A STUDY ON IMPACT OF CLIMATE CHANGE ON WHEAT CROP YIELD AND DEVELOPMENT OF STATISTICAL MODELS FOR PRE- HARVEST FORECAST OF CROP - YIELD IN AYODHYA DISTRICT OF EASTERN U.P

THESIS
SUBMITTED TO THE

ACHARYA NARENDRA DEVA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

IN FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
AGRICULTURAL STATISTICS

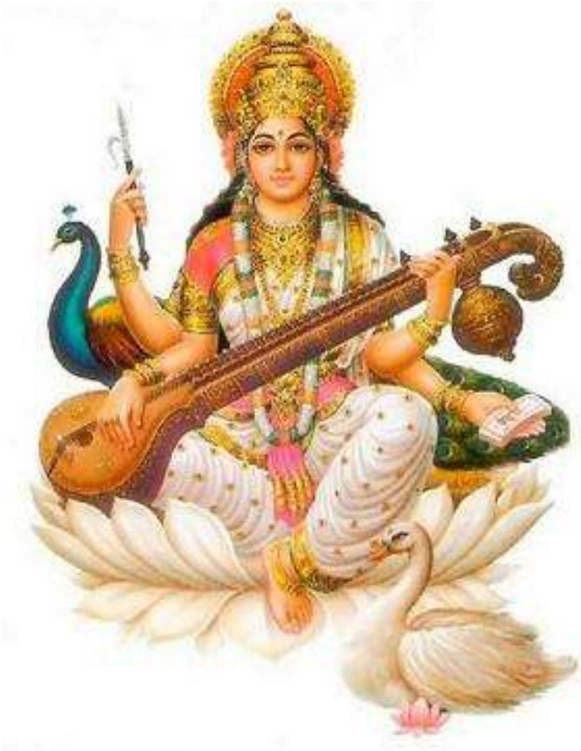
By

RAVI PRAKASH GUPTA

I.D. No. A-8957/15/17

DEPARTMENT OF AGRICULTURAL STATISTICS
ACHARYA NARENDRA DEVA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
NARENDRA NAGAR (KUMARGANJ) AYODHYA-224 229 (U.P.),
INDIA

2021



DEDICATED
To My
Beloved Parents

Mr. Satya dev

Mrs. Basanta Devi

Ravi…. .

Dr. V.N RAI
Professor/HEAD/DEAN

Deptt. Of Agricultural Statistics
A.N.D. University of Agriculture &Technology,
Narendra Nagar (Kumarganj),
Ayodhya- 224229, (U.P.) INDIA

CERTIFICATE-I

This is to certify that the thesis entitled "*A STUDY ON IMPACT OF CLIMATE CHANGE ON WHEAT CROP YIELD AND DEVELOPMENT OF STATISTICAL MODELS FOR PRE- HARVEST FORECAST OF CROP - YIELD IN AYODHYA DISTRICT OF EASTERN U.P*" submitted for the degree of '**Doctor of Philosophy**' in subject of '**Agricultural Statistics**' of the Acharya Narendra Deva University of Agriculture & Technology, Narendra Nagar (Kumarganj), Ayodhya (U.P.) is a bonafide research work carried out **Mr. Ravi Prakash Gupta**, I.D No. **A-8957/15/17**, under my supervision and that no part of this thesis has been submitted for any other degree.

The assistance and help received during the course of investigation have been duly acknowledged.

Narendra Nagar
…………. 2021

(Dr. V.N Rai)
Major Advisor & Chairman

CERTIFICATE–II

This is to certify that the thesis entitled "*A STUDY ON IMPACT OF CLIMATE CHANGE ON WHEAT CROP YIELD AND DEVELOPMENT OF STATISTICAL MODELS FOR PRE- HARVEST FORECAST OF CROP - YIELD IN AYODHYA DISTRICT OF EASTERN U.P*" submitted by **Ravi Prakash Gupta,** I.D. No. **A-8957/15/17** to the Acharya Narendra Deva University of Agriculture & Technology, Narendra Nagar (Kumarganj), Ayodhya (U.P.) in partial fulfillment of the requirement for the degree of '**Doctor of Philosophy**' in the subject of '**Agricultural Statistics**' has been approved by the Student's Advisory Committee after an oral examination on the same in collaboration with an external examiner.

(External Examiner)

Advisory Committee

Major Advisor & Chairman	24/9/21 **(Dr. V.N.Rai)** Professor Deptt. of Agricultural Statistics
Member (Major)	24-09-2021 **(Dr. Manish Kumar)** Assistant Professor Deptt. of Agricultural Statistics
Member (Minor)	24.09.2021 **(Dr. Supriya)** Assistant Professor Deptt. of Agricultural Economics
Member (Related Field)	**(Dr. B.N Singh)** Assistant Professor Deptt. of Agricultural Agronomy
Dean's Nominee	**(Dr. K.K Singh)** Assistant Professor Deptt. of Agricultural Economics
Dean, College of Agriculture	24/9/21 **Head of Department**

ACKNOWLEDGEMENT

*All and every praise be for "**GOD**", the most magnificent, the most beneficent. I pray to him for his guidance and protection throughout my life. It is with his guidance and help that I have been able to complete this huge task.*

Let me first of all, I feel great pleasure in taking the opportunity to express my heart-felt thanks and sense of gratitude to **Dr. B.V.S Sisodia**, *Professor, Agril. Statistics for his incredible support, valuable guidance, continued patience, encouragement, inspiration, constructive criticism and suggestions during the course of investigation. Without his stimulating guidance, meticulous counsel and tremendous encouragement, this study would not have been possible.*

I feel golden opportunity and proud privilege to work under the talented and inspiring guidance of **Dr. V.N Rai**, *Professor Agril. Statistics, and Chairman of advisory committee, for his imperative and facile suggestion, conspicuous ability, constructive criticism and above all the inculcating in me a spirit of independent thinking with confidence which made this endeavor possible. I am extremely indebted to him for being meticulous through.*

I am highly indebted to **Dr. Supriya**, *Asstt. Professor, Agril. Economics,* **Dr. K.K Singh,** *Asstt. Professor Department of Agril. Economics and,* **Dr. B.N Singh** *Asstt. Professor Department of Agronomy, who have paid special attention to the constructive criticism of this manuscript and they have also taken pain in giving proper suggestions, decent advice and liberal help rendered throughout the period of study.*

I also offer gratefulness to all teaching staff of my department especially to **Dr. Manish Kumar** *Assitt. Prof. and* **Dr Vishal Mehta** *Assitt. Prof. for providing necessary facilities and fruitful suggestions during preparation of this manuscript and entire courses of my study.*

I feel paucity of words in expressing my outmost gratefulness and reverence to ***Dr. Bijendra Singh***, *Hon'ble Vice-Chancellor and* ***Dr. V.N Rai***, *Dean, College of*

Agriculture of this university for providing all sorts of facilities to carry out the present investigation.

And thanks for non-teaching staff members (Sri K.K Yadav, Smt. Sashi Prabha Sri Mehdi, Sri Vinod, Sri Tarun Singh Sri Jagram, sri Ram manorath Mishra) Department of Agriculture Statistics for providing necessary facilities and cordial help whenever needed.

I am over whelmed with rejoice to avail this rare opportunity to devote this work evince my profound sense of veneration and gratitude to my father **Shri Satya Dev**, my mother **Smt. Basanta Devi**, my di **Mrs. Deepika Gupta** Jiju **Mr. Pannalal Gupta** , and my lovely nephew **Mitesh Gupta** ,**Ritesh Gupta** , and **Lovee**, **Yuvi** and **Harsh** and **Arav gupta** who have always moral support chi-tiring activities for whole of the day that remain a source of inspiration and engorgement to recharge me to work hard and exhilaration has always animated me to rise against problems and face them smiling.

I am highly thankful to my elder brother **Mr. Rajesh Gupta ,Drx. Mukesh Gupta ,Drx. Rahul Gupta ,Mr. Rohit Gupta** & bhabi **Smt. Alka Gupta & Smt. Mamta Devi** for their sweetest love, affection and moral support.

I would like to express my affection to my loving friends **Savan Kumar, Ramesh Bind,Rahul Jaiswal, Pramod Maurya, vijay Pal, R.D.**

I also wish to express my indebtedness to my guest faculty **Dr. Annu**, Dr. Ekta pandey, and my super seniors **Dr.K.K.Mourya** Dr. Amar Singh, Dr. Dhyanesh Shukla , **Dr. Manoj Sharma** Dr. Hargovind Singh , Dr. Azfar, Dr. Monika, **Dr. Snehdeep**, Dr.Sweta ,**Dr. Ajay**, Dr. Sandhya Verma, **Dr Sarvesh Maurya**, and my Classmate Naveen, Brijesh kumar (Biochemistry), Sarvesh Kumar Dubey and My seniors Elango s, Amit Singh , Ashish Shukla, Amit Mishra because their patience, inspiration, love and co-operation, this task would have been a more difficult challenge to me And my lovely juniors, Ajeet kumar, Pavan, Ravi Prakash Chaudhary, Ashish Tiwari & Vijay Kannojiya.

Besides these, many more people helped me. If I don't list them all it is not for lack of gratitude but for lack of space which compels me to hold my pen down and therefore, I give them all my thanks.

Acknowledgment are generally written as the preface of a thesis in order to thank those who have been helpful in either ways to complete it but I am going to keep those all people here so that wherever I go through it, I should be very honest to accept that it was ***not only me*** *who did it.*

Again I thank ***'God'*** *and my family for his blessing bestowed upon me.*

Narendra Nagar, Ayodhya

.... 2021 ***(Ravi Prakash Gupta)***

CONTENTS

LIST OF TABLES

Chapter-I

INTRODUCTION

Wheat (*Triticum aestivum* L.) is a second most important cereal crop after rice, play significance role in food security. It is a type of grasses family cultivated for millions of peoples as a staple food. Wheat is major cereal crop in temperate region being utilised as a human food and livestock feed. It can be grown in tropical, sub -tropical and temperate region. Wheat shows highest yield potential in regions favoured with cool, moist weather and followed by dry condition. Nutritionally wheat is rich source of essential amino acids, vitamins, minerals, phytochemicals and other dietary fibre component of diets. It provides delicious food for human like bread, biscuit, noodles, pasta, dalia and others food product. The cereal is one of the cheapest sources of energy, provides a major share of protein (20%) and calorie intake (19%) from consumption.

In India, wheat is the second most cultivated food crop after rice in production as well as in area. The crop has been under cultivation in about 30 million hectares (14% of global area) to produce the all-time highest output of 99.70 million tonnes of wheat (13.64% of world production with a record average productivity of 3371 kg/ha (T.M and G. et.al 2017-18). In India mostly wheat grown in northern region during winter season. It covered 26% areas of total cereals area of country. Uttar Pradesh, Punjab, Rajasthan, Haryana, Madhya Pradesh, Bihar and Gujarat are the major wheat growing states in the country. Uttar Pradesh ranks first in area (36.58%) as well as in production (36.27%). It is the highest wheat growing state of the country and produces 28 million tonnes of wheat. Being as a second country in population, it is also the second wheat consuming country after China, with a huge wheat demand. It can be grown not only in the tropical and sub-tropical zones but also in the temperate zone and the cold tracts of the far north, beyond even 60 degree latitude. The best wheat are produced in areas favoured with cool, moist weather during the major portion of the growing period followed by dry, warm weather to enable the grain to ripe properly. India raises almost exclusively winter wheat. The major wheat growing areas in India are located in the northern region of the country. Uttar Pradesh holds the first rank in wheat area (9.75 million ha) as well as production (31.88 million tonnes) during 2017-18 However productivity wise, Uttar Pradesh (3.27 tonnes per ha, hereafter 't/ha') is not the leading state which is even less than the national average (3.37 t/ha) in the same period (Ministry of Agriculture & Farmers' Welfare, 2019).Uttar Pradesh produces maximum wheat in India, accounting for about 33 percent of India's total wheat production (2017-18). The Punjab is the second largest producer of wheat in India, accounting about 18.50 percent of the nation's total wheat production (2017-18). The sowing of winter wheat begins about the first of October and runs through the end of December. Wheat usually begins to head in January, with harvest following in last week of March and April.

Forecasting opens menu window on to future. It plays a vital role in most of our activities and in all we do concern about future. Establishing a functional crop yield/ production forecasting system is an extremely important component of a national agricultural statistical system. Crop forecasts are used for a number of policy decisions such as those relating to procurement, trade, storage, transportation, distribution, import and export of food grains in general, and for implementing food security programme in particular. Crop production forecasts are not only meant to serve the interests of governments but other stakeholders in the agricultural sector also find use for crop forecast data in their day-to-day decision functions. The value of the various policy and business decisions could be enhanced if these are supported by a strong system of food crop production forecasting.

Weather variables affect the crop at different stages of crop development. Thus the extent of weather influence on crop yield depends not only on magnitude of weather variable but also on the distribution pattern of weather over the crop season which as such calls for the necessity of dividing the whole crop season into finer intervals. However, doing so will increase number of variable in the models and in turn a large number of observations may not be available for precise estimation of these parameters. Suitable dimension technique such as like transformation say forming indices or multivariate technique such as principle component, discriminate function etc. can be employed to tackle such problem.

Understanding the impact of climate variability and change on crop yields is fundamental to the success of such research. It is also an essential step towards the development of key adaptive strategies to cope up with climate change. Pre-harvest forecasting is extremely useful in formulation of polices regarding stock, distributions and supply of agricultural produce to various part of the country. Pest and diseases are one of the major causes of reduction in crop yield. Thus, there is need to develop forewarning model which provides advance information for outbreak of pest and diseases attack so that remedial measures can be implemented before the actual onset of the damage. Timely application of control measures may reduce the yield losses. This information would be obtained through modelling qualitative data (Agrawal et al. 2007).

The crop yield depends on many types of variables which are discussed above viz., weather factors, plant characters during crop growth stages, pest and diseases attack, agricultural inputs etc. That will provide better forecast if all these variables are considered while forecasting crop yield. But it may not be realistic to develop a single model based on different type of the data. In such case, separate models may be developed based on different group of variables.

The prime object of the study is to make use of crop-weather relationships to develop pre-harvest forecast models. Keeping in view the above facts following objectives have been framed for the present study.

1. To study the effect of weather variables on the yield of wheat crop in Ayodhya district of Uttar Pradesh.
2. To develop statistical models for pre-harvest forecast of yield of wheat crop based on weather indices obtained using weekly data on weather variables.
3. To develop statistical models for pre-harvest forecast of yield of wheat crop by applying discriminant function analysis of weather indices as well as weekly data on weather variables.
4. To develop statistical models for pre-harvest forecast of yield of wheat crop by applying principle component analysis of weather indices as well as weekly data on weather variables.
5. To make a comparative study of the models developed in (2) (3)& (4).

Chapter-II

REVIEW OF LITERATURE

Several studies have been carried out to examine the effect of different weather variables on crop production. The weather variables affect the crop yield during various stages of its development. Keeping in view the objective of the study, a review of the available literature is presented here with a view to survey the various methodologies employed by the researchers for developing the forecasting models. The review is organized into the following heads according to different approaches used for developing pre-harvest forecast models.

2.1 Regression Technique

2.2 Discriminant Function Analysis

2.3 Principal Component Analysis.

2.1 Regression Technique

Fisher (1924) developed first time a statistical model to study the crop weather relationship. He assumed that the effect of change in weather variables in successive weeks would not be an abrupt or erratic change but an orderly one that follow some mathematical law. He assumed that these effects are composed of the terms of polynomials function of time. Further, the value of weather variables in w^{th} week, X_w was also expressed in terms of orthogonal function of time. Substituting these in usual regression equation

$$Y = A_0 + A_1 X_1 + A_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots A_n X_n$$

where Y denotes yield and X_w rainfall in wth week (w = 1, 2, …n,) and utilizing the properties of orthogonal polynomial a normalized function was obtained as

$$Y = A_0 + a_0 \rho_0 + a_1 \rho_2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots a_k \rho_k$$

where $A_0, a_0, a_1 \ldots a_k$ are constant to determined and $\rho_i \left(i = 1, \ldots k\right)$ are distribution constant of X_w. Fisher suggested to use k=5, for most of the practical situations. In fitting equation for k=5, the number of constants to be evaluated will remain 7, no matter how finely growing season is divided. This model was used by Fisher (1924) for studying the influence of rainfall on the yield of wheat.

Hendricks and Scholl (1943) modified Fisher's technique and divided the crop season into n weekly intervals and have assumed that a second degree polynomial in week number would be sufficiently flexible to express the relationship. Mathematically, it can be expressed as

$$A_w = a_o + a_1 w + a_2 w^2$$

In particular $A_1 = a_0 + 1.a_1 + 1^2.a_2$

$$A_2 = a_0 + 2.a_1 + 2^2.a_2$$

$$\vdots$$

$$A_n = a_0 + n.a_1 + n^2.a_2$$

Substitute the expression for A_w in regression equation the model was obtained as

$$Y = A_0 + a_0 \sum_w X_w + a_1 \sum_w W X_w + a_2 \sum_w W^2 X_w$$

In this model, number of constant to be determined reduces to 4, irrespective of n. Equation was extended for two weather variables to study joint effects.

The model obtained was

$$\begin{aligned} Y = A_0 + a_0 \sum_w X_{1w} + a_1 \sum_w W X_{1w} + a_2 \sum_w w^2 X_w + \\ b_0 \sum_w X_{2w} + b_1 \sum_w W X_{1w} + b_2 \sum_w W^2 X_{2w} + \\ c_0 \sum_w X_{1w} X_{2w} + c_1 \sum_w W X_{1w} X_{2w} + c_2 \sum_w W^2 X_{1w} X_{2w} + \delta T \end{aligned}$$

Since the data for such studies extended over a long period of years, an additional variate T representing the year was included to make allowance for the trend.

Sarkar (1965) studied the weather yield relationship during tillering and elongation phase of sugarcane separately. It was observed that 50 per cent of the total variation was accumulated by various weather factors (i.e. maximum and minimum temperature, rainfall and sunshine) during tillering phase while only 25% is accounted by weather factor during elongation phase.

Runge (1968) applied Hendricks and Scholl (1943) model to study joint effects of rainfall and temperature on corn yield and found that maximum daily temperature and rainfall had a large effect on corn yield from 25 days before to

15 days after anthesis. It was concluded that high temperature can be beneficial to corn yield if moisture available to the plant is adequate.

Sardna *et al.* (1972) suggested another approach in study of Jute crop. They studied biometrical character of Jute crop as Plant population/plot, plant height, and girth of internodes used as regressor and found that all are significantly correlated with yield.

Huda *et al.* (1975) studied the intensity and distribution pattern of weather parameters at different stages of growth of the rice yield at Pantnagar. They reported that second degree multiple regression can be employed to study the relationship between rice yield and weather variables. It was reported that above average total rainfall is beneficial during nursery stage but has adverse effect at the vegetative phase. It was also found that maximum and minimum daily temperature has beneficial effect during nursery period but showed adverse effect during maturity phase. However, reproductive phase is susceptible to a change of 1% maximum and minimum relative humidity.

Agrawal *et al*. (1980) developed the forecasting models for the rice yield in Raipur district based on weekly data of weather parameters. Stepwise Regression Analysis was applied to develop the models. It was found that forecasting of rice yield using weather variables is best possible only two and half months before the harvest.

Jain *et al*. (1980) applied multiple Regression Technique for developing rice yield forecasting models in Raipur district of Madhya Pradesh. It was reported that effect of increase in total weekly rainfall and average relative humidity was beneficial throughout the crop seasion. He found that developed model was reliable to forecast rice yield two and half months before the harvest.

Choudhary and Sarkar (1981) used the correlation technique for identifying weather factors and their duration, which exert influence on the crop growth and the yield for rice grown in the south eastern parts of Madhya Pradesh. Relationship between crop yields (during its various growths Phase) with important weather parameters was obtained using multiple linear regression models. The value of coefficient of multiple determinations was obtained as 0.81 indicating 81% of variation in the yield due to explanatory variable included in the study.

Agrawal and Jain (1982) developed forecasting models for the rice yield in Raipur district of Madhya Pradesh based on weekly data of weather Parameters. It was observed that weather variables with time trend accounted more than 70% of the total variation in yield at about 2.5 months before harvest.

Agrawal and Mehta (1982) studied the several weather based forecasting models for crop yield of rice, wheat, sorghum, maize and sugarcane at selected districts/agro-climatic zones/ state of India using regression analysis, discriminant function analysis and water balance technique. It was reported that reliable forecast of crop yield could be provided before harvest. They also developed models for forewarning of important pests/ diseases in rice, mustard, pigeon pea, sugarcane, groundnut, mango, potato and cotton using regression analysis and Artificial Neural Network (ANN) Technique. It was found that the reliable forewarning of important pests/diseases could be achieved at least one week in advance.

Jain *et al.* (1985) used this method for forecasting with growth indices as explanatory variable and found better than conventional linear regression model where weather variables used as explanatory variables.

Singh and Bapat (1988) developed a suitable model for pre-harvest forecast of sugarcane yield. A pilot study was carried out during 1977 to 1980. Stratified multi-stage random sampling design was used for selection of fields. The multiple-regression technique was adopted for fitting the regression models. For selection of yield attributes to be entered finally in the forecast model, the technique of step-wise regression was used. It was observed that 3 yield attributes (number of canes, plant height and girth of cane) could be used to forecast the sugarcane yield about 4 months before the harves**t.**

Rao and Singh (1990) studied distributions pattern of weather variables and developed methodology for forecasting extreme values of weather parameters using time series data at Pantnagar. It was found that square root model ($Y=a+b\sqrt{x}+cx$) could be used to predict wheat yield based on meteorological observations.

Khan *et al.* (1995) applied multiple regression model to predict the yields of winter rice, and rape and mustard [Brassica juncea] in different districts of West Bengal using the rainfall distribution before sowing and during the growing season. The models developed were able to account for more than 70% of the total variation in rice yield in most of the districts. The range was 23-98%. For rapeseed and mustard, agronomic inputs (cultivar choice in particular) affected the prediction equation greatly without which the coefficient of determination (R^2) was reduced. Out of the 9 districts, the regression models developed for 6 district and accounted for more than 70% variability in yield and their R^2 values were significant at 5% level. The models could be used for pre-harvest forecast of winter rice and mustard in West Bengal by 21 October and 31 December respecting, which are quite earlier than the final forecasts made on 31 January and 13 April, respectively, by the State agencies.

Pal *et al.* (1996) applied the linear multiple regression equation for predicting sugarcane productivity in western Uttar Pradesh based on weather parameters. It was found that weather conditions perverting from germination to elongation phase accounted for 61% variation followed by 60% by preventing weather conditions during growth, tillering to elongation. The relationship between estimated and actual cane productivity was found significant.

Pal and Saini (1997) used multiple linear regression equations for estimating sugarcane productivity at Pantnagar and Muzaffarnagar of Uttar Pradesh state using weather parameters. Correlation between weather parameter and cane productivity were obtained. It was found that weather parameters caused 88.4% of variations in the cane productivity at Pantnagar and 88.6% at Muaffarnagar. The relationship between estimated and actual cane yield was found highly significant at both Places.

Marcelis and Gijizen (1998) developed model for predicting the weekly fresh weight harvest of cucumber fruits and their quality in Wageningen using average weather data and green house parameters. The model was validated by comparing simulation result with production data of 10 commercial growers in 1996 and 14 growers in 1997 (January-May). It was found that average standard error of the weekly prediction of the fresh weight yield was 14.9%, while the standard error of the annual yield was 2.8%. It showed that average weekly standard errors of the predicted average fruit size and per centage of second class fruit were 6.5% and 5.3%, respectively.

Rai *et al.* (1998) developed a yield forecasting model for wheat (*Triticum aestivum*). Two types of functions i.e. linear and quadratic were fitted. It was found that the explanatory variables could explain the variation in yield up to 62% if they were considered to be linearly related but 68% provided the function was of quadratic in nature. The linear type of model was more valid than a quadratic one. It is suggested that this forecast model could successfully be used for obtaining advance estimates of wheat yield for a period of 4 years.

Saha *et al*. (1998) revealed that the most important yield affecting attributes in lac for the summer season crop of Rangeeni lac on Butea monosperma, 14 factors were studied. Significant and positive correlation was found with yield and quantity of brood lac used for raising lac insect (*Kerria lacca*) culture, number of host shoots with lac culture, height of host crown, length of lac insect settlement per shoot, volume occupied by 100 female lac insects, diameter of host canopy and number of stumps per tree in decreasing order of magnitude. The forecast model, retaining quantity of brood lac used, number of host shoots with lac culture per tree, length of lac insect settlement per shoot and number of living lac insect per cm^2, explained 51% variation in yield. The model developed showed that forecasts of stick lac yield per tree are possible once lac insect culture is established on shoots (3-4 weeks after raising the culture). This corresponds to 20-22 weeks before crop harvesting.

Chandrahas (1999) used the data on crop yield and biometrical character and developed the following models for predicting crop yields.

He attempted the following linear multiple regression model.

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_{2.}+........+e.$$

where Y and $X_{i's}$ are yield and plant characters respectively. β_0 and $\beta_{i's}$ are constant to be estimated and e is random error term. He also developed the model based on growth indices of biometrical characters to improve the aforesaid model following the work of Jain *et al.*. (1985).

Kokate *et al.* (2000) developed a statistical model for forecasting the yield of rice in Ratnagiri district of Maharashtra. They applied correlation analysis to know the association of plant characters and climatological parameter with the yield. They applied step-wise regression analysis technique to know the dominant variables in the study. It was found that integrating plant characteristics and climatological factors together in the model provide better estimates for forecasting rice yield.

Ramirez (2000) studied the multiple linear regression analysis for the weather based crop yield models in Laguna and Nueva Ecija (Philippines). It was observed that the trend line of the weather data greatly affects the predicted yield. The predicated verses the actual yield showed absolute per cent error ranging from 1.48 to 45%. The results of the regression analysis indicated that coefficient of determination was highly significant.

Sarkar (2000) applied correlation and regression analysis to develop the agro-meteorological model for forecasting wheat yield in Gujarat. The model could explain about 94% of the variation in the yield due to weather parameter but the prediction was found reasonably outside the analyzed samples.

Agrawal *et al.* (2001) revealed the forecasting model for wheat in Vindhyan Plateau zone of Madhya Pradesh. For developing forecasting models for rice, the two zones viz. Chhattisgarh Plain and Baster Plateau were grouped together. Time series data on weather variables and agricultural inputs were used and multiple linear regression analysis was applied for developing models. It was reported that reliable forecasting yield could be obtained when both the crops were 12 weeks old i.e. about 2 month before harvest.

Sarmah and Handique (2001) developed linear regression models for the pre-harvest forecasting of winter rice yield in Jorhat district of Assam. It was found that 69.10% of the total variability in rice yield is due to weather variables. The deviation of forecasted yield from that of observed yield was ranging from 0.06 to 21.30%. The best time for rice yield forecasting was observed in 43^{rd} standard meteorological week.

Saksena *et al.* (2001) used a different approach which was based on water balance technique. Model for rainfed crops using weighted stress indices have been developed for Rice-Raipur, Sorgham–Delhi & Prbhani and Maize-Delhi. In this approach, water deficit/surplus has been worked out at different phases of crop growth and using suitable weights, accumulated weighted stress index has been developed for each year which was used as regressor in the forecast model. This technique provide forecast six weeks before harvest for rice.

Kandinnan *et al.* (2002) developed crop-weather model for prediction of rice yield in Coimbatore, Tamil Nadu. It was found that model without solar radiation recorded less R^2 (0.63) as compared to model with inclusion of solar radiation recorded high R^2 (0.95). Step-wise regression analysis was used taking seven weather variables which finally included four weather variables in the model with a R^2 value of 92%.

Ozkan and Akcaoz (2002) revealed the relationship between the climatic variables and yields of three crops i.e. wheat (*Triticum aestivum L.*), Maize (*Zea mays L.*) and Cotton (*Gossypium hirsitum L.*) in the Cukurova region of Turkey. In the study, time series data were used to analyze crop yield across various climate factors for the period 1975 to 1999. The climatic variables were arranged according to phonological periods of the examined crops such as planting, flowering and harvesting time. A linear perturbation modes (LPM) was used for the identification of the role of climate variables. 27 climate factors were considered as explanatory variables in the model. Step-wise regression analysis was used for developing the model. The result of the linear perturbation model (LPM) showed that the R^2 values for Wheat, Maize and Cotton were found to be 46.1%, 57.2% and 74.5% respectively. The highest coefficient of variation (CV) was found in Maize production (43.4%) followed by Cotton (23.14%) and wheat (15.29%). The most significant climatic factor affecting deviation in crop yields was found to be temperature at planting, flowering and harvesting time.

Sarkar and Thapliyal (2003) developed the state-wise Agro meteorological model for forecasting wheat yield in Uttar Pradesh. They applied correlation and regression analysis to identify significant correlation between yield and the meteorological parameters. Step-wise regression analysis was done for developing models. It was found that 98% of the variation in the wheat yield was due to meteorological parameters.

Kandiannan *et al.* (2004) studied the crop weather models for prediction of turmeric yield in Coimbatore district of Tamil Nadu State. Significance of correlation coefficient between the monthly climate variables and turmeric yield was tested. The developed multiple regression models gave a reliable forecast of the dry turmeric yield with a coefficient of determination (R^2) value as 89%.

Nain *et al.* (2004) revealed that the methodology for crop yield forecast using a simulation model. The regression coefficients were generated using 10 years' data (1984/85-1994/95) and reliability of the approach was tested on a data set of 5 years' independent data (1995-96 to 1999-2000). The results showed that this approach could capture year to year variability in large area of wheat with reasonable accuracy. The Root Mean Square Error (RMSE) between observed and predicted yield was reported as 0.098 t/ha for the mean yield of 2.072 t/ha (4.72%). The pre-harvest forecast were made using seasonal weather data up to the end of February and climate normal for the rest of the wheat growing season, which showed good agreement with observed wheat yields.

Sharma *et al.* (2004) developed agrometeorological models based on weather parameter for forecasting wheat yield in 6 major districts of Himachal Pradesh. It was observed that rainfall significantly affected the wheat yield which was decreased by 6.8-32.3% compared to previous years, mainly due to moisture stress caused by delayed and insufficient rainfall.

Hansena *et al.* (2004) developed a method using GCM based seasonal rainfall forecast with a wheat simulation model for forecasting district and state aggregate yields in queen-stand, Australia. It predicted yield by linear regression of simulated yield, transformed to correct departures from normality, against GCM predictors optimized by a linear transformation. Cross-validation of predictor selection and regression ensured conservation assessment of prediction accuracy. Yield forecast made, prior to planting, accounted for a significant portion of the variability of simulated yield averaged across the state (r =0.518) and in most wheat, producing districts (r =0.497, area-weighted average among districts). The advantage of the GCM-based forecasts was greatest at the longest lead time.

Varmola *et al.* (2004) used a pre-harvest forecast model for wheat yield based on weather parameters (maximum and minimum temperature, morning and evening relative humidity, wind speed and bright sunshine hours) using step-wise regression technique. Models based on original weather variables with week-wise and crop stage-wise approach and the generated variables taking week number and correlation coefficient as weight. Among the different approaches, the model based on generated weather variables (correlation coefficient as weight) was the best on the basis of R^2 values (0.943) and simulated forecast errors (<6%), which can predict the wheat yield 12 weeks after sowing

Kumar and Bhar (2005) developed the multiple linear regression model for forecasting yield of Indian mustard (*Brassica juncea* L.) at Hisar district of Haryana. They developed models for each growing phase of Mustard. It was reported that the reliable earliest forecasting could be achieved 6 weeks after sowing and the latest forecasting could be done 4-5 weeks before harvesting.

Khan *et al*. (2006) developed weather-based regression equations for predicting seed yield before harvest. Using accumulated rainfall during 36 to 39 standard weeks (X1), 40 to 43 weeks (X2), 44 to 52 weeks (X3) and 1 to 5 weeks (X4); accumulated actual evapotranspiration during 42 to 3 weeks (X5) and 42 to 52 weeks (X6); soil moisture content on the 45th week (X7); water requirement satisfaction index on the 5th week (X8); and technological trend (T) as independent variables and seed yield (Y) as dependent variable from 1952-53 to 1992-93, regression analysis was carried out for development of yield prediction models. The model for 24-Parganas district was able to explain 66% of the total variation in seed yield. The regression model for Midnapore district accounted for 65% of the total variation in seed yield. The models developed for West Dinajpur and Cooch Behar districts explained 54% and 30% of the total variation in seed yield, respectively. The multiple linear regression models developed for 24-Parganas, Midnapore, West Dinajpur and Cooch Behar districts can be useful in the pre-harvest forecast of the seed yields of rapeseed and mustard by the first week of February, which is quite earlier than the final forecast made in April by the State Department of Agriculture in West Bengal.

Lobell *et al.* (2006) studied the weather-based forecasting models for state-wide yield for twelve major crops in California. It was found that the most successfully modeled crop was almonds, with 81% of yield variance captured by the forecast. It was found that the prediction of the most crops relied on weather measurement well before harvest time.

Shamasundaran and Venugopalan (2006) observed that the crop yield forecast plays a vital role in arriving at pre-harvest yield estimate of a standing crop and to identify the stage at which reliable forecasting could be made before final harvest. In this paper, an attempt has been made to apply the regression technique for prediction of yield in rose. Rose, is an important flower crop not only for internal market but is also intended for export, and since it shrivels, estimation of yield of a standing crop before its actual harvest is essential. A model was developed, which showed that information from the first two pickings of a standing crop could be used to forecast rose yield to an extent of 77% two months before final harvest. It was also suggested to have a minimum sample size of 20% to develop such a forecast model

Patel *et al.* (2007) developed the pre-harvest forecasting model of rice yield using weather variables and technological advances using 33-years yield data of Kheda district, Gujarat, India from 1967/68 to 2001/02. The weekly averages of weather variables, namely bright sunshine hour’s, rainfall maximum and minimum temperature and morning relative humidity from 23rd to 42nd standard meteorological weeks of the respective years were considered. The week-wise, crop stage-wise and generated weather variables (weighted) criteria were used. Time trend was also included as an independent variable in all the model. To provide pre-harvest forecasts, different week-intervals were considered. A step-wise regression technique was employed for all forecasting models. The model based on week-wise criteria using original weather variables of 16 weeks provided a reliable pre-harvest forecasting of rice in the Kheda district. The forecasting can be done four weeks before the expected harvest (i.e. 3rd week of September or at the end of the 38th SMW). The average absolute mean forecast errors of the selected model for the subsequent three years were 10.97%. The pre-harvest proposed forecast model accounted for 75.4% variation in rice yield

Aneja *et al.* (2008) proposed statistical methodology for the pre-harvest estimates of cotton yield by taking biometrical characters as explanatory variables. The data were procured through a pilot survey in Hisar (Haryana, India), and the biometrical characters such as height, girth, total number of bolls, number of opened and unopened bolls and yield of first pick were observed at different stages of the crop growth. The yield of first picking at approximately five months after sowing along with number of unopened bolls was found to be fairly adequate for use in building advance estimates of yield of cotton with the help of multiple linear regression models. There was no appreciable gain in transforming the data into different scales and considering the simplicity in operation, the data have been analyzed in the original scale. The estimated yield of cotton (H-1098) in mid October was 1124, 1245, 1409 and 1382 kg/ha for the years 2003-04, 2004-05, 2005-06 and 2006-07, respectively.

Chattopadhyay *et al.* (2008) studied the effect of meteorological parameter on yield of different varieties of cotton (AHH-468, MCU-9 and MCU-10) in Akola district of Maharashtra. It was observed that minimum temperature at vegetative and flowering stage was favorable and decrease in maximum temperature at flowering and boll development stage was conducive for the yield of variety of AHH-468. It was also found that relative humidity was positively correlated with the yield of varieties of AHH-468 and MCU-10. The rainfall at the beginning of the season was favorable for yield of the crop.

Priya and Radhakrishnan (2008) studied both linear and non-linear regression models for forecasting yields of five food crops (rice, cholam, cumbu, ragi and maize) in Tamil Nadu, India. Secondary data for each crop for the period 1954-2004 were used in the analysis. Comparison of the results of the models indicated that the non-linear regression model was better for prediction and was better fitted compared to the linear regression model due its high R^2 value. Among the five food crops, rice had the highest R^2 value. Rice yield was found to be influenced by wholesale price, area and cropping pattern

Sharma *et al.* (2008) developed an agro climatic model for the expected quantum of the rice production in Andhra Pradesh. Growing degree day or heat unit theories of the crops and Integrated Normalized Difference Vegetable Index along with the rainfall are made use of in obtaining the relation with the crop yield of rice. The atmospheric and oceanic indices such as Southern Oscillation Index and sea Surface Temperature of Nino 3 region were also incorporated in developing the multiple regression models for the estimation of rice yield. The rice yield is not varied linearly with rainfall of AP but maintained good relations with the growing degree day units and the satellite derived vegetation Index INDVI. The impact of Sea Surface temperature of Nino 3 region is more on rice yield compared to that of SOI. The suggested statistical agro-meteorological model imbibed the five variables that have significant impact on yield.

Yadav and Patil (2008) revealed the influences of agro-climatic indices on fruit yield of cucumber during *Kharif* in Dapoli, district Ratnagiri, Maharashtra. They obtained the relationship between fruit yield of cucumber and agro-climatic indices. It was found that early sowing of cucumber i.e.

immediately after onset of monsoon produced significantly highest fruit yield over the late sowing.

Garde (2009) used Linear and Non Linear Regression Analysis and linear correlation for crop yield forecasting. Studies have been carried out to forecast crop yield using weather parameters. He used time series data for 27 years (1981-82 to 2007-08) of yield of rice & wheat and weather parameters obtained from G. B. Pant University of Agriculture and Technology, Pantnagar, District Udham Singh Nagar, Uttarakhand, India, for developing forecast models. He also used technical and statistical indicators in the models.

Kumar *et al.* (2009) studied the changes in long-term climate parameters of Kullu valley on the basis of annual as well as monthly rainfall, maximum and minimum temperatures recorded during AD 1962 to 2004. Further, an analysis of the impact of climatic changes during 1981 to 2004 was done on the seed yield of cabbage var. Golden Acre in the upper Kullu valley in Hindukush Himalayas. For this, average monthly maximum temperature and rainfall data of May (pod setting period) and monthly average maximum temperature data of March-April (period of bolting and flowering) were analyzed. It was observed that the average maximum temperature of May rose by 1.58°C. The minimum temperatures for the months of April and August rose by 2.03 and 2.165°C, respectively. From 1981 to 2004, around 40% reduction in seed production per unit area was noted. The relative humidity during the month of May did not have any significant effect on seed yield. Correlation coefficients between mean monthly rainfall during May and seed yield (r= -0.49), mean maximum temperature during April and seed yield (-0.36) and maximum temperature during May and seed yield (-0.39) indicate that when temperature rise, it affects seed production of cabbage adversely. Also, if rainfall increases during May, the seed yield is reduced. It has also been observed that the rainfall during August has decreased and during September it has increased resulting in late onset of autumn thereby suggesting that the planting of cabbage should also be delayed at least by a fortnight to avoid incidence of soft rot and increased seed yield.

Osowski (2010) estimated that the influence of weather conditions (temperature, rainfall) in the period June-August and of the level of resistance of 32 selected potato cultivars on the rate of late blight development, time of 50% haulm destruction and the percentage of haulm destruction at the end of a growing season. The highest values of estimated parameters were recorded for the group of susceptible cultivars. The values of parameters decreased with increasing level of potato resistance to late blight.

Auffhammer *et al.* (2011) indicates that monsoon rainfall became less frequent but more intense in India during the latter half of the Twentieth Century, thus increasing the risk of drought and flood damage to the country's wet-season (kharif) rice crop. Our statistical analysis of state-level Indian data confirms that drought and extreme rainfall negatively affected rice yield (harvest per hectare) in predominantly rainfed areas during 1966–2002,with drought having a much greater impact than extreme rainfall.

Joshi *et al.* (2011) studied the effect of observed climate variables on yield of major food-crops in Nepal, namely rice, wheat, maize, millet, barley and potato based on regression model for historical (1978-2008) climatic data and yield data for the food-crops. The yield growth rate of all the food crops was found to be positive. However, the growth rate for all crops, except potato and wheat, was below population growth rate during the period. Climate variables like temperature and precipitation were the important determinants of crop yields. Trend of precipitation was neither increasing nor decreasing significantly during this period. However, temperature was increasing by 0.7 °C during the period. Climate variables showed some influences on the yield of these major food-crops in Nepal. Increase in summer rain and maximum temperature has contributed positively to rice yield. Also, increase in summer rain and minimum temperature has positive impact on potato yield. However, increase in summer rain and maximum temperature adversely affected the yield of maize and millet. Increase in wheat and barley yield was contributed by current trend of winter rain and temperature. Consideration of spatial variation in similar type of study in Nepal that will be helpful in identifying the region more vulnerable to climate change in terms of crop yield is highly recommended.

Garde *et al.* (2012) explained techniques for development of weather indices which were used as explanatory variables (Predictors) in the multiple regression model. The technique was further modified by incorporating technical and statistical indicators along with developed predictors. The study proposed that modified model incorporating technical and statistical indicators are effectively used for early pre-harvest forecasting for crop yield particularly up to two and half month before harvest.

Garde *et al.* (2012) used the multiple regression model for estimation of wheat productivity for the district Ghazipur in eastern Uttar Pradesh. Weather indices were computed using varied weather parameters for the year 1982-83 to 2005-06. The cross- validation of the developed forecast models were tested their accuracy using the 2006-07. Based on a forecast error percentage it was found that the forecasting model produced the most accurate forecast for 15th week of the crop growing season. The relationship between actual and forecast wheat yield was highly significant being R^2 varied from 0.72 to 0.89 for the different weeks.

Sharma and Singh (2012) studied on processing varieties of potato and its correlation with weather parameters, at Central Potato Research Institute Campus, Modipuram in early, main and spring crops. Comparison of mean leaf hopper population during three different crop seasons revealed that early crop planted during September favoured highest development of leaf hoppers followed by main (Oct-Feb) and spring crops (Dec-April). A distinct varietal difference was observed in appearance and buildup of leaf hoppers. Multiple regression equation based on temperature, relative humidity, wind velocity, sunshine duration and rainfall could explain leaf hopper population ranging from 50-96%.

Azfar *et al.* (2015) used the time series data on rice crop yield in Eastern Uttar Pradesh, pertaining to the period 1990-91 to 2009-10 for fitting the models for studying the effect of weather variables. He used the relevant statistical tools of stepwise regression analysis.

Poonam *et al.* (2017) stepwise multiple regression technique was applied with yield as dependent variable and weather parameters (artificial variables generated from weekly weather values) as independent variables. Another six years' data (2010-11 to 2015-16) were used for the validation of the developed models.

2.2 Discriminant Function Analysis

Goudie (1987) described an appropriate framework for the identification of firms of the corporate sector that might be expected to suffer from financial problems, leading to failure. He used the discriminative model for this purpose.

Carter and Eisner (1997) described the methodology of forecasting rainfall over Puertorico with the help of discriminant function analysis and factor analysis.

Cannor (1999) developed quadratic discriminant function model to forecast the number of salmon fish that would survive in dam to help managers to effectively time the release of reservoir water to mitigate the problem of reduction in survival . He also developed a multiple regression model using discriminant score to predict the time during which survivors would pass the dam.

Rai and Chandrahas (2000) developed pre-harvest forecast model for rice based on weather variables using techniques of discriminant function analysis in Raipur district of chhatishgarh, India. These techniques have been used to find weather score for each year at different phases of crop growth. It was found that forecast to rice crop can be made about two month before harvest.

Lee *et al*. (2002) developed the methodology for forecasting the credit scores using discriminant function technique. They have also showed that how discriminant analysis had improved the credit scoring over other methods.

Sinclair *et al.* (2005) forecasted the sea lion survival in Alaska region using discriminant function analysis and multiple linear regression models.

Yadav (2010) used the technique of discriminant function analysis for the development of pre-harvest forecast model for wheat yield in Faizabad district of U.P., India. Time series data of wheat yield and weather variables were used for the study. In all seven model were developed. The best model showed R^2 value as 97 per cent and the model –I has turned out to be best followed by the model-VI and model-III. However, there has not been much different in the value of adjust R^2 corresponding to model –I, III and VI (97.0, 93.3 and 93.3 per cent respectively). Moreover, it requires to generated number of weather indices for the development of the model-I, which require more labour. On the other hand, model-III and VI are easy to develop which used discriminant function. RMSE of model-VI is little less (0.570) as compared to model-III (0.572). Validation of the models and forecasting of the wheat yield based on the model-VI for the years 2008-09and 2009-10 have favoured the model –VI. Therefore, the model-VI can be recommended for pre- harvest forecast of the wheat yield in practice. However, the model-III is as good as model-VI.

Pandey (2011) developed different models for pre-harvest forecast of rice yield based on some weather variables, agriculture input. He also applied the technique of discriminant function analysis. The effects of individual as well as joint effect of weather variable have also been studied. He used the time series data of rice yield for Faizabad district of Eastern Uttar Pradesh, pertaining to the period 1989-90 to 2009-10, along with weekly data on weather variables for the same period.

Agrawal *et al.* (2012) used of discriminant function analysis for developing wheat yield forecast model for Kanpur, U.P., (India). The quantitative forecasts of yield have been obtained using multiple regression technique taking regressors as weather scores obtained through discriminant function analysis. The approach provided reliable yield forecast about two months before harvest.

Annu (2013) demonstrated the application of discriminant function analysis for development of pre-harvest forecast models for rice and wheat yield using biometrical characters. He used experimental data obtained from the experiments conducted at Main Experiment station N.D.U.A.T. Kumarganj, Faizabad during 2010-11. The relation between actual and forecast yield of the wheat and rice was significant as R^2 was found to be 61.74 % and 61.64 % for respectively.

Tripathi (2013) also made use of technique of discriminant function analysis for the development of pre-harvest forecast model for potato yield in Faizabad district of U.P., India.

Pandey *et al.* (2013) developed models for forecasting rice yield at district level on the basis of weather variables. Stepwise regression was used to screen out the important weather variables and multiple regression approach was subsequently employed to estimate model parameters.

Sisodia et al. (2014) have used discriminant function analysis of meteorological parameter for developing suitable statistical model for forecast of wheat yield in Faizabad district of eastern Uttar Pradesh, India. Time series data on wheat yield and weekly data on weather variables for 20 year (1990-91 to 2009-10) have been used in the study. The discriminant score obtained from discriminant function analysis have been used as regressor variables along with trend variable (T) in development of statistical models. In all six models have been developed. The model developed have been used for forecasting the wheat yield for year 2008-09 and 2009-10 which were not included in the development of models. It has been found that most of the models provide reliable pre-harvest forecast of the wheat yield about two months before the harvest.

Alayande, S. *et al.* (2015) In different areas of applications the term "discriminant analysis" has come to imply distinct meanings, uses, roles, etc. In the fields of learning, psychology, guidance, and others, it has been used for prediction. It is sometimes preferable than logistic regression especially when the sample size is very small and the assumptions are met.

2. 3 Principal Component Analysis

Jain *et al.* (1984) gave a new approach for forecasting. They constructed principal components of biometrical characters for Sorghum crop. These were plant population/plot, plant height, length of earhead, number of healthy earhead/plant, circumferences of earhead, , number of green leaves per plant and length and width of leaves. Under this study they found that plant population/plot, plant height and number of green leaves per plant were significantly correlated with yield. The principal components were obtained by using data of one or more than one point of time.

Jain *et al.* (1985) also constructed growth indices of biometrical characters measured over specific time interval. The growth indices were used as regressor variables in the multiple regressor model. They showed that the procedure also provided reliable forecast of crop yield.

Kumar *et al.* (1998) revealed that the principle component regression gives better precision for the estimates of the ordinary least square regression analysis. Three meteorological variable namely chill unit, number of hours with temperature and number of hours with relative humidity of 40-60 percent during the flowering time were used for forecasting of yield at apple. The variable were chosen on the basis of simple correlation with yield for regression analysis.

Annu (2013) developed the pre-harvest crop yield forecast models on the basis of plant biometrical characters. The relevant statistical tools and techniques like multiple regression analysis, discriminant function analysis, principle component analysis, growth indices and composite forecast have been used for the purpose of development of the models.

Paul (2013) forecasting of crop yield based on historical data and pertinent external climatic information is considered. To this end, Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) time-series model along with its estimation procedure is studied. In the present investigation, five models at five important stages of wheat growth are developed by including the most important weather variables.

Yadav *et al.* (2014) used principal component analysis for developing statistical models for forecasting crop yield. The time series data on wheat yield and weekly weather variable viz. minimum and maximum temperature, relative humidity, wind- velocity and sun- shine hours pertaining to the period 1999 to

2010 in Faizabad district of Uttar Pradesh have been used in this study. Weather indices have been constructed using weekly data on weather variable (Agrawal *et al*.1983). Principal component analysis has been carried out taking different combination of weather indices. Four models have been developed using principal component analysis as regressor variables including time trend and wheat yield as regressand.

Chapter-III

MATERIALS AND STATISTICAL METHODOLOGY

This Chapter comprises of the material used and the methodology utilized for developing models to study the association between crop yield and weather variables, and to develop models for making pre-harvest forecast of yield. In order to facilitate systematic presentation, the chapter is divided into following sub-category:

3.1 General information of the study area

3.2 Sources and description of data

3.3 Statistical methodology used for the development of models

3.4 Measures for the comparison and validation of different models

3.1 General information of the study area and crop covered:

The study has been conducted for Ayodhya district of Eastern Uttar Pradesh, India, which is situated between 26.92^0 N latitude, 81.20^0 E longitude and 26.26^0N latitude 82.07^0 E longitudes respectively. It lies in the Eastern plain zone of Uttar Pradesh with an annual rainfall of 1002 mm, 90% of which is received during mid-June to mid-October.

3.2 Sources and description of data:

Yield data:

Time series data on yield of wheat of Ayodhya district of Uttar Pradesh for 27 years (1990-91 to 2016-17) have been collected from the

Bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh.

Weather data:

Weekly weather data for the same period on seven weather variables viz., Minimum Temperature, Maximum Temperature, Rainfall, Number of rainy days, Relative Humidity at 8.30 and 17.30 hrs and Wind-Velocity have been used in the study. The weekly data on these weather variables have been obtained from the Department of Agricultural meteorology A.N.D.U.A & T Kumarganj Ayodhya U.P. India.

Wheat crop growing season:

Wheat is generally sown in the month of October when average daily temperature falls around 20-25^{0}c. High temperature at the time of sowing causes poor germination, reduced tillering and early onset of flowering which leads to depress the crop yield. The different wheat crop growth stages are explained below:

(i) Seed germination and CRI stage

Seed germination and early growth stage starts from sowing to last week of November. Germination takes 6-7 days after sowing of the crop. The crown root initiation (CRI) occurs in wheat in 25-30 days after sowing of near about 3 weeks from germination. During CRI stage, plant require a huge amount of moisture i.e. irrigation is required during this stage.

(ii) Tillering and Jointing stage

This stage starts after the CRI stage and last up to 40-45 days after sowing or near about 2-3 weeks after CRI stage. At end of tillering stage i.e. 45-60 days after sowing the jointing stage is started which is the peak plant growth condition. During this period temperature tends to be the lowest in the year 1-4^{0}c.

(iii) Reproductive stage

At jointing stage the vegetative growth of the crop is completed and the crop then goes to reproductive stage. This stage lasts 60-85 days after sowing. At this the average daily maximum temperature shows increasing tendency.

3.3 Statistical methodology used for the development of models:

The weekly weather data covering full crop season from 15th April (about a fortnight before sowing) to harvest were utilized for studying the effects of weather variables on yield while data of partial crop season were utilized for developing forecast models. Data for a fortnight before sowing were considered as this period is expected to have effect on establishment of the crop.

3.3.1 Statistical models for studying the relationship between weather variables and crop yield

The statistical models have been proposed by expressing effect of changes in weather variables on yield in w^{th} week as a linear function of respective correlation coefficients between yield and weekly weather data (Agrawal *et al*. 1986). Trend effect on yield is also removed from yield while calculating correlation coefficients of yield with weather variables to be used as weights. Effects of second degree terms of weather variables are also studied. In all, we have considered seven models.

Model I: The first model given by Agrawal *et al.* (1986) is

$$Y = a + b_0 \sum_{w=1}^{n} X_w + b_1 \sum_{w=1}^{n} r_{xy(w)} X_w + b_2 \sum_{w=1}^{n} r_{xy(w)}^2 X_w + cT + e$$

$$= \mathrm{a} + \mathrm{b}_0 \mathrm{Z}_0 + \mathrm{b}_1 \mathrm{Z}_1 + \mathrm{b}_2 \mathrm{Z}_2 + \mathrm{cT} + \mathrm{e} \qquad \text{.... (3.3.1.1)}$$

where Y is crop yield (q/ha), a, b_j (j=0, 1, 2) and c are model parameters, n is number of weeks up to the time of harvest, w is the week identification, X_w is the value of weather variable under study in the w^{th} week and $r_{xy(w)}$ is correlation coefficient between yield and weather variable in w^{th} week. T is trend variable (Time index) and e is error terms assumed to follow independently normal distribution with mean zero and constant variance σ^2.

Model II: In fact, the model II is deduced from the model I by deleting the term b_2Z_2. The model becomes as follows.

$$Y = a + b_0Z_0 + b_1Z_1 + cT + e \qquad ....(3.3.1.2)$$

Model III & IV: These models are same as models I and II, respectively except that $r_{xy(w)}$ is obtained using yield adjusted for trend effect.

Model V: This is obtained by including quadratic terms of weather variables and correlation coefficients in model I, as such the model becomes.

$$Y = a + b_0 \sum_{w=1}^{n} X_w + b_1 \sum_{w=1}^{n} r_{xy(w)} X_w + b_2 \sum_{w=1}^{n} r^2_{xy(w)} X_w + b_{00} \sum_{w=1}^{n} X_w^2$$
$$+ b_{11} \sum_{w=1}^{n} r_{x^2y(w)} X_w{}^2 + b_{22} \sum_{w=1}^{n} r^2_{x^2y(w)} X_w^2 + cT + e$$

$$= a + b_0Z_0 + b_1Z_1 + b_2Z_2 + b_{00}Z_{00} + b_{11}Z_{11} + b_{22}Z_{22} + cT + e \ ...(3.3.1.3)$$

Model VI: This is obtained by deleting the quadratic terms b_2Z_2 and $b_{22}Z_{22}$ in model V, as the model becomes:

$$Y = a + b_0Z_0 + b_1Z_1 + b_{00}Z_{00} + b_{11}Z_{11} + cT + e \qquad ....(3.3.1.4)$$

Model VII & VIII: The models are same as models V and VI, respectively, except that correlation coefficients are obtained using adjusted yield for trend effect.

Data on relative humidity have been transformed into arc-sine root proportion as they were in percentage. Stepwise regression was used to select significant generated weather indices (independent variables in the models).

The effects on yield per unit change in weather variables in w^{th} week have been calculated by differentiating the models with respect to X_w.

Forecast model and time of forecast

The following model including single and various combinations of weather variables to exhibit their interaction effects on crop yield was fitted using partial crop season data.

$$Y = a + \sum_{i=1}^{p}\sum_{j=0}^{1} b_{ij} Z_{ij} + \sum_{i \neq i'=1}^{p}\sum_{j=0}^{1} b_{ii'j} Z_{ii'j} + cT + e \qquad ....(3.3.1.5)$$

where $Z_{ij} = \sum_{w=1}^{m} r_{i\,w}^{j}$

$$Z_{ii'j} = \sum_{w=1}^{m} r_{ii'w}^{j}$$

Y is yield, p is number of weather variables used, m is number of weeks considered for the model building, $r_{iw} / r_{ii'w}$ are the correlation coefficients of yield and the i^{th} weather variable (X_i) the product of the two weather variables (X_i and $X_{i'}$) in w^{th} week $(i, i' = 1, 2, ...6)$ T is time trend variable. a, b_{ij}, $b_{ii',j}$ and c are model parameters. e are error terms assumed to follow independently normal distribution with mean zero and variance σ^2. The model has been fitted for different values of m (m=13, 14, ..., 17). The data after 15th week have not been used as the idea was to forecast yield well in advance of harvest.

3.3.2 Development of statistical forecast models applying discriminant function analysis of data on weather variables

The primary objective of this technique is to develop a suitable statistical methodology for obtaining pre-harvest forecast models for yield of Ayodhya crop on the basis of discriminant function analysis. For achieving this, relationships between meteorological data and corresponding de-trended yield data have been established by developing regression models considering discriminant scores as the regressors along with trend.

Discriminant function analysis and Discriminant scores

Discriminant analysis is a multivariate technique concerned with separating distinct sets of objects (or sets of observations) and allocating new objects (or observations) to the previously defined groups. As a classificatory model, it is often employed in order to investigate observed differences when causal relationships are not well understood.

The steps of discriminant function analysis are

1. To develop an equation or a function using variables under consideration for computing a new variable or index that will parsimoniously represent the differences between various groups.
2. Use of discriminant function to classify future observations into any of the pre-defined groups.

The technique of discriminant function analysis is used to identify an appropriate function that discriminates best between set of observations from two or more groups and classifying the function observations into one of the previously defined groups. This technique is a multivariate technique viz. discriminant function analysis discussed in many standard books, to mention a few, Anderson (1984), Hair *et al*. (1995), Sharma (1996), Johnson and Wichern (2006) etc. However, few conceptual aspects of technique is given below.

Discriminant function in case of two groups

Consider a linear function of the form

$$DF = \sum_{i=1}^{p} l_i X_i \qquad \text{....(3.3.2.1)}$$

where X_i is the ith weather variable used to discriminate the groups and l_i is the corresponding discriminant coefficient, p being the number of weather variables. DF is a scalar quantity which is generally designated as discriminant score based on X_i*'s* and l_i*'s*.

The optimal value for the weighing coefficient, l_i, is determined in such a way that the variation between the two groups gets maximized relative to the variation within the groups. Maximizing the ratio yields simultaneous equations of the following form

$$\begin{aligned}
&l_1\sum x_1^2 + l_2\sum x_1x_2 + \ldots\ldots\ldots + l_p\sum x_1x_p = d_1\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&l_1\sum x_1x_p + l_2\sum x_1x_p + \ldots\ldots\ldots + l_p\sum x_p^2 = d_p
\end{aligned}$$

where x_1, $x_{2,\ldots,}$ x_p are deviations from their respective group means and $d_i = \overline{X}_{i1} - \overline{X}_{i2}$ for i=1,2,...,p. $\overline{X}_{i1}$ and $\overline{X}_{i2}$ are means of ith variable in two population. In matrix notation

$$\boldsymbol{S\,l} = \boldsymbol{d}$$

where S is the discriminant dispersion matrix given by

$$\mathbf{S}=\begin{pmatrix} \Sigma x_1^2 & \Sigma x_1x_2 & \ldots\ldots\ldots & \Sigma x_1x_p \\ & \Sigma x_2^2 & \ldots\ldots\ldots & \Sigma x_1x_p \\ \ldots\ldots & \ldots\ldots & \ldots\ldots\ldots & \ldots\ldots \\ & & & \Sigma x_p^2 \end{pmatrix} = (S_{ij})$$

where S_{ij} is the value corresponding to the ith and jth variables in matrix **S.** $\boldsymbol{l}$ and $\boldsymbol{d}$ are column vectors of discriminant coefficients and differences in group means respectively. The value of $\boldsymbol{l}$ can be obtained as

$$\boldsymbol{l} = \boldsymbol{S}^{-1}\boldsymbol{d}$$

The adequacy of the function is tested with the help of F test as

$$F_{(p, n_1+n_2-p-1)} = \frac{n_1 n_2 (n_1+n_2-p-1)}{(n_1+n_2)p} D^2 \qquad ....(3.3.2.2)$$

where, $D^2 = \sum_{i=1}^{p} l_i d_i$ (known as Mahalanobis D^2 statistics), d_i is the difference in means i.e. $\bar{X}_{i1} - \bar{X}_{i2}$ for the ith variable, n_1, n_2 are the number of units in the two groups and $\bar{X}_{i1}$ & $\bar{X}_{i2}$ are means of the ith variable for the two groups.

The discriminatory point S_c for classifying individual years into two groups is calculated as

$$S_c = \frac{\left(n_1\bar{S}_1 + n_2\bar{S}_2\right)}{\left(n_1 + n_2\right)} \qquad ....(3.3.2.3)$$

where $S_1 = \sum_{i=1}^{p} l_i X_{i1}$ and $S_2 = \sum_{i=1}^{p} l_i X_{i2}$ and $\bar{S}_1$ & $\bar{S}_2$ are the corresponding group cancroids. 1 and 2 denote the groups.

Fisher's discriminant function in the case of several groups:

Consider a linear function of the form

$$DF = \sum_{i=1}^{p} l_i X_i = l'X \qquad ....(3.3.2.4)$$

where *DF* is discriminant function, $\boldsymbol{l'} = \left(l_1, l_2, ..., l_p\right)$, $\mathbf{X'}$ = (X_1, X_2, ...,X_p), X_i is the ith weather variables under interest used to discriminate the groups and l_i is the corresponding discriminant coefficient, p is the number of weather variables. Consider that there are g groups. Assume that variance-covariance matrix $\sum_j (j=1,2,....,k)$ are same i.e. $\Sigma_1 = \Sigma_2 = ... = \Sigma_k = \Sigma$.

Let n_j be the size of j^{th} group (j = 1,….., k) and X_{ijm} be the m^{th} observation of i^{th} variable in j^{th} group (m=1, 2, …, n_j). Then mean of j^{th} group for i^{th} variable is $\overline{X}_{ij} = \frac{1}{n_j} \sum_{m=1}^{n_j} X_{ijm}$ and overall average for i^{th} variable is given by

$$\overline{X}_i = \frac{\sum_{j=1}^{k} n_j \overline{X}_{ij}}{\sum_{j=1}^{k} n_j} = \frac{\sum_{j=1}^{k}\sum_{m=1}^{n_j} X_{ijm}}{\sum_{j=1}^{k} n_j}$$

Let $\overline{X}' = \overline{X}_1, \overline{X}_2, ...\overline{X}_p$ and $\overline{X}'_j = \left(\overline{X}_{1j}, \overline{X}_{2j}, ...\overline{X}_{pj}\right)$

Let $B = \sum_{j=1}^{k} \left(\overline{X}_j - \overline{X}\right)\left(\overline{X}_j - \overline{X}\right)'$ be between group matrix of sum of squares and cross products and let $A = \sum_{j=1}^{k}(n_j - 1)S_j = \sum_{j=1}^{k}\sum_{m=1}^{n_j}\left(X_{jm} - \overline{X}_j\right)\left(X_{jm} - \overline{X}_j\right)'$ be the pooled matrix of sum of squares and sum of products, where S_j is a matrix of sample sum of squares and cross product within j^{th} group. The matrices *A* and *B* are independent. Consequently $A/(n_1+n_2+ … +n_k-k) = S_{pooled}$ is an estimated of $\sum$.

The vector ***l*** is obtained by maximizing the ratio $\frac{l' B l}{l' A l}$ and substituting the value of ***l*** in equation (3.3.1.4) the discriminant function is obtained.

In terms of eigen-values and eigenvectors, the discriminant function will be as follows:

Let $\lambda_1, \lambda_2, \ldots, \lambda_s$ denote the $s \leq \min(g-1, p)$ non- zero eigen-values of $A^{-1}B$ and $e_1, e_2, \ldots, e_s$ be the corresponding eigenvectors (scaler so that $e' S_{pooled}\, e=1$).

Then vector of coefficients *l* that maximize the ratio

$$\frac{l'Bl}{l'Al} = \frac{l'\left[\sum_{j=1}^{k}(\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})'\right]l}{l'\left[\sum_{j=1}^{k}\sum_{xn}^{m}(x_{jm} - \bar{x}_j)(x_{jm} - \bar{x})'\right]l}$$

is given by $\hat{l}_1 = \hat{e}_i$. The linear combination $\hat{l}_i'x$ is called the sample first discriminant function. The choice $\hat{l}_2 = \hat{e}_i$ produces the sample second discriminant function $\hat{l}_{2k}'$ .Continuing $\hat{l}_k'x = \hat{e}_k x$ is the sample kth discriminant function, $k \leq s$. For example, if g=3 and p=4, then minimum of (g-1, p)= minimum (2, 4), that is 2. Therefore, there will be 2 discriminant functions for three groups.

Determination of number of discriminant scores

Consider that observations are classified into k non-overlapping groups on the basis p variables. The technique identifies linear functions where the coefficients of the variables are determined in such a way that the variation between the groups gets maximized relative to the variation within the groups. The maximum number of discriminant functions that can be obtained is equal to minimum of (k-1) and p. These functions are used to classify the observations into different groups.

Development of statistical forecast models

In order to apply discriminant function analysis for modeling yield using weather variables, crop years under consideration have been divided into three groups, namely adverse, normal and congenial on the basis of crop yield adjusted for trend effect. Data on weather variables in these three groups were used to develop linear discriminant functions and the discriminant scores were obtained for each year. These discriminant scores were used along with year index (trend variable) as regressors and crop yield as regressand in developing the forecast models. In the present study the number of groups is three and number of weather variables is six, therefore only two discriminant functions can be obtained which are sufficient for discriminating a crop years into either of the three groups.

Three groups of crop year s, viz. adverse, normal and congenial have been obtained as follows: Let $\bar{y}$ and s be the mean and standard deviation of the adjusted crop yields of n years. The adjusted crop yields less than or equal to $\bar{y} - s$ would form adverse group, the adjusted crop yields between $\bar{y} - s$ and $\bar{y} + s$ would form normal group and adjusted crop yields above or equal to $\bar{y} + s$ would form congenial group. The adjusted crop yields were assigned codes 1, 2 and 3 if they belong to adverse, normal and congenial groups, respectively.

It is also known that weather variables effect the crop differently during different phases of crop development. Its effect depends not only on its magnitude but also on its distribution pattern over the crop season. Therefore, using weekly weather data as such in developing the model process a problem as number of independent variables in the regression model would increase enormously. To solve this problem, following weather indices have been developed using the procedure of Agrawal *et al*. (1983, 1986).

$$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j}$$

$$Z_{ii',j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=1}^{n} r_{ii'w}^{j} \text{ , j=0,1 and i=1,2,…,p.} \quad (3.3.2.5)$$

where Z_{ij} is un-weighted (for j=0) and weighted (for j=1) weather indices for i^{th} weather variable and $Z_{ii',j}$ is the un-weighted (for j=0) and weighted (for j=1) weather indices for interaction between i^{th} and i'^{th} weather variables. $\boldsymbol{X_{iw}}$ is the value of the i^{th} weather variable in w^{th} week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with i^{th} weather variable/product of i^{th} and i'^{th} weather variable in w^{th} week, n is the number of weeks considered in developing the indices and p is number of weather variables. Here, p=7 and n=15, i.e. 15 weeks data from 44th week to 52nd week of a year and 1st week to 6th week of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. Here only the first 24 years data from 1990-91 to 2016-17 have been utilized for model fitting and remaining three years were left for the validation of the model. The growth process of the crop has various phases and weeks within phases. In the development of pre-harvest model based on discriminant function analysis the entire 15 weeks data from 44th SMW to 52nd SMW of a year and 1st SMW to 6th SMW of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. In all 56 indices (28 weighted and 28 un-weighted) consisting of 7 weighted and 21 weighted interaction weather indices and 7 un-weighted and 21 un-weighted interaction weather indices have been constructed. For quantitative forecasting, linear regression models are fitted by taking the discriminant scores and the trend variable as the regressors and crop yield as the regressand. The following models are considered.

Model-A:

This model is based on the method given by Agrawal *et al.* (1986) for developing forecast model using weather indices. This is not based on discriminant function analysis. In this model the entire 15 weeks data from 44th SMW to 52nd SMW of a year and 1st SMW to 6th SMW of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions following the formula given equation (3.3.2.5). In all 56 weather indices (28 weighted and 28 un-weighted) consisting of 7 weighted weather indices and 21 weighted interaction indices, and 7 un-weighted weather indices and 21 un-weighted interaction indices have been obtained. Considering these 56 indices and trend variable (T) as regressors and yield as dependent variable, forecast model has been developed. The model fitted is

$$Y = a_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'-1}^{p} \sum_{j=0}^{1} a_{ii'j}\, z_{ii'j} + cT + \varepsilon$$

where $$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j} \qquad j = 0,1$$

$$Z_{ii'j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=1}^{n} r_{ii'w}^{j}$$

Y is the original crop yield, $\boldsymbol{X_{iw}}$ is the value of the i^{th} weather variable in w^{th} week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with i^{th} weather variable/product of i^{th} and i'^{th} weather variable in w^{th} week, n is number of weeks considered in developing the indices respectively and p is number of weather variables used. a_0, a_{ij}, $a_{ii,j}$ and c are the model parameters. T is trend variable (T=1,2,3,……,n) and

ε is error term assumed to follow independently normal distribution with mean zero and variance σ^2. The step-wise regression analysis was employed to develop the forecast model.

Model-D$_1$:

This model is the 2nd model of Agrawal *et al.* (2012). This model utilizes the complete data over 15 weeks and also considers relative importance of weather variables in different weeks. Using seven weighted weather indices of seven weather variables as discriminating variables, discriminant function analysis has been carried out and two discriminant functions have been obtained. Two sets of discriminant scores for the years under consideration from these two discriminant functions were obtained. For developing forecast model, these two sets of discriminant scores along with the trend variable were utilized as the regressors and the yield as the regressand. The form of model considered is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where βo = intercept of the model,

y = crop yield

$\beta_{i's}$ $(i = 1,2,3)$ = the regression coefficients

ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3,……,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D_2:

This model is 4^{th} model of Agrawal *et al.* (2012).Two discriminant functions and therefrom two sets of discriminant scores have been obtained using the first week data (44^{th} SMW) on seven weather variables. Next, two sets of discriminant scores obtained from first week and the second week (45^{st} SMW) data on seven weather variables data have been used as discriminating variables, so in all there were 9 discriminating variables, and based on these 9 discriminating variables the discriminant function analysis has been done and, therefore, two sets of discriminant scores have been obtained. This process was repeated up to the last week till the time of forecast (6^{th} SMW or 15^{st} week) and finally two sets of discriminant scores have been obtained. Based on these two sets of discriminant scores, the forecasting model taking yield as the regressand and the discriminant scores and the trend variable as the regressor variables has been fitted. The form of model considered is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where y= crop yield

βo = intercept of the model

$\beta_{i's}$ $(i=1,2,3)$= the regression coefficients

ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3,……,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D3:

In this procedure, all 56 (weighted and un-weighted including interaction indices) have been used as discriminating variables in discriminant function analysis and two sets of discriminant scores from two discriminant functions have been obtained. Forecasting model has been fitted taking un-trended yield as the regressand variable and the two sets of discriminant scores and the trend variable (T) as the regressor variables. The form of the model fitted is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where y = crop yield

βo = intercept of the model

$\beta_{i's}$ $(i=1,2,3)$= the regression coefficients

ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3,……,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D4:

In this procedure, 7 weighted and 7 un-weighted weather indices have been used as discriminating variables. Now, based on these 14 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The fitted model here is

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where y= crop yield

βo = intercept of the model

$\beta_{i's}$ $(i=1,2,3)$= the regression coefficients

ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3,……,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_5$:

In this procedure, discriminant function analysis has been carried out using the un-weighted and weighted average (weather indices) for the first weather variable (here, discriminating factors will be only two). Using the two sets of discriminant scores on the basis of first weather variable, and un-weighted and weighted average (weather indices) for the second weather variable, discriminant function analysis has been further carried out (here, the discriminating factors will be 4). This process is continued up to seven weather variables, and finally we get two sets of discriminant scores ds_1 and ds_2. Using crop-yield as regressand and discriminant scores ds_1 & ds_2 and the time trend T as regressor variables, the following model is fitted for the development of forecast model.

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where y= crop yield

βo = intercept of the model

$\beta_{i's}$ $(i=1,2,3)$= the regression coefficients

ds_1 and ds_2 are the two discriminant scores

T is trend variable (T=1,2,3,……,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_6$:

In this procedure, discriminant function analysis have been carried out using weekly data of the first weather variable spread over 15 weeks as discriminating variable. Using two sets of discriminant scores obtained from two estimated discriminant functions based on data of the first weather variable and 15 weeks data of the second variable, discriminant function analysis has been again performed and two sets of discriminant scores are obtained (here discriminating variable will now become 17). Using these two sets of discriminant scores and 15 weeks data of third weather variable have been again used to carry out discriminant function analysis and subsequently two sets of discriminant scores have been obtained. This process is continued up to seven weather variables, and ultimately we get two sets of discriminant scores ds_1 and ds_2. These two sets of scores and the trend variable (T) as the regressor variables and crop-yield as the regressand were utilized to develop fitting forecast model by the following model.

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where y= crop yield

βo = intercept of the model

$\beta_{i's}$ $(i=1,2,3)$= the regression coefficients

ds_1 and ds_2 are the two discriminant scores

T is trend variable (T=1,2,3,…… ,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

3.3.3 Development of statistical forecast models applying principal component analysis

It may be possible that the auxiliary variables used in the regression model at district level may be highly correlated and may valet the final results. In such situations, principal component analysis can be performed and first few principal components may be used as the independent variables in the regression model at the district level for the development of pre-harvest forecast model. We first briefly describe the principal component analysis.

Principal component analysis

Principal component analysis primarily deals with explaining the variance and covariance structure through a few linear combinations of original variables. The objectives are (1) data reduction (2) interpretation. In fact, Principal Component Analysis is more of a means to an end rather that an end in itself because this frequently serves as intermediate steps in much large investigation by reducing the dimensionality of the problem and providing easier interpretation. It is a mathematical technique, which does not require user to specify the statistical model or assumption about distribution of original variables. It may also be mentioned that principal components are artificial variables and often it is not possible to assign physical meaning to them. Further, since principal component analysis transforms original set of variables to new set of uncorrelated variables, it is worth stressing that if original variables are uncorrelated, and then there is no point in carrying out principal component analysis.

Let the random vector $X' = (X_1, X_2, ..., X_p)$ have the variance - covariance matrix $\sum$ with eigen values $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$. These p components variables are required to reproduce the total system of variability. But, quite often this variability can be accounted for by a small number of k principal components (k<p). The k principal component can then replace the initial p variables, and the original data

set, consisting n measurements on p variables is reduced to one consisting of n measurement on k principal component (k<p).

Consider the linear combinations

$$\begin{aligned} PC_1 &= \ell'_1 X = \ell_{11} X_1 + \ell_{21} X_2 + \ldots + \ell_{p1} X_p \\ PC_2 &= \ell'_2 X = \ell_{12} X_1 + \ell_{22} X_2 + \ldots + \ell_{p2} X_p \\ &\vdots \\ PC_p &= \ell'_p X = \ell_{1p} X_1 + \ell_{2p} X_2 + \ldots + \ell_{pp} X_p \end{aligned} \qquad \ldots(3.3.3.1)$$

We also have

$$Var(PC_i) = \ell'_i \sum \ell_i \qquad i = 1, 2, \ldots, p \qquad \ldots(3.3.3.2)$$

$$Cov(PC_i, PC_k) = \ell'_i \sum \ell_k \qquad i, k = 1, 2, \ldots, p \qquad \ldots(3.3.3.3)$$

The principal components are uncorrelated linear combinations $PC_1, PC_2, \ldots, PC_p$, whose variances in (3.3.3.2) are as large as possible.

The first principal component is the linear combination with maximum variance. i.e, it maximizes $Var(PC_1) = \ell'_1 \sum \ell_1$. It is clear that $Var(PC_1) = \ell'_1 \sum \ell_1$ can be increased by multiplying any ℓ_1 by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length. Therefore, we can define

First principal component	=	linear combination $\ell'_1 X$ that maximizes $Var(\ell'_1 X)$ subject to $\ell'_1 \ell_1 = 1$
Second principal component	=	linear combination $\ell'_2 X$ that maximizes $Var(\ell'_2 X)$ subject to $\ell'_2 \ell_2 = 1$ and $Cov(\ell'_1 X, \ell'_2 X) = 0$

i^{th} principal component	=	linear combination $\ell'_i X$ that maximizes $Var(\ell'_i X)$ subject to $\ell'_i \ell_i = 1$ and $Cov(\ell'_i X, \ell'_k X) = 0$

Let e_1, e_2,…, e_p be the Eigen-vectors corresponding to Eigen-values $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$ of variance-covariance matrix $\sum$ $X'=(X_1, X_2, \ldots, X_p)$. The i^{th} principal component is given by

$$PC_i=e'_1X= e_{1i}X_1+e_{2i}X_2+ \ldots + e_{pi}X_p; \; i= 1, 2, \ldots, p \qquad \ldots(3.3.3.4)$$

with the variance of PC_i given by

$$Var(PC_i) = e'_i \sum e_i = \lambda_i, \; i =1, 2, \ldots, p \qquad \ldots(3.3.3.5)$$

and

$$Cov(PC_i, PC_k) = e'_i \sum e_k = 0, \; i \neq k= 1, 2, \ldots, p \qquad \ldots(3.3.3.6)$$

If some λ_i are equal, the choices of the corresponding coefficient vectors e_i, and hence PC_i, are not unique. The principal components are uncorrelated and have variance equal to eigen-values of $\sum$.

We have other important results. Let σ_i^2 be the variance of X_i, i=1,2,…, P., then we have

$$\sigma_{1i}^2 + \sigma_{2i}^2 + \cdots + \sigma_p^2 = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} Var(Y_i) \qquad \ldots(3.3.3.7)$$

That is the total population variance $\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2 = \lambda_1 + \lambda_2 + \cdots + \lambda_p$ and consequently, the proportion of total variance due to (explained by) the k^{th} principal component is

$$\begin{pmatrix} \text{Proportion of total} \\ \text{population variance} \\ \text{due to k}^{\text{th}} \text{ principal} \\ \text{component} \end{pmatrix} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \ k = 1, 2, \ldots, p \qquad \ldots(3.3.3.8)$$

If most (for instance, 80 to 90%) of the total population variance, for large p, can be attributed to the first k (k<p) components, then these components can 'replace' the original p variables without much loss of information.

Each component of the coefficient vector $e'_i=[e_{i1}, \ldots, e_{ki}, \ldots, e_{pi}]$ also merits inspection. The magnitude of e_{ki} measures the importance of the k^{th} variable to the i^{th} principal component, irrespective of the other variables. In particular, e_{ki} is proportional to the correlation coefficient between Y_i and X_k.

So we have another important results that if $Y_1=e'_1X$, $Y_2=e'_2X, \ldots$, $Y_p= e'_2X$ are the principal components obtained from the covariance matrix $\sum$, then

$$\rho_{Y_i X_k} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \ ; i, k = 1, 2, \ldots, p \qquad \ldots(3.3.3.9)$$

are the correlation coefficients between the components Y_i and the variables X_k. Here (λ_1, e_1), $(\lambda_2, e_2), \ldots$, (λ_p, e_p), are the eigen values- eigen vector pairs for $\sum$.

Proposed models for pre- harvest forecast of crop yield

Let PC_1, PC_2, ..., PC_k be first k (k< p) principal components explaining variability up to 90 percent. Using these k principal components as regressor variables and crop yield (y) as regressand, the following pre-harvest forecast models are proposed.

Model-P_1:

In this procedure, 7 un-weighted weather indices of Seven weather variables have been used as p=7 variables in the principal component analysis. Suppose that principal component analysis has identified some first k principal components as most significant ones as per loading have explained more than about 75 percent of the total variance. Therefore, these first k principal components have been used along with the trend variable (T) were utilized as the regressors and the yield as the regressand in multiple linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where Y is the crop yield, $\beta_i^{'s} (i = 0,1,2,.....k)$ and δ are model parameter, PC_1, PC_2,......PC_k are principal components, T is the trend variable and e is error term assumed to follow normal distribution with mean zero and variance σ^2.

Model-P_2:

In this model, 7 weighted weather indices of seven weather variables have been used in principal component analysis. The principal component analysis has identified first k principal component as more significant once as per loading have explained more than 75 percent of the total variance. Therefore, these first k principal components have been used along with the trend variable (T) were utilized as the regressors and the yield as the regressand in multiple linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where the notations stand as usual as described in model-1.

Model-P_3:

In this procedure, all 56 weather indices (28 weighted and 28 un-weighted) consisting of 7 weighted weather indices and 21 weighted interaction indices, 7 un-weighted weather indices and 21 un-weighted interaction indices have been used as 56 variable (p=56) in principal component analysis. If first k (k<p) principal components as most significant once as per loading have explained more than about 75 percent of the total variance, then these first k principal components have been used along with the trend variable (T) were utilized as the regressors

and the yield as the regressand in multiple linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where the notations stand as usual as described in model-1.

Model-P_4:

In this model, 7 weighted and 7 un-weighted weather indices of seven weather variables have been used as p=14 variables in the principal component analysis. Suppose that principal component analysis has identified some first k principal components as more significant ones as per loading have explained more than about 75 percent of the total variance. Therefore, these first k principal components have been used along with the trend variable (T) were utilized as the regressors and the yield as the regressand in multiple linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where the notations stand as usual as described in model-1.

Model-P_5:

In this model, 7 un-weighted and 21 un-weighted interactions weather indices of seven weather variables have been used in principal component analysis. Suppose that principal component analysis has identified some first k principal components as more significant once as per loading have explained more than 75 percent of the total variance. Therefore, these first k principal components have been used along with the trend variable (T) were utilized as the regressors and the yield as the regressand in multip le linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where the notations stand as usual as described in model-1.

Model-P_6:

In this model, 7 weighted and 21 weighted interactions weather indices of seven weather variables have been used in principal component analysis. Suppose that principal component analysis has identified some first k principal components as more significant once as per loading have explained more than 75 percent of the total variance. Therefore, these first k principal components have been used along with the trend variable (T) were utilized as the regressors and the yield as the regressand in multiple linear regression model. The form of model considered is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k + \delta T + e$$

where the notations stand as usual as described in model-1.

3.4 Measures for the comparison and validation of different models

Different models have been used in the present study for the comparison and the validation of the models developed. These models are given below:

R^2 (Coefficient of Determination):

It is in generally used for checking the adequacy of the model. R^2 is given by the following formula;

$$R^2 = 1 - \frac{ss_{res}}{ss_t}$$

where ss_{res} and ss_t are the residual sum of square and the total sum of square respectively.

R^2 never decreases when a regressor is added to the model, regardless of the value of the contribution of the variable in the model. Therefore, it is difficult to judge whether an increase in R^2 is really important. So, it is preferable to use Adjusted R^2 when models to be compared are based on different number of regressors. Adjusted R^2 is given by the following formula

$$R^2_{adj} = 1 - \frac{ss_{res}/(n-p)}{ss_t/(n-1)}$$

where n is the number of observation and p is the number of regressor variables. The total mean square is constant regardless of how many variables are in the model. On adding a regressor in the model Adjusted R^2 increases only if the addition of the regressor reduces the residual mean square. It also penalizes for adding terms that are not helpful, so it is very important in evaluating and comparing the regression models.

Percent Deviation:

This measures the deviation (in percentage) of forecast from the actual yield data. The formula for calculating the percent deviation of forecast is given below:

$$\textit{Percentage deviation} = \frac{\text{(actual yield- forecasted yield)}}{\text{(actual yield)}} \times 100$$

Percent Standard Error of the Forecast (CV):

Let $\hat{y}_f$ be forecast value of crop yield and X_0 be the vector of selected values for regressor variables for the yield is forecasted.

The variance of $\hat{y}_f$ as given in (Draper and Smith, 1998) is obtained as

$$V(\hat{y}_f) = \hat{\sigma} X_0{}' (X'X)^{-1} X_0$$

Where $X'X$ is the dispersion matrix of the sum of square and cross products of regressor variables used for the fitting the model and $\hat{\sigma}^2$ is the estimated residual variance.

The percent standard error (PSE) of forecast yield $\hat{y}_f$ is given by

$$\text{PSE} = \frac{\sqrt{V(\hat{y}_f)}}{\text{Forecast yield}} \times 100$$

Infact, the PSE is the coefficient of variation (C.V.) of forecast yield.

Root Mean Square Error (RMSE):

It is also a measure of comparing two models. The formula of RMSE is given below

$$\text{RMSE} = \left[\left\{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2\right\}\right]^{\frac{1}{2}}$$

where O_i and the E_i are the observed (actual) and forecasted value of the crop yield respectively and n is the number of years for which forecasting has been done.

Chapter-IV

RESULTS AND DISCUSSION

This chapter deals with results, salient findings and discussion of the study undertaken. Various pre-harvest forecast models as described in the preceding chapter have been developed. The results, salient findings and discussion are presented in the following order:

4.1 Statistical models for studying the relationship between weather variables and crop yield.

4.2 Pre-harvest forecast models using discriminant function analysis of weekly data of weather variables.

4.3 Pre-harvest forecast models using principal component analysis of weekly data of weather variables.

4.1 Statistical models for studying the relationship between weather variables and crop yield

We have studied eight models to explore best possible relationship between crop yield and individual weather variable. The fitted models by step-wise regression analysis for each weather variables along with values of R^2/ adj R^2 are given in Appendix-1. A summary of fitted models in terms of the values of the coefficient of determination (R^2) for the eight models studied are presented in the Table 4.1.1. Perusal of the results of the Table-4.1.1 indicate that (i) models using correlations based on yield adjusted for trend effect were better than the once using simple correlations, (ii) inclusion of quadratic terms of weather variables and also second power of correlation coefficient did not improve the model in general. However, the model V & VI have come out to be the exactly same after fitting them with data using step-wise regression analysis. We have chosen the model- V to study effects of weather variables on wheat yield.

Table 4.1.1: Coefficient of determination (R^2) under different models

Weather Variables	**Model**							
	I	II	III	IV	V	VI	VII	VIII
Minimum Temperature	0.54 (1)	0.54 (1)	0.44 (1)	0.44 (1)	0.62 (1)	0.62 (1)	0.44 (2)	0.44 (2)
Maximum Temperature	0.54 (1)	0.54 (1)	0.38 (1)	0.38 (1)	0.54 (1)	0.54 (1)	0.38 (3)	0.38 (3)
Relative Humidity at 7 hour	0.53 (3)	0.53 (3)	0.38 (3)	0.38 (3)	0.55 (2)	0.55 (2)	0.38 (2)	0.38 (2)
Relative Humidity at 14 hour	0.61 (1)	0.61 (1)	0.35 (3)	0.35 (3)	0.69 (2)	0.69 (2)	0.51 (2)	0.51 (2)
Rainfall	0.66 (1)	0.66 (1)	0.51 (1)	0.51 (1)	0.63 (1)	0.63 (1)	0.51 (2)	0.51 (2)
Rainy Day	0.63 (1)	0.63 (1)	0.47 (1)	0.47 (1)	0.69 (1)	0.69 (1)	0.50 (2)	0.50 (2)
Wind Velocity	0.37 (2)	0.37 (2)	0.37 (2)	0.37 (2)	0.42 (2)	0.42 (2)	0.37 (3)	0.37 (3)

NB: The figures in parenthesis indicate the number of explicative variables included in the models.

4.1.1 Effect of climatic variables on Wheat yield

The effects of one unit increase in weather variables over the average yield at different growth stages of the crop have been computed and are given

Table 4.1.2. The effects of one unit decrease below the average can be obtained by reversing the vertical scale.

4.1.1.1 Effect of minimum temperature

The multiple regression equation (Model-V) obtained for minimum temperature is given below.

$$Y = 37.03 + 0.77Z_1 \ (R^2=0.62)$$

The effects were obtained from

$$\frac{\partial Y}{\partial X_{(w)}} = 0.77r$$

It can be observed from the Table 4.1.2 that during preparation, sowing, emergence, and initial growth stages, the effect of 1^0c above the average have been found to be detrimental but 4th week has no effect. However, during milking and dough formation stages, the effects were fluctuating. The effects have been found to be beneficial during ripening and harvesting stages.

4.1.1.2 Effect of maximum temperature

The multiple regression equation (Model-V) obtained for maximum temperature is given below.

$$Y = 34.05 + 0.463Z_1 \ (R^2=0.54^{**})$$

The effects were obtained from

$$\frac{\partial Y}{\partial X_{(w)}} = 0.463r$$

During germination, preparation, sowing, emergence and initial growth stages, the effect of 1^0c above the average has been found to be detrimental. However, during vegetative growth stage, the effects were beneficial in general. The effects have been found to be beneficial during milking stage but during reproduction and dough formation stages, the effects were detrimental in general.

4.1.1.3 Effect of relative humidity (7 hr)

The multiple regression equation (Model-V) obtained for relative humidity (7 hr) is given below.

$$Y = 17.57 + 0.002Z_{11} \ (R^2=0.55^{**})$$

The effects were obtained from

$$\frac{\partial Y}{\partial X_{(w)}} = 0.002 \times 2r_{x^2y(W)X_W}$$

The effects were in general beneficial on the Ayodhya yield throughout the crop growth period except in the harvesting stage.. Hence, arise in humidity at 7 hr during maximum vegetative growth stage to ears stage could be beneficial to the Wheat yield but the effect could be detrimental during last five weeks i.e. harvesting stage.

4.1.1.4 Effect of relative humidity (14 hr)

The multiple regression equation (Model-V) obtained for relative humidity (14 hr) is given below.

$$Y = -3.04 + 1.47Z_1 + 0.016Z_{11} \ (R^2=0.69^{*})$$

The effects were obtained from

$$\frac{\partial Y}{\partial X_{(w)}} = 1.47 r_{xy(w)} + 0.016 * 2 r_{x2\,y(w)} X_W$$

The effects were in general detrimental on the Wheat yield throughout the crop growth period except in the harvesting stage. Hence, arise in humidity at 14 hr during harvesting stage could be small beneficial to the Wheat yield. The effects were pronounced in 1st and 3rd weeks.

4.1.1.5 Effect of wind velocity

The multiple regression equation (Model-V) obtained for wind velocity is given below.

$Y = 20.44 + 0.255 Z_{11}$ (R^2=0.42**)

The effects were obtained from

$$\frac{\partial Y}{\partial X_{xy(w)}} = 0.255 \times 2 r_{x^2} y(w) X_w$$

One unit increase above the average weekly wind velocity has shown detrimental effect on the crop yield during entire period of Wheat production except in the vegetative growth stage.

4.1.1.6 Effect of Rainfall

The multiple regression equation (Model-V) obtained for rainfall is given below.

$Y = 22.65 + 0.216 Z_1$ (R^2=0.63**)

The effects were obtained from

$$\frac{\partial Y}{\partial X_{(w)}} = 0.216 r_{xy(w)}$$

The effect of increase of one unit in rainfall above the average weekly rainfall have been found to be detrimental in general on crop yield during entire period of wheat production except in reproduction, ears formation and ripening stages. Hence, rainfall during reproduction to ripening stages could be some beneficial to the Wheat yield.

4.1.1.6 Effect of Rainy Day

The multiple regression equation (Model-V) obtained for Rainy day is given below.

$$Y = 22.44 + 0.024 Z_0 - 1.099 Z_2 \quad (R^2=0.69^{**})$$

The effects were obtained from

$$\frac{\partial Y}{\partial X_{xy(w)}} = 0.024 - 1.099 r^2{}_{xy(w)}$$

The effect of increase of rainy days above the average weekly rainy days have been found to be detrimental in general on crop yield during entire period of Wheat production except in reproduction, ears formation and ripening stages. Hence, rainy days (mm) during reproduction to ripening stages could be beneficial to the Wheat yield.

Table 4.1.2 Percent change in yield per unit increase in weather variable over its average value

Phases	SMW	Week No.	Temp.		R.H.		Wind Velocity	Rainfall	rainyday
			Min.	Max.	7 hr	14 hr			
I	44	1	-0.528	-0.17	0.138	0.001	0.001	0.095	0.047
	45	2	-0.394	-0.08	0.025	0.003	0.054	-0.049	0.012
	46	3	0.051	0.01	0.024	0.011	0.000	0.112	0.066
	47	4	0.141	-0.21	0.030	-0.006	0.143	-0.088	0.041
	48	5	0.152	-0.25	0.001	-0.004	0.282	0.114	0.068
	49	6	0.437	-0.28	0.006	-0.024	0.676	0.068	0.019
II	50	7	0.459	-0.23	0.107	-0.031	0.165	-0.185	0.064
	51	8	-0.114	-0.26	0.000	0.002	0.722	-0.232	0.282
	52	9	-0.259	-0.03	0.025	-0.005	0.037	0.025	0.017
	1	10	-0.020	-0.22	0.074	-0.007	0.094	0.011	0.000
	2	11	-0.564	-0.04	0.012	0.016	0.900	-0.269	0.342
	3	12	0.170	0.39	0.043	0.026	0.651	-0.118	0.073
	4	13	-0.277	-0.04	0.019	0.001	0.132	0.054	0.164
III	5	14	-0.723	0.12	0.000	0.013	0.207	-0.392	0.527
	6	15	-0.705	0.08	0.004	0.014	0.001	0.113	0.237
	7	16	-0.028	-0.07	0.012	0.003	0.021	0.261	0.358
	8	17	0.241	0.23	0.002	0.011	0.745	-0.058	0.019
	9	18	-0.296	0.25	0.090	0.011	0.958	-0.379	0.755
	10	19	-0.015	0.11	0.016	0.009	0.001	-0.104	0.167
	11	20	-0.588	-0.15	0.055	0.010	0.000	0.135	0.006

4.1.2 Forecast models and time of forecast

The forecast model given in Sub Section- 3.3.1.2 has been fitted with the data by applying step-wise regression analysis for m=13, 14, 15, 16 and 17. The fitted models along with the values of the coefficient of determination (R^2/ R^2 adj) for the models at different points of time are given in Table 4.1.2.1.

Table 4.1.2.1 Details of forecast models fitted.

M	Model	R^2	R^2 adj
13	$Y = 14.982 - 0.191Z_{31} - 0.031Z_{11} - 0.005Z_{231} - .007Z_{241} - .006Z_{251} + .237T$ (6.889) (0.083) (0.007) (.002) *(1.002)* *(.003)* *(.031)*	88.8	85.4
14	$Y = 24.792 - .632Z_{11} - 0.007Z_{231} - 0.007Z_{241} + .246T$ (4.138) (0.119) (0.001) (0.002) (.030)	87.4	85.1
15	$Y = 2.144 - .285Z_{31} - 0.031Z_{121} - 0.011Z_{261} + 0.222T_{561}$ (4.86) (0.072) (0.006) (.003) (.032)	85.9	83.4
16	$Y = 40.306 + 0.232Z_{31} + 0.008Z_{141} + 0.008Z_{261}$ (5.348) (0.072) (0.002) (0.003)	74.5	71.1
17	$Y = 22.947 - 0.988Z_{11} + 0.007Z_{141} - 0.006Z_{231} - .009Z_{241} - .023Z_{371} + .260T$ (3.669) (0.191) (0.003) (0.001) (.001) (.008) (.023)	92.9	90.8

Perusal of the Table 4.1.2.1 reveals that (4th week of March) is the appropriate time for forecasting Wheat yield in Ayodhya district as there is no improvement in the value of R^2 by including data of later periods. Finally, the model fitted for m= 17 comes out to be

$$Y = 22.947 - 0.988Z_{11} + 0.007Z_{141} - 0.006Z_{231} - .009Z_{241} - .023Z_{371} + .260T \quad (R^2=0.92)$$

It is obvious from the above model that un-weighted and weighted weather indices of minimum temperature, weighted weather interaction of wind velocity and rainy days including time trend variable (T) have appeared as the important explicative variables.

Yield forecast for 2014-15, 2015-16 and 2016-17 have been computed using the fitted forecast model as stated above and are presented in the Table 4.1.2.2.Perusal of the results of the Table 4.1.2.2 indicate that forecast yields were relatively close to the actual yield. The percent deviation of forecast varied between 14.32 and 18.61. The percent standard errors of forecast yield were found to be reasonably low for first two years but it is high for 2016-17. Therefore, it can be concluded that a reliable forecast of Wheat yield can be made by the aforesaid model one and half months before the harvest.

Table 4.1.2.2 Comparison between actual and forecast yields along with statistical measures of Wheat in different years for M=24

Year	Actual Yield (q/ha)	Predicted Yield (q/ha)	Percent Deviation	PSE (CV)	RMSE
2014-15	20.600	26.219	27.278	14.320	6.516
2015-16	24.400	21.645	21.003	18.613	
2016-17	32.890	24.972	24.072	16.942	

4.2 Pre-harvest forecast models using discriminant function analysis of weekly data of weather variables

Statistical models for pre harvest forecast of the wheat yield in Ayodhya district of Eastern Uttar Pradesh have been developed on the basis of weekly data on weather variables *viz.*, minimum temperature, maximum temperature, relative humidity at 7 hour, relative humidity at 14 hour, rainfall, rainy days and wind velocity using discriminant function analysis.

Following the seven procedures described in Chapter-III, seven models have been developed. Sowing of wheat starts generally from the first week of November in Ayodhya district. Therefore, weekly data on the weather variables have been considered from pre-sowing the 44nd SMW of crop which fall during the first week of November. It has been proposed to make pre-harvest forecast of the wheat yield at the stage of milking / dough, about two months before the harvest. Milking and dough stages generally start after about 100 days of sowing. Therefore, 6th SMW of the next year (Feb.5-Feb.11) has been considered the week of pre-harvest forecast. Thus, in all 15 weeks data on the weather variables (44th SMW of the previous year to 6th SMW of the next year) have been utilized to develop the statistical models.

In order to carry out discriminant function analysis, the wheat yields are adjusted for trend effect. The crop years have been divided into three groups namely adverse, normal and congenial. The actual wheat yields, adjusted wheat yield and the groups indicated by 1, 2, and 3 as adverse, normal and congenial, respectively, are given in the Table 4.2.1.

Table 4.2.1 Actual and adjusted yield of wheat

Year	Actual yield (Q/ha)	Adjusted yield (Q/ha)	Groups
1990-91	20.34	27.28	2
1991-92	21.42	26.42	2
1992-93	23.00	25.06	2
1993-94	20.89	27.39	2
1994-95	23.44	25.07	2

1995-96	25.30	23.43	1
1996-97	25.02	23.93	1
1997-98	26.76	22.41	1
1998-99	23.97	25.42	2
1999-00	22.30	27.31	2
2000-01	26.26	23.57	1
2001-02	23.92	26.14	2
2002-03	21.79	28.49	2
2003-04	26.37	24.13	1
2004-05	23.77	26.95	2
2005-06	26.07	24.87	1
2006-07	26.83	24.34	1
2007-08	27.58	23.81	1
2008-09	28.1	23.51	1
2009-10	26.05	25.78	2
2010-11	28.14	23.91	1
2011-12	29.56	22.72	1
2012-13	29.63	22.87	1
2013-14	23.34	29.38	2
2014-15	20.6	32.34	3
2015-16	27.4	25.76	2
2016-17	32.89	20.50	1

The models have been developed by utilizing 27 years data of wheat yield (1990-91 to 2016-17) and remaining three years were left for the validation of the model. The models developed are described below.

Model –A

As already illustrated in Chapter-III that technique of discriminant function analysis has not been used in this model. Considering the actual wheat yield as regressand and 56 weather indices generated (given in Appendix-2) and time trend (T) as regressor variables, the model was fitted using step-wise regression analysis. The results are presented below in the Table 4.2.2.

Table 4.2.2 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.87**	0.998	34.9^{**}
2	Z_{10}	0.041**	0.233	
3	Z_{20}	0.027	0.160	
4	Z_{120}	0.058	0.359	
5	Trend	0.238**	0.062	

* P< 0.05, ** P<0.1

The Durbin-Watson result is non-significant according to tabulated value and the table value for n=24 and k=3 is,

Forecast model- A

$$\mathbf{Y = 21.87+0.041\ Z_{10} +0.027\ Z_{20} +0.058\ Z_{120} +0.238T}$$

where, Z_{10} = unweighted average of minimum temperature.

Z_{11} = weighted average of minimum temperature.

Z_{670} = weighted interaction between wind velocity and rainfall.

T = Time trend (1, 2,.......22)

These above three weather indices have been found significant variables for forecasting the pre-harvest wheat yield at 6th SMW of crop-production Milking/Dough Stage.

The model is validated by forecasting the wheat yield for the years 2014-15, 2015-16 and 2016-17. The results of validation are given in the Table-4.2.3. The values of per cent deviation of forecast yield from actual yield, % RMSE and %SE were also computed and are presented in the Table-4.2.3.

Table 4.2.3 Validation of the Model-A

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	31.44	37.4** (34.9**)	52.65	4.35	6.94
2015-16	27.40	26.98		1.51	5.62	
2016-17	32.89	27.70		15.75	5.93	

** P<0.1

It can be observed from the results of the Table-4.2.2 and Table-4.2.3 that finally entered regressor variables were significant. The value of adjusted R^2 is also quite high to the extent of 34.9%. The per cent deviation, %SE and RMSE are also quite low indicating thereby that the model is best fitted and it has high power to pre-harvest forecast wheat yield at Milking/dough stage, about two months before the harvest.

Model-D_1

The discriminant function analysis was carried out to find out the discriminant functions and discriminant scores. The estimated discriminant functions, The classification results based on estimated discriminant functions are given in Table 4.2.4.

Table 4.2.4 Classification Results

	CAT	**Predicted Group Membership**			**Total**
Original Count		**1**	**2**	**3**	
	1.00	8	3	2	13
	2.00	3	10	0	13
	3.00	0	0	1	1
%	1.00	61.5	23.1	15.4	100.0
	2.00	23.1	76.9	0	100.0
	3.00	.0	.0	100.0	100.0

(95.0% of original grouped cases correctly classified).

It is obvious from the results of the Table 4.2.4 from that 95.00 per cent of original grouped cases have been correctly classified indicating thereby that discriminant factors used in discriminant analysis are sufficient to classify the groups.

The discriminant score ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual Wheat yield as regressand for fitting the regression model (Model-D_1). The results of the fitted model are presented in the Table 4.2.5.

Table 4.2.5 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	22.119**	0.923	55.8^{**}
2	ds_1	0.671	0.327	
3	ds_2	0.203	0.423	
4	Trend	0.232	0.065	

* P< 0.05, ** P<0.1,

The Durbin-Watson result is non-significant according to tabulated value for n=24 and k=3.

Forecast model- D_1

$$\mathbf{Y = 22.119+0.671\ ds_1 +0.203\ ds_2+0.232\ T}$$

The model is validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in the Table 4.2.6.

Table 4.2.6 Validation of the model – D_1

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	26.28	61.5** (55.8**)	27.58	4.06	4.36
2015-16	27.40	28.43		3.77	3.57	
2016-17	32.89	28.00		14.86	3.09	

** P<0.1

Model-D_2

Following the procedure in Model-D_2 as described in Chapter-III, the discriminant function analysis was carried out. The estimated discriminant functions, discriminant scores and other relevant. The classification results based on estimated discriminant function are presented in the Table 4.2.7.

Tale 4.2.7 Classification Results

	CAT	**Predicted Group Membership**			**Total**
Original Count		**1**	**2**	**3**	
	1.00	10	2	1	13
	2.00	2	11	0	13
	3.00	0	0	1	1
%	1.00	76.9	15.4	7.7	100.0
	2.00	15.4	84.6	.0	100.0
	3.00	.0	.0	100.0	100.0

(100.0% of original grouped cases correctly classified).

In this case 100 percent of original grouped cases have been correctly classified.

The discriminant scores ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual Wheat yield as regressand for fitting the forecast model. The results of the fitted model are presented in the Table 4.2.8.

Table 4.2.8 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	22.242**	.737	65.4**
2	ds_1	-.816**	.241	
3	ds_2	0.166	.350	
4	Trend	0.219	.052	

* P< 0.05, ** P<0.1

The Durbin-Watson result is non-significant according to tabulated value for n=24 and k=3.

Forecast model- D_2

$$\mathbf{Y = 22.242 - 0.816 ds_1 + 0.166\, ds_2 + 0.219T}$$

The model is validated by forecasting the wheat yield for the years 2014-15, 2015-16 and 2016-17. The results of validation are given in the Table-4.2.9

Table 4.2.9 Validation of the model- D_2

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	26.11	69.9** (65.4**)	26.79	4.86	4.25
2015-16	27.40	28.08		2.50	3.32	
2016-17	32.89	28.06		14.67	2.87	

** P<0.1

Model-D_3:

Discriminant function analysis was carried out following the procedure in Model-D3 as described in Chapter-III. Two discriminant functions were obtained and using these functions two sets of discriminant score were computed for each year under consideration (1990-91 to 2016-17). The estimated discriminant functions, discriminant scores and other relevant results. The classification results obtained by estimated discriminant function are presented in Table-4.2.10.

Table 4.2.10 Classification Results

	CAT	**Predicted Group Membership**			**Total**
Original Count		**1**	**2**	**3**	
	1.00	12	1	0	13
	2.00	0	13	0	13
	3.00	0	0	1	1
%	1.00	92.3	7.7	.0	100.0
	2.00	.0	100.0	.0	100.0
	3.00	.0	.0	100.0	100.0

(100.0% of original grouped cases correctly classified).

It is obvious from the above table that 100 percent of original grouped cases have been correctly classified.

Forecast model was obtained by fitting regression model where discriminant scores ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual Wheat yield as regressand. The results of the fitted model are presented in the Table 4.2.11.

Table 4.2.11Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	22.654**	.613	78.2**
2	ds_1	0.680**	.122	
3	ds_2	0.022	.281	
4	Trend	0.181	.042	

* P< 0.05, ** P<0.1

The Durbin-Watson result is non-significant according to tabulated value for n=24 and k=3.

Forecast model- D_3

$Y = 22.654+0.680ds_1+0.022ds_2+0.1817T$

The Model-D_3 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results are presented in the Table 4.2.12.

Table 4.2.12 Validation of the model- D_3

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	28.77	81.0** (78.2**)	39.68	7.14	8.68
2015-16	27.40	29.44		7.46	2.32	
2016-17	32.89	26.16		20.43	2.25	

** P<0.1

Model-D_4

Following the procedure in Model–D4 as described in the Chapter-III, discriminant function analysis was carried out to find out discriminant functions and discriminant scores for each year under consideration. The two estimated discriminant functions, two sets of discriminant score (ds_1 and ds_2) and other relevant results.

The classification results based on the estimated discriminant functions are presented in Table-4.2.13.

Table 4.2.13 Classification Results

	CAT	**Predicted Group Membership**			**Total**
Original Count		**1**	**2**	**3**	
	1.00	61.5	23.1	15.4	100.0
	2.00	23.1	76.9	.0	100.0
	3.00	.0	.0	100.0	100.0
%	1.00	61.5	23.1	15.4	100.0
	2.00	23.1	76.9	.0	100.0
	3.00	.0	.0	100.0	100.0

(100% of original grouped cases correctly classified).

It can be seen from the results of the Table 4.2.17 that 100 percent of original grouped cases have been correctly classified.

The forecast model was obtained by fitting regression model using discriminant scores ds_1 and ds_2 and time trend as regressor variables and actual Wheat yield as regressand. The results of the fitted model are presented in the Table 4.2.14.

Table 4.2.14 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	22.118**	.924	55.8^{**}
2	ds_1	-.671	.327	
3	ds_2	-.204	.423	
4	Trend	0.232	.065	

* P< 0.05, ** P<0.1

The Durbin-Watson result is inconclusively according to tabulated value for n=24 and k=3.

Forecast model- D_4

$Y = 22.118 - 0.671\, ds_1 - 0.204\, ds_2 + 0.232\, T$

In order to validate the model, Wheat yields for the year 2014-15 to 2016-2017 were forecasted using the Model-D_4. The results are presented in the Table 4.2.15.

Table 4.2.15 Validation of the model- D_4

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	26.23	61.6** (55.8**)	27.35	4.07	4.36^{**}
2015-16	27.40	28.38		3.57	3.57	
2016-17	32.89	27.95		15.00	3.09	

** P<0.1

Model-D_5:

Two discriminant functions and two sets of discriminant scores (ds_1&ds_2) have been estimated by carrying out discriminant function analysis using the procedure in Model-D5 as described in Chapter–III . The estimated discriminant functions, discriminant scores and other relevant results. The classification results based on the estimated discriminant functions are presented in the Table-4.2.16

Table 4.2.16 Classification Results

	CAT	**Predicted Group Membership**			**Total**
Original Count		**1**	**2**	**3**	
	1.00	8	3	2	13
	2.00	3	10	0	13
	3.00	0	0	1	1
%	1.00	61.5	23.1	15.4	100.0
	2.00	23.1	76.9	.0	100.0
	3.00	.0	.0	100.0	100.0

(100.0% of original grouped cases correctly classified).

It is evident from the results of the above table that only 100.0 percent of original grouped cases have been correctly classified.

The regression model was fitted with discriminant scores ds_1 & ds_2 and time trend (T) as regressor variable and actual wheat yield as regressand. The results of the fitted model are presented in the Table 4.2.17.

Table 4.2.17 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	**22.118****	.924	55.8**
2	ds_1	-0.671	.327	
3	ds_2	-0.204	.423	
4	Trend	0.232	.065	

* P< 0.05, ** P<0.1

The Durbin-Watson result is inconclusively according to tabulated value for n=24 and k=3.

Forecast model- D_5

$$\mathbf{Y = 22.118 - 0.671\ ds_1 - 0.204\ ds_2 + 0.232\ T}$$

The forecast Model-5 has been validated by forecasting the wheat yield for the year 2014-15, 2015-16 and 2016-17. The results are presented in Table-4.2.18

Table 4.2.18 Validation of the model- D_5

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	26.23	61.6** (55.8**)	27.35	4.07	4.36**
2015-16	27.40	28.38		3.57	3.57	
2016-17	32.89	27.95		15.00	3.09	

** P<0.1

Model-D_6:

Two discriminant functions and two sets of discriminant scores (ds_1 & ds_2) have been obtained by carrying out discriminant function analysis using the procedure in Model-D_6 as described in Chapter –III. The classification results based on the estimated discriminant functions are presented in the Table 4.2.19.

Table 4.2.19 Classification Results

		CAT	Predicted Group Membership			Total
Original	**Count**		**1**	**2**	**3**	
		1.00	12	1	0	13
		2.00	0	13	0	13
		3.00	0	0	1	1
	%	1.00	92.3	7.7	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

(100.0% of original grouped cases correctly classified).

It is obvious from the above results that 100.0 percent of original grouped cases have been correctly classified.

The forecast model was obtained by fitting the regression model where discriminant scores ds_1 & ds_2 and time trend (T) were considered as regressor variables and actual Wheat yield as regressand. The result of the fitted model are presented in the Table 4.2.20.

Table 4.2.20 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	22.654***	.613	78.2^{**}
2	ds_1	0.680***	.122	
3	ds_2	0.022	.281	
4	Trend	0.181**	.042	

* P< 0.05, ** P<0.1

The Durbin-Watson result is inconclusively according to tabulated value for n=24 and k=3.

Forecast model- D_6

$$Y = 22.654 + 0.680ds_1 + 0.022\ ds_2 + 0.181\ T$$

The forecast Model-D_6 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results are presented in the Table 4.2.21.

Table 4.2.21 Validation of the model-D_6

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	22.60	28.77	81.0** (78.2**)	39.68	7.14	6.22
2015-16	27.40	29.44		7.46	2.80	
2016-17	32.89	26.16		20.43	2.25	

** P<0.1

Comparison of the models:

Table 4.2.22 show that the actual and forecast Wheat yield based on different forecast models for different years i.e. 2014-15, 2015-16 and 2016-17.The values of percent deviation of forecast from actual yield, PSE (CV), R^2, R^2 Adj. and RMSE are also presented in this table. The values of Adj. R^2 (%) and RMSE, the Model-D_6 (78.2%) besides Model-A has been found to be the best followed by the Model-D_2 (65.4%) and the Model-D_3 (78.2%). The average percent standard error (PSE) value of the Model-A is 5.3 while average PSE value of Model-D_2, D_3 and D_6 are 3.68, 3.90 and 4.06, respectively, which shows that these models are better for forecast. This indicates that forecast of Model-D_3, D_2 and D_6 have been most satisfactory among all the models in Ayodhya district of Uttar Pradesh. Hence, a reliable forecast of Wheat yield about one and half months before the harvest can be obtained from the Model- D_3, D_2 and D_6. The results of the Table 4.2.22 have also been graphically presented in Fig. 1 to highlight the forecasting power of the models.

Table 4.2.22 Estimate of important parameters for the comparison of models.

Model	Year	Actual yield (Q/ha)	Predicted yield (Q/ha)	Percent deviation	Percent Standard Error	R^2 (%)	Adjusted R^2 (%)	RMSE
A	2014-15	22.60	31.44	52.65	4.35	37.4	34.9	6.94
	2015-16	27.40	26.98	1.51	5.62			
	2016-17	32.89	27.70	15.75	5.93			
D_1	2014-15	22.60	26.28	27.58	4.06	61.5	55.8	4.36
	2015-16	27.40	28.43	3.77	3.57			
	2016-17	32.89	28.00	14.86	3.09			
D_2	2014-15	22.60	26.11	26.79	4.86	69.9	65.4	4.25
	2015-16	27.40	28.08	2.50	3.32			
	2016-17	32.89	28.06	14.67	2.87			
D_3	2014-15	22.60	28.77	39.68	7.14	81.0	78.2	8.68
	2015-16	27.40	29.44	7.46	2.32			
	2016-17	32.89	26.16	20.43	2.25			
D_4	2014-15	22.60	26.23	27.35	4.07	61.6	55.8	4.36
	2015-16	27.40	28.38	3.57	3.57			
	2016-17	32.89	27.95	15.00	3.09			
D_5	2014-15	22.60	26.23	27.35	4.07	61.60	55.8	4.36
	2015-16	27.40	28.38	3.57	3.57			
	2016-17	32.89	27.95	15.00	3.09			
D_6	2014-15	22.60	28.77	39.68	7.14	81.0	78.2	6.22
	2015-16	27.40	29.44	7.46	2.80			
	2016-17	32.89	26.16	20.43	2.25			

4.3 Pre-harvest forecast models using principal component analysis of weekly data of weather variables

Model-P_1:

In this procedure, 7 un-weighted weather indices of seven weather variables have been used in principal component analysis. The principal component analysis has identified first two principal components as more significant. Hence, these first three principal components have been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.1.

Table 4.3.1 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.731***	1.049	43.1
2	PC_1	.078	.430	
3	PC_2	-.047	.524	
4	Trend	.261	.075	

* P< 0.05, ** P<0.1

Forecast model- P_1

Y = $21.73 + 0.078\ PC_1 - 0.047\ PC_2 + 0.261T$

The forecast Model-P_1 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results are presented in the Table 4.3.2

Table 4.3.2 Validation of the model – P_1

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	28.21	50.5 (43.1)	36.96	3.63	5.05
2015-16	27.4	28.63		4.52	3.84	
2016-17	32.89	28.74		12.59	3.18	

** P<0.1

Model-P_2

In this procedure, 7 weighted weather indices of 7 weather variables have been used in principal component analysis. The principal component analysis has identified two principal component as more significant. Hence two principal component has been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.3.

Table 4.3.3 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.731***	1.049	43.1
2	PC_1	.078	.430	
3	PC_2	-.047	.524	
4	Trend	.261	.075	

* P< 0.05, ** P<0.1

Forecast model- P_2

$$Y = 21.731 + 0.078\,PC_1 - 0.047\,PC_2 + 0.261\,T$$

The forecast Model-P_2 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in Table 4.3.4.

Table 4.3.4 Validation of the model- P_2

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	28.08	50.5 (43.1)	36.31	3.64	4.97
2015-16	27.4	28.56		4.26	3.85	
2016-17	32.89	28.77		12.52	4.18	

** P<0.1

Model-P_3

In this procedure, all 56 weather indices of Seven weather variables have been used in principal component analysis. The principal component analysis has identified first seven principal components as most significant. Hence, these first seven principal components have been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.5.

Table 4.3.5 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.475	1.218	44.3^{**}
2	PC_1	.178	.418	
3	PC_2	-.119	.588	
4	PC_3	-.583	.421	
5	PC_4	.016	.461	
6	PC_5	.008	.470	
7	PC_6	.627	.393	
8	PC_7	-.342	.424	
	Trend	.277	.087	

* P< 0.05, ** P<0.1

Forecast model- P_3

Y = $21.47+0.178\ PC_1 - 0.119\ PC_2 - 0.583\ PC_3 + 0.016\ PC_4 + 0.008\ PC_5 + 0.627\ PC_6 - 0.342 PC_7 + 0.277T$

The forecast Model-P_3 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in Table 4.3.6.

Table 4.3.6 Validation of the model – P_3

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	28.30	63.7 (44.3)	37.40	4.34	5.14
2015-16	27.4	27.99		2.17	4.51	
2016-17	32.89	28.43		13.53	6.88	

** $P<0.1$

Model-P_4

In this procedure, 7 weighted and 7 un-weighted weather indices of Seven weather variables have been used in principal component analysis. The principal component analysis has identified first three principal components as more significant. Hence, these first three principal components have been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.7.

Table 4.3.7 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.965***	1.108	43.00
2	PC_1	.108	.439	
3	PC_2	.148	.539	
4	PC_3	.101	.479	
5	PC_4	-.587	.439	
6	Trend	.240	.079	

* P< 0.05, ** P<0.1

Forecast model- P_4

$Y= 21.96+0.108PC_1+0.148PC_2-0.101PC_3-0.587PC_4+0.240\ T$

The forecast Model-P_4 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in Table 4.3.8.

Table 4.3.8 Validation of the model – P_4

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	28.16	(55.4) (43.00)	36.73	4.28	5.40
2015-16	27.4	28.29		3.27	4.34	
2016-17	32.89	27.44		16.56	6.61	

** P<0.1

Model-P_5

In this procedure, 7 un-weighted and 51 un-weighted interactions weather indices of Seven weather variables have been used in principal component analysis. The principal component analysis has identified first five principal components as more significant. Hence, these first five principal components have been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.9.

Table 4.3.9 Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.88***	1.220	40.7
2	PC_1	.187	.431	
3	PC_2	.028	.595	
4	PC_3	-.623	.432	
5	PC_4	.075	.472	
6	PC_5	.038	.485	
7	Trend	.246	.086	

* P< 0.05, ** P<0.1

Forecast model- P_5

$Y = 21.88 + 0.187\ PC_1 + 0.028\ PC_2 - 0.623 PC_3 + 0.075 PC_4 + 0.038\ PC_5 + 0.246\ T$

The forecast Model-P_5 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in Table 4.3.10.

Table 4.3.10 Validation of the model – P_5

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	28.09	(56.2) (40.7)	36.40	4.47	5.30
2015-16	27.4	28.31		3.33	4.45	
2016-17	32.89	27.64		15.94	6.64	

** P<0.1

Model-P_6

In this procedure, 7 weighted and 15 weighted interactions weather indices of Seven weather variables have been used in principal component analysis. The principal component analysis has identified first four principal components as more significant. Hence, these first four principal components have been used as regressors in the development of forecasting model. The results of the fitted model are presented in the Table 4.3.11.

Table 4.3.11Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	21.884***	1.220	40.7
2	PC_1	.186	.431	
3	PC_2	.027	.595	
4	PC_3	-.623	.432	
5	PC_4	-.074	.472	
6	PC_5	.038	.485	
7	Trend	.246	.086	

* P< 0.05, ** P<0.1

Forecast model- P_6

$Y = 21.884 + 0.186\ PC_1 + 0.027\ PC_2 - 0.623\ PC_3 - 0.074\ PC_4 + 0.038 PC_5 + 0.246\ T$

The forecast Model-P_6 has been validated by forecasting the Wheat yield for the year 2014-15 to 2016-17. The results of validation are given below in Table 4.3.12.

Table 4.3.12 Validation of the model – P_6

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2014-15	20.6	27.35	(56.2)	32.76	4.60	
2015-16	27.4	29.03	(40.7)	5.95	4.34	5.47
2016-17	32.89	26.43		19.63	6.95	

** P<0.1

Comparison of the models:

The entire previous results have been summarized in Table 4.3.13 in order to compare the developed forecast models.

Table: 4.3.13 Comparison of models according to number of principal components

Model	No. of Weather Indices	R^2 (%)	Adjusted R^2 (%)	No. of Principal Components (K)
P_1	**7**	50.5	43.1	**2**
P_2	**7**	50.5	43.1	**2**
P_3	**56**	63.7	44.3	**7**
P_4	**14**	55.4	43.0	**4**
P_5	**28**	55.4	43.0	**5**
P_6	**28**	56.2	40.7	**5**

On the basis of number of weather indices, Adj. R^2 and number of principal components, the Model-P_3 has been found to be the best followed by the Model-P_1 and P_2. The value of Adj. R^2 of Model-P_3 has been found around 44.3 percent included 56 weather indices and obtained 7 principal components. The Model-P_1, P_3, and P_2 were also recommanded by Yadav *et al.* (2014) and highest value of Adj. R^2 was around 43 percent of Model-P_1 (which was best model) for Wheat crop.

Table 4.3.14 presents the actual and predicted wheat yield based on different forecast models for different years i.e. 2014-15, 2015-16 and 2016-17. The values of percent deviation of forecast from actual yield, PSE (CV), R^2, Adj R^2 and RMSE are also presented in this table. On the basis of Adj. R^2 (%), PSE and RMSE, the Model-P_3 has been found to be the best followed by the Model-P_2 and the Model-P_1. However, it may be noted that values of R^2 Adj for the models have not been found relatively high in comparison to the models developed by an application of discriminant function analysis (Agrawal *et al.*, 2012 and Sisodia *et al.*, 2014). The average percent standard error (PSE) value of the Model-P_3 is 5.24 while PSE value of Model-P_1 and P_2 are 3.55 and 3.89 respectively which shows that these models are better for forecast. This indicates that forecast of Model-P_3, P_2 and P_1 have been most satisfactory among all the models in Ayodhya district of Uttar Pradesh. Hence, a reliable forecast of Wheat yield about one and half months before the harvest can be obtained from the Model- P_3, P_2 and P_1. The results of the

Table 4.3.14 have also been graphically presented in Fig. 2 to highlight the forecasting power of the models.

Table: 4.3.14 Estimate of important parameters for the comparison of models.

Model	Year	Actual yield (Q/ha)	Predicted yield (Q/ha)	Percent deviation	Percent Standard Error	R^2 (%)	Adjusted R^2 (%)	RMSE
P_1	2014-15	20.6	28.21	36.96	3.63	50.5	43.1	5.05
	2015-16	27.4	28.63	4.52	3.84			
	2016-17	32.89	28.64	12.59	3.18			
P_2	2014-15	20.6	28.08	36.31	3.64	50.5	43.1	4.97
	2015-16	27.4	28.56	4.26	3.85			
	2016-17	32.89	28.77	12.52	4.18			
P_3	2014-15	20.6	28.30	37.40	4.34	63.7	44.4	5.14
	2015-16	27.4	27.99	2.17	4.51			
	2016-17	32.89	28.43	13.53	6.88			
P_4	2014-15	20.6	28.16	36.73	4.28	55.4	43.00	5.40
	2015-16	27.4	28.29	3.27	4.34			
	2016-17	32.89	27.44	16.56	6.61			
P_5	2014-15	20.6	28.09	36.40	4.47	56.2	40.7	5.30
	2015-16	27.4	28.31	3.33	4.45			
	2016-17	32.89	27.64	15.94	6.64			
P_6	2014-15	20.6	27.35	32.76	4.60	56.2	40.7	5.47
	2015-16	27.4	29.03	5.95	4.34			
	2016-17	32.89	26.43	19.63	6.95			

4.4 Discussion and Conclusion

4.4.1 Relationship between crop yield and weather variables

The effect of individual weather variable on the crop yield viz. Rice and wheat has been studied by various research workers Fisher (1924), Hendricks and scholl (1943), Huda *et al.* (1975), Jain *et al.* (1980), Agrawal *et al.* (1980, 1986), Yadav *et al*. (2015) etc. They have obtained varying effects of weather variable on the crop yield in different phases of crop production. Similar results in case of wheat have been observed and specific conclusion these regards are as follows.

As for as the effect of individual weather variable on wheat yield is concerned, it can be concluded that the per unit increase in the magnitude of most of the weather variables has made adverse effect on the yield during the entire crop season except during certain phases & crop growth. For example, the beneficial effect on the yield has been generally obtained during fourth phase due to unit increase in minimum temperature. Per unit increase in maximum temperature has made beneficial effect during vegetative phase and early phase of reproduction etc. Similarly, beneficial effect has been observed due to unit increase in relative humidity at seven hours almost during second and third phase of crop growth. Wind velocity has shown beneficial effect during vegetative phase if unit increase to observed in it.

4.4.2 Forecast models:

Various research workers have made attempt to develop forecast models for rice and wheat yield. Notably among them are Agrawal *et al.* (1986), Agrawal *et al.* (2012), Jain *et al.* (1980), Sisodia *et al.* (2014), Yadav *et al.* (2015) etc. Jain *et al.* (1984) have developed forecast models for rice yield using principal component analysis of biometrical characters. The present work related to wheat yield in Ayodhya district (U.P.). The various methodologies such as discriminant function analysis and principal component analysis of weekly weather data have been used to develop forecast model. The best three forecast models found from such technique are summarized in the Table 4.4.1.

Table 4.4.1 Description of the best three models

Methods based on	Model	Year	Actual yield (Q/ha)	Predicted yield (Q/ha)	Percent deviation	PSE	R^2 (%)	Adj. R^2 (%)	RMSE
Discriminant Function Analysis	D_2	2014-15	22.60	26.11	26.79	4.86	69.9	65.4	4.25
		2015-16	27.40	28.08	2.50	3.32			
		2016-17	32.89	28.06	14.67	2.87			
	D_3	2014-15	22.60	28.77	39.68	7.14	81	78.2	8.68
		2015-16	27.40	29.44	7.46	2.3			
		2016-17	32.89	26.16	20.43	2.25			
	D_6	2014-15	22.60	28.77	39.68	7.14	81.00	78.2	6.22
		2015-16	27.40	29.44	7.46	2.80			
		2016-17	32.89	26.16	20.23	2.25			
Principal Component Analysis	P_3	2014-15	20.6	28.30	31.40	4.34	63.7	44.4	5.14
		2015-16	27.4	27.99	2.17	4.51			
		2016-17	32.89	28.43	13.53	6.88			
	P_1	2014-15	20.6	28.21	36.96	3.63	50.5	43.1	5.05
		2015-16	27.4	28.63	4.52	3.84			
		2016-17	32.89	28.74	12.59	3.18			
	P_2	2014-15	20.6	28.08	36.31	3.64	50.5	43.1	4.97
		2015-16	27.40	28.56	4.26	3.85			
		2016-17	32.89	28.77	12.52	4.18			

The Table 4.4.1 represents the forecast yields by proposed technique from the three best models for the year 2014-15 to 2016-17 along with corresponding actual yield of wheat crop. On the basis of comparison of results that model based on application of discriminant function analysis of weekly weather data have performed well over the methods based on application of principal component analysis of weekly weather data. It may be pointed out here that the Model-D_2 was found best for forecasting rice yield in Kanpur district of Uttar Pradesh (Agrawal *et al.* 2012) and the Model-D_6 was found best for wheat yield in Faizabad district of Uttar Pradesh (Sisodia *et al.*, 2014). Application of principal component analysis of weekly weather data for development of forecast model for wheat yield in Faizabad district of Uttar Pradesh has results Model-P_3 as the best model (Yadav *et al.*, 2015).

The aforesaid developed models can be used for pre harvest forecast of wheat yield in neighboring district of Ayodhya district provided they have almost similar agro-climatic conditions.

Chapter-V

SUMMARY AND CONCLUSION

In this chapter summaries the entire work from all the previous chapter and valid conclusions are made out from the results.

The Chapter-I has briefly outlined the importance of the present investigation. The problem of research undertaken has been introduced reasonably and specific objectives of the study have been described here.

The relevant literature related to study the relationship between crop yield, pre- harvest forecast model based on weather variable, biometrical characters etc. have been revised critically presented in chronological order in chapter-II

The Chapter-III deals with materials and statistical methodology used for effect of developing pre-harvest forecast model of yield in Ayodhya districts. Time series data on Wheat yield and weekly data from 44^{th} SMW of previous year to 11^{th} SMW of the following year on seven weather variables *viz*, minimum temperature, maximum temperature, relative humidity at seven hour, relative humidity at fourteen hour, wind velocity rainfall and rainy days covering the period from 1990-91 to 2016-17 have been utilized to study the effect of climatic variables on crop yield and development of pre-harvest forecast model. The application of discriminant function analysis and principle component analysis for developing of pre-harvest forecast models have been described in this chapter. Various statistical measures for validation and comparison of the model have been presented in this chapter.

The Chapter-IV dealt with result and discussion. In all, Eight models have been developed to study the relationship between crop yield and weather variables. The Model-II and VI have been found at par and also have been adjusted the best among all Eight models on the basis of values of R^2. However, the fitted Model- III Model- V for the individual variable has been chosen to the study effect of individual weather variable on the yield.

As far as the effect of individual weather variables on wheat yield is concerned, it can be concluded that the per unit increase in the magnitude of most of the weather variable has made adverse effect on the yield during the entire crop season except during certain phases and crop growth. For example, the beneficial effect on the yield has been generally obtained during fourth phase due to unit increase in minimum temperature. Per unit increase in maximum temperature has made beneficial effect during vegetative phase and early phase of milking , reproduction etc. similarly beneficial effect has been observed due to unit increase in relative humidity at seven hours all most during second and third phase of crop growth. Wind velocity has shown beneficial effect during vegetative phase if unit increase is observed in it. Increase in sun-shine hours has been found to be beneficial during the phase of milking, reproduction etc.

It may, however be calculated that the weather variables play important and different role in influencing the wheat yield during different phases of crop production.

As far as in concerned with the development of forecast models. In all 7 Models based on the application of discriminant function analysis and 6 Models based on the application of principle component analysis have been developed. The best three models obtained by the application of discriminant and principle component analysis of weekly weather data have been given below.

Model D-2: $Y = 22.242 - 0.816 ds_1 + 0.166\, ds_2 + 0.219T$ $R^2=69.9\%$, $R^2_{adj}=65.4\%$

Model D-3: $Y = 22.654+0.680ds_1+0.022ds_2+0.1817T$ $R^2=81.0\%$, $R^2_{adj}=78.2\%$

Model D-6: $Y = 22.654+ 0.680ds_1+0.022\, ds_2 + 0.181\, T$ $R^2=81.0\%$, $R^2_{adj}=78.2\%$

Model P-3: $Y =21.47+0.178\, PC_1 - 0.119\, PC_2- 0.583\, PC_3+ 0.016\, PC_4+0.008\, PC_5+ 0.627\, PC_6 -0.342PC_7 + 0.277T$, $R^2=63.7\%$, $R^2_{adj}=44.4\%$

Model P-1: $Y =21.73+ 0.078\, PC_1 - 0.047\, PC_2+ 0.261T\ T$, $R^2=50.5\%$, $R^2_{adj}=43.1\%$

Model P-2: $Y =21.731+ 0.078\, PC_1 - 0.047PC_2+ 0.261\, T$

, $R^2=50.5\%$, $R^2_{adj}=43.1\%$

The forecast yields for the years 2014-15 to 2016-17 obtained from the aforesaid models, actual yield and various statistical measures for validation and comparison of the models are presented in the Table 5.1.

Table 5.1 Details of best three forecast models

Methods based on	Model	Year	Actual yield (Q/ha)	Predicted yield (Q/ha)	Percent deviation	PSE	R^2 (%)	Adj. R^2 (%)	RMSE
Discriminant Function Analysis	D_2	2014-15	22.60	26.11	26.79	4.86	69.9	65.4	4.25
		2015-16	27.40	28.08	2.50	3.32			
		2016-17	32.89	28.06	14.67	2.87			
	D_3	2014-15	22.60	28.77	39.68	7.14	81	78.2	8.68
		2015-16	27.40	29.44	7.46	2.3			
		2016-17	32.89	26.16	20.43	2.25			
	D_6	2014-15	22.60	28.77	39.68	7.14	81.00	78.2	6.22
		2015-16	27.40	29.44	7.46	2.80			
		2016-17	32.89	26.16	20.23	2.25			
Principal Component Analysis	P_3	2014-15	20.6	28.30	31.40	4.34	63.7	44.4	5.14
		2015-16	27.4	27.99	2.17	4.51			
		2016-17	32.89	28.43	13.53	6.88			
	P_1	2014-15	20.6	28.21	36.96	3.63	50.5	43.1	5.05
		2015-16	27.4	28.63	4.52	3.84			
		2016-17	32.89	28.74	12.59	3.18			
	P_2	2014-15	20.6	28.08	36.31	3.64	50.5	43.1	4.97
		2015-16	27.40	28.56	4.26	3.85			
		2016-17	32.89	28.77	12.52	4.18			

The Table 5.1 represented the forecast yields by proposed technique from the three best models for the year 2014-15 to 2016-17 along with corresponding actual yield of wheat crop. On the basis of comparison of results the models based on application of discriminant function analysis of weekly weather data have performed well over the methods based on application of principle component analysis of weekly weather data. It may be pointed out here that the Model-D_2 was found best for forecasting rice yield in Kanpur district of Uttar Pradesh (Agrawal et al. 2012) model- D_6 was found best for wheat yield in Faizabad district of Uttar Pradesh (Sisodia *et al*. 2014). Application of principle component analysis of weekly weather data for development of forecast model for wheat yield in Faizabad district of Uttar Pradesh has results Models-P_3 as the best model (Yadav *et al*. 2015).

However, in case of wheat yield, the Models D_2, D_3 and D_6 based on disciminant function analysis have been found at par for pre-harvest forecasting of yield. The Model-P_3 has also been found best among all the models based on principle component analysis.

Finally, the model- D_2 and D_6 based on discriminant function analysis and model- P_3 based on principle component analysis can be recommend for pre-harvest forecasting of wheat yield one and half month before the harvest.

The aforesaid developed models can be used for pre-harvest forecast of wheat yield in district of ayodhya provided they have almost similar agro-climatic condition.

Finally, the thesis ends with the Chapter-V of summary and conclusions.

BIBLIOGRAPHY

Agrawal, R. Jain, R.C. Jha M. P. and Singh, D. (1980). Forecasting of rice yield using climate variables. *Ind. J. Agri. Sci.,Vol.* 50 (9): pp. 680-684.

Agrawal, R. and Jain, R.C. (1982). Composite model for forecasting rice yield. *Ind. J. Agri. Sci*. Vol. 52 (3), pp.177-181.

Agrawal, R. and Mehta, S.C. (1982). Weather based forecasting of crop yield, pest and diseases – IASRI Models. *J. Ind. Soc. of Agril. Stat*. Vol.61 (2), pp. 255-263.

Agrawal, R.; Jain, R.C. and Jha, M.P. (1983). Joint effects of weather variables on rice yields. *Mausam,* Vol.34 (2), pp. 189-194.

Agrawal, R. Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variable and agriculture inputs on agro – climate zone basis. *Ind. J. Agri. Sci.* Vol.71 (7), pp. 487 -490.

Agrawal R.,Chandrahas and Kaustav A. (2012). Use of discriminant function analysis for forecasting crop yield, *Mausam.* Vol.63(3) (July 2012), pp. 455-458.

Agricultural Statistics at a Glance (2012). Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India. pp. 88-90.

Ahlawat, I.P.S. and Rana, DS (2005). Concept of efficient water use in pulses. Agrotech Publishing Academy, Udaipur, India.pp. 313-339.

Aneja, D. R. Lajpat Rai, Deepak Grover, Batra, S. D. (2008). Pre- harvest forecast models for cotton yield. Journal of Cotton Research and development, vol.22 (1), pp. 129-134.

Auffhammer, M., Ramanathan, V. and Vincent, J.R. (2011). Climate change, the monsoon, and rice yield in India. *Climatic Change* 111(2), 411–24.

Annu (2013). A study on pre-harvest crop yield forecast based on biometrical characters. Ph.D. (Agril.Statistics) thesis, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

Azfar, M., Sisodia, B.V.S., Rai, V.N. and Kumar, S. (2015). A Study on Impact of Changes in Climatic variables on Rice Production in Faizabad district of Uttar Pradesh. *Mausam*. Vol.66(2). 281-286.

Alayande, S. Ayinla, Bashiru Kehinde, Adekunle (2015). An Overview and Application of Discriminant Analysis in Data Analysis. *Journal of Mathematics (IOSR-JM)*. Volume 11, Issue 1 Ver. V (Jan - Feb. 2015), PP 12-15.

Cannor, W.P. (1999). Forecasting survival and passage of migratory Juvenil Salmonids. *North American Journal of Fishries Management,* Vol. 20, pp. 651-660.

Carter, M.M. and Eisner, J.B. (1997). A Statistical methods for forecasting Rainfall over Puerto Rico. Weather and Forecasting, Department of Meteorology, the Florida State University, Vol.12, pp. 515-525.

Chandrahas (1999). Pre harvest crop yield forecast based on plant biometrical characters reference manual *IASRI Library Avenue , Pusa , New delhi*.

Chattopadhyay, N. Samui, R.P. and Banerjee, S.K. (2008). Effect of weather on growth and yield of cotton grown in the dry farming tract of peninsular India. *Mausam*. Vol.59 (3), pp.339-346.

Choudhary, A. and Sarkar, M. B. (1981). Estimation of rice yield though weather factors in a dry sub- humid region. *Mausam*. Vol.32(4), pp. 393-396.

Draper, N. R. and Smith, H. (1981). Applied regression analysis, second edition John Willy and sons, New York.

Fisher, R. A. (1924). The influences of rainfall on the yield of wheat at Rothamsted. *Philosophical Transaction of Royal Society of London*, Series B, Vol. 213,pp. 89-142.

Goudie, A. W. (1987). Forecasting Corporate Failure: The Use of Discriminant Analysis within a Disaggregated Model of the Corporate Sector. *Journal of the Royal Statistical Society, Series A General*, 150 (1), 69-80.

Garde, Y. A. (2009). Some statistical Models for crop yield forecasting, M.Sc. (Agril.Stat.). Thesis G.B.P.U.A.T. Pantnagar, Uttarakhand.

Garde, Y. A., Shukla, A. K. and Singh, S. (2012). Pre-harvest forecasting of rice yield using weather indices in Pantnagar, (Uttarakhand), Int.J.Agril. Statistical Sci. Vol. 8(1), pp. 233-241.

Garde, Y. A., Singh, S., Mishra G. C. and Singh, T. (2012). Weather based pre- harvest forecasting of wheat at Ghazipur Uttar Pradesh. *International Journal of agricultural sciences*. Vol.8(2), pp. 325-328.

Hansena, J. W., Potgieterb, and Tippett, M.K. (2004). Using a general circulation model to forecast regional wheat yield in northeast Australia, *Agricultural and Forest Meteorology*, Vol.127, pp.77-92.

Hendricks, W.A. and Scholl, G.C. (1943). Technique in measuring joint relationship: The joint effects of temperature and precipitation on crop yield. *N. Carolina Agric. Exp. Stat. Tech. Bull.,* pp. 74.

Huda, A. K. S., Ghildyal, R. A.,Tomar, V. S. and Jain, R.C. (1975). Contribution of climatic variable in predicting rice yield. *Agric. Met.* Vol.15 pp. 71-86.

Jain, R. C. and Agrawal, R. (1992 a). Probability model for crop yield forecasting. *Biom. J*. Vol.34(4), pp. 501-511.

Jain, R. C. and Agrawal, R. (1992 b). A within year growth model for crop yield forecasting. *Biom. J*. Vol.34(7)**,** pp.789-799.

Jain, R.C., Jha, M.P. and Agrawal, Ranjana, (1985). Use of growth indices in yield forecast. Biometrical Journal, Vol.27(4), pp. 435-439.

Jain, R. C., Sridharan, H. and Agrawal, R. (1984) Principal component technique for forecasting of Sorgham yield. . *Ind. J. Agri. Sci.,* Vol.56 (6), pp. 467 -470.

Jain, R. C., Agrawal, R. and Jha, M. P. (1980). Effect of climate variables on rice yield and its forecast. *Mausam*, Vol. 31 (4) pp. 591-596.

Johnson, R.A. and Wichern, D.W. (2001). Applied Multivariate Statistical Analysis . 3rd edition, Prentice-Hall of India**.**

Joshi, N.P. Maharjan, K.L. Piya Luni (2011). Effect of Climate Variables on Yield of Major Food-crops in Nepal. *Journal of Contemporary India Studies: Space and Society, Hiroshima University.* Vol.1. 19-26.

Kandianan, K.,Karthikeyan, R., Krishnan, R., Kailasam, C. and Balasubramanian, T. N. (2002). A crop –weather model for prediction of rice (*Oryza sativa L.*) yield using an empirical – statistical technique. *Journal of agronomy and crop science*, Vol.188 (1) pp. 59-62.

Kandianan, K.,Chandaragiri, K. K.,Sankanr, N., Balasubramanian, T. N. and Kailasam, C. (2004). Crop- weather model for turmeric yield forecasting for Coimbatore district, Tamil Nadu, India. *Agricultural and forest meteorology*. Vol.112 (4), pp.133 - 137.

Khan, S. A. Majumdar, A. Mandal, B. K. Pramanik, M. and Chaudhuri, S. (1995). Agrometeorological study for forecasting the yields of winter rice and mustard in West Bengal. *Environment and Ecology*. Vol.13 (3) pp. 716-720.

Khan, S. A. Bhowmick, M. Das, L. (2006). Predicting seed yield of rape seed and mustard by agrometeorological parameters .*Journal of Interacademicia*. Vol. 10(1), pp. 54-59.

Kokate, K. D., Sananse, S. L. and Kadam, J. R. (2000). Pre- harvest forecasting of rice yield in Konkan region of Maharashtra. *J. Maha. Agric. Univ*. Vol. 25(3), pp. 289 -293.

Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*. Vol. 27, pp. 238-251.

Kumar, A. and Bhar, L. (2005). Forecasting model for yield of Indian mustard (*Brassica juncea*) using weather parameter. *Ind. J. Agri. Sci*. Vol. 75(10), pp. 688-690.

Kumar, P.R., Yadav, S.K., Sharma, S.R., Lal, S.K. and Jha, D.N. (2009). Impact of Climate Change on Seed Production of Cabbage in North Western Himalayas. *World Journal of Agricultural Sciences*, 5 (1). 18-26.

Kumar, Shashi, Awasthi, R. P. and Kumar, S. (1998). Yield forecasting in apple based on meteorological parameters. *Ind. J. Horti*. Vol. 55(4), pp. 190-195.

Lobell, D. B Cahill, K. N. and Field, C. B. (2006). Weather – based yield forecast developed for 12 California crops. *California agriculture*. Vol. 60(4), pp. 211-215.

Lee, T.S., Chih, C.C. and Chi, J.L. (2002). Forecasting of credit scores using discriminant function method. Department of Business Administration, Fu-Jen Catholic University, Taipei, Taiwan.

Marcelis, L. F. M. and Gijizen, H. (1998). A model for prediction of yield and quality of cucumber fruits. *Acta horticulture.* Vol. 47(6), pp .337- 342.

Mohd.,Azfar, (2011) A study of effect of climate change on crop production in Faizabad District, Uttar Pradesh. M. Sc. (Agril.) thesis, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

Nain, A.S, Dadhwal, V. K. and Singh, T. P. (2004). Use of CERES- wheat model for wheat yield forecast in central Indo-Gangetic Plains of India. *Journal of Agri.Sci*. Vol.142, pp. 59-70.

Ozkan, B. and Akcaoz, H. (2002). Impacts of climate factor on yield for selection crop in Southern Turkey, *Mitigation and Adaption Strategies for Global Change*. Vol. **7**, pp. 367-380.

Pal, P. and Saini, S.K. (1997). Estimated sugarcane yield on the basis of weather variable in North – West India. Indian Sugar. Vol. 47, (7), pp. 495-498.

Pal, Y, Sharma, H. C. and Kumar, A. (1996). Prediction of sugarcane yield in weather variable parameters. India Sugar. Vol. 46(7), pp. 513-517.

Pandey, K. K. (2011). A study on statistical models for forecasting crop production. M. Sc. (Agril.) thesis, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

Pandey, K.K., Rai, V.N., Sisodia, B.V.S., Bharti, A.K., Gairola, K.C. (2013). Pre -Harvest Forecast Models Based On Weather Variable And Weather Indices For Eastern U.P. *Adv. Biores*., Vol. 4 (2). 118-122.

Patel, G. B. Vaishnav, P. R. Patel, J. S. Dixit, S. K. (2007). Pre-harvest forecasting of rice (Oryza sativa L.) yield based on weather variables and technological trend. *Journal of Agrometeorology*. Vol. 9(2), pp. 167-173.

Priya, and S.K.K.Radhakrishnan, (2008). Multiple regression model for food crop yield forecast. *Agricultural Situation in India*. Vol.65 (1), pp. 11-13

Poonam, D.R.Aneja, M.L.Khichar, Mamta and Nitin Tanwar (2017). Statistical models for wheat yield forecasting using weather parameters for western agroclimatic zone of Haryana. *Bull. Env. Pharmacol. Life Sci.*, Vol 6 Special issue (4) 2017: 424-427.

Rai, T. and Chandrahas, (2000). Use of discriminant function of weather parameters for developing forecast model of rice crop.(*IASRI Publication)*.

Ramirez, A.M. (2000). Testing and evaluation weather based model in crop yield forecasting. College Laguna (Philippines). 72 leaves.

Rai, T., Onkar Swarup and Chandrahas (1998). yield forecast model for wheat (Triticum aestivum) crop. India. *Journal of Agri.Sci*. Vol. 68 (4), pp. 206-208.

Rao, A . and Singh, J. B. (1990). A study on the distribution of weather parameters and their influence on forecasting the yield of wheat. *J. Ind. Soc. of Agri. Stat.* Vol. XIII, (3), pp. 338.

Reddy S.J. and Vermin S.M. (1980). Pigeonpea and its climatic environment. In: Proceeding of the International Workshop on Pigeonpea , ICRISAT, Patancheru, India, 15-19 December ,1980. Pp: 259-270.

Runge, E.C.A. (1968). Effect of rainfall and temperature interactions during growing season on crop yield. *Agron. Jour*. Vol. 60, pp. 503- 507.

Paul,R.K, Prajneshu, and Ghosh, H.(2013) Statistical modelling for forecasting of wheat yield based on weather variables. *Indian Journal of Agricultural Sciences* 83 (2): 180–3,

Saha, S. K., Jaiswal, A.K. and Singh, B. H. (1998). Pre-harvest forecasting of sticklac yield from culture of lac insect, kerria Lacca (kerr), on Butea monosperma (Lam.) Taub. *Journal of Entomological Research.* Vol.22(3), pp. 273-282.

Sarkar, J. and Thapliyal, V. (2003). Wheat yield forecasting over Uttar Pradesh using agro-meteorology model. *J. Maha. Agric.Univ*. Vol.28.(3), pp.299-302.

Sarkar, R.P. (1965). A curvilinear study of yield with reference to weather-sugarcane. Ind. J.Met. Geophys. Vol.16, pp. 103-110.

Sarkar, J. (2000). Wheat yield forecasting over Gujarat using agro-meteorology model. *J. Maha.Agric. Univ*. Vol. 25(3), pp. 294-297.

Sardna, M.G., Khosla, R.K., Ohri, N.K., Mitra, P.C. (1972). Pre-harvest forecasting of yield rate of jute. Jute Bull. Vol.35, pp. 7-8.

Sinclair, E.H., Moore, S.E., Friday, N.A., Zeppelin, T.K. and Waite, J.M. (2005). Do patterns of Stellar sea lion (Eumetopias jubatus) diet, population trend and cetacean occurrence reflect oceanographic domains from the Alaska Peninsula to the central Aleutian Islands. *Fisheries Oceanography,* 14, 223-42.

Sarma, A. A. L. N., Kumar, TV. L. and Koteswararao, K. (2008). Development of an agro-climate mode for the estimation of rice yield. *Journal.Ind. Geophys, Union*, Vol, 12(2), pp.89-96.

Shamasundaran, K. S. Venugopalan, R. (2006). Statistical modelling for pre-harvest forecast: an illustration with rose .*Journal of Horticultural Sciences*. Vol.1(1), pp. 68-70.

Sharma, A. K. (1996). *Applied multivariate technique*. John Wiley & Sons, New Delhi.

Sharma, A., Sood, R. K. and Kalubarme M. H. (2004). Agrometeorological wheat yield forecast in Himachal Pradesh. Journal of Agrometeorology, 6 (Special Issue): pp.153-160.

Shinha AC, Mandal BB and Jana PK. (1988). Physiology analysis of yield variation in irrigated pigeonpea in relation to time of sowing, row spacing and weed control measures. Indian Agriculturist Vol. 32, pp.177-185.

Singh, B. H. and Bapat, S. R. (1988) Pre-harvest forecast models for prediction of sugarcane yield. Indian Journal of Agricultural Sciences, vol. 58(6), pp. 465-469.

Singh, H., Singh, K. N. Hassan, B. and Khan, A. A. (2007). Agro- climatic models for prediction of growth and yield of rice (*Oryza sativa*) under temperature Kashmir condition. *Ind. J. Agri.Sci. vol.*80(3), pp.254-257.

Sisodia, B. V. S., Yadav, R. R., Kumar, S. and Sharma, M. K. (2014). Forecasting of Pre- harvest crop yield using discriminant function analysis of meteorological parameter. *Journal of Agrometeorology*. Vol.16(1), pp.121-125.

Tripathi Ramesh (2013). A Study on Statistical Models for Pre-harvest forecasting of potato yield based on weather variables. . M. Sc. (Agril.) thesis, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

Varmola, S. L. Dixit, S. K. Patel, J. S. Bhatt, H. M. (2004). Forecasting of wheat yields on the basis of weather variables. Journal of Agrometeorology. Vol.6(2), pp. 223-228.

Vogel, Fred, A. (1985). Crop forecasting methodology current and future procedure in the U.S. proceeding of the 45th session of the *International Statistics Institute held at Amsterdam:* Vol.13(1), pp. 1-13.

Wery J, Ture O and Lecoeur J. (1993).Mechanisms of resistance to cold, heat and drought in cool season legumes with special reference to chickpea and pea. In KB singh and saxena (Eds), Breeding for stress tolerance in cool-season food legumes. John Wiley & Sons.,chichester, UK. pp: 271-291.

Yadav, R. R. (2010). A study on statistical models for pre- harvest forecasting of wheat yield based on discriminant function, M. Sc.

(Agril.) thesis, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

Yadav, S. T. and Patil, B. P. (2008). Influences of agro-climate indices on fruit yield of cucumber. *J. Maha. Agric. Univ.*, Vol. 33(2), pp. 237-239.

Yadav, R.R., Sisodia, B.V.S., Kumar, S. (2014). Application of Principal Component Analysis in Developing Statistical Models to Forecast Crop Yield Using Weather Variables. *Mausam*, Vol.65(3).357-360.

Appendix – 1

Fitted models for individual weather variables

S.N0.	Weather variables	Model No.	Model	R^2	R^2 adj
1	**Minimum Temperature**	I	$Y = 37.039 + 0.777Z_1 + 0.129T$ (3.876) (0.194) (0.056)	**0.62**	**0.59**
		II	$Y = 37.039 + 0.777Z_1 + 0.129T$ (3.876) (0.194) (0.056)	**0.62**	**0.59**
		III	$Y = 22.386 + 0.609Z_1$ (0.742) (0.136)	**0.44**	**0.42**
		IV	$Y = 22.386 + 0.609Z_1$ (0.742) (0.136)	**0.44**	**0.42**
		V	$Y = 37.039 + 0.777Z_1 + 0.129T$ (3.876) (0.194) (0.056)	**0.62**	**0.59**
		VI	$Y = 37.039 + 0.777Z_1 + 0.129T$ (3.876) (0.194) (0.056)	**0.62**	**0.59**
		VII	$Y = 37.039 + 0.777Z_1 + 0.129T$ (3.876) (0.194) (0.056)	**0.44**	**0.42**
		VIII	$Y = 4.967 + .031Z_{11}$ (.562) (.006)	**0.44**	**0.42**
2	**Maximum Temperature**	I	$Y = 34.059 + 0.463Z_1 + 0.163T$ (4.09) (.152) (0.059)	**0.54**	**0.51**
		II	$Y = 34.059 + 0.463Z_1 + 0.163T$ (4.09) (.152) (0.059)	**0.54**	**0.51**
		III	$Y = 27.024 + 0.524Z_1$ (0.581) (0.133)	**0.38**	**0.35**
		IV	$Y = 27.024 + 0.524Z_1$ (0.581) (0.133)	**0.38**	**0.35**
		V	$Y = 34.059 + 0.463Z_1 + 0.163T$ (4.09) (.152) (0.059)	**0.54**	**0.51**
		VI	$Y = 34.059 + 0.463Z_1 + 0.163T$ (4.09) (.152) (0.059)	**0.54**	**0.51**
		VII	$Y = 25.125 + 0.12Z_{11}$ (0.380) (0.003)	**0.38**	**0.36**
		VIII	$Y = 25.125 + 0.12Z_{11}$ (0.380) (0.003)	**0.38**	**0.36**

3	**Relative Humidity at 7 hour**	**I**	$Y = 13.707 + 0.271Z_1 + 0.1544T$ (2.950) (0.094) (0.062)	**0.53**	**0.49**
		II	$Y = 13.707 + 0.271Z_1 + 0.1544T$ (2.950) (0.094) (0.062)	**0.53**	**0.49**
		III	$Y = 46.178 + 0.388Z_1$ (5.319) (0.099)	**0.38**	**0.35**
		IV	$Y = 46.178 + 0.388Z_1$ (5.319) (0.099)	**0.38**	**0.35**
		V	$Y = 17.574 + 0.002Z_{11} + 0.150T$ (1.740) (.001) (.061)	**0.55**	**0.51**
		VI	$Y = 17.574 + 0.002Z_{11} + 0.150T$ (1.740) (.001) (.061)	**0.55**	**0.51**
		VII	$Y = 36.496 + 0.003Z_{11}$ (2.846) (.001)	**0.38**	**0.36**
		VIII	$Y = 36.496 + 0.003Z_{11}$ (2.846) (.001)	**0.38**	**0.36**
4	**Relative Humidity at 14 hour**	**I**	$Y = 29.969 + 0.197Z_1 + 0.211T$ (2.217) (0.051) (0.50)	**0.61**	**0.58**
		II	$Y = 29.969 + 0.197Z_1 + 0.211T$ (2.217) (0.051) (0.50)	**0.61**	**0.58**
		III	$Y = 9.569 + 0.115Z_1$ (4.228) (0.031)	**0.35**	**0.33**
		IV	$Y = 9.569 + 0.115Z_1$ (4.228) (0.031)	**0.35**	**0.33**
		V	$Y = -3.042 - 1.471Z_1 + 0.026Z_{11} + 0.219T$ (13.37) (0.670) (0.007) (0.045)	**0.69**	**0.65**
		VI	$Y = -3.042 - 1.471Z_1 + 0.026Z_{11} + 0.219T$ (13.37) (0.670) (0.007) (0.045)	**0.69**	**0.65**
		VII	$Y = 67.358 - 0.688Z_1 + 0.008Z_{11}$ (21.105) (.0290) (0.003)	**0.51**	**0.47**
		VIII	$Y = 67.358 - 0.688Z_1 + 0.008Z_{11}$ (21.105) (.0290) (0.003)	**0.51**	**0.47**

5	**Rainy Days**	**I**	$Y = 27.703 + 0.188Z_1 + 0.203T$ (0.850) (0.053) (0.052)	**0.37**	**0.34**
		II	$Y = 27.703 + 0.188Z_1 + 0.203T$ (0.850) (0.053) (0.052)	**0.37**	**0.34**
		III	$Y = 24.581 + 0.179Z_1$ (0.379) (0.038)	**0.47**	**0.45**
		IV	$Y = 24.581 + 0.179Z_1$ (0.379) (0.038)	**0.47**	**0.45**
		V	$Y = 22.440 + 0.024Z_0 - 1.099Z_2 + 0.261T$ (0.817) (0.011) (0.230) (0.045)	**0.69**	**0.65**
		VI	$Y = 22.440 + 0.024Z_0 - 1.099Z_2 + 0.261T$ (0.817) (0.011) (0.230) (0.045)	**0.69**	**0.65**
		VII	$Y = 24.955 + 0.005Z_{11}$ (0.344) (0.001)	**0.50**	**0.48**
		VIII	$Y = 24.955 + 0.005Z_{11}$ (0.344) (0.001)	**0.50**	**0.48**
6	**Rainfall**	**I**	$Y = 22.659 + 0.216Z_1 + 0.204T$ (0.795) (0.52) (0.049)	**0.63**	**0.60**
		II	$Y = 22.659 + 0.216Z_1 + 0.204T$ (0.795) (0.52) (0.049)	**0.63**	**0.60**
		III	$Y = 24.723 + 0.190Z_1$ (0.351) (0.037)	**0.51**	**0.49**
		IV	$Y = 24.723 + 0.190Z_1$ (0.351) (0.037)	**0.51**	**0.49**
		V	$Y = 22.659 + 0.216Z_1 + 0.204T$ (0.795) (0.052) (0.049)	**0.63**	**0.60**
		VI	$Y = 22.659 + 0.216Z_1 + 0.204T$ (0.795) (0.052) (0.049)	**0.63**	**0.60**
		VII	$Y = 24.723 + 0.190Z_1$ (0.351) (0.037)	**0.51**	**0.49**
		VIII	$Y = 24.723 + 0.190Z_1$ (0.351) (0.037)	**0.51**	**0.49**

7	**Wind Velocity**	**I**	$Y=21.878 + 0.238T$ (0.988) (0.062)	**0.37**	**0.34**
		II	$Y=21.878 + 0.238T$ (0.988) (0.062)	**0.37**	**0.34**
		III	$Y=21.135 + 2.43Z_1$ (0.608) (0.627)	**0.37**	**0.35**
		IV	$Y=21.135 + 2.43Z_1$ (0.608) (0.627)	**0.37**	**0.35**
		V	$Y=20.445 + 0.255Z_{11}$ (1.209) (0.60)	**0.42**	**0.39**
		VI	$Y=20.445 + 0.255Z_{11}$ (1.209) (0.60)	**0.42**	**0.39**
		VII	$Y=27.135 + 2.434Z_1$ (0.608) (0.627)	**0.37**	**0.35**
		VIII	$Y=27.135 + 2.434Z_1$ (0.608) (0.627)	**0.37**	**0.35**

Appendix – 2

Detail of generated weather indices (Sum)

Years	Yield	Adjusted	Z10	Z11	Z20	Z21	Z30	Z31	Z40	Z41	Z50	Z51	Z60	Z70	Z71
1990-91	20.34	27.28	10.11	-273.17	27.87	123.41	65.81	310.38	44.29	70.12	1.42	7.30	0.85	0.85	11.37
1991-92	21.42	26.42	9.61	-259.85	23.94	105.99	70.42	332.15	47.50	75.20	1.84	9.45	1.37	0.48	6.40
1992-93	23.00	25.06	9.69	-261.99	24.87	110.09	73.89	348.52	46.80	74.10	1.58	8.12	1.19	1.19	15.82
1993-94	20.89	27.39	10.23	-276.63	25.82	114.33	77.91	367.44	43.23	68.45	0.79	4.04	3.58	3.58	47.71
1994-95	23.44	25.07	9.26	-250.39	23.97	106.11	73.57	347.00	42.52	67.31	1.78	9.14	1.17	1.17	15.64
1995-96	25.30	23.43	10.30	-278.41	24.86	110.04	73.56	346.97	44.43	70.35	1.67	8.59	8.35	8.35	111.33
1996-97	25.02	23.93	8.52	-230.37	24.92	110.35	69.66	328.56	39.34	62.28	1.03	5.31	0.19	0.19	2.49
1997-98	26.76	22.41	10.53	-284.70	21.69	96.04	68.17	321.53	51.56	81.62	1.62	8.32	6.24	6.24	83.17
1998-99	23.97	25.42	9.87	-266.68	21.75	96.28	68.32	322.25	47.21	74.74	1.80	9.23	2.09	1.71	22.75
1999-00	22.30	27.31	9.64	-260.55	25.13	111.24	66.77	314.93	49.30	78.06	2.10	10.79	0.97	0.97	12.88
2000-01	26.26	23.57	9.57	-258.75	25.71	113.81	64.68	305.08	42.49	67.27	2.27	11.65	0.42	0.42	5.60
2001-02	23.92	26.14	10.34	-279.60	24.49	108.41	72.27	340.86	49.56	78.46	2.11	10.82	2.17	2.17	28.97
2002-03	21.79	28.49	9.33	-252.26	24.50	108.47	69.30	326.83	49.86	78.94	1.67	8.60	1.99	2.12	28.26
2003-04	26.37	24.13	9.37	-253.16	23.28	103.07	69.81	329.28	50.10	79.32	2.37	12.19	4.11	4.11	54.73
2004-05	23.77	26.95	9.37	-	24.93	110.39	70.00	330.15	47.78	75.65	1.93	9.93	1.25	1.89	25.23

				253.16											
2005-06	26.07	24.87	8.13	-219.65	25.85	114.43	68.51	323.15	40.01	63.35	1.82	9.35	0.00	0.00	0.00
2006-07	26.83	24.34	9.49	-256.41	25.79	114.20	67.79	319.72	43.80	69.34	2.40	12.33	2.06	2.07	27.54
2007-08	27.58	23.81	9.19	-248.48	25.05	110.89	72.44	341.68	42.56	67.38	1.69	8.67	0.00	0.00	0.00
2008-09	28.1	23.51	10.27	-277.67	25.74	113.96	71.87	338.96	45.60	72.19	1.68	8.63	0.00	0.80	10.66
2009-10	26.05	25.78	9.09	-245.78	23.68	104.84	69.05	325.67	46.93	74.30	2.11	10.86	7.48	1.66	22.12
2010-11	28.14	23.91	8.69	-234.79	24.55	108.68	69.17	326.25	48.53	76.83	2.29	11.75	0.17	0.17	2.22
2011-12	29.56	22.72	9.59	-259.11	23.67	104.81	71.35	336.54	51.35	81.30	2.45	12.61	5.32	5.32	70.91
2012-13	29.63	22.87	7.98	-215.69	23.33	103.27	70.95	334.62	45.87	72.63	2.20	11.30	1.94	1.94	25.86
2013-14	23.34	29.38	9.64	-260.55	23.32	103.25	71.33	336.44	49.89	78.98	2.31	11.89	4.31	4.31	57.49
2014-15	20.6	32.34	8.92	-241.09	21.91	96.99	72.16	340.32	51.69	81.84	2.60	13.36	3.88	3.88	51.71
2015-16	27.4	25.76	8.89	-240.19	26.51	117.35	67.95	320.50	45.73	72.40	2.12	10.89	0.00	0.00	0.00
2016-17	32.89	20.50	9.74	-263.26	23.69	104.90	77.95	367.66	48.05	76.08	2.15	11.06	0.45	1.12	14.93

DEPARTMENT OF AGRICULTURAL STATISTICS
ACHARYA NARENDRA DEVA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
NARENDRA NAGAR (KUMARGANJ), AYODHYA-224 229 (U.P.)

Title: “A Study on impact of climate change on Wheat crop yield and development of statistical models for pre- harvest forecast of crop - yield in Ayodhya district of eastern U.P**”**

Major Advisor & Chairman:
Dr. V. N. Rai
Professor/Head/Dean

Name of Student:
Ravi Prakash Gupta
I. D. No.: A-8957/15/17
Ph. D. Agricultural Statistics

ABSTRACT

The present investigation entitled “A Study on impact of climate change on Wheat crop yield and development of statistical models for pre- harvest forecast of crop - yield in Ayodhya district of eastern U.P” consists of five chapters including summary and conclusion. The purpose of the study is to develop statistical models for studying the relationship between weather variables and crop yield and to develop different forecast models based on discriminant function and principal component analysis.

Time series data on wheat crop yield and weekly data from 44^{th} SMW of previous year to 11^{th} SMW of the following year on seven weather variables *viz.*, minimum temperature, maximum temperature, relative humidity at 7 hour, relative humidity at 14 hour, wind velocity rainfall and rainy days covering the period from 1990-1991 to 2016-2017 have been utilized to study the relationship wheat crop yield and weather variables and development of pre-harvest forecast model. In all, eight models have been developed to study the relationship between crop yield and weather variables. The model-V has been found to be the best for studying the relationship between crop yield and weather variables.

Statistical methodology using multiple regression, discriminant functions and principal component analysis for developing pre-harvest forecast model has been described. In all, 13 models (one based on regression, seven from discriminant function and six from principal component) have been developed for pre-harvest forecast model. The model-A is based on weather indices, D_2 to D_6 based on discriminant function and P_1 to P_2 based on principal component analysis have been developed. On the basis of Adjust R^2, RMSE and PSE, the best three models for both technique obtained by the application of discriminant function and principal component analysis of weekly weather data are D_2, D_3 & D_6 and P_1, P_2 & P_3 respectively. These models can be used to get the reliable forecast of wheat one and half months before the harvest.

Raviprakash Gupta

(V. N. Rai) *(Ravi Prakash Gupta)*