the data are not random. To identify an appropriate time series model, the autocorrelations are usually plotted for many lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero.

B. Underlying Trends in Data

Once we've know that the data possesses non-randomness, the next step involves finding the underlying trends in data and identifying outliers. For this a scatter plot is plotted in between the data and it's lagged copy. This would help understand the characteristics of data in a linear space. For e.g. whether the data in increasing linearly or is it scattered randomly?

C. Autoregressive Model

Autoregressive models and processes operate under the premises that past values have an effect on current values, which make the statistical technique popular for analyzing nature, economics and other time-varying processes. An AR (1) autoregressive process is the first order process, meaning that the current value is based on the immediately preceding values. TT Index at a particular lat-lon (Ahmedabad) for the month of April 2017 were taken. This would be the training data set. The data would be regressed using OLS (ordinary least square) method. It is a method for estimating the unknown parameters in a linear regression model.

This method chooses its parameters of linear function by minimizing the sum of square of the difference between the observed variables and those predicted by linear function. We can define the linear relation as: y = β0 + β1x.

The equation is similar to $y = mx + c$ where slope (m) = β1 and intercept c = β0. x is the independent dataset and y is the dependent dataset. Calculation of the parameters can be done as followed:

$$\beta_1 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} \qquad \text{……………….2.1}$$

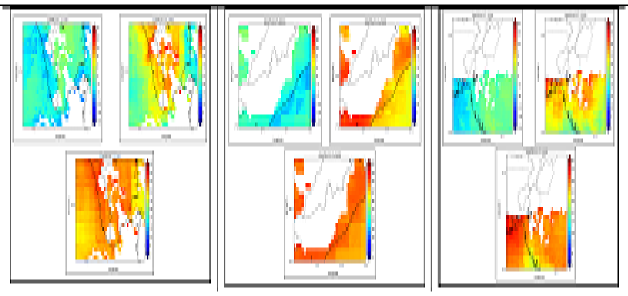
IV. OUR PROPOSED ALGORITHM

Input: latitude and longitude, data path, file name etc. were passed through the URL itself using a feature called url routing provided by flask
Output: the encoded image URL to the client

Read HDF File.
Find the nearest sounding distance and get the indices of the entered latitude and longitude.
Extraction of dataset of required parameters, for plotting for all the 40 pressure levels and store them into an ASCII file.
Plotting and Generating:
Removing of all the fill values (filtering the dataset).
Converting humidity into dewpoint temperature.
Calculating various convective parameters.
Encode the image, return it and delete it from the buffer.
Returns the encoded image URL to the client.
The required inputs such as latitude and longitude, data path, file name etc. were passed through the URL itself using a feature called url routing provided by flask.

V. EXPERIMENTS

Before moving over to the result, a case study was taken into consideration to understand the spatial distribution of various indices over Indian region to do the experimentation. For this, 3 different regions where thunderstorm was reported were selected namely: Thiruvananthapuram, Visakhapatnam and Mizoram. Plots for previous 6 hours were also studied to understand the changes in the index values over time. The study provides the ground, to understand the distribution of stability indices over various regions of India. Also, through the plots we can check for the quality of the data.



Thiruvananthapuram	Visakhapatnam	Mizoram

VI. RESULTS

Once the average rainfall data was plotted it was examined to get heavy rainfall at a particular latitude-longitude. Time series graphs were plotted using HEM data, for that particular latitude-longitude for the whole month to get the date, time and peak value accounting it for the case of thunderstorm.

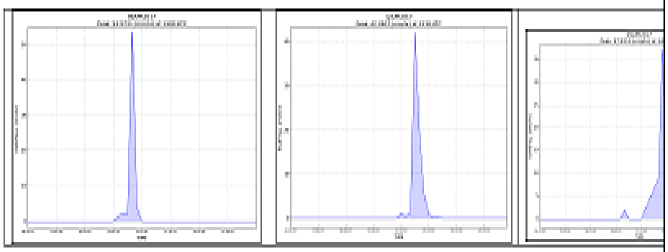


Figure 5. 1 Rainfall (HEM) Time Series Plots

Probability distribution curves (PDF) of active and null cases were plotted for a particular Index to check whether the index has any threshold and if it has then what that threshold is. Below is the resultant PDF plots for all the three indices: Lifted, K and Total totals.

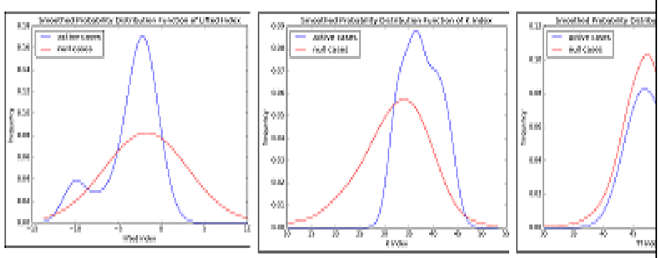


Figure 5.2 (a) Probability distribution curve for LI	(b) Probability distribution curve for KI	(c) Probability distribution curve for TT

VII. CONCLUSION

Over the year's thunderstorms have been one of the major causes of death and one of the most catastrophic natural calamities in the country and the most challenging part is the prediction of thunderstorm beforehand, because of the random nature of our atmosphere. Through this study the attempt was made to do analyze the atmospheric stability indices (atmospheric instability causes thunderstorms) using INSAT-3D sounder data to predict thunderstorms by predicting their values. Results shows that the indices possess a threshold value and also there is predictability in the data which can be used to predict their future values.

REFERENCES

1. M. Li, Q. Zhang, and F. Zhang, "Hail Day Frequency Trends and Associated Atmospheric Circulation Patterns over China during 1960-2012", vol. 29. 2016.
2. M. Acreman,"Extreme Rainfall in Calderdale," 19 May 1989, vol. 44. 1989.
3. A. KIEFFER WEISSE and P. H. BOIS, "A comparison of methods for mapping statistical characteristics of heavy rainfall in the French Alps: the use of daily information / Comparaison de méthodes de cartographie de paramètres statistiques des précipitations extrêmes dans les Alpes françaises:" vol. 47. 2002.
4. J. Thielen and A. Gadian, "Influence of topography and urban heat island effects on the outbreak of convective storms under unstable meteorological conditions: A numerical study", vol. 4. 2006.
5. D. S. Wilks, "Chapter 7 - Statistical Forecasting," in Statistical Methods in the Atmospheric Sciences, vol. 100, D. S. B. T.-I. G. Wilks, Ed. Academic Press, 2011, pp. 215–300.
6. Z. Meng and F. Zhang, "DATA ASSIMILATION AND PREDICTABILITY | Ensemble-Based Data Assimilation," G. R. North, J. Pyle, and F. B. T.-E. of A. S. (Second E. Zhang, Eds. Oxford: Academic Press, 2015, pp. 241–247.
7. S. Siegert and D. B. Stephenson, "Chapter 15 - Forecast Recalibration and Multimodel Combination," A. W. Robertson and F. B. T.-S.-S. to S. P. Vitart, Eds. Elsevier, 2019, pp. 321–336.
8. Z. Toth and R. Buizza, "Chapter 2 - Weather Forecasting: What Sets the Forecast Skill Horizon?" A. W. Robertson and F. B. T.-S.-S. to S. P. Vitart, Eds. Elsevier, 2019, pp. 17–45.
9. "Uncertain Forecasts From Deterministic Dynamics," Stat. Postprocessing Ensemble Forecast., pp. 1–13, Jan. 2018.
10. D. S. Wilks and S. Vannitsem, "Chapter 1 - Uncertain Forecasts From Deterministic Dynamics," S. Vannitsem, D. S. Wilks, and J. W. B. T.-S. P. of E. F. Messner, Eds. Elsevier, 2018, pp. 1–13.

AUTHORS PROFILE

Dr. Kaushik K. Rana completed M.E in Computer Engineering from Dharmsinh Desai University, Nadiad, India, Dr. Kauhsik K. Rana achieved Ph.D. in Computer Engineering from Gujarat Technological University in the year 2017. The author's major field of study covers Service Oriented Architecture, Software Testing and Cloud Computing. He has been working as an Assistant Professor in VGEC Chandkheda, Gujarat, India.

Prof. Ketan Sarvakar is working as an Assistance Professor, Information Technology department at U. V. Patel College of Engieering, Ganpat University, Kherva, Mehsana, and Gujarat, India. He has completed his Mtech CSE in Engineering and Technology from the Nirma University. His primary areas of interest are Machine Learning, Data Mining, Cloud Computing, Mobile application Development and Network Security.

Prof. Anamika Mittal is working as an Assistance Professor, Information Technology department at Government Engineering College, Gandhinagar, and Gujarat, India. He has completed her ME IT from the. L.D. College of Engineering, Ahmedabad. His primary areas of interest are Machine Learning, Software Engineering, Data Structure and Cryptography