PRE-HARVEST FORECASTING MODELS AND TRENDS IN PRODUCTION OF BANANA (*MUSA* SPP.) IN KERALA

by

SHARATH KUMAR M P

(2014-19-101)

THESIS

**Submitted in partial fulfilment of the
requirement for the degree of**

MASTER OF SCIENCE IN AGRICULTURE

**Faculty of Agriculture
Kerala Agricultural University**

**DEPARTMENT OF AGRICULTURAL STATISTICS
COLLEGE OF AGRICULTURE
VELLAYANI, THIRUVANANTHAPURAM- 695 522
KERALA, INDIA
2016**

DECLARATION

I, hereby declare that this thesis entitled **"Pre-harvest forecasting models and trends in production of banana (*Musa* spp.) in Kerala."** is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title, of any other University or Society.

Place: Vellayani
Date : 09/09/2016

SHARATH KUMAR M P
(2014 -19-101)

CERTIFICATE

Certified that this thesis, entitled "**Pre-harvest forecasting models and trends in production of banana (Musa spp.) in Kerala.**" is a record of research work done independently by **Mr. Sharath Kumar M P** under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship or associate ship to him.

Place: Vellayani
Date: 09/09/2016

Dr. Vijayaraghava Kumar
(Chairman, Advisory Committee)

Professor and Head
Department of Agricultural Statistics,
College of Agriculture, Vellayani
Thiruvananthapuram -695 522

CERTIFICATE

We, the undersigned members of the advisory committee of **Mr. Sharath Kumar M. P.,** for the degree of **Master of Science in Agricultural Statistics**, agree that the thesis entitled **"PRE-HARVEST FORECASTING MODELS AND TRENDS IN PRODUCTION OF BANANA (*Musa* spp.) IN KERALA,"** may be submitted by **Mr. Sharath Kumar M. P.**, in partial fulfilment of the requirement for the degree.

9/9/16

Dr. Vijayaraghava Kumar
(Chairman, Advisory Committee)
Professor and Head,
Department of Agricultural Statistics,
College of Agriculture, Vellayani.

Smt. Brigit Joseph
(Member, Advisory Committee),
Associate Professor,
Department of Agricultural Statistics.
College of Agriculture, Vellayani.

Dr. Elsamma Job
(Member, Advisory Committee)
Professor ,
Department of Agricultural Economics,
College of Agriculture, Vellayani.

Dr. Babu Mathew P
(Member, Advisory Committee),
Professor and Head,
Instructional Farm,
College of Agriculture, Vellayani.

Dr. Mallikarjuna. G. B,. 9/9/2016
(External Examiner)
Professor and Head,
Department of Agricultural Statistics,
University of Agricultural Sciences,
GKVK,,Bengaluru-560065.

ACKNOWLEDGEMENTS

"Gratitude is the memory of the heart"

With enchanting and ever refreshing memories...................... my academic course has approached and encouraged .It's my bouden duty to remember all those personalities and moments that, as the seeming beautiful petals of fragrant flower. I have to express my heartfelt gratitude to a plethora of personality whose untiring inspiration and invaluable assistance on the other way, brought a sort of perfection to this little effort of mine, which I feel has attained.

My enthusiasm was the rock rolling down the mountain which the Sisyphean, ***Dr. Vijayaraghava Kumar*** *the Chairman of my advisory committee and Professor and Head, Department of Agricultural Statistics. His Excellency guided me on throughout the study, his encouragement, inspiring guidance, valuable suggestions and constructive criticisms were always a source of inspiration to me. To have worked under his valuable guidance, I must confess, was a rare privilege. This work would not have been complete without his valuable help and support so I express my profound and heartfelt gratitude to him.*

I wish to express my sincere gratitude to ***Smt. Brigit Joseph****, Associate Professor, Department of Agricultural Statistics and member of advisory committee, for the help rendered for the smooth conduct of research work, co-operation and critical evaluation of thesis.*

I am grateful to ***Dr. Elsamma Job****, Professor, Department of Agricultural Economics and member of advisory committee for her valuable suggestions, timely support and critical evaluation during the course of this work.*

I have been very much fortunate in having ***Dr. Babu Mathew P****, Associate Professor and Head, Instructional Farm, College of Agriculture, Vellayani, in my advisory committee. I acknowledge with sincere thanks, the remarkable help and cooperation extended by him.*

I am obliged to ***Dr. Ranganath H.K****, ARS scientist, IASRI, New Dehli and* ***Sommanna , P.hd ,IASRI, New Delhi*** *for their valuable guidance, teaching, inspiration to get a bit of knowledge in statistics and moral support during my M.Sc. study.*

I thank Department of Economics and Statistics, Thiruvananthapuram for offering me help during my data collection.

The field experiment was a novel experience. I am only glad to mention that the response from the farm labors and the officials of the instructional farm has been encouraging to conduct the experiment. I have great appreciation for their righteousness and readiness to extend any amount of cooperation throughout the experiment. I thank them for the gestures of goodwill and friendliness.

How can I wink my eyes over the spirit, enthusiasm and earnestness of my dear friends ***Suni,Jitendra,Tamil, Indraja, Dharmendra,Ambi, Raghu, Giri, Shiva, Alur, Siji, Elsa, Anjali*** *and* ***Anju(Golden Group).*** *I owe my gratitude to all these people because of whom my post graduate experience has been one that I will cherish forever.*

I also have been highly fortunate in having many affectionate seniors whose hands were evident at every moment of tension, anxiety and achievements. I am ever grateful to ***Ravi kishore , Darshan S, Siddesh N, Gangadhar, Ramling, Srikant , Seenu, Sainath , Prithin, Bibishan and Kranti patil***

I thank to my department staff ***Leji, Smitha, Vinitha,Chitra, Mohandas sir*** *and my lovely Juniors* ***Raki, Jaslam, Vishnu , Neethu, Amogh, Salma, Manu, Madan*** *for their help and support. I am also indebted to my many student colleagues for providing a stimulating and fun filled environment in KAU.*

My diction is too poor to translate into words the sense of gratitude, heartfelt respect and affection to my noble mother ***Smt. Gayathri****, my father* ***Sri. Puttaraju*** *(Agriculturist) and my sweetest sister* ***Ramya****, without these personalities everything was incomplete to me,*

Since for their affection, encouragement and love bestowed upon me throughout the course of my study and research for their encouragement to complete my M.Sc. programme.I would also like to express my sincere apologies, if ever I failed to mention any names. I am intensely grateful to one and all for being a part in the triumphant completion of the study.

The sprawling complex and sylvan surroundings of KAU have served to assuage the agonies of my heart at times of despair. My heart is with it, for I liked it; I loved it.

Darkness never found me or my path and I know why....The Lord Sai Baba is my shepherd, I shall not want....

Sharath kumar M P

TABLE OF CONTENTS

LIST OF TABLES

9

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
ANNs	Artificial neural networks
CAGR	Compound Annual Growth Rate
df	Degrees of freedom
et al.	Co- workers/ co- authors
Fig.	Figure
FAO	Food and Agriculture Organization
i.e.	that is
IQR	Inter Quartile Range
KAU	Kerala Agricultural University
kcal	Kilo calories
MSL	Mean sea level
MLR	Multiple linear regression
PCA	Principal component analysis
PED	Percent error deviation
R^2	Coefficient of Determination
$\overline{R^2}$	Adjusted R^2
RMSE	Root mean square error
SMW	Special marine warning
S.D	Standard deviation
SSE	Error sum of square
VIF	Variance inflation factor
viz.	Namely

Introduction

1. INTRODUCTION

Banana (*Musa* spp), is the most important fruit crop of many tropical and subtropical regions in the world particularly, India. Banana contributes significantly towards nutrition, fiber, supplement, nourishment, vitamin B_6, potassium, protein and caloric intake providing 89 kcal per 100 g of fruit.

In world scenario, banana was first cultivated in Southeast Asia around 5000 years ago and according to Food and Agriculture Organization (FAO, 2015) world production is around 140 million metric tonnes of which, India contributes 30.18 % followed by china (9.35%) and Philippines (8.59 %). India is the largest producer of banana in 2015. The country has exported 72,689 tonnes of banana during the year 2014-15. The major destinations are UAE, Saudi Arabia, Iran, Kuwait and Bahrain. India produces 30 per cent of the world's banana, but export is negligible at less than 1 per cent. Other countries such as the Philippines and Ecuador, that were earlier ranked 8 and 10 respectively in production, had shot ahead and are currently the top exporters. Moreover, India's post –harvest facilities are still in infancy, especially as there is poor cold chain infrastructure (Agribusiness, 2015).

Under national scenario, Banana occupies 20% of the total area under crops in India. Tamilnadu has the first rank in the list of banana producing states followed by Gujarat, Maharashtra, and Karnataka. However, more than 27 % of total banana production takes place in Maharashtra and Gujarat states. Banana is cultivated in sufficient acreage and in different agro climatic conditions and thus in a position to meet the large demands from importing countries on a continuous basis, provided planting and cultivation is well planned. Kerala stands 11th position in production and fifth position in area under cultivation. Banana is one of the major fruit crops of the state occupying first position in area (61 thousand ha) with production of 5.45 lakh tonnes (GOK, 2016) together with other plantains.

However, the cultivation of banana has occupied pride of place in agrarian economy of Kerala. Banana, basically a tropical crop, grows well in a temperature range of 15°C to 35°C with relative humidity of 75-85 per cent. It prefers tropical humid lowlands and is grown from the sea level to an elevation of 2000m above mean sea level.

Kerala is a deficient state in production and productivity of banana. In 1991-92 productivity was 13034 kg ha^{-1}. It was decreased substantially for past twenty five years to 8806 kg ha^{-1} in 2015. There was an overall decrease of 32.4 % in banana production from 1991-92 to 2014-15. Though even there is an increase in area under cultivation every year, production is decreasing year by year due to imbalance in environmental factors, poor post harvest facilities, price changes, weather characters, agronomic management and biometric characters.

Crop yield forecasting methods and early warning systems are strategies to improve agriculture and rural statistics across the globe. Reliable forecasts of crop production before the harvest constitute a problem of topical interest. Such forecasts are needed by the Government for making policy decisions regarding procurement, distribution, buffer stocking, import - export, price fixation and marketing of agricultural commodities, agro-based industries and for proper planning of their operations. Such information comprises pre-harvest crop forecasting models of total production estimates. This involves building up suitable forecast models using growth and yield characters of crop. Plant characters are integrated effects of all the factors affecting the production. By taking observations throughout the crop growing season, deterministic crop model(s) can be developed to give a precise scientific and independent measure of yield as early as possible during the crop growing season. Difference between forecast and final estimates are analyzed after harvest. Linear and non linear models by regressing yield as a function of the above characters can be used for predicting yield which indirectly influences optimization of crop yield, price stability and insurance coverage. Models developed based on primary data were scarce in banana and this attempt is aimed at developing pre harvest models in four of the popular cultivars grown in Kerala.

1.1 Objectives of the study

This study was taken up with the following objectives

1. To develop models for early forecasting of yield in four major banana cultivars grown in Kerala *viz.*, Nendran, Robusta, Redbanana and Njalipoovan.
2. To carry out the time series analysis of the trend in area and production of banana in Kerala.

1.2. NATURE AND SCOPE OF THE STUDY

The present work is an attempt to make pre harvest forecasting of yield of banana, based on which suitable management practices can be adopted to get better returns to farmers compared to old practices. The time series analysis on area, production and productivity over a long period incorporating additional information on price and climatic factors, in any crop will give models at state level indicating the behavior of the system quantitatively. Existing data will be used to identify the rate of change in the production system in relation to the changes of different related factors.

1.3 Limitation of the study

The study has been done as part of the M. Sc. programme and is limited by time and resource constraint.

1.4 Organization of study

For analytical convenience and clear exposition of the results of the present study, the thesis has been organized into six chapters including the introduction chapter, which highlights the importance of the topic, objectives, scope and limitations of the study. The second chapter deals with the review of literature including the findings of related studies. The third chapter highlights the methodology adopted including description of the study area, nature and sources of data and the analytical techniques employed in the study. The results of the study are presented in the fourth chapter. Discussion and summary with policy implications of the study are presented in the fifth and sixth chapter respectively.

Review of Literature

2. REVIEW OF LITERATURE

A reliable estimate of a crop yield well before the harvest is of considerable importance in formulating policy regarding planning of crop procurement, storage, distribution, price fixation, movement of agricultural commodity, import-export plans and marketing. Usually such pre-harvest estimates of the yield are based on the biometrical characters. The objective of this study is selection of suitable characters and to identify in which growth stage these variables explain a large part of variation in the yield. This would enable the researcher to steer the study in the right direction, collect the appropriate data and to draw meaningful results through best models out of it. Keeping in view the objective of the study, the reviews are presented under the following headings.

2.1 Pre-harvest forecasting models.

2.2 Trend analysis for area and production of banana

2.1 Pre-harvest forecasting models

A number of studies have been undertaken to study the pre-harvest forecasting models, which comprises of correlation, multiple linear regression studies and principal component analysis. The crop forecasting studies mainly use associations among biometric measurements of the crops on yield as well as effect of long term climatic variations on yield. In the present study an attempting has been made for prediction of crop yield based on biometric characters.

2.1.1 **Correlation and multiple linear regression analysis**

Singh *et al.* (1976) in their study, selected biometrical characters (number of tillers per plant, height of main tiller, length of ears, number of green leaves and diameter of the main tiller at the base) for pre-harvest forecasting of wheat yield. A multistage random sampling technique was adopted for selection of the plots. They observed that the correlation between yield of the crop and number of tillers were positive and highly significant, and found that 26 to 36 per cent of variation in the yield is explained by biometrical characters by multiple linear regression analysis.

Singh *et al.* (1979) carried out a pilot study for four years in the state of Bihar and west Bengal for pre-harvest forecasting of yield of jute crop using biometrical characters. They adopted multiple linear regression models between yield and biometrical characters. Multiple correlation coefficients gave high significance between yield and explanatory variables (number of plants, height of plants, and diameter of plants). They concluded that plant density and plant height contribute 30 to 50 per cent variation explained in jute yield and showed that appropriate time of forecast is about 2-3 months after sowing.

Jain *et al.* (1981) conducted study on forecast model for obtaining pre- harvest estimates of sugarcane yield in Meerut district (Uttar Pradesh). Correlation coefficient between yield and number of canes and their height were positive and highly significant for all periods except for the first stage. They could explain about 70 to 80 per cent of the variation using regression technique when the crop was about seven to eight months old. Further they concluded that it was possible to forecast the crop yield about two to three months before harvest.

Chandrahas *et al* (1983) analyzed three year data of Prakasham District (A.P) for tobacco crop. The biometrical characters chosen were plant population per plot, number of curable leaves per plant, plant height and breadth of tenth leaf. Correlation of cured leaf yield with plant population and number of curable leaves were significant in all cases when the crop was more than 10 weeks old, with height it was significant when the crop was 12 to 17 weeks old. The results of regression analysis showed that these biometrical characters explain about 40 to 70 per cent of variation in flue cured yield of tobacco grown in black soil and about 50 per cent of that grown in red soil. They found that a pre-harvest forecast for tobacco yield is feasible when the crop is 12 weeks old.

Aneja and Chandrahas (1984) in their study used regression model for forecasting of cotton yield. The variables selected are plant population and number of bolls, these variables explained only 35 to 40 per cent of variation in crop yield before harvest but with the inclusion of first picking yield as an explanatory variable in regression model, the extent of explained variation in crop yield increased to about 80 per cent. The first picking yield is available about 4 months after sowing the cotton crop, thus the cotton yield forecast on this basis is feasible about two months before harvest.

Agrawal *et al.* (1984) conducted an experiment to develop an integrated model for forecasting the yield of rice at New Delhi (*Oryza sativa.* L) Using a spectral data. A split plot design was followed; soil treatments were taken as main plots and nitrogen levels in sub plots. Observations were recorded from 18 plots and correlation and regression analysis were carried out as per standard procedure. The characters could explain about 95 per cent of the variation in yield. Using the spectral data, rice yield could be forecasted a month before harvest.

Jain *et al.* (1985) used growth indices of plant biometrical characters based on two or more periods for forecasting of hybrid jowar yield in Sangli district of Maharashtra. Here growth indices are the weighted accumulation of observations on plant characters in different periods, weight being respective correlation coefficient between yield and the biometrical characters. The biometrical characters used were number of plants, height, of plants, number of leaves per plant, length and breadth of flag leaf and length of ear head. The variation explained in case of local jowar by simple linear model in which biometrical characters have been used varied from 14 to 34 per cent. However, in hybrid jowar the variation explained was about 50 per cent when the crop was 10 to 12 weeks old. To improve further the explained variation in hybrid jowar, a new model based on growth indices was used in which biometrical characters of two or more periods have been utilized simultaneously. (The period two and three correspond to flowering and milking stages of the crop). By comparison on the basis of R^2 (multiple correlation coefficient) showed that the new model was superior to simple linear model in hybrid jowar. The forecast of yield using this model was 2.12 kg /plot as against 2.4 kg / plot (observed yield rate). The standard error of forecast of yield rate was worked out at 12 per cent. A pre-harvest forecast of hybrid jowar was thus feasible about one month before harvest

Chandrahas (1992) in a study on "modeling for crop yield forecasting using biometrical characters" found that cost incurred and time spent in the collection of data on different biometrical characters also should be taken into consideration for economically feasible conditions. He conducted regression approach under 3 models, namely linear regression, growth indices and principal components. The following two models were estimated in his study:

Multiple Linear regression models

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + e$$

Where, Y and X_i are yield and plant characters respectively

Model based on growth indices

$$Y = b_0 + \sum b_iG_i + e$$

Where $G_i = \sum_{w=n_1}^{n_2} r_{iw} X_{iw}$ is the index of the i^{th} character, w is period

r_{iw} is simple correlation coefficient between yield and i^{th} character in w^{th} period

He concluded that model using growth indices as explanatory variables was found better than conventional linear regression model

Rao *et al.* (1993) used regression models to predict yield of banana from biometrical characters for ten cultivars (genome group) including Njalipoovan. The morphological characters which were found to be influencing yield were mainly leaf characters such as leaf length and leaf breadth. The other characters were stem girth and days to flowering. It was found that prediction of yield was possible even in early stages of the crop.

Groten (1993) gives distinguishing features of conventional and remote sensing type of yield estimation methods. He found that remote sensing data has the potential and capacity to provide spatial information at global scale than conventional methods.

Matthews *et al.* (1994) developed mathematical models for a range of cultivars of cassava. Good agreement was observed between the simulated and measured values.

Onkar *et al.* (1998) used multiple linear regression models to forecast wheat yield based on the data collected from experimental plots in Junagarh (Gujarat) on cropping system. The biometrical characters included are numbers of tillers, height of plants and length of ear head. The experiment was laid out in a spilt – plot design consisting of two treatments, *viz.*, normal and reduced tillage in main plots and four treatments in sub-plots. Two types of functions, *i.e.* linear and quadratic were fitted. It was found that the explanatory variables could explain the variation in yield up to 62 per cent if they considered to be linearly related but 68 per cent provided the function was of quadratic one. The results also showed that forecast models with longer time series are preferred, since the forecasted yield was very close to the observed yield.

Forecasting models developed at Indian Agricultural Statistics Research Institute (IASRI), on different crops results that amount of variation in crop yield explained by fitted regression equations in terms of R^2 values differed from crop to crop, the R^2 values were too low to high. For wheat crop, about 50 to 60 per cent of variation at the stage of two to three months after sowing can be forecasted. For cotton, it was 30 to 40 per cent of variation explained by plant population and number of bolls and for sugarcane, characters like number of millable canes, their height and girth explained over 70 per cent of variation in yield at the stage of six to seven months old crop (Rai and Chandrahas ,1998)

Shamasundaran *et al.* (2002) used regression models to predict yield from plant characters for seven commercial varieties of banana at five stages of growth. The biometrical characters used for prediction of yield were plant height, plant girth and number of suckers. Precision of prediction was higher at later stages.

Kandala and Prajneshu (2002) demonstrated fuzzy linear regression methodology for crop yield forecasting using remote sensed data. It was shown that widths of prediction intervals in respect to Fuzzy linear regression model were much less than those for multiple linear regression models. Their result was that traditional statistical methodologies are not capable of handling data, when two explanatory or response variables are highly correlated.

Priya *et al.* (2003) made an attempt to suggest suitable model for forecasting the yield in cardamom plantations through intensive management. Thirteen growth characters namely tillers per clump, tiller height, leaves per tiller, vegetative buds per clump, bearing tillers per clump, panicle per clump, panicle length, racemes per breadth and recovery per centage were chosen as explanatory variables and they exhibited a precision of about 82 per cent of variation in total tield. Step down regression analysis resulted in the retention of only four characters namely, panicles per clump, racemes per panicle, capsules per raceme and leaf breadth with which yield can be estimated at 77 per cent precision.

Tittonell *et al.* (2005), considered multiple linear regression models including plant height, ear length and ear diameter as explanatory variables. Simple linear regression including only plant height was found to be good to estimate yield of cereals crops.

Anuradha *et al.* (2007) in their study used thirty genotypes of pigeonpea grown during kharif 2005-2006. Correlation studies noticed that number of pods per plant, number of secondary branches per plant; harvest index and dry matter per plant have significant positive association with seed yield per plant. Harvest index had a high positive and significant effect on seed yield per plant. In addition, characters like dry matter per plant, seed per pod, primary branches exerted positive effect. Hence, emphasis should be given on these characters while breeding for seed yield, in pigeon pea.

Ram *et al.* (2007) made an attempt on yield prediction in plantain using multiple regression models. The regression studies on ten plantain varieties revealed that vegetative growth and fruit characters can efficiently be used for yield prediction. In their study growth and fruit characters were used separately as predictor variables for developing yield prediction models. Observations on plant growth and fruit characters were recorded at the time of harvesting. Leaf width, plant girth, number of leaves and plant height predicted yield up to 89.29 per cent. While fruit characters, *viz.*, fingers per bunch, finger diameter, hands per bunch and finger length could predict yield up to 99.57 per cent in main crop. On ratoon crop, plant height, plant girth and number of leaves were responsible for prediction of yield up to 78.69 per cent, whereas fingers per bunch and hands per bunch could predict the yield up to 86.54 per cent.

The wheat yield prediction is better performed when agro meteorological indices combined with spectral index Normal difference vegetation index (NDVI) and trend estimated yield, rather than when they individually in the multiple linear regression model. It was found that agro met-spectral-trend yield model could explain 96 per cent of variation in total wheat yield (Bazgeer *et al.* 2008).

Soares *et al.* (2013), in their study employed Artificial Neural Networks (ANNs) for predicting the bunch weight of banana using yield descriptors like bunch weight , number of fruits per bunch , weight of second hand , weight of fruit, diameter of fruits, length of fruits and weight of rachis. Several multiplayer models of ANNs are developed in order to determine which would best represent weight of the bunch. Among them 10-10-1 neural architecture having best value for R^2 of 84 per cent.

Soares *et al.* (2014) made comparison of ANNs and multiple linear regression (MLR) analysis in prediction of yield in banana plants. Independent variables were selected by stepwise procedures, which select only the significant variables such as number of green leaves at harvest, number of fruits, average weight of the fruit, length of stalks, weight of the rachis. Total number of hands per bunch, perimeter of the pseudo stem, diameter of the fruit and diameter of the stalk were non-significant in the model. The coefficient of determination of the MLR was considered relatively low (71 per cent) to ANN models (91 per cent). According to the analysis, the neural network proved to be more accurate in forecasting the weight of the bunch compared to MLR in terms of the mean prediction error(1.40), mean square deviation (2.29) and R^2 (91 per cent)

2.1.2 Principal component analysis

Raghupathi (2005) conducted a study on pre-harvest forecasting of yield based on biometrical characters in three varieties of cotton (CPD-1, Lakshmi and Jayadhar). Principal component analysis was adopted to develop forecast models which, resulted that first principal component extract alone about 40 per cent of total variability in all varieties. Plant height, number of nodes, number of branches and green boll were contributes 85 per cent of variation in total yield.

Yadav *et al*. (2014) demonstrated the application of principal component analysis in developing statistical models for forecasting crop yield. The time series data on wheat yield and weekly weather variables, *viz.,* minimum and maximum temperature, relative humidity, wind velocity and sun shine hours pertain to the period 1990 to 2010 in Faizabad district of Uttar Pradesh have been used in this study. Weather indices have been constructed using weekly data on weather variables. Four models have been developed using principal component analysis as regressor variables including time trend and wheat yield as regressand. Temperature and wind- velocity have been found to be most appropriate on the basis of adjusted R^2, per cent deviation and Root mean square error (RMSE) (%) for the forecast of wheat yield, two months before the harvest of the crop.

2.2. Trend analysis for Area, production, Productivity, cost of cultivation, price and climatic parameters

Banana is one of the most produced fruit of India, in respect of area it ranks second and first in production. Therefore it was felt that is would be gainful to analyses in detail the growth of area, production and productivity of banana. So review on the following studies which would help to understand the performance of banana production for past years was made.

Subramaniyam (1982) analysed the growth of area, production and productivity of banana in different states of India based on secondary data from the period 1965-66 to 1978-79. According to this study there was an overall increase in area under banana by about 25 thousand hectares between the two periods, which was nearly 25 per cent more over the triennium ending 1967-68.

Raju *et al.* (1989) analysed the Growth and Distribution of Fruit crops in Andhra Pradesh. This study was carried out for Banana, Mango, Citrus, and Grapes, Guava, Papaya, Cashew and other total fruits. The period of study was from 1970-71 to 1982-83. The distribution of fruit crops in Andhra Pradesh over space and time was studied over three-time periods, namely, 1972-73, 1977-78 and 1982-83 and over three regions (Costal Andhra, Rayalaseema and Telangana). For the analysis purpose, compound growth rate and simple regression equation were used.

Indira *et al.* (1990) used the three nonlinear models *viz.*, quadratic, semi log and exponential for analyzing trends in area, production and productivity of banana in Kerala for the period from 1970-71 to 1986-87. Among them quadratic model was significant and could explain satisfactorily the trend in production during the period. They also estimated the output response behavior of banana growers in Kerala in both yield and area response by using Nerlovian lagged adjustment models of linear and double log forms.

Kaushik (1993) analysed the growth of agricultural production in the Kerala state in terms of its determinants, *viz.*, area, productivity and prices. Changes in the gross cropped area, productivity and level of prices are the important components influencing the growth of aggregate agricultural output.

A study by Lalitha (1993) showed that the agricultural income of the State has been growing since the mid eighties. Compared to this, in the period between mid seventies to mid eighties, it showed a mere stagnation. This revealed that growth in agriculture was mainly due to the increase in yield and shift in cropping pattern to high valued crops

Dhindsa and Sharma (1995) analysed the growth rates of area, production and yield of various crops in relation to the cropping pattern changes in the Punjab State during the period 1965-66 to 1990-91, by fitting the exponential function $Y = ab^X$ where Y is the variable for which growth rate is calculated and x is the time, $CGR=(antilog\ b - 1)\times 100$ where CGR is the compound growth rate and concluded that the cropping pattern has changed in favor of only those crops like wheat, Paddy, Moong, American cotton which have shown very high growth rates of production on account of increase in both area and yield growth rates.

Maheshwari (1996) studied the growth pattern of paddy, jowar, ragi, tur and groundnut during the period 1955-56 to 1989-90 in Karnataka. To analyse the influence of green revolution the period of study has been divided into two, *viz.*, 1955-56 to 1966-67 as first period and 1967-68 to 1979-80 as second period. From period I to period II, the exponential trend rates for yield have gone up in all the cases except for groundnut, for production. They increased for all crops except for paddy and jowar, and for area they declined except for tur.

By using forecasted data of total food grain production of our country for 1980 to 2001 Chandran and Prajneshu (2004), reported that growth models like monomolecular model were inappropriate for describing the data set under consideration, compared to Logistic and Gompertz model in studying compound growth rates.

Bal *et al.* (2004) used multiple regression technique to develop models to forecast wheat yield over central Punjab (Ludhiana District) by using weather and production data for (1999-2002). The sensitive periods of statistical and phonological significance were selected for regression analysis. Regression models were highly significant at 5 per cent and explained 69 per cent of variation in yield, based on only weather parameters. The inclusion of technological trend in the model improved the prediction considerably ($R^2 = 0.87$).

Varghese (2007) conducted a study on trend analysis in area production and productivity and price behavior of cardamom in Kerala. He reported that the percentage annual trend growth rate of area, production and productivity of cardamom are -1.21, 4.14 and 5.51 respectively.

Akaike's information criterion(AIC), Bayesian information criterion(BIC), Root mean square error (RMSE) and Mean Absolute Error are the basis for goodness of fit in non-linear Statistical model for trend analysis of area, production, productivity of crops as explained by (Rajarathinam and Parmar, 2011).

Jayasree *et al.* (2012) made an attempt on analysis of price behavior of cassava in Kerala during the period from January 1999 to June 2011. The growth in cassava prices was slow, but with high instability of 36.97 per cent. Time series decomposition analysis was carried out to isolate different components present in the price series. The single exponential smoothening model was found to capture the underlying trend in cassava prices with a low Mean Absolute Percentage Error (MAPE) value of 1.67. The seasonal decomposition analysis revealed that the months of March and April was typically characterized by a peak phase for cassava prices in Kerala. The period from September to November was also marked by a buoyant phase for cassava due to lean market arrival during this period. Even though no pronounced long term cycles were evident, three short cycles were identified during the study period. There were random effects in the price-but it oscillated more or less evenly around the mean value of 100, indicating that random factors were evenly occurring in cassava trading. As the shelf life of cassava tubers are too short, product development and diversification through processing and value addition is suggested to overcome the adverse seasonal and cyclical price movements.

Datarkar, *et al*. (2015) examined the region wise compound growth rates in area, production and productivity of Kharif groundnut in Maharashtra state over different time period-I (1990-91 to 2001-02), period-II (2002-03 to 2012-13)and overall period (1990-91 to 2012-13). The growth in the area, production and productivity of kharif groundnut was estimated by using the compound growth function of the non linear form *i.e.* exponential curve. The study analyzed that area, production and productivity of Kharif groundnut had decreased during the study period. The increase in production of kharif groundnut in Maharashtra during overall period was relatively more as compared to period-I and II. This was due to the increase in acreages under kharif groundnut.

Malhi and Kiran (2015) examined the trends in area, production; productivity of important crops *viz*., cotton, castor and banana of India from 2000-01 to 2011-2012, (12 years). They resulted that banana production showed annual instability at state level and productivity also showed decreasing trend at district level of Gujarat in last 12 years.

According to Agrawal (2015), plant characters can be taken as the integrated effects of various weather parameters and ultimately determine crop yield. She attempted two types of model approaches *i.e.* between year and within year models and said that in India, yield is directly regressed on plant counts and yield contributed characters to obtain forecast model. Finally she found that logistic model having some yield components as dependent variable and an independent 'time' variable generally fits well to the growth process of crop yield components

2.2.1 Growth rates based on climatic factors

Studies by (Cummings and Ray, 1969) proved that there is a relationship between rainfall and crop output. In order to take into account the effect of rainfall on the growth of production area and yield, rainfall adjusted growth rates were worked out employing transformed exponential growth function with "time" as independent variable.

$$Log\ Y_t = \beta_0 + \beta_1 t + \beta_2 log\ R_t + u_t$$

Where,

Y_t = Area/production/productivity

t = time in years 1, 2, 3… etc

R_t = annual rainfall/ rainy days

u_t = random disturbance term

$(Antilog\ \beta_1 - 1) \times 100$ give the rainfall and rainy days adjusted growth rates to the banana production.

Khistaria *et al.* (2004) made an attempt to suggest most pre-harvest forecasting model for Rajkot district of Gujarat state. Weather data of 29 years (1970-71 to 1998-99) used for study. The total rainfall from 25th Special marine warning (SMW) and time trend were also included ad independent variables. The data on average wheat yield considered as discriminant variable. The stepwise regression procedure was adopted by using 26 years data. The prediction equations and forecast of subsequent years were made separately for 22 to 26 years data set. The positive and significant effect of time trend (29 kg ha^{-1} $year^{-1}$) and the beneficial effect of total rainfall (10 kg ha^{-1} $year^{-1}$) on wheat productivity were observed in Rajkot district (Gujarat). The week wise approach using original weather variables was found to be superior to other approaches. This approach provided suitable pre-harvest forecasting model predicting yield 6 weeks before expected harvest (3rd week of February) time and explained more than 95 per cent variation in wheat yield. The deviations between the predicted and observed yields ranged from 0 to 7.51 per cent.

Kaur and Hundal (2009) employed ten yearly moving averages in trend line obtained by regression against time for annual, kharif, summer, rabi and winter season rainfall for past 108 years (1991-2008). It resulted that annual, kharif and summer seasons reveled an increasing trend and no significant trend in rabi and winter season rainfall was observed.

Krishnakumar *et al.* (2009) studied the temporal variation in monthly, seasonal and annual rainfall over Kerala, India for the period of 1871 to 2005 using Man – Kendall rank Statistics and linear trend. Analysis revealed significant decrease in southwest monsoon rainfall while increase in post –monsoon season over the state. Rainfall during winter and summer season showed significant increase in trend, but during June and July significant decline trend, while increasing trend in January, February and April. Increase in the frequency of tropical cyclones during the post-monsoon season may be one of the important reasons for increasing post-monsoon rainfall over Kerala. Majority of plantation crops are likely to benefit due to increase in rainfall during the post-monsoon season.

Trend detection of rainfall for the period 1901–2000 over India was carried out using Sen's non-parametric estimator of slope, the magnitude of trend in hydro-meteorological time series and its statistical significance was assessed by the Mann–Kendall test. (Jain and Kumar, 2012)

2.2.2 Linear and Non-linear regression models

Prajneshu (2002) used deterministic and stochastic models by adding an error term for making appropriate assumptions about non-linear statistical models. To minimize the residual sum of squares of normal equation, he employed iterative procedures like linearization method, steepest descent method & Liebenberg –

Marquardt's method (LM-method). Finally he interprets that LM-method represents a compromise between the other two methods. It is good in the sense that is almost always converges rapidly providing the vicinity of the true parameters.

Nandi and Kanth (2004) adopted different linear and non-linear growth models for the purpose of estimating growth rate and fitting the best model. They used one sample run test and Shapiro- Wilk test for randomness and normality of time series data. They mentioned that R^2 is not an adequate measure for choice of nonlinear models; because the reduced linear model obtained through linearization method or LM method is not having intercept term. In case of model without intercept term, sum of residuals is not zero. So they suggested $\overline{R^2}$ or *RMSE* is a better measure of goodness of fit than R^2.

Sunilkumar and Prajneshu (2008) found superiority of a wavelet procedure over Box-Jenkins approach and non parametric regression method on the basis of mean square errors for the data of country's marine fish production from 1971-2002.

Real world observations are analyzed and integrated into an explanatory model to stimulate behavior of a system. In quantitative models a dynamic system like a crops rate of change in growth is approximated by rate of change of different processes involved in it (Santhoshmithra, 2011).

Arunachalam and Balakrishnan (2012), analysed trends in area, production and productivity of wheat crop using different non-linear models and non parametric models. None of the non-linear model was found suitable to fit the trends in area data. Nonparametric model was found to be a suitable measure to estimate the growth rates of the wheat production. Finally they concluded that Sinusoidal model was found suitable to fit the trends in production as well as productivity of wheat crop grown in India.

Materials and Methods

3

3. MATERIALS AND METHODS

This study mainly utilizes data based on a field experiment on four commonly grown banana cultivars *viz.,* Nendran, Robusta, Redbanana and Njalipoovan in Kerala. This chapter gives briefly the characteristics of the study location, materials, various statistical methods and techniques utilized in analyzing the available data. The details are presented under the following headings.

3.1 Description of the experimental site

3.2 Materials: Collections of primary and secondary data

3.3 Methods: Statistical tools and technique employed.

3.1 Description of the experimental site

A general awareness about the characteristics of the study area is vital to understand the background of research. The field experiment was conducted in the instructional farm, College of Agriculture, Vellayani. The site is situated at $8^{0}25'$ N latitude and 76^{0} $59'$ E longitude at an altitude of 29 m above MSL.

3.2 MATERIALS: COLLECTIONS OF PRIMARY AND SECONDARY DATA

3.2.1. Collections of primary data:

3.2.1.1Growth and yield characters of banana plants and their measurement

Simple random sample of about 50 suckers were selected from 300 sucker samplings of each of the four cultivars. All banana cultivars were planted during October - November 2014. Recommended practices were followed during the cultivation. The following biometrical characters were recorded from each plant.

3.2.1.2. Growth characters and notations used

1. **Weight of suckers [*SW*]**-The sucker weight was recorded in kg at the time of planting.

2. **Girth of suckers [*SG*]** -The girth of sucker was measured in cm.

3. **Length of suckers [*SL*]**-Length of sucker was measured from base of sucker to the tip of sucker in cm.

4. **Plant height (** for i^{th} month H_i $i = 2, 3, \ldots 9$**)** - Height of plant was measured in cm from base of the stem at soil level to the axial of the youngest leaf of plant.

5. **Girth of plant (**for i^{th} month G_i $i = 2, 3, \ldots 9$**)** - Girth of pseudo stem at base level above the soil level was measured in cm.

6. **Number of leaves (** for i^{th} month L_i $i=2, 3, \ldots 9$)-The number of leaves which were photo synthetically active (green) were counted in each plant at every month.

7. **Leaf area (of D leaf) [*LA*]** -Leaf area of D-leaf (fully opened top 3rd leaf in 4th month of growing) was measured and calculated using equation (by Murray, 1960)

$$LA = L \times W \times 0.8$$

Where,

LA: leaf area in cm^2
L: length of lamina measured in cm from base of leaf up to tip
W: width of leaf was measured from broadest part of lamina

3.2.1.3. Bunch Observations

1. **Bunch weight (*BW*)** – the whole bunch structure was weighted (bunches, stalk, and rachis) and recorded in kg
2. **Weight of second hand (*WSH*)** – this is the reference hand of bunch. Weight of second hand was recorded in kg.
3. **Weight of the fruit (*FW*)** – the weight of the fruit was measured in grams (g) of central finger in the external row of the second hand (reference hand) was obtained individually from each bunch.
4. **Length of the fruit (*FL*)** – measured in cm over the external curvature of the fruit or central finger in the external row of fruits in the second hand from the base to the apex of the fruit (disregarding the peduncle and the apex)
5. **Girth of the fruit (*FG*)** – measured in cm at the median part of the central fruit in the outer row in the second bunch.
6. **Numbers of fruits (*NF*)**-Number of fruits were counted from each bunch

3.2.1.4. New added parameter

Volume of sucker (*SV*) - sucker volume approximated from formula of volume of cone and measured in cm^3

$$SV = \frac{1}{3}\pi \times ({}^{SG}\!/_{2})^2 \times SL$$

3.2.2. Collections of secondary data:

The data pertaining to area, production, productivity, cost of cultivation, whole sale price and climatic factors were collected for the period of 25 years (1990 to 2015) from the publications of Department of Economics and Statistics, Thiruvananthapuram, Kerala.

Area (ha), production (mt), productivity (kg ha^{-1}), whole sale price (Rs qtl^{-1}), rainfall (mm) and rainy days were collected for the study period. Available data on cost of cultivation of banana (Rs ha^{-1}year^{-1}) was collected for last 15 years (2000-2015).

3.3. METHODS: STATISTICAL TOOLS AND TECHNIQUES EMPLOYED.

The methods of statistical analysis and their details are presented below

3.3.1. Detection and removal of outliers and leverages

Generally data are influenced by outliers and extreme values, so the foremost step in the analysis is scrutiny of data. To overcome this problem outlier diagnostics is carried out using box plots. The box plot uses the median and the lower and upper quartiles (defined as the 25th and 75th percentiles). If the lower quartile is Q_1 and the upper quartile Q_3, then the difference $(Q_3 - Q_1)$ is called the inter quartile range (IQR) and it is length of box plot. An outlier is any value that lies more than one and half times the length of the box from either end of the box. *i.e* if a data point is below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$, it is viewed as being too far from the central values (Benjamini, 1988). Similarly leverage is a measure of how far away the independent variable values of an observation are from those of the other observations. If calculated leverage value is less than $(2 \times p \, / \, n)$ (where p is the number of characters and n number of observations) then the observations among predictors are within safe limit (Goodall, 1993).

3.3.2. METHODS OF STATISTICAL ANALYSIS AND DEVELOPMENT OF FORECAST MODEL FOR PRIMARY DATA.

3.3.2.1. Correlation analysis

Simple correlation coefficients were computed among bunch weight of Banana and the above described Biometrical characters (Growth and Bunch observations) separately for each cultivar obtained as below:

$$r = \frac{\sum (x_i - \bar{x})\ (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2\ \sum (y_i - \bar{y})^2}}$$

Where,

$\frac{\sum (x_i - \bar{x})\ (y_i - \bar{y})}{n}$ = Covariance between biometrical variables (x) and bunch weight(y).

$\frac{\sum (x_i - \bar{x})^2}{n}$ = Variance of biometrical variables (x)

$\frac{\sum (y_i - \bar{y})^2}{n}$ = Variance of the bunch weight (y).

Correlation coefficients were compared with table values at *(n-2)* degrees of freedom (df) at the probability at 0.05 and 0.01 level of significance; *'n'* is the number of observations .

3.3.2.2 Multiple linear regressions (MLR)

Linear regression is a method of estimating the conditional expected value of variable *Y* given the value of variable *X*. The variable '*Y*' is called the dependent variable (bunch weight) and variable '*X*' is called explanatory variable (biometrical characters). Multiple Linear regressions analysis provide a statistical tool that allows to examine how multiple independent variables are related to a dependent variable and use it to make much more accurate predictions and identify best model. In most situations, more than one predictor variable will be available. This leads to the following process is called "Multiple linear Regressions Analysis".

$$E(Y|X) = \alpha + \beta X \qquad \text{(Simple linear regression)}$$

$$E(Y|X) = \alpha + \beta_1 X_1 + \ldots + \beta_p X_p \qquad \text{(Multiple linear regression)}$$

Generally the model

$$Y_i = X_{i1}\beta_i + X_{i2}\beta_2 + \ldots. + X_{ip}\beta_p + \varepsilon$$

Where random noise variables $\varepsilon_1 \ldots \varepsilon_n$ are independent and identical distributed $N(0,\sigma^2)$. We can write this in a matrix form as below

$$Y = X\beta + \varepsilon,$$

$Y = (Y_1, Y_2 \ldots Y_n)$, is the vector of the dependent Variable (of bunch weight)

$$X = \begin{bmatrix} 1 & X_{11} & . & . & X_{1P} \\ 1 & X_{21} & . & . & X_{2P} \\ . & . & . & . & . \\ 1 & X_{n1} & . & . & X_{nP} \end{bmatrix}_{n\times(p+1)}$$

is n×(p+1) matrix of explanatory variables

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_p \end{bmatrix}_{p\times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \ldots \\ \varepsilon_p \end{bmatrix}_{n\times 1}$$

The least square estimation of the multiple linear regression coefficients are given as

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \text{and Its sampling variance is} \quad V(\beta) = (X'X)^{-1}\sigma^2$$

3.3.2.2.1 ***Stepwise regression***

After fitting the regression models, an important scaling issue is selection of explanatory variables which really influence yield. Many methods are available for selection of influencing explanatory variables. In deciding on the "best" set of explanatory variables for a regression model, most common method is stepwise regression. In this method introducing the X variables one at time (stepwise forward selection) or by including all the possible X variables in one multiple regression and rejecting them one at a time depending on its influences on Y (stepwise backward elimination) or combination of both forward and backward *i.e* stepwise regression method. The decision to add or drop a variable is usually made on the basis of the contribution of that variable to the error sum of square (SSE), as judged by the F test. (Montgomery, *et al.* 2003)

To test the significance of partial regression coefficients using analysis of variance

$$H_0: \beta_i=0 \quad vs. \quad H_1: \beta_i \neq 0$$

For $i = 1, 2, \ldots, q$

Table.1 Analysis of variance for significance of stepwise regression

Source variation	df	Sum of square	Mean sum of square	F- value
Regression	q	SSR	MSR	MSR/MSE
Residuals	$N-q-1$	SSE	MSE	
Total	$N-1$	SST		

Where,

N is number of observations,
q is the number of independent variables

$SSE = SST-SSR$
$MSR = SSR/q$
$MSE = SSE/N-q-1$

The F statistic is compared with the critical $F_{(\alpha,q,N-q-1)}$, if observed F-value is greater than the critical F, then H_0 will be rejected.

3.3.2.2.2 Goodness of fit statistics

3.3.2.2.2.1Coefficient of determination

The Coefficient of determination (denoted by R^2) is a key part of regression analysis. It is interpreted as the proportion of the variance in the dependent variable is predictable from the independent variables.

The R^2 lies in the interval [0, 1]

- An R^2 of 0 means that the dependent variable cannot be predicted from the independent variables.
- An R^2 of 1 means that the dependent variable can be predicted without error from the independent variables.
- An R^2 between 0 and 1 indicates the extent to which the dependent variable is predictable.

How well the estimated model fits the data can be measured by the value of R^2. In general, the R^2 measures percentage of the variation explained by regressors in the total variability, which is explained by

$$R^2 = SSR/SST = 1 - SSE/SST$$

Significance of the R^2 have been tested using F test

$F = (n - q - 1)R^2/q(1 - R^2)$ followed with $F_{(q,n-q-1)}$ degrees of freedom

3.3.2.2.2.2. Adjusted R^2 ($\overline{R^2}$)

$$\overline{R^2} = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1}$$

As we see from this formula, $\overline{R^2}$ should be always less than R^2, for comparative purpose, $\overline{R^2}$ is a better measure than computed R^2. A adjusted R^2 is a modified version of R^2 that has been adjusted for the number of predictors in the model. The adjusted R^2 increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor variable improves the model by less than expected by chance. Here, P is the number of regressors variables included in the equation, n is number of observations (Theil, 1971).

3.3.2.2.2.3. Detection of multicollinerity

The presence of high linear correlation existing between explanatory variables, causes serious problem in estimation, prediction and inference is called multicollinerity. Variance inflation factor (*VIF*) is an appropriate tool used for the detecting the multicollinerity. VIF measures how much the variance of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related(Montgomery, *et al.* 2003).

Let's consider a model with correlated predictors:

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_kx_{ik}+\cdots+\beta_{p-1}x_{i,p-1}+\varepsilon_i$$

The variance inflation factor for the k^{th} predictor. That is:

$$VIF_k = \frac{1}{1-R_k^2}$$

Where, R_k^2 is the R^2-value obtained by regressing the k^{th} predictor on the remaining predictors.

If, VIF = 1, there is no correlation among the k^{th} predictor and the remaining predictor variables.

If, VIF < 4, there is no such serious multicollinerity among the k^{th} predictor and the remaining predictor variables.

If, VIF >10, signs of serious multicollinerity among the k^{th} predictor and the remaining predictor variables and requiring correction.

3.3.2.2.2.4. Mallow's C_P Criterion

Model consisting of *k* regressors, from that of p regressors (*p*<*k*) were choosen, then whether the model with *p* regressors is adequate or that it does not suffer from lack of fit is studied by Mallow's C_P Criterion. Following criterion for model selection, known as the *Cp* criterion:

$$C_p = \frac{RSS_p}{\sigma^2} - (n - 2p)$$

Where,

n is the number of observations.

RSS_p is the residual sum of squares of p regressors

In choosing a model according to the *Cp* criterion, we would look for a model that has a low Cp value, about equal to p.

3.3.2.2.2.5. Percent error deviation (PED)

PED is the per cent of deviation of estimated value from actual value. The per cent error can be determined when the actual value is compared to estimated value according to the equation below:

$$PED = \frac{| Estimatedvalue - Actualvalue |}{Actualvalue} \times 100$$

3.3.3. PRINCIPAL COMPONENT ANALYSIS

The principal component analysis (PCA) is one powerful statistical method widely applied to find the best low-dimensional representation of the variation in a multivariate data set. We carry out a PCA to capture most of the variation using a small number of new variables (principal components). PCA reduces the original variables into a new set of uncorrelated variables known as principal components (PCs).The PRINCOMP procedure performs PCA in SAS System 9.3 software. As input can use raw data, a correlation matrix, a covariance matrix, or a sum-of-squares-and-cross products (SSCP) matrix. The output data sets containing Eigen values, eigenvectors, and standardized or unstandardized principal component scores (Rao, 1964).

Steps in PCA

1. From the data matrix X of independent variables (biometrical characters) of size K×N (K is the number of biometrical characters and N is the number of observations
2. Calculate the covariance matrix(S) based on X
3. Based on characterstic equation, eigen values (λ_i) and eigen vectors (e) are obtained

$$X = \begin{bmatrix} X_{11} & X_{21} & . & . X_{K1} \\ X_{12} & X_{22} & . & . X_{K2} \\ . & . & . & . \\ X_{1N} & X_{2N} & . & . X_{KN} \end{bmatrix}_{K \times N} \qquad S = \begin{bmatrix} \mathrm{cov}(1,1) & \mathrm{cov}(1,2) & . & . \mathrm{cov}(1,k) \\ \mathrm{cov}(2,1) & \mathrm{cov}(2,2) & . & . \mathrm{cov}(2,k) \\ . & . & . & . \\ \mathrm{cov}(k,1) & \mathrm{cov}(k,2) & . & . \mathrm{cov}(k,k) \end{bmatrix}_{K \times K}$$

Principal components were calculated using eigen values and eigen vectors. First principal component (PC_1) which, eigenvalue with the largest absolute value will indicate that the data have the largest variance along its eigenvector, the direction along which there is greatest variation. PC_2 had the maximum variation left in data, orthogonal to the PC_1. In general, first two- three directions manage to capture most of the variability in the data

3.3.4. SECONDARY DATA ANALYSIS

3.3.4.1 Linear and nonlinear models

Linear and non linear models were fitted for each parameter and identified the best models to study trends of banana production. The following growth models are attempted and their estimation procedures used for developing suitable growth models for banana production.

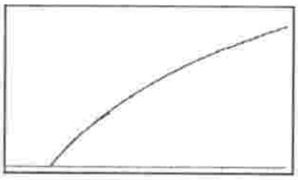
1. **Linear model:** The form of the model is

$$Y = b_0 + b_1(t) + \varepsilon$$

2. Logarithmic model: The curve of the model is,

$$Y = b_0 + b_1 \ln(t) + \varepsilon$$

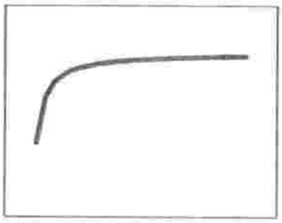
The curve of the model is presented below



3. Inverse model: The equation of the model is,

$$Y = b_0 + \frac{b_1}{t} + \varepsilon$$

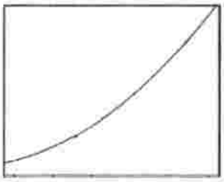
The curve of the model is presented below



4. Quadratic model: The form of the equation is,

$$Y = b_0 + b_1 t + b_2 t^2 + \varepsilon$$

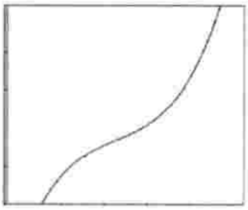
The quadratic model can be used to model a series that "takes off" or a series that dampens, the curve of the model is presented below



5. Cubic model: Here the equation is,

$$Y = b_0 + b_1 t + b_2 t^2 + b_3 t^3 + \varepsilon$$

The representation of the model is given below



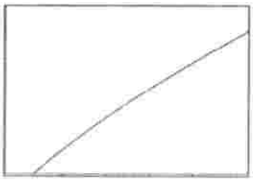
6. Power model: It has the form,

$$Y = b_0 t^{b_1} \varepsilon$$

On transformation, linear form with excluded error term

$$\ln(Y) = \ln(b_0) + b_1 \ln(t)$$

The shape of the curve in this model is



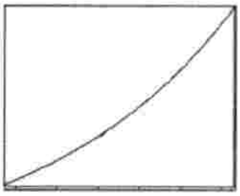
7. Compound model: It has the form

$$Y = b_0 b_1^{t} \varepsilon$$

On transformation, linear form with excluded error term

$$\ln(Y) = \ln(b_0) + t \ln(b_1)$$

The shape of the curve in this model is

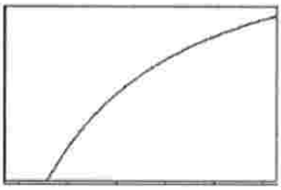


8. S-Curve: It has the form,

$$Y = e^{b_0 + \left(\frac{b_1}{t}\right)} \varepsilon \text{ or}$$

$\ln (Y) = b_0 + b_1 (1/t)$ which is a linear form of the model with excluded error term

The shape of the curve in this model is

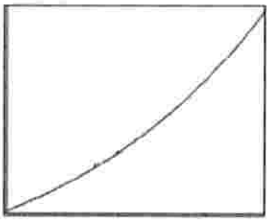


9. Exponential model: The form of the equation is,

$$Y = b_0 e^{(b_1 \times t)}$$

Linear form is , $\ln(Y) = \ln(b_0) + (b_1 t)$

The Exponential curve is given below



Where,

Y = Area /production /productivity / wholesale price / cost of cultivation,

t = years (1991-2015),

b_0, b_1, b_2 and b_3 are constants to be estimated

ln = Natural logarithm

3.3.4.2. COMPOUND ANNUAL GROWTH RATE (CAGR)

Trend analysis in area, production, productivity, whole sale price and cost of cultivation on banana was studied using compound annual growth rate. Student's *t* test was adopted to find significance of regression coefficients (Dhindsa and Sharma, 1995).

Compound Growth Rate in area, production, productivity, farm harvest price, procurement price, scale of procurement of paddy was estimated using exponential function and it is given as

$$\ln(Y) = \ln(b_0) + (b_1 t)$$

Where, t is the time variable,

Y is the variable for which growth rate is calculated

b_1 is the regression coefficient of t on Y.

The Compound Annual Growth Rate (CAGR) is obtained as

$$CAGR\ (\%) = (Antilog\ b_1 - 1) \times 100$$

3.3.4.3. RAINFALL AND RAINY DAYS ADJUSTED GROWTH RATES

In order to take into account the effect of climatic factors on the growth of area, production and productivity of banana, rainfall and rainy days adjusted growth rates were worked out employing the following equation (Cummings and Ray, 1969).

$$\log Y_t = \beta_0 + \beta_1 t + \beta_2 \log R_t + u_t$$

Where,

Y_t = Area/Production/Productivity

t = time in years 1, 2, 3… etc

R_t = annual rainfall/ rainy days

u_t = random disturbance term

(Antilog β_1 *-1) x 100* gives the rainfall and rainy days adjusted growth rates to the banana production

3.3.4.4.Residual analysis

After fitting non- linear model and getting predicted values, checking the assumption of residuals are essential. So two important assumptions made in the model are

1. Errors are randomly distributed

2. Errors are normally distributed

3.3.4.4.1*Test of Randomness: One Sample Run Test*

Randomness of residuals can be tested by using non-parametric one sample run test (Siegel and Castellan, 1988). A run is defined as a succession of identical symbols in which the individual scores or observations originally were obtained. For example, suppose a series of binary events occurred in this order:

+ + - - -+ - - - -+ + - -+ then number of runs =7 for n= 15

If very few runs occur, a time trend or some bunching owing to lack of independence is suggested. If a great many runs occur, systematic short-period cyclical fluctuations seem to be influencing the scores.

Let 'm' be the number of elements of one kind, and 'n' be the number of elements of the other kind in a sequence of $N = m + n$ binary events. If both m and n are less than or equal to 20 (small sample), then the number of runs, r if falls between the critical values, we cannot reject null hypothesis.

For large samples if either 'm' or 'n' is large than 20, a good approximation to the sampling distribution of 'r' is the normal distribution with.

Mean= $(\mu_r) = \frac{2mn}{N} + 1$

and standard deviation $= \sigma_r = \sqrt{\frac{2mn(2mn - N)}{N^2(N - 1)}}$

then H_0 may be tested by $Z = \frac{r - \mu_r}{\sigma_r}$

The significance of any observed value of Z computed by using the equation may be determined from a normal distribution table.

3.3.4.4.2. *Test of Normality: Shapiro-Wilk Test*

Three popular tests for normality are, the Shapiro-Wilk (Shapiro and Wilk, 1965), the Kolmogorov – Smirnov test, Cramer Von Misses test and Anderson-Darling test. If the sample size is less than 2000, the Shapiro-Wilk test is better.

Shapiro-Wilk statistic 'W' is given as

$$W = \frac{\left[\sum_{i=1}^{n} a_i X_{(i)}\right]^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Where,

$i = 1, \ldots, n$

$X_{(i)}$ = ordered sample values

a_i = constants generated from mean, variance and covariance of the order statistics of a sample of size 'n' from a normal distribution.

If the p-value is smaller than the critical value, normally 0.05, H_0 is rejected and we conclude that the population is not normal. The distribution of Shapiro-Wilk statistic is W; ranges in between 0 and 1, highly skewed to the right.

3.4.4 Statistical Software Used for the analysis

SAS.9.3 (licensed by IASRI), IBM SPSS. 20, STATA.12 and MS Excel-07 statistical packages are used for tabulation, correlation analysis, multiple linear regression, principal component analysis, and trend analysis.

Results

4. RESULTS

Keeping in view the objectives of the study, data collected from field and published sources were analysed employing appropriate statistical techniques. This chapter presents the results in line with the objectives of the study under the following headings.

4.1. Summary Statistics
4.2. Correlation analysis
4.3. Multiple linear regression analysis (Stepwise)
4.4. Principal component analysis
4.5. Trend analysis for banana production
4.6 Compound annual growth rate (CAGR)
4.7. Rainfall and rainy days adjusted growth rates to banana production

4.1 Summary Statistics

4.1.1. World banana production

Bananas are the fifth largest agricultural commodity in the world trade after cereals, sugar, coffee and cocoa. India, Equador, Brazil and China alone produce half of total bananas of the world. The world scenario of area, production and productivity of banana for world and India is given in Table.2

Table.2 Area, Production and productivity of banana in world and India during 2012-13

	Area('000*ha*)	Production('000*t*)	Productivity(kg ha^{-1})
World	5577	100823	18078
India	811	30247	36856

Source : (FAO, 2015)

4.1.2. State level banana production

The major banana producing states of India are Tamilnadu, Maharashtra, Karnataka, Gujarat, Andhra Pradesh, Assam and Madhya Pradesh. Table. 3 provides state wise area, production and productivity of banana in India during 2012-13

Table.3 State level Area, Production and productivity of banana for 2012-13

States	Area(in '000*ha*)	Production(in '000*mt*)	Productivity ($mt\ ha^{-1}$)
Tamilnadu	113.7	5136.2	45.4
Gujarat	61.9	4523.4	73.0
Maharashtra	85.0	3600.0	42.3
Andra Pradesh	80.6	3242.7	40.2
Karnataka	104.4	2529.0	24.2
Bihar	31.5	1702.4	54.0
Madhya Pradesh	33.0	1701.0	51.5
West Bengal	41.0	1077.8	26.2
Assam	53.4	837.2	15.6
Kerala	61.6	515.6	8.3
Orissa	24.7	521.3	21.1
Other states	175.3	1873.1	10.7

Source: (NHB, 2014)

4.1.3 Characteristics of banana cultivars

Based on the observations of all growth and yield characters from four banana cultivars at different periods. The mean and standard deviation (S.D) were worked out and the Tables 4, 5, 6 and 7 depicts the mean and S.D of growth and yield characters at different growth period of banana cultivars.

Figure.1. Average growth performance of cultivars at maturity stage

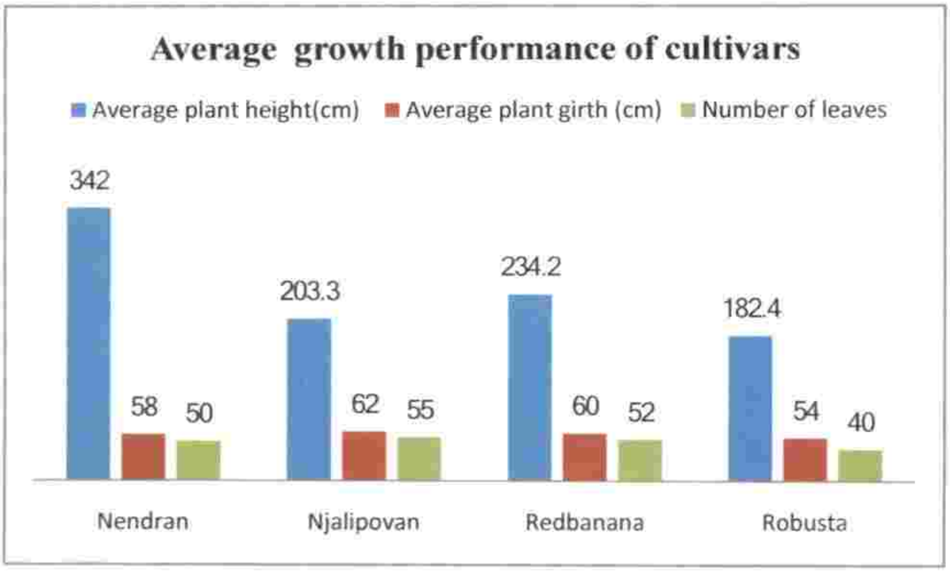


Figure.2. Average bunch weight of cultivars

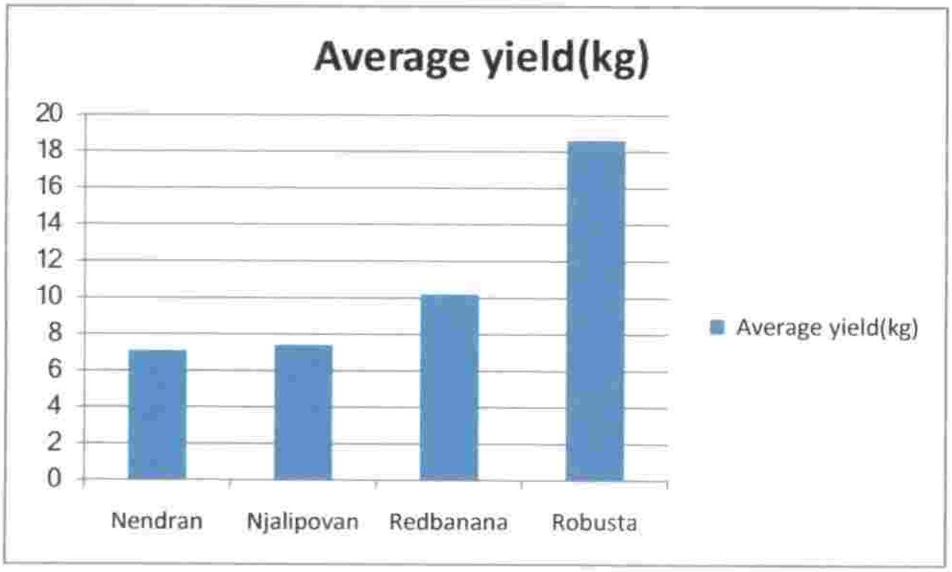


Table .4 Summary statistics of biometric parameters for Nendran

Parameters	BW(kg)	SW(kg)	SL(cm)	SG(cm)	NF	LA(cm^2)	WSH(kg)	FW(g)	FL(cm)	FG(cm)
Mean ± S.D	7.1±1.2	1.4± 0.5	31.8± 6.1	40.1± 5.4	37± 5.1	5159± 11	1.21± 0.3	174± 10	18.1± 2.9	14± 0.87

Summary statistics of growth parameters

Months	2nd	3rd	4th	5th	6th	7th	8th	9th
No. of leaves	7± 2	7± 1	10± 2	8± 1	9± 2	9± 1	7± 1	7± 2
Plant height(cm)	73± 12	80± 12	154± 21	191± 21	258± 22	281± 23	318± 23	342± 22
Plant girth(cm)	20± 3	24± 3	36± 6.1	43± 6	50± 1	59±3	59± 5	60± 7

Table .5 Summary statistics of biometric parameters for Njalipoovan

Parameters	BW(kg)	SW(kg)	SL(cm)	SG(cm)	LA(cm^2)	NF	WSH(kg)	FW(g)	FL(cm)	FG(cm)
Mean ± S.D	7.4±1.3	1.1± 0.34	25.8± 3	36.1± 5.6	4267±736	102± 12	0.8± 0.08	60± 10	15± 1.9	11± 0.64

Summary statistics of growth parameters

Months	2nd	3rd	4th	5th	6th	7th	8th
No. of leaves	4± 1	6± 2	9± 2	9± 2	10± 2	9± 2	10± 3
Plant height(cm)	30± 10.2	45.1± 10.2	79.8± 15.2	90.8± 16	131.4± 21.9	161.4± 29.5	203.3± 37
Plant girth(cm)	12.2± 3.7	15.3± 3.7	18.4± 3.42	23.3± 4	30.5± 4.9	39.0±5.6	59.5± 4.6

Table .6 Summary statistics of biometric parameters for Redbanana

Parameters	BW(kg)	SW(kg)	SL(cm)	SG(cm)	NF	LA(cm^2)	WSH(kg)	FW(g)	FL(cm)	FG(cm)
Mean ± S.D	10.2±3.6	3.9± 1.18	37.3± 4.8	36± 6.29	92± 12	6302±84	1.93± 0.3	180± 38	15± 0.73	14± 0.71

Summary statistics of growth parameters

Months	3rd	4th	5th	6th	7th	8th
No. of leaves	4± 1	6± 1	7± 2	7± 1	7± 1	6 ± 1
Plant height(cm)	42.89± 12.9	52.08± 12.9	74..34± 14.2	134.93± 12.7	178.15± 25.6	234.2± 34.7
Plant girth(cm)	15.4± 4.3	19.4± 4.1	24.8± 4.29	35.6± 7.05	50±8.8	59.3± 9.4

Table .7 Summary statistics of biometric parameters for Robusta

Parameters	BW(kg)	SW(kg)	SL(cm)	SG(cm)	LA(cm^2)	NF	WSH(kg)	FW(g)	FL(cm)	FG(cm)
Mean ± S.D	18.6±2.9	1.5± 1.14	26± 7.5	32.1± 10.1	3827±442	87± 9	2.4± 0.25	150± 15	21.2± 1.6	10.3± 0.9

Summary statistics of growth parameters

Months	3rd	4th	5th	6th	7th	8th
No. of leaves	6± 1	7± 2	6± 1	6± 1	5± 1	6± 1
Plant height(cm)	26.9± 7.2	44.84± 16	50.1± 18.2	86.1± 13.1	174.5± 20.2	182.7± 21.2
Plant girth(cm)	14.9± 4.4	16.2± 3.1	17.1±4.2	31.1± 5.6	50.3±5.6	55.2± 5.6

4.1.4. **Detection of outliers, leverages and multicollinerity.**

Detection of outliers and leverages in data set were done for each cultivar at every month. Calculated leverage values were less than $2 \times p / n$ (0.16) shows that observations among p predictors are within safe limit. Some observations were found as more than this limit and such observations are discarded to carry out the good prediction. Four to five outliers were detected in respective month using box plot and removed to carry out error free regression analysis. In Table 8, 0.02 indicates 2 per cent out of hundred observations are too away from the central values.

Table.8 Measure of outliers and leverages for each month of biometrical characters

Measures	Nendran : months								
	1	2	3	4	5	6	7	8	9
Outliers	0.02	0.02	0.00	0.00	0.00	0.02	0.02	0.04	0.04
Leverage	0.19	0.20	0.18	0.14	0.14	0.18	0.18	0.20	0.20

Measures	Njalipoovan : months							
	1	2	3	4	5	6	7	8
Outliers	0.02	0.04	0.00	0.02	0.00	0.02	0.04	0.04
Leverage	0.14	0.18	0.18	0.16	0.14	0.18	0.18	0.20

Measures	Redbanana: months						
	1	3	4	5	6	7	8
Outliers	0.04	0.02	0.00	0.02	0.02	0.02	0.04
Leverage	0.20	0.18	0.14	0.16	0.18	0.20	0.18

Measures	Robusta : months						
	1	3	4	5	6	7	8
Outliers	0.04	0.04	0.04	0.04	0.000	0.02	0.00
Leverage	0.16	0.16	0.14	0.20	0.18	0.18	0.20

4.2. Correlation analysis

The knowledge regarding association of various characters among themselves is necessary to understand the nature and degree of relationship. The Pearson's product moment correlations between yield and biometrical characters at different plant growth stages (month) are calculated for Nendran, Njalipoovan, Redbanana and Robusta cultivars and the results are presented as follows.

4.2.1 Nendran

The estimate of correlation coefficient between bunch weight and biometric characters are presented in Table 9. All the estimated correlation coefficients between bunch weight and sucker characters were non-significant, which is an indication of non significant relation between bunch weight and suckers characters.

Vegetative characters of Nendran namely plant height, girth and number of leaves were recorded at second to ninth month of growth period. It is evident from the Table that final yield was significantly correlated with plant height at 4th, 5th, 6th and 7th month, plant girth at 3rd, 4th, 5th and 6th month respectively. This clearly indicates that the yield (bunch weight) mainly depends upon the plant height and plant girth attained at different growth stages of the plant. The highest association to the yield was from the plant height (0.787) at 4th month and plant girth (0.832) at 5th month. Fruit characters like number of fruits (0.558) and fruit girth (0.578) at harvesting stage had positive significant correlations with yield. Hence, these bunch characters also can be effectively used for estimation of yield

4.2.2 Njalipoovan

The product moment relationship between yield and biometric characters were presented in Table 10. The results shows that, in the case of Njalipoovan there didn't exists any significant relationship between yield and sucker variables, but SW (0.346) and SG (0.359) had some positive correlation with bunch weight. Growth characters of Njalipoovan namely plant height, plant girth and number of leaves were recorded at second to eight month of growth period. Here also results suggests that bunch weight had significant positive correlation with plant height at 3rd, 4th, 5th and 6th month of growth period. Similar relationship was noticed in the case of plant girth. In addition to this, L_4 and LA were also positive and significant correlation with yield at 1 % significance level. Variables of 4th month was having more correlation than previous months for plant height and girth with yield. Bunch characters namely number of fruits per bunch (0.678), weight of second hand (0.678), fruit weight (0.687) and fruit girth (0.535) also had significant positive relationship with yield at 1 per cent level

4.2.3 Redbanana

The significant product moment correlation coefficient between yield and biometrical characters were calculated for different growth stages are presented in Table 11. Correlation coefficients were negative and non significant in the case of all sucker characters with bunch weight. Vegetative characters of Redbanana namely plant height, plant girth and number of leaves were recorded from third to eighth month of growth period. It is evident from the table that final yield was significantly correlated with plant height and plant girth at 4th, 5th , 6th and 7th month at one per cent significance level. As observed by the correlation analysis in 5th month of growth, plant height (0.913) and plant girth (0.857) had high correlation with bunch weight. So these variables would be the best predictor variables in Redbanana. For harvest stage, number of fruits per bunch (0.777), weight of second hand (0.636) and fruit weight(0.684) had positive relationship with bunch weight and contribute for most part of yield.

Table.9 Correlation between bunch weight and biometrical parameters of nendran

	SW	SL	SG	SV	LA	NF	WSH	FW	FL	FG
BW	0.151	-0.136	0.283	0.168	0.292	0.556**	0.166	0.193	-0.483**	0.578**
Number of leaves for ith month = L_i (i=2,3,…9)										
	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9		
BW	-0.084	0.111	0.011	-0.020	0.184	-0.070	0.043	-0.151		
Plant height for ith month = H_i (i=2,3,…9)										
	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9		
BW	0.053	0.691**	0.787**	0.701**	0.670**	0.830**	0.223	0.249		
Plant girth for ith month = G_i (i=2,3,…9)										
	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9		
BW	0.198	0.589**	0.767**	0.832**	0.705**	0.263	0.091	0.204		

Table. 10 Correlation between bunch weight and biometrical parameters of Njalipoovan

	SW	SL	SG	SV	LA	NF	WSH	FW	FL	FG
BW	0.341	-0.318	0.358	0.247	0.416**	0.675**	0.670**	0.689**	-0.087	0.536**
Number of leaves for ith month = L_i (i=2,3,…9)										
	L_2	L_3	L_4	L_5	L_6	L_7	L_8			
BW	0.037	-0.001	0.669**	0.083	-0.012	-0.071	-0.162			
Plant height for ith month = H_i (i=2,3,…9)										
	H_2	H_3	H_4	H_5	H_6	H_7	H_8			
BW	0.353	0.714**	0.883**	0.793**	0.704**	0.661**	0.143			
Plant girth for ith month = G_i (i=2,3,…9)										
	G_2	G_3	G_4	G_5	G_6	G_7	G_8			
BW	0.346	0.271	0.298	0.390*	0.382*	0.301	-0.038			

* Significant at 5%, ** significant at 1% level

Table .11 Correlation between bunch weight and biometrical parameters of Redbanana

	SW	SL	SG	SV	LA	NF	WSH	FW	FL	FG
BW	-0.132	-0.023	-0.108	-0.128	0.192	0.770**	0.638**	0.688**	0.150	0.298

Number of leaves for ith month = L_i (i=2,3,…9)

	L_3	L_4	L_5	L_6	L_7	L_8
BW	0.273	0.086	0.266	0.160	0.071	-0.164

Plant height for ith month = H_i (i=2,3,…9)

	H_3	H_4	H_5	H_6	H_7	H_8
BW	0.327	0.806**	0.914**	0.570**	0.448**	0.470*

Plant girth for ith month = G_i (i=2,3,…9)

	G_3	G_4	G_5	G_6	G_7	G_8
BW	0.213	0.813**	0.859**	0.464**	0.388*	0.341

Table. 12 Correlation between bunch weight and biometrical parameters of Robusta

	SW	SL	SG	SV	LA	NF	WSH	FW	FL	FG
BW	-0.108	-0.055	-0.103	0.167		0.763**	0.540**	0.610**	0.169	0.367*

Number of leaves for ith month = L_i (i=2,3,…9)

	L_3	L_4	L_5	L_6	L_7	L_8
BW	0.014	0.115	-0.008	0.200	0.165	0.168

Plant height for ith month = H_i (i=2,3,…9)

	H_3	H_4	H_5	H_6	H_7	H_8
BW	0.299	0.739**	0.688**	0.544**	0.426*	0.278

Plant girth for ith month = G_i (i=2,3,…9)

	G_3	G_4	G_5	G_6	G_7	G_8
BW	0.680**	0.795**	0.499**	0.624**	0.336	0.358

* Significant at 5%, ** significant at 1% level

4.2.4 Robusta

The nature and degree of association of biometric characters with yield is presented in Table 12. An attempt on correlating the biometric characters with yield for robusta, provided, significant results in case of plant height, plant girth and some fruit characters. Correlation coefficient (r) was observed for plant height (0.739) and plant girth (0.795) at 4th month indicates highly significant correlation with yield and these variables were contributing more to yield compared to other months. It indicates that the increase in plant height and girth resulted significant rise in bunch weight. Also it is noticed that good relationship between bunch weight and plant height, plant girth with bunch weight in fifth and sixth month. It clearly indicates that plant height and plant girth are the two important characters help to increase the bunch weight. With regard to fruit characters same result was noticed in Robusta such as NF (0.769), WSH (0.545) and FW (0.616) were significantly correlated with yield.

4.3. Multiple Linear Regressions

In agriculture, statistical models play a vital role in prediction of crop yield before harvest of the crops. An effort was made to develop the prediction models for the crop yield with biometrical characters for every month; one which is found out to be significant was selected. In case, there is more than one; the best models were selected on the basis of R^2 values, Adjusted R^2 and Mallow's C_p. Multiple linear regression technique was used to find the linear effect of biometrical characters at different stages on bunch weight of banana cultivars namely nendran, Njalipoovan, Red banana and Robusta. So regression models were fitted for each month of banana growth using explanatory variables.Prediction models for yield were carried out with the help of the biometrical characters. To know which of the predictor variables among included in the model are most significantly contributing to the yield, the stepwise regression was carried out and the models obtained.

4.3.1 Nendran

Prediction models for the crop yield with biometrical characters for every month were attempted. The best models were selected on the basis of R^2 values, Adjusted R^2 and Mallow's Cp criteria and are depicted in the Table.13. In the first month, sucker characters were used as predictor variables in multiple regressions. The results of analysis indicated that, only 14 % of variation was explained by these variables in the first month. None of the sucker characters had significant effect on yield. In the second month, along with sucker variables, plant height, plant girth and number of leaves were used as explanatory variables. But there was no considerable increase in R^2 and adjusted R^2 values in the model. From 3rd month onwards, the predictor variables showed significant increase in the R^2 and adjusted R^2 values. For fourth and fifth month the R^2 and Adjusted R^2 were higher as compared to the previous models with minimum number of explanatory variables, *i.e.* in the 5th month, about 80 per cent of variation was explained by the model with help of two independent variables namely H_4 and D_5. The result of the prediction models confirms that H_4 and G_5 had significant role in predicting the yield prior to harvest of crop.

Yield attributing characters like number of fruits, fruit weight and fruit girth had predominant role in determing the final yield. It contributes to 55 per cent of variation in total yield of Nendran cultivar and these three yield characters can be used for estimation of yield at harvesting stage. The following prediction models were selected as the best for early prediction from regression analysis for nendran

For 4th month, $Y = -0.47 + 0.03H_3 - 0.08L_4 + 0.01H_4 + 0.07G_4$ $(\overline{R^2} = 0.740)$

$(R^2 = 0.775)$ (0.96) (0.009)** (0.064) (0.007)** (0.02)**

For 5th month, $Y = -1.37 + 0.025\,H_4 + 0.10\,G_5$ $(\overline{R^2} = 0.778)$

$(R^2 = 0.805)$ (0.64) (0.005)** (0.01)**

Stepwise forward regression analysis was conducted using all independent variables to identify the best predictor variables and the result of the analysis suggests that H_4 and G_5 were best explanatory variables to predict the yield (Table 17 and 18).

4.3.2 Njalipoovan

Prediction models for bunch weight with growth and yield characters for every month were attempted. Multiple regression models were done in SAS 9.3 software and are presented in Table 14. In the first month, all sucker characters were used as explanatory variables. Among sucker characters, weight and girth explained 16 per cent variation in the total yield. But none of the sucker characters had significant effect on yield. In the second month, along with sucker variables, plant height, plant girth and number of leaves of second month were used as explanatory variables. Here also, all regression models didn't reveal any considerable increase in R^2 and adjusted R^2 values in the model. For second month, model with three variables namely L_2, H_2 and SW were explained 17 per cent R^2 by the model, which was highest in second month. For the 3rd month, the predictor variables showed reasonable increase in the R^2 and adjusted R^2 values. H_3 and G_3 made significant effect on bunch weight with R^2 of 61 per cent, which was rational, compared to first two month. Added new variables in the fourth month made a significant increase in the adjusted R^2 of 80 per cent and results that 81 per cent variation by the model. L_4 and H_4 variables made significant effect on bunch weight, which was highest compared to all months. Subsequently, these two variables has made dominant role in all prediction models with increase in R^2 and adj.R^2. Therefore it was clearly understandable that, fourth month is reasonable and best for prediction of yield prior to harvest in Njalipoovan cultivar. Number of leaves and plant height of fourth month are chosen as best predictor variables form multiple regression analysis.

Yield characters like number of fruits, fruit weight, fruit girth, fruit length and weight of second hand were used for multiple regression analysis. Results of analysis shows that fruit weight and fruit girth were made significant role in estimation of bunch weight at harvesting stage and contributes to 71 per cent of variation in total yield of Njalipoovan cultivar and these two characters s are significant at 1 per cent significance level. The following prediction models were selected as the best for early prediction from regression analysis

For 4th month, $Y= 2.78-0.04G_3+0.07L_4+0.04H_4$ $(\overline{R^2} =0.795)$
$(R^2 = 0.817)$ (0.30) (0.015) ** (0.039)** (0.04)**

For 5th month, $Y=2.98- 0.07G_3+0.08L_4+0.04H_4+0.04G_4-0.0001LA$
$(\overline{R^2} =0.805)$
$(R^2 = 0.834)$ (0.37) (0.03)** (0.03)** (0.004)** (0.03) (0.0007)

Stepwise forward regression analysis was conducted using all independent variables to identify the best predictor variables and the result of the analysis suggests that H_4 and L_4 were best explanatory variables to predict the yield, which are significant at one per cent level of significant(Table 19 and 20)..

4.3.3. Redbanana

Table 15 presents prediction models for bunch weight of Redbanana with growth and yield characters for every month. As done in last two cultivars, in redbanana also in the first month all sucker characters were used as explanatory variables. But results were not satisfactory with sucker characters. None of the sucker characters had significant effect on bunch weight. In redbanana growth characters were recorded from third month to eight month. Plant height, plant girth and number of leaves of third month were used as explanatory variables along with sucker characters. Here also, were attempted regression models for each month. Plant height, plant girth and number of leaves for the third month were not showed sufficiently increase in R^2 and adjusted R^2 values in the model. Model with three variables namely L_3, H_3 and G_3 were explained R^2 of 14 per cent by the model, which was highest in that month. Addition of new variables in the fourth month made a significant increase in the adjusted R^2 of 73 per cent and R^2 of 74 per cent variation was explained by the model. G_4 and H_4 variables had significant effect on bunch weight, because addition of these variables help in increase R^2 as compared to past months. From the fifth and subsequent months, these two variables namely, H_4 and G_4 are played dominant role in all prediction models with an increase in R^2 and adjusted R^2. Therefore it was clearly evident that , fourth month is perfect and best for prediction of yield prior to harvest in Redbanana cultivar with maximum variation in the total yield. Plant height and plant girth of 4th month are chosen as best predictor variables form multiple regression analysis.

Yield characters like number of fruits, fruit weight, fruit girth, fruit length and weight of second hand were used for multiple regression analysis. Results of analysis found that number of fruits and weight of second hand were made significant role in estimation of bunch weight at harvesting stage and contributes to 71 per cent of variation in total yield of the cultivar. The following prediction models were selected as the best for early prediction from regression analysis

For 4th month, $Y=0.12+0.09\,H_4+0.02\,G_4\ (\overline{R^2}=0.738)$

$(R^2 = 0.748)$ (0.83) $(0.024)^{**}$ $(0.06)^{**}$

For 5th month, $Y=0.19-0.02H_3+0.04H_4+0.07H_5+0.06G_4\ \ (\overline{R^2}=0.862)$

$(R^2 = 0.873)$ (0.63) (0.012) $(0.019)^{**}$ $(0.014)^{**}$ $(0.04)^{**}$

Stepwise forward regression analysis was conducted using all independent variables to identify the best predictor variables and the result of the analysis suggests that H_4 and G_4 were best explanatory variables to predict the yield, whichare significant at 1 per cent and 5 per cent level of significance respectively (Table 21 and 22).

4.3.4. Robusta

Table 16 presents prediction models for bunch weight on growth and yield characters at every month of Robusta cultivar. As similar to the other three cultivars, in robusta also all sucker characters were used as explanatory variables. None of the sucker characters had significant effect on bunch weight. In Robusta cultivar growth characters were recorded from third month to eighth month. H_3, G_3 and L_3 were used as explanatory variables in the prediction models from third month onwards. Using of these variables resulted an increase in Adjusted R^2 (0.677) and maximum variation (R^2 of 0.691) in total yield explained by model. From 3rd month onwards, Adjusted R^2 and R^2 was geared up to a rational. Addition of new variables in the fourth month made a significant increase in the adjusted R^2 (of 72 per cent) and resulted that 75 per cent variation explained by the model. From the fifth month, two variables H_4 and G_3 along with fifth month data had dominant role in all prediction models with instant increase the R^2 and adjusted R^2. Therefore it was clearly evident that , fourth month was perfect and best for forecasting of yield prior to harvest in Robusta cultivar with maximum variation in the total yield. Subsequently these two variables made significant role in all months. Plant height at 4th and plant girth at 3rd month are chosen as best predictor variables form multiple regression analysis.

Yield characters like number of fruits, fruit weight, fruit girth, fruit length and weight of second hand were used for multiple regression analysis. Results of analysis found that number of fruits and fruit girth were made significant role in estimation of bunch weight at harvesting stage and contributes to 96 per cent of variation in total yield of cultivar .which was maximum among all cultivars.

The following prediction models were selected as the best for early forecasting from regression analysis

For 3th month, $Y=6.82+0.06H_3+0.27G_3 \quad (\overline{R^2}=0.677)$

($R^2 = 0.691$) (0.62) (0.02)** (0.03)**

For 4th month, $Y=5.37+0.09SL-0.141L_3+0.08H_4+0.259G_3$ ($\overline{R^2} =0.721$)

($R^2 = 0.752$) (1.18) (0.03) (0.14) (0.022)** (0.03)**

Stepwise forward regression analysis was conducted using all independent variables to identify the best predictor variables and the result of the analysis suggests that H_4 and G_3 were best explanatory variables to predict the yield, which are significant at 1 per cent and 5 per cent level of significance respectively (Table 23 and 24).

Table.13Multiple linear regressions of bunch weight on biometric characters for each month in Nendran

MONTH	Multiple linear regression		Model fit criteria		
	No.	Best model selection	Adj. R^2	R^2	Cp
1ST	3	$Y = 9.41-1.04SW - 0.11SL+0.0002SV$ (1.33) (0.55) (0.04) (0.007)	0.119	0.145	5.00
2nd	4	$Y= 3.14 - 0.76SW- 0.03SL +0.10SG +0.06G_2$ (2.32) (0.62) (0.03) (0.05) (0.22)	0.119	0.141	3.11
3rd	5	$Y= 1.15 +0.003 SV- 0.06G_2 - 0.16L_3+ 0.06H_3 +0.12 G_3$ (1.13) (0.03) (0.03) (0.11) (0.01)** (0.04)**	0.582	0.625	5.55
4th	4	$Y=-0.47+0.03 H_3-0.08L_4+0.01H_4+0.07G_4$ (0.96) (0.009)** (0.064) (0.007)** (0.02)**	0.740	0.775	0.08
	5	$Y=-0.39-0.02+0.03H_3-0.07L_4+0.02H_4+0.07G_4$ (0.96) (0.02) (0.009)** (0.06) (0.02)**(0.025)**	0.746	0.779	1.36
5th	2	$Y= -1.37+0.025H_4+0.10G_5$ (0.64) (0.005)** (0.01)**	0.778	**0.805**	2.21
	4	$Y= - 1.6+0.01H_3+0.01H_4+0.03G_4+0.07G_5$ (0.64) (0.009) (0.006)** (0.022) (0.019)**	0.809	0.830	3.80
6th	5	$Y= -2.71+0.015H_3+0.016H_4+0.03G_4+0.06G_5+0.006H_6$ (1.01) (0.009) (0.006)** (0.022) (0.022)** (0.005)	0.798	0.838	-3.87
7th	2	$Y= -5.35 + 0.08 G_4+0.03 H_7$ (0.97) (0.015)** (0.004)**	0.825	0.833	3.48
	4	$Y= -3.95 -0.07 L_4 + 0.07 G_4 + 0.04 G_5 + 0.02 H_7$ (0.99) (0.04) (0.016)** (0.017) (0.004)**	0.860	0.872	-2.53
8th	4	$Y= -6.08+0.06G_4+0.05G_5+0.02H_7+0.02G_8$ (0.81) (0.15)** (0.018) (0.016)** (0.004)	0.862	0.875	4.22
9th	5	$Y= -6.05+0.006G_4+0.05G_5+0.02H_7+0.03G_8-0.02G_9$ (0.88) (0.15)** (0.018)** (0.016)** (0.004) (0.05)	0.878	0.892	4.37
Harvest stage	4	$Y = - 5.90 +0.14NF+ 0.17WSH+ 2.61FW+ 0.41FG$ (2.26) (0.031)** (0.39) (1.32) (0.180)**	0.514	0.556	4.30

* Significant at 5%, ** significant at 1% level, ( ) values in parenthesis indicates standard errors for respective estimates

Table.14 Multiple linear regressions of bunch weight with biometric characters at each month of Njalipoovan

MONTH	Multiple linear regression		Model fit criteria		
	No.	Best model selection	Adj. R^2	R^2	Cp
1ST	2	$Y= 7.09+0.64SW-0.05SG$ (1.24) (0.34) (0.039)	0.120	0.162	2.53
2nd	3	$Y = 6.9-0.05L_2 +0.02H_2 +0.32SW$ (1.2) (0.39) (0.03) (0.01)	0.125	0.178	2.75
3rd	5	$Y = 3.64+0.01H_2 +0.11L_3 +0.06H_3 -0.15G_3 +0.10G_2$ (0.65) (0.010) (0.07) (0.09)** (0.05)** (0.05)	0.566	0.610	5.55
4th	3	$Y=2.78-0.0408G_3 +0.078L_4 +0.045H_4$ (0.30) (0.015)** (0.039)** (0.04)**	0.795	**0.817**	-3.48
	6	$Y= 3.09+0.03G_2 -0.09G_3 +0.08L_4 +0.04H_4 +0.04G_4$ (0.39) (0.03) (0.04)** (0.03)** (0.004)** (0.03)	0.809	0.838	0.21
5th	5	$Y=2.98- 0.07G_3 +0.08L_4 +0.04H_4 +0.04G_4 -0.0001LA$ (0.37) (0.03)** (0.03)** (0.004)** (0.03) (0.0007)	0.805	0.834	-2.90
	6	$Y=3.32+0.03G_2 -0.09G_3 +0.08L_4 +0.04H_4 +0.04G_4 -0.02L_5$ (0.37) (0.03) (0.04)**(0.004)** (0.03)** (0.03) (0.05)	0.817	0.840	-0.10
6th	7	$Y= 2.80 + 0.03 G_2 - 0.09 G_3 +0.09 L_4 +0.03 H_4 + 0.047 G_4$ (0.46) (0.03) (0.043)** (0.039)** (0.006)** (0.037) $-0.0001 LA_4 +0.004 H_6$ (0.0007) (0.003)	0.819	0.843	-0.88
7th	11	$Y=3.58+0.041G_2 -0.11 G_3 +0.07 L_4 +0.03 H_4 +0.057G_4$ (0.39) (0.03) (0.043)** (0.039)** (0.006)** (0.037) $-0.0001 LA_4 -0.02L_5 -0.064L_6 +0.009 H_6 +-0.02 G_6 +0.011G_7$ (0.0001) (0.05) (0.02) (0.003) (0.02) (0.05)	0.816	0.857	2.01
8th	15	$Y=5.22-0.017SL+0.048G_2 +0.042L_3 +0.0117H_3 -0.12$ $G_3 +0.101L_4 +0.03H_4 +0.05G_4 -0.0001LA_4 -0.05L_5 -$ $0.059L_6 +0.010H_6 -0.037G_6 +0.0183G_7 -0.0040H_7$	0.824	0.878	8.54
Harvest Stage	4	$Y = -2.78 + 0.02 NF - 0.04 FL+ 0.58 FW + 0.33FG$ (0.389) (0.06) (0.09) (0.036)** (0.02)**	0.693	0.718	8.27

* Significant at 5%, ** significant at 1% level, ( ) values in parenthesis indicates standard errors for respective estimates

Table.15 Multiple linear regressions of bunch weight on biometric characters at each month of Redbanana

MONTH	Multiple linear regression		Model fit criteria		
	No.	Best model selection	Adj. R^2	R^2	Cp
1^{ST}	3	$Y= -3.53 + 0.18SL + 0.37SG - 0.0005\ SV$ (11.8) (0.16) (0.34) (0.0004)	-0.01	0.04	3.04
3^{rd}	3	$Y= 5.59 + 0.712L_3 + 0.079H_3 - 0.1320G_3$ (1.90) (0.52) (0.04) (0.15)	0.088	0.144	1.73
4^{th}	2	$Y=0.12+0.09H_4+0.027G_4$ (0.834) (0.024)** (0.0584)**	0.738	**0.748**	4.06
	5	$Y=0.64+0.54L_3-0.15G_3+0.07H_4+0.34G_4-0.0001LA$ (1.02) (0.27) (0.06)** (0.025)** (0.06)** (0.0001)	0.750	0.789	2.1
5^{th}	3	$Y=-0.22+0.053H_4+0.087H_4+0.035G_4$ (0.59) (0.05)** (0.087)** (0.05)**	0.853	0.859	1.92
	4	$Y=0.19-0.025H_3+0.04H_4+0.07H_5+0.06G_4$ (0.63) (0.012) (0.019)** (0.014)** (0.04)*	0.862	0.873	1.28
6^{th}	5	$Y=1.086 - 0.02\ H_3 + 0.05\ H_4 + 0.07\ H_5 + 0.08\ G_4 - 0.04\ G_6$ (3.5) (0.012) (0.01)** (0.014)** (0.06) (0.02)	0.856	0.880	9.74
7^{th}	9	$Y = -9.23 + 0.29\ SW + 0.14\ SG - 0.0001S\ V-$ (3.68) (0.15) (0.06) (0.0001) $0.021H_3+0.053H_4+0.069H_5+0.07G_4+0.07H_6-0.142G_7$ (0.012) (0.019)** (0.014)** (0.06)** (0.003) (0.308)	0.891	0.911	5.55
8^{th}	10	$Y=-10.46+0.30SW+0.13SG-0.00016SV-$ $0.021H_3+0.045H_4+0.073H_5+0.079G_4+0.191L_6+0.076H_6-0.014G_6$	0.892	0.914	6.30
Harvest Stage	4	$Y= -13.89+ 0.13NF+ 1.71WSH +12.40FW+ 0.37FG$ (2.15) (0.08)** (0.085)** (0.05) (0.01)	0.687	0.712	4.37

* Significant at 5%, ** significant at 1% level, ( ) values in parenthesis indicates standard errors for respective estimates

Table.16 Multiple linear regressions of bunch weight on biometric characters at each month of Robusta

MONTH	Multiple linear regression		Model fit criteria		
	No.	Best model selection	Adj. R^2	R^2	Cp
1st	3	$Y= 11.50 - 0.33\,SW + 0.08\,SL - 0.02\,SG$ (1.20) (0.43) (0.09) (0.06)	0.043	0.025	4
3rd	2	$Y=6.82 + 0.06\,H_3 + 0.27\,G_3$ (0.62) (0.02) (0.03)**	0.677	0.691	8.55
	4	$Y=6.98 + 0.01\,SL - 0.18\,L_3 + 0.07\,H_3 + 0.27\,G_3$ (1.06) (0.01) (0.15) (0.02)** (0.03)**	0.683	**0.711**	9.40
4th	4	$Y= 5.37 + 0.09\,SL - 0.14\,L_3 + 0.08\,H_4 + 0.25G_3$ (1.18) (0.03) (0.14) (0.022)** (0.03)**	0.721	**0.752**	1.63
5th	6	$Y=4.70 - 0.65\,SW + 0.11\,SL + 0.15H_4 + 0.08\,H_3 + 0.02\,G_3 + 0.01H_5$ (1.28) (0.23) (0.03) (0.014)** (0.022) (0.03)** (0.008)	0.726	0.762	5.31
	7	$Y=3.8 - 0.74\,SW + 0.12\,SL + 0.09\,H_3 + 0.25\,G_3 + 0.01\,H_5 + 0.03\,H_4 + 0.09\,G_5$ (1.32) (0.23) (0.03) (0.01) (0.03)** (0.015) (0.013)** (0.069)	0.747	0.792	4.62
6th	6	$Y= 5.46 - 0.19\,L_3 + 0.06\,H_3 + 0.28\,G_3 + 0.013\,H_4 + 0.137\,L_6 + 0.022\,H_6$ (1.85) (0.14) (0.014)** (0.03)** (0.013)** (0.012) (0.003)	0.746	0.780	9.56
7th	12	$Y=2.97-0.82SW+0.1094SL-0.181L_3+0.075H_3+0.282G_3$ $-0.024H_4+0.018H_5+0.156L_6+0.0173H_6-0.048G_6+0.029SG$	0.747	0.814	6.57
8th	12	$Y=3.28-0.78SW+0.0984SL-0.193+0.081H_3+0.266G_3-0.016H_4$ $+0.024H_5+0.195L_6+0.053H_8-0.062G_6+0.038SG-0.045H_7$	0.785	0.842	5.12
Harvest Stage	4	$Y = -15.31 + 0.16NF + 0.30WSH + 90.09FW - 0.07FG$ (0.97) (0.005)** (0.30) (0.61) (5.9)**	0.914	0.967	4.16

* Significant at 5%, ** significant at 1% level, ( ) values in parenthesis indicates standard errors for respective estimates

Table.17 ANOVA for significance of stepwise regression for nendran at 5th month

Source of variation	df	Sum of square	Mean sum of square	F- value	p- value
Regression	4	57.6	14.4	49.08	<0.0001
Residuals	40	11.7	0.29		
Total	44	69.3			

Table.18 Summary of stepwise regression for nendran (5th month)

Variable entered	Variables in model	Model R^2	Standard error	p-value
G_5	1	0.68**	4.60	<0.0001
H_4	2	0.80**	2.10	<0.0001
H_3	3	0.81	0.95	0.1060
G_4	4	0.83	0.90	0.0873

Table.19 ANOVA for significance of stepwise regression for Njalipoovan at 4th month

Source of variation	df	Sum of square	Mean sum of square	F- value	p- value
Regression	4	26.81	6.70	54.61	<0.0001
Residuals	40	5.52	0.12		
Total	44	32.33			

Table.20 Summary of stepwise regression for Njalipoovan (4th month)

Variable entered	Variables in model	Model R^2	Standard error	p-value
H_4	1	0.780**	0.004	<0.0001
G_3	2	0.801**	0.015	0.028
L_4	3	0.817	0.038	0.0524
LA	4	0.829	0.000	0.0861

Table.21 ANOVA for significance of stepwise regression for redbanana at 4th month

Source of variation	df	Sum of square	Mean sum of square	F- value	p- value
Regression	5	306.32	61.26	33.06	<0.0001
Residuals	40	81.53	1.85		
Total	44	387.86			

Table.22 Summary of stepwise regression for redbanana cultivar (4th month)

Variable entered	Variables in model	Model R^2	Standard error	p-value
G_4	1	0.661**	0.058	<0.0001
H_4	2	0.748**	0.024	0.0002
G_3	3	0.762	0.06	0.1029
L_3	4	0.777	0.27	0.0835
LA	5	0.789	0.000	0.1207

Tabl.23 ANOVA for significance of stepwise regression for robusta at 4th month

Source ofvariation	df	Sum of square	Mean sum of square	F- value	p- value
Regression	4	86.70	21.67	25.31	<0.0001
Residuals	41	35.11	0.85		
Total	45	121.81			

Table.24 Summary of stepwise regression for robusta cultivar (4th month)

Variable entered	Variables in model	Model R^2	Standard error	p-value
G_3	1	0.687**	0.03	<0.0001
H_4	2	0.752*	0.022	0.035
L_3	3	0.765*	0.15	0.024
SL	4	0.772	0.01	0.0968

4.3.5. Detection of multicollinerity

Detection of multicollinerity among independent variables were done for each cultivar at every month. The calculated value of variance inflation factor (VIF) for the growth and yield characters were below the specified level (VIF<10), which indicates that no such severe multicollinerity was present among the regressors of cultivars Nendran, Njalipoovan, Redbanana and Robusta.

4.3.6. Prediction of bunch weight

Predicted bunch weights were calculated using selected best regression models, in which nendran cultivar had less percentage error when compared to observed yield. Njalipoovan had more PED (21.7 per cent) among four cultivars.

Table.29 Observed and predicted bunch weight for all cultivars

Cultivars	Nendran	Njalipoovan	Redbanana	Robusta
Observed yield(kg)	7.1	7.5	10.2	18.6
Predicted yield(kg)	6.78	6.16	8.58	16.40
PED (%)	4.7	21.7	18.8	13.41

Table .25 Variance inflation factor to biometrical parameters for nendran

Parameters	SW	SL	SG	LA	NF	WSH	FW	FL	FG
VIF	2.38	1.003	2.38	1.772	1.311	1.821	1.939	1.124	1.323
variance inflation factor of growth parameters									
Months	2nd	3rd	4th	5th	6th	7th	8th	9th	
No. of leaves	1.119	1.257	2.220	1.045	1.098	1.035	1.107	1.004	
Plant height	1.634	2.499	8.586	3.456	2.292	3.954	1.296	2.575	
Plant girth	1.491	2.817	6.943	3.455	2.376	3.938	1.264	2.574	

Table .26 Variance inflation factor to biometrical parameters for Njalipoovan

Parameters	SW	SL	SG	LA	NF	WSH	FW	FL	FG
VIF	3.466	1.247	3.413	1.245	1.336	1.97	1.090	1.003	1.247
variance inflation factor of growth parameters									
Months	2nd	3rd	4th	5th	6th	7th	8th		
No. of leaves	1.26	1.023	1.903	1.088	1.318	1.029	1.482		
Plant height	1.850	1.625	1.971	1.715	2.089	1.247	2.020		
Plant girth	1.87	1.599	1.428	1.731	2.098	1.223	2.681		

Table .27 Variance inflation factor to biometrical parameters for Redbanana

<table>
<tr><td>Parameters</td><td>SW</td><td>SL</td><td>SG</td><td>LA</td><td>NF</td><td>WSH</td><td>FW</td><td>FL</td><td>FG</td></tr>
<tr><td>VIF</td><td>3.33</td><td>1.38</td><td>2.91</td><td>1.055</td><td>1.427</td><td>2.965</td><td>1.264</td><td>3.238</td><td>1.277</td></tr>
<tr><td colspan="10">variance inflation factor of growth parameters</td></tr>
</table>

<table>
<tr><td>Months</td><td>3rd</td><td>4th</td><td>5th</td><td>6th</td><td>7th</td><td>8th</td></tr>
<tr><td>No. of leaves</td><td>1.639</td><td>1.059</td><td>1.354</td><td>1.054</td><td>1.003</td><td>1.058</td></tr>
<tr><td>Plant height</td><td>2.070</td><td>2.501</td><td>3.509</td><td>4.957</td><td>4.769</td><td>1.315</td></tr>
<tr><td>Plant girth</td><td>2.761</td><td>2.586</td><td>3.950</td><td>5.046</td><td>4.762</td><td>1.271</td></tr>
</table>

Table .28 Variance inflation factor to biometrical parameters for Robusta

<table>
<tr><td>Parameters</td><td>SW</td><td>SL</td><td>SG</td><td>LA</td><td>NF</td><td>WSH</td><td>FW</td><td>FL</td><td>FG</td></tr>
<tr><td>VIF</td><td>4.24</td><td>7.81</td><td>7.78</td><td>1.02</td><td>1.03</td><td>2.26</td><td>1.056</td><td>2.4</td><td>1.58</td></tr>
<tr><td colspan="10">variance inflation factor of growth parameters</td></tr>
</table>

<table>
<tr><td>Months</td><td>3rd</td><td>4th</td><td>5th</td><td>6th</td><td>7th</td><td>8th</td></tr>
<tr><td>No. of leaves</td><td>1.33</td><td>1.18</td><td>1.29</td><td>1.047</td><td>1.047</td><td>1.199</td></tr>
<tr><td>Plant height</td><td>3.62</td><td>3.39</td><td>4.11</td><td>1.712</td><td>1.043</td><td>1.279</td></tr>
<tr><td>Plant girth</td><td>3.81</td><td>3.60</td><td>4.14</td><td>1.719</td><td>1.046</td><td>1.124</td></tr>
</table>

4.4. Principal component analysis:

Principal component analysis has been applied in this study as a statistical approach to identify major variance components, their contributions and reduction of dimensionality. This method assists in reducing the number of characters in the data collection and also helps in reducing the cost and time during field experiment. In this method, fewer variables explaining variations among individuals can be screened among various characters

The multiple linear regressions fitted in section 4.3 assume the independency among those explanatory variables. However this is not true as most of the biometrical characters at different stages of the plant growth are highly correlated this may leads to biased estimates. In order to overcome this multicollinearity, principal component analysis has been used. The details of this technique are presented in chapter 3. The principal component analysis adopted for all the four cultivars are given below.

4.4.1. Nendran cultivar

For nendran cultivar eight biometrical characters at different growth stages are treated as independent variables and subjected to principal component analysis. Correlations among the biometric characters are presented in the Table 30. It Results that biometrical characters were highly correlated themselves. The results of principal component analysis are presented in Table.34. It was observed that the first principal component alone has extracted about 66 per cent of the total variability and first three principal components extracted about 83 per cent. Thus the first linear function itself is enough so that a simple linear regression of the bunch weight with linear function of the variable can indicate the contributing biometrical characters to the yield.

Highest contributions of variables in first principal component are H_4 (0.37), G_4 (0.36), H_5 (0.35), G_5 (0.37) and NF (0.36). However all eight components equally contributing to the total yield. When we compare the first two components it can be seen that the contribution of the variables of second component is mainly characterized by D_4 (0.43), H_4 (0.33) and FW (0.36). Scree plot for eigen values are depicted in Fig 3. which shows that only first principal component had eigen value more than one, indicates we can get maximum amount of information on yield from independent variables by using PC1 itself (Kaiser , 1960).

4.4.2. Njalipoovan cultivar

Eight biometrical characters at different stage are treated as independent variables and subjected to principal component analysis for Njalipoovan cultivar. Correlations among the biometric characters are presented in the Table 31. It Results that biometrical characters were highly correlated themselves. The results of principal component analysis are presented in table.35. It was found that the first principal component alone has extracted about 55 per cent of the total variability and first four principal components extracted about 85 per cent. Thus the first linear function itself is enough so that a simple linear regression of the bunch weight with linear function of the variable can indicate the contributing biometrical characters to the yield.

Highest contributions of variables in first principal component are H_4 (0.44), H_5 (0.45), H_6 (0.44) and NF (0.31). When we compare the first two components it can be seen that the contribution of the variables of second component is mainly characterized by LA (0.90), H_4 (0.19) and WSH (0.25). Scree plot for eigen values are depicted in Fig 4. Which shows that first two principal component had eigen value more than one; indicates we can get maximum amount of information on yield from independent variables by using PC1 and PC2.

4.4.3. Redbanana cultivar

Eight biometrical characters at different stage are treated as independent variables and subjected to principal component analysis for Redbanana cultivar. Correlations among the biometric characters are presented in the Table 32. It Results that biometrical characters were highly correlated themselves. The results of principal component analysis are presented in Table.36. It was found that the first principal component alone has extracted about 57 per cent of the total variability and first three principal components extracted about 81 per cent. Thus the first linear function itself is enough so that a simple linear regression of the bunch weight with linear function of the variable can indicate the contributing biometrical characters to the yield.

Highest contributions of variables in first principal component are G_5 (0.41), H_5 (0.38), H_4 (0.38) and FW (0.35). When we compare the first two components it can be seen that the contribution of the variables of second component is mainly characterized by NF (0.38) and WSH (0.54). Scree plot for eigen values are depicted in Fig 5. Which shows that first two principal component had eigen value more than one; indicates we can get maximum amount of information on yield from independent variables by using PC1 and PC2.

4.4.4. Robusta cultivar

Five biometrical characters at different stage are treated as independent variables and subjected to principal component analysis for Robusta cultivar. Correlations among the biometric characters are presented in the Table 33. It Results that biometrical characters were highly correlated themselves. The results of principal component analysis are presented in Table.37. It was found that the first prlincipal component alone has extracted about 59 per cent of the total variability and first two principal components extracted about 80 per cent. Thus the first linear function itself is enough so that a simple linear regression of the bunch weight with linear function of the variable can indicate the contributing biometrical characters to the yield.

Highest contributions of variables in first principal component are H_4 (0.55), G_3 (0.47) and G_4 (0.49). When we compare the first two components it can be seen that the contribution of the variables of second component is mainly characterized by NF (0.96) and H_5 (0.16). Scree plot for eigen values are depicted in Fig 6. Which shows that first two principal component had eigen value more than one; indicates we can get maximum amount of information on yield from independent variables by using PC1 and PC2.

Table .30 Correlation matrix for biometric characters for Nendran

Correlation coefficients	H_4	G_4	H_5	G_5	H_6	NF	WSD	FW
H_4	1							
G_4	0.776**	1						
H_5	0.702**	0.621**	1					
G_5	0.630**	0.656**	0.675**	1				
H_6	0.485**	0.489**	0.672**	0.708**	1			
NF	0.706**	0.753**	0.664**	0.658**	0.564**	1		
WSH	0.676**	0.547**	0.590**	0.726**	0.705**	0.615**	1	
FW	0.571**	0.563**	0.399*	0.514**	0.455**	0.582**	0.543**	1

Table.31 Correlation matrix for Biometrical variables of Njalipoovan

Correlation coefficients	H_4	G_4	H_5	G_5	H_6	NF	WSH	FW
H_4	1							
LA	0.174	1						
H_5	0.840**	-0.006	1					
H_6	0.702**	0.010	0.830**	1				
H_7	0.669**	-0.066	0.655**	0.787**	1			
NF	0.474**	0.032	0.493**	0.539**	0.396**	1		
WSH	0.673**	0.042	0.529**	0.410**	0.238	0.289	1	
FW	0.315	0.024	0.328*	0.3	0.355*	0.213	0.259	1

*significant at 5% level; **significant at 1% level

Table.32 Correlation matrix for biometric characters of Redbanana

Correlation coefficients	H_4	G_4	H_5	G_5	H_6	NF	WSH	FW
H_4	1							
G_4	0.769**	1						
H_5	0.479**	0.585**	1					
G_5	0.592**	0.714**	0.920**	1				
H_6	0.320*	0.453**	0.676**	0.652**	1			
NF	0.402**	0.272	0.336*	0.380**	0.404**	1		
WSH	0.311	0.502**	0.365*	0.457**	0.364**	0.426**	1	
FW	0.342*	0.544**	0.480**	0.540**	0.449**	0.503**	0.785**	1

*significant at 5% level; **significant at 1% level

Table.33 Correlation matrix for Biometrical variables of Robusta

Correlation coefficients	H_4	G_4	H_5	G_5	H_6
H_4	1				
G_4	0.774**	1			
H_5	0.746**	0.555**	1		
G_3	0.746**	0.613**	0.492**	1	
NF	-0.061	0.054	0.088	-0.103	1

Table.34 Principal Component scores and their proportion of variances of Nendran

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
H_4	0.371	0.336	-0.158	0.441	0.378	-0.038	0.120	-0.608
G_4	0.361	0.430	-0.206	0.121	-0.271	-0.271	0.574	0.386
H_5	0.356	-0.187	-0.421	0.259	0.616	-0.052	-0.270	0.369
G_5	0.372	-0.252	0.033	-0.007	-0.384	-0.671	-0.401	-0.191
H_6	0.337	-0.592	0.130	0.343	0.006	0.197	0.510	-0.316
NF	0.360	-0.279	0.275	-0.683	-0.057	0.227	0.010	0.440
WSH	0.365	0.199	-0.272	0.169	-0.449	0.616	-0.361	-0.108
FW	0.295	0.365	0.765	0.332	0.222	0.022	-0.171	0.046
Eigen value	5.31	0.746	0.626	0.353	0.339	0.279	0.214	0.125
% of variance	66.4	9.3	7.8	4.4	4.2	3.4	2.6	1.5
Cumulative variance	66.4	75.7	83.5	88.0	92.2	95.7	98.4	100

Table .35 Principal component scores and their proportion of variances of Njalipoovan

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
H_4	0.447	0.192	-0.078	-0.168	-0.150	0.121	-0.594	-0.578
LA	0.026	0.900	-0.029	0.390	-0.091	0.006	0.102	0.130
H_5	0.450	-0.043	-0.109	-0.070	-0.140	-0.663	-0.237	0.511
H_6	0.440	-0.129	-0.160	0.197	-0.142	-0.238	0.673	-0.440
H_7	0.393	-0.257	-0.016	0.353	-0.381	0.621	-0.026	0.372
NF	0.312	-0.026	-0.261	0.220	0.874	0.121	-0.068	0.062
WSH	0.311	0.258	0.090	-0.757	0.088	0.281	0.347	0.213
FW	0.230	-0.019	0.937	0.164	0.172	-0.090	-0.012	-0.057
Eigen value	4.098	1.065	0.840	0.818	0.652	0.261	0.208	0.053
% of variance	51.2	13.3	10.5	10.2	8.1	3.2	2.6	0.6
Cumulative variance	51.2	64.5	75.0	85.2	93.4	96.7	99.3	100

Table .36 Principal component scores and their proportion of variances of Redbanana

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
H_4	0.328	-0.273	0.598	0.294	0.118	0.106	0.589	-0.013
G_4	0.385	-0.198	0.438	-0.195	0.261	-0.199	-0.664	0.184
H_5	0.385	-0.327	-0.332	-0.086	-0.451	0.050	0.110	0.637
G_5	0.411	-0.285	-0.149	-0.096	-0.370	0.104	-0.126	-0.739
H_6	0.336	-0.128	-0.551	0.090	0.736	0.020	0.113	-0.045
NF	0.271	0.382	-0.061	0.814	-0.158	0.012	-0.289	0.044
WSH	0.324	0.549	0.097	-0.355	0.046	0.672	0.024	0.064
FW	0.351	0.482	-0.003	-0.246	-0.082	-0.694	0.292	-0.058
Eigen value	4.563	1.109	0.860	0.705	0.355	0.205	0.140	0.059
% of variance	57.0	13.8	10.7	8.8	4.4	2.5	1.7	0.7
Cumulative variance	57.0	70.9	81.6	90.4	94.9	97.5	99.2	100

Table .37 Principal component scores and their proportion of variances of Robusta

Variables	PC1	PC2	PC3	PC4	PC5
H_4	0.552	-0.049	-0.06	-0.019	-0.829
G_4	0.496	0.076	0.317	-0.735	0.318
H_5	0.467	0.1693	-0.775	0.153	0.359
G_3	0.479	-0.171	0.505	0.640	0.274
NF	-0.007	0.966	0.197	0.144	-0.086
Eigen value	2.97	1.04	0.492	0.363	0.125
% of variance	59.5	20.8	9.8	7.2	2.5
Cumulative variance	59.5	80.3	90.2	97.5	100

Figure.3 Scree plot of eigen values for principal components in Nendran

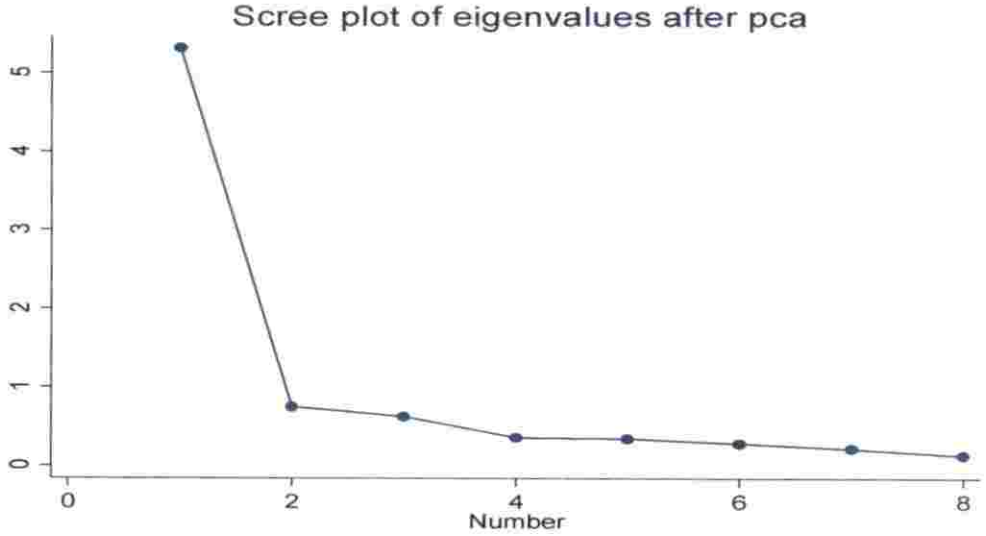


Figure.4 Scree plot of eigen values for principal components in Njalipoovan

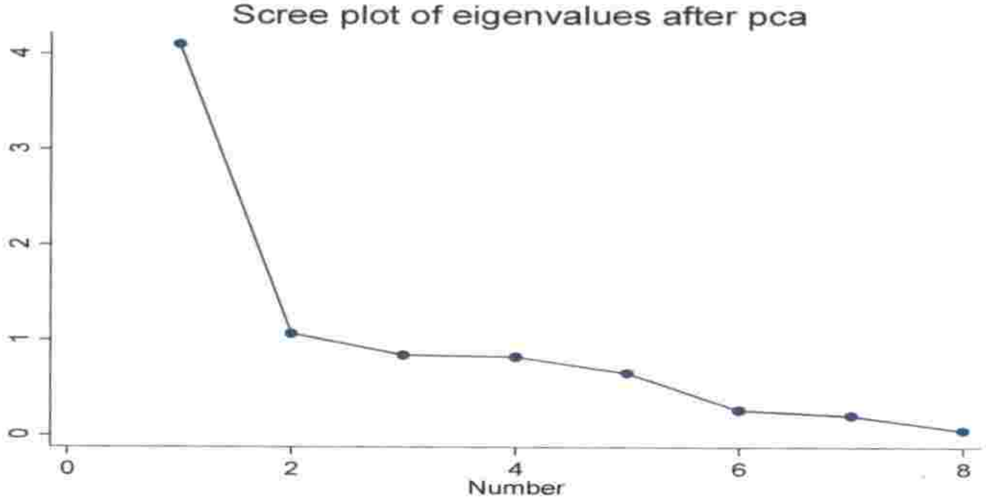


Figure.5 Scree plot of eigen values for principal components in Redbanana

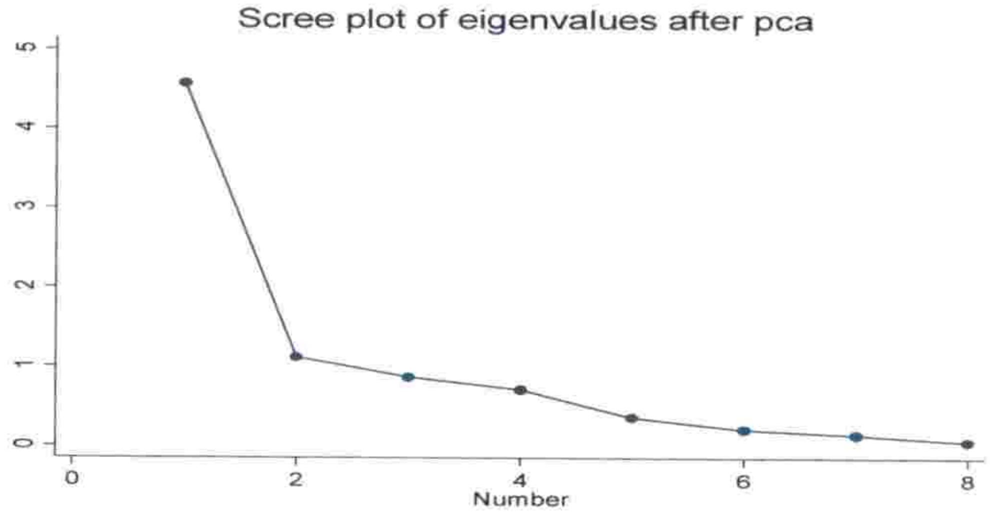
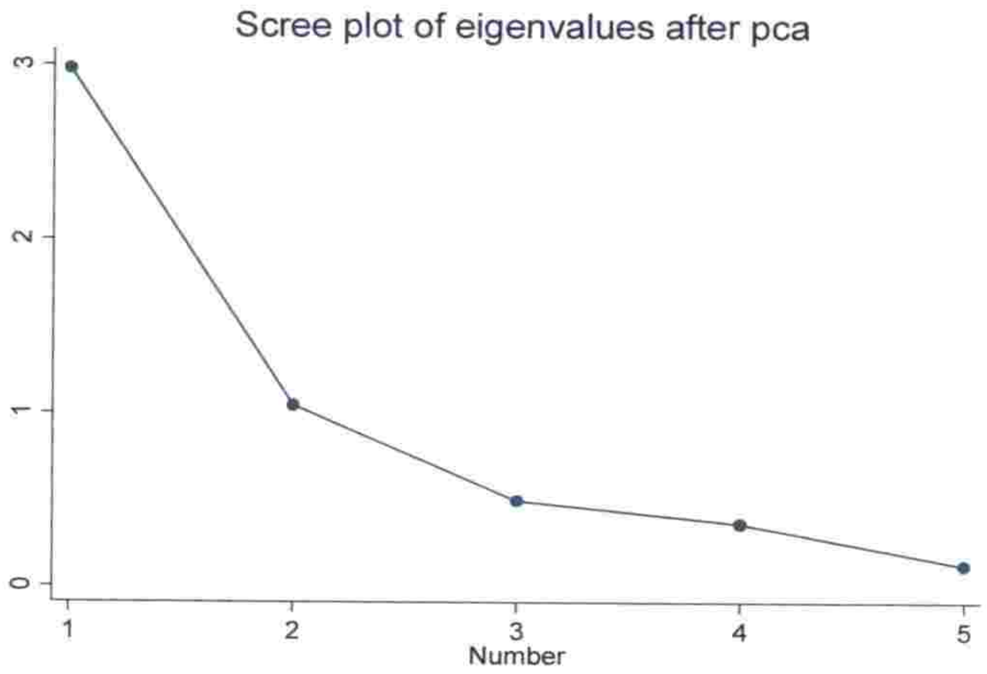


Figure. 6 Scree plot of eigenvalues for principal components in Robusta



4.5. TREND ANALYSIS OF BANANA PRODUCTION IN KERALA

Linear and nonlinear growth models are developed for the purpose of estimating the growth rates and fitting the best model. The use of R^2, criteria of randomness and normality of time series data were used as measure of goodness of fit for residuals.

The results of linear and nonlinear regression models were developed to achieve trend analysis of area, production, productivity, price and cost of cultivation data for the past two and half decade (1990-91 to 2014-2015) and the results are presented individually for area, production, productivity, price and cost of cultivation of banana for Kerala state.

Straight line, logarithmic, exponential, quadratic, cubic, compound , power ,S-curve and Exponential growth models were considered for studying area, production, productivity, price and cost of cultivation of banana crop in Kerala. They were also used for developing an appropriate growth model which would be able to describe the area, production, productivity, price and cost of cultivation of banana and also to forecast the area, production and productivity. Linear and non-linear polynomial growth models were fitted by using IBM.SPSS.20 Software.

4.5.1. Trends in area

The data presented in Table.38. Area under the banana crop revealed that among the linear and nonlinear growth models fitted to the area under the banana crop , the maximum R^2 of 84 per cent was observed in case of cubic model in comparison to that of other models. The Shapiro wilk test (test for normality) was found to be non –significant indicating that the residuals due to this model were found to be normally distributed and also the run test (test for randomness) value was non-significant suggests that residuals were not correlated, randomly in nature. Seven runs were observed in run test for residuals of area under banana. Next to the cubic model, quadratic model had the maximum R^2, but not satisfying both randomness and normality criteria. Hence, cubic model was selected to fit the trends for area under banana crop in Kerala.

For cubic model, $Y = 35214.6 - 2611.2 \times t + 427.9 \times t^2 - 11.51 \times t^3$ (R^2=84 %)

The graph of the actual and fitted trends for the area under banana cultivation using quadratic and cubic models are depicted in the Fig.7

4.5.2. Trends in production

In case of the production of banana crop, the quadratic model had maximum R^2 of 71 per cent as compared to other models. Cubic had R^2 of 67 per cent, nearer to the quadratic model. Among the fitted growth models, these two models explained maximum variation in dependent variable. Moreover, the run tests as well as the Shapiro-wilk test values were non –significant in both models indicating that the residuals were independently normally distributed. All the estimated v values were in 5 per cent significant level, indicating that the characters values were significant (Table .39).

For Quadratic model, $Y = 338364.9 - 370.6 \times t + 374.3 \times t^2$ (R^2=71%)

For Cubic model, $Y = 396906.7 - 3088.9 \times t + 316.5 \times t^2 + 1.47 \times t^3$ (R^2=67 %)

The graph of the actual and fitted trend for the production of banana crop using the quadratic and cubic models are depicted in the Fig.8

4.5.3. Trends in productivity

Characteristics of fitted linear and non-liner models for productivity of banana in Kerala presented in Table 40 reveals that, cubic and quadratic models were the best models which had maximum variation in model with R^2 of 79 per cent and 70 per cent respectively, while lowest variation was noticed for that of sigmoid –curve with 28 per cent variation. Almost all fitted models satisfying the assumption of normality. Moreover, cubic function and quadratic models had met the both assumption of residuals like normality and randomness. Cubic function as well as quadratic function are selected as the best model to fit the trends of productivity and also used for forecasting the productivity for future period. All estimated values of characters in the models were found to be within the 95 % confidence interval indicating that the characters were significant at 5% level of significance. Shapiro-wilks test and run test values were found to be non- significant indicating that the residuals fulfilled model selection criteria. The selected models are given below.

The quadratic model, $Y = 15896.1 - 764.5 \times t + 18.5 \times t^2$ (R^2=70%)

The cubic model, $Y = 13200.2 + 370.02 \times t - 88.4 \times t^2 + 2.74 \times t^3$ (R^2=79 %)

The graph of the actual and fitted trend for the productivity of banana crop using the quadratic and cubic models are depicted in the Fig.9

4.5.4. Trend analysis in whole sale price of banana

Whole sale price of banana are measured in Rs / quintal in Kerala. So to find trend and prediction model for price of banana, here also linear and non-linear models employed, Table 41 presents fitted linear and non-linear models for price. Cubic model with maximum R^2 of 91 per cent followed by quadratic function with 89 per cent were significant, but cubic model failed to fulfills the assumption of normality, Shapiro-wilks test was found to be significant indicating the residuals are not normally distributed. However, quadratic model fulfill the both assumptions like normality and randomness and values of test statistic were found to be non-significant.The selected model was given below.

For quadratic model, $Y= 1355.7 - 127.6\times t+ 8.3\times t^2$ (R^2=89%)

The graph of the actual and fitted trend for the price of banana crop using the quadratic and cubic models are depicted in the Fig.10

4.5.5. Trend analysis on cost of cultivation

Data on cost of cultivation of banana was obtained from Department of Economics and Statistics, Thiruvananthapuram for fifteen years (2000-2015). For this data liner and non-linear models were fitted to found trend on cost of cultivation. Table 42 revealed that all models are significant at 5 per cent level, however cubic model has maximum multiple R-square of 85 per cent. Meanwhile the lowest with s-curve and inverse function had 11 per cent and 13 per cent respectively. All models fulfill the assumption of randomness through run test, which results that run test values were non-significant. But some models failed in fulfilling the assumption of normality. Shapiro –Wilk's test statistics was significant in the case of Logarithmic, inverse and power functions suggest that the assumption of normality was not satisfied in these models. Hence here also cubic and quadratic models were found suitable for fit the trends of cost of cultivation for last twenty five years. The selected models were given below.

For quadratic model, $Y= 90485.09 - 3479.2\times t+ 471.5\times t^2$ (R^2=79%)

For cubic model, $Y= 112258.7 -17585.5\times t +2606.2\times t^2 -88.9\times t^3$ (R^2=85 %)

The graph of the actual and fitted trend for the cost of cultivation of banana crop using the quadratic and cubic models are depicted in the Fig.11

Table .38 Fitted linear and non-linear models for area under banana

	Models	Regression coefficient				Goodness of fit		
		b_0	b_1	b_2	b_3	R^2	Shapiro-wilk test (W)	Run test (Z)
1	Linear	26470.02** (2677.910)	1585.4** (180.136	-	-	0.771	0.9321 [0.097]	-3.87836 [0.0021]
2	Logarithmic	17779.4** (4978.3)	12628.8** (2023.8)	-	-	0.629	0.8856 [0.009]	-3.47011 [0.0005]
3	Inverse	52268.3** (2887.7)	-33988.7** (11394)	-	-	0.279	0.8749 [0.005]	-4.28 [0.0001]
4	Quadratic	23869.0** (4266.2)	2163.3 ** (756.1)	-22.23 (28.2)	-	0.77	0.9238^{NS} [0.062]	-3.878 [0.001]
5	Cubic	35214.6** (5250.1)	-2611.2 (1714.5)	427.9** (151.6)	-11.54** (3.8)	**0.844**	0.9570^{NS} [0.359]	-1.783NS [0.0987]
6	Compound	28194.6** (1908.1)	1.037** (0.005)	-	-	0.732	0.911 [0.033]	-3.061 [0.0021]
7	Power	23090.4 ** (2786.8)	0.288** (0.049)	-	-	0.600	0.8688 [0.004]	-3.47011 [0.0005]
8	S-Curve	10.834** (0.068)	-0.773 ** (0.269)	-	-	0.264	0.8537 [0.002]	-4.28 [0.0001]
9	Exponential	28194.6 ** (1908.1)	0.036 ** (0.005)	-	-	0.732	0.9118 [0.033]	-3.061 [0.002]

*significant at 5% level; **significant at 1% level; values in brackets ( ) indicates standard error; values in square brackets [ ] indicates p-values , NS- non significant

Table.39 Fitted linear and non-linear models for production under banana

	Models	Regression coefficient				Goodness of fit		
		b_0	b_1	b_2	b_3	R^2	Shapiro-wilk test (W)	Run test (Z)
1	Linear	354560.9** (15928.2)	6031.5** (1071.4)	-	-	0.579	0.9434NS [0.1781]	-3.4758 [0.0034]
2	Logarithmic	334758.9** (28532.4)	42330.3** (11598.9)	-	-	0.367	0.9721NS [0.7002]	-3.7415 [0.0016]
3	Inverse	448532.5 (13918.4)	-101946.6 (54919.2)	-	-	0.130	0.883 [0.025]	-3.8701 [0.0069]
4	Quadratic	338364.9 (22728.2)	-370.6 (4028.2)	374.3 (150.3)	-	**0.715**	0.9675^{NS} [0.5828]	-1.837^{NS} [0.061]
5	Cubic	396906.7 (33450.7)	-3088.9 (10924.2)	316.5 (966.2)	1.478 (24.4)	**0.672**	0.9680^{NS} [0.5729]	-1.745^{NS} [0.068]
6	Compound	359564.7 (13603.9)	1.014 (0.003)	-	-	0.555	0.9359NS [0.119]	-3.948 [0.066]
7	Power	343525.3 (22852.6)	0.096 (0.027)	-	-	0.354	0.9012 [0.006]	-2.984 [0.0007]
8	S-Curve	13.005 (0.032)	-0.231 (0.127)	-	-	0.125	0.8925 [0.0002]	-4.2381 [0.0086]
9	Exponential	359564.7 (13603.9)	0.014 (0.003)	-	-	0.555	0.9359NS [0.119]	-4.8568 [0.003]

*significant at 5% level; **significant at 1% level; values in brackets ( ) indicates standard error; values in square brackets [ ] indicates p-values, NS- non significant

Table .40 Fitted linear and non-linear models for productivity under banana

	Models	Regression coefficient				Goodness of fit		
		b_0	b_1	b_2	b_3	R^2	Shapiro-wilk test (W)	Run test (Z)
1	Linear	13728.3 (723.4)	-282.7 (48.6)	-	-	0.595	0.9565^{NS} [0.3493]	-3.87836 [0.00001]
2	Logarithmic	15723.7 (1086.5)	-2444.5 (441.7)	-	-	0.571	0.9593^{NS} [0.4021]	-3.47011 [0.0005]
3	Inverse	9003.45 (587.5)	6869.7 (231.2)	-	-	0.276	0.8383 [0.001]	-4.28 [0.0001]
4	Quadratic	15896.1 (1004.7)	-764.5 (178)	18.5 (6.4)	-	0.701	0.9595^{NS} [0.4121]	-1.878^{NS} [0.001]
5	Cubic	13200.2 (1231.7)	370.02 (402.2)	-88.4 (35.5)	2.74 (0.901)	**0.792**	0.9706^{NS} [0.6608]	-1.65^{NS} [0.072]
6	Compound	13731.1 (971.1)	0.974 (0.005)	-	-	0.595	0.9705^{NS} [0.6580]	-3.061 [0.0021]
7	Power	16675.4 (1737.5)	-0.232 (0.042)	-	-	0.567	0.9720^{NS} [0.6968]	-3.4701 [0.0005]
8	S-Curve	9.081 (0.056)	0.664 (0.220)	-	-	0.283	0.8699 [0.004]	-4.28 [0.0001]
9	Exponential	13731.1 (971.1)	-0.027 (0.005)	-	-	0.575	0.9705 [0.6580]	-3.061 [0.002]

*significant at 5% level; **significant at 1% level; values in brackets ( ) indicates standard error; values in square brackets [ ] indicates p-values , NS- non significant

Table .41 Fitted linear and non-linear models for whole sale price of banana

	Models	Regression coefficient				Goodness of fit		
		b_0	b_1	b_2	b_3	R^2	Shapiro-wilk test (W)	Run test (Z)
1	Linear	378.8 ** (200.5)	89.3** (13.4)	-	-	0.656	0.9585^{NS} [0.3868]	-3.47 [0.0005]
2	Logarithmic	142.5** (392.0)	602.6** (159.3)	-	-	0.383	0.9015 [0.001]	-4.28 [0.0005]
3	Inverse	1766.2 (192.6)	-1476.4 (760.0)	-	-	0.141	0.8183 [0.0004]	-4.28 [0.0001]
4	Quadratic	1355.7 ** (180.4)	-127.6 ** (31.9)	8.3 (1.19)	-	0.893	0.9312^{NS} [0.092]	-1.837^{NS} [0.06]
5	Cubic	959.6 ** (236.7)	38.9 (77.3)	-7.3* (6.8)	0.403 (0.173)	**0.915**	0.9069 [0.028]	-0.2014^{NS} [0.838]
6	Compound	714.7** (70.06)	1.052 (0.07)	-	-	0.722	0.9069 [0.028]	-2.061 [0.0071]
7	Power	604.9** (121.0)	0.358 (0.081)	-	-	0.456	0.8434 [0.001]	-4.28 [0.0001]
8	S-Curve	7.378** (0.102)	-0.937 (0.401)	-	-	0.192	0.7848 [0.0002]	-3.92 [0.0001]
9	Exponential	714.7** (70.06)	0.051 (0.007)	-	-	0.722	0.9083 [0.028]	-2.65 [0.007]

*significant at 5% level; **significant at 1% level; values in brackets ( ) indicates standard error; values in square brackets [ ] indicates p-values , NS- non significant

Table .42 Fitted linear and non-linear models for cost of cultivation of banana

	Models	Regression coefficient				Goodness of fit		
		b_0	b_1	b_2	b_3	R^2	Shapiro-wilk test (W)	Run test (Z)
1	Linear	69107.1**	4065.9** (7323.5)	- (805.4)	-	0.662	0.9293^{NS} [0.2664]	-0.8017^{NS} [0.4228]
2	Logarithmic	68426.3** (12422.6)	17854.2** (6188.0)	-	-	0.390	0.8822 [0.05]	-0.8017^{NS} [0.4228]
3	Inverse	108920.4** (7625.3)	-32936.2** (23491.0)	-	-	0.131	0.8956 [0.081]	-0.8017^{NS} [0.4228]
4	Quadratic	90485.09** (9771.9)	-3479.2** (2810.4)	471.5* (170.8)	-	0.793	0.9088^{NS} [0.12]	-0.2672^{NS} [0.7892]
5	Cubic	112258.7* (12874.4)	-17585.8* (6743.6)	2606.2* (963.2)	-88.9* (39.6)	**0.858**	0.9794^{NS} [0.96]	-0.8017^{NS} [0.4228]
6	Compound	72847.8** (5784.5)	1.042** (0.09)	-	-	0.628	0.9069^{NS} [0.126]	-0.8017^{NS} [0.4228]
7	Power	72662.5** (9056.1)	0.168** (0.062)	-	-	0.361	0.8573 [0.022]	-0.8017^{NS} [0.4228]
8	S-Curve	11.57** (0.078)	-0.299* (0.233)	-	-	0.113	0.9027^{NS} [0.1048]	-1.870^{NS} [0.061]
9	Exponential	72847.8** (5484.5)	0.039** (0.008)	-	-	0.628	0.9080^{NS} [0.1266]	-0.8017^{NS} [0.4228]

*significant at 5% level; **significant at 1% level; values in brackets ( ) indicates standard error; values in square brackets [ ] indicates p-values , NS- non significant

Figure 7: Actual and estimated trends in area under cultivation

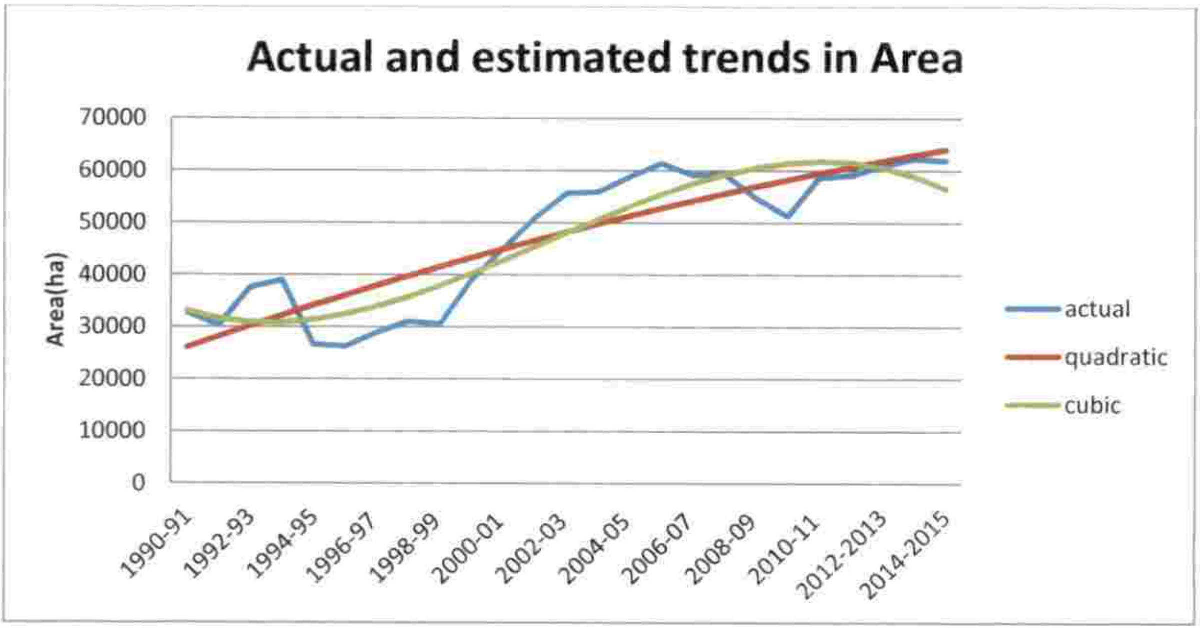


Figure 8: Actual and estimated trends in production

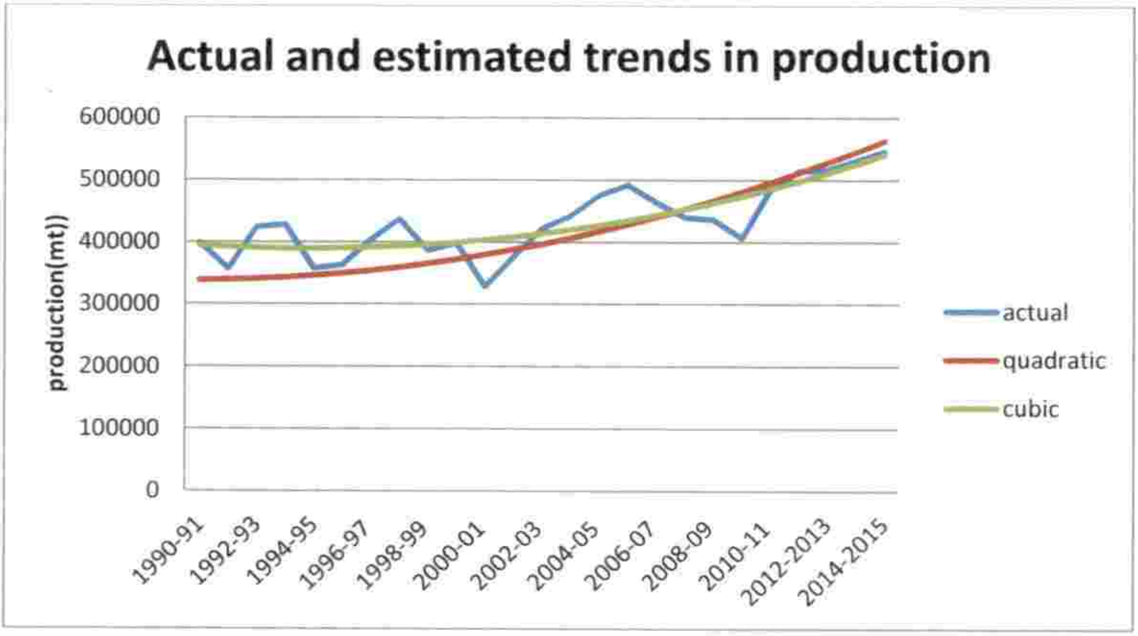


Figure 9: Actual and estimated trends in productivity

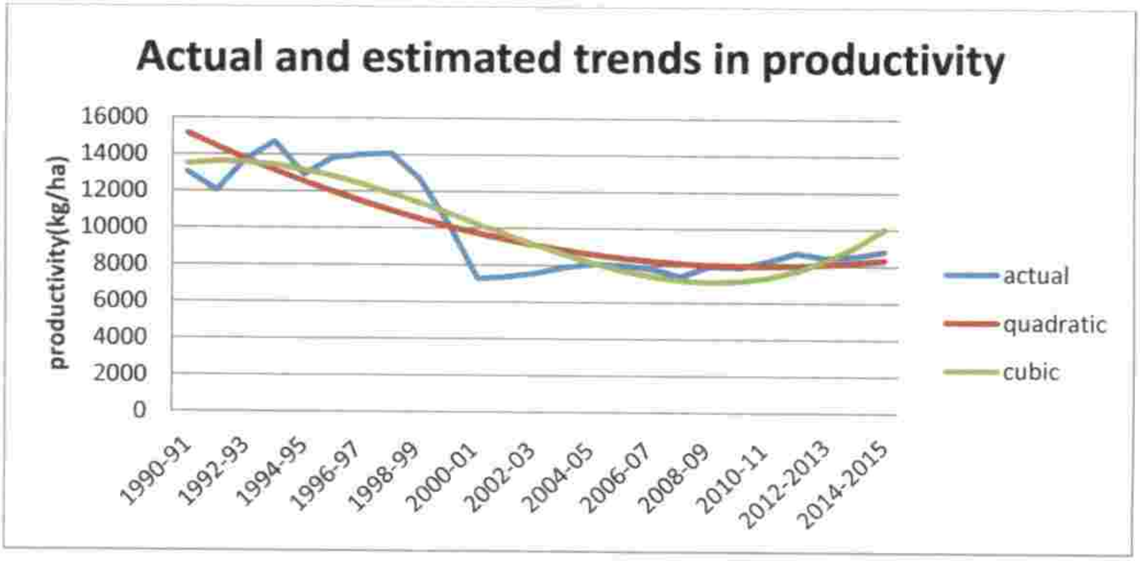


Figure 10: Actual and estimated trends in whole sale price of banana

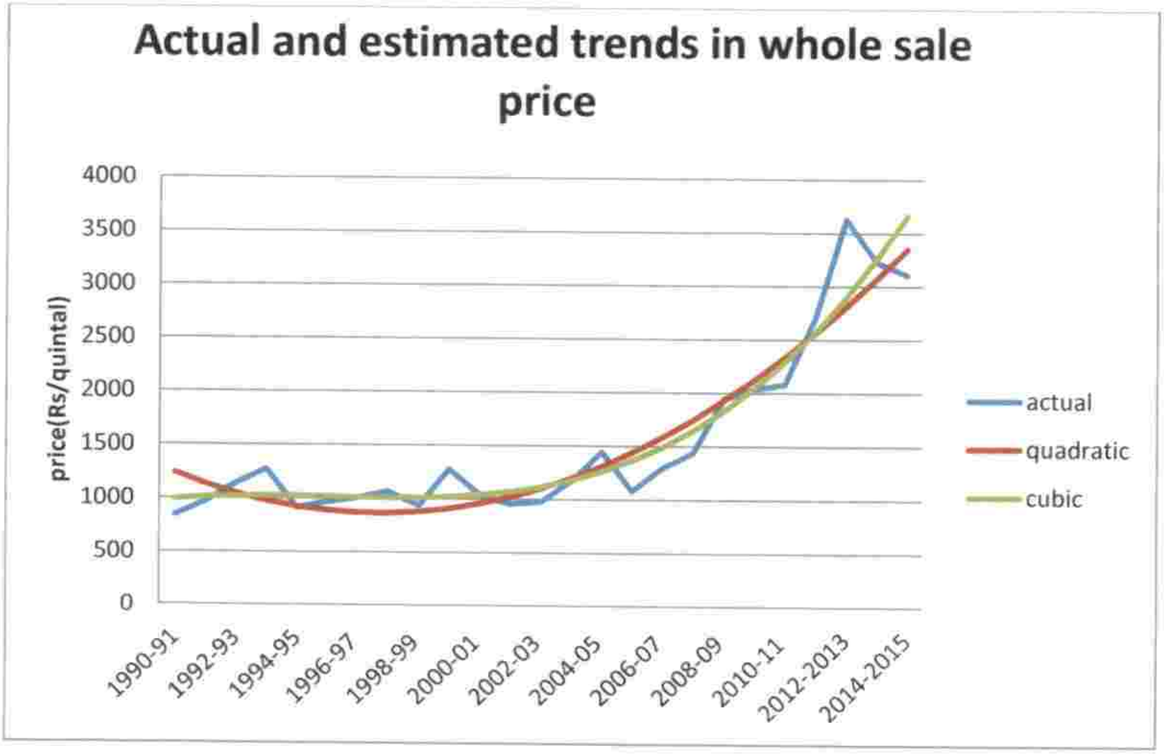


Figure 11: Actual and estimated trends in cost of cultivation

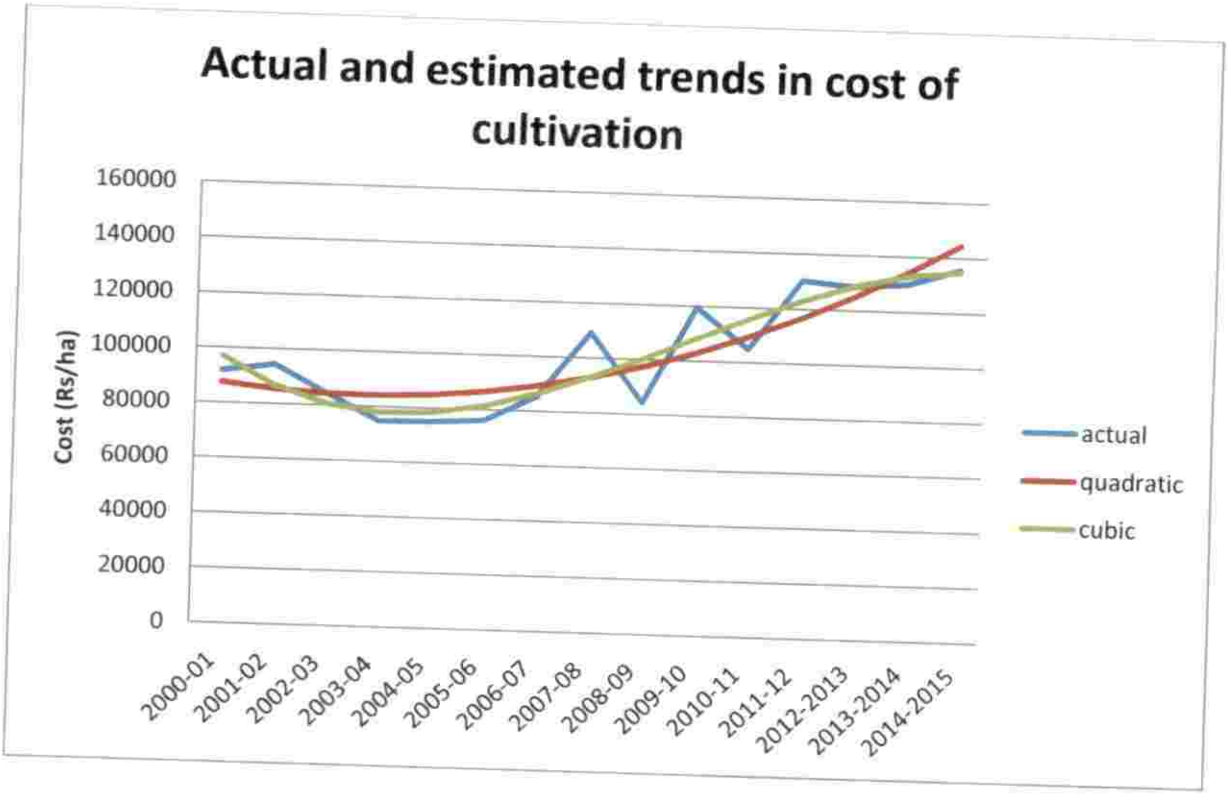


Table.43 Estimated values of area, production and yield of banana considering the actual recorded values of 25 years (1991-2015) using cubic model

	Actual			Estimated		
Year	Area (ha)	Production (T)	Yield (kg/ha)	Area (ha)	Production (T)	Yield (kg/ha)
1990-91	32687	398745	13034	33019	394135	13484
1991-92	30274	356891	12023	31611	392006	13608
1992-93	37569	423675	13727	30921	390528	13588
1993-94	38979	428153	14713	30878	389710	13440
1994-95	26589	357857	12912	31415	389560	13181
1995-96	26267	362917	13816	32461	390088	12828
1996-97	28855	403673	13990	33947	391302	12397
1997-98	31001	436717	14087	35805	393212	11903
1998-99	30521	386588	12666	37964	395826	11365
1999-00	39046	398145	10197	40357	399153	10798
2000-01	45059	327955	7278	42912	403202	10218
2001-02	50871	375903	7389	45561	404983	9643
2002-03	55668	421809	7577	48236	413503	9088
2003-04	55906	442220	7910	50866	419772	8570
2004-05	58866	475371	8075	53382	426798	8106
2005-06	61400	491823	8010	55715	434591	7712
2006-07	59143	463766	7841	57796	443160	7404
2007-08	59341	439803	7411	59555	452513	7199
2008-09	54739	435979	7965	60924	462659	7114
2009-10	51275	406242	7923	61832	473608	7164
2010-11	58671	483667	8244	62211	485368	4366
2011-12	59069	514054	8703	61992	497947	7738
2012-13	61015	515607	8451	61105	511356	8294
2013-14	62261	531299	8553	59481	525602	9052
2014-15	61936	545431	8806	57050	540695	10028

Table.44 Comparison of forecasted values (2015-2020) of area, production and productivity of banana using quadratic and cubic models

	Quadratic			Cubic model		
Year	Area (ha)	Production (T)	Yield (kg/ha)	Area (ha)	Production (T)	Yield (kg/ha)
2015-16	64057	581756	6712	54459	556526	9613
2016-17	65087	601223	5966	50296	573326	9897
2017-18	66072	621439	5220	45126	590998	10182
2018-19	67013	642403	4474	38880	609552	10467
2019-20	67909	664116	3728	31488	628995	10751

Table.45 Percentage Error Deviation of area, production and productivity for 2013-14 and 2014-15.

Year	Area	Production	Productivity
2013-14	4.46	1.07	5.83
2014-15	7.88	0.86	13.8

4.6. Compound annual growth rate (CAGR) of banana production in Kerala:

In order to analyze the changes in area, production, productivity, price and cost of cultivation of banana in Kerala during period of last twenty five years 1990-91 to 2014-2015) Compound annual growth rates were estimated. In the present study, the compound growth rates in area, production , productivity, price and cost of cultivation of banana in Kerala were estimated by fitting exponential type of equation($Y = \beta_0\beta_1^t$).

Table 46 shows that the compound annual growth rates were got positive in area, production, price and cost of cultivation and negative in productivity for the overall period (1991-92 to2014-15). During the overall period considered for analysis, it is revealed that area had been accelerating at the rate of 3.6 per annum with estimates are significant at 1 % level of significance. In 1995-96 area of banana was 26 thousand hectare, and then it was steadily increases to 61 thousand hectare in 2014-15. This is an indication of change in the cropping pattern in Kerala towards banana production.

Growth rate of production under banana was positive and significant (1.4 per annum) for last twenty-five years. Even though production is slowly increasing, but it fluctuates from year to year. Productivity had negative growth rate (-2.76 per annum) due decrease in production with respect to area under cultivation during the study period. Estimates of production and productivity are not significant at 5 % level of significance. Price had positive growth rates, price of banana was increasing in year to year in study year. In 1991-92, banana price was 842 Rs per quintal; there was rapid rise in the banana price 3107 Rs per quintal in 2014-2015. Similarly, in cost of cultivation of banana production also, here observed positive and increasing trend (growth rate of 3.9 per annum) in past study period.

Table .46 Compound annual growth rate (%) of banana in Kerala from 1991-92 to 2014-2015

characters	Regression equation: $\ln Y= \ln \beta_0+\beta_1 t+u_t$		Model-R^2	Compound annual growth rate(CAGR)
	β_0	β1		
Area	28194.6 (1908.1)	0.036 ** (0.005)	0.732	**3.6**
Production	359564.7 (13603.9)	0.014** (0.003)	0.555	**1.4**
Productivity	13731.1 (971.1)	-0.027** (0.005)	0.575	**-2.66**
Whole sale price	714.7 (70.06)	0.051** (0.007)	0.722	**5.1**
Cost of cultivation#	72847.8 (5484.5)	0.039** (0.008)	0.628	**3.9**

Figures in brackets are the standard errors

*significant at 5%, **significant at 1% level

CAGR calculated for 15 years (2001-2015)

4.7. Rainfall and rainy days adjusted growth rates to banana production:

Rainfall and rainy days adjusted growth rates were estimated in order to take into account the effect of rainfall and rain days on growth of area, production and productivity of banana crop in Kerala for last twenty-five years. Table 47 and 48 gives the regression results of the estimated models of area, production and productivity of banana using exponential function and calculated the adjusted growth rates. According to these two tables, the time variable is significant for all characters namely area, production and productivity. The estimated coefficients of rainfall and rainy days found to be non-significant but rainfall had positive effect on production, productivity and area, while number of rainy days had negative effect on banana production for available secondary data. For the whole period of study the rainfall and rainy days adjusted growth rates were worked out to be 3.59 per cent, 1.36 per cent and -2.6 per cent in area, production and productivity respectively. The results clearly indicate that there was no significant effect of rainfall and rain days on banana production during study period.

Table.47 Rainfall adjusted growth rate (%) of banana for Kerala from 1991-2015

characters	Regression equation: ln Yt= $\beta_0 + \beta_1 t + \beta_2 \ln R_t + u_t$			Model-R^2	Rainfall adjusted growth rate
	β_0	β1	β2		
Area	9.76 (3.96)	0.0368** (7.363)	0.0611 (0.195)	0.731	3.59
Production	11.76 (8.65)	0.0125** (4.84)	0.1304 (0.7571)	0.562	1.36
Productivity	9.06 (3.52)	-0.0298** (-5.26)	0.058 (0.179)	0.573	-2.57

Figures in brackets are the t-values

*significant at 5%, **significant at 1% level

Rt =Annual rainfall (mm) in **t**th year

Table.48 Rainy days adjusted growth rate (%) of banana for Kerala from 1991-2015

characters	Regression equation: $\ln Yt = \beta_0 + \beta_1 t + \beta_2 \ln R_t + u_t$			Model-R^2	Rainydays adjusted growth rate
	β_0	β1	β2		
Area	10.62 (8.44)	0.0312** (7.48)	-0.07 (-0.29)	0.736	3.52
Production	13.42 (19.41)	0.0135** (4.97)	-0.12 (-0.91)	0.578	1.358
Productivity	9.62 (7.30)	-0.0274** (-5.31)	-0.018 (-0.07)	0.572	-2.69

Figures in brackets are the t-values

*significant at 5%, **significant at 1% level

Rt =number of rainy days (mm) in $\mathbf{t}^{th}$ year

Discussion

5. DISCUSSION

The present chapter is on discussing of results obtained in the previous chapter. The discussion has been made under the following headings.

5.1 CORRELATION STUDIES

Correlation studies gave way to know the association prevailing between banana yield and biometric characters which give better understanding of the contribution of each biometric parameter to crop yield, with this view the crop yields and biometric characters were subjected to correlation analysis and the results are discussed.

An attempt on correlating the biometrical characters with bunch weight of nendran cultivar revealed significant results in case of plant height and plant girth at earlier stages of growth. The highest association to the yield was from the plant height (0.787) at 4th month and plant girth (0.832) at 5th month. It was observed that, rapid increase in plant growth in terms of plant height and plant girth in earlier growth stages. The results of the correlation analysis of growth characters and banana yield level gain support from the study conducted by Rao *et al.* (1993), who observed that prediction of yield was possible in early stages of the crop with the help of plant height and stem girth.

Yield characters like number of fruits (0.558) and fruit girth (0.578) at harvesting stage had positive significant correlations with yield. However, correlation coefficients between bunch weight and sucker characters were non-significant, was an indication of non significant relation between bunch weight and suckers characters. These findings are in concurrence with studies of (Sheela and Nair, 2001) concluding that tissue culture banana plants gave a more significant yield as compared to sucker banana plants.

In case of Njalipoovan cultivar, plant height was significantly correlated with yield at fourth and fifth month, in addition the number of leaves and leaf area of fourth month also had significant correlation with yield. These results are in linear agreement with studies of Rao *et al.* (1993), which remarked that morphological characters were found to be influencing yield mainly leaf characters –leaf length and leaf breadth

In redbanana and robusta, correlations of biometric characters with bunch weight were positive and significant. Plant height and plant girth had high association in fourth and fifth month of growth stage. Fruit characters namely number of fruits per bunch , weight of second hand and fruit weight were had significant correlations with bunch weight .Similar results was found by (Ram and Achal,2007). In their study, fruit characters, *viz.*, fingers bunch, finger diameter, hands bunch and finger length could predict yield up to 99.57%

5.2 MULTIPLE REGRESSION ANALYSIS

The influence of plant growth characters on banana yield established by correlation analysis indicated that there was a significant relationship between yield and biometrical characters. Models were built in order to predict yield with the help of minimum number of variables as possible in early stages of growth. Best prediction models were selected based on the significance of the model, R^2 (coefficient of determination), Adjusted R^2 and Cp , which explains the variation in dependent variable due to independent variables.

5.2.1. Nendran cultivar

The best yield prediction models estimated using biometrical characters independent variables and bunch weight as dependent variables and the best estimated model was obtained at fifth month is

$$Y = -1.37 + 0.025\ H_4 + 0.10\ G_5$$

R^2= 80.5 per cent

where, H_4= plant height at 4th month

G_5= plant girth at 5th month

It was observed that partial regression coefficient using data of previous months in the fifth month (H_4 and G_5) were significant. The percentage of variation explained in total yield by the above biometrical characters was 80.5 per cent, which was maximum compared to past months and also partial regression coefficients corresponding to the H_4 and G_5 were contributing for successive growth stages for the predicting the total yield.

5.2.2Njalipoovan cultivar

In the case of Njalioovan cultivar the partial regression coefficients corresponding to plant height and number of leaves at fourth month were statistically significant and highly contributing to total yield in consecutive growth stages from fourth month onwards. Among the different stages of crop growth, the multiple regression equation fitted to biometrical characters as explanatory variables and bunch weight as dependent variables, 4th month was the best time for prediction of bunch weight. The fitted equation is given as

$$Y = 2.78 - 0.0408G_3 + 0.078L_4 + 0.045H_4$$

R^2 = 81.7 per cent

where, H_4= plant height at 4th month

L_4 = number of leaves at 4th month

G_3 = plant girth at 3rd month

The coefficient of determination in yield explained by these three variables was 81.7 per cent. That was observed, the strong relationship between growth characters of fourth month and bunch weight in correlation analysis.

5.2.3. Redbanana cultivar

The results revealed that fourth month onwards the partial regression coefficients corresponding to plant height and plant girth was highly significant. Among the different fitted prediction models by using biometrical characters as independent and bunch weight as dependent variable fourth month was the best stage for early yield prediction. The best prediction model is given as

$$Y=0.12+0.09\,H_4+0.02\,G_4$$

R^2 = 74.8 per cent

Where, H_4= plant height at 4th month,

G_4 = plant girth at 4th month.

H_4 and G_4 were the most influential biometrical characters on total yield during early stage of the plant growth and these variables accounted for 74 per cent of explained variation. It was again observed that these variables had pre-dominant role in successive stages even other characters are included in the model. So we can conclude that forecast based on biometrical characters is possible about five to six months before the harvest with higher precision.

5.2.4. Robusta cultivar

Results of prediction models for robusta showing that, pre-harvest prediction of banana yield is possible in fourth month itself with help of plant height and plant girth as explanatory variables with high precision. Plant height at fourth month and plant girth of third month were chosen as best predictor variables for early prediction. From fourth month onwards these two variables made significant effect on bunch weight along with other biometrical characters. The best fitted model was obtained at fourth month and it is given as follows

$$Y=5.37+0.096SL-0.141L_3+0.088H_4+0.259G_3$$

$R^2 = 75.2$ per cent

Where,

H_4= plant height at 4th month,
G_3 = plant girth at 3rd month
SL = sucker length
L_3 = number of leaves at 3rd month

Coefficient of determination (R^2) of 75 per cent was explained by fourth month by these explanatory variables to total yield.

Yield characteristics like number of fruits, weight of second hand and fruit weight were combined influence on total yield and contributing to estimating the total yield. Fruit length has less influence on the total yield in all four cultivars.

Results of multiple regression analysis were in conformity with the similar findings of (Kumar *et al* , 2007) who reported that growth and yield characters were used separately for developing yield prediction models and also found that plant girth, number of leaves and plant height were helped in predicting yield up to 89.29 per cent. Study conducted by Chandahas and Agrawal (1984) also reported that, models using data at one point of time do not provide the idea of growth pattern of plant characters. Growth indices based on plant characters observed on two or more points of time during crop growth obtained as weighted accumulations of observations on plant characters in different periods.

5.3 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis has been applied in this study as a statistical approach to identify major variance components, their contributions and reduction of dimensionality. This method assists in reducing the number of characters and also helps in reducing the cost and time during field experiment. In this method, fewer variables explaining variations among individuals can be screened among various characters.

From the results, it was revealed that biometrical characters at different growth stages are treated as independent variables and we obtained eigen values and proportion of variance corresponding to each principal components. it was observed that first principal component alone has extracted about 55 per cent of total variability in all four cultivars.

In Nendran, first three principal components extracted about 83 per cent variation and H_4 and G_5 becomes highly contributing characters to the total yield.

In Njalipoovan, first four components (PC1, PC2, PC3 and PC4) explained up to 85 per cent of the total variation and plant height was equally contributing to the total yield in first principal component and LA and WSH are highly contributors in second principal component.

In case of Redbanana, first three principal components accounted for 81 per cent variation in the total yield. G_5, H_5 and H_4 were highly contributing characters in PC1. In second principal component, NF and WSH were more contributing to the bunch weight.

In Robusta, first two components (PC1 and PC2) are explained upto 80 per cent variation in the total yield. H_4 and G_4 in PC1 and NF in PC2 were the more contributing variables.

In general, 5-6 variables are sufficient to get maximum information on total yield in all cultivars. Plant height and plant girth at different growth stages are sufficient to explain maximum information and can be used for data collection in pre-harvest prediction of banana yield.

5.4 TREND ANALYSIS OF BANANA PRODUCTION IN KERALA

Statistical information on cropped area, production and productivity form the backbone of agricultural statistical system. Regional data analysis is extremely vital since it form the basis for economic and policy planning by the state and central governments. It is easy to formulate and initiate appropriate policy measures, if the data with regard to the trend (either increasing or decreasing) of area, production and productivity are obtained and analysed in advance.

Keeping above as view point, the present study was aimed to develop appropriate statistical growth models to describe the trends in banana production for the period of 25 years (1991-20105) in Kerala. The results obtained from the present study are discussed as follows.

Different linear and non-linear regression models were fitted to the available data sets. Results of fitting trends revealed that among attempted models, quadratic and cubic models were showed a higher R^2 and these two models were fulfills the assumptions of residuals like normality and randomness. For area (R^2 of 84 %), productivity (R^2 of 79%), whole sale price (R^2 of 91 %) and cost of cultivation (R^2 of 85%) cubic model was best and for production, quadratic model was best suitable (R^2 of 71 %).

Similar findings were found by (Indiradevi *et al,* 1990), who reported that quadratic model was significant and could explain satisfactorily the trend in production during period. In another study , conducted by (Rajan *et al*., 2015), they concluded that that cubic regression model was found to be best fit for estimating area, production and productivity of cotton .

5.5 GROWTH RATE ANALYSIS

Results of compound annual growth rate analysis are presented in last chapter, hence a brief discussion about growth of respective characters are presented below.

Both area and production of banana cultivation was positive and increasing over the study period from 1990-91 to 2014-15. Area of banana under cultivation was 32687 hectares during 1990-91 and it increased to 47.45 per cent over the period and it stood at 61936 hectares in 2014-15. Growth rate of area under cultivation of banana was 3.6 per cent during the period. Banana production fluctuates from year to year, so there was no constant increase or decrease in production and its growth rate was 1.4 per cent due to pest and disease incidences like pseudo stem weevil and sigatoka leaf spot. These two made vast damage in banana cultivation during last two decade. Productivity had negative trend during study period (-2.76 per cent) due to variation in production with respect to increasing in area under banana cultivation.

The cost of cultivation for last decade has shown increasing in trend (3.9 per cent), because of change in the production technology, scarce of labor resources and increasing in the price of agricultural inputs *etc*. The wholesale price of banana also has shown increasing trend(5.1 per cent) during last two and half decade , due to increasing in demand for nutrition, value added products of banana and increase in prices of other fruits *etc*.

5.6 Rainfall and rainy days adjusted growth rates to banana production

Climatic factors like rainfall and rainy days were found to be non significant for the last two and half decade. Means, change in area, production and productivity trend was not due to climatic factors on banana cultivation during study period.

FUTURE LINE OF WORK

- Similar multiple regression strategy may also used for different crops for pre-harvest prediction.
- Most of the work on pre-harvest forecasting of yield on biometrical characters is fitted using linear models *viz*., Regression models and principal component analysis. Since some of the biometrical characters may be related to yield in non linear form, hence it is better to explore non linear models for pre- harvest forecasting.
- For effective and essence of results in deciding the appropriate growth models, data for a long period can be made use

Summary

6. SUMMARY

The present research study entitled with "Pre-harvest forecasting models and trends in production of banana (*Musa* spp.) in Kerala." was formulated with the following objectives.

- To develop models for early forecasting of yield in four major banana cultivars grown in Kerala *viz.*, Nendran, Robusta, Redbanana and Njalipoovan
- To carry out the time series analysis of the trend in area and production of banana in Kerala

The study was mainly based on plant growth and yield characters of four commonly grown banana cultivars *viz.,* Nendran, Njalioovan, Redbanana and Robusta for early forecasting. Field experiment was conducted by selecting simple random sample of about fifty suckers from 300 sucker samplings of the four cultivars. All banana cultivars were planted during October -November 2014. Observations were taken on cultivars in monthly interval. Correlation and multiple regression techniques were adopted to develop the best forecast models. Secondary data on area, production, productivity, whole sale price and cost of cultivation for last twenty-five years were collected from Department of Economics and Statistics, Govt. of Kerala. Linear and non-linear models were used to study the trend analysis of banana production.

The salient findings of the study are given below

- Growth characters *viz.,* plant height and plant girth at earlier growth stage of all cultivars showed a positive and significant correlation with bunch weight,
- Correlation coefficients between bunch weight and sucker characters were non-significant in all cultivars, was an indication of non significant relation between bunch weight and suckers characters.

- Yield characters namely number of fruits, weight of second hand, fruit weight and fruit girth had positive correlation with bunch weight in all cultivars, but fruit length was negatively correlated with bunch weight.
- In prediction models of Nendran, plant height at 4th month and plant girth at 5th month were observed as the best predictor variables with R^2 of 80.5 per cent and best stage was 5th month for early forecasting of yield.
- In case of Njalipoovan, plant height and number of leaves at fourth month were statistically significant and highly contributing to total yield and chosen as best forecasting biometrical characters with R^2 of 81.7 per cent. Fourth month was selected as best time for early yield prediction.
- Multiple regression models for Redbanana showed that plant height and plant girth at fourth month were statistically significant and highly contributing to total yield and chosen as best forecasting biometrical characters with R^2 of 76 per cent. Fourth month was selected as best time for early yield prediction.
- Similar results of MLR were found in Robusta also, plant height at fourth month and plant girth at third month were chosen as best predictor variables for early prediction. From fourth month onwards these two variables made significant effect on bunch weight along with other biometrical characters (R^2 of 75 per cent). The best fitted model was obtained at fourth month for pre-harvest forecasting the bunch weight.
- Results of PCA summarized that, first principal component alone has extracted about 55 per cent of total variability in all four cultivars, first three PCs are explained 80 per cent of total variability in yield.
- In general, 5-6 variables are sufficient to get maximum information on total yield in all cultivars using PCA. Plant height and plant girth at different growth stages are sufficient to explain maximum information and can be used for data collection in pre-harvest prediction of banana yield.

- Results of fitting trends revealed that among attempted models, quadratic and cubic models showed a higher R^2 and these two models fulfill the assumptions of residuals like normality and randomness. For area (R^2 of 84 %), productivity (R^2 of 79%), whole sale price (R^2 of 91 %) and cost of cultivation (R^2 of 85%) cubic model was best and for production, quadratic model was best suitable model (R^2 of 71 %)
- Growth rate of area under cultivation of banana was 3.6 per cent for during the period. Banana production was fluctuates from year to year, so there was no constant increase or decrease in production. Productivity had negative trend during study period (-2.76 per cent) due to decline in production with respect to increasing in area under banana cultivation. Cost of cultivation and whole sale price has shown increasing trend during last two and a half decade.
- Climatic factors like rainfall and rainy days were found to be non significant for the last two and a half decade. Means, change in area, production and productivity trend was not due to climatic factors on banana cultivation during study period.

References

7. **REFERENCES**

Agrawal, R. 2015. Weather based pre-harvest crop forecasting – an overview. *Indian. J. Agric. Sci.*58 (19):2005-2014

Agrawal,R.,Jain,R.C.,Garg,R.N.,Singh and Gurucharan .1984.Model to forecast yield of rice (*oryza sativa*) using agro-spectral data. *Indian .J. Agric. Sci.*64(5):320-323.

Agribusiness. 2015. Going bananas: India's export zooms 45 %. Business line home page[online].Available:http://www.thehindubusinessline.com/. Agribusiness [9 April 2015].

Aneja ,B. and Chandrahas . 1984. Pre-harvest cotton crop yield based on plant biometrical characters. *Silver Jubilee Souvenir*, 41(2):48-58.

Anuradha, B., Koteswara, R.Y., Rama, K.P.V. and Srinivasa, R.V., 2007. Correlation and path analysis for seed yield and yield contributing characters in pigeon pea. *The Andra Agric. J.,* 54(1&2):9-12

Arunachalam, R. and Balakrishnan, V. 2012. Statistical Modeling for Wheat (Triticum *Aestivum*) Crop Production . *Int. J. of. Statist. and Appl.* 2(4): 40-46

Bal, S.K., Mukhergee, J., Mallick, K. and Hundal, S.S., 2004. Wheat yield forecasting models for Ludhiana district of Punjab state. *J. of. Agro. Met.*, 6; 161-165.

Bazgeer, S., Khamali, A., Eslamian, S.S., and Moradi.2008. Pre-Harvest Wheat Yield prediction using Agro meteorological Indices for different Regions of Kordestan Province, Iran. *Res.J. of. Environ. Sci.* 2(4):275-280.

Benjamini, Y.1988.Opening the Box a box plot. *The Am. Statist.* 42(4):257-262

Chandrahas, and Agrawal,R . 1984. Preharvest crop Production forecast methodologies- IASRI studies. IASRI- An era of excellence. Project Report, IASRI, New Delhi.pp.45-78

Chandrahas, B.H., Singh., and Jha, M.P. 1983. Final report of pilot studies on pre-harvest forecasting of yield of tobacco in Prakasam District (A.P), Annual Report, IASRI, New Delhi.pp.27-36.

Chandran, K. P. and Prajneshu, (2004) Computation of growth rates in agriculture: Nonparametric regression approach . *J. the Ind. Soc. of Agric. Stat.,* **57:** 382-392.

Chandrahas, 1992. Efforts of errors of measurement on estimates of characters of forecast model. *Annu. of. Agril.Res.*19 (2):139-143

Cummings, R. W. and Ray, S.K. (1969) 1968-69 Food grain production: Relative contribution of weather and new technology, *Econ and Political Weekly*, (Review of Agriculture) September.

Datarkar, S., Pagire, B. V., Darekar, A., and Hile, R. B. 2015. Region wise compound growth rates in area, production and productivity of Kharif ground nut in Maharashtra. *Int. J. of .Trop. Agric.* 33(2):1101-1106

Dhindsa, K.S. and Sharma,A. 1995, Analysis of cropping pattern change in Punjab during 1965-66 to 1990-91, *Indian* Economic Review 30(1): 69-87.

FAO [Food and Agricultural Organization]. 2015. [On-line].Available: http://www.fao.org/home/en.com[20 Jan2016]

GOK [Government of Kerala]. 2016. *Agricultural Statistics* 2014-2015. Department of Economics & Statistics,Thiruvananthapuram.46p.

Goodall, C. R. 1993. "Computation Using the QR Decomposition." *Handbook in Statistics.* Vol. 9, Amsterdam: Elsevier/North-Holland, 558p.

Groten, S. M. E. 1993. NDVI–Crop Monitoring and Early Yield Assessment of Burkina Faso. *Int. J. Rem. Sen* .,**14**(8):1495–1515

Indira, D. P., Thomas, E.K., and Thomas, J.K. 1990. Growth and supply response of Banana in Kerala. *Agril. Situ. India*. 45 (4) 239-242.

Jain,R.C., Jha,M.P., and Singh,D. 1981. Pre-harvest forecasting of sugar cane yield. *Indian. J. Agric.Sci.*51(11):757-761.

Jain,R.C., Agarwal, R. and Jha,M.P., 1985. Use of growth indices in yield forecast. *Biom. J.*27 (4):435-439.

Jain, S.K. and Kumar, V. 2012. Trend analysis of rainfall and temperature data for India. *Current . Sci.* 102(1): 37-50.

Jayasree, K. Resmi, P and Satheesh, B. K.2012. Price Behavior of Cassava in Kerala: An Econometric Analysis. *J. of .Root .Crops.* 38(2):168-173.

Kaiser, H.F. 1960. The application of electronic computers to factorial analysis.*Educ. and Psychol. measurement*. 20(2):141-151.

Kandala, V. M. and Prajneshu 2002. Fuzzy regression forecasting using remotely sensed data. *J . Ind. Soc. Rem. Sen.,* **30**: 191-95.

Kaur, P. and Hundal,.S.S,. 2009. Production potential of cereal crops in relations to climate changes under Punjab conditions. *J. of. Agron.Met.*11:18-23

Kaushik, K.K., 1993. Growth of Agricultural Production: a component analysis, Agricultural Transformation in a Developing Economy, Kanishka Publishers and Distributors, Delhi, pp: 216-238.

Khistaria, M.K., Vamora, S.L. Dixit, S.K., Kalola,A.D, and Rathod ,D.N.2004. Pre-harvest forecasting of wheat yield from weather variables in Rajkot district of Gujarat. *J. of. Agron. Met.*6:197-203.

Kumar, R., Rajan, S. and Singh, A., 2007. Multiple regression models for yield prediction in plaintain. *Indian. J. of .Hort.* 64 (2):159-162

Krishnakumar, K.N., Prasada Rao, C.S. and Gopakumar.2009. Rainfall trends in twentieth century over Kerala, India. *Atmos. Environ.* 43: 1940–1944.

Lalitha, A.1993. Cropping Pattern in Andhra Pradesh 1957 to 1977. M.Phil, Thesis of Osmania University, Hyderabad. 135p.

Maheswari, A .1996. Agricultural growth in semi-Arid Area - The Case of Karnataka. *Indian .J.Agric. Econ* .51 (3):315.

Malhi, R. K .M. Kiran, G. S. 2015. Analysis of trends in area, production and yield of important crops of India. *Int. J. Agri. & Agri. R.* 7 (1): 86-92,

Matthews, R.B., Hunt, l..A. and Gumcas, A. 1994. A model describing the growth of cassava (*Manihot esculenta* L. Crantz). *Field Crops Res.*, **36** : 69-84

Montgomery, D.C., Peck ,E.A.and Vinning , G.2003. *Introduction to linear regression Analysis* (3^{rd} Ed.). John Wiley and Sons (Asia) pvt. Ltd.258p.

Murray, D.B.1960. Shade and fertilizer relations in the banana. *Trop. Agric*, (Trinidad), 38:123-132.

Nandi, P.K. and Kanth, G.K., 2004.Study on Growth Models: A Critical Analysis with Reference to Andhra Pradesh and India *J. Indian .Soc. Agric. Statist.* 45 (1): 106

NHB[National Horticultural Board].2014. [On-line].Available: http://www.nhb.gov.in/[15 July 2015]

Onkar, S., Rai,T. and Chanrahas, 1998, Yield forecast model for wheat(*Triticum Aestivum*) crop. *Indian. J. Agric. Sci.,* 68(4):206-208.

Prajneshu .2002.Non –linear regression models and their applications. *J. Indian .Soc.Agric.Statist.*51:73-81

Priya, P.M., Radhakrishnan, V.V., Madhusoodanan, K.J. and Kuruvilla, K.M. 2003. Yield forecasting in cardamom plantations under intensive management. *J. Spices. and Aromat.crop.*12 (1):55-56.

Raghupathi, N.2005. *A statistical model for Pre-harvest forecasting of yield based on biometrical characters in cotton.* M.Sc Thesis, University of Agricultural Sciences, Bengaluru: p. 23

Rai, T. and Chandrahas .1998. Use of discriminant function of weather characters for developing forecast model of rice crop. Project Report, IASRI, New Delhi.

Rajan , M.S., Palanivel, M. and Mohan, K .S.2015. Forecasting of Cotton Area, Production and Productivity using Trend analysis. *Int.j. for. Appl.sci.& Eng.tec* .12(3):516-520

Rajarathinam, A. and Parmer, R.S. 2011.Applications of Parametric and nonparametric regression models for Area, Production and Productivity Trends of Castor *(Ricinus communis L.)* Crop. *Asian. J. of. Appl.sci.* 4(1):42-52.

Raju, V.T., Darsi, Rao, V.S. and Ravishankar , C. 1989. Spatio - Temporal, Growth and Distribution of Fruit Crops in Andhra Pradesh, *India . J. of. Agril. Econ.,* .42(3):438-445.

Ram, K. Rajan,.S. and Achal, S.2007. Multiple regression models for yield Prediction in plantain. *Indian. J. of. Hort.*64 (2):159-162.

Rao, G.S.P., Rekha, A. R.D. Rawal and M.B.N.V.Prasad. 1993. Yield prediction in ten commercial varieties of banana. *Hort. Res.* Golden Jubilee Symposium on horticultural research. Abstr. 24: 421.

Rao, T.V. 1964. Multivariate analysis of agronomic traits of new corn hybrids (*Zea maize* L.). *Int. J. Agric. Sci.* 1(16):314:322

Santhoshmithra, V.S.2011 *Computer simulation of Sweet Potato growth* Ph.D. Thesis Gandhigram Rural University, Tamilnadu: p. 120

Shamasundaran, K.S., Venugopalan. R . and Rekha, A. 2002.Statistical model for yield prediction in banana using morphological characters *.Proc. of Glob. Conf. in banana& plant*. Annual report 2003-2004. Indian Institute of Horticultural Research (IIHR). Bangalore. 95 p.

Shapiro, S.S. and Wilk, M.B.1965. An Analysis of Variance Test for Normality *.Biometrika*.52: 229-240.

Sheela.,V. L. and Nair, S.R. 2001. Growth , flowering , and yield potential of tissue culture banana (*Musa* AAB cv. Nendran). *Journal of Tropical Agriculture.* 39:1-4

Siegel, S. and Cestellen, N.J.1988. Nonparametric Statistics for Behavioral Sciences, *McGraw-Hill.* : 256p

Singh,D., Singh,H.P and Singh ,P.,1976. Pre-harvest forecasting of wheat yield *.Indian. J. Agric. Sci.,* 46(10):445-450.

Singh, D., Singh,H.P., Singh ,P and Jha, M.P. 1979. A study of Pre-harvest forecasting of yield of Jute. *.Indian. J. Agri. Res.,* 13(3):167-179.

Soares, J.D.R., Pasqual, M., Lacerda.W .S, Silva, S.O. Donato, S.L.R. 2013. Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. *Sci. Horti.* 155: 24–29

Soares, J.D.R., Pasqual, M., Lacerda.W .S, Silva, S.O. Donato, S.L.R..2014. Comparison of techniques used in the prediction of yield in banana plants. *Sci. Horti.* 167: 84–90

Subramaniyam, K.V. 1982. Analysis of Growth of Area, Production and Productivity of banana in India, 1965-66 to 1978-79. *Agric. Situ. India,.* 37 (6): 365-372.

Sunilkumar, G., and Prajneshu 2008. Modeling and forecasting India's marine fish production data Using wavelet thresholding for autocorrelated errors. *Ind. J. Fish.*, **55:**291-94.

Tittonell, P., Vanlauwe, B., Leffelaar, P.A. and Giller, K.E.2005. Estimating yields of tropical maize genotypes from non-destructive, on-farm plant morphological measurements. *Agric. Ecosyst. and Environ.* **105**: 213-220.

Theil, H. 1971. *Principles of Econometrics,* John Wiley & Sons, New York. pp.605–606.

Varghese, P.K. 2007. Economics of cardamom cultivation in Kerala, *Indian. J. of .Agril. Econ*. 62(1): 99-112.

Yadav, R.R, Sisodia, B.V.S., and Sunil Kumar .2014. Application of Principal component analysis in developing statistical models to forecast crop yield weather variables. *Mausam*.65 (3):357-360

\

Abstract

PRE-HARVEST FORECASTING MODELS AND TRENDS IN PRODUCTION OF BANANA (*MUSA* SPP.) IN KERALA

by

SHARATH KUMAR M P

(2014-19-101)

ABSTRACT

Submitted in partial fulfilment of the
requirement for the degree of

MASTER OF SCIENCE IN AGRICULTURE

Faculty of Agriculture
Kerala Agricultural University

DEPARTMENT OF AGRICULTURAL STATISTICS
COLLEGE OF AGRICULTURE
VELLAYANI, THIRUVANANTHAPURAM- 695 522
KERALA, INDIA

2016

ABSTRACT

The study entitled "Pre-harvest forecasting models and trends in production of banana (*Musa* spp.) in Kerala" was conducted at Instructional farm, College of Agriculture, Vellayani. The objectives of the study were to develop models for early forecasting of yield in four major banana cultivars grown in Kerala *viz*., Nendran, Robusta, Redbanana and Njalipoovan and also to carry out the time series analysis of the trend in area and production of banana in Kerala.

The study was based on both primary and secondary data. Initial and monthly observations on growth habits and yield of commonly grown banana cultivars were used for forecasting. Secondary data on area, production and productivity over a period of twenty five years (1991-2015) were collected from published sources of Directorate of Economics and Statistics, Govt. of Kerala and State Department of Agriculture. Additional information on price change and climatic factors were also incorporated in state level time series analysis.

Pre-harvest forecasting models developed for the first three months, using sucker characters and numbers of leaves were not found to be sufficient in forecasting yield and best models were identified from the fourth month onwards in all cultivars. Correlation analysis of yield (bunch weight) with biometrical characters in all four cultivars showed that correlation is positive and significant in 4th, 5th and 6th months of growing. Among biometrical characters, plant height and plant girth showed significant relationship with yield in all cultivars. In Njalipoovan, in addition to plant height and plant girth, number of leaves and leaf area (D-leaf) had some positive relationship with the ultimate yield. Meanwhile fruit characters like number of fruits, weight of second hand, fruit weight had significant correlations with yield in all cultivars.

Stepwise multiple linear regressions were attempted to primary data at every month. The statistically most suited forecasting models were selected on the basis of coefficient of determination (R^2), adjusted R^2 and mallow's Cp criteria. It resulted that, in nendran variety, plant height and plant girth were contributing to yield with highest R^2 of 0.80 in the 5th month (model $Y= -1.37+0.025\ H_4+0.10\ G_5$). Fruit characters were statistically significant to making of a 55 per cent of variation in total yield.

In Njalipoovan, models from 4th month onwards were found good for early forecasting of yield. Number of leaves, plant height, and leaf area and plant girth could predict yield with R^2 of 81.7%, while fruit characters, *viz.*, number of fruits, fruit length, fruit girth and fruit weight could predict yield with an R^2 of 71.88 %.

In Red banana, it was found that plant height and plant girth at fourth gave suitable prediction with an R^2 of 0.762, meanwhile fruit characters could predicted yield with an R^2 of 71 .28%.

In Robusta variety, prediction can be made from 4th month onwards as best predictor as plant height and girth (with an R^2 of 75.24 %). At harvesting stage, fruit characters could predict the maximum yield up to 96.76 %. Principal component analysis resulted that first three principal components are sufficient for getting maximum information from explanatory variables in all four cultivars with 75 % explained variation.

Linear and nonlinear growth models were developed for the purpose of estimating the growth rate and fitting the best model. The use of R^2, criteria of randomness and normality of time series data were used as a measure of goodness of fit. Cubic model was found as best fit for estimated trends in area, productivity, whole sale price and cost of cultivation under banana production. Quadratic function was selected as best suited for production trend. However, rainfall and rainy days were found to have less effect on changing in area, production and productivity of banana. Area, production, wholesale price and cost of cultivation showed a positive trend during past twenty- five years.

Hence, reliable estimate of a crop yield, well before harvest can be made of from 4th month onwards in all cultivars studied. Policy decisions regarding planning of crop procurement, storage, distribution, price fixation, movement of agricultural processing commodity, import-export plans, marketing can be formulated based on these forecasts.