Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 7 [1] December : 59-65

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.876
Universal Impact Factor 0.9804
NAAS Rating 4.95

ORIGINAL ARTICLE **OPEN ACCESS**

Effect of Climatic Variables on Different Stages of Wheat with the help of powerful Statistical tool CART

Subhash Kumar*, S. P. Singh and Mahesh Kumar
Department of SMCA, Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar - 848125
Email*:* *subhashkr97@gmail.com

ABSTRACT
The impact of climate change is studied in many aspects in different locations in the country and it is concluded that there is high impact on agriculture compared to any other sector in the country. Many studies have been conducted to illustrate the changes in annual temperature, relative humidity, evaporation and rainfall are becoming evident on a global scale.This study examines the effect of climatic factor e.g. Temperature (Maximum and Minimum), Relative humidity (Morning and Evening), Evaporation and Rainfall variation on the yield of different stages of wheat in Samastipur district of Bihar by using CART(Classification And Regression Tree) statistical methods. The data of wheat yield of 29 Years (1984-2013) was taken from Department of Agricultural Economics, RAU, Pusa and Weather Variables (1984-2013) was taken from Agro-metrology Unit, RAU, Pusa. The time series information of yield and seasonal meteorological data (e.g., Temperature (Maximum and Minimum), Relative humidity (Morning and Evening), Evaporation and Rainfall will used CART will used to estimate the impact of climate variables on the yield. With the help of regression and classification and regression trees (CART) we can also find out the importance of different climatic variables at different stages of wheat growth well identified. CART analysis allowed to: (i) unravel interactions and combined effects in a complex dataset; (ii) identify thresholds in the relationship between wheat yield and different weather variables. The approach provided insight into the structure of interrelationships within the dataset more easily as compared to multiple regression modeling.
Key Words: Climate change, Wheat yield, CART, Effect of weather on crop yield.

Received 26.08.2017 Revised 29.09.2017 Accepted 21.11. 2017

INTRODUCTION

CART stands for Classification and Regression Tree. CART analysis is a tree-building technique which is different from traditional data analysis methods. In addition, CART is often able to uncover complex interactions between predictors which may be difficult or impossible using traditional multivariate techniques. It is now possible to perform a CART analysis with a simple understanding of each of the multiple steps involved in its procedure. Classification tree methods such as CART are convenient way to produce a prediction rule from a set of observations described in terms of a vector of features and a response value. Tree based classification and regression procedure have greatly increased in popularity during the recent years. Tree based decision methods are statistical systems that mine data to predict or classify future observations based on a set of decision rules and are sometimes called rule induction methods because the reasoning process behind them is clearly evident when browsing the trees.

MATERIAL AND METHODS

This study deals with methods, procedures and measurement techniques followed for carrying out the research work entitled "statistical analysis of wheat yield and climatic change in Samastipur district of Bihar". The present study has been carried out to focus on the overall impact of climate change on the wheat yield.

The methodology includes:

1. Locale of the study
2. Data and variables
3. Various statistical tools

General description of the study area

The study was carried out in Samastipur district of Bihar in India. This is situated in Agro- climatic zone I (Northern West). The traditional agricultural practice is prevalent in this district. Then latitude and longitude is 25° 51′47.48″ N and 85° 46′48.04 0″ E respectively. It is situated at an elevation of about 52 m above mean sea level. The climate of the site is characterized by hot and humid summers and cold winters with an average rainfall of 1200 mm, 70 percent (941 mm) of which occurs during July - September and average temperature is maximum 36.6°C and minimum temperature is 7.7°C. Frequent droughts and floods are common in the region.

Data and variables

Wheat productivity data is collected from Dept. of Agricultural Economics, RAU, Pusa, Samastipur, Bihar. We take data of wheat productivity and climatic variable from 1984-2013. We consider the average amount of wheat productivity in tonnes/hectare. The direct impact of climatic variables on wheat yield. The data regarding the climatic variables is collected data source from the Agro-meteorology Unit, RAU, Pusa, Samastipur, Bihar.

Following are the climatic factor and their units which are taken in this research:

1. Maximum temperature (°C)
2. Minimum Temperature (°C)
3. Relative Humidity (morning) (%)
4. Relative Humidity (evening) (%)
5. Rainfall (mm)
6. Evaporation (mm/m^2)

Now we can also analysis the wheat production after the effect of climate change on each stages of wheat. Generally we take eight stages of wheat. Sowing time of wheat is mid of November i.e. 15^{th}Nov-20^{th}Nov and harvesting time is start from first week of April.We use eight stage i.e.Seedling emergence (20^{th} Nov – 26^{th} Nov), Tillering (27^{th} Nov- 15^{th} Dec), Node stage (16^{th} Dec- 5^{th} Jan), Boot stage (6^{th} Jan-30^{th} Jan),Ear head emergence (31^{st} Jan – 20^{th} Feb), Milk stage (21^{st}Feb – 5^{th} March), Dough stage (6^{th} March – 15^{th} March) and Maturity stage (16^{th} March- 31^{st} march).

Here we discussed each stage. In this discussion climatic condition of each stage are mentioned. So first of all we are discuss the first stage i.e. Seedling emergence.

Classification & Regression Tree (CART) Approach:

The CART methodology have found favour among researchers for application in several areas such as agriculture, medicine, forestry, natural resources management etc. as alternatives to the conventional approaches such as discriminate function method, multiple linear regression, logistic regression etc. In CART, the observations are successively separated into two subsets based on associated variables significantly related to the response variable; this approach has an advantage of providing easily comprehensible decision strategies. CART can be applied either as a classification tree or as a regressive tree depending on whether the response variable is categorical or continuous. Tree based methods are not based on any stringent assumptions. These methods can handle large number of variables, are resistant to outliers, non-parametric, more versatile, can handle categorical variables, though computationally more intensive. CART can be a good choice for the analysts as they give fairly accurate results quickly, than traditional methods. If more conventional methods are called for, trees can still be helpful if there are a lot of variables, as they can be used to identify important variables and interactions. These are also invariant to the monotonic transformations of the explanatory variables and do not require the selection of the variable in advance as in regression analysis.

CART methodology

The CART methodology developed by Breiman *et al*. [1] is outlined here. For building decision trees, CART uses so-called learning set - a set of historical data with pre-assigned classes for all observations. An algorithm known as recursive partitioning is the key to the nonparametric statistical method of CART. It is a step-by-step process by which a decision Tree is constructed by either splitting or not splitting each node on the tree into two daughter nodes. An attractive feature of the CART methodology is that because the algorithm asks a sequence of hierarchical questions, it is relatively simple to understand and interpret the results. The unique starting point of a classification tree is called a root node and consists of the entire learning set L at the top of the tree. A node is a subset of the set of variables, and it can be terminal or non-terminal node. A non-terminal (or parent) node is a node that splits into two daughter nodes (binary split). Such a binary split is determined by a condition on the value of a single variable, where the condition is either satisfied or not satisfied by the observed value of that variable. All observations in L that have reached a particular (parent) node and satisfy the condition for that variable drop down to one of the two daughter nodes; the remaining observations at that (parent) node that do not satisfy the condition drop down to the other daughter node. A node that does not split is called a terminal node and

is assigned a class label. Each observation in L falls into one of the terminal nodes. When an observation of unknown class is "dropped down" the tree and ends up at a terminal node, it is assigned the class corresponding to the class label attached to that node. There may be more than one terminal node with the same class label.

To produce a tree-structured model using recursive binary partitioning, CART determines the best split of the learning set L to start with and thereafter the best splits of its subsets on the basis of various issues such as identifying which variable should be used to create the split, and determining the precise rule for the split, determining when a node of the tree is a terminal one, and assigning a predicted class to each terminal node. The assignment of predicted classes to the terminal nodes is relatively simple, as is determining how to make the splits, whereas determining the right-sized tree is not so straightforward. In order to explain these in details, procedure of growing a fully expanded tree and obtaining a tree of optimum size is explained subsequently.

Tree growing procedure

Let (Y, X) be a multivariate random variable where X is the vector of K explanatory variables (both categorical and continuous) and Y is the response variable taking values either in a set of classes C (=1,...,j,...,J) or in real line.

Splitting strategy

In determining how to divide subsets of L to create two daughter nodes from a parent node, the general rule is to make, with respect to the response variable, the data corresponding to each of the daughter nodes "purer" in the sense that the data in each of the daughter nodes is obtained by reducing the number of cases that has been misclassified. For a description of splitting rules, a distinction between continuous and categorical variables is required.

Continuous or numerical variable

For a continuous variable, the number of possible splits at a given node is one less than the number of its distinctly observed values.

Nominal or categorical variable

Suppose that a particular categorical variable is defined by M distinct categories, l_1, l_2,..., l_M. The set of possible splits at that node for that variable is the set of all subsets of $\{l_1, l_2,...,l_M\}$. Denote by τ_L and τ_R the left daughter-node and right daughter-node, respectively, emanating from a (parent) node τ. In general there will be 2_{M-1-1} distinct splits for an M-categorical variable.

Node impurity function

At each stage of recursive partitioning, all of the allowable ways of splitting a subset of L are considered, and the one which leads to the greatest increase in node purity is chosen. This can be accomplished using what is called an "impurity function", which is nothing but a function of the proportion of the learning sample belonging to the possible classes of the response variable. To choose the best split over all variables, first the best split for a given variable has to be determined. Accordingly, a goodness of split criterion is defined. The impurity function should be such that it is maximized whenever a subset of L corresponding to a node in the tree contains an equal number of each of the possible classes (since in that case it is not possible or, is too difficult to sensibly associate that node with a particular class). The impurity function should assume its minimum value for a node that is completely pure, having all cases from the learning sample corresponding to the node belonging to the same class.

Pruning procedure

A specific way to create a useful sequence of different-sized trees is to use "minimum cost-complexity pruning". In this process, a nested sequence of sub trees of the initial large tree is created by "weakest-link cutting". With weakest-link cutting (pruning), all of the nodes that arise from a specific non-terminal node are pruned off (leaving that specific node itself as terminal node), and the specific node selected is the one for which the corresponding pruned nodes provide the smallest per node decrease in the re-substitution misclassification rate. If two or more choices for a cut in the pruning process would produce the same per node decrease in the re-substitution misclassification rate, then pruning off the largest number of nodes is preferred. In some cases (minimal pruning cases), just two daughter terminal nodes are pruned from a parent node, making it a terminal node. But in other cases, a larger group of descendant nodes are pruned all at once from an internal node of the tree. Barnabas *et al.* [2] studied about wheat production affected by climate change are mainly concerned with future CO2 concentrations and its Analysis showed that a more serious problem associated with global warming might be an increase in the frequency of heat stress around flowering, which represents a greater risk for sustainable wheat production. Goswami *et al.* [3] diagnosed the causes of reduction in wheat yield in Ludhehana Province of India while the visible crop condition was the best. They pointed out that the occurrence of mild heat wave (13 days above normal (2-3^{0}C) temperatures) in early spring at reproductive stage caused 28% reduction in the grain yield of wheat. Kumar and Singh [4] analysed the climate change and its

impact on wheat production. This studied was related to an analysis of crop-climate relationships, using historic production statistics for wheat crops. An overview of the state of the knowledge of possible effect of the climate variability and change on wheat production indicated that an increase in 1^0C mean temperatures' associated with CO_2 increase, would not cause any significant loss to wheat production, if simple adaptation strategies such as change in planting date and varieties are used. They examine data on the mitigation potential of agro forestry in the humid and sub-humid tropics.Peng *et al.* [5] observed that due to higher night temperatures during 2003, the respiration over ruled the photosynthesis causing reduction in net gain. Rice grain yield declined 10% for each 1^0C increase in minimum temperature.Singh and Thornton, [6] studies over Asia and widely used DSSAT for yield gap analysis, decision making and planning, strategic and tactical management decisions and climate change.

RESULTS AND DISCUSSION

Regression tree approach of weather variables at different growth stages

Seedling emergence (20th Nov – 26th Nov.)

The results indicate that evaporation is the most important variable determining yield variability out of the total time period of 29 years, the average yield was 2.4 t/ha in 11 years when evaporation was more than 2.2 mm/m^2 and average yield was 1.9 t/ha in 18 years. It means when evaporation was more than 2.2 mm/m^2 the average yield was increased by 0.5 t/ha. It is clear that at seedling emergence evaporation more than 2.2 mm/m^2 is beneficial (fig-1).

Tillering (27th Nov- 15th Dec)

Morning Relative humidity is observed to be the most important variable at second stage of wheat growth which explained the maximum variability in the yield out of the total time period of 29 years. The threshold value of Relative humidity is 90% or more than the average yield was 2.4 t/ha in 13 years and average yield was 1.8 t/ha in 18 years. It means when Morning Relative humidity was more than 90% the average yield was increased by 0.6 t/ha. It is clear that at tillering stage Morning Relative humidity more than 90% is beneficial (fig.-2).

Node stage (16th Dec- 5th Jan)

Maximum Temperature is observed to be the most important variable at third stage of wheat growth which explained the maximum variability in the yield out of the total time period of 29 years. When maximum temperature was 23^{o}C or more than the average yield was 1.9 t/ha in 12 years and Maximum Temperature is less than 23^{o}C average yield was 2.2 t/ha in 17 years.It means when Maximum Temperature was less than 23^{o}C the average yield was increased by 0.3 t/ha. It is clear that at node stage Maximum Temperature more than 23^{o}C is not beneficial (fig.-3).

Boot stage (6th Jan – 30th Jan)

The optimum regression tree at boot stage had two splits with terminal node the first split in the tree occurred when evaporation was 1.4 mm/m^2 which suggested that the evaporation was the most important factor affecting yield. The split produced two groups of data. One was with lower evaporation level and other with higher evaporation level with average yield of 2.5 t/ha in 9 years. Morning Relative humidity was the second most important variable impacting yield only when evaporation was less than 1.4 mm/m^2 then it splits into Morning Relative humidity. Morning Relative humidity again splits into two nodes when Morning Relative humidity was less than 91% then yield was 1.8 t/ha in 13 years. When Morning Relative humidity was more than 91% then yield was 2.1 t/ha in 7 years (fig-4).

Ear head emergence (31st Jan – 20th Feb)

The optimum regression tree at ear head emergence stage had two splits with terminal node the first split in the tree occurred when Maximum Temperature was 24^{o}C or more which suggested that the Maximum Temperature was the most important factor affecting yield. The split produced two groups of data. One was with lower Maximum Temperature level and other with higher Maximum Temperature level with average yield of 2.5 t/ha in 8 years when Maximum Temperature was less than 24^{o}C. Evaporation was the second most important variable impacting yield only when Maximum Temperature was more than or equal to 24^{o}C then it splits into Evaporation. Evaporation again splits into two nodes when Evaporation was less than 2 mm/m^2 then yield was 1.7 t/ha in 12 years. When Evaporation was more than 2 mm/m^2 then yield was 2.2 t/ha in 9 years. It means at ear head emergence Maximum Temperature was less than 24^{o}C then it was beneficial for the wheat yield (fig-5).

Milk stage (21st Feb – 5th March)

The optimum regression tree at milk stage had two splits with terminal node the first split in the tree occurred when evaporation was more than 3.4 mm/m^2 which suggested that the evaporation was the most important factor affecting yield. The split produced two groups of data. One was with lower evaporation level and other with higher evaporation level, with average yield of 2.5 t/ha in 7 years when evaporation was less than 3.4 mm/m^2. Evening Relative humidity was the second most important

variable impacting yield only when Evaporation was less than 3.4 mm/m^2 then it splits into Relative humidity. Evening Relative humidity again splits into two nodes when Evening Relative humidity was less than 48% then yield was 1.8 t/ha in 13 years. When Evening Relative humidity was more than 48% then yield was 2.2 t/ha in 9 years. It means at stage-6 Evaporation was more than 3.4 mm/m^2 then it was beneficial for the wheat yield (fig-6).

Dough stage (6th March – 15th March)

Morning Relative humidity is observed to be the most important variable at seven stage of wheat growth which explained the maximum variability in the yield out of the total time period of 29 years. The threshold value of Relative humidity was more than 79% then the average yield was 2.3 t/ha in 18 years and average yield was 1.8 t/ha in 11 years. It means when Morning Relative humidity was more than 79% the average yield was increased by 0.5 t/ha. It is clear that at dough stage Morning Relative humidity more than 79 % is beneficial(fig-7).

Maturity stage (16th March -31st March)

The optimum regression tree at maturity stage had two splits with terminal node the first split in the tree occurred when Morning Relative Humidity was more than 82% which suggested that the Morning Relative Humidity was the most important factor affecting yield. The split produced two groups of data. One was with lower Morning Relative Humidity level and other with higher Morning Relative Humidity level, with average yield of 2.5 t/ha in 9 years when Morning Relative Humidity was less than 82%. Morning Relative Humidity was the second most important variable impacting yield only when Morning Relative Humidity was less than 82% then it splits into again Morning Relative Humidity. Morning Relative Humidity again splits into two nodes when Morning Relative Humidity was less than 76% then yield was 1.6 t/ha in 7 years. When Morning Relative Humidity was more than 76% then yield was 2.1 t/ha in 13 years. It means at maturity Morning Relative Humidity was more than 2.5% then it was beneficial for the wheat yield(fig-8).

Fig. 1- Stage 1: Seedling emergence (20th Nov - 26th Nov).

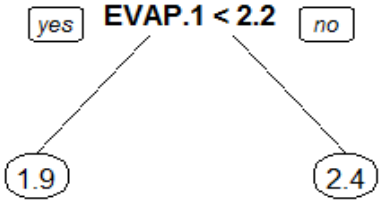


Fig. 2- Stage 2: Tillering (27th Nov- 15th Dec).

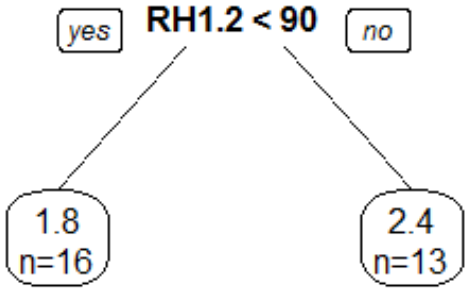


Fig.3- Stage 3: Node stage (16th Dec- 5th Jan)

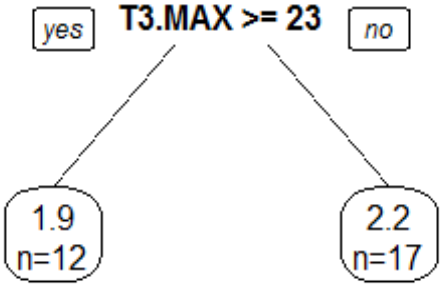


Fig. 4- Stage 4: Boot stage (6thJan – 30th Jan

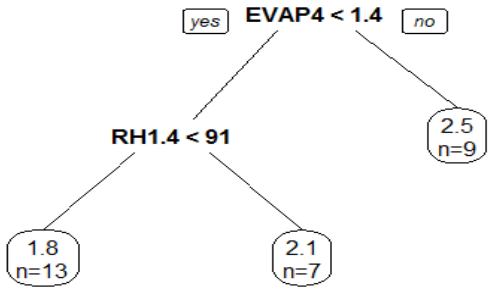


Fig. 5- Stage 5: Ear head emergence (31st Jan – 20th Feb)

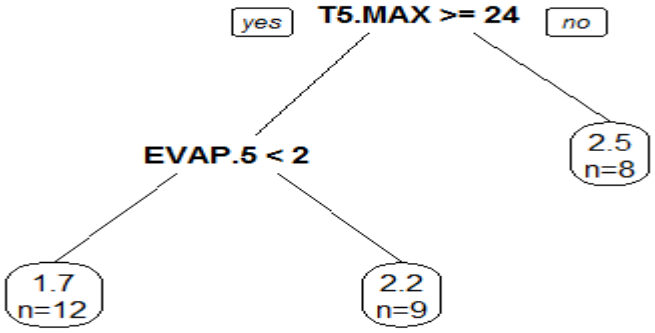


Fig.6- Stage 6: Milk stage (21st Feb – 5th March)

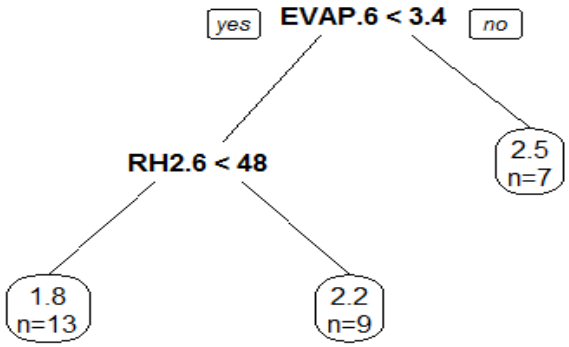


Fig.7- Stage 7: Dough stage (6th March – 15th March)

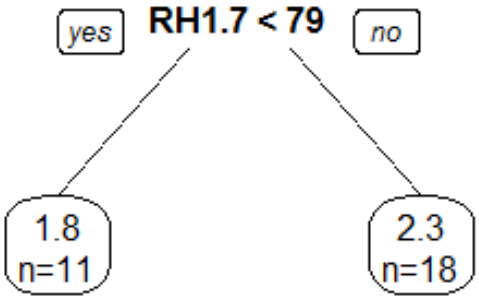
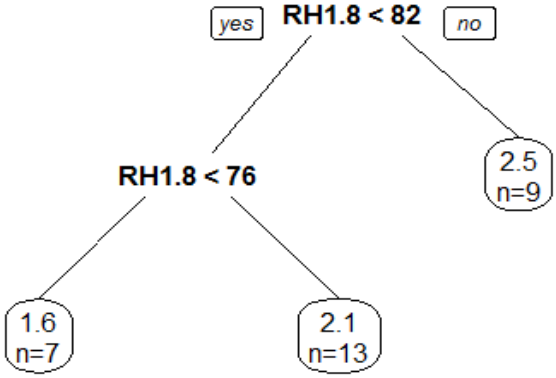


Fig.8- Stage 8: Maturity stage (16th March -31st March)



CONCLUSION

CART analysis allowed to: (i) unravel interactions and combined effects in a complex dataset; (ii) identify thresholds in the relationship between wheat yield and different weather variables. The approach provided insight into the structure of interrelationships within the dataset more easily as compared to multiple regression modeling. Evaporation was found to be the most important variable in determining the yield variability at the stages: seedling emergence, boot stage and milk stage. At tillering, dough and maturity stages, morning relative humidity was identified as the most important variable in explaining the yield variability. Maximum temperature was identified as the most important variable at node and ear head emergence. Regression tree presented a simple decision rule to identify the importance of different weather variables at different stages which is expected to help in assessing the estimated impact of the variables at different growth stages of the crop which in turn will help to adopt preventive measures to deal with the adverse effects of change in weather.

REFERENCES

1. Breiman, L., Freidman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and regression trees. *Wadsworth, Belmont CA*.
2. Barnabas, B., Jager, K. and Feher, A.(2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environment*.**31**: 11-38.
3. Goswami, A. K., Chauhan, R.S. and Dalawat, D.S. (2005): Revs of Hydroxytriazene. *Anal. chem.***24**:75– 102.
4. Kumar and Singh. (2014). Climate change and its impact on wheat production and mitigation through. agroforestry technologies. *International Journal on Environmental Sciences*, Vol. **5** (Issue 1), pp. 73-90.
5. Peng, S.B. (2004). Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci. U. S. A.***101** (27), 9971–9975.
6. Singh, U. and Thornton, P.K.(1992). Using crop models for sustainability and environmental quality assessment**.** *Outlook Agriculture*. **21**:209–218.

Citation of this Article
Subhash Kumar, S. P. Singh and Mahesh Kumar. Effect of Climatic Variables on Different Stages of Wheat with the help of powerful Statistical tool CART. Bull. Env. Pharmacol. Life Sci., Vol 7 [1] December : 59-65