International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Attribute Selection for Earth's Climate Prediction

R. K. Tiwari[1], Shailendra Singh[2]

[1,2] Govt. New Science College, Rewa, Madhya Pradesh (India)

Abstract : ***The strong association between monthly average rainfall and monthly mean temperature over the 30-year period from 1976 to 2006 for Vindhya Region is analyzed. This study shows climatological characteristics and fluctuation of climate with rainfall and found temperature as a most significant attribute. Both local and field significances have been tested by using CFS subset evaluator. Greedy algorithm applies on the normalized data sets of selected three solar cycles. From the analysis we found that the merit of best subset are 0.394 (SC-21), 0.274 (SC-22) and 0.348 (SC-23) for selected solar cycles. We have observed that for all the solar cycles temperature is the most significant attribute among the selected solar activity parameters.***

Keywords: Greedy algorithm, temperature, CFS evaluator, wind velocity

1. Introduction

Climate scientists are actively involved in weather forecasting and exploring possible links between solar activity and earth's climate. Response of the atmospheric dynamics to solar variability has been the subject of considerable investigation over many years. Variations in the total solar irradiance (TSI) incident on Earth's atmosphere control the average weather and the global temperature change. The temperature of Earth is controlled by the balance between the energy received from the sun and earth's thermal emission to space. The potential perturbation of climate due to solar variability involves two mechanisms. One of them is associated with disturbances produced in the upper atmosphere and resulting from ozone variations generated by changes in short wave solar radiation. The other mechanism is linked to the ocean-surface response to 11-year changes in the total solar irradiance. The effect of solar variability on the Earth's climate has been investigated by many researchers [1,2,3]. They have developed a good understanding of the mechanism to interpret the earth's climate system, and how the different parts of the climate system interact with one another. This understanding is helpful in developing model for weather forecasting.

It is known that the sun has influenced temperatures on different time scales. The radiative forcing considerations along with the results of general circulation model (GCM) and energy balance models suggest that there are other processes in the middle atmosphere that may amplify solar impact.

Indirect climate effect of the sun through modulation of the mean circulation structure has been studied by Casper Amman [4]. He suggested that when Earth's radiative energy balance is altered due to change in the solar cycle forcing, although the global mean temperature change may be small, regional signature in moisture, pressure and temperature offer a consistent role.

2. Literature Survey

The data mining techniques are often used to locate and identify events in climate datasets. The long range forecasting of weather is very complex due to its chaotic behavior. Karmakar et.al.[5] and Guhathakurta [6], have studied the application of artificial neural network in long range weather forecasting. Rajeevan et.al [7] have developed new models for long range forecast of summer monsoon rainfall over North West and Peninsular India. BPN model for range forecast of monsoon rainfall over a very small geographical region has been developed by Gyanesh et.al. [8]. they have tested the performance of the model and found excellent.

A large number of features are computed for each time series in a training set and the most informative class structure are taken using greedy forward feature selection with a linear classifier. The resulting feature-based classifiers find the differences between classes using a reduced number of time-series properties. This helps to calculate distances between time series. The features selected provide an understanding of the properties of the dataset, which helps for further scientific investigation [9].

In this study the feature selection process is applied in given datasets. Features selection evaluates a subset of features as group for suitability. Correlation feature selection evaluates subsets of features on the basis of the given data sets. Good feature subsets contain highly correlated features with classification, which are not correlated to each other [10, 11, 12]. Evaluation of the subsets requires a scoring metric that grades a subset of features. The general greedy selection algorithm attempts to find the globally best subset of selected attributes by making local greedy decisions for changing the current subset. Greedy stepwise forward algorithm follows the problem solving methodology of making the locally optimal choice at each stage in search of finding a global optimum [12].

3. Data Collection and Methodology

The data are collected from NOAA, climate prediction center (www.nws.noaa.gov) and IMD Pune.The pre-processing methods applied on the selected period 1976-2006. Pre-processed data are divided into three solar cycles – C-21 : 1976-1986, C-22 : 1986-1996 and C-23 : 1996-2006.

In the present study, we have applied best attribute finder method, which helps to find most significant climate attribute that affect the rainfall. CFS subset evaluator is used to find best attribute. Correlation feature selection method is applied to get these parameters.

Table 1(a): Normalized values of solar cycle-21 (1976-1986)

Year	Nv	Nap	Nssn	Nt	Nr
1976	0.8544	0.5536	0.0836	0.9139	1.0000
1977	0.7898	0.4840	0.1786	0.9437	0.8606
1978	0.8063	0.7014	0.5952	0.8940	0.8938
1979	0.7882	0.6216	1.0000	0.9338	0.9389
1980	0.7415	0.4684	0.9945	0.9040	0.6852
1981	0.8827	0.9553	0.9037	0.8940	0.9165
1982	0.9057	0.8052	0.7378	0.9205	0.8746
1983	0.9165	0.8222	0.4134	0.9934	0.7469
1984	0.8846	0.5950	0.2749	0.9073	0.9617
1985	0.8703	0.5507	0.0936	0.8940	0.8712
1986	0.8091	0.4738	0.0672	0.9669	0.8110

Table 1(b): Normalized values of solar cycle-22 (1986-1996)

Year	Nv	Nap	Nssn	Nt	Nr
1986	0.8091	0.4738	0.0672	0.9669	0.8110
1987	0.8151	0.5724	0.1540	0.8940	0.7387
1988	0.8356	0.6825	0.5588	0.9007	0.7008
1989	0.8789	1.0000	0.9591	0.8775	0.9922
1990	0.8220	0.7037	0.8714	0.9106	0.8458
1991	0.8664	0.6020	0.9237	0.9338	0.9073
1992	0.9838	0.7877	0.6043	0.8808	0.8260
1993	0.8123	0.5367	0.3458	0.9338	0.7973
1994	0.8006	0.3950	0.2040	0.9603	0.8582
1995	0.7221	0.3560	0.1140	1.0000	0.9670
1996	0.7781	0.5088	0.0527	0.9470	0.8651

Table 1(c): Normalized values of solar cycle-23 (1996-2006)

Year	Nv	Nap	Nssn	Nt	Nr
1996	0.7781	0.5088	0.0527	0.9470	0.8651
1997	0.8322	0.5306	0.1313	0.9669	0.8709
1998	0.8493	0.6357	0.4012	0.9371	0.8682
1999	0.8328	0.5923	0.6193	0.9669	0.9202
2000	0.8547	0.6421	0.7901	0.8411	0.8401
2001	0.7729	0.5000	0.5043	0.8748	0.7861
2002	0.7482	0.3947	0.4725	0.9030	0.6832
2003	0.7729	1.0000	0.2894	0.8785	0.8791
2004	1.0000	0.2632	0.1836	0.8927	0.7707
2005	0.7940	0.4737	0.1354	0.8957	0.8683
2006	0.8275	0.1316	0.0691	0.8778	0.8859

Greedy stepwise algorithm is used to find the best subset. This algorithm executes attribute subset evaluator to get significant attribute of climate which is most effective to the rainfall of selected solar cycles.

4. Result and Discussion

Greedy algorithm applies on the normalized data sets of selected three solar cycles 21, 22 and 23. CFS subset evaluator is used to evaluate significant attributes among the selected parameters i.e. SSN, Ap, V, Temp. For solar cycle 21 merit of best subset is 0.394 and selected attributes are 1 and 4 i.e. wind velocity (v) and temperature (T). Similarly for solar cycle 22 merit of best subset is 0.274 and temperature (T) is found as a significant attribute. For the solar cycle 23 similar results were observed. Most significant attribute is temperature found at 0.348 best merit evaluations as shown in table-2.From the figure-1 it is observed that the temperature has a strong association to the rainfall for all the solar cycles under study.

Table 2: Result of CFS subset evaluator based on Greedy stepwise algorithm

S No.	Period	Method	Merit of best subset	No. of selected attribute	Name of attribute
1.	1976-86 (C-21)	CFS Subset Evaluator	0.394	02	V, Temp (1,4)
2.	1986-96 (C-22)	CFS Subset Evaluator	0.274	01	Temp. (4)
3.	1996-2006 (C-23)	CFS Subset Evaluator	0.348	02	Temp. (4)

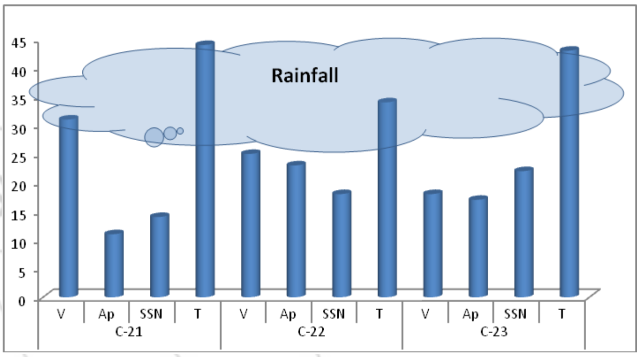

Figure 1: Impact of solar attributes in different solar cycles (SC-21, SC-22 and SC-23)

We found that the temperature is evaluated as most effective attribute among the selected parameters, which significantly affect the rainfall of the selected periods. This study interpret that the mean global temperature is closely linked to the radiative balance of the Earth and its climate. The work will be extended to evaluate the regional climate from the observation by the solar radiation and climate experiment (SORCE). Similar study will be carried out for the current solar cycle 24. The other parameters will be tested to find the closest link between solar activity and climate change on long term basis.

5. Acknowledgement

The authors are thankful to MPCST Bhopal for providing financial support under minor research project

References

[1] Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Avery, M. Tignor, and H.L. Miller (eds.). (2007): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA.

[2] J.D. Haigh "The effects of Solar variability on the Earth's climate", Phil. Trans. R. Soc. Lond. A(2003), 95-111.

[3] Marsh D.R. and Garcia R.R. "Attribution of decadal variability in lower stratospheric tropical ozon" GeoPhysical Research letters, 34: L2 1807(2007).

[4] Casper Amman workshop report "The effect of solar variability on Earth's climate" National Academic Press 2012.
[5] Karmakar S, Kower M.K. and Guhathakurta P. "Artificial neural network skeleton in deterministic forecast to recognize pattern of TMRF, Ambikapur district, CSVTV Res. J. 2(2), 41-45.
[6] Guhathakurta P. "Long range monsoon prediction of 2005 for the districts and sub division Kerala with artificial neural network", Current science 90(6):773-779 · March 2006 .
[7] Rajveen M., Guhathakurta P. and Thapliyal V. " New models for long range forecasts of summer monsoon rainfall over North West and Peninsular India" Metro ref. Atomos. Phys., 73, 211-225.
[8] Shrivastava G., Karmakar S., Kower M.K. and Guhathakurta P. "BPN model for long range forecast of monsoon rainfall over a very small geographical region and it's verification for 2012", GEOFIZIKA, Vol. 30, 49-66 (2013).
[9] Daniel Levy, "Spatio Temporal Pattern Detection in Climate Data", University of Toronto, dlevy@cs.toronto.edu, 2013.
[10] Dash M., Liu H., "Feature Selection for Classification", Intelligent Data Analysis, Vol. 1, no. 3, 1997 Elsevier Science B.V. 1, 131–156.
[11] Hall M., "Correlation –based feature selection for machine learning", Ph.D. Thesis, University of Waikato, NewZealand 1999.
[12] Baris S., "Fast correlation based Filter (FCBF) with a different search strategy." Computer and information sciences, ISCIS'08, IEEE, 2008, 4.
[13] Guyon, I and Andre E., "An introduction to variable and feature selection", JMLR, 2003, 3.