

**DEVELOPMENT OF YIELD PREDICTION MODELS
FOR OILSEEDS BASED ON WEATHER PARAMETERS
IN INDIA**

M.Sc. (Ag) Thesis

by

SURENDRA A R

**DEPARTMENT OF AGRICULTURAL STATISTICS AND
SOCIAL SCIENCE (L.)
COLLEGE OF AGRICULTURE, RAIPUR
FACULTY OF AGRICULTURE
INDIRA GANDHI KRISHI VISHWAVIDYALAYA
RAIPUR (Chhattisgarh)
2019**

**DEVELOPMENT OF YIELD PREDICTION MODELS
FOR OILSEEDS BASED ON WEATHER PARAMETERS
IN INDIA**

Thesis

Submitted to the

Indira Gandhi Krishi Vishwavidyalaya, Raipur

by

SURENDRA A R

**IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF**

Master of Science

In

Agriculture

(Agricultural Statistics)

V.V. ID No. 20171827352

ID No. 120117069

OCTOBER, 2019

CERTIFICATE - I

This is to certify that the thesis entitled "Development of yield prediction models for oilseeds based on weather parameters in India" submitted in partial fulfillment of the requirements for the degree of Master of Science in Agriculture of the Indira Gandhi Krishi Vishwavidyalaya, Raipur, is a record of the bonafide research work carried out by SURENDRA A R under our guidance and supervision. The subject of the thesis has been approved by the Student's Advisory Committee and the Director of Instructions.

No part of the thesis has been submitted for any other degree or diploma or certificate course. All the assistance and help received during the course of the investigations have been duly acknowledged.

Co-Chairman K. Alivelu

Pandey
Chairman

Date:

THESIS APPROVED BY THE STUDENT'S ADVISORY COMMITTEE

Chairman : (Dr. K K Pandey)

Pandey

Co-Chairman : (Dr. K. Alivelu)

K. Alivelu

Member : (Dr. (Smt.) S. Shukla)

Shukla

Member : (Dr. Praveen Verma)


Verma

Member : (Shri Umesh)

Umesh

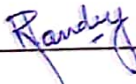
CERTIFICATE – II

This is to certify that the thesis entitled “Development of yield prediction models for oilseeds based on weather parameters in India” submitted by SURENDRA A R to the Indira Gandhi Krishi Vishwavidyalaya, Raipur, in partial fulfillment of the requirements for the degree of Master of Science in Agriculture in the Department of Agricultural Statistics and Social Science (L.) has been approved by the external examiner and Student’s Advisory Committee after oral examination.



Signature External Examiner
(Name.....)

Date: 23/11/19

Major Advisor



Head of the Department


_____ 23-11-19

Dean / Faculty Dean

Approved/Not approved

Director of Instructions

ACKNOWLEDGEMENT

With ever regardful memories.....

With all my humble and wholehearted prostration before Almighty and Omnipresent "God" who always showered blessing on me to reach the acme of success and always with me for improving the confidence and determination in my whole life. How can I express my thanks to "God" because there is no any word to express it? So, my lord, please realize and accept my feelings.

I wish to express my deep sense of reverence to my guide Dr. K K Pandey Assistant Professor, Department of Agriculture Statistics and Social Science (L.), College of Agri. & Res. station, IGKV, Bhatapara. I am profoundly thankful to him for his constant encouragement, keen interest, constructive criticism and everlasting patience throughout the endeavor and devoting his time amidst his busy schedule. His fatherly way of treatment, patient counseling and wise direction will never be forgotten.

With stupendous ecstasy and profundity of complacency, I pronounce utmost of gratitude to my Co-Major Advisor, Dr. K. Aivelu, Principle scientist, ICAR- Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad. For her constant attention, care, support, constructive criticism, and proper guidance during my stay there and course of investigation. I am lucky to work under the superb guidance of her.

It is with great respect and devotion I place on record, my deep sense of gratitude and indebtedness to my advisory committee members Dr. (smt).S.Shukla, Professor, Department of Agricultural Statistics & Social Science (L.), College of Agriculture, Raipur (C.G.), Dr. Praveen Verma, Department of Agricultural Economics, College of Agri. & Res.station, IGKV, Bhatapara. Shri Umesh, Department of Genetics and Plant Breeding, College of Agri. & Res. station, IGKV, Bhatapara. for their sustained and valuable guidance, constructive suggestions, unfailing patience, friendly approach, constant support and encouragement throughout the period of my research work.

I am highly obliged to Hon'ble Vice- Chancellor Dr. S.K. Patil, Dr. O.P. Kashyap, Dean, College of Agriculture, Raipur (C.G.), Dr. S.S. Rao, Director of Research Services, IGKV, Raipur (C.G.) for providing necessary facilities to conduct the investigation.

I do, express my heartiest thanks to the teachers of my department Dr. A.K Singh, Dr. M.L. Lakhera, Dr. Ravi R. Saxena, Dr. Gayatri Chandrakar, Dr. Sindhu Shukla and

Dr. Sweta Ramole for their constant co-operative suggestion, encouragement and help during my investigation.

I would like to express my sincere gratitude to Dr. M. Pandey (Librarian, Nehru Library, Raipur) for giving me their kind helps during my present study.

I am highly thankful to all the non-teaching staff of my department namely, Shri P.B. Pradhan, Shri Subhash Agarwal, Sushma Raj, Kajai Bai and Chero for their help and co-operation.

Friends always knew how to pep me up during my bad days and how to celebrate when I achieve something. It is difficult to express the gratitude I owe to all my friends especially to Revanasiddesh, Umesh, Suresh, Sanjay, Basavaraj, Dileep, Linga, Nagaraj, Kumar, Pradeep, Anji, Ramanjalu, Ashok, Mani, Ramana, Vinod. I extend a very heartfelt thanks to them for their whole hearted solidarity. There are many friends and well-wishers who helped me in various ways towards the present study and they deserve my sincere thanks. I also wish my warmest thanks to my respectful seniors Manju sir, Srikanth sir, Yellaswamy sir, Phani sir who spared their valuable time and energy and extended their cooperation to me.

Words fail to express my indebtedness to my divine parents. I am eternally grateful to my beloved father Shree Ravanappa and Mother Smt. Padmavathamma for their unconditional love, personal sacrifices, encouragement and giving me liberty to choose what I desired. They have been selfless in giving me the best of everything. I consider myself the luckiest in the world to have such parents. I am also eternally grateful to my beloved uncle Shree Somashekhar and Auntysmt. Gayathri for their unconditional love, encouragement and Support. I extend my respect to brother, Srinivasa, Anjaneya, Gangadhara, Nagendra Babu, Manjunath, Charan, Srikanth, Gajendra and Sister, Sumadevi, Manimala, Putty, Shirisha, Shalini for their love and support. I thank my entire family for their constant moral support, love and stand by me in difficult path to complete this work.

At last, I want to reserve a special word of thanks for all who helped me during my research work.

Department of Agricultural Statistics and Social Science (L.),

College of Agriculture,

IGKV, Raipur (C.G.)

Date: 1.7.10.1.2019.


Surendra A R

TABLE OF CONTENTS

Chapter	Title	Page
	ACKNOWLEDGEMENT	i
	TABLE OF CONTENTS	iii
	LIST OF TABLES	iv
	LIST OF FIGURES	ix
	LIST OF ABBREVIATIONS /NOTATIONS	xi
	ABSTRACT	xii
I	INTRODUCTION	1
II	REVIEW OF LITERATURE	6
III	MATERIALS AND METHODS	20
	3.1 Data Collection	20
	3.2 Software Used	21
	3.2.1 R-Software	21
	3.3 Analytical tools and techniques used	21
	3.3.1 Multiple Linear Regression Model	21
	3.3.2 LASSO Stepwise Regression	22
	3.3.2.1 Lasso methodology	23
	3.3.2.2 Lasso Regression, Cross-Validation and Consistency	25
	3.3.3 Models based on weather indices	26
	3.3.4 Model using Composite Weather Variables	28
	3.3.5 Artificial Neural Networks (ANN)	30

3.3.5.1	ANN Modelling	31
3.3.5.2	Back-Propagation algorithm	33
3.3.5.3	ANN Illustration	34
3.4	Selection of Model	35
3.4.1	Coefficient of Determination (R2 criteria)	35
3.4.2	Root Mean Square Error (RMSE)	35
3.4.4	Mean Absolute Percentage Error (MAPE)	35
IV	RESULTS AND DISSCUSSION	36
4.1	Collection of time series of both crop yield and meteorological data	36
4.1.1	Time Series of Crop Yield	36
4.1.1.1	Results From Analysis Of The Trends In Oilseeds Yield	36
4.1.2	Time Series of Meteorological Parameters	38
4.1.2.1	Results From Analysis Of The Trends In Maximum Temperature	38
4.1.2.2	Results From Analysis Of The Trends In Minimum Temperature	39
4.1.2.3	Results From Analysis Of The Trends In Annual Rainfall Of India	41
4.1.2.4	Results From Analysis Of The Trends In Annual RH1 Of India	42
4.1.2.5	Results From Analysis Of The Trends In Annual RH2 Of India	43
4.2	Determine the statistical relationship between weather parameters and yield	45
4.2.1	Results From The Correlation Analysis	45

4.2.2	Inter correlation between weather parameters	46
4.2.3	Nature and relationship between individual weather variables and yield	47
4.3	Develop prediction models to predict yield from climate variables over the area of study	50
4.3.1	Multiple Linear Regression Model	50
4.3.2	LASSO Stepwise Regression	54
4.3.2.1	Preparing the data	54
4.3.2.2	Computation of lasso regression	54
4.3.2.3	Finding the optimal value of lambda that minimizes the cross-validation error	57
4.3.2.4	computing the final lasso model	59
4.3.2.5	Predictions of oilseeds yield using lasso	61
4.3.3	Models based on weather indices.	64
4.3.3.1	Weather indices calculation	64
4.3.3.2	Model based on simple weather indices	65
4.3.3.3	Model based on weighted weather indices	66
4.3.3.4	Predictions of oilseeds yield based on weather indices	68
4.3.4	Models using composite weather variables	70
4.3.4.1	data requirement	71
4.3.4.2	Model development	71
4.3.4.2	Predictions of oilseeds yield using composite weather variables	72

4.3.5	Artificial Neural Network (ANN)	75
4.3.5.1	Preparing to fit the neural network	75
4.3.5.2	Predicting oilseeds yield using the neural Network	79
4.4	Validation of the developed models	81
4.4.1	R ² and Root Mean Square Error (RMSE)	81
4.4.2	Yield data comparison between predicted and observed values	82
4.4.3	Cross-comparison of the developed models	85
V	SUMMARY AND CONCLUSIONS	87
	REFERENCES	94
	RESUME	102

LIST OF TABLES

Table	Title	Page
3.3.3	Notations for simple weather indices and weighted weather indices	27
4.1.1	Descriptive statistics of total oilseeds yield in india from 1970-2017	36
4.1.2.1	Descriptive statistics of annual maximum temperature in india from 1970 - 2017	37
4.1.2.2	Descriptive Statistics of Annual Minimum Temperature in India from 1970 – 2017	39
4.1.2.3	Descriptive statistics of annual rainfall in india from 1970-2017	40
4.1.2.4	Descriptive Statistics of Annual RH1 in India from 1970-2017	42
4.1.2.5	Descriptive Statistics of Annual RH2 in India from 1970-2017	43
4.2.1	Correlation between weather parameters and yield	45
4.2.2	Inter correlation between weather parameters	45
4.3.1	Summary of the multiple linear regression model	50
4.3.2	ANOVA for multiple regression model	50
4.3.3	Coefficients of the model	50
4.3.2.3	Regression coefficients using lambda.min as best lambda	57
4.3.2.4	Regression coefficients using lambda.lse as best lambda	58
4.3.2.5	Lasso model for yield forecasting	58
4.3.2.6	Model Summary	59
4.3.2.7	ANOVA of the model	59

4.3.3.1	Notations for simple weather indices and weighted weather indices	58
4.3.3.2	Simple weather indices-based regression model	64
4.3.3.3	Model Summary	64
4.3.3.4	ANOVA for the developed model	64
4.3.3.5	Coefficients for the developed model	65
4.3.3.6	Weighted weather indices-based regression model	65
4.3.3.7	Model Summary	65
4.3.3.8	ANOVA for weighted weather indices-based model	66
4.3.3.9	Coefficients for the developed model	66
4.3.4.1	Model using composite weather variables	70
4.3.4.2	Model Summary	71
4.3.4.3	ANOVA for model using composite weather variable	71
4.3.4.4	Coefficients for the developed model	71
4.3.5.1	Min-Max normalization to scale the data	75
4.3.5.2	Estimated Neural Network Weights	76
4.4.1	Validation of Artificial Neural Network (ANN) model	82
4.4.2	Validation of LASSO Stepwise Regression Model	82
4.4.3	Validation of composite weather variables model	83
4.4.4	Validation for weather indices-based models	83
4.4.5	Comparison of developed models	84
4.4.7	Forecast of oilseeds yield using best fitted model	85

LIST OF FIGURES

Figure	Title	Page
3.3.2	Visual interpretation comparing OLS and LASSO regression	23
3.3.5.1	Neural network modeling procedure	31
3.3.5.2	Structure of basic artificial neural network model	31
4.1.1	Time series of total oilseeds yield in india from 1970 to 2017	36
4.1.2.1	Time series of annual maximum temperature in india from 1970 to 2017	37
4.1.2.2	Time series of annual maximum temperature in india from 1970 to 2017	39
4.1.2.3.1	Time series of annual rainfall in india from 1970 to 2017	40
4.1.2.3.2	Seasonal Rainfall in India from 1970 to 2017	40
4.1.2.4	Time series of annual RH1 in india from 1970 to 2017	42
4.1.2.5	Time series of annual RH2 in india from 1970 to 2017	43
4.2.3.1	Scatterplot of yield verses maximum temperature	47
4.2.2.2	Scatterplot of yield verses minimum temperature	47
4.2.2.3	Scatterplot of yield verses rainfall	48
4.2.2.4	Scatterplot of yield verses RH1	48
4.2.2.5	Scatterplot of yield verses RH2	49
4.3.1	Histogram for the multiple linear regression model	51
4.3.2	Actual vs Predicted yield using regression model	52
4.3.3	Normal P-P Plot of Regression Standardized Residuals	52
4.3.2.2	LASSO Coefficients	55

4.3.2.3	Cross-validation error according to the log of lambda	56
4.3.2.4	Actual yield verses predicted yield using LASSO regression	60
4.3.3.4	Actual yield and predicted yield using LASSO regression	61
4.3.2.5	Histogram of the model	61
4.3.2.6	Normal P-P Plot of Regression Standardized Residuals	62
4.3.3.4	Actual verses predicted yield based on weather indices	67
4.3.3.4	Actual and predicted yield based on weather indices	68
4.3.3.5	Histogram of the developed model	68
4.3.3.6	Normal P-P Plot of Regression Standardized Residual	69
4.3.4.2	Actual verses predicted yield using composite weather variables	72
4.3.4.3	Actual and predicted yield using composite weather variables	72
4.3.4.4	Histogram of the developed model	73
4.3.4.5	Normal P-P Plot of Regression Standardized Residual	73
4.3.5.1	Artificial neural network model for yield based on weather parameters	77
4.3.5.2	Actual yield verses predicted yield using neural network method	78
4.3.5.3	Actual yield and predicted yield using neural network method	79
4.3.5.4	Residual by Predicted Chart	79

LIST OF ABBREVIATIONS/NOTATIONS

%	Percent
<i>et.al</i>	And others
i.e.,	That is
etc.	Many more
kg	Kilogram
ha	hectare
Mt	Million tons
°C	Degree Celsius
mm	millimeter
viz.	videlicet (namely)
λ	Lambda
S. E	Standard Error
r	Correlation coefficient
ANOVA	Analysis of Variance
df	degrees of freedom
PE	Percentage Error
R ²	Multiple Coefficient of Determination
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
OLS	Ordinary Least Square
LASSO	Least Absolute Shrinkage Selection Operator
CWV	Composite Weather Variables
ANN	Artificial Neural Network

THESIS ABSTRACT

- a) Title of the Thesis : Development of yield prediction models for oilseeds based on weather parameters in india
- b) Full Name of the Student: Surendra A R
- c) Major Subject : Agricultural Statistics
- d) Name and Address of the Major Advisor : Dr. K. K. Pandey, Assistant professor
Department of Agricultural Statistics and Social Science(L), DKS, College of Agriculture and Research Station, Bhatapara.
- e) Degree to be Awarded : M.Sc. (Ag.), Agricultural Statistics

Signature of the Major Advisor

Signature of the Student

Date:

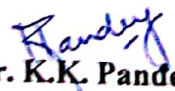
Signature of Head of Department

ABSTRACT

Oilseeds cultivation is undertaken across the country in about 26.2 million hectares, of which 72% is confined to rainfed farming. The diverse agro-ecological conditions in the country are favourable for growing nine annual oilseed crops, which include seven edible oilseeds (groundnut, rapeseed & mustard, soybean, sunflower, sesame, safflower and niger) and two non-edible oilseeds (castor and linseed). Any changes in weather parameters might affect the oilseeds yield. So, the crop yield prediction based on weather parameters will help farmers, policy makers and administrators to manage activities. The present study examines the development of yield prediction models for oilseeds based on weather parameters in india. The study was undertaken based on secondary data, the data on yield of total oilseeds in india for 48 years data from 1970-71 to 2017-18 has been collected from www.indiastat.com. Weather data for five weather parameters (Max & Min Temperature, Rainfall, RH1 and RH2) were collected from India Meteorological

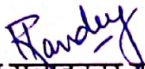
Department (IMD), Pune and www.indiawaterportal.org. Four models were developed for oilseeds yield namely LASSO Stepwise Regression, Models based on weather indices (MWI), Models using composite weather variables (CWV) and Artificial Neural Networks (ANN). A comparison study of the results obtained from LASSO, MWI, CWV and ANN is also performed. The results were compared using the R^2 , RMSE statistic and percentage prediction error. The R^2 & RMSE of the models varied between 0.72 & 66.66, 0.69 & 79.39, 0.74 & 62.45 and 0.81 & 56.99 respectively. The average percentage prediction error for LASSO, MWI, CWV and ANN models on the test set were found to be 5.99, 7.20, 5.55 and 5.18 respectively. The results indicate that the ANN model had a higher R^2 value, lower RMSE value and a lower percentage prediction error when compared to other models. The ranking of the model reveals that Artificial Neural Networks (ANN) was the best performing model followed by models using composite weather variables, LASSO Stepwise Regression model and models based on weather indices. The reason behind the good performance of ANN model is that, in a real sense, neural networks are one of the best solutions in search for a few agriculture problems, especially when it comes to predict crop yield. A conclusion has been made that ANN has better explained yield variability rather than other methods. Undeniably, the application of ANN to precision agriculture plays a crucial role in future evaluation of the concept of precision agriculture as a sustainable means of meeting crop yield demands. However, further research about the ANN impacts towards crop yield production must be conducted to ensure sustainability of future needs.

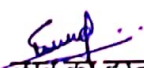
Department of Agricultural Statistics and Social Science (L.)
College of Agriculture, Raipur (Chhattisgarh)


Dr. K.K. Pandey
Major Advisor


शोध सारांश

- क) थीसिस का शीर्षक : भारत में मौसम के मापदंडों के आधार पर तिलहन के लिए उपज मॉडल का विकास
- ख) छात्र का पूरा नाम : सुरेन्द्र ए.आर.
- ग) प्रमुख विषय : कृषि सांख्यिकी
- घ) प्रमुख सलाहकार का नाम और पता : डॉ के के पाण्डेय, सहायक प्राध्यापक
कृषि सांख्यिकी और सामाजिक विज्ञान विभाग (एल),
दाऊ कल्याण सिंह कृषि महाविद्यालय एवं
अनुसंधान केंद्र, भाटापारा
- च) उपाधि : स्नाकोत्तर (कृषि), कृषि सांख्यिकी


प्रमुख सलाहकार का हस्ताक्षर


छात्र का हस्ताक्षर

दिनांक: 29.11.19


विभागाध्यक्ष का हस्ताक्षर

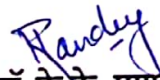
सारांश

तिलहन की खेती देश भर में लगभग 26.2 मिलियन हेक्टेयर में की जाती है, जिसमें से 72% वर्षा आधारित खेती तक ही सीमित है। देश में विविध कृषि-पारिस्थितिक स्थितियां नौ वार्षिक तिलहन फसलों को उगाने के लिए अनुकूल हैं, जिसमें सात खाद्य तिलहन (मूंगफली, रेपसीड और सरसों, सूरजमुखी, तिलहन, तिल, कुसुम और नीगर) और दो गैर-खाद्य तिलहन (अरंडी और अलसी) हैं। शामिल हैं।)। मौसम के मापदंडों में कोई भी बदलाव तिलहन की उपज को प्रभावित कर सकता है। इसलिए, मौसम के मानकों के आधार पर फसल की उपज की भविष्यवाणी किसानों, नीति निर्माताओं और प्रशासकों को गतिविधियों का प्रबंधन करने में मदद करेगा। वर्तमान अध्ययन भारत में मौसम के मापदंडों के आधार पर तिलहन के लिए उपज प्रॉक्सी मॉडल के विकास की जांच करता है। अध्ययन माध्यमिक आंकड़ों के आधार पर किया गया था, भारत में 1970-71 से 2017-18 तक 48 वर्षों के कुल तिलहन की उपज का डेटा www.indiastat.com से एकत्र किया गया है। भारत के मौसम विभाग (IMD), पुणे और www.indiawaterportal.org से पांच मौसम मानकों (अधिकतम और न्यूनतम तापमान, वर्षा, आरएच। और

आरएच 2) के लिए मौसम डेटा एकत्र किए गए थे। तिलहन की उपज के लिए चार मॉडल विकसित किए गए, जैसे कि लिस्ट एब्सोल्यूट थ्रिंकेज सिलेक्शन ऑपरेटर (LASSO), मौसम संकेतक (MWI) पर आधारित मॉडल, मिश्रित मौसम चर (CWV) और कृत्रिम तंत्रिका नेटवर्क (ANN) का उपयोग करने वाले मॉडल। LASSO, MWI, CWV और ANN से प्राप्त परिणामों का तुलनात्मक अध्ययन भी किया जाता है। परिणामों की तुलना R^2 , RMSE स्टेटिस्टिक और प्रतिशत पूर्वानुमान त्रुटि का उपयोग करके की गई। मॉडल के आर² और आरएमएसई क्रमशः 0.72 और 66.66, 0.69 और 79.39, 0.74 और 62.45 और 0.81 और 56.99 के बीच भिन्न थे। परीक्षण सेट पर LASSO, MWI, CWV और ANN के लिए औसत प्रतिशत भविष्यवाणी मॉडल की त्रुटि क्रमशः 5.99, 7.20, 5.77 और 5.18 पाई गई। परिणाम बताते हैं कि ANN मॉडल में उच्च R^2 मान, RMSE मूल्य कम और LASSO की तुलना में कम प्रतिशत रेटिंग त्रुटि थी, मौसम संकेतकों पर आधारित मॉडल और मिश्रित मौसम चर का उपयोग करके मॉडल। मॉडल की रैंकिंग से पता चलता है कि आर्टिफिशियल न्यूरल नेटवर्क (ANN) सबसे अच्छा प्रदर्शन करने वाला मॉडल था, जिसके बाद से कम से कम पूर्ण संकोचन चयन ऑपरेटर (LASSO), कंपोजिट वेदर वेबल्स और मेट्रो संकेतकों पर आधारित मॉडल का उपयोग करने वाले मॉडल थे। एएनएन मॉडल के अच्छे प्रदर्शन के पीछे कारण यह है कि, वास्तविक अर्थों में, तंत्रिका नेटवर्क कुछ कृषि समस्याओं की तलाश में सबसे अच्छे समाधानों में से एक हैं, खासकर जब फसल की उपज की भविष्यवाणी करने की बात आती है। एक निष्कर्ष निकाला गया है कि एएनएन ने अन्य तरीकों के बजाय उपज परिवर्तनशीलता को बेहतर बताया है। निर्विवाद रूप से, सटीक कृषि के लिए ANN का प्रश्न, फसल उपज मांगों को पूरा करने के एक स्थायी साधन के रूप में, सटीक कृषि की अवधारणा के भविष्य के मूल्यांकन में महत्वपूर्ण भूमिका निभाता है। हालांकि, फसल की पैदावार उत्पादन की दिशा में ANN प्रभावों के बारे में और अधिक शोध भविष्य की आवश्यकताओं की स्थिरता सुनिश्चित करने के लिए आयोजित किया जाना चाहिए।

कृषि सांख्यिकी और सामाजिक विज्ञान विभाग (एल)

कृषि महाविद्यालय, रायपुर (छत्तीसगढ़)


डॉ के के पाण्डेय
प्रमुख सलाहकार

CHAPTER - I

INTRODUCTION

India is one of the major oilseed's grower and importer of edible oils. India's vegetable oil economy is world's fourth largest after USA, China & Brazil. The oilseed accounts for 14% of the Gross Cropped Area, 5% of the Gross National Product and 10% value of all agricultural commodities (2017-18).

The diverse agro-ecological conditions in the country are favourable for growing nine annual oilseed crops, which include seven edible oilseeds (groundnut, rapeseed & mustard, soybean, sunflower, sesame, safflower and niger) and two non-edible oilseeds (castor and linseed). Oilseeds cultivation is undertaken across the country in about 26.2 million hectares of which 72% is confined to rainfed farming.

During the last few years, the domestic consumption of edible oils has increased substantially and has touched the level of 25.25 million tonnes in 2017-18 and is likely to increase further. For a projected population of 1.34 billion, the total vegetable oils demand is likely to touch 33.20 million tonnes by 2020 with per capita consumption of vegetable oils at the rate of more than 22 kg/year/person.

The area and production of oilseeds are mainly concentrated in the states of Gujarat, Rajasthan, Madhya Pradesh and Maharashtra. Among different oilseeds, soybean (10.98 Mt), groundnut (9.18 Mt) and rapeseed and mustard (8.32 Mt) together account for about 86.41 per cent of oilseeds area and 90.96 per cent of oilseeds production in the country (2017-18).

Nine oilseeds are the primary source of vegetable oils in the country, which are largely grown under rainfed conditions. Among these, soybean (35%), groundnut (29%), rapeseed & mustard (29%) contributes to more than 90% of total oilseeds production and more than 80% of vegetable oil with major share of mustard (35%), soybean (23%) and groundnut (25%). In addition to nine oilseeds, 3 million tonnes of

vegetable oil is being harnessed from secondary sources like cottonseed, rice bran, coconut, Tree Borne Oilseeds (TBOs) and Oil Palm.

With competing demands on agricultural land from various crops and enterprises, the oilseeds production can be increased only if productivity is improved significantly and farmers get remunerative and attractive prices and assured market access. However, farmers face various constraints in the production of oilseeds. Several biotic, abiotic, institutional, technological, and socio-economic constraints also inhibit exploitation of the yield potential of crops and need to be addressed. By considering facts, viz. changing policy environment, slow growth in domestic production, increasing demand and rising imports, an attempt has been made to develop the statistical and artificial intelligence models to analyze the past performance, its trend and future forecast of oil seed production of India.

By late 1970's, India became a net importer of edible oils due to increased demand and stagnation in production. Until then India was a net exporter of oilseeds, oil meals, edible oils and extractions. In spite of having second largest area under oilseeds in the world, India's import quantity of edible oils reached around 30 per cent of the total imports by mid-1980's which made edible oils a major item of import, only after petroleum products. By 1990's, government decided to launch various policies and technological interventions in order to achieve self-sufficiency in edible oilseeds by promoting new production and processing technologies of oilseeds through various schemes sponsored by the centre.

1.1 National Mission on Oilseeds and Oil Palm (NMOOP):

India has a serious import dependency in edible oil. One of the biggest constraints to raising oilseed output has been that production is largely in rain-fed areas. Only one fourth of the oilseed producing area in the country remains under the irrigation.”

In 1986, government had launched a Technology Mission on Oilseeds to improve productivity. This resulted in some growth but then growth in this field has been

sluggish only. Current Government is promoting NMOOP during 2012-17. This mission has some clear-cut objectives such as:

- Increasing seed substitution ratio (SSR) in oil crops focusing on varietal substitution.
- Increasing oilseed irrigation coverage from 26% to 36%
- Area diversification from low yielding cultivation of cereals to oilseeds; pulses /intercropping of cereals / sugarcane oilseeds.
- Use of fallow land after cultivation of paddy / potato.
- Oil Palm and tree borne oilseeds cultivation expansion in watersheds and wastelands.
- Increased availability of quality planting material to improve the procurement and collection of oilseeds. and
- Tree borne oilseeds Processing.
- Intercropping oil palm and tree-borne oilseeds during the gestation period would provide farmers with financial return when there is no production.

The mission is implemented with funds in 75:25 Centre-state ratios. In this scheme, there are three mini missions. I, II and III mini-mission. Mini Mission-I is focused on oilseeds; Mini Mission II is focused on oil palms and Mini Mission III is focused on tree-borne oilseeds (TBOs).

Yield prediction is one of the most critical issues faced in the agricultural sector. lack of knowledge about harvest glut for farmer's, uncertainties in the weather conditions and seasonal rainfall policies, depletion of nutrition level of soils, fertilizer availability and cost, post-harvest loss, pest control and other factors leads to decrease in the production of the crops.

Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The data when become information is highly useful for many purposes. Crop yield forecasting is

important for national food security providing timely information for optimum management of growing crops. It helps farmers to develop a better plan for next year's crop and government to plan import/export strategies. Crop forecasting at various spatial scale, district, state and national level, is fundamental to provide decision makers with timely information for rapid and sound decisions during the growing season.

In agricultural production, weather plays a significant role. It has a deep impact on a crop growth, development and yield, incidence of pests and diseases, water needs and fertilizer requirements in terms of nutrient mobilization variations due to water stress and timeliness and efficiency of prophylactic and cultural activities on crops. Weather forecast is referred to the weather prediction in advance. For agricultural activities, weather forecast assumes considerable importance for effective planning of modern agricultural practices such as sowing, management of crop diseases and pests, highly efficient irrigation, and harvest planning.

Agriculture is an extremely dependent economic sector on climatic conditions. Crop models are often used to assess weather forecasts ' capacity to guide crop management practices. In statistical model approach, one or more variables (representing climate or weather) are related to crop responses such as yield and yield contributing characters. Crop yield is the integrated impact of several physical and physiological processes occurring during the period of crop growth. These processes are influenced by the characteristics of the crop, weather, soil and time management factors. Statistical models provide simple alternative to process- based models. The major advantages of statistical models are their restricted dependence on information from field calibration and their transparent evaluation of model uncertainties. Shankar and Gupta (1988) related the meteorological parameters and yield data using the phenological periods, in which weather elements illustrate significant correlation with crop yield, by using multiple linear regressions.

Models based on weather parameters can provide reliable forecast of crop yield in advance of harvest. The largely unpredictable weather therefore plays a major role in determining the agricultural performance. Understanding the role of climatic factors on

crop yields is essential in predicting the future impact of climate change. It helps the researchers to fix the bench mark of yields.

Extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season. So, there is need to develop statistically sound objective forecasts of crop yield based on weather variables so that reliable forecasts can be obtained.

The present study was undertaken with the following specific objectives:

1. Collection of time series of both crop yield and meteorological data.
2. Determine the statistical relationship between weather parameters and yield.
3. Develop prediction models to predict yield from climate variables over the area of study.
4. Validation of the developed models.

CHAPTER – II

REVIEW OF LITERATURE

In order to evaluate the objectives of the study, it was considered desirable to have an understanding of the researchers ' results of some of the previous studies and the methodology they adopted.

Leyla Muradkhanli (2018) Used Neural Networks for Prediction of Oil Production. Neural network MLP was used in his research to predict oil production. The neural network model that has been created is well fitted, has better predictions and greater precision. The results were contrasted with the evaluation of regression and it demonstrates that the model of the neural network accomplished better efficiency of prediction. Great matching was found between the real production of data oil and the expected data. The results showed that the SOCAR oil production series MLP neural network model has better forecast performance. This advanced model would help to predict the output of oil in SOCAR.

Geetika and Rohit Bajaj (2018) Studied crop yield prediction using data mining: an efficient data modeling. This paper addresses wheat crop yield prediction using approaches to data mining. They used k-means of clustering and extraction of features using analysis of independent components, and arrangement is completed by means of linear discriminating analysis and simulation is done using MATLAB environment. It can be observed from the outcome and conversations that our suggested strategy is capable of achieving less probabilities of error rate and an elevated signal to noise ratio. The method suggested is capable of achieving a mean 0.5 square error rate and a 33.18 db. maximum signal to noise ratio.

Vuong M. Ngo. *et.al* (2018) developed an efficient data warehouse for crop yield prediction. In this paper, they described a data warehouse solution for crop farming. The data warehouse (DW) supports Big Data analytics and storage. The recent applications of precision agriculture systems and major challenges into building a successful

precision agriculture data warehouse were reviewed and analysed. The DW architecture includes necessary modules to deal with large scale and efficient analytics.

Andrew Crane-Droesch (2018) made use of machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Crop yields are critically weather-dependent. An increasing empirical literature is modeling this connection to project effects on the industry of climate change. They define a yield modeling strategy that utilizes a semi-parametric version of a profound neural network that can simultaneously account for complicated nonlinear interactions in high-dimensional datasets, as well as recognized parametric composition and unnoticed cross-sectional heterogeneity. Using scenarios from a suite of climate models showed major adverse effects of climate change on corn yield, but less serious than those projected using conventional statistical methods.

Panwar. S. et.al (2018). Forecasting of crop yield using weather parameters—two step nonlinear regression model approach. The Concept of paper is firstly to remove the trend of crop yield and then to develop the forecasting models using detrended yield. Not much work is available or development of forecast models or modelling due to their non-linear behavior. For that, in this paper, methodology developed for forecasting using nonlinear growth models, which will help in forecasting yield, pest and disease incidences etc. with high accuracy. Crop yield forecast models for wheat crop have been developed for different districts of Uttar Pradesh (UP). Weather Indices (WI) were obtained using above two approaches. Weather index-based regression models were developed using weather indices as independent variables, while under-study character such as crop yield was used as dependent variable for wheat crop, i.e. two non-linear forecast model steps. Technique of forecasting using non-linear approach and using weather indices will enrich the knowledge in developing customized models on forecasting for different types of crops and for different locations. The approach provided about two months before harvest a reliable yield forecast.

Satvinder Kour et.al (2018) developed a forecasting of rice yield based on weather parameters. in kheda district of Gujarat, India. The present study was

undertaken to investigate the feasibility of estimating the yield of rice based on combined effects of weather parameters and technological advancement, using past weather records for Kheda district of middle Gujarat state. Two models are being created at the tillering point of rice development by including the most significant weather factors in the current inquiry. ARIMAX models are used as an illustration to forecast rice yield in the Gujarat district of Kheda. Comparative study of the fitted models is carried out from the viewpoint of Relative mean absolute percentage error (RMAPE), Mean absolute deviation (MAD) and root mean square error (RMSE) values.

Arun Kumar, Naveen Kumar and Vishal Vats (2018) have been done comparative study which tells the accuracy of training proposed model and error rate. The training model's precision should be greater and the error rate should be minimal and the model proposed can provide the actual cost of estimated crop yield and is labeled as LOW, MID, and HIGH. This work basically offers the comparative analysis of different algorithms when applying these algorithms to data sets and demonstrates the precision of each algorithm to train the data sets and also means squared error during the sample data cross-validation stage. This work is independent from a domain. It means that we can build other domain systems, such as medical, product comparison, retail, etc. We simply need to pass through this system the datasets, but the dataset should be consistent.

Surya. P. and Laurence A. I (2018) crop yield prediction in agriculture using data mining predictive analytic techniques. This paper dealt with various regression techniques for agriculture crop yield prediction. In their proposed work mainly focused on to get predictor model by using regression techniques. Predictor formula is most useful in the crop prediction of Agriculture crop Production in Tons. Highest rate of yield production in tamilnadu state particularly in North Western zone is sugarcane and tapioca. Banana, Maize, Ragi, Turmeric, Coconut, Cotton, Jowar are having next highest yield of production rate compared to the other crops depends on the results of predictor model.

Manjula. E and Djodiltachoumy. S (2017) developed model for crop yield prediction. This article provides a short evaluation of crop yield prediction based on data mining method for the chosen area i.e. Tamil Nadu district in India. The purpose of this study is to suggest and execute a rule-based system to predict crop yield production from previous data. This was achieved by applying the association rule mining from 2000 to 2012 to agricultural data. The experimental results show that crop yield production is effectively predicted by the proposed work.

Santosha Rathod *et.al* (2017) studied modeling and forecasting of oilseed production of India through artificial intelligence techniques. A nonlinear artificial intelligence technique like neural networks and support vector machines can be an effective way to improve forecasting performance. Based on the results obtained in this work one can infer that application of artificial intelligence techniques like time delay neural networks and nonlinear support vector regression techniques in modeling and forecasting of time series can increase the forecasting accuracy, in particular, the nonlinear support vector regression model performed better for forecasting oilseed production of India as compared to other models.

Betty J. Sitienei, Shem G. Juma and Everline Opere (2017) on the use of regression models to predict tea crop yield responses to climate change: A Case of Nandi East, Sub-Country of Nandi Country, Kenya. They tried using a regression model to predict tea yield based on changes in maximum temperature, minimum temperature, and precipitation over the study area. To determine the quality of climate information, a single mass curve method was used. The study defined statistically the mean and skewedness of the variables under investigation. Attempts have been created to develop an equation of multiple linear regression that best reflects the connection between variables. The model was verified and specified using a contingency table. The annual rainfall time series showed a positive, though not statistically significant, trend. This can be attributed to the two major rainfall seasons ' opposite trend. Nevertheless, The annual rainfall showed enhanced variability and high negative skewness, implying higher annual drought frequencies.

Prajakta Prashant Bhangale et.al (2017) conducted a study on improved crop yield prediction using neural network. Data mining can be implemented in many fields with the enhancement of data mining techniques, particularly those without subjective premises or humans. To estimate crop yield assessment with current information and their use in data mining, some data mining methods were introduced. This introduces fresh study opportunities for applying contemporary methodologies of classification to the yield prediction problem. In agriculture, there are an increasing number of applications of data mining methods and an increasing quantity of data presently accessible from many resources. Several methods of data mining used in the area of agricultural research. Various weather changes are evaluated using SVM. To classify the soil and crops, K means approach is used. He found that ANN is a useful instrument for predicting crops. This involves their regional soil parameter. Using feed forward propagation ANN, it is then analyzed. Analyzed to create it more effective in the matlab ANN approach.

Menaka. K, N. Yuvaraj M.E, Miste (2017) studied an ANFIS based crop yield prediction model. The application of crop yield prediction model-ANFIS with different vital input characteristics is discussed in this paper. using feature selection algorithm (PSO), These attributes are pre-processed. By comparing the results with predictive models such as SVM and ANN, the efficiency of the suggested model can be predicted. The model's predictive results are contrasted with some predictive algorithms like SVM and ANN based on the values of precision, recall, F-measure and precision. From the results of the comparison, we conclude that the classification precision will improve many times over the ordinary prediction models by adding and processing all the essential characteristics. ANFIS can therefore be used more closely than any other model to predict crop yield.

Jin-Ki Park, Amrita Das and Jong-Hwa Park (2017) developed an integrated model for predicting rice yield with climate change. This article discusses a new approach to rice yield prediction using artificial neural network, spatial interpolation, remote sensing, and GIS techniques. Here, the yield variation is ascribed to climatic

parameters and crop health, and the standardized vegetation difference index from MODIS is used as a plant health and development indicator. It was provided due significance to scaling up the input parameters using spatial interpolation and GIS and minimizing the sources of mistake in each modeling step. The low percentage error (2.91) and high correlation (0.76) signifies the robust performance of the proposed model. This easy but efficient method is then used to assess the impact of climate change on the production of South Korean rice. As suggested in the RCP 8.5 scenario, temperature increases can boost rice production across South Korea.

Nishad *et.al* (2016) conducted a study by fitting estimated models for area, production and yield of rapeseed-mustard in Chhattisgarh and its constituent districts. It has been noted that more regions in the Chhattisgarh plain are being brought under rapeseed-mustard at a substantial partial compound growth rate of 21.08% after the fifth year i.e. 2002-03 and at a non-significant annual partial compound growth rate within each specified period. Except for Rajnandgaon, the entire plain area and the districts of its constituent showed non-significant yield patterns, suggesting a lack of effect of new techniques in those areas. It was concluded that the production is more influenced by the yield (60-80%) than by the area. It has been predicted that by the turn of a decade, the rise in region and decline in productivity under rapeseed-mustard will be (10.89 percent and -15.66 percent per year, respectively) compared to the growing population's oil requirement. Therefore, it was suggested that gaps and constraints be identified and that new enhanced techniques be adopted to improve productivity at a much quicker rate.

Sellam. V and Poovammal. E (2016) studied prediction of crop yield using regression analysis. Regression Analysis is used in this study to evaluate the factors of the environment and their impact on crop yield. Regression Analysis is a method of multivariate analysis that analyzes the factors grouping them into variables of explanatory and response and helps to reach a decision. They found that in Annual Rainfall, Area under Cultivation and Food Price Index, there is a steady growth or slight difference that obviously affects crop production. The influenced value $R^2 = 0.7$ is

obtained by implementing the Regression Analysis for the data. This R² value clearly states that AR, AUC and FPI have an average of 70% influence in the crop yield. Regression Analysis is thus used in this study to create a connection between a set of AR, AUC and FPI factors and their impacts on rice crop yield. This research can be expanded by considering more variables such as Minimum Support Price (MSP), Weather Conditions, Soil Parameters etc. that affect crop yield and by using different data mining, statistical techniques to analyze factors that influence yield.

Pandey *et.al* (2016) conducted a study on development of forecasting model on weather parameters for eastern U.P. The present study has been carried out for the development of pre-harvest forecast model of rice at district level based on weather variables. Weekly data of 14 meteorological weeks on seven weather variables under rice crop season over a period of 21 years from 1989-90 to 2009-10 have been used along with the annual data of rice production for Faizabad district of eastern Uttar Pradesh. Generated weather variable (56 weighted, unweighted and joint +1 Time trend) has used under the stepwise regression to screen out the important weather variables and multiple regression approach was subsequently employed to estimate model parameters. The proposed model contains combination of weighted and unweighted weather variables and explains 76% (significant at 1% level of significance) of the variability of rice production in terms of R². The model has been developed for 19 years data and validated for 2 years data. Predicted yield for the 2009 is 23.765 and 2010 is 24.31 with deviation 0.021% and -5.69 % respectively, RMSE (1.498) for the model is also calculated for the validation.

Shreya S. Bhanose *et.al* (2016) conducted a study on Crop and Yield Prediction Model. The research demonstrates in crop conjecture about altered k- Means clustering by raising the count of quality and accuracy. The altered clustering algorithm k-Means is evaluated by comparing k-Means and k-Means++ algorithms, achieving the highest number of high-quality clusters, correct crop prediction and the highest accuracy count. Info mining plays a significant role in improving harvest prediction in the agricultural sector. The suggested research is performed on the Maharashtra Condition crop dataset.

Our future work involves considering geographic region for worldwide harvest prediction using the world's geographic information system.

Sri Preethaa M.E. (2016) conducted a study on crop yield prediction. This article can enhance the crop's net yield rate by comparing the parameters in the datasets and analyzing them. By using the Bayesian algorithm, where precision can also be predicted, analysis and prediction can be performed. This makes farmers take the correct choice for the correct crop so that creative ideas will develop the agricultural sector.

Kavita Bhatt, K. K. Gill and Sandeep Singh Sandhu (2015) conducted a study on Comparison of Different Regression Models to Predict Mustard yield in Central Punjab. An attempt was made to predict mustard yield over central Punjab (Ludhiana, Amritsar and Patiala) by regression models. For Ludhiana, among all the three models, Basic model explained up to 84 % variation in yield due to weather parameters while modified model explained highest i.e. 87% variation in mustard yield due to weather parameters. In case of Amritsar district, basic, modified and SPSS model explained 57, 59 and 62 % variation in mustard yield, respectively. In case of Patiala, the modified model was the best fit model to explain mustard yield as it explained up to 82 % variation in yield followed by SPSS (76 %) and basic model (49 %). The results reveal that modified model is best fit for Ludhiana and Patiala region, while SPSS model fits better for Amritsar region as far as mustard yield is concerned.

Balakrishnama Naidu. V, Siva Sankar. A and Leelavathi. C (2014) conducted a study on trends in area, production and productivity of selected oil seed crops in Andhra Pradesh. Oilseed crops contribute a significant proportion to the agricultural GDP. However, Indian oilseed yields are among the lowest in the globe, as most are cultivated on marginal non-irrigated land with low quality seed. The yields per hectare in kilograms of different oilseeds are groundnut 913, Soybean 1008, rapeseed-mustard 875 and linseed 344. From the analysis, it is clear that the area, production and yield under Groundnut and sesame have been decreasing during the study period from 1996-97 to 2011-12 in Andhra Pradesh. It is also concluded that there is a negative growth is

registered under area of all the above said oilseeds crops. The yield of groundnut is almost stagnant, but the yield of sesame and sunflower are much better than groundnut.

Vikas Lamba and V.S. Dhaka (2014) developed a wheat yield prediction using artificial neural network and crop prediction techniques. This survey paper reflects the Agricultural (Wheat Crop) forecasting methods. Paper demonstrates all previous forecasting research development in all areas. The main predictive models in the field of agricultural yield are Statistical, Meteorological, Simulation, Agronomic, Remote Satellite Sensed, Synthetic and Mathematical. This article demonstrates a compact mixture of all these models and demonstrates why Neural Network Model is essential for nonlinear information behavior system models such as wheat crop yield forecast from other models. ANN tool is user-friendly and very effective for nonlinear information as well. Many researches have been conducted so far in forecasting using ANN tool and they have been very precise.

Snehal. S. Dahikar and Sandeep V. R (2014) studied an agricultural crop yield prediction using artificial neural network approach. By considering different climate phenomena scenarios influencing local weather conditions in different areas of the globe. These weather conditions have a direct effect on crop yield. Different research has been conducted to explore the associations between climate events on a large scale and crop yield. It has been shown that artificial neural networks are strong tools for modeling and predicting to improve their effectiveness. Methodology of crop prediction is used to predict the appropriate crop by detecting different soil parameters and atmospheric parameters. Parameters such as soil type, PH, nitrogen, phosphate, phosphorus, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, humidity. We use artificial neural network (ANN) for this purpose. They found that ANN is a useful instrument for predicting crops. The parameter of their regional soil parameter is included in this paper. It is then analyzed through the use of ANN feed forward propagation. Analysis to make it more efficient in matlab ANN approach.

Amit K. Singh et.al (2014) Conducted a study on development of multi stage district level wheat yields forecast models. Multistage crop yield forecasts were done by developing district level statistical yield models at midseason (45 days before harvest, DBF) and pre-harvest stage (25 DBF). Crop yield forecasts were generated for wheat crop in *rabi* season in Amritsar, Bhatinda and Ludhiana districts of Punjab. Composite weather variables were used to develop the model statistical yield prediction. These models developed at midseason (F2) and pre-harvest (F3) stage of crop growing period. Simple and weighted weather indices have been designed for individual climate factors as well as for the season-wide interaction of two factors at a time. Statistical models based on weather indices can successfully simulate multi-stage yield forecast of wheat. The nRMSE values were 11.6, 12.0 and 12.5 for Amritsar, Bhatinda and Ludhiana respectively in mid-season forecast, while for pre-harvest forecast, values were 6.0, 8.6 and 8.4, respectively for Amritsar, Bhatinda and Ludhiana districts.

Tirtha R. Ranjeet and Leisa J. Armstrong (2014) developed an artificial neural network for predicting crop yields in Nepal. In this article, to predict crop yields in Nepal, the efficiency of an artificial neural network is examined. The neural network model was a mild mistake in the training stage that demonstrates the precision of this model. Furthermore, the comparison between the expected and the neural net result demonstrates the comparable result which further proves the precision of the prediction values using this model of the neural network. The expected result can be used to boost paddy yield in Siraha district as well as in other areas where topography and vegetation are comparable to Siraha district. Future research involves investigating and comparing their performance with other neural network architectures including feed forward, acyclic and modular neural networks. In addition, by using search algorithms such as the genetic algorithm, the appropriate input parameters, appropriate amount of concealed layers and number of neurons can be chosen to improve forecast precision.

Ghosh K. et.al (2014) Conducted a study on development of crop yield forecast models under FASAL- a case study of *kharif* rice in West Bengal. For establishment of methodology of crop yield forecast under FASAL project, modified Hendrick and

Scholl model (Agrawal *et.al.* 1986) using composite weather indices has been used for developing the District level yield forecast models and their performance is studied. Simple and weighted weather indices are ready for individual weather factors as well as two at a time interaction considering throughout the crop growing season during the model development stage. The model's performance in predicting district-level yields for various major crops in various country states is quite satisfactory. The studies carried out for *kharif* rice of West Bengal in the paper also reflect the similar result. The prediction of State level *kharif* rice yield for West Bengal is also near accurate for both the years.

Askar Choudhury (2014) developed a crop yield prediction using time series models. This paper makes several important literature contributions. It provides additional evidence of the component of a time series ' crop yield cycle in most districts. It also indicates evidence of the unnoticed impact of external variables on crop yield creating the crop yield cycle. However, this research does not explore any crop yield connections that may occur with the non-observed external variables. While important, these findings are not unexpected considering the vibrant modifications resulting from external variables such as weather (rainfall, temperatures, etc.), land management (including district re-division), pests, and diseases. The unexpected finding is the original continuous decrease in crop yield that has been going on in most districts for several years without any interference from management.

William W. Guo and Heru Xue (2014) conducted a study on crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models. This article introduces the recent study results from the use in crop forecasting of spatial as well as temporal neural network models. Our simulation demonstrates that the spatial NN model can predict wheat yield with a high precision compared to the temporal NARNN and NARXNN models with regard to a specified plantation area. However, in crop yield forecasting, the elevated precision of the spatial NN model is restricted to crop yield forecasting only within ordinary ranges.

Ramakrishna R & Naveen Kumar Boiroju (2013) have done the forecasting of yield per hectare of Rice in Andhra Pradesh using Box-Jenkins methodology and feed forward neural networks are presented. The forecasting performance of the models evaluated using mean absolute error, mean absolute percentage error and root mean squared errors. Modeling of neural networks outperforms the model of Box-Jenkins. It provides the rice yield (in Kg) error measurements and predictions using the two models. We computed the forecasts for the given data using both the models and computed the root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Pandey *et.al* (2013) conducted a study on Pre -Harvest Forecast Models Based on Weather Variable and Weather Indices for Eastern U.P. In the present study, attempts were made to develop models based on weather variables to forecast rice yield at district level. Weekly data (of 19 weeks) on seven weather factors over a 21-year period (1989-90 to 2009-10) was used in conjunction with annual rice production data for Faizabad district (eastern UP). Stepwise regression was used to screen major weather variables and subsequently multiple regression approaches were used to estimate model parameters. The model –I was developed as best for yield forecasting out of five models evaluated, this model's R² and RMSE is 71.2%, 0.733 respectively followed by model V with R² and RMSE 51.25%, 1.2524 respectively.

Lobell D.B. and Burke, M.B. (2010). Using statistical models to predict reactions to climate change from crop yields. Here they used an ideal model strategy to examine the capacity of statistical models to predict yield reactions to mean temperature and precipitation changes, as a process-based crop model simulates. The agreement between the projections of the process-based and statistical models was then evaluated as a measure of how well statistical models can capture crop reactions to changes in heating or precipitation. Statistical models performance varied by climate variable and spatial scale, with time-series statistical models ably reproducing site-specific yield reaction to changes in precipitation but performing less well for temperature reactions. On the other hand, statistical models relying on various site data, namely panel and cross-sectional

models, were better at predicting temperature change reactions than changes in precipitation. Climate predictions are also the most accessible and reliable at these wider scales, and therefore statistical models are likely to continue to play a significant part in predicting future climate change effects.

Xiuhong *et al* (2010) carried out a study on tea yield prediction model based on SPSS statistical software in Rizhao city. According to the data of meteorological and tea production from 1993 to 2008 in Rizhao city in Shandong province, correlation analysis of tea production and climatic factors was conducted by using the SPSS statistical software. The results showed that tea yield can be forecasted and fits a higher rate of actual yield, and the maximum prediction accuracy can be up to 100%, and the smallest accuracy is 86 per cent with the average accuracy of 94 per cent. The prediction model has high accuracy and practicality, and can be used as an effective tool for prediction of tea production.

Christopher J. Martinez *et.al* (2009) studied a use of climate indices to predict corn yields in southeast USA. PCA was used to summarize Alabama, Florida, and Georgia rainfall, surface temperature, and county maize yield residuals to decrease data sets dimensionality. Linear correlation was assessed for the connection of corn yield residuals with precipitation and surface temperature. The association of corn yield residuals, seasonal precipitation, and surface temperature (which showed the greatest correlation with yield residuals) was then assessed by linear correlation with weather indices. Residual corn yields showed statistically significant correlation with each of the evaluated indices. MJJ precipitation showed only important correlation with three of the five indices and JJA surface temperature showed significant correlation with all five indices. The timing of maximum climate indices correlation with precipitation of MJJ and surface temperature of JJA often did not match the timing of maximum indices correlation with residual corn yield. The magnitudes and distributions of precipitation within a specified month can differ significantly within a county, leading in distinct yields that are not solved using yield information at county level. Nevertheless, the methods used can be helpful in the southeastern United States to forecast corn yields.

Chethana (2009) studied on prediction models in teak based agro forestry systems in northern transitional zone of Karnataka. Different prediction models namely Linear, Quadratic, Cubic, Exponential, Growth, Sigmoid, vapour pressure, Hassel, MMF compound, Logarithmic, Logistic, Weibull, Gompertz, Power etc., have been tried to predict the diameter and height growth of teak tree. MMF model was found better followed by Gompertz and Weibull for diameter prediction whereas for height growth prediction MMF model followed by logistic and Richards were found to be better.

Ranjana Agrawal and S.C. Mehta (2007) conducted a study on Weather Based Forecasting of Crop Yields, Pests and Diseases - IASRI Models. To sum up, IASRI created various kinds of models to forecast crop yields and warn pests and diseases. Model performance was found to be good. The models provided accurate crop yield forecasts well ahead of harvest and early warning of the severity of the pest population / disease could be given. Different other employees and organizations used the methodologies effectively. National Research Center for Rapeseed Mustard, Bharatpur used the models created for mustard to provide forewarning to the fanners for three years in a row, enabling them to optimize the resources of plant protection measures.

CHAPTER - III

MATERIALS AND METHODS

The aim of this chapter is to provide a brief description of the materials and methods which provide the necessary data base for this study and to highlight the important analytical tools and techniques employed in the analysis. The methodology is presented under the following headings:

3.1 Data Collection.

3.2 Software Used.

3.3 Analytical tools and techniques used.

3.4 Selection of Model.

3.1 Data Collection. The study was undertaken based on secondary data. To meet the objectives of this study, data on yield of total oilseeds which includes groundnut, rapeseed & mustard, soybean, sunflower, sesame, safflower, niger, castor and linseed in India for 48 years from 1970-71 to 2017-18 has been collected from Department of Agriculture and Economics, Ministry of Agriculture, Government of India and www.indiastat.com. Weather data were collected from India Meteorological Department (IMD), Pune and www.indiawaterportal.org from 1970-71 to 2017-18. The meteorological data on weather parameters namely; maximum temperature (Max.Temp, °C), minimum temperature (Min.Temp, °C), Rainfall (Rainfall, mm) and Relative Humidity (RH1 and RH2, %) were considered in this study.

3.2 Software Used: The statistical software's, viz. R-Software and Microsoft Excel, 2016, has been used for modelling and forecasting of oilseed yield of India.

3.2.1 R-Software: R is a command-driven statistical package available as open source software at <https://www.r-project.org/> for this reason, at first sight, this can make it rather daunting to use. However, there are several reasons to learn statistics using this computer software. The most important reasons may be stated as:

- R is free; anyone can download it and install it onto computer with Windows or Linux or Mac operating systems.
- R allows us to do all the statistical analyses, from simple to highly advanced ones, providing advanced scalability in level of analyses of data.
- An additional bonus is that R has excellent graphics and programing capabilities, so can be used as an aid to be teaching and learning.
- In addition to the features mentioned above, R has the capacity of developing packages for analyses in any field of exploratory analyses, theoretical or practical.
- R has been developed in the University of Auckland, in a university environment, as a project and is being groomed further in different universities environments, due to which this has been adopted and practiced in all the top universities of the world including Ivy Universities.

3.3 Analytical tools and techniques used:

The methodology for forecasting models viz. Multiple Linear Regression (MLR), LASSO Regression, models based on weather indices, models using composite weather variables and Artificial Neural Networks (ANN) and their performance evaluation criteria is also discussed in this section.

3.3.1 Multiple Linear Regression Model

Multiple regression analysis is to include several independent parameters at the same time for predicting the significance of a dependent parameter, (*Snedecor and Cochran, 1967*). In the study, the multiple linear regression equation fitted to the yield

treating as dependent variable and weather parameters treating as independent variables and the generalized form of the multiple regression model is given below.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Where:

Y = Dependent variable.

β_0 = Intercept.

β_i = regression coefficient of i^{th} independent parameters, ($i = 1, 2, \dots, n$).

ϵ = error term.

X_i = i^{th} weather parameter.

To identify the significant parameters for predicting the dependent parameter based on the independent parameters, stepwise regression analysis was used. Stepwise process starts with a simple regression model in which most extremely correlated one independent parameter was only incorporated at first in the company of a dependent parameter. Correlation coefficient is further examined in the practice to find an additional independent parameter that explains the major portion of the error remaining from the initial regression model. Until the model includes all the significant contributing parameters, the procedure keeps on repeating. The possible bias in the stepwise regression procedure fallout from the consideration of only one parameter at a time.

3.3.2 LASSO Stepwise Regression

LASSO (Least Absolute Shrinkage Selector Operator). LASSO regression analysis is a shrinkage and variable selection method for linear regression models. The goal of LASSO regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The LASSO does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients

variables are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical or both.

3.3.2.1 LASSO methodology.

Shrinkage methods aim to reduce (or shrink) the values of the coefficients to zero compared with Ordinary Least Squares (OLS). The advantage of shrinkage methods is that the estimated models exhibit less variance than least squares estimate. In addition, some shrinkage methods also reduce the number of covariates included in the regression model by yielding coefficient estimates exactly zero, facilitating the model selection process.

$$\text{LASSO Regression: } \min_{\beta_0, \beta_j} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad \text{subject to } \|\beta\|_1 \leq s,$$

where, y is the vector of observations of the independent variable, x denotes the covariates, β are the corresponding coefficients, $\|\cdot\|_1$ are the L_1 norms and s is a user-specified parameter. The langrangian formulation of the LASSO regression:

$$\text{LASSO Regression: } \min_{\beta_0, \beta_j} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \|\beta\|_1 \leq s,$$

The least squares estimation corresponds to an unconstrained minimization problem, the LASSO regression adds a convex but non-smooth ℓ_1 constraint. Least squares favor including as many covariates as possible since it helps reducing the sum of squares. In LASSO regression, non-null coefficients carry a penalty, which helps reducing the value of the LASSO coefficients or reducing the number of the covariates included in the LASSO model.

The following diagram is the visual interpretation comparing OLS and LASSO regression.

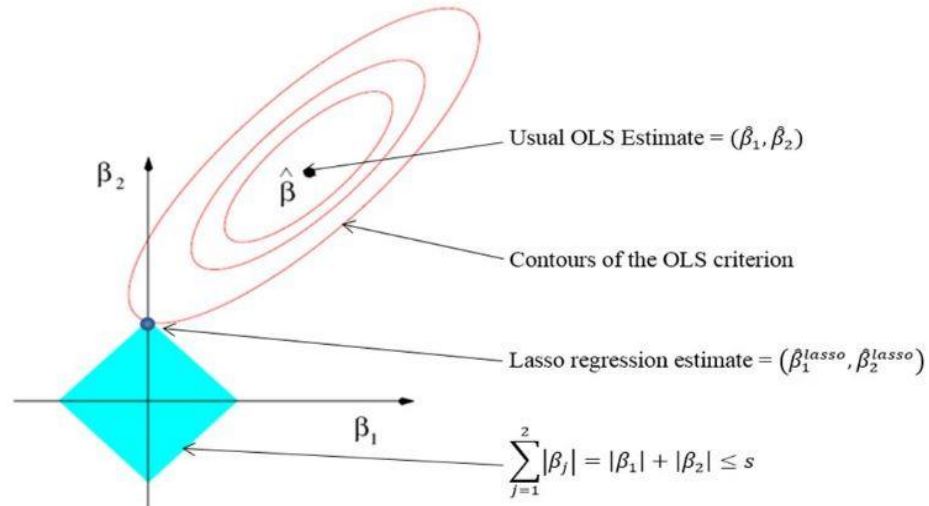


Fig: 3.3.2 Visual interpretation comparing OLS and LASSO regression.

Figure 3.3.2 illustrates the geometric intuition underlying the differences between least squares and the LASSO regressions, with the ellipses representing the contours of the residual sum of squares. The least squares coefficients, which correspond to the solution of the unconstrained optimization problem, are large relative to those produced by the shrinkage methods. In contrast, the presence of corners in the equation in figure constraint increases the chances the ellipses would intersect at the corner, yielding exactly zero coefficients and reducing the number of covariates included in the forecasting model. The LASSO regression, hence, appears well suited for addressing the model selection challenge posed by the forecasting requirements of yield. Note that the LASSO regression performs both the variable selection and parameter estimation simultaneously. Finally, compared with subset selection methods, LASSO exhibits lower variability and cheaper computational costs, especially for high dimensional problems.

3.3.2.2 LASSO Regression, Cross-Validation and Consistency

In Figure:3.3.2, the value of the user-specified parameter ‘ s ’ is small enough such that the unconstrained optimization solution is not contained within the regions bounded by the $L1$ constraints. Had this parameter been sufficiently large, the LASSO regression solutions would have coincided with the least square’s solution. In a LASSO regression, the value of the parameter controls both the size and the number of coefficients, with higher values leading to a greater number of covariates to be included in the linear model. This increases the flexibility of the model and reduces its variance but at the cost of a higher model bias.

Cross-validation, a resampling technique, helps to find a parameter value that ensures a proper balance between bias and variance (or flexibility and interpretability). Crossvalidation selects the best parameter value as the one that minimizes the estimated test error rate of the estimator, in this case the LASSO regression. In the absence of a test set, it is not possible to calculate the test error – the average error from using a method to predict the response of a new observation. In cross-validation a subset of the data observations, the training set (75% of dataset), is used to estimate or train the model, and the remainder observations are held to serve as test set or validation set (25% of dataset). The selected test sets serve to provide an estimate of the test error rate. Typically, the measure of the test error is the mean square error (MSE).

The K -fold cross-validation method divides the data set randomly into K different subsets, typically ten in the present study. Keeping one of the subsets as the validation set, the model is trained (estimated) over the remaining $K-1$ sets for a range of values of the parameter s (or λ , if the Langrangian formulation is used). We repeat this process using each of the K subsets as a validation set, yielding K estimates of the MSE for each parameter value, and its K -fold estimate is simply the average value of the K estimates.

The best parameter value is the one yielding the lowest K -fold estimate, which we denote as λ -min in the Langrangian formulation. It is also typical to report results corresponding to the minimum parameter value such that its K -fold estimate does not exceed the minimum K -fold estimate by more than one standard error using a lower

number of covariates. This parameter estimate is the one-standard error rule parameter, λ -1se. The Leave-One-Out cross-validation method is a special case of the K -fold cross-validation method. In this case, a sample containing N observations is partitioned exactly into N subsets, which is equivalent to N -fold cross-validation.

Since the LASSO biases the coefficients towards zero, the estimates are not consistent. There are ways to address this issue using two-step estimation procedures. In the first step, the LASSO regression selects the covariates in the model. In the second step, only the selected covariates are included in a linear model estimated by applying the LASSO.

3.3.3 Model based on weather indices

The present study uses model suggested by Hendricks and Scholl has been modified by Agarwal *et al.* 1980,1983 and Jain *et al.* 1980 by expressing effects of changes in weather variables on yield in the successive periods as second degree polynomial in respective correlation coefficients between yield and weather variables. This explains the relationship in a better way as it gives appropriate weightage to different periods. Under this assumption, the models were developed for studying the effects of weather variables on yield using complete crop data whereas forecast model utilized partial crop data.

These models were further modified by expressing the effects of changes in weather variables on yield in successive periods as a linear function of respective correlation coefficients between yield and weather variables. As trend effect on yield was found to be significant, its effect was 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 were also studied. The results indicated that (i) the models using correlation based on yield were better than the ones using simple correlations, (ii) inclusion of quadratic terms of weather variables and the second power of correlation coefficients did not improve the model. This suggests that the following models can be used to study effects of weather on yield and its forecasts.

$$Y = A_0 + a_0 z_0 + a_1 z_1 + cT + e$$

Where,

$$Z_j = \sum_{w=1}^n r_w^j X_w \quad ; \quad j=0, 1$$

Here, Y is yield r_w is correlation coefficient between the weather variable in w-th period (X_w) with yield, T is the year number and e is error term. The models were further extended for studying joint effects.

The forecast model has been developed using crop data considering all weather variables simultaneously. The model finally recommended was of the form

$$Y = A_0 + \sum_{i=1}^p \sum_{j=0}^1 a_{ij} Z_{ij} + \sum_{i \neq i'}^p \sum_{j=0}^1 a_{ii'j} Z_{ii'j} + cT + e$$

Where,

$$Z_{ij} = \sum_{w=1}^m r_{iw}^j X_{iw}$$

$$Z_{ii'j} = \sum r_{ii'w}^j X_{iw} X_{i'w} \quad j = 0, 1; i \neq i' = 1, 2, \dots, 5$$

$r_{iw}/r_{ii'w}$ is correlation coefficient of Y with i-th weather variable/product of i-th and i'-th weather variable in w-th period. m is period of forecast and p is number of weather variables used. the product of the two weather variables (X_i and $X_{i'}$) in w-th period. (Here $i = 1, 2, \dots, 5$ corresponding to maximum temperature, minimum temperature, rainfall, relative humidity morning and relative humidity evening).

In this approach, for each weather variable, two types of indices were developed, one as simple total values of weather variable in different periods [un-weighted index or simple weather indices - Z_{i0}] and the other one as weighted total [weighted weather index Z_{i1}] weights being correlation coefficients between yield and weather variable in respective periods. On similar lines, for studying joint effects, un-weighted & weighted indices for interactions were computed with products of weather variables (taken two at a time). Stepwise regression technique was used to select important indices in the model.

Table: 3.3.3 Notations for simple weather indices and weighted weather indices

Parameter	Simple weather indices					Weighted weather indices				
	Tmax	Tmin	Rf	RH1	RH2	Tmax	Tmin	Rf	RH1	RH2
Tmax	Z10					Z11				
Tmin	Z120	Z20				Z121	Z21			
Rf	Z130	Z230	Z30			Z131	Z231	Z31		
RH-I	Z140	Z240	Z340	Z40		Z141	Z241	Z341	Z41	
RH-II	Z150	Z250	Z350	Z450	Z50	Z151	Z251	Z351	Z451	Z51

3.3.4 Models using Composite Weather Variables

The Composite Weather Variable (CWV) is a single measure of daily weather in each Local Distribution Zone (LDZ) and is a function of actual temperature, wind speed, effective temperature and seasonal normal effective temperature. The definition of CWV includes provision for summer cut-offs and provision for cold weather upturn during low temperature extremes, defined such that a linear relationship applies between daily demand in the LDZ and the composite weather variable.

Data of different weather variables viz. Maximum temperature ($^{\circ}\text{C}$), minimum temperature ($^{\circ}\text{C}$), rainfall (mm) and relative humidity (%) were used to get weighted and un-weighted indices for regression analysis in the prediction of oilseeds yield. Yield forecast models have been worked out through stepwise regression method.

For each weather variable, two indices were worked out.

1. Un-weighted weather index = Sum (each weather variable)
2. Weighted weather index = Sum (each weather variable x correlation Coefficient between yield and weather variable).

Weather indices denoted as Z ; un-weighted indices are 0 and weighted indices are 1. For instance, maximum temperature taken as 1st variable, hence weather index of un-weighted maximum temperature is Z_{10} and for weighted Z_{11} . In the same way, other indices were worked out for other weather variables. To study the combined effect of weather variables, un-weighted and weighted indices were also computed. For instance, combination of maximum and minimum temperature is obtained by multiplying values of maximum and minimum temperature.

For selecting the best regression equation among number of independent variables, stepwise regression procedure was adopted.

The stepwise regression technique was used to find out the most significant variable. The multiple regression model was used based on yield as regressor and significant generated variable as regress and for the forecasting of yield. The model is given below.

$$Y = A_0 + \sum_{i=1}^p \sum_{j=0}^1 a_{ij} Z_{ij} + \sum_{i \neq i'} \sum_{j=0}^1 a_{ii'j} Z_{ii'j} + cT + e$$

where,

$$Z_{ij} = \sum_{w=1}^m r_{iw}^j X_{iw}$$

$$Z_{ii'j} = \sum r_{ii'w}^j X_{iw} X_{i'w}$$

Where,

$A_0, a_{ij}, a_{ii'j}$	constant
Z_{ij}	generated Variable (individual)
$Z_{ii'j}$	generated Variable (interaction term)
r_{iw}	is correlation coefficient of yield with i-th weather variable in w-th period
$r_{ii'w}$	is correlation coefficient (adjusted for trend effect) of yield with product of i-th and i'-th weather variables in w-th period
m	is period of forecast
p	is number of weather variables used
T	time trend
e	is random error distributed as $N(0, \sigma^2)$

Models using composite weather variables were developed using combination of simple and weighted weather indices as shown in table: 3.3.2 as independent variables while character under study such as yield was used as dependent variable. Stepwise regression technique has been used for selecting significant variables in the model.

3.3.5 Artificial Neural Networks (ANN)

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes. ANNs are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found. ANN takes data samples rather than entire data sets to arrive at solutions, which saves both time and money. ANNs are considered simple mathematical models to enhance existing data analysis technologies.

Artificial neural networks (ANNs) have recently received a great deal of attention in many fields of study, like engineering, medical science, and economics. Excitement stems from the fact that these networks are attempts to model capabilities of human brain which has approximately ten billion neurons acting in parallel. Neurons, basic computational unit of brain, are highly interconnected, with a typical neuron being connected to several thousand other neurons. As opposed to this, ANNs rarely have more than a few hundred or a few thousand neurons. So, networks comparable to a human brain in complexity are still far beyond the capability of the fastest, most highly parallel computers in existence.

An ANN is a set of simple computational units (also called nodes) that are highly interconnected. ANNs have been used for a wide variety of applications where statistical methods are traditionally employed. They have been used in classification problems, such as identifying underwater sonar currents, recognizing speech, and predicting heart problems in patients. In the present study, ANNs have been used in predicting total

oilseeds yield based on weather parameters in india. It is, therefore, time to recognize ANN as a powerful tool for data analysis.

The general form of an ANN is a “black box” model of a type that is often used to model high-dimensional, nonlinear data. However, most ANNs are used to solve prediction problems for some system, as opposed to formal model-building or development of underlying knowledge of how the system works.

3.3.5.1 ANN Modelling

The development and the training of the network model were carried out using Neural Network in R software. The ANN modeling comprises of two phases: the first phase is to train the network model, while the second phase is to test or validate the network model with new data which were not used for training. The flow of neural network modeling procedure is shown in figure 3.3.5.1 Choosing the optimum network architecture is one of the challenging steps in neural network modeling. Figure 3.3.5.2 shows the neural network architecture with back-propagation neural network model. The network has three layers: input layer, hidden layer and output layer. As there are five input and one output, the numbers of neurons in the input and output layer had to be set to 1 and 5, respectively. In many ANN applications, the propagation architecture with one hidden layer is enough. In order to find an optimal architecture, different numbers of neurons in the hidden layer were considered and prediction error for each network was calculated. The operation of the neural network model can be divided into two main steps: forward computing and backward learning. In the forward computing, the input patterns applied to the neurons of the first layer are just a stimulus to the network. In fact, there is no computation in the input layer. each neuron in the hidden layer determines a net input value based on all its input connections. These nodes are connected to each other so that the value of one node will affect the value of another. The relative influence that one node has on another one is specified by the “weight” that is assigned to each connection.

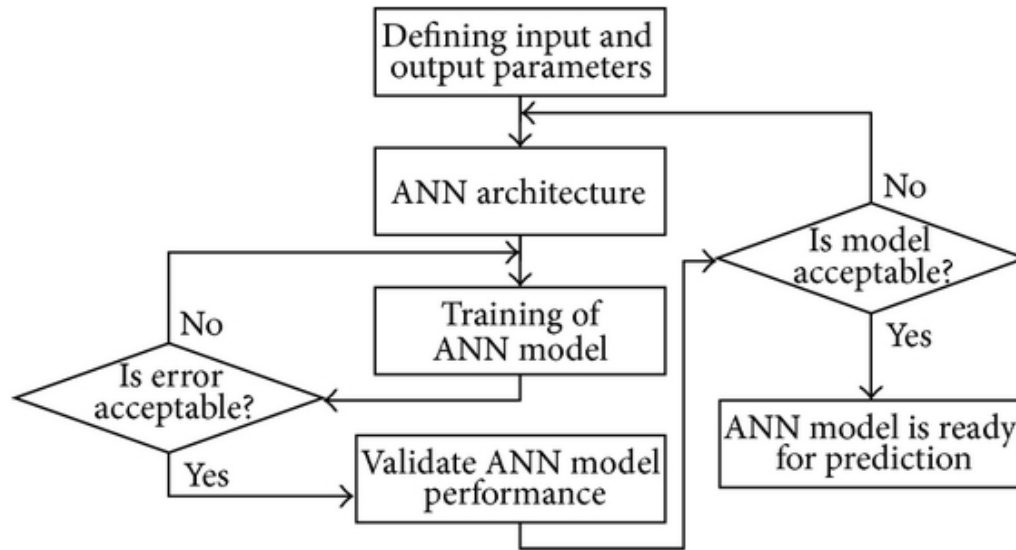


Fig: 3.3.5.1 Neural network modeling procedure.

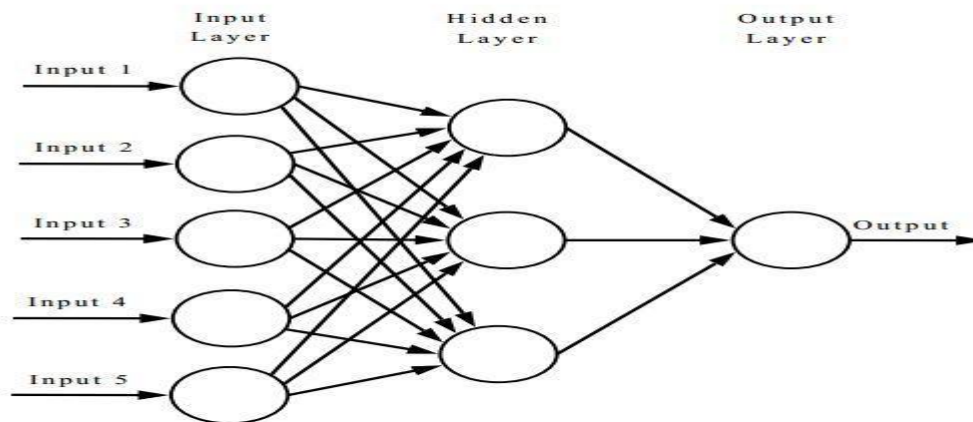


Fig: 3.3.5.2 Structure of basic artificial neural network model

In the figure: 3.3.5.2 above, each circular node represents an artificial neuron, and an arrow represents the input from one neuron output to another neuron. Here, the hidden layer is the inner processing unit layer of the neural network and the output layer is used to generate the output mode.

The net input is calculated by summing the input values multiplied by their corresponding weight. Once the net input is calculated, it is converted to an activation value. The weight on the connection from the i -th neuron in the backward layer to the j th neuron is indicated as w_{ij} . The output value of neuron j is computed by the following equation:

$$\mathbf{net}_j = \sum_{i=0}^n w_{ij} \mathbf{x}_i + \mathbf{x}_0, \quad \mathbf{Y}_j = \mathbf{f}_{\text{act}}(\mathbf{net}_j)$$

Where \mathbf{net}_j is the linear combination of each of the x_i values multiplied by w_{ij} , x_0 is a constant known as the bias, n is the number of inputs to the j -th neuron, and f_{act} is the activation of neuron j .

3.3.5.2 Back-Propagation algorithm

In backward learning, the generated output of the network is compared to the desired output, and an error is computed for each output neuron. The error vector E between desired values and the output value of the network is defined as:

$$\mathbf{E} = \sum \mathbf{E}_j = \sum 1/2 (\mathbf{T}_j - \mathbf{Y}_j)^2$$

Where Y_j is the output value of the j^{th} output neuron and T_j is the desired value of the j^{th} output neuron. Errors are then transmitted backward from the output layer to each neuron in the forward layer. The process is repeated layer by layer. Connection weights are updated by each neuron to cause the network to converge. The network was trained with the Levenberg-Marquardt training algorithm. This training algorithm was chosen due to its high accuracy in similar function approximation. The adjustments of weights and biases are done according to transfer following function:

$$\Delta \mathbf{w}_{ij} = - (\mathbf{J}^T \mathbf{J} + \mu \mathbf{I})^{-1} \mathbf{J}^T \mathbf{E}$$

Where J is Jacobian matrix of derivation of each error, μ is a scalar and E is error function.

The input and output parameters data sets could not be trained by neural network in their original form due to the wide range of values among them. In order to become feasible neurons, all the values in the input neurons had to be preprocessed by normalizing and transformed within the range of -1 and +1.

3.3.5.3 ANN Illustration

In the present study, we have used three layers namely input layer, hidden layer and an output layer depending on the function and each layer is functionally connected. Each layer consists of neurons or nodes interconnected with each other. The number of nodes in input is fixed by the dataset used. The input layer connects the external input mode and is transmitted in units of hidden layers according to the input unit. The main problem in the implementation of ANN is to find the optimum number of hidden nodes by 'train' function of the 'caret' package using the method 'nnet' in R software. In the present study we use yield as the dependent variable and weather parameters namely maximum temperature, minimum temperature and rainfall as the independent variables. We divide the data into training and test set. Training set is used to find the relationship between dependent and independent variables while the test set assesses the performance of the model. In the present study uses 75% of the dataset to develop the model and 25% of the dataset to validate the developed model.

3.4 Selection of Model

The R^2 , MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) have been widely used as a criterion for model evaluation.

3.4.1 Coefficient of Determination (R^2 criteria)

R^2 is a statistic that gives some information on overall variation explained by the model out of total variation. It is also known as coefficient of multiple determination and it has been used by many researchers for model selection. Mathematically,

$$R^2 = 1 - \frac{\text{Error sum of squares}}{\text{Total sum of squares}} = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2}$$

3.4.2 Root Mean Square Error (RMSE)

It is the standard deviation of the residuals (prediction errors). The RMSE tells you how concentrated the data is around the line of best fit. It has also been used for model selection. The lesser value of RMSE the better model fit is. Mathematically,

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

3.4.4 Mean Absolute Percentage Error (MAPE)

It is a measure of prediction accuracy of a forecasting method in statistics. It expresses the mean of the percentage errors over different points of time, and is mathematically expressed as:

$$\text{MAPE} = \frac{100}{N} \times \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

R^2 values close to 1 and RMSE values close to 0 indicate better model performance.

CHAPTER – IV

RESULTS AND DISCUSSION

This chapter includes the discussion of the results obtained from the analysis of the data to achieve the objectives. The results obtained from the present investigation have been divided into following sections.

- 4.1 Collection of time series of both crop yield and meteorological data.
- 4.2 Determine the statistical relationship between weather parameters and yield.
- 4.3 Develop prediction models to predict yield from climate variables over the area of study.
- 4.4 Validation of the developed models.

4.1 Collection of time series of both crop yield and meteorological data.

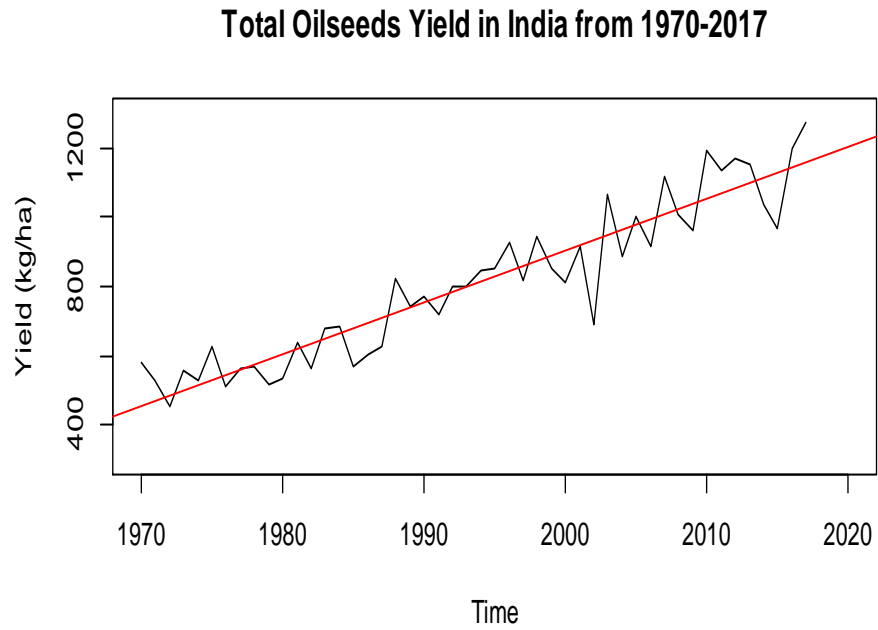
4.1.1 Time Series of Crop Yield

4.1.1.1 Results From Analysis Of The Trends In Oilseeds Yield.

Nine major oilseeds *viz.*, groundnut, soybean, rapeseed-mustard, sesamum, sunflower, safflower, linseed, niger and castor together contribute to the total productivity of oilseeds in India. In 1970-71 total yield under oilseeds was 579 kg/ha whereas, in 2017-18 the yield under oilseeds in India increased to 1284 kg/ha. Over the past few decades productivity of oilseeds in India has shown a continuously increasing trend. The time series of the annual oilseeds yield is shown in the Fig. 4.1.1 there is a general increasing trend in the oilseeds yield, which is an indication of improving yield over time. From the Table: 4.1.1, The results show that the mean of annual oilseeds yield in India was about 806.15 kg/ha with a standard error of 32.26. The skewness of the data is positive (0.317), meaning that the distribution is positively skewed.

Table: 4.1.1 Descriptive statistics of total oilseeds yield in India from 1970-2017

Mean	S. E	Median	Skewness	95% Conf. Interval	
806.15	32.26	804.5	0.317	871.047	741.253

**Fig. 4.1.1 Time series of total oilseeds yield in India from 1970 to 2017.**

4.1.2 Time Series of Meteorological Parameters

4.1.2.1 Results From Analysis Of The Trends In Maximum Temperature.

The time series of annual maximum temperature is shown in the Fig. 4.1.2.1. The figure indicates that the annual maximum temperature in India showed significant increasing trend of 0.16, 0.11 and 0.14 °C per decade since 1981. Maximum warming trend is seen during post monsoon season. In this figure the red line depicts the linear trend based on 1970-2017 are statistically significant with magnitude of 0.4655. Annual Maximum Temperatures averaged over India shows significant warming trend. The descriptive statistics of annual maximum temperature is shown in the table 4.1.2.1 The results show that the mean of annual maximum temperature in India was about 29.65 °C with a standard error of 0.07. The table indicates that by considering skewness of the data, the maximum temperature is positively skewed (2.01) and it is greater than + 1 describes that the distribution of annual maximum temperature is highly skewed.

Table: 4.1.2.1 Descriptive statistics of annual maximum temperature in India from 1970 - 2017

Mean	S. E	Median	Skewness	95% Conf. Interval	
29.65	0.07	29.59	2.01	29.80	29.50

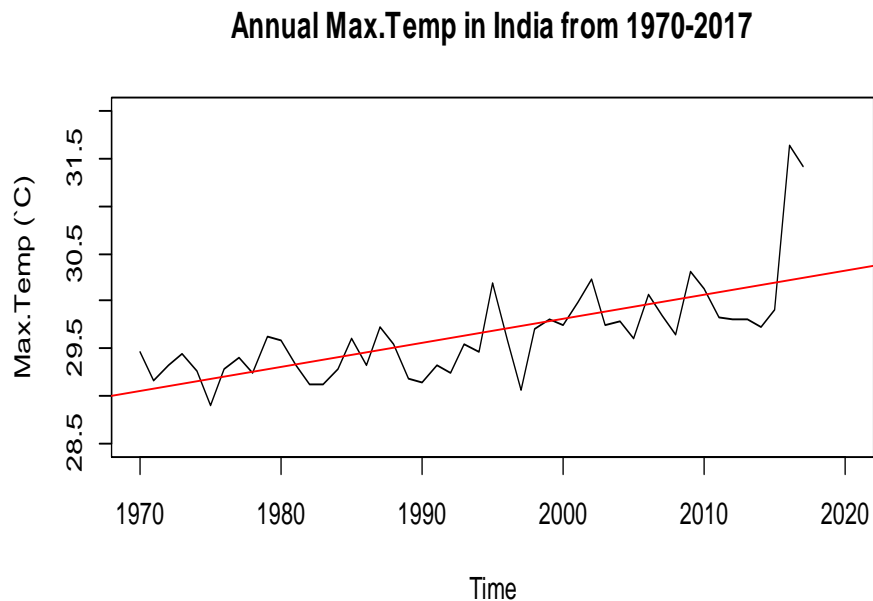


Fig. 4.1.2.1 Time series of Annual Max. Temp in India from 1970 to 2017.

4.1.2.2 Results From Analysis Of The Trends In Minimum Temperature.

The time series of the annual minimum temperature is shown in Fig. 4.1.2.2 as the figure indicates that the annual minimum temperature in India showed significant increasing trend. Maximum warming trend is seen during post monsoon season. In this figure the red line depicts the linear trend based on 1970 – 2017 data are statistically significant with magnitude of 0.5509. Annual minimum temperatures averaged over India shows significant warming trend. The descriptive statistics of the annual minimum temperature is presented in the Table: 4.1.2.2 The results show that the mean annual minimum temperature in India was about 19.91 °C with a standard error of 0.07. The skewness is positive (1.48), the data are positively skewed or skewed right, meaning that the right tail of the distribution is longer than the left tail and in the present study the skewness is greater than +1, meaning that the distribution is highly skewed.

Table: 4.1.2.2 Descriptive statistics of annual minimum temperature in India from 1970 - 2017.

Mean	S. E	Median	Skewness	95% Conf. Interval	
19.51	0.07	19.48	1.48	19.66	19.36

Annual Min.Temp in India from 1970-2017

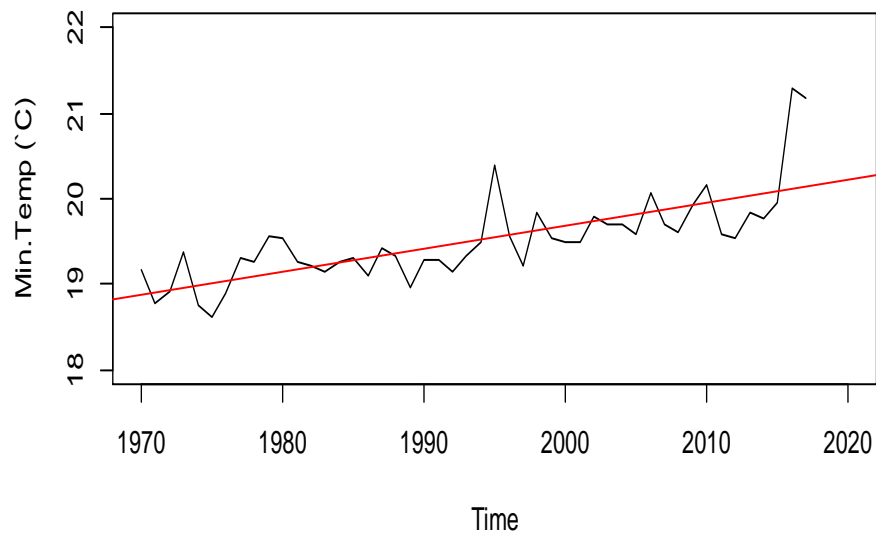


Fig. 4.1.2.2 Time series of Annual Min. Temp in India from 1970 to 2017.

4.1.2.3 Results From Analysis Of The Trends In Annual Rainfall Of India.

The Time series of annual rainfall of overall India is shown in the 4.1.2.3 It represents the annual rainfall in India shows significant decreasing trend (Negative trend). The Table: 4.1.2.3 represents the descriptive statistics of annual rainfall in India. The mean annual rainfall of the of the study period was 1155.22mm with standard error of 14.06. The minimum and maximum ever recorded rainfalls were 947.1 mm in 1972 (the driest year) and 1294.8 in 1975 (the wettest year) per year respectively.

Table: 4.1.2.3 Descriptive statistics of annual rainfall in India from 1970-2017

Mean	S. E	Median	Skewness	95% Conf. Interval	
1155.22	14.06	1157.50	0.13	1183.51	1126.93

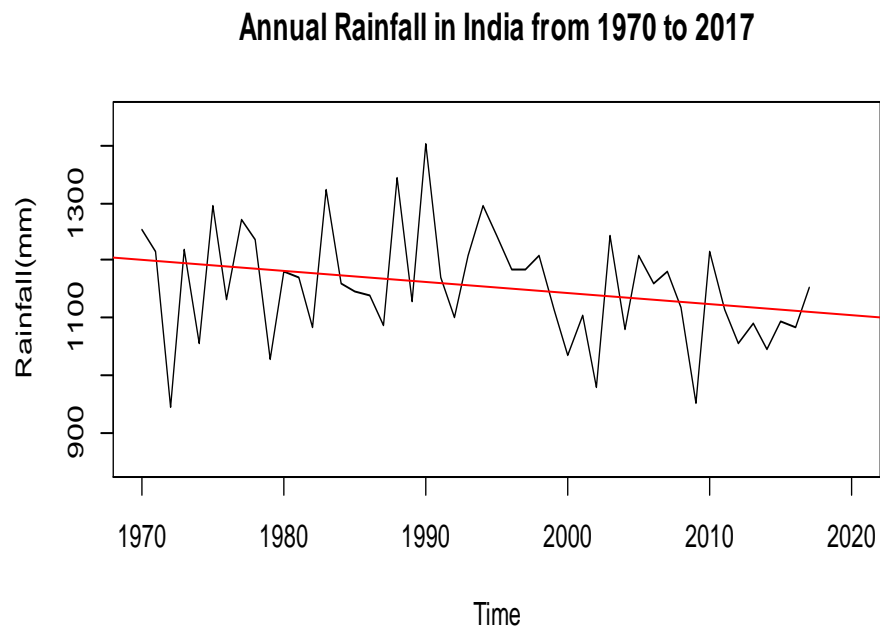


Fig. 4.1.2.3.1 Time series of Annual Rainfall in India from 1970 to 2017.

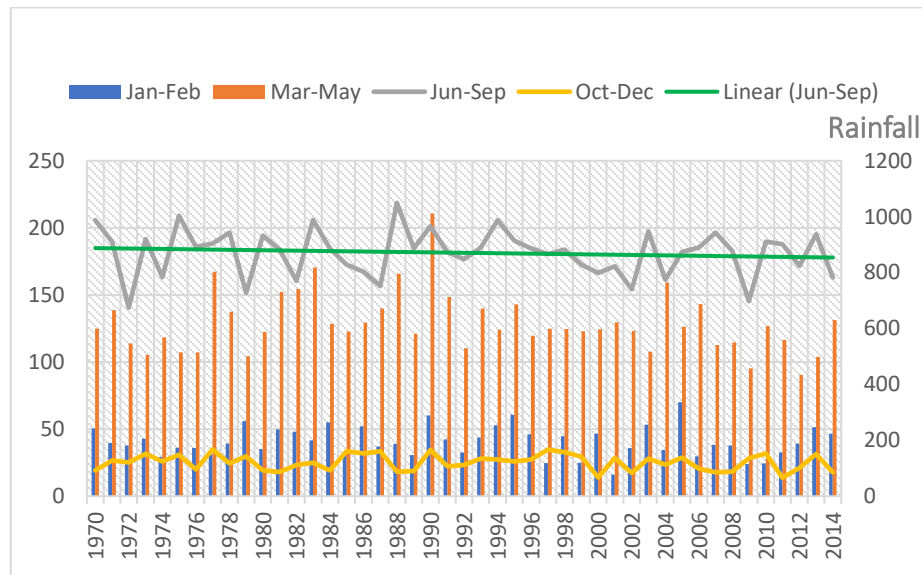


Fig. 4.1.2.3.2 Seasonal rainfall in India from 1970 to 2017.

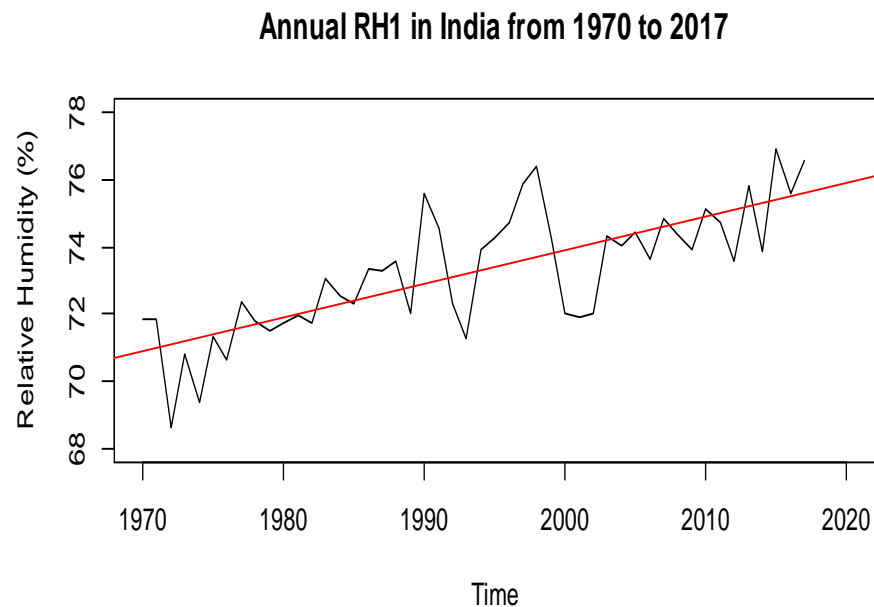
The Fig.4.1.2.3.1 describes that the annual rainfall in India showed significant decreasing trend. India receives 75 % of Annual rainfall from June –sept summer monsoon season. Annual as well as seasonal (June-sept) monsoon rainfall over India showed significant decreasing trend as shown in Fig.4.1.2.3.2

4.1.2.4 Results From Analysis Of The Trends In Annual RH1 Of India.

The time series of the annual morning relative humidity is shown in Fig. 4.1.2.4 As the figure indicates that the annual morning relative humidity in India showed significant increasing trend. In this figure the red line depicts the linear trend based on 1970-2017 are statistically significant. Annual RH1 averaged over India shows significant warming trend. The descriptive statistics of the annual RH1 is presented in the Table: 4.1.2.4 The results show that the mean of annual RH1 in India was about 73.25 percent with a standard error of 0.27. The skewness of the data is negative (-0.14), meaning that the distribution is negatively skewed.

Table: 4.1.2.4 Descriptive statistics of annual RH1 in India from 1970-2017

Mean	S. E	Median	Skewness	95% Conf. Interval	
73.25	0.27	73.45	-0.14	73.75	72.72

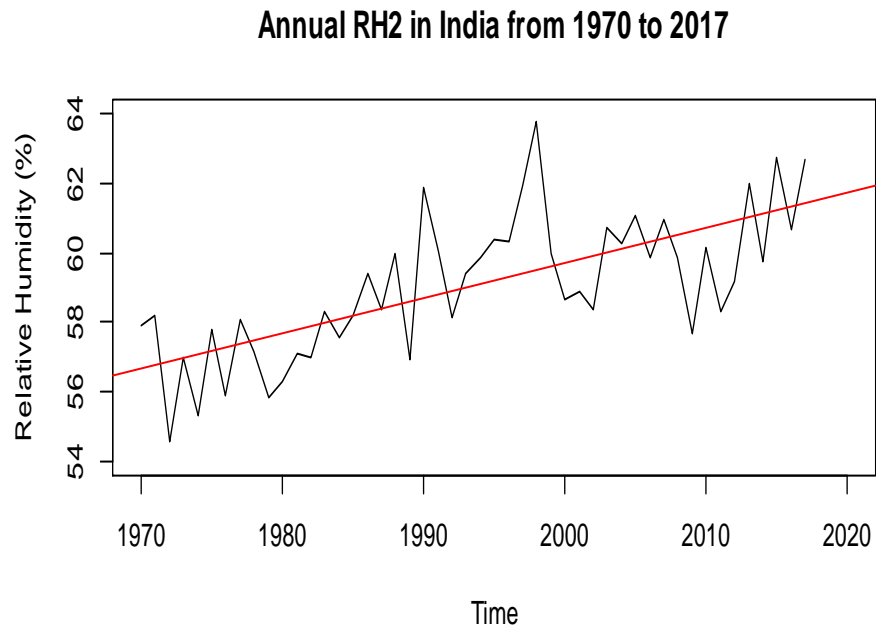
**Fig. 4.1.2.4 Time series of Annual RH1 in India from 1970 to 2017.**

4.1.2.5 Results From Analysis Of The Trends In Annual RH2 Of India.

The time series of the Annual evening relative humidity is shown in Fig. 4.1.2.5 As the figure indicates that the annual RH2 in India showed significant increasing trend. In this figure the red line depicts the linear trend based on 1970-2017 are statistically significant. Annual RH2 averaged over India shows significant warming trend. The descriptive statistics of the annual evening relative humidity is presented in the Table: 4.1.2.5 The results show that the mean of annual RH2 in India was about 59.04 percent with a standard error of 0.29. The skewness of the data is positive (0.07), meaning that the distribution is positively skewed.

Table: 4.1.2.5 Descriptive statistics of annual RH2 in India from 1970-2017

Mean	S. E	Median	Skewness	95% Conf. Interval	
59.04	0.29	59.03	0.07	59.63	58.45

**Fig. 4.1.2.5 Time series of annual RH2 in India from 1970 to 2017.**

4.2 Determine the statistical relationship between weather parameters and yield.

The results from the analysis of the relationship between the weather parameters and yield are discussed in this section. The knowledge regarding association of various parameters among themselves is necessary to know the nature and degree of relationship.

4.2.1 Results From The Correlation Analysis

To determine the relationship between yield and weather parameters, correlation analysis was carried out. Correlation studies pave way to know the association prevailing between crop yield and weather parameters which was given better understanding of the contribution of each weather parameter to the crop yield, with this view the yield and weather parameters were subjected to correlation analysis and the results were discussed. Correlation analysis of yield against maximum temperature, minimum temperature, rainfall, morning relative humidity and evening relative humidity are shown in the table:4.2.1. The correlation coefficients were computed between the oilseeds yield and the weather parameters of India. From the table, an attempt on correlating the yield with maximum temperature, gained significant result *i.e.* The correlation coefficient (r) was found to be 0.654156 indicated that the yield was significantly and positively correlated with maximum temperature. Correlating the yield with minimum temperature, gained significant result *i.e.* The correlation coefficient (r) was found to be 0.721703 indicated that the minimum temperature showed significant positive correlation with yield. The correlation between yield and rainfall gained, non-significant results *i.e.* The correlation coefficient (r) was found to be -0.06537 indicates that the rainfall having negative correlation with yield. Correlating the yield with relative humidity morning, gained significant result *i.e.* The correlation coefficient (r) was found to be 0.7773441 indicated that the RH1 showed significant positive correlation with yield. Correlating the yield with relative humidity evening, gained significant result *i.e.* The correlation coefficient (r) was found to be 0.723408 indicated

that the RH2 showed significant positive correlation with yield. The highest correlation values were observed with the relative humidity at evening, while the rainfall showed the least correlation values.

Table: 4.2.1 Correlation between weather parameters and yield.

	MaxTemp	MinTemp	Rainfall	RH1	RH2
Yield	0.6537**	0.7221**	-0.0697	0.7720**	0.7225**

** Correlation is significant at the 0.01 level

4.2.2 Inter correlation between weather parameters.

The inter correlation coefficient between different weather parameters based in 48 years data set is computed and presented in the in table 4.2.2 it was observed that maximum temperature is positively correlated with morning relative humidity and evening humidity but highly positively correlated with minimum temperature. Morning relative humidity is also highly positively correlated with evening relative humidity. There is a negative correlation between maximum temperature and minimum temperature while rainfall is positively correlated with morning and evening relative humidity. It can be observed that there is highest correlation between maximum and minimum temperature and lowest correlation among rainfall and minimum temperature.

Table: 4.2.2 Inter correlation between weather parameters

	MaxTemp	MinTemp	Rainfall	RH1	RH2
MaxTemp	1				
MinTemp	0.936418767	1			
Rainfall	-0.304227278	-0.149444672	1		
RH1	0.466616175	0.644916058	0.192472572	1	
RH2	0.400415174	0.565960059	0.287880594	0.915805325	1

4.2.3 Nature and relationship between individual weather variables and yield.

The following scatter diagrams (scatterplots) showing how yields are vary with independent variables *i.e.* weather parameters. Scatterplots are used to present relationships between quantitative variables when the variable on the x -axis (typically the independent variable) has a large number of levels. Each point in a scatterplot represents an individual rather than the mean for a group of individuals, and there are no lines connecting the points. The graphs below is an APA-style version, which illustrates a few additional points. First, when the variables on the x -axis and y -axis are conceptually similar and measured on the same scale as here, where they are measures of the same variable on two different occasions this can be emphasized by making the axes the same length. Second, when two or more individuals fall at the same point on the graph, one way this can be indicated is by offsetting the points slightly along the x -axis. Other ways are by displaying the number of individuals in parentheses next to the point or by making the point larger or darker in proportion to the number of individuals. Finally, the straight line that best fits the points in the scatterplot, which is called the regression line, can also be included. The Fig.4.2.3.1 displays a scatter diagram showing how oilseeds total yield is varying with maximum temperature. At the end yield increases with an increasing temperature. Fig.4.2.3.2 displays a scatter diagram showing how oilseeds total yield is varying with minimum temperature. It shows that yield increases with an increasing temperature. It was noted that yield has increased over most recent years despite increasing temperature. The Fig. 4.2.3.3 displays a scatter diagram showing how oilseeds total yield is varying with rainfall. It shows that there is an inverse relationship between oilseeds yield and rainfall in India. Fig.4.2.3.4 displays a scatter diagram showing how oilseeds total yield is varying with morning relative humidity. It shows that yield increases with an increase in humidity. Fig.4.2.3.5 displays a scatter diagram showing how oilseeds total yield is varying with evening relative humidity. It shows that yield increases with an increasing humidity.

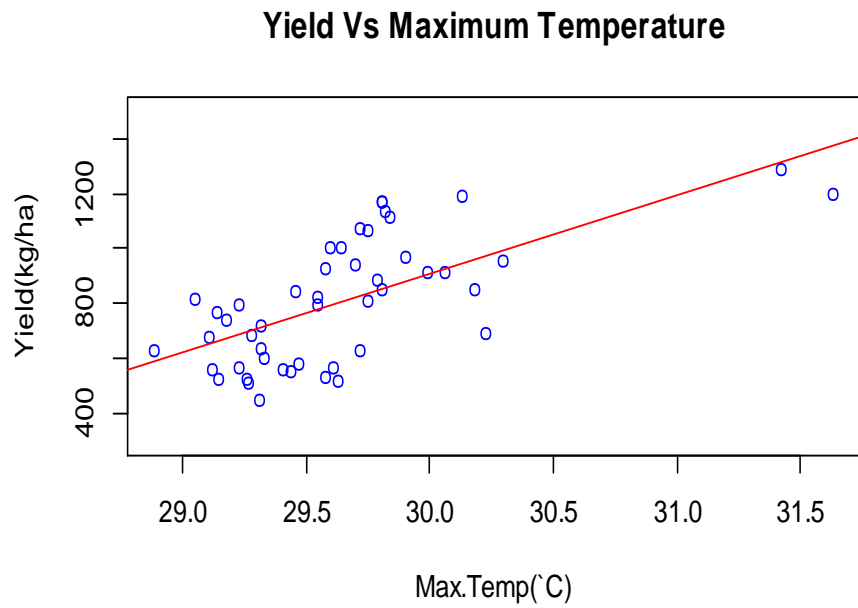


Fig.4.2.3.1 Scatterplot of yield versus maximum temperature.

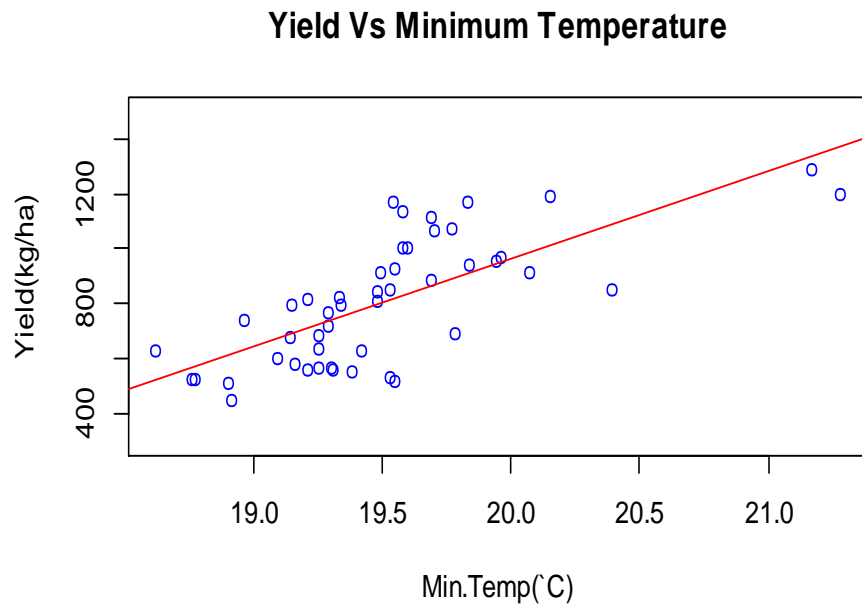
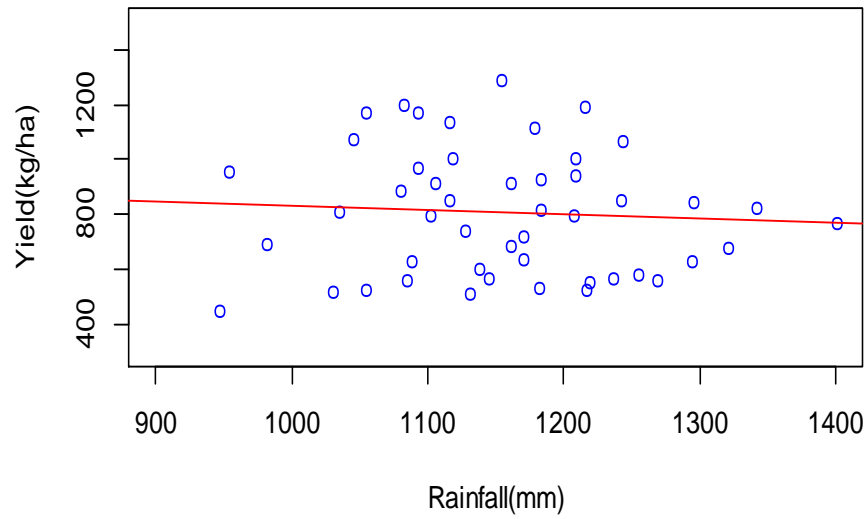
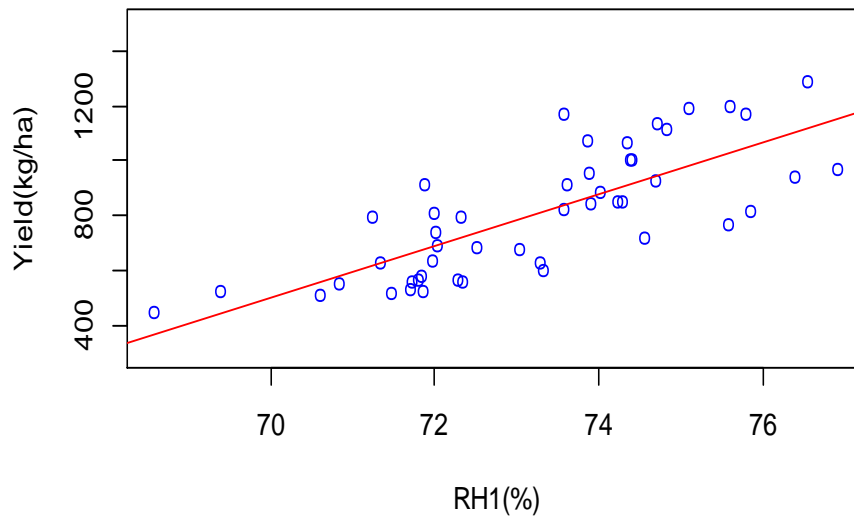


Fig.4.2.2.2 Scatterplot of yield versus minimum temperature.

Yield Vs Rainfall**Fig.4.2.2.3 Scatterplot of yield verses rainfall.****Yield Vs RH1****Fig.4.2.2.4 Scatterplot of yield verses RH1.**

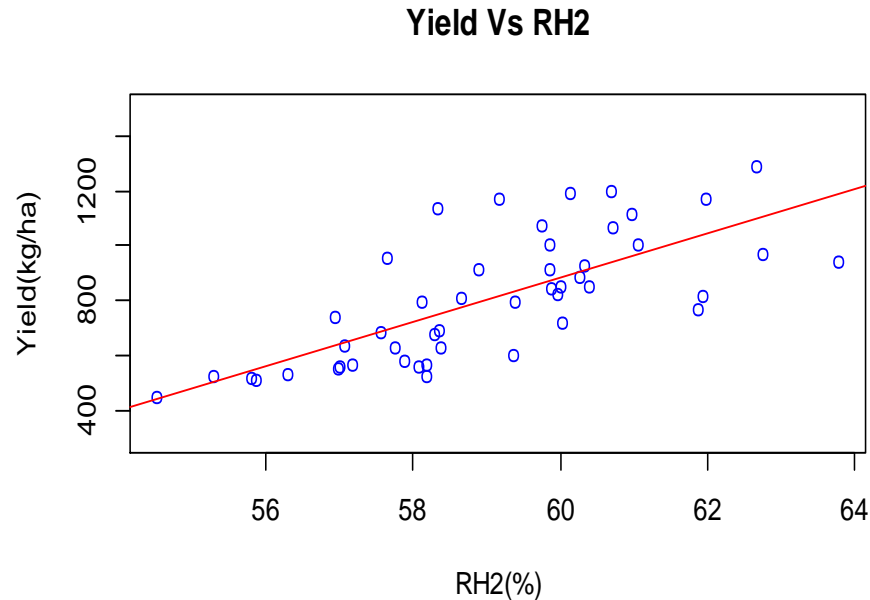


Fig.4.2.2.5 Scatterplot of yield verses RH2.

4.3 Develop prediction models to predict yield from climate variables over the area of study.

4.3.1 Multiple Linear Regression Model

The multiple linear regression model is fitted to predict the total yield as dependent variable taking the other independent variables as weather parameters viz., maximum temperature, minimum temperature, rainfall, morning relative humidity and evening relative humidity. From the results, we can determine that the intercept is -7220.278, the coefficient for the maximum temperature is 67.823, the coefficient for minimum temperature is -90.529, coefficient for the rainfall is -0.185, coefficient for the morning relative humidity is 38.008 and coefficient for the evening relative humidity is 28.446 and the best fitted model is presented.

The best fitted multiple linear regression model is given below:

$$\text{Yield} = -7220.278 + 67.823 \cdot X_1 - 90.529 \cdot X_2 - 0.185 \cdot X_3 + 38.008 \cdot X_4 + 28.446 \cdot X_5$$

Where X_1 , X_2 , X_3 , X_4 and X_5 represents Maximum Temperature, Minimum Temperature, Rainfall, Relative Humidity (RH1) and Relative Humidity (RH2) respectively.

Table: 4.3.1 Summary Of Multiple Linear Regression Model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0.865	0.748	0.718	119.866

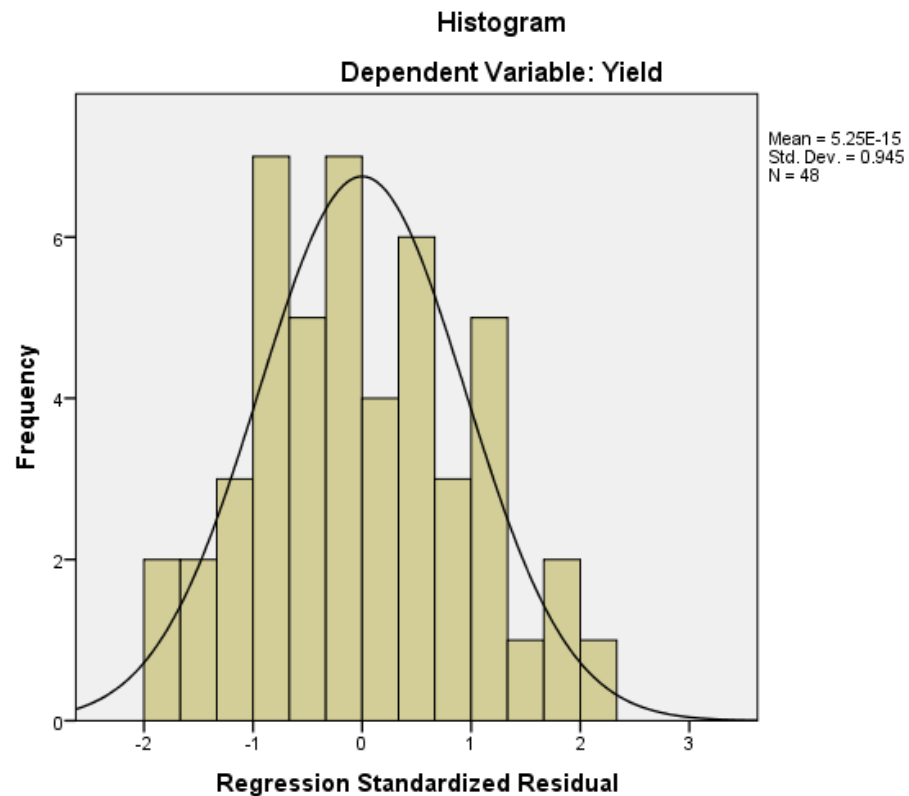
Table: 4.3.2 ANOVA of the developed model

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	1786981.210	5	357396.242	24.875	.000
Residual	603450.707	42	14367.874		
Total	2390431.917	47			

Table: 4.3.3 Coefficients of the developed model

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	-7220.278	1063.458		-6.789	0.000
MaxTemp	67.823	51.570	0.191	1.315	0.196
MinTemp	90.529	61.740	0.234	1.466	0.150
Rainfall	-0.185	0.217	-0.080	-0.852	0.399
RH1	38.008	22.733	0.320	1.672	0.102
RH2	28.446	19.894	0.261	1.430	0.160

The multiple regression model was developed by yield taken as dependent variable and weather parameters as independent variables.

**Fig: 4.3.1 Histogram for the multiple linear regression model**

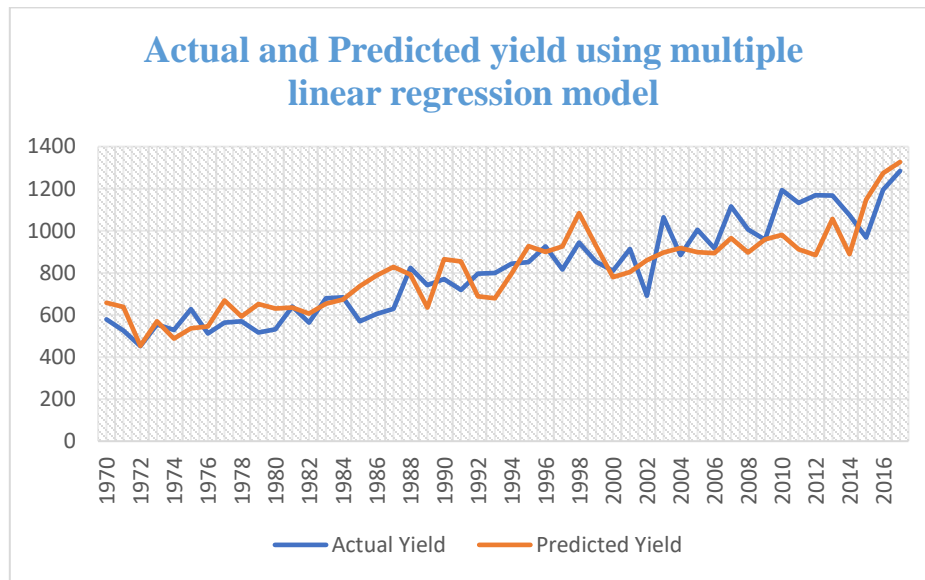


Fig. 4.3.2 Actual and Predicted yield using regression model

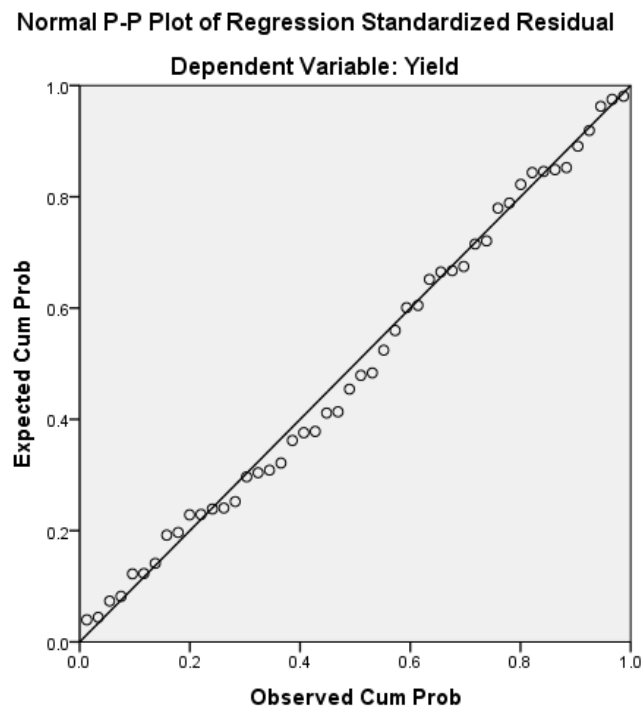


Fig. 4.3.3 Normal P-P Plot of Regression Standardized Residuals

4.3.2 LASSO Stepwise Regression.

The standard linear model (or the ordinary least squares method) performs poorly in a situation, where you have a large multivariate data set containing several variables superior to the number of samples. A better alternative is the penalized regression allowing to create a linear regression model that is penalized, for having too many variables in the model, by adding a constraint in the equation (James *et al.* 2014, P. Bruce and Bruce (2017)). This is also known as shrinkage or regularization methods. The consequence of imposing this penalty, is to reduce (i.e. shrink) the coefficient values towards zero. This allows the less contributive variables to have a coefficient close to zero or equal zero. The shrinkage requires the selection of a tuning parameter (λ) that determines the amount of shrinkage.

The present study describes the most commonly used penalized regression method called LASSO regression. It shrinks the regression coefficients toward zero by penalizing the regression model with a penalty term called **L1-norm**, which is the sum of the absolute coefficients. In the case of LASSO regression, the penalty has the effect of forcing some of the coefficient estimates, with a minor contribution to the model, to be exactly equal to zero. This means that, LASSO can also be seen as an alternative to the subset selection methods for performing variable selection in order to reduce the complexity of the model.

Selecting a good value of λ for the LASSO is critical. One obvious advantage of LASSO regression is that it produces simpler and more interpretable models that incorporate only a reduced set of the predictors.

Generally, LASSO might perform better in a situation where some of the predictors have large coefficients, and the remaining predictors have very small coefficients. Cross-validation methods can be used for identifying which of these two techniques is better on a particular data set.

4.3.2.1 Preparing the data

In the present study uses the total oilseeds yield in India of 48 years and annual data of weather parameters viz., maximum temperature, minimum temperature, rainfall and relative humidity (RH-I and RH-II) dataset from 1970-71 to 2017-18. Split the data into training set (75% for building a predictive model) and test set (25% for validating the model).

4.3.2.2 Computation of LASSO regression

It shrinks the regression coefficients toward zero by penalizing the regression model with a penalty term called **L1-norm**, which is the sum of the absolute coefficients. The penalty term has the effect of forcing some of the coefficient estimates, with a minor contribution to the model, to be exactly equal to zero. Generally, LASSO might perform better in a situation where some of the predictors have large coefficients, and the remaining predictors have very small coefficients. Different values of alpha return different estimators, alpha=1 is the LASSO. To develop the LASSO regression model, the present study uses yield as the dependent variable and maximum temperature, minimum temperature, rainfall, RH1 and RH2 as independent variables. Divide the data into training and test set. Training set is used to find the relationship between dependent and independent variables while the test set assesses the performance of the model and used 75% of the dataset as training set and 25% of the dataset to validate the model.

The results of the analysis are shown in figure: 4.3.2.2 below. On the x- axis the different values of lambda are shown. Each line represents one of the explanatory variables (weather variables) and its role in the model. In the plots we can see when each variable entered in the model and to which extent, they influenced the response variable yield. Analyzing the plot, we can say that the variable that most influence the model is maximum temperature, because it enters the model first and steadily positively affect the yield. The second most important variable is RH2 that enters later in the model and affect the model positively and also affect the trend of maximum temperature variable once it enters in the model. Furthermore, we can select other important

variables, looking at their trends such as minimum temperature and RH1 and rainfall seems to be less significant for the model.

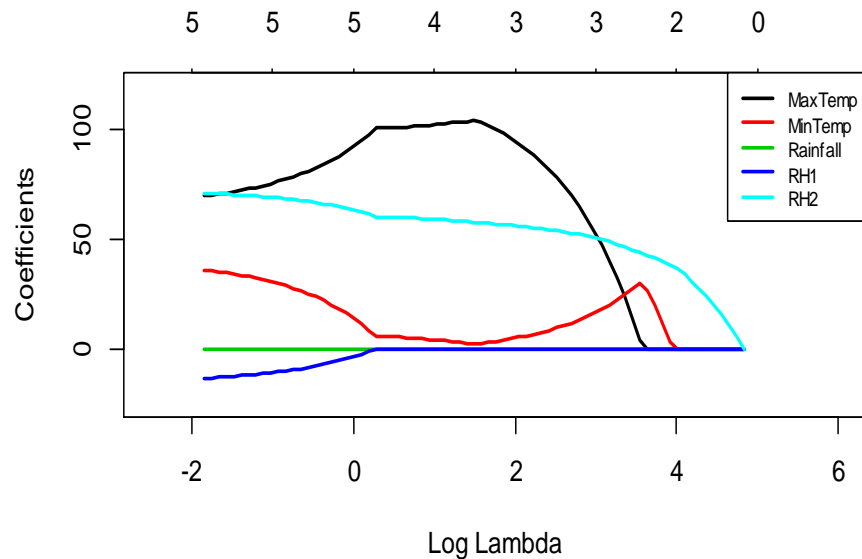


Figure.4.3.2.2 LASSO Coefficients.

To interpret this plot, recall the optimization problem the LASSO solves

$$\hat{\beta}^{\text{LASSO}} = \min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_j |\beta_j|$$

so λ is the penalty or the Lagrange multiplier if you prefer and is always positive. Setting $\lambda=0$ yields the familiar minimization of squared residuals while for greater values, some of the coefficients will be set to zero. As $\lambda \rightarrow \infty$, all the coefficients will be set to zero.

The next step is to choose the which value of lambda to consider. Even if for this small dataset we will not get very informative results, we can still use the cross validation.

4.3.2.3 Finding the optimal value of lambda that minimizes the cross-validation error

In the present study it was found that all the values of lambda are positive so, there is a maximum shrinkage and now we have to select one of the values of lambda and visualize the inspection is good enough. We do this by the crossvalidation function of the fitted model by using 'cv.glmnet' package in R- software. It is a predictive criterion that evaluates the sample performance by splitting the sample into training and validation sets and choosing the value of lambda with which the error of prediction is minimum.

In order to choose the most appropriate value for lambda the LASSO method extracts different values for lambda, such as lambda minimum (λ_{\min}) that is first vertical dotted line in the figure:4.3.2.3 that gives the minimum mean cross-validated error and lambda lse (λ_{lse}) second vertical dotted line in the same figure, that gives a model such that error is within four standard of the minimum.

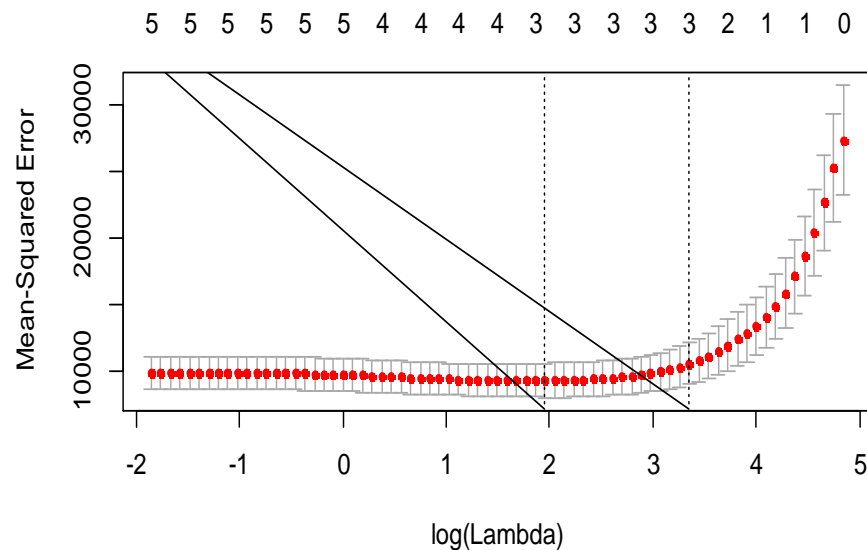


Figure.4.3.2.3 Cross-validation error according to the log of lambda.

The Figure.4.3.2.3 displays the cross-validation error according to the log of lambda. The left dashed vertical line indicates that the log of the optimal value of lambda is approximately two, which is the one that minimizes the prediction error. This lambda value will give the most accurate model. The exact value of lambda minimum in the present study is 7.088

Generally, the purpose of regularization is to balance accuracy and simplicity. This means, a model with the smallest number of predictors that also gives a good accuracy. To this end, the function `cv.glmnet` finds also the value of lambda that gives the simplest model but also lies within one standard error of the optimal value of lambda. This value is called `lambda.1se.` and is about 28.61 in our study.

In the present study, using `lambda.min` value as the best lambda, gives the following regression coefficients and to get the list of important variables we just need to investigate the coefficients of final best model and the results are presented in the table below.

Table:4.3.2.3 Regression coefficients using lambda.min as best lambda

(Intercept)	-5501.48
MaxTemp	95.86
MinTemp	4.59
Rainfall	.
RH1	.
RH2	56.26

Using lambda.lse value as the best lambda, gives the following regression coefficients:

Table:4.3.2.4 Regression coefficients using lambda.lse as best lambda

(Intercept)	-3189.18
MaxTemp	23.66
MinTemp	3.88
Rainfall	.
RH1	.
RH2	56.29

Using lambda.min as best lambda, the results from the table 4.3.2.3 shows that only three variables have non-zero coefficients. The coefficients of other variables i.e. rainfall and RH1 is set to zero by the LASSO algorithm, reducing the complexity of the model. Using lambda.lse as best lambda value, the results shows that rainfall and RH1 is set to zero by the LASSO algorithm, but it is less accurate compared to lambda.min. Setting lambda=lambda.min produces a simpler model compared to lambda.lse, but the model might be a little more accurate than the one obtained with lambda.lse.

4.3.2.4 Computing the final LASSO model

The present study computes the final LASSO model using lambda.min value as the best lambda and it is found to be 7.088 in the present study with percentage deviation of 0.6921 and the best fitted LASSO regression model is presented.

Table: 4.3.2.5 LASSO model for yield forecasting.

Model	LASSO Equation
LASSO	Yield = -5218.285+ 71.615* Max.Temp + 25.552* Min.Temp + 56.802* RH2

Table: 4.3.2.6 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
LASSO	0.85	0.72	0.70	87.533

Table: 4.3.2.7 ANOVA of the developed model

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	649190.037	3	216396.679	28.243	.000
Residual	245184.519	32	7662.016		
Total	894374.556	35			

Table: 4.3.2.8 Coefficients for the developed model

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	-5218.285	878.732		-5.938	.000
MaxTemp	71.615	41.271	0.226	1.735	.092
MinTemp	25.552	45.604	0.073	.560	.579
RH2	56.802	7.885	0.713	7.204	.000

4.3.2.5 Predictions of oilseeds yield using LASSO

The present study uses 48 years data from 1970-71 to 2017-18 to predict the oilseeds yield using weather parameters. we predict the yield using the LASSO method. The predicted yield should be transformed in order to make a comparison with actual yield and also compared the predicted yield with actual yield using visualization. The actual verses predicted yield using LASSO method is presented in the figure: 4.3.2.4 below.

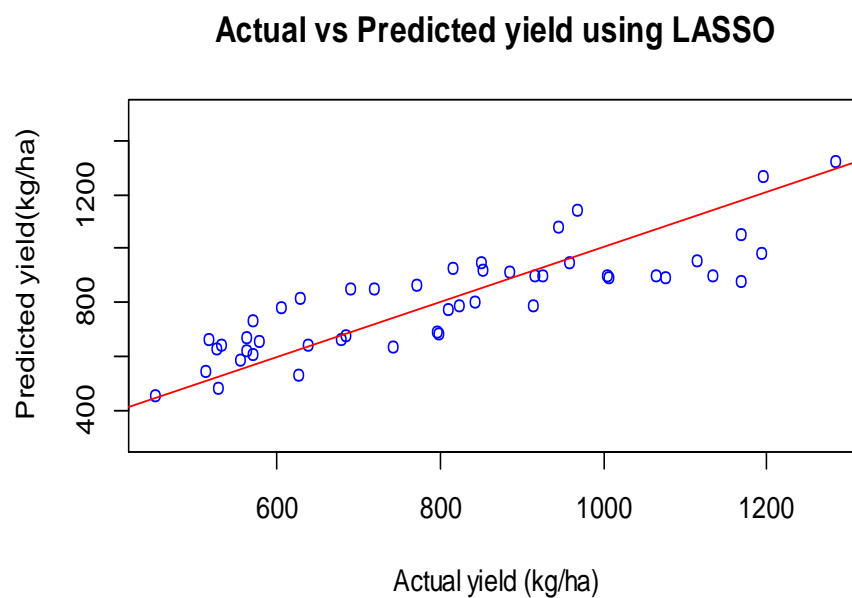


Figure.4.3.2.4 Actual yield verses predicted yield using LASSO regression

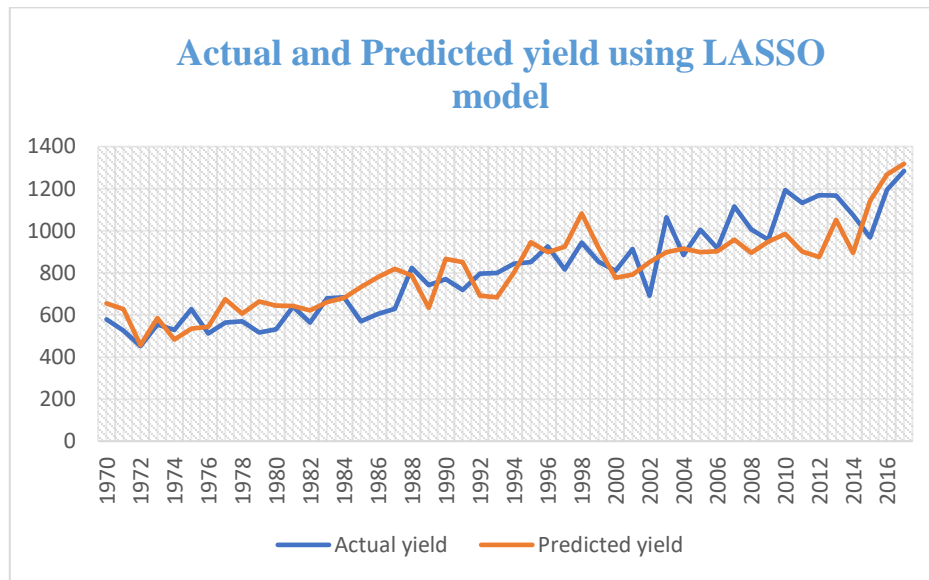


Figure.4.3.2.4 Actual yield and predicted yield using LASSO regression

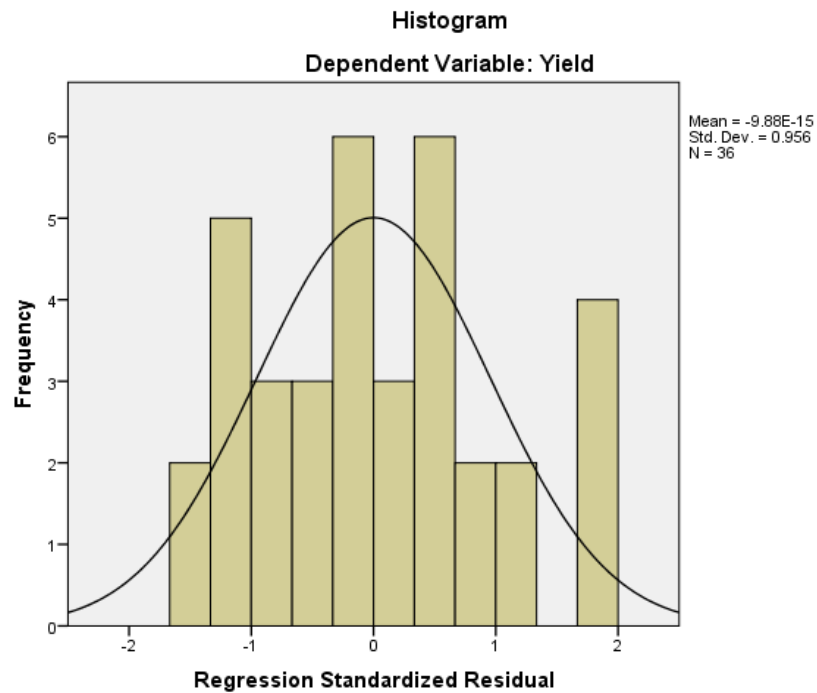


Figure.4.3.2.5 Histogram of the model

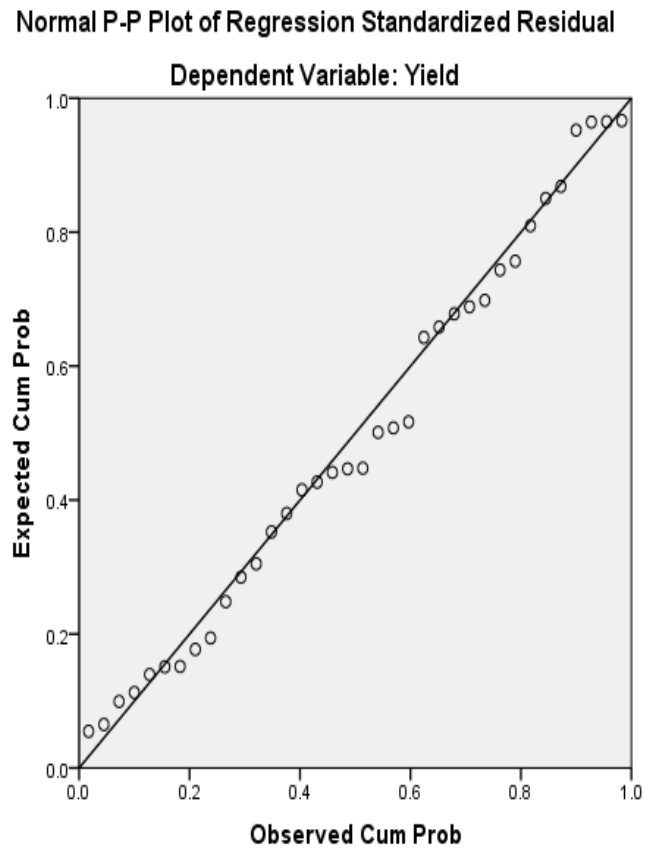


Figure.4.3.2.6 Normal P-P Plot of Regression Standardized Residuals

4.3.3 Models based on weather indices.

4.3.3.1 Weather indices calculation

Two types of weather indices were developed for each weather variable, i.e. simple and weighted weather indices. Simple weather indices were generated by summing the individual or interaction of weather parameters by taking annual data of two weather variables at a time. Weighted weather indices were generated from sum product of individual or interaction of weather variables and it is in correlation with crop yield

Data of different weather variables viz. Maximum temperature ($^{\circ}\text{C}$), minimum temperature ($^{\circ}\text{C}$), rainfall (mm) and relative humidity (%) were used to get simple and weighted weather indices for regression analysis in the prediction of oilseeds yield. Yield forecast models have been worked out through stepwise regression method.

The formation of weather variables for weather indices, thus, simple and weighted weather indices are presented in Table 4.3.3.1 below.

Table: 4.3.3.1 Notations for simple weather indices and weighted weather indices

Parameter	Simple weather indices					Weighted weather indices				
	Tmax	Tmin	Rf	RH1	RH2	Tmax	Tmin	Rf	RH1	RH2
Tmax	Z10					Z11				
Tmin	Z120	Z20				Z121	Z21			
Rf	Z130	Z230	Z30			Z131	Z231	Z31		
RH-I	Z140	Z240	Z340	Z40		Z141	Z241	Z341	Z41	
RH-II	Z150	Z250	Z350	Z450	Z50	Z151	Z251	Z351	Z451	Z51

For selecting the best regression equation among number of independent variables, stepwise regression procedure was adopted. A regression model was fitted considering the entered variables obtained from individual stepwise regression analysis to predict the yield of oilseeds for the subsequent years.

4.3.3.2 Model based on simple weather indices

Weather indices-based regression models were developed using simple weather indices as independent variables while character under study such as yield was used as dependent variable. Stepwise regression technique has been used for selecting significant variables in the model. Weather indices-based regression model is presented in the table: 4.3.3.2

Table: 4.3.3.2 Simple weather indices-based regression model.

Model	Regression Equation
SWI	Yield = -2774.73 + 2.017* Z150 + 31.59* Time

SWI = Simple Weather Indices

Table: 4.3.3.3 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
SWI	0.835	0.697	0.688	89.244

Table: 4.3.3.4 ANOVA for the developed model

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	623580.155	1	623580.155	78.295	.000
Residual	270794.400	34	7964.541		
Total	894374.556	35			

Table: 4.3.3.5 Coefficients for the developed model

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-2774.733	394.175		-7.039	.000
Z150	2.017	0.228	0.835	8.848	.000

4.3.3.3 Model based on weighted weather indices

Weather indices-based regression models were developed using weighted weather indices as independent variables while character under study such as yield was used as dependent variable. Stepwise regression technique has been used for selecting significant variables in the model. Weather indices-based regression model is presented in the table: 4.3.3.6

Table: 4.3.3.6 Weighted weather indices-based regression model

Model	Regression Equation
WWI	Yield = -2774.73 + 2.465* Z151 + 13.59* Time

WWI = Weighted Weather Indices

Table: 4.3.3.7 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
WWI	0.835	0.697	0.688	89.244

Table: 4.3.3.8 ANOVA for weighted weather indices-based model

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	623580.155	1	623580.155	78.295	.000
Residual	270794.400	34	7964.541		
Total	894374.556	35			

Table: 4.3.3.9 Coefficients for the developed model

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	-2774.733	394.175		-7.039	.000
Z151	2.017	.228	.835	8.848	.000

4.3.3.4 Predictions of oilseeds yield based on weather indices

The present study uses 48 years data from 1970-71 to 2017-18 to predict the oilseeds yield using weather parameters. we predict the yield using weather indices. The predicted yield should be transformed in order to make a comparison with actual yield and also compared the predicted yield with actual yield using visualization. The actual verses predicted yield using weather indices method is presented in the figure: 4.3.3.4

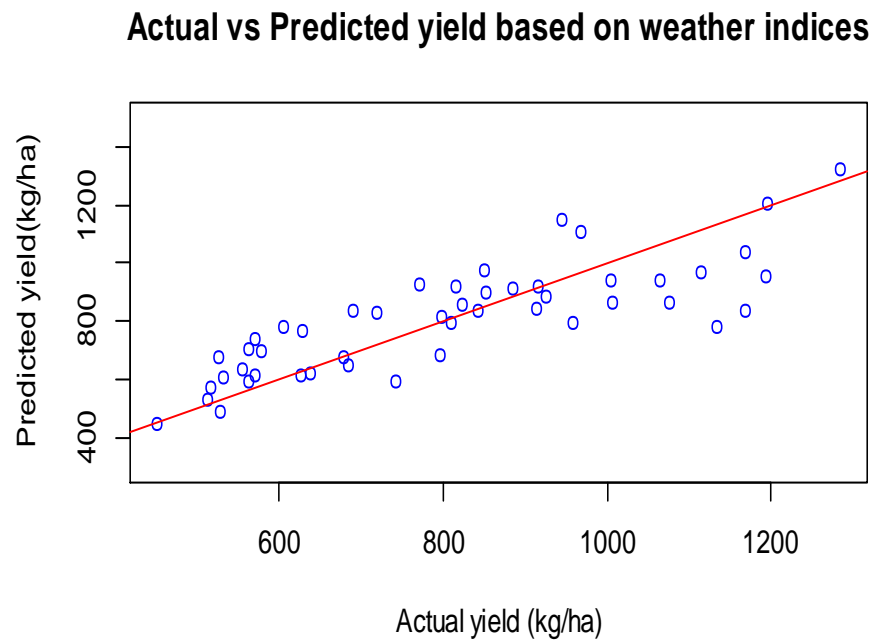


Fig: 4.3.3.4 Actual verses predicted yield based on weather indices

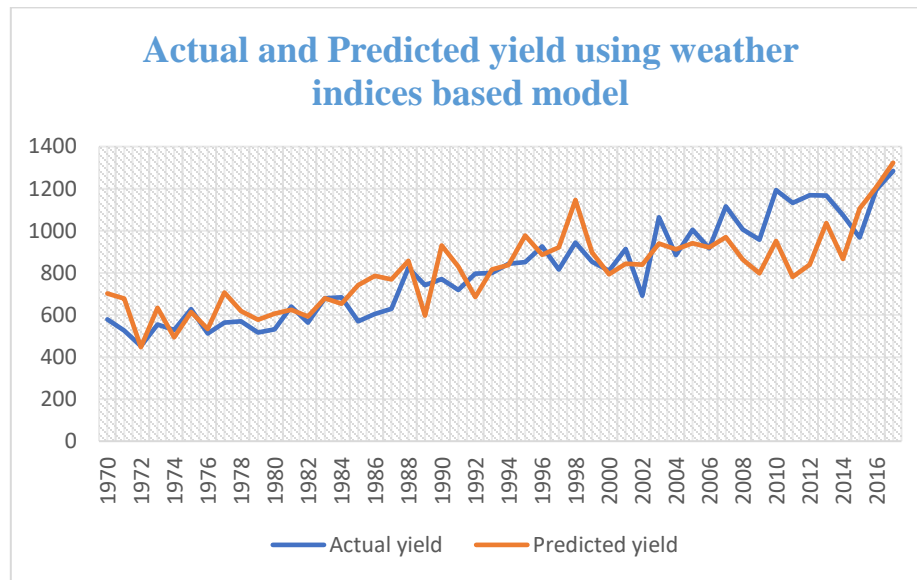


Fig: 4.3.3.4 Actual and predicted yield based on weather indices

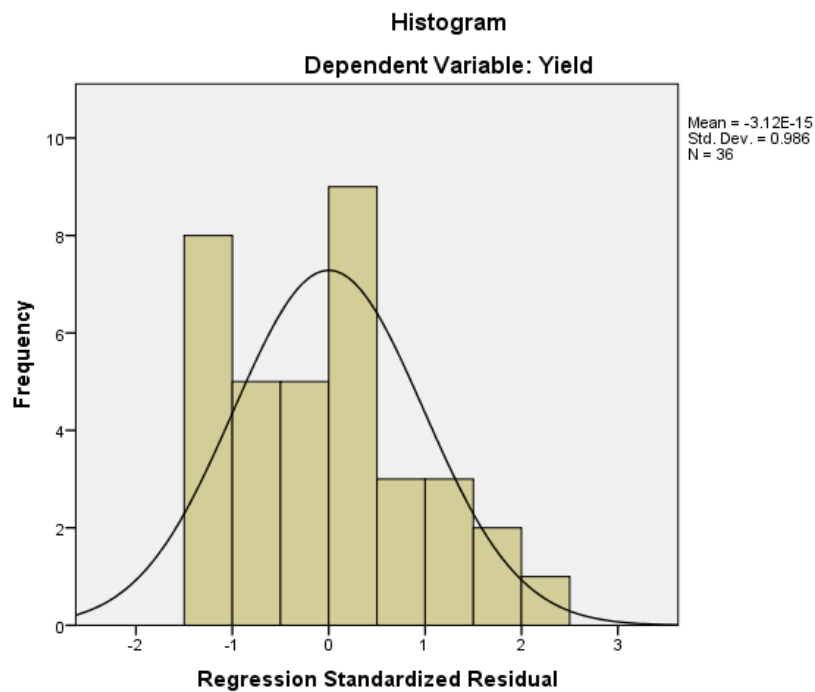


Fig: 4.3.3.5 Histogram of the developed model

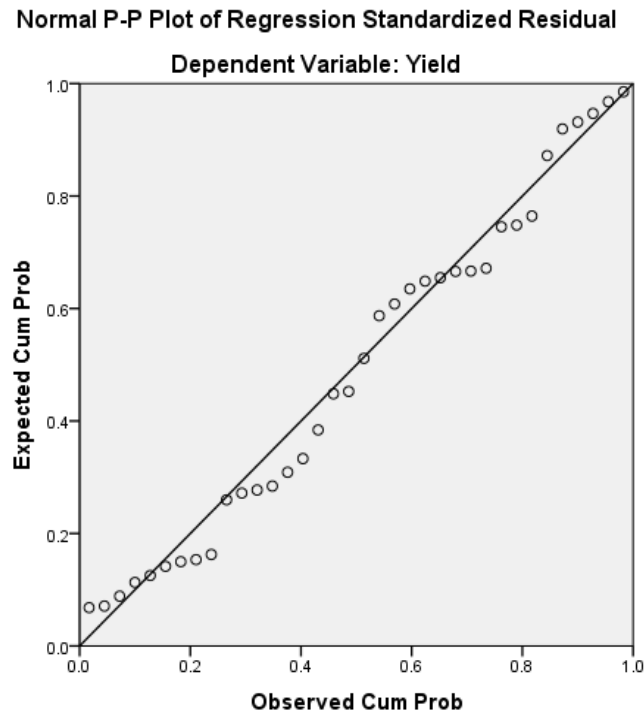


Fig: 4.3.3.6 Normal P-P Plot of Regression Standardized Residual

4.3.4 Models using composite weather variables.

Data of different weather variables viz. Maximum temperature ($^{\circ}\text{C}$), minimum temperature ($^{\circ}\text{C}$), rainfall (mm) and relative humidity (%) were used to get weighted and un-weighted indices for regression analysis in the prediction of oilseeds yield. Yield forecast models have been worked out through stepwise regression method.

For each weather variable, two indices were worked out.

1. Un-weighted weather index = Sum (each weather variable)
2. Weighted weather index = Sum (each weather variable x correlation Coefficient between yield and weather variable).

Weather indices denoted as Z; un-weighted indices are 0 and weighted indices are 1. For instance, maximum temperature taken as 1st variable, hence weather index of un-weighted maximum temperature is Z10 and for weighted Z11. In the same way, other

indices were worked out for other weather variables. To study the combined effect of weather variables, un-weighted and weighted indices were also computed. For instance, combination of maximum and minimum temperature is obtained by multiplying values of maximum and minimum temperature. For selecting the best regression equation among number of independent variables, stepwise regression procedure was adopted. A regression model was fitted considering the entered variables obtained from individual stepwise regression analysis to predict the yield of oilseeds for the subsequent years. The stepwise regression technique was used to find out the most significant variable. The multiple regression model was used based on yield as regressor and significant generated variable as regress and for the forecasting of yield.

4.3.4.1 Data requirement

Long period data of weather parameters viz., maximum temperature, minimum temperature, rainfall, RH1 and RH2 were collected from *Indiastat* and India Meteorological Department (IMD), Pune. Long period annual data of total oilseeds yield collected from *Indiastat* from 1970-71 to 2017-18.

4.3.4.2 Model development

Models using composite weather variables were developed using simple and weighted weather indices as shown in table: 4.3.4.1 as independent variables while character under study such as yield was used as dependent variable. Stepwise regression technique has been used for selecting significant variables in the model. Composite weather variables-based regression model is presented in table: 4.3.4.1 below.

Table: 4.3.4.1 Model using composite weather variables

Model	Regression Equation
CWV	$\text{Yield} = -1998.920 + 1.759 * Z151 + 0.016 * Z130 + 13.51 * \text{Time}$

CWV= Composite weather variables

Table: 4.3.4.2 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
WWI	0.86	0.74	0.73	82.440

Table: 4.3.4.3 ANOVA for model using composite weather variable

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	670095.725	2	335047.863	49.298	.000
Residual	224278.831	33	6796.328		
Total	894374.556	35			

Table: 4.3.4.4 Coefficients for the developed model

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	-1998.920	469.601		-4.257	.000
Z151	1.759	.373	.596	4.712	.000
Z130	.016	.006	.331	2.616	.013

4.3.4.2 Predictions of oilseeds yield using composite weather variables

The present study uses 48 years data from 1970-71 to 2017-18 to predict the oilseeds yield using weather parameters. we predict the yield using composite weather variables. The predicted yield should be transformed in order to make a comparison with actual

yield and also compared the predicted yield with actual yield using visualization. The actual versus predicted yield using composite weather variables method is presented in the figure: 4.3.4.2 below.

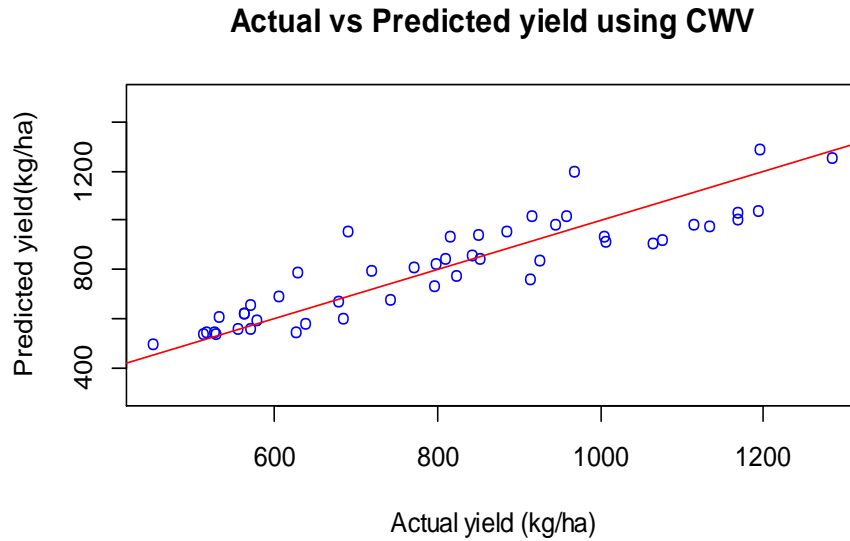


Fig: 4.3.4.2 Actual versus predicted yield using composite weather variables

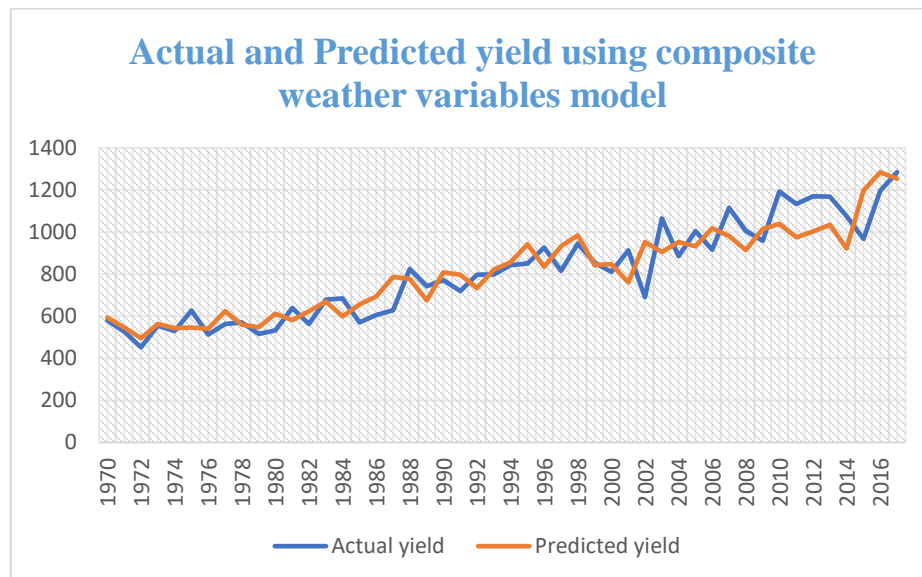


Fig: 4.3.4.3 Actual and predicted yield using composite weather variables

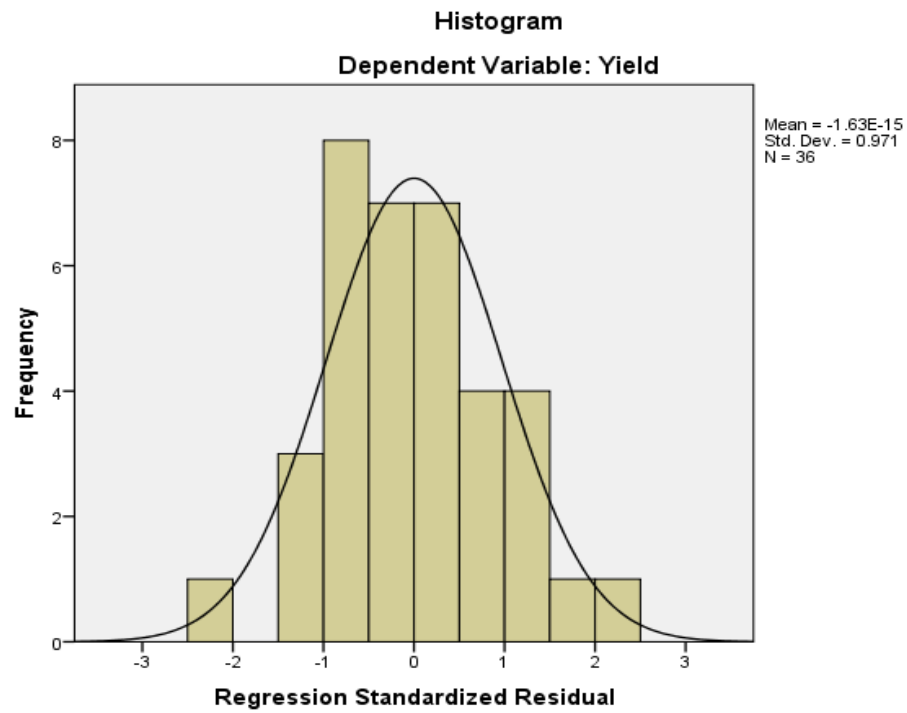


Fig: 4.3.4.4 Histogram of the developed model

Normal P-P Plot of Regression Standardized Residual

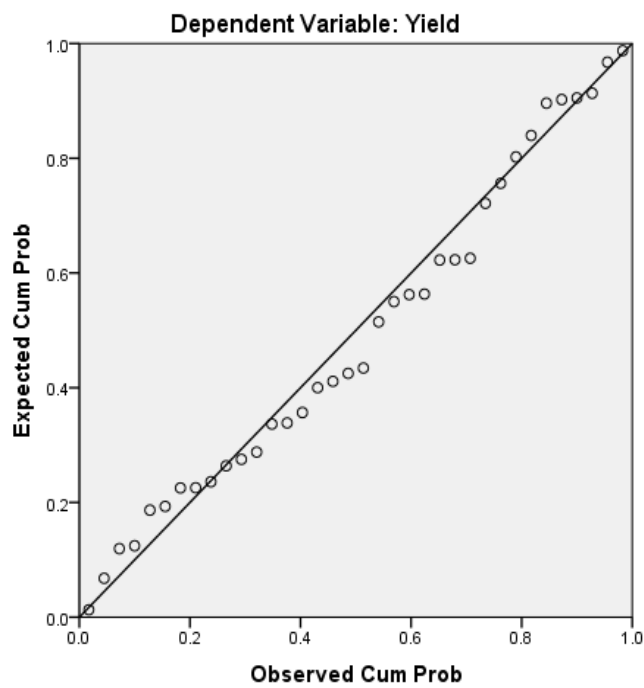


Fig: 4.3.4.5 Normal P-P Plot of Regression Standardized Residual

4.3.5 Artificial Neural Network (ANN)

Neural networks have always been one of the most fascinating machine learning models in my opinion, not only because of the fancy backpropagation algorithm, but also because of their complexity (think of deep learning with many hidden layers) and structure inspired by the brain. Neural networks have not always been popular, partly because they were, and still are in some cases, computationally expensive and partly because they did not seem to yield better results when compared with simpler methods such as support vector machines (SVMs).

In the present study we are going to fit a simple neural network using the neuralnet package in R-software. The data set used to develop the model is total oilseeds yield data based on weather parameters in India from 1970-71 to 2017-18. First, we need to check that no datapoint is missing, otherwise we need to fix the dataset. In the present study there is no missing data, good. We proceed by splitting the data into a train and a test set, then we fit a neural network model and test it on the test set.

4.3.5.1 Preparing to fit the neural network.

Before fitting a neural network, some preparation needs to be done. Neural networks are not that easy to train and tune. As a first step, we are going to address data preprocessing. It is a good practice to normalize our data before training a neural network. Depending on the dataset, avoiding normalization may lead to useless results or to a very difficult training process (most of the times the algorithm will not converge before the number of maximum iterations allowed). Choose different methods to scale the data (z-normalization, min-max scale, etc...). In this study we chose to use the min-max method and scale the data in the interval (0,1). Usually scaling in the intervals (0,1) or (-1,1) tends to give better results. Therefore, scale and split the data before moving on: Note that the scale returns a matrix that needs to be coerced into a data frame. We use min-max normalization to scale the data and is presented in the table:4.3.5.1 below.

Table: 4.3.5.1 Min-Max normalization to scale the data

Scale	Yield	Max.Temp	Min.Temp	Rainfall	RH1	RH2
Maximum	1064	30.23	20.39	1401.40	76.40	63.78
Minimum	452	28.89	18.62	947.10	68.59	54.53

To develop the neural network model, the present study uses yield as the dependent variable and maximum temperature, minimum temperature, rainfall, RH1 and RH2 as independent variables. We divide the data into training and test set. Training set is used to find the relationship between dependent and independent variables while the test set assesses the performance of the model. We use 75% of the dataset as training set and 25% of the dataset as test set. The assignment of the data to training and test set is done using random sampling. We perform sampling on R using *sample* function and used *set.seed* to generate same sample everytime and maintain consistency. We will use the *index* variable while fitting neural network to create training and test data sets.

As far as there is no fixed rule as to how many layers and neurons to use although there are several more or less accepted rules of thumb. Usually, if at all necessary, one hidden layer is enough for a vast number of applications. As far as the number of neurons is concerned, it should be between the input layer size and the output layer size, usually $2/3$ of the input size. At least in my little experience testing again and again is the best solution since there is no guarantee that any of these rules will fit our model best. In the present study we are going to use 2 hidden layers with this configuration: 5:3:2:1. The input layer has 5 inputs, the two hidden layers have 3 and 2 neurons and the output layer has, of course, a single output since we are doing regression. The scaled data is used to fit the neural network. We visualize the neural network with weights for each of the variable. The results are presented in the table.4.3.5.2 below.

Table: 4.3.5.2 Estimated Neural Network Weights.

Neural Network Result Matrix	
error	0.223179683
reached.threshold	0.009957663
steps	319.000000000
Intercept.to.1layhid1	0.691191479
MaxTemp.to.1layhid1	1.017489101
MinTemp.to.1layhid1	0.291578953
Rainfall.to.1layhid1	-0.278354573
RH1.to.1layhid1	-0.871739338
RH2.to.1layhid1	-2.998366726
Intercept.to.1layhid2	0.535520279
MaxTemp.to.1layhid2	-3.737142676
MinTemp.to.1layhid2	-2.019762697
Rainfall.to.1layhid2	-1.553614596
RH1.to.1layhid2	-0.466008790
RH2.to.1layhid2	9.838276597
Intercept.to.1layhid3	3.050740267
MaxTemp.to.1layhid3	-2.678509246
MinTemp.to.1layhid3	-4.814368559
Rainfall.to.1layhid3	0.248077286
RH1.to.1layhid3	2.100508356
RH2.to.1layhid3	-1.759492181
Intercept.to.2layhid1	0.568593044
1layhid1.to.2layhid1	-2.453804462
1layhid2.to.2layhid1	1.358967839
1layhid3.to.2layhid1	-2.482077152
Intercept.to.2layhid2	-1.592719978
1layhid1.to.2layhid2	1.374810710
1layhid2.to.2layhid2	1.528530930
1layhid3.to.2layhid2	0.108425030
Intercept.to.Yield	-0.197697988
2layhid1.to.Yield	1.258267124
2layhid2.to.Yield	0.372639761

Figure: 4.3.5.1 visualizes the computed neural network. Our model has (3,2) neurons in its hidden layer. The black lines show the connections between each layer and the weights on each connection while the blue lines show the bias term added in each step. The bias can be thought as the intercept of a linear model. The net is essentially a black box so we cannot say that much about the fitting, the weights and the model. Suffice to say that the training algorithm has converged and therefore the model is ready to be used.

This is the graphical representation of the model with the weights on each connection:

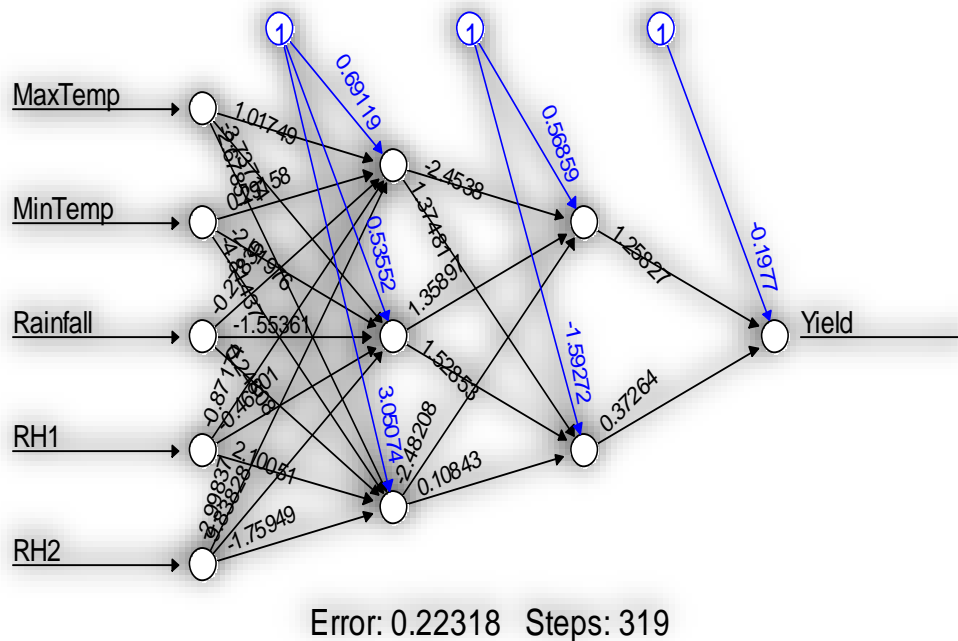


Fig: 4.3.5.1 Artificial neural network model for yield based on weather parameters.

4.3.5.2 Predicting oilseeds yield using the neural network

The present study uses 48 years data from 1970-71 to 2017-18 to predict the oilseeds yield using weather parameters. we predict the yield using the neural network method. The predicted yield should be transformed in order to make a comparison with actual yield and also compared the predicted yield with actual yield using visualization. The actual verses predicted yield using neural network method is presented in the figure: 4.3.5.2 below.

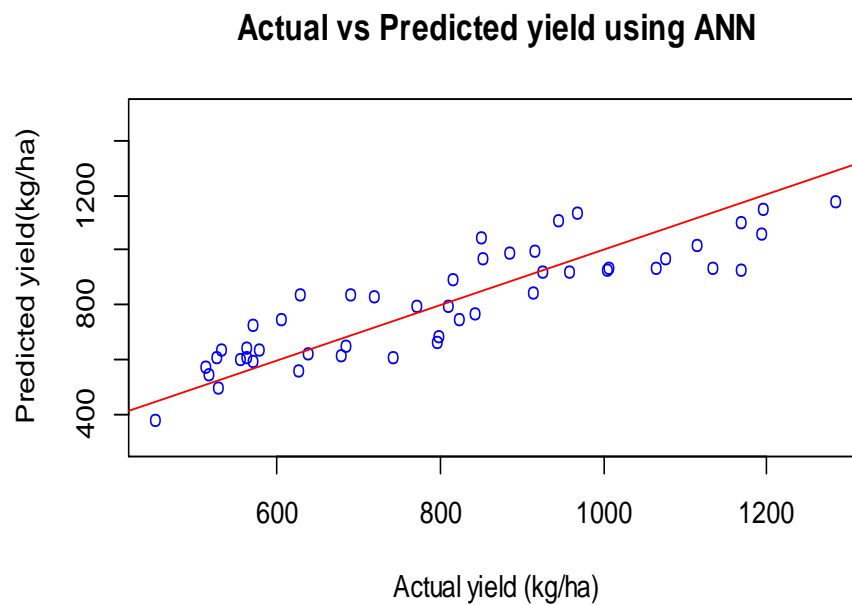


Fig: 4.3.5.2 Actual yield verses predicted yield using neural network method.

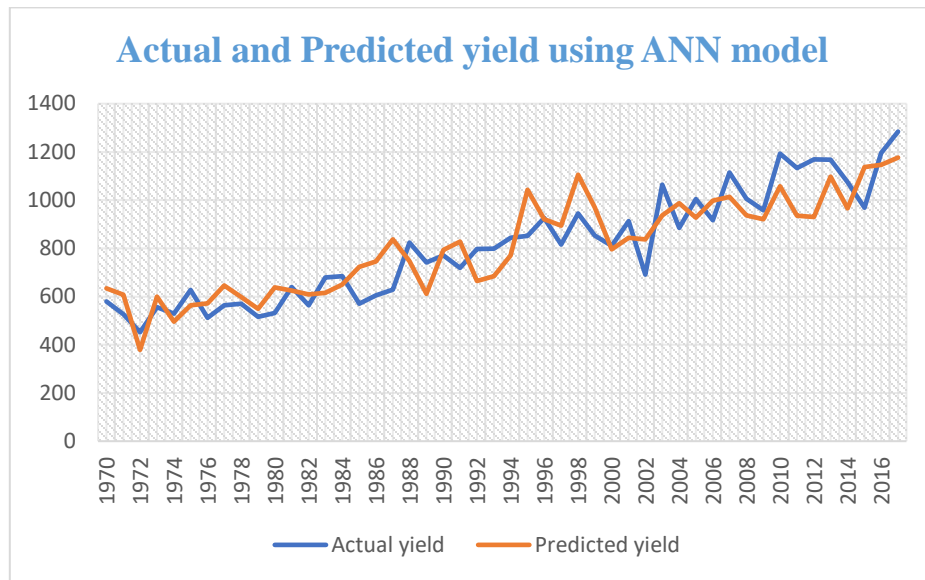


Fig: 4.3.5.3 Actual yield and predicted yield using neural network method.

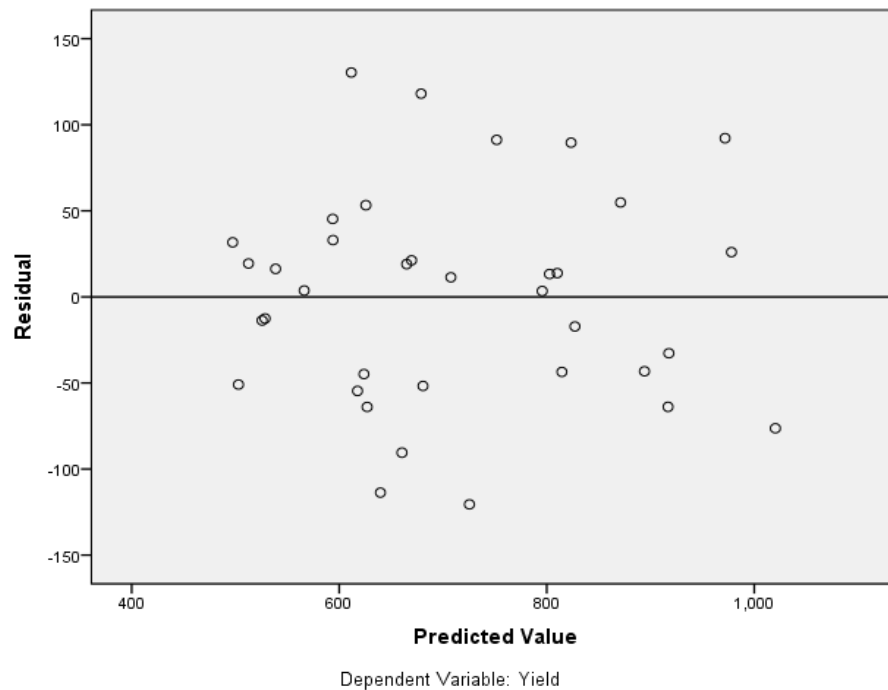


Fig: 4.3.5.4 Residual by Predicted Chart

4.4 Validation of the developed models

In order to assess the performance of the model and to ascertain the accuracy, it is essential to undertake validation as the second stage of development of crop yield forecast model. Validation of yield forecast models is carried out mainly using the yield data and weather parameters of India for the interest of latest 12 years data *i.e.* 25% of the total data. Once crop yield forecast models are developed using long term data, crop yield forecast for latest 12 years is estimated and compared with the actual yield for the country. Percentage error for the forecasted yield against actual yield for the latest 12 years is computed. Level of acceptable limits of percentage error at various levels is mentioned below.

District level ± 10

State level ± 5

National level ± 2

If crop yield forecast models developed for a particular nation does not seem to be accurate with respect to percentage error, there is a need to improve the models so that the percentage error falls within acceptable limits.

Validation is determination of the truthfulness of the model with independent data sets *i.e.*, whether model accurately predicts growth yield and processes. Validation is the comparison between the simulated data and the observed dataset. Model can be considered valid and useful even though there may be some difference between observed data and simulated output. If the simulated values lie within the projected confidence band, then the model can be considered valid. Thus, validation is used as an evaluation of model for its usefulness. Validation of the model was carried out by comparing the model output with the observed values of oilseeds crop yield.

4.4.1 R^2 and Root Mean Square Error (RMSE)

R-Square is the proportion of variance in the dependent variable (yield) which can be explained by the independent variables (weather variable like Tmax, Tmin, RH1, RH2 and Rainfall). This is an overall measure of the strength of association to

independent variable with the dependent variable. R^2 between 0 and 1 indicates the extent to which the dependent variable is predictable. For example, if $R^2 = 0.90$ means that 90 percent of the variance in yield is predictable from weather variable and so on.

An analysis of the degree of coincidence between simulated and observed values was carried out by using Root Mean Square Error. The RMSE has been widely used as a criterion for model evaluation and has been computed as:

$$\text{RMSE (\%)} = \left\{ \frac{1}{n} \sum (O_i - S_i)^2 \right\}^{1/2} \times \frac{100}{\bar{o}}$$

observed values and n is the number of observations within each treatment. The lower is the value of RMSE, higher is the accuracy of the model predictions. A zero value is ideal. Normalized RMSE gives a measure (%) of the relative difference of simulated versus observed data. The simulation is considered excellent with a normalized RMSE less than 10%, good if the normalized RMSE is greater than 10% and less than 20%, fair if normalized RMSE is greater than 20 and less than 30%, and poor if the normalized RMSE is greater than 30% (Jamieson *et al.*, 1991).

4.4.2 Yield data comparison between predicted and observed values

Comparison between Observed and Predicted yield for the different statistical techniques used to develop the yield predictive models based on the weather parameters. The following results shows the validation of 25% dataset which were not included in the model development along with percentage deviation. The summary measures describe the quality of simulation while the difference measures try to locate and quantify errors. The latter include the Root Mean Square Error (RMSE).

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

RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between forecasted and observed. PE is defined as ratio of RMSE to mean observed value expressed as percentage.

$$\text{PE} = \frac{\text{RMSE}}{\bar{o}} \times 100$$

Table: 4.4.1 Validation of Artificial Neural Network (ANN) model

Year	Observed	Predictions	%deviation	RMSE	Error (%)
2006	916	997.41	-8.89	40.71	3.71
2007	1115	1013.92	9.07	50.54	4.60
2008	1006	936.50	6.91	34.75	3.16
2009	958	920.88	3.88	18.56	1.69
2010	1193	1056.91	11.41	68.04	6.20
2011	1133	935.71	17.41	98.65	8.98
2012	1169	929.60	20.48	119.70	10.90
2013	1168	1097.04	6.08	35.48	3.23
2014	1075	965.10	10.22	54.95	5.00
2015	968	1137.18	-17.48	84.59	7.70
2016	1195	1146.61	4.05	24.20	2.20
2017	1284	1176.36	8.38	53.82	4.90

Table: 4.4.2 Validation of LASSO Stepwise Regression Model

Year	Observed	Predictions	%deviation	RMSE	Error (%)
2006	916	902.1798	1.509	6.910	0.754
2007	1115	957.8305	14.096	78.585	7.048
2008	1006	895.5114	10.983	55.244	5.491
2009	958	948.9731	0.942	4.513	0.471
2010	1193	984.3002	17.494	104.350	8.747
2011	1133	901.7171	20.413	115.641	10.207
2012	1169	875.9067	25.072	146.547	12.536
2013	1168	1051.535	9.971	58.233	4.986
2014	1075	894.7009	16.772	90.150	8.386
2015	968	1141.181	-17.891	86.591	8.945
2016	1195	1267.322	-6.052	36.161	3.026
2017	1284	1318.124	-2.658	17.062	1.329

Table: 4.4.3 Validation of composite weather variables model

Year	Observed	Predicted	%deviation	RMSE	Error (%)
2006	916	1017.762	-11.1094	50.8812	5.5547
2007	1115	980.944	12.02296	67.0280	6.0115
2008	1006	914.2242	9.122847	45.8879	4.5614
2009	958	1014.406	-5.88785	28.2028	2.9439
2010	1193	1038.31	12.96648	77.3451	6.4832
2011	1133	974.681	13.97343	79.1595	6.9867
2012	1169	1002.912	14.20772	83.0441	7.1039
2013	1168	1033.699	11.49837	67.1505	5.7492
2014	1075	921.7354	14.25717	76.6323	7.1286
2015	968	1196.618	-23.6176	114.3090	11.8088
2016	1195	1283.467	-7.40308	44.2334	3.7015
2017	1284	1017.762	2.436236	15.6406	1.2181

Table: 4.4.4 Validation for weather indices-based models

Year	Observed	Predicted	%deviation	RMSE	Error (%)
2006	916	921.4422	-0.59413	2.7211	0.2971
2007	1115	969.2997	13.0673	72.8502	6.5336
2008	1006	862.7468	14.23988	71.6266	7.1199
2009	958	797.1568	16.78947	80.4216	8.3947
2010	1193	951.9877	20.2022	20.5061	10.1011
2011	1133	780.4007	31.12086	76.2997	15.5604
2012	1169	838.8944	28.23829	65.0528	14.1191
2013	1168	1036.824	11.23082	65.5880	5.6154
2014	1075	866.3374	19.41047	04.3313	9.7052
2015	968	1104.408	-14.0917	68.2038	7.0458
2016	1195	1206.25	-0.94139	5.6248	0.4707
2017	1284	1323.134	-3.0478	19.5668	1.5239

The observed and predicted yields for period (2006-2017) and various error analysis of independent data of ANN, LASSO, CWV and MWI have been presented in Table: 4.4.1, Table: 4.4.2, Table: 4.4.3 and Table: 4.4.4 respectively and the regression models were validated with 12 years data from 2006 to 2017 of independent dataset. The data exposed that the oilseeds yield forecasting models for India showed their reliability by producing error below 10% as showed in the tables earlier. The error structures for the models ANN, CWV, LASSO and MWI were viz., RMSE of 56.99, 62.45, 66.66 and 79.39 kg/ha with percentage error (PE) of 5.18, 5.77, 5.99 and 7.20 respectively.

4.4.3 Cross-comparison of the developed models

The ranking of the models based on R^2 and RMSE revealed that the Artificial Neural Networks (ANN) was the best performing model followed by least absolute shrinkage selection operator (LASSO), models using composite weather variables and models based on weather indices as presented in the table: 4.4.5 below. The order of performance of model using R^2 was as follows: ANN (0.81) > CWV (0.74) > LASSO (0.72) > SWI (0.69). The models were also ranked using RMSE which was found as follows: ANN (56.99) < CWV (62.45) < LASSO (66.66) < MWI (79.39). The models were also ranked using MAPE and the order of the models are ANN < CWV < LASSO < MWI.

Table: 4.4.5 Comparison of developed models.

Model	R²	RMSE	MAPE	%deviation	% Error
ANN	0.81	56.99	10.35	5.95	5.18
CWV	0.74	62.45	11.54	7.53	5.77
LASSO	0.72	66.66	11.98	7.55	5.99
MWI	0.69	79.39	14.41	11.30	7.20

ANN = Artificial Neural Network

LASSO = Least Absolute Shrinkage Selection Operator

CWV = Composite Weather Variables

MWI = Model Based on Weather Indices

Table:4.4.6 Forecast of oilseeds yield using best fitted model.

Year	Actual yield	Forecast
2010	1193	1056.910
2011	1133	935.707
2012	1169	929.598
2013	1168	1097.036
2014	1075	965.096
2015	968	1137.176
2016	1195	1146.605
2017	1284	1176.364
2018		1176.642
2019		1191.708
2020		1206.773

CHAPTER - V

SUMMARY AND CONCLUSION

The entire work carried out in the present investigation has been summarized and valid conclusions have been drawn in this Chapter.

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

The relevant literature related to forecast models based on weather variables, biometrical characters, input of crop-production etc. has been reviewed critically and presented in chronological order in Chapter-II.

The Chapter-III dealt with the material and statistical methodologies used for developing forecast models of yield. Time series data on rice yield, In this work total oilseeds yield prediction is done by considering various parameters like Maximum temperature, Minimum temperature, Rainfall, RH1 and RH2 and historic oilseeds yield using 48 years covering the period from 1970-71 to 2017-18 have been utilized for development of pre-harvest forecast model and for studying the effect of weather variables on oil seed yield data. data for development of LASSO Stepwise Regression, Models based on weather indices, Models using composite weather variables and Artificial Neural Networks (ANN). A comparison study of the results obtained from ANN, LASSO, MWI and CWV is also performed. The results were compared using the R^2 , RMSE, MAPE statistic and percentage prediction error.

Techniques of development of pre-harvest forecast models based on weather variables and inputs have been illustrated. Regression models have been used to study the individual and joint effect of weather variables.

The results of the investigation have been presented in the Chapter-IV, and have been discussed critically. The results are summarized as under.

Effect of individual weather variables: -

- 1. Maximum Temperature:** The time series results of annual maximum temperature revealed a positive trend, indicating that the maximum temperatures are increasing. Annual maximum temperatures averaged over India shows significant warming trend. The results show that the mean of annual maximum temperature in india was about 29.65 °C with a standard error of 0.07.
- 2. Minimum Temperature:** A positive trend was observed in the time series of annual minimum temperature, indicating that the minimum temperatures are increasing. Annual minimum temperatures averaged over India shows significant increasing trend. The results show that the mean annual minimum temperature in india was about 19.91 °C with a standard error of 0.07.
- 3. Rainfall:** Time series results of annual rainfall revealed a negative trend *i.e.* It represents the annual rainfall in india shows decreasing trend (Negative trend). The results show that the mean annual rainfall of the of the study period was 1155.22mm with standard error of 14.06. The minimum and maximum ever recorded rainfalls were 947.1 mm in 1972 (the driest year) and 1294.8 in 1975 (the wettest year) per year respectively.
- 4. Relative Humidity (RH1):** The time series results of annual RH1 revealed a positive trend, indicating that the RH1 in india showed significant increasing trend. The results show that the mean of annual RH1 in india was about 73.25 percent with a standard error of 0.27.
- 5. Relative Humidity (RH2):** A positive trend was observed in the time series of annual RH2, indicating that the relative humidity is increasing. Annual RH2 averaged over India shows significant increasing trend. The results show that the mean annual RH2 in india was about 59.04 Percent with a standard error of 0.29.

Nine major oilseeds viz., groundnut, soybean, rapeseed-mustard, sesamum, sunflower, safflower, linseed, niger and castor together contribute to the total productivity of oilseeds in India. There was a general increasing trend in the oilseeds yield, which is an indication of improving yield over time. The results show that the mean of annual oilseeds yield in india was about 806.15 kg/ha with a standard error of 32.26.

To determine the relationship between weather parameters and yield, correlation analysis was carried out. Correlation studies pave way to know the association prevailing between crop yield and weather parameters which was given better understanding of the contribution of each weather parameter to the crop yield, with this view the yield and weather parameters were subjected to correlation analysis and the results were discussed. The highest correlation values were observed with the relative humidity (RH1), while the rainfall showed the least correlation values.

Forecast models: -

LASSO Stepwise Regression:

- Feature selection is crucial and challenging in this field of study, mainly because the desired output varies for different set of data, and it is hard to find a model that works for every kind of problem. For these reasons researchers always try to find feature selection model that are well adaptable for the dataset they want to analyze.
- The task becomes even more challenging when dealing with high-dimensional datasets. In the present study we decided to face the feature selection problem using the LASSO method. We tested this method using deferent setups, mainly we focused on statistical model: Generalized linear model.
- For the GLM we considered the Logistic regression model for a small n- large-P dataset. Finally, we can say that in our examples the LASSO method helps us to choose a model with the most relevant features in it.

- According to the bias-variance trade-off, all things equal, simpler model should be always preferred because it is less likely to overfit the training data. For variable selection, an alternative to the penalized logistic regression techniques is the stepwise logistic regression.

Model based on weather indices:

- Forecasting models were developed based on modified Hendrick and Scholl technique for oilseed crops by using past years yield and weather data. In this techniques time trend, various weather variables and their simple and weighted weather indices were utilized.
- Weather indices-based regression models were developed using individual effect of simple weather indices and weighted weather indices as independent variables while character under study such as yield was used as dependent variable.
- Stepwise regression technique has been used for selecting significant variables in the model. The models are relying on various weather parameters during their growing period.
- The two models gave the least estimates for yield forecast by giving lower regression coefficient and higher error per cent during validation period compared to other models.

Model using composite weather variables:

- Models were developed based on composite weather variables for oilseed crops by using yield and weather data. In this model time trend, various weather variables and their combined effect of simple and weighted weather indices were utilized.
- The combined effect of weather variables viz., maximum temperature, minimum temperature, rainfall, RH1 and RH2 for oilseeds played crucial role in yield determination.

- The models were developed using combination of simple weather indices and weighted weather indices as independent variables while character under study such as yield was used as dependent variable.
- Stepwise regression technique has been used for selecting significant variables in the model.
- The model using composite weather variables gave the better estimates for yield forecast by giving higher regression coefficient and lower error per cent during validation period compared to model based on weather indices and least compared to LASSO and ANN models.

Artificial Neural Network (ANN):

- Based on the presented analyzes, it should be stated that the prediction of oilseeds yield using artificial neural networks effects in good forecasting results. However, in order to optimize the model, further research should be undertaken, which will aim at obtaining data from more fields and further analysis of the number of independent factors within the model.
- Utilizing artificial neural network with the MLP 5:3:2:1 structure to produce a forecast model of oilseeds yield allows to use of this method in agricultural applications for yield simulation.
- Obtaining RMSE percentage prediction errors at the level of 6.20% should be considered satisfactory. Further work should be undertaken on the optimization of the neural model, i.e., the selection of an appropriate number of independent features affecting the oilseeds yield.
- Hence combination of weather and yield data is appropriate and consistent option for yield forecasting.

Discussion of the models and comparisons:

Errors are the difference between observed and expected values and they are assumed to be very less. Among the models tested for the present problem in predicting oilseeds yield, prediction obtained from ANN with the

lowest expected errors, out performs the other models. The R^2 & RMSE of ANN, CWV, LASSO and MWI models varied between 0.81 & 56.99, 0.74 & 62.45, 0.72 & 66.66 and 0.69 & 79.39 respectively. The average percentage prediction error for the models on the test set were found to be 5.18, 5.77, 5.99 and 7.20 respectively.

Conclusions:

- Different prediction models namely LASSO Stepwise Regression, Models based on weather indices, Model using composite weather variables and Artificial Neural Networks have been tried to predict the oilseeds yield based on weather parameters in India.
- The results indicate that the ANN model had a higher R^2 (0.81) value and a lower percentage prediction error (5.18) followed by model using composite weather variables, LASSO Stepwise Regression and model based on weather indices. This shows that the ANN model was able to predict the oilseeds yield better than the other models for the given data set.
- The reason behind the good performance of ANN model is that, in a real sense, neural networks are one of the best solutions in search for a few agriculture problems, especially when it comes to predict crop yield. A conclusion has been made that ANN has better explained yield variability rather than other methods. Undeniably, the application of ANN to precision agriculture plays a crucial role in future evaluation of the concept of precision agriculture as a sustainable means of meeting crop yield demands.
- The outcome of this work may assist the agricultural agencies in providing crop strategies for improving oilseeds yield.

- In future, a generalized prediction model for various crops by considering other parameters like bright sunshine and solar radiation can be developed.

Future Developments and Current Trends in the Application of Neural Network to Predict Yield

- Thus, future developments of ANN in predicting crop yield are needed to better focused on current research in order to see at what extend that this approach can go further in precision agriculture. It remains unknown whether ANN is applicable for different intended goals such as to optimize agriculture productivity, minimize the use of natural or man-made resources and also reduce harmful environmental impacts. Furthermore, the factors affecting yield may be different from year to year.
- Current trends in crop yield prediction can be found to focus on soil characteristics parameters, electrical conductivity (ECa) and genotype of seed which affected crop growth and yield performances. The past decades of observational papers have shown that these focus points were related to yield but less attention was given whether those properties were really influencing those crop's yield or not. Lately, the trend is toward a greater interpretation of the relationship between these properties to a crop's yield at a specific location and point in time. With the existence of ANN, more predictions and interpretations of achieving higher yield can be realized.
- Sensitivity analysis must be conducted to detect the robustness of every ANN model because different choices of function and parameters in ANN models would influence the performance of simulation

The thesis is ended with Chapter-V of Summary and Conclusions.

REFERENCES

- Andrew, C. D. 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Published by IOP Publishing Ltd Environmental Research Letters, Volume 13, Number 11.
- Amit, K. Singh. et al. 2014. Conducted a study on development of multi stage district level wheat yields forecast models. Journal of Agricultural Physics ISSN 0973-032X. Vol. 14, No. 2, pp. 189-193.
- Arun, K., Naveen, K., Vishal, V. 2018. Efficient Crop Yield Prediction Using Machine Learning Algorithms. *International Research Journal of Engineering and Technology (IRJET)* Volume: 05 Issue: 06 | June-2018.
- Askar, Choudhury. 2014. developed a crop yield prediction using time series models. *Journal of Economics and Economic Education Research, Volume 15, Number 3, 2014.*
- Bappa Das, Bhakti Nair, Viswanatha. K., Reddy and Paramesh Venkatesh (2018) Evaluation of multiple linear, neural network and penalized regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology* <https://doi.org/10.1007/s00484-018-1583-6>
- Betty, J., Shem, G and Everline, Opere. 2017. On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County, Kenya.
- Chethana, H. M. 2009. Prediction models in teak based agro forestry systems in Northern transitional zone of Karnataka. M. Sc Thesis, University of Agriculture Science, Dharwad.

- Christopher, J. M., Guillermo, A. B., and James, W. Jones. 2009. Use of climate indices to predict corn yields in southeast USA. *Int. J. Climatol.* **29**: 1680–1691.
- Demšar, J. 2006 Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7:1–30
- Dhekale, B.S, Mahdi SS, Dalvi TP, Sawant PK .2014. Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. *J Agrometeorology*16:238–239
- Dhivya B. H, Manjula R, Siva Bharathi S, Madhumathi., R. 2017. A Survey on Crop Yield Prediction based on Agricultural Data. *www.ijirset.com* Vol. 6, Issue 3, March 2017.
- Diwan, U.K., H.V. Puranik, G.K. Das and Chaudhary, J.L. 2018. Yield Prediction of Wheat at Pre-Harvest Stage Using Regression Based Statistical Model for 8 District of Chhattisgarh, India. *Int.J.Curr.Microbiol.App.Sci.* 7(01): 2180-2183.
- Friedman J, Hastie T, Tibshirani R. 2009. glmnet: Lasso and elastic-net regularized generalized linear models. R Package Version 2009:1
- Geetika and Rohit, B. 2018. Crop yield prediction using data mining: an efficient data modeling approach. *International Journal of Engineering & Technology*, 7 (2.27) (2018) 128-131.
- Ghosh, K., et.al 2014. Development of crop yield forecast models under FASAL- a case study of *kharif* rice in West Bengal. *Journal of Agrometeorology* 16 (1): 1-8 (June 2014).
- Goyal, M. and Verma, U. 2015. Development of weather-spectral models for pre-harvest wheat yield prediction on agro-climatic zone basis in Haryana, *International J. of Agricultural and Statistical Sciences*, 11(1), 73-79.

- H.M. Meena, K.S. Jadon and D.V. Singh. 2017. Significance of weather forecasting on oilseeds crop production in India: A-review. *Agricultural Reviews*, 39(2) 2018: 157-162 Print ISSN:0253-1496 / Online ISSN:0976-0539
- Jamme, Z., *et.al.* 2014. Validation of prediction models based on lasso regression with multiply imputed data. <http://www.biomedcentral.com/1471-2288/14/116>
- K. Menaka., and N. Yuvaraj. 2016. A survey on crop yield prediction. *Indian Journal of Innovations and Developments* Vol 5(12), October 2016.
- Kavita Bhatt, K. K. Gill and Sandeep Singh Sandhu. 2015. conducted a study on Comparison of Different Regression Models to Predict Mustard yield in Central Punjab. *School of Climate Change & Agricultural Meteorology Punjab Agricultural University, Ludhiana.*
- Katerina, Tsakiri., Antonios, Marsellos and Stelios, Kapetanakis. 2018. Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York. *Water* 2018, 10, 1158
- Ki-Young, Lee., *et.al.* 2017. Comparison and Analysis of Linear Regression & Artificial Neural Network. *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 12, Number 20 (2017) pp. 9820-9825.
- Koide N, Robertson AW, Ines AVM, Qian J-H, Dewitt DG. 2012. Prediction of rice production in the Philippines using seasonal climate forecasts. *J. Appl. Meteorol. Climatol.* **52**: 552–569.
- Kuhn M. 2008. Building predictive models in R using caret package. *J Stat Software* 28:1–26
- Jin-Ki Park, Amrita, D., and Jong-Hwa, Park. 2017. Developed an integrated model for predicting rice yield with climate change. *Int. Agrophys.*, 2018, 32, 203-215 doi: 10.1515/intag-2017-0010.

- Leyla, Muradkhanli. 2018. Neural Networks for Prediction of Oil Production. *IFAC PapersOnLine* 51-30 (2018) 415–417.
- Lobell, D.B., Burke, M.B. 2010. On the use of statistical models to predict crop yield responses to climate change. *Agric. Forest Meteorol.* (2010), doi: 10.1016/j.agrformet.2010.07.008.
- Manjula, E., and Djodiltachoumy, S. 2017. A Model for Prediction of Crop Yield. *International Journal of Computational Intelligence and Informatics*, Vol. 6: No. 4, March 2017.
- Menaka, K., N. Yuvaraj, M.E. and MISTE. 2017. ANFIS based crop yield prediction model. *International Journal of Science, Engineering and Technology Research (IJSETR) Volume 06, Issue 05.*
- Monali Paul, Santosh K. Vishwakarma, Ashok Verma. Analysis of soil behavior and prediction of crop yield using data mining approach. *Computational Intelligence and Communication Networks (CICN)*. 2015; 766-771.
- Mrs.K.R. Sri., Preethaa., M.E. 2016. Crop Yield Prediction. *International Journal on Engineering Technology and Sciences – IJETSTM* ISSN(P): 2349-3968, ISSN (O): 2349-3976 Volume III, Issue III, March- 2016.
- Naidu, V.B., Sankar, A. S. and Leelavathi, C.2014. Trends in area, production and productivity of selected oilseed crops in Andhra Pradesh. *Int J Multidisciplinary Res Develop* **1**: 366-69.
- Nari Kim, Kyung-Ja Ha, No-Wook Park, Jaeil Cho, Sungwook Hong and Yang-Won Lee. (2019) A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015. *ISPRS Int. J. Geo-Inf.* 2019, 8, 240; doi:10.3390/ijgi8050240

- Nikolaos E. Karkalos *et.al.* 2019. A Comparative Study between Regression and Neural Networks for Modeling Al6082-T6 Alloy Drilling. *Machines* 2019, 7, 13; doi:10.3390/machines7010013
- Nishad, D. *et.al* 2016. Estimated models for the area, production and yield of rapeseed mustard in Chhattisgarh and its constituent districts. *Inter J Agric Sci* **8**: 2146-49.
- Pandey, *et al.* 2016. Development of forecasting model on weather parameters for eastern U.P. ISSN: 0974-4908. *Res. Environ. Life Sci.* 9(11) 1370-1372 (2016).
- Pandey, *et al.* 2013. Pre -Harvest Forecast Models Based on Weather Variable and Weather Indices for Eastern U.P. *Journal's URL*: <http://www.soeagra.com/abr/abr.htm>.
- Panwar, S *et.al* 2018 Forecasting of crop yield using weather parameters–two step nonlinear regression model approach. *Indian Journal of Agricultural Sciences (TSI)*, 88 (10). pp. 1597-1599. ISSN 0019-5022.
- Prajakta, P. B., Yogesh, S. Patil. And Dinesh, D. Patil. Improved Crop Yield Prediction Using Neural Network. *IJARIE-ISSN (O)-2395-4396 Vol-3 Issue-2*.
- Ramakrishna, R. and Naveen, K.B. 2013. Forecasting Yield Per Hectare of Rice In Andhra Pradesh. *International Journal of Mathematics and Computer Applications Research (IJMCAR)* ISSN 2249-6955 Vol. 3, Issue 1, 9-14.
- Ramesh, D., Vishnu Vardhan, B. 2015. Analysis of crop yield prediction using data mining techniques. *IJRET Int. J. Res. Eng. Technol.*
- Ranjana Agrawal and S.C. Mehta.2007 conducted a study on Weather Based Forecasting of Crop Yields, Pests and Diseases - IASRI Models. *J. Ind Soc. Agril Statist.* 61(2), 2007: 255-263.

- Ratna., R, L., and Amrender, K. 2011. Weather based forecasting model for crops yield using neural network approach. *Statistics and Applications* Volume 9, Nos. 1&2, 2011 (New Series), pp. 55-69`.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Sanjeev, Puneet Verma and Urmil Verma. 2018. Development of Weather Based Wheat Yield Forecast Models in Haryana. *Int.J.Curr.Microbiol.App.Sci.*7(12): 2973-2978. doi: <https://doi.org/10.20546/ijcmas.2018.712.340>
- Santosha, Rathod., et al.2017. Modeling and forecasting of oilseed production of India through artificial intelligence techniques. *Indian Journal of Agricultural Sciences* 88 (1): 22–27.
- Satvinder, Kour. et al. 2018. developed a forecasting of rice yield based on weather parameters. in kheda district of Gujarat, India. *Int. J. Agricult. Stat. Sci.* Vol. 14, No. 2, pp. 611-615, 2018.
- Sellam, v. and Poovammal, E. 2016. Prediction of Crop Yield using Regression Analysis. *Indian Journal of Science and Technology*, Vol 9(38).
- Shreya., S. Bhanose. et al. 2016. conducted a study on Crop and Yield Prediction Model. *International journal of advance scientific research and engineering trends*. || Volume 1 ||Issue 1 ||April 2016||ISSN (Online) 2456-0774.
- Snehal, S., Dahikar, Sandeep, and Rode, V. 2014. Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. *International journal of innovative research in electrical, electronics, instrumentation and control engineering* vol. 2, issue 1.

- Subhadra Mishra, Debahuti Mishra, GourHariSantra. Applications of machine learning techniques in agricultural crop production: a review paper. *Indian Journal of Science and Technology*.2016, 9(38), 1-14.
- Surya., P. Dr. Laurence A. I. 2018. crop yield prediction in agriculture using data mining predictive analytic techniques. *International Journal of Research and Analytical Reviews (IJRAR)* www.ijrar.org (E-ISSN 2348-1269, P- ISSN 2349-5138).
- Veenadhari, S., Misra, B., Singh, C.D. 2014. Machine learning approach for forecasting crop yield based on climatic parameters. In: 2014 International Conference on Computer Communication and Informatics, Coimbatore, India (2014).
- Vikas, L., and Dhaka, V.S. 2014. Wheat Yield Prediction Using Artificial Neural Network and Crop Prediction Techniques. *International journal for research in applied science and engineering technology (i jras et) Vol. 2 Issue IX, September 2014 ISSN: 2321-9653*.
- Vuong, M., Ngo.et al. 2018. Studied An Efficient Data Warehouse for Crop Yield Prediction. *A paper from the Proceedings of the 14th International Conference on Precision Agriculture* June 24-June27, 2018.
- Tirtha, R., Ranjeet and Armstrong, L.J. 2014. Developed an artificial neural network for predicting crop yields in Nepal. ECU Publications.
- William, W., Guo and Heru Xue. 2014. conducted a study on crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models. *Mathematical Problems in Engineering* Volume 2014, Article ID 857865.

Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua, Wang Juyun. Prediction of crop yield using big data. *8th International Symposium on Computational Intelligence and Design (ISCID)*.2015;1, 255-260.

Xiangtuo Chen, Paul-Henry Cournède. (2017) Model-Driven and Data-Driven Approaches for Crop Yield Prediction: Analysis and Comparison. *International Journal of Mathematical and Computational Sciences* Vol:11, No:7, 2017

Xiuhong, Z., Zheng, M., Yao, W. and Yu, H. 2010. Tea yield prediction model based on SPSS statistical software in Rizhao city. *Journal of Henan Agricultural sciences*. (7): 31-33

RESUME

Name : Surendra A R
Date of Birth : 06/07/1995
Present Address : Room No 114, Sundaram Boys Hostel, College of Agriculture, IGKV, Krishak Nagar, Zora, Raipur, Chhattisgarh 492012
Phones : (+91) 8762681772
E. mail : surisuren2@gmail.com
Permanent Address : S/o Ravanappa, Ammagaripete (v), Jinkalavaripalli (post), Srinivasapur (Tq), Kolar (Dist.), karnataka-572137

Academic Qualification:

Degree	Year	University / Institute
10 th	2011	Govt. P U. College Rayalpad Srinivasapur Taluk and Kolar District.
12 th	2013	Venkatadri Ind. P.U College, Chintamani.
B.Sc. (Seri.)	2017	UAS, Bengaluru.
M.Sc. (Ag.)	2019	IGKV, Raipur.

Professional Experience (If any) : RAWEP 2016

Membership of Professional Societies : Nill

Award / Recognitions : NTS (National Talent Scholarship)

Publication if any : No


Signature



suri suren <surisuren2@gmail.com>

Acknowledgement of your article and assign manuscript number-201911020-reg

1 message

Ijcmass Editor-in-chief <editorijcmass@gmail.com>
To: surisuren2@gmail.com

Wed, Oct 16, 2019 at 6:11 PM

Dear Sir/ Madam,

We hereby acknowledge the receipt of the manuscript of the paper entitled "**Comparison of ANN and LASSO models for prediction of oilseeds yield based on weather parameters in India**" submitted by your good self.

The paper will be processed and after obtaining the approval from the* *reviewers of the Editorial Board and it will be published in the **Vol 8 (11) of International Journal of Current microbiology and applied sciences (IJCMAS) to be released in November- 2019. doi: 10.20546 /ijcmass.**

Kindly note down Your manuscript no*(**No:IJCMAS-201911020**) for further communication.

***We thank you very much for your patronage with *IJCMAS*!**

PAYMENT OF PROCESSING FEES

Since the Processing fee of your paper, you are required to pay a sum of **Rs.4000/-** towards publication charges.

<http://www.ijcmass.com/guideline.php#FC>

For the said amount has to be send immediately

Mode of payment

1) Transfer the money to the following account

Shift Code : SBININBB226

Account name : EXCELLENT PUBLISHERS

Account no. : 00000037797246206

Bank Name : State Bank of India,

**Branch Name : Kancheepuram branch,
Tamilnadu, India.**

Bank code : 853,

IFSC Code : SBIN0000853

ACCOUNT TYPE : CURRENT ACCOUNT

After transfer the money, send the Transaction details (Scanned copy of receipt) or payment details to this e.mail ID (editorijcmass@gmail.com) with in 48 hours.

(If available State bank of India use wire)

Please note that the consideration of publication of your paper in the current issue is subject to the receipt of publication charges. Therefore you are informed to transfer the amount or send the amount without any delay.

National Academy of Agricultural Sciences (NAAS) :

NAAS Score: *5.38 (2019) [Effective from January 1, 2019]

[For more details click here](#)

Index Copernicus ICI Journals Master List 2017 - IJCMAS Results

I would like to inform you that the ICV value (2017) has been issued by Index Copernicus, you may click the following link for the IJCMAS Index Copernicus's value (ICV - 2017).

Information about the indexation in 2017 can be found at: <http://journals.indexcopernicus.com/International+Journal+of+Current+Microbiology+and+Applied+Sciences+IJCMAS,p2935,3.html>

ICV 2017: 100.00

Thank you for your cooperation.

--

Regards

Dr.M.Prakash

Editor-in-chief

International Journal of Current Microbiology and Applied Sciences

+91-9842641794

www.ijcmas.com

www.ijcrar.com

www.excellentpublishers.com