PREDICTION OF INCIDENCE OF MAJOR PESTS IN JUTE CROP UNDER COOCH BEHAR DISTRICT OF WEST BENGAL

A Thesis

Submitted to the

Uttar Banga Krishi Viswavidyalaya

in partial fulfillment of the requirements for the

Degree of Master of Science (Agriculture)

in

AGRICULTURAL STATISTICS

By

Chinmaya Subhrajyoti Panda

(Registration No.: A-2020-016-M)

Department of Agricultural Statistics

Faculty of Agriculture

Uttar Banga Krishi Viswavidyalaya

Pundibari, Cooch Behar -736 165, West Bengal, India

2022

Dedicated

To

My Beloved

Family

UTTAR BANGA KRISHI VISWAVIDYALAYA
Pundibari, Cooch Behar, West Bengal-736165
Department of Agricultural Statistics
FACULTY OF AGRICULTURE
www.ubkv.ac.in

From:
Dr. D. S. Gupta
Professor and Head

Cell No.: +91-9434685759
e-mail ID: dsguptaubkv@gmail.com
: hod.agril.stat.ubkv@gmail.com

Certificate

This is to certify that the work incorporated in the thesis entitled "**PREDICTION OF INCIDENCE OF MAJOR PESTS IN JUTE CROP UNDER COOCH BEHAR DISTRICT OF WEST BENGAL**" submitted by Mr. Chinmaya Subhrajyoti Panda (Reg. No.: A-2020-016-M) in partial fulfillment of the requirements for the Degree of Master of Science (Agriculture) in Agricultural Statistics of Faculty of Agriculture of Uttar Banga Krishi Viswavidyalaya, is a faithful and bona-fide research work carried out under my supervision and guidance. Results of the investigation reported in the thesis have not been submitted anywhere for any other Degree or Diploma. The assistance and help received from different sources during the course of investigation have been duly acknowledged.

Debsankar Gupta

Date: 29-08-2022

Place: UBKV, Pundibari, Cooch Behar

Prof. (Dr.) Deb Sankar Gupta

(Chairman, Advisory Committee)

UTTAR BANGA KRISHI VISWAVIDYALAYA
Pundibari, Cooch Behar, West Bengal-736165
Department of Agricultural Statistics
FACULTY OF AGRICULTURE
www.ubkv.ac.in

From:
Dr. D. S. Gupta
Professor and Head

Cell No.: +91-9434685759
e-mail ID: dsguptaubkv@gmail.com
: hod.agril.stat.ubkv@gmail.com

Certificate

Certified that the thesis entitled **"PREDICTION OF INCIDENCE OF MAJOR PESTS IN JUTE CROP UNDER COOCH BEHAR DISTRICT OF WEST BENGAL"** submitted by Mr. **Chinmaya Subhrajyoti Panda (Reg. No.: A-2020-016-M)** in partial fulfillment of the requirements for the award of the degree of Masters of Science (Agriculture) in Agricultural Statistics of the Uttar Banga Krishi Viswavidyalaya has been checked against plagiarism through **URKUND (Ouriginal)** software on 24th August 2022 and that the similarity index has been achieved as **1%**, which is below the maximum tolerable range as per stipulation of this Viswavidyalaya. The thesis of Mr. **Chinmaya Subhrajyoti Panda** may be accepted for the award of the Degree of Masters of Science (Agriculture) in Agricultural Statistics of the Uttar Banga Krishi Viswavidyalaya.

Debsankar Gupta

Date: 29-08-2022

Place: UBKV, Pundibari, Cooch Behar

Prof. (Dr.) Deb Sankar Gupta

(Chairman, Advisory Committee)

Uttar Banga Krishi Viswavidyalaya

Certificate of Originality

To

The Chairman, Advisory Committee

Department of Agricultural Statistics

Faculty of Agriculture

Uttar Banga Krishi Viswavidyalaya

Sub: Undertaking regarding checking for similarity index of manuscript of submitted thesis

Respected Sir,

I wish to submit the following:

i. Name of the student (in block letters): **CHINMAYA SUBHRAJYOTI PANDA**
ii. Registration Number: **A-2020-016-M**
iii. Name of Chairman of Advisory Committee: **Prof. (Dr.) Deb Sankar Gupta**
iv. Name of the Department: **Agricultural Statistics**
v. Title of the Thesis: **Prediction of Incidence of Major Pests in Jute Crop Under Cooch Behar District of West Bengal**
vi. Total Pages of the thesis: **77** (main text) + **VII** (Reference)

I do hereby certify that the above-mentioned thesis was checked against plagiarism / for similarity detection. The process and outcome are stated below:

vii. Software used: **URKUND Software**
viii. Date: 24/08/2022
ix. Similarity Index: 1%

The complete report is submitted for review.

Thanking You.

Yours faithfully,

Chinmaya Subhrajyoti Panda
29-08-2022

Signature of student

Name of the student (in block letters): **MR. CHINMAYA SUBHRAJYOTI PANDA**

Only software for checking similarity index output is certified.

29/08/2022

(Signature of In-charge, Central Library with official seal)

Contd…/- (PTO)

Contd… from pre-page (Overleaf)

The complete report of the above-mentioned thesis has been reviewed by the Advisory Committee of the Student (following to be ticked as appropriate).

☑ The similarity index of the above-mentioned thesis is below the accepted norms.

☐ The similarity index of the above-mentioned thesis is above the accepted norms, because of the following reasons:

1.……………………………………………………………………………………
2.……………………………………………………………………………………
3.……………………………………………………………………………………
4.……………………………………………………………………………………
5.……………………………………………………………………………………

The thesis may be considered for the award of Degree (relevant software generated document attached).

Forwarded for consideration.

Debasankar Gupta
29-08-2022
Signature of Chairman, Advisory Committee

Forwarded

29-08-2022

Head
Department of Agricultural Statistics
Faculty of Agriculture
Uttar Banga Krishi Viswavidyalaya
P.O. Pundibari, Dist. Cooch Behar
West Bengal, Pin-736165

(Signature and Official Seal of Head of the Department)

The process for submission of thesis is allowed / ~~not allowed~~ to commence.

29/8/22

DEAN
Post-Graduate Studies
Uttar Banga Krishi Viswavidyalaya
Pundibari Cooch Behar W B

(Signature and Official Seal of the Dean, Post Graduate Studies)

APPROVAL OF EXAMINERS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (AGRICULTURE)

IN

AGRICULTURAL STATISTCS

We, the undersigned, having been satisfied with the performance of Mr. Chinmaya Subhrajyoti Panda (Reg. No.: A-2020-016-M), in the viva-voce examination on final evaluation of the thesis entitled **"PREDICTION OF INCIDENCE OF MAJOR PESTS IN JUTE CROP UNDER COOCH BEHAR DISTRICT OF WEST BENGAL"**, conducted today on 2nd September 2022, recommend that the thesis be accepted for the award of the Degree of Master of Science (Agriculture) in Agricultural Statistics.

Name	**Signature**
1. Prof. (Dr.) Deb Sankar Gupta Chairman, Advisory Committee	02-09-2022
2. Dr. Mrinmoy Ray Scientist, ICAR-IASRI, New Delhi (External Examiner)	Mrinmoy Ray 02/09/2022
3. Dr. Pradip Basak Member, Advisory Committee	02.09.2022
4. Dr. Manoj Kanti Debanth Member, Advisory Committee	MK Debnath 02/09/2022
5. Dr. Sabita Mondal Member, Advisory Committee	Sabita Mondal 02/9/22
6. Prof. (Dr.) Deb Sankar Gupta Head of the Department	02-09-2022
7. Prof. (Dr.) Prabhat Kumar Pal Dean, Post Graduate Studies	2/9/22
8. Prof. (Dr.) Dibyendu Mukhopadhyay Dean, Faculty of Agriculture	02/09/2022

Acknowledgement

With reverence I express my deep sense of gratitude to **Prof. (Dr.) Deb Sankar Gupta,** Professor and Head, Department of Agricultural Statistics, UBKV, Pundibari and Chairman of my Advisory Committee for his initiative, benevolence, endurance, constructive criticism and constant monitoring during the period of my investigation and also in the preparation of this thesis.

I am equally obligated to **Dr. Pradip Basak,** Assistant Professor, Department of Agricultural Statistics, UBKV, Pundibari and member of my Advisory Committee for his counsel, valuable suggestions and ready to help nature for accomplishment of my research work by guiding every possible moment throughout the whole journey. I consider myself fortunate in having the privilege of being guided by him.

I am extremely thankful to **Dr. Manoj Kanti Debnath**, Assistant Professor, Department of Agricultural Statistics, UBKV, Pundibari and member of my Advisory Committee for his moral support and useful suggestions during my investigation.

I am highly grateful to **Dr. Sabita Mondal**, Assistant Professor, Department of Agricultural Extension, UBKV, Pundibari and member of my Advisory Committee for her great help during my thesis work.

I would like to express my gratefulness to **Prof. (Dr.) Arunava Ghosh**, **Prof. (Dr.) Satyananda Basak, Dr. Sankalpa Ojha** and **Mr. Pratap Basak** for their genuine efforts in offering suitable advice on course work and assisting me in learning new ideas and clearing existing concerns.

I express my extreme gratitude to the esteemed Dean, Post Graduate Studies, **Prof. (Dr.) Prabhat Kumar Pal**, to our respected Dean, Faculty of Agriculture, **Prof. (Dr.) Dibyendu Mukhopadhyay** and to Honorable Deputy Registrar, **Prof. (Dr.) Ayon Roy** for extending their help, assistance, and guidance throughout the period of study.

*I want to express my respect and gratitude to **my family** for their sacrifices they did for me. These were the love, affection, inspirations and moral boosting of them which gave me sufficient energy to complete this thesis in time.*

*Special mention to my friend **Akruti** for being there and supporting me psychologically throughout the journey. Also, I am very grateful to have a friend like **Mithu, Jeeban** and senior like **Naffees didi** who were always there for any academical or non-academical circumstances.*

I do not have any words to express the help and support given by my seniors specially Mihir vai, Manish vai, Subhra vai, Siddharth Vai, Kaushik dada, Nishtha didi, Vaidheki didi, Ajith dada, Prem dada, Chowa Ram dada, Pavithra didi, Neethu didi, Vinay dada and all my juniors specially Chandan, Manas, Jeeban, Jageswar, Pravat, Karthik, Reesma and Keerthana.

I am thankful to my classmates Soumyasesh, Vivekananda, Dipta, Smaraki and Ragini for their friendly approach and moral support throughout my entire M.Sc. tenure.

I am also thankful to all my seniors, juniors and friends in UBKV for their moral support and inspirations.

I take this opportunity to appreciate the help rendered by the staffs of UBKV and Department of Agricultural Statistics. Special thanks are also due to Ratan Dada and Ratna Didi in making the official tasks smooth and easy for us.

Last but not least, I am thankful to Indian Agricultural Research Institute (IARI) for the financial assistance provided to me in the form of NTS Fellowship during the tenure of my study.

Chinmaya Subhrajyoti Panda

Date: 29-08-2022

Chinmaya Subhrajyoti Panda

Place: UBKV, Pundibari, Cooch Behar

(Reg. No.: A-2020-016-M)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREDICTION OF INCIDENCE OF MAJOR PESTS IN JUTE CROP UNDER COOCH BEHAR DISTRICT OF WEST BENGAL

ABSTRACT

Being one of the most important commercial crops, Jute is cultivated extensively in Cooch Behar district of West Bengal. The primary reason behind this hefty cultivation is the presence of dominant share of small and marginal farmers within the farming community. Cooch Behar suffers a substantial amount of physical and economical loss every year in jute cultivation due to several major insect pests infestation such as Yellow Mite (*Polyphagotarsonemus latus* Banks) and Jute Semilooper (*Anomis sabulifera* Guen) at different stages of the crop growth. Thus, seasonal plots have been constructed for a better understanding of the seasonal incidence of major pests in the study area and findings reveal that in case of Yellow Mite pest incidence is maximum at 55 DAS, while in case of Semi Looper it is observed at 45 DAS. Based on Pearson's correlation analysis, it is found that the weather parameters such as minimum temperature at current week, maximum RH at one week lag, minimum temperature, minimum and maximum RH at two week lag are significantly correlated with the incidence of Yellow Mite, while in case of Jute Semilooper maximum temperature, minimum and maximum RH at two week lag are significantly correlated. Different forecasting models like ARIMA, ARIMAX, SARIMA, SARIMAX and SVR have been fitted on the training dataset followed by model validation on the testing dataset using RMSE values. In case of Jute Semilooper, SARIMAX model is found to be the best fitted model on the basis of least RMSE value followed by SVR and SARIMA. Similarly, for Yellow Mite ARIMA, ARIMAX and SVR models have been fitted. It is observed that ARIMAX model produces the least RMSE value followed by SVR and ARIMA. Following successful model validation, forecasting of the jute pest incidence is done for the year 2022 at 35, 45, 55 and 65 DAS using the best fitted models.

***Keywords*:** **Jute, Pest Incidence, Forecasting, ARIMA, ARIMAX, SARIMA, SARIMAX, SVR.**

Debsankar Gupta
29-08-2022
Prof. (Dr.) Deb Sankar Gupta

(Chairman, Advisory Committee)

Chinmaya Subhrajyoti Panda
29-08-2022
Chinmaya Subhrajyoti Panda

(Reg. No.: A-2020-016-M)

1. Introduction

Chapter 1

INTRODUCTION

Jute fibre has a long history as an important commercial crop. Jute (*Chorchorus spp.*) is a dicotyledonous fibre crop that belongs to the family Tiliaceae and genus Chorchorus. The word Corchorus seems to be derived from the ancient Greek word ‘Korkhoros’, which was used for this pot herb Jute. There are also references to this crop in ancient events and literature like Kautilya’s Arthashastra, "Aini Akbari" and "Kabikinan Chandi". In pre-biblical times, it was used for cordage and fabrics. In earlier days of European countries, rope was made out of hemp, which was quite expensive. So, in the 18th century, Roxburgh found a suitable replacement for the plant yielding fibre and coined the term "Jute". On the basis of commercial importance, jute fibre is extracted from two well-known species, *C. capsularis* (White Jute) and *C. olitorious* (Tossa Jute). White jute is also known as "tita" or "bitter jute" due to the presence of glucoside (corchorin) which is bitter in taste, while tossa jute is known as "mitha" or "sweet jute" due to its tastelessness.

Next to Cotton (*Gossypium spp.*) jute is the most economical and vital of all textile fibres and is used extensively in manufacturing several types of packaging materials for various agricultural and industrial products. It is also nick named as “Golden fiber” due to its financial advantages. Jute and Mesta crop are together known as raw jute as their uses in trade and industry are almost same. It makes a significant contribution to the nation's economy. Previously it was considered as a raw material for packaging industry. But now-a-days it has become a versatile raw material for various industries, such as, textile, paper, building and automotive industries, used as soil saver, decorative and furnishing materials, etc. Due to its biodegradable and annually renewable nature, for environmental concerns, raw jute is considered as an eco-friendly and sustainable packaging material, against the synthetic fibers, which are claimed hazardous for nature by the environmentalists (Sen *et. al.*, 2008). Considering its environmental significance, the year 2009 has been declared as the International Year for Natural Fibers by Food and Agriculture Organization (FAO) (Chapke, 2013).

Cultivation of jute mainly occurs in humid tropical climate under rainfed condition, summer season and well-drained, sandy loam soil. Nearly 85% of total jute cultivation in all over the world is concentrated in the fertile geographic region known as Ganges delta which is shared by both India and Bangladesh. Being a commercial fiber crop, jute plays an important role in the economy of various South-East Asian countries like, India, China, Bangladesh, Nepal, Thailand and Myanmar. In 2019-20 total production of raw jute in worldwide was 165.43 lakh bales (of 180 kg each) out of which 79 lakh bales from India, 85.78 lakh bales from Bangladesh and 0.65 lakh bales from other jute producing countries like Myanmar, Nepal and Thailand (FAO, 2022). In 2019-20 Bangladesh exported around 60% of its total jute production while India exported only 10-15% to meet its domestic demand. Being the largest consumer of raw jute, India imports a good chunk of raw jute from Bangladesh (Commission for Agricultural Costs and Prices, Govt. of India).

Considering the agro-climatic requirement of this fibre crop, it is cultivated particularly in the Eastern and North Eastern states of India like West Bengal, Bihar, Assam, Odisha, Tripura, Meghalaya and Nagaland. Earlier days, it was mostly grown for domestic use and fuel. But currently the jute fibres are mainly used to manufacture products like hessian cloths, sacking cloths, jute yarn, bags, twines, tarpaulins, shopping bags, geotextiles, floor mats, carpets, canvas, and so on (Singh *et. al.*, 2019). It has been grown in the undivided Bengal area of India from ancient times. Around 1790s, mainly as an export item to the Western Hemisphere commercial jute cultivation was introduced in the country. In 1855, India started its own jute-processing units, converting Calcutta as a major centre. After partitioning in 1947, much of the jute-producing land remained in Bangladesh, which caused great damage for Indian jute industry. But fortunately, most of the jute mills were situated in India. Since then, raw jute productivity has increased by twofold, owing to the development of high-yielding varieties, improved production technologies and government policy support (Sinha *et. al*., 2009). This can be supported by the fact that during partition area and production of raw jute was 2.6 lakh hectares and 16 lakh bales, which during 2019-20 reached 6.73 lakh hectares and 98.76 lakh bales respectively. India is also world's largest consumer of jute and jute goods, with around 85% of production going to meet domestic demand, primarily for sacking (nearly 80%) under the Jute Packaging Materials Act (Compulsory Use in Packing Commodities),1987. Also, in 2019-20 India exported around 24.5 thousand tonnes of raw jute next to Bangladesh (Directorate of Economics and Statistics, MoAFW, Govt. of India).

In India nearly 4 million small and marginal farmers, 0.25 million workers and 0.5 million traders find profitable employment in this raw jute sector (Chapke, 2013). From the above data it gives evidence that being a highly labour-intensive crop, raw jute sector ensures job for substantial section of rural population. So, considering the socio-economic importance of jute cultivation, any positive impact in this sector will directly benefit the small and marginal farmers involved in this field.

Raw jute industry has social, economic and physical importance on 33-35 lakh small and marginal farmers involved in jute cultivation (Sarkar *et. al.*, 2016). In physical aspect, majority of cultivated land under jute comes under West Bengal and almost 75% of the functional jute mills are situated in West Bengal. So, jute produced in this region meets the requirement of the raw materials needed for these mills, which eventually fulfills the need of binding materials for rural people. In economic aspect, it is an important commercial crop in West Bengal as it ensures hard cash to poor and marginal farm families. It is also gaining importance in international market due to its biodegradability, sustainability and its positive residual effects which eventually improve soil properties and quality. Considering the social importance, harvesting time of raw jute fibre coincides with major festivals of different Bengali communities such as Hindus, Muslims etc. Marketing of these raw fibre somewhat fulfils the monetary need for festivals, essentials and entertainment purpose. Besides these utilities mainly in April-May, young jute leaves and tender twigs are used as leafy vegetables by Bengalis due to its nutritional importance (Choudhary *et. al*., 2013). Also, the by-products known as jute sticks can be used in various farm and household purposes like fencing, support in vegetable fields and beetle vines and as fuel.

Jute has been assimilated into the socio-economic life of rural in West Bengal, not merely as a major commercial crop also as a component of lifestyle. Raw jute industry exerts tremendous economic and trade importance predominantly in West Bengal. Currently there are 100 jute mills situated all over India out of which 73 are in West Bengal alone. West Bengal takes the lion's share of raw jute production in India accounting for over 5.54 lakh hectares cultivated area and 77.77 lakh bales of production during 2015-16, which has reached 5.18 lakh hectares cultivated area and 80.67 lakh bales of production during 2019-20 (Directorate of Economics and Statistics, MoAFW, Govt. of India).

One of the major raw jute producing district in West Bengal is Cooch Behar, situated in the northern part of the state just below the Himalayas. It comes under a special category of agro-ecology for its geographical location which is also known as Terai zone. It has adequate natural resources like forest, river, fertile soil etc. Besides these, other remarkable features include adequate rainfall, a wide variety of crops/ plants, livestock, skilled human resource (labour), etc. In this region agriculture is the chief source of livelihood and employment. Here in the farming community more than 90% of farmers are either small or marginal and the number of marginal farmers has an increasing trend (Roy, 2016). Mainly grown in pre-kharif season as a cash crop, Jute cultivation covered 67.92 thousand hectares area and production was 7.88 lakh bales during 2015-16, which was later 60.28 thousand hectares area and 8.21 lakh bales production during 2019-20 (Directorate of Economics and Statistics, MoAFW, Govt. of India).

Jute cultivation carried out in pre-kharif season, suffers huge losses every year due to a number of insect pests infestation in different stages of plant growth. However, in global scenario there are around 40 insect pests identified related to jute crop (Sadat *et. al.*, 2015). Out of which some of the major pests causing economic damage in India are jute semilooper (*Anomis sabulifera* Guen.), Bihar hairy caterpillar (*Spilarctia obliqua* Wlk.), indigo caterpillar (*Spodoptera exigua* Hubner), stem girdler (*Nupserha bicolor* Dutt), stem weevil (*Apion corchori* Marshall), grey weevil (*Myllocerus discolor* Bohemus), yellow mite (*Polyphagotarsonemus* latus Banks) and red mite (*Oligonychus coffeae* Nietner) (Sadat *et. al.*, 2015). It is estimated that the avoidable loss in fibre yield was found to be around 31-34% in West Bengal (Rahman *et. al.*, 2012).

The jute semilooper (*Anomis sabulifera* Guen), a cosmopolitan pest, belongs to the family Noctuidae and order Lepidoptera. It is monophagous in nature and it has high degree of destruction. It is reported from mostly entire jute cropping region all over the world and

Fig. 1.1.1: Jute crop cultivated in UBKV research field, Pundibari, Cooch Behar

Fig. 1.1.2: Yellow Mite infestation in jute crop

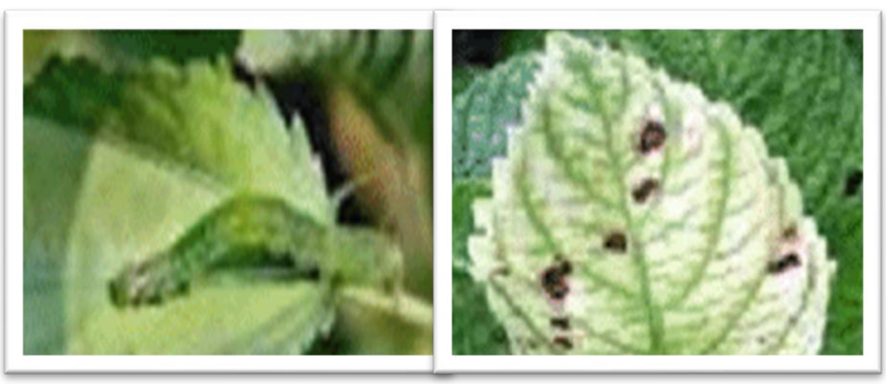

Fig. 1.1.3: Semi Looper infestation in jute crop

in some cases, it causes damage up to 90% of the leaves of jute plant, which results in poor plant growth and ultimately lowers the fiber quantity (Sadat *et. al.*, 2015). According to Dutta (1958), pre-monsoon rains followed by drought conditions are ideal for semilooper outbreaks, which can result in crop losses up to 50%. Generally, the second generation of the pest is more destructive than the parent generation (Sadat *et. al.*, 2015). The damage is caused by the larvae, which feed voraciously on the leaves, later resulting in defoliation of the plant.

The bihar hairy caterpillar (*Spilarctia obliqua* Wlk.) belongs to family Arctiidae and order Lepidoptera. It is polyphagous in nature and its degree of destruction is moderate. It is distributed over the major jute producing countries like India, Bangladesh and China. Once it was considered as a sporadic pest on jute, but now it has become a major threat to jute plant from last two decades. In this pest attack young caterpillar causes leaf skeletonization, older insects may defoliate the plants. It generally reduces the fibre yield up to 30 % causing severe economic damage to jute crop (Kumari, 2019). The larval population of bihar hairy caterpillar will be more due to high evening humidity than the other weather parameters (Bhosale *et. al.*, 2019).

Yellow Mite (*Polyphagotarsonemus latus* Banks) belongs to the family Tarsonemidae and order Acarina, is a destructive pest for jute production. It is highly polyphagous in nature and its nature of damage is moderate. It is distributed almost all over the world. It feeds on the apical leaves and sucking sap causes leaf browning and curling and loss of nutrition in younger plants due to sucking, causes stunted growth and apical stem become twisted (Rahman *et. al.*, 2012). Early sown jute is more prone to damage than lately sown crop (Gotyal *et. al.,* 2018). Infestation generally occurs in mid-May and gets its peak in June and in late July. Suitable condition for rapid growth of mites is dry weather with moderately high temperature, while damp weather is unfavourable for the infestation (Islam *et. al.*, 2020). The highest losses in fibre yield due to yellow mite infestation was found to be 74.71% (Kamruzzaman *et. al.*, 2013).

Agricultural production losses can’t be controlled fully as it is affected by numerous uncertain factors, but it can be minimized up to some extent by following certain management practices. One of the major crop losses is generally caused by infestation of insect pests, resulting deterioration in terms of both quality and quantity. Rahman et. al. (2012) stated that particularly in jute crop cultivated in West Bengal, the preventable losses caused by major insect pests was found to be 31-34%. Certainly, this percentage can be reduced by adopting some sustainable plant protection measures such as integrated pest management system, use of biological control, mechanical methods and minimum use of pesticides etc. Also, from the studies conducted by different scientists like Rahman *et. al.* (2012), it was evident that incidence and existence of insect pests has a significant relationship with the weather parameters. So, influence of weather parameters can't be overlooked while conducting effective plant protection measures. But this combined effect causing crop loss has not been investigated that expansively, especially in jute crop cultivated in West Bengal. For studying this complex association between weather parameters and occurrence of insect pests we should have prior knowledge about the seasonal incidence of that particular insect pest. Generally, incidence of insect pests is affected by weather parameters like precipitation (mm), maximum and minimum temperature (˚C) and maximum and minimum relative humidity (%).

Various mathematical, statistical and simulation models can be used for timely and accurate forecasting of pest incidence which will help the farmers in proper management of the pests. This eventually benefit the farmers in minimizing the losses caused by insect pests. Models specifically taking the effects of weather parameters into account can be of much use for precise prediction of pest incidence, so that effective plant protection measures can be followed. A model is a simple and symbolic representation of a complex system in reality, where the salient features of the system are described. Traditional models are based on data set fitting, regression and approximation theory. But now-a-days using advanced computational power, complex models can be designed for an accurate and precise forecasting. Some of the significant models helpful in prediction of incidence of insect pests by statistical techniques include simple and multiple linear regression technique, non-linear models and time series analysis.

Being one of the most important methods in statistical modelling, time series forecasting method predicts the future values of a variable based on the past observations obtained in a sequential order over a period of time, while also explaining the underlying principles. Generally available data and the trend observed in the data determine the selection of forecasting method. For insect pest forecasting model often time series data for insect count is taken into account. In count data, data can take only non-negative integer values. It is very important to maintain discreteness of data and satisfy the properties of time series data, while dealing with such time series count data. While forecasting insect pests, important environmental elements and several key traits of insects should be taken into account. According to Prasad *et. al.* (2012) data of weather parameters should be collected, analyzed and incorporated in models for reliable forecasting. From several research works it is evident that at least 30 standard meteorological weeks are needed for a valid and suitable prediction model formation (Bhat *et. al.*, 2018). Also, it was found out that incidence of insect pests was correlated with current time period as well as 1 to 4 lead times (Katke *et. al.*, 2009, Balikai *et. al.*, 2019).

However, the aspects regarding incidence of major pests in jute crop in Cooch Behar district and the role of weather factors affecting it have not been explored extensively. Therefore, considering all the discussions carried out above, the present study mainly focusses in the following objectives.

OBJECTIVES:

i. To study the seasonal incidence of major pests in Jute in Cooch Behar district of West Bengal.
ii. To predict the incidence of major pests in Jute in Cooch Behar district of West Bengal.

2. Review of Literatures

Chapter 2

REVIEW OF LITERATURES

In the present chapter, a detailed review of published literature has been conducted in order to track the previously carried out research activities and to gain knowledge about the recent developments that have occurred relevant to the current study. It also helps in creating a base for detecting and developing appropriate research methodologies pertinent to the current investigation. However, past studies assist in identifying gaps and new prospects in any research.

In light of the investigation's purpose, the primary objective is to build a suitable model to recognize the seasonal incidence of major pests in jute crop in Cooch Behar district. The secondary objective focuses on understanding the trend followed in the past years and accordingly, prediction is done for the upcoming years.

In this chapter, an attempt was made to review the literature relevant to the study, keeping in mind the aforementioned objectives. Current chapter covers areas regarding major pests in jute crop, their seasonal incidence and prediction of those major pests using different statistical methods.

Autoregressive Integrated Moving Average (ARIMA) has been considered as the most popular model in linear modelling for a long time, but some recent developments in non-linear models have gained attention. Also, time series data are not purely linear or non-linear in nature. So, Zhang *et. al.* (2001) proposed hybrid model as a combination of ARIMA and Artificial Neural Network (ANN) models and forecasting accuracy got improved.

Agrawal *et. al.* (2007) developed weather-based forecasting models using techniques like regression models, complex polynomial through GMDH method and ANN method and found out satisfactory and timely forewarning of pest population/disease severity.

Katke *et. al.* (2009) studied seasonal incidence of grape mealy bug and found that incidence of mealy bug for all the seasons was significantly negatively correlated with minimum and maximum RH, and rainfall during current time to 4-week lead time and for minimum temperature also same pattern was observed except at the week of observation where it was non-significant. For maximum temperature at current and LT1 the correlation was positive and significant, but at LT2 and LT4 it was positive and at LT3 it was negative and non-significant.

For pest prediction, conventional approaches like time series, regression analysis and Autoregressive Moving Average (ARMA) models have a number of assumptions and exogenous data are used. While in one hand neural networks proved to be cost-efficient methods, but in other hand it has its drawback in training process. Therefore, Lv *et. al.* (2009) proposed a hybrid model using neural network with practical swarm optimization learning algorithm and found that the hybrid model predicted accurately than the traditional BP- based perception and ARMA model in terms of performance indicators like Root Mean Square Error (RMSE) and MAXIMAL.

Damos *et. al.* (2010) constructed a three parameter Boltzman and a four parameter logistic non-linear regression model to study the emergences and seasonality of major moth pests of peach. The four parameter logistic non-linear regression model fitted better in most data sets as evaluated by the criteria like Adjusted R^2, the Akaike Information Criteria (AIC) and Bayes-Schwartz Information Criteria (BIC).

Prasad *et. al.* (2012) found that pest monitoring was the base for early warnings, development and validation of pest forecast models which eventually led to successful implementation of IPM modules. They also stated that several key characteristics of insect pests and important environmental elements should be collected and analyzed in models for an accurate and reliable forecasting.

Rahman *et. al.* (2012) conducted experiments on olitorius jute and found that incidence of jute semilooper (*Anomis sabulifera* Guen) was negatively correlated (r = -0.795 to -0.725) with the maximum temperature but positively correlated with minimum temperature (r = 0.528 to 0.715), morning relative humidity (r = 0.579 to 0.857) and afternoon relative humidity (r = 0.261 to 0.876). Regarding yellow mite (*Polyphagotarsonemus latus* Banks), its incidence showed positive association with morning relative humidity (r = 0.563 to 0.679) and afternoon relative humidity (r = 0.526 to 0.618) but rainfall had a negative impact on it. Thus, it was concluded that weather parameters particularly temperature, relative humidity and rainfall played a crucial role on occurrence and existence of different insect pests on jute crop.

Rahman *et. al.* (2012) found that certain plant characteristics of jute plant had effect on incidence of major pests. Plant height and leaf characteristics had positive correlation with incidence of some of the major pests. Incidence of yellow mite showed positive and significant correlation with leaf area and moisture content of leaves. Also, characteristics like leaf thickness and chlorophyll content of leaves and stem, fibre thickness and moisture content of stem had some significant effect on incidence of pests.

Kamruzzaman *et. al.* (2013) estimated that highest fibre yield losses due to yellow mite (*Polyphagotarsonemus latus* Banks) infestation in jute crop cultivated in net house condition was 74.71%, highest seed yield losses was 64.34% and highest stick weight losses was 57.18%. Further, high mite population caused significant losses in fibre and stick weight in different varieties of jute crop.

Several analytical and other techniques on pest surveillance had been carried out by scientists to get the information about the occurrence of peak activities of insects in order to develop suitable pest management system. For this purpose, Patil *et. al.* (2013) designed one standard feed forward Multi-Layer Perceptron (MLP) neural network and found that it was better in comparison to the existing models.

Paul *et. al.* (2013) considered historical data and external climatic information as base for forecasting of crop yield. In this investigation they developed five different models using Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) time-series methodology at five important stages of wheat growth. The ARIMAX model was able to generate better pre-harvest projections based on weather variables.

Permanasari *et. al.* (2013) fitted seasonal ARIMA (SARIMA) model for prediction of number of malaria incidence in US and got SARIMA (0, 1, 1) $(1, 1, 1)_{12}$ as best fitted model with 21.6% MAPE value.

Similarly, de Oliveira *et. al.* (2014) carried out hybrid model composed of ARIMA-Support Vector Regression (SVR) methodology optimized by the particle swarm optimization (PSO) algorithm and found out that PSO-ARIMA-SVR model outpaced other models.

Mishra *et. al.* (2014) studied population dynamics and incidence of mites on potato. They found that population of mite was at peak during later part of November. Further, correlation of mite population with weather parameters like maximum and minimum temperatures was negative and non-significant, but with morning and evening RH it was positive and significant.

Arya *et. al.* (2015) applied ARIMAX time-series model inducting historical data and external climatic information like weekly maximum temperature, minimum temperature, rainfall, maximum RH and minimum RH and evaluated on the basis of relative mean absolute prediction error (RMAPE). Using the aforesaid combined data set, forecasting was carried out in order to help the farmers in reducing the losses caused by pest infestation.

Sadat *et. al.* (2015) concluded that losses due to insect pests was one of the main constraints for less jute productivity in India and they have suggested to adopt effective Integrated Pest Management (IPM) practices.

Lee *et. al.* (2017) installed IoT systems near orchards and analyzed the correlation between appearance of insect pests and weather data such as temperature and humidity, which was later used in formulation of pest prediction models.

Sometimes assumption of homoscedasticity in errors is not satisfied and non-linear portion of the time series data is ignored. So, Mitra *et. al.* (2017) proposed two hybrid methodologies such as ARIMA-GARCH and ARIMA-ANN for model building and forecasting of potato price in Agra market. Brock-Dechert-Scheinkman (BDS) test was also carried out for checking non-linearity in the residuals. On the basis of mean absolute percentage error (MAPE) and RMSE ARIMA-ANN model outdid other models.

Narayanasamy *et. al.* (2017) stated that weather-based forecasting models were of much importance in sustainable integrated pest management system, as occurrence of insect pests were correlated to weather elements viz., temperature, rainfall, relative humidity, sunshine hours and wind speed. For prediction of the population of Yellow Stem Borer (YSB), Brown Plant Hopper (BPH) and Rice Leaffolder (RLF), a Generalized Linear Model was developed and from the results of the chi-square test, it was evident that, besides the induction of weather parameters, some other factors like variety, soil, fertilizer application, etc., could have been used to increase the predictability of the model.

Aswathi *et. al.* (2018) took weekly data of incidence of aphid, thrips, jassid and whitefly of cotton and weather parameters for development of statistical models like multiple linear regression, ARIMA and ARIMAX for forecasting purpose. The weather parameters included were rainfall, maximum temperature, minimum temperature, morning humidity and evening humidity. After comparing these models on the basis of root mean square error, it was found that ARIMAX model was more accurate than the other models.

Bhat *et. al.* (2018) found that mean and maximum Percent Disease Index (PDI) of early blight had significant and positive correlation with morning and evening relative humidity (RH) at one week and two week lags. For maximum temperature it showed negative and non-significant correlation and for minimum temperature it showed negative and significant correlation for current week but maximum temperature lagged by two weeks had significant negative correlation with maximum severity of early blight. Rainfall had a positive and non-significant association with mean and maximum severity of early blight. After validation it was concluded that beyond 30 SMW prediction models were suitable.

Durgabai *et. al.* (2018) stated that for upliftment of conventional farming system different machine learning algorithms such as K-Means Clustering, ANN, Support Vector Machine (SVM) etc. should be used for effective pest management as these were found to have higher accuracy than traditional statistical models such as regression analysis.

Paul *et. al.* (2018) carried out studies on field incidence of Sterility Mosaic Disease (SMD) of Pigeon pea and found that incidence of SMD was significant and negative to the influence of evening humidity at current week; significant and positive to the influence of sunshine of current to two lagged weeks. Whereas, for maximum incidence of SMD, significant negative association was found with minimum temperature and evening humidity at current, one week and two-week lags; significant positive association with sunshine at current to two lagged weeks. For forecasting the incidence of SMD, SVR was found out to be better than multiple linear regression, ARIMAX and ANN models, in terms of RMSE and mean square error (MSE).

Suyal *et. al.* (2018) investigated the seasonal incidence of insect pests of soybean crop and observed that there was a significant association between the population of pests and weather parameters, which provided an opportunity for modification of pest monitoring systems and integrated pest management systems by taking into account climatic factors.

Field trials conducted by Balikai *et. al.* (2019) on rabi sorghum found that, there was a significant negative correlation between shoot fly deadhearts and minimum temperature at current time and at lead times (LD) 1 to 3 but with maximum temperature the association was positive at all lead times except the current one and non-significant with all lead times 0 to 3. Leaf sugary exudates were positively and significantly correlated with minimum temperature at LT0 to LT2. Aphid population was positively and significantly associated with minimum and maximum temperature at LT0 and LT1, but with morning RH at LT0 and LT1 association was negative and significant. From the results it can also be concluded that for estimating the incidence of any insect pests or damage, prior weather data are as important as the current weather data.

Chiu *et. al.* (2019) developed ARIMA model using greenhouse whitefly count and also ARIMAX model using whitefly count and environmental data to predict the possible increase in whitefly population. They found that ARIMAX model with temperature and humidity taken as exogenous factors was better than ARIMA model.

Das *et. al.* (2019) proposed several time series parametric regression models viz. Linear model, Quadratic model, Exponential model, Logarithmic model, Power model and ARIMA to investigate the trend of jute production in West Bengal from 1950 to 2016. Based on some performance judgement criteria viz. RMSE, MAPE, Mean Absolute Error (MAE), and R-squared values, ARIMA (1,1,2) produced the most accurate result among other models.

Xiao *et. al.* (2019) carried out Aprioro algorithm to get the information about the association between weather factors and occurrence of cotton pests and found that moderate temperature, humid air, low wind speed and rain fall in autumn and winter were suitable for occurrence of cotton pests and diseases. Also concluded that based on this association Long Short Term Memory Network performed better than traditional machine learning models like SVM and Random Forest.

Almeyda *et. al.* (2020) developed two machine learning models: Logistic regression (LR) and SVM for predicting pest incidence in organic banana crop. It was found that a LR model, when compared to other ML algorithms, had the following advantages like it did not require a large amount of computer capacity for training of the model, simple interpretation, and practical to implement. However, SVM model has a strong mathematical formulation that is able to classify correctly with high accuracy as was tested in this research work.

Aparecido *et. al.* (2020) found that the efficiency of disease and pest alert models could be increased by taking the weather parameters into account. The Random Forest model proved to be the most accurate than neural networks and k-Nearest neighbor's algorithm in terms of the criterion Willmott's 'd', RMSE and R^2 by using weather parameters.

According to Dutta *et. al.* (2020), use of advanced Artificial Intelligence technologies such as Expert System, ANN and Fuzzy logic proved to be more beneficial in diagnosis and management of insect pests. But this progressive technology faced some problems during implementation as it could not make creative responses like human experts and there was also language barrier because most of the farmers were from rural area unable to understand languages of AI applications.

Islam *et. al.* (2020) concluded that meteorological factors like rainfall, temperature and relative humidity (RH) influenced the population of yellow mite. It was also found that prevalent dry conditions having moderately high temperature was favourable for incidence of pests and rainfall had an adverse effect on population density of yellow mite.

Kumar *et. al.* (2021) fitted mustard yield data in ARIMA models and fortnightly weather data in ARIMAX models collected over a period of 1980-81 to 2011-12 and forecasted it for 2012-13 to 2016-17. From the results it was concluded that, on the basis of average absolute percent deviations and RMSEs, ARIMAX models performed better than ARIMA models.

Rainfall being a complex phenomenon, model building and forecasting for it was difficult. So, Lama *et. al.* (2021) fitted parametric models like SARIMA and exponential autoregressive (EXPAR) models along with non-parametric model like time delay neural network (TDNN) model. It was found out that on the basis of RMSE and MAPE, TDNN model was best fitted followed by SARIMA and EXPAR.

Marković *et. al.* (2021) concluded that appearance of some pests such as *Helicoverpa armigera* were affected by environmental conditions such as temperature and relative humidity. They also proposed a Machine Learning model using those easily accessible weather parameters in order to help farmers for effective pest management planning in beforehand.

Ma *et. al.* (2022) concluded that model predictions could be improved by considering the mitigation capacities and responses of invasive crop pests which used microclimates to tackle the impacts of climate change.

3. Materials and Methods

Chapter 3

MATERIALS AND METHODS

In this current chapter, various materials and methodologies used for forecasting models concerning the present investigation are discussed concisely. In the first part, the data sets used in the current study are briefly described. In the subsequent part, the outline of the parametric models that has been utilized for modelling aspects and forecast evaluation methods are found. Accordingly, this chapter is divided into the following subsections,

3.1 Data Description

3.2 Seasonal Plot

3.3 Seasonal Indices

3.4 Seasonality Test

3.5 Two-way Analysis of Variance (ANOVA)

3.6 Correlation Analysis

3.7 ARIMA Model

3.8 ARIMAX Model

3.9 SARIMA Model

3.10 Support Vector Regression (SVR)

3.11 Nonlinearity Test

3.12 Forecast Evaluation Methods

3.1 Data Description

Data on incidence of major pests and diseases of Jute in different location of Terai region are available from 2013 to 2021 under All India Network Project (AINP) on Jute & Allied Fibres, Uttar Banga Krishi Viswavidyalaya (UBKV) Centre. From this data, incidence of major pests of Jute at UBKV, Pundibari, Cooch Behar are used for the present study. Jute Semilooper and Yellow Mite are the major pests that have been considered for the present study. During each year, incidence of pests are available at 25, 35, 45, 55, 65 and 75 days after sowing (DAS). For Yellow Mite, incidence is measured in terms of its count per square cm of 2nd unfold leaf whereas for Semilooper it is measured as percentage infestation.

In addition to this, daily weather data have been collected from Gramin Krishi Mausam Sewa (GKMS), Agrometeorological Field Unit (AMFU), Pundibari, UBKV for the year 2013 to 2021. The weather parameters that have been considered are:

X_1: Rainfall (mm)

X_2: Maximum Temperature (^{0}C)

X_3: Minimum Temperature (^{0}C)

X_4: Maximum Relative Humidity (%)

X_5: Minimum Relative Humidity (%)

The weather data have been considered for Standard Meteorological Weeks (SMWs) by considering the date of survey of pest incidence. For example, during 2015, the survey date of 25 DAS of jute is 05/05/2015 which falls in 18 SMW. Therefore, daily weather data needs to be converted into SMWs weather data, by considering the average of seven days of the SMWs except for the parameter rainfall. For rainfall, total of seven days rainfall is used as SMWs rainfall data.

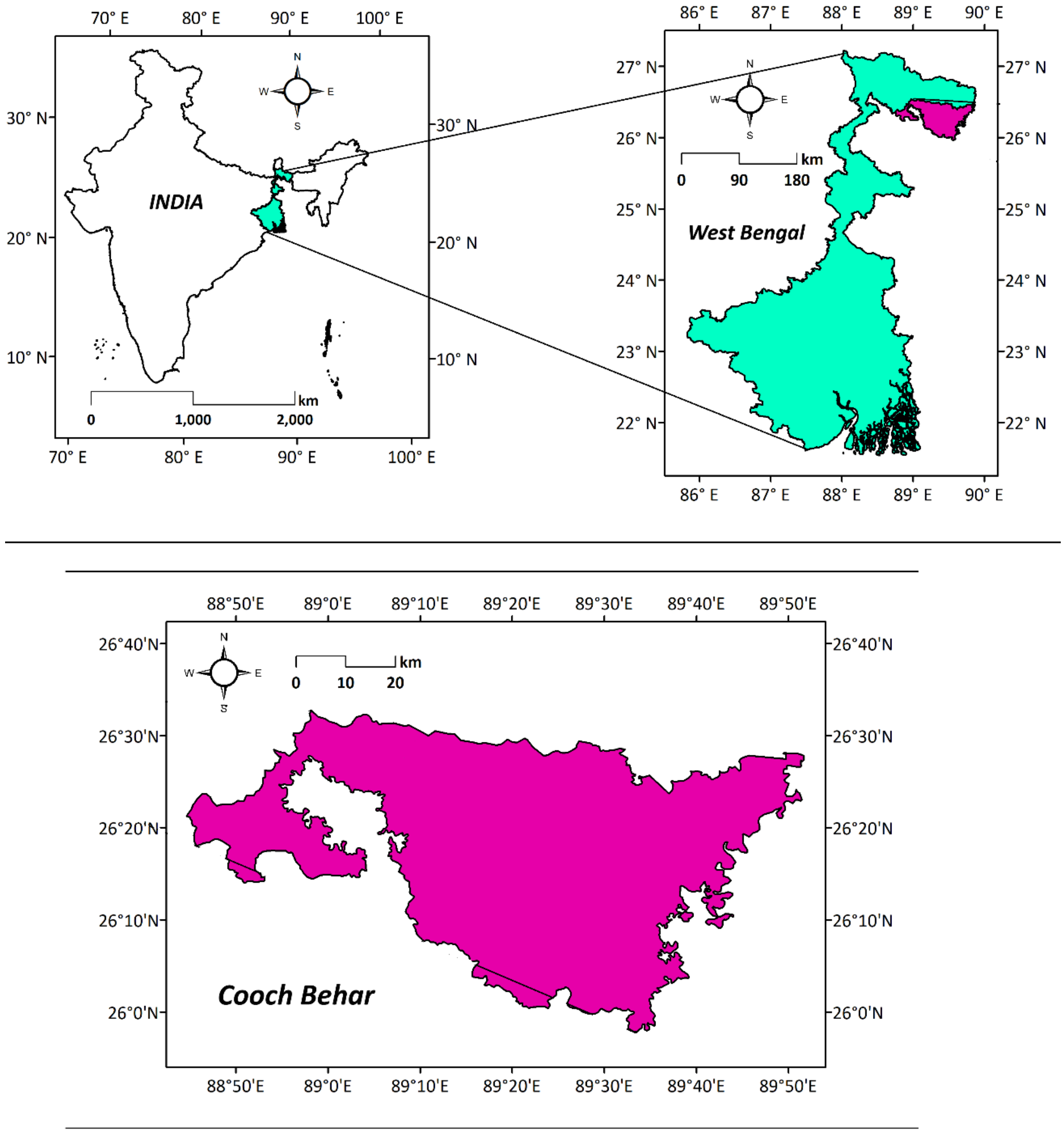

Fig. 3.1.1: Location of Study

During 2013-21, total 54 data points are available for analysis purpose. But it is observed that mostly in every year pest incidence on 25 DAS and 75 DAS is zero. These values may cause anomalies in model fitting and forecasting process. Further, we are mainly interested on pest incidence which is beyond economic threshold level. Therefore, in this current study pest incidence on 25 DAS and 75 DAS are not considered and analysis has been carried out on the remaining 36 data points.

3.2 Seasonal Plot

In time series data any expected variation or trend that occurs repeatedly over a one-year period is considered seasonal. To visualize this seasonality graphically, seasonal plots are developed over a period of time. In contrast to a time plot, a seasonal plot depicts data against the various "seasons" when the data were observed. The data's seasonal incidence can be graphically displayed via the seasonal plot.

3.3 Seasonal Indices

Seasonal indices are also another method to identify seasonality in the data set. Seasonal index is an average value which is used for comparing an actual observation with what it would be in the absence of seasonality. Seasonal index values being more than 1 indicates seasonal variation in the time series data set.

3.4 Seasonality Test

Another way of checking the presence of seasonality is to conduct Webel-Ollech (WO) test. Webel-Ollech (WO) seasonality test integrates the findings of the QS-tests and kwman-tests (Webel and Ollech, 2021). Both tests are performed on the residuals of an automated non-seasonal ARIMA model. Thus, when the p-value of the QS-test and kwman-test is less than 0.01 and 0.002 respectively, the WO-test will confirm presence of seasonality in the associated time series.

3.5 Two-way Analysis of Variance (ANOVA)

The analysis of variance also popularly known as ANOVA, is the organized algebraic procedure of disintegrating the overall variance into different components in order to compare the effects of different treatments. In statistical analyses, the Two-way ANOVA is an extension of the One-way ANOVA. It examines the influence of two different categorical independent variables on one quantitative dependent variable. Here in our present investigation, effect of season (DAS) and Years are considered as two independent factors which may affect the dependent factor i.e., the mean incidence of pest and these effects are examined through Two-way ANOVA. The levels of the significant independent factors are compared through multiple comparison procedure using least significant difference (LSD). Analyses are carried out using SAS software.

3.6 Correlation Analysis

In statistical analyses Pearson Correlation Coefficient is used to determine the degree of association between two variables in a linear fashion. It is generally denoted by '*r*' and expressed as,

$$r_{xy} = \text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \tag{1}$$

where, cov (x, y) = Covariance between variable X and variable Y

σ_x = Standard Deviation of variable X

σ_y = Standard Deviation of variable Y

Pearsonian correlation coefficient is an unitless quantity and its value ranges from -1 to +1. Here in the current study correlation analysis is carried out to know the relationship between incidence of pest (Y) and weather parameters (X) *viz.* total rainfall (mm), maximum and minimum temperature (MaxT & MinT) (^{0}C) and maximum and minimum relative humidity (MaxRH & MinRH) (%) at current week, one and two week lag.

3.7 ARIMA Model

The act of ARIMA modelling gained its popularity from Box and Jenkins 1976, (Box and Jenkins, 1976) and prevalently known as Univariate Box-Jenkins (UBJ) Models due to its use in univariate time series data for analysis and forecasting. There are certain conditions for use of UBJ Models which are discussed afterward. These ARIMA models generally suitable for short-term forecasting are based only on past values of the variable being forecast. Both discrete data as well as continuous data series can be used in these models however, data available at equally spaced discrete time intervals are preferable. To build a good ARIMA model at least 35-40 time series observations and a fairly large sample size is looked for while dealing with seasonal data. The most important requirement is that the concerned time series data should be stationary which means it should have mean, variance and auto-correlation functions constant over time. However, in many practical situation time series data sets are non-stationary in nature which can be transformed to stationary series by simple mathematical operations like differencing. Generally, first difference is enough for creating stationary series.

Theoretically ARIMA model includes three components: Auto-Regressive (AR), Moving-Average (MA), and Integrated (I) terms. The first two components are expressed in equation (2):

$$\nabla^d y_t = \underbrace{\phi_1 \nabla^d y_{t-1} + \ldots + \phi_p \nabla^d y_{t-p}}_{\text{AR terms}} + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}}_{\text{MA terms}} \tag{2}$$

where ϕ is a number strictly between −1 and +1, and θ are the weights, and p is the order of the AR model, and q is the order of the MA model. Here, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 $\forall$ t =1,2, …, n. Usually, the values of p and q lie between 0 and 3. Then again, the I component which is known as differencing is defined in equation (3), with values of d often set into 0, 1, or 2.

$$\begin{cases} \text{if } d = 0,\ \nabla y_t = y_t \\ d = 1,\ \nabla^1 y_t = y_t - y_{t-1} \\ d = 2,\ \nabla^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \end{cases} \tag{3}$$

Alternatively, the model can be written as,

$$\phi_p(B)\nabla^d y_t = \theta_q(B)\varepsilon_t \tag{4}$$

where, $\phi_p(B) = 1 - \phi_1(B) - \phi_2(B^2) - ... - \phi_p(B^p)$,

B is the backshift operator such that $(B^p)\nabla^d y_t = \nabla^d y_{t-p}$

$\nabla^d y_{t-p} = (1-B)^d y_{t-p}$

$\theta_q(B) = 1 - \theta_1(B) - \theta_2(B^2) - ... - \theta_q(B^q)$

$(B^q)\varepsilon_t = \varepsilon_{t-q}$

For finding a good forecasting model UBJ methodology follows mainly three stages, viz. identification, estimation and diagnostic checking, described as follows.

(i) Model Identification Stage

At this stage, two graphical measures such as estimated auto-correlation function (acf) and estimated partial auto-correlation function (pacf) are calculated which describes the statistical relationships within a data series in a somewhat crude way but helps in detecting the underlying pattern in the available data. These functions serve as a guide to selecting one or more ARIMA models that appear to be suitable. Whatever model is chosen in this stage, is only a tentative candidate for the final model.

The primary step in this stage is to check for the stationarity of the series which is done by Augmented-Dickey-Fuller (ADF) and Phillips-Perron (PP) tests. These tests are used to detect the presence of non-seasonal unit root.

D.A. Dickey and W.A. Fuller (1979) have proposed a test for presence of non-seasonal unit root as mentioned below:

$$\nabla^1 y_t = \rho y_{t-1} + \alpha_1 \nabla^1 y_{t-1} + \varepsilon_t \tag{5}$$

where ∇^1 denotes the differencing operator i.e., $\nabla^1 y_t = y_t - y_{t-1}$. The relevant null hypothesis is y_t has a unit root (ρ = 0) i.e., the original series is nonstationary and the alternative hypothesis is ρ < 0 i.e., the original series is stationary. Here the error terms i.e., ε_t follows $N\left(0,\sigma^2\right)$.

Phillips and Perron (1988) developed a test which differs from ADF tests primarily in how it handles serial correlation and error heteroscedasticity. The test regression for the PP test is,

$$\nabla^1 y_t = \beta' D_t + \pi y_{t-1} + u_t \tag{6}$$

where, u_t may be heteroscedastic. Under the null hypothesis H_0: $\pi = 0$, the PP statistic have the same asymptotic distributions as the ADF t-statistic.

Usually, after detecting non-stationarity, differencing is applied to achieve stationarity until the autocorrelation function (acf) and partial autocorrelation function (pacf) shows an interpretable pattern with only a few significant autocorrelations.

(ii) Estimation Stage

The tentative models chosen in the previous stage are specified and precise estimation of the model parameters are carried out on the basis of available data. The parameters are evaluated using the maximum likelihood method in order to reduce the total measure of errors or to maximize the likelihood function. This stage can also be used as a cautionary indicator if the estimated coefficients do not satisfy certain mathematical inequality conditions.

(iii) Diagnostic Checking Stage

This step is basically for diagnostic checking of the errors related to several model assumptions. To choose the actual ARIMA model so that the residuals calculated from this model are white noise, much skill is necessary. This is achieved by performing portmanteau test which determines the model residuals are white noise or not. The null hypothesis tested is that the current set of residuals is white noise.

The Ljung-Box statistic is given as,

$$Q = n(n+2)\sum_{k=1}^{h}(n-k)^{-1}r_k^{\,2} \tag{7}$$

where, h is the maximum lag, n is the number of observations, k is the number of parameters in the model. If the null hypothesis is accepted i.e., the data are white noise, the Ljung-Box Q statistics has a chi- square distribution with (h-k) degrees of freedom.

Selecting the best suitable model

The most suitable ARIMA model is selected using the smallest Akaike Information Criterion (AIC) or Schwarz-Bayesian Criterion (SBC). AIC is given as,

$$AIC = (-2\log L + 2m) \tag{8}$$

where, $m = p + q$ and L is the likelihood function.

SBC can be used as an alternative to AIC, which is given by,

$$SBC = \log\sigma^2 + (m\log n)/n \tag{9}$$

After carrying out diagnostic checking, if the model is not statistically adequate, a new tentative model should be identified, which is again followed by the above discussed stages. These steps of model building process are typically repeated several times until a satisfactory final model is selected. At the end, the final model can be used for forecasting purposes. This iterative nature of the UBJ modeling procedure is the main advantageous fact in application purpose.

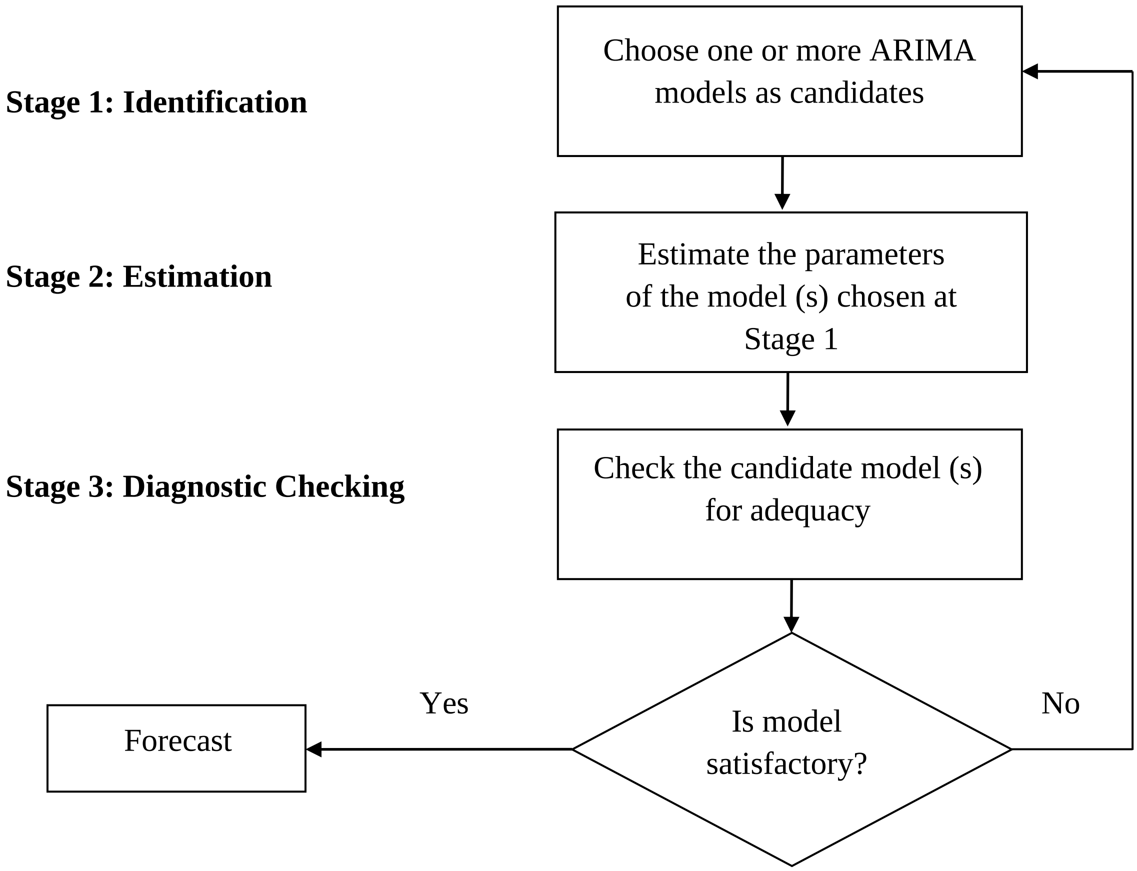

Fig. 3.7.1: Stages in the Box-Jenkins iterative approach to ARIMA model building

3.8 ARIMAX Model

When ARIMA models are insufficient to offer a model with an acceptable level of overall explanatory power, then other possible choices are explored. One of them is ARIMAX model, which is abbreviated as autoregressive integrated moving average with exogenous variables (Bierens, 1987). It is the logical generalization of pure ARIMA model where exogenous independent variable (X) is added to increase the explanatory value of the model. This incorporation of an external input variable transforms the ARIMA model into a multiple regression model. ARIMAX models are preferred over ARIMA models when there is presence of additional variables which have possible influence over the predicted values.

Let's a time series process $\{(y_t, x_t)\}$ having $(k + 1)$ terms, where y_t and k values of x_t are real valued random variables. The ARIMAX model can be formulated as,

$$\nabla^d y_t = \phi_1 \nabla^d y_{t-1} + ... + \phi_p \nabla^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q} + \beta_0 + \beta_1 x_{1t} + ... + \beta_k x_{kt} \quad (10)$$

ε_t 's are the errors.

But here interpreting β is difficult. So, it is expressed as,

$$\nabla^d y_t = \beta_0 + \beta_1 \nabla^d x_{1t} + ... + \beta_k \nabla^d x_{kt} + \nabla^d \eta_t \quad (11)$$

$$\nabla^d \eta_t = \phi_1 \nabla^d \eta_{t-1} + ... + \phi_p \nabla^d \eta_{t-p} + z_t - \theta_1 z_{t-1} - ... - \theta_q z_{t-q} \quad (12)$$

where, η_t = eta at t and

z_t = error.

ARIMAX model also undergoes the similar processes as ARIMA model with an addition of checking stationarity of exogenous variable before modelling. In the next step the transformed variable is added to the ARIMA model, in which the lag length *r* is estimated. Lastly the parameters of ARIMAX model are estimated by the nonlinear least squares estimation process.

3.9 SARIMA Model

In order to improve the performance of conventional ARIMA model seasonal data patterns are added to develop Seasonal ARIMA (SARIMA) model (Box & Jenkins, 1976). The SARIMA model is parsimonious by nature and has a lot of backing in the literature for modelling seasonal series. Generally, SARIMA models are expressed as ARIMA (p, d, q) $(P, D, Q)^s$, where (p, d, q) and $(P, D, Q)^s$ are the non-seasonal and seasonal part of the model, respectively and s is the notation of number of periods per season. The seasonal component of the model is fairly similar to the non-seasonal component, although it is involved in backshifts of the seasonal period. The ARIMA model is finalized by modifying the parameters of p, d, and q using the supplied dataset.

Let us assume a time series y_t (t = 1, 2, …, T) which follows the SARIMA process can be formulated as,

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D\nabla^d y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \tag{13}$$

where, ε_t the residual at time *t* follows $N\left(0,\sigma^2\right)$ and *B* is the backward shift operator.

The polynomials $\phi_p(B)$ and $\theta_q(B)$ represents the non-seasonal autoregressive and moving average terms with orders *p* and *q*, respectively:

$$\phi_p(B) = 1 - \phi_1(B) - \phi_2(B^2) - ... - \phi_p(B^p) \tag{14}$$

$$\theta_q(B) = 1 - \theta_1(B) - \theta_2(B^2) - ... - \theta_q(B^q) \tag{15}$$

Similarly, the seasonal autoregressive and moving average terms of order *P* and *Q*, respectively are represented by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ polynomials:

$$\Phi_P(B^s) = 1 - \Phi_1(B^s) - \Phi_2(B^{2s}) - ... - \Phi_P(B^{Ps}) \tag{16}$$

$$\Theta_Q(B^s) = 1 - \Theta_1(B^s) - \Theta_2(B^{2s}) - ... - \Theta_Q(B^{Qs}) \tag{17}$$

Also, the seasonal and non-seasonal differencing terms are represented by, respectively:

$$\nabla_s^D = (1 - B^s)^D \text{ and} \tag{18}$$

$$\nabla^d = (1 - B)^d \tag{19}$$

SARIMA models also follow the same three stages as conventional ARIMA model which are identification, estimation and diagnostic checking phase. After which forecasting is carried out using the best selected model on the basis of different evaluation criteria which are discussed in the last section of this chapter.

3.10 Support Vector Regression (SVR)

The SVR method is a nonlinear modelling procedure which utilizes the principle of structured risk minimization, in which the upper bound of the generalization error is minimized (Vapnik, 2000). Let $D = \{(x_i, y_i)\}$ (i = 1 to N) be a training set, where $x_i \in R^n$ is input vector, $y_i \in R$ is scalar output and N is the size of the dataset. Then the general representation of nonlinear SVR estimating function is:

$$f(x) = w^T \varphi(x) + b \tag{20}$$

where $\varphi(.) : R^n \rightarrow R^{n_h}$ is a nonlinear mapping function from original input space into a higher dimensional feature space, $w \in R^{n_h}$ is weight vector, b is bias term and the superscript T is transpose.

The coefficients in the equation (20) i.e., w and b are estimated by minimizing the regularized risk function represented below:

$$R(\theta) = \frac{1}{2}\|w\|^2 + C\left[\frac{1}{N}\sum_{i=1}^{N} L_s\big(y_i, f(x_i)\big)\right] \tag{21}$$

where, $\frac{1}{2}\|w\|^2$ is known as regularized term. It measures the flatness of the function, and

$\frac{1}{N}\sum_{i=1}^{N} L_s\big(y_i, f(x_i)\big)$ is known as empirical error.

This above-mentioned regularized risk function employs the structural risk minimization (SRM) principle to prevent under and over fitting of training data and simultaneously minimizes both empirical error and regularized term. The empirical error term present in the equation (21) is estimated by Vapnik ε-insensitive loss function which represents the radius of the tube of accuracy located around the regression function given by:

where, y_i is the actual value

$f(x_i)$ is the estimated value.

C is known as regularization factor.

ε is called tube size which is also the approximation accuracy in training data and both C and ε are hyper-parameters determined by users.

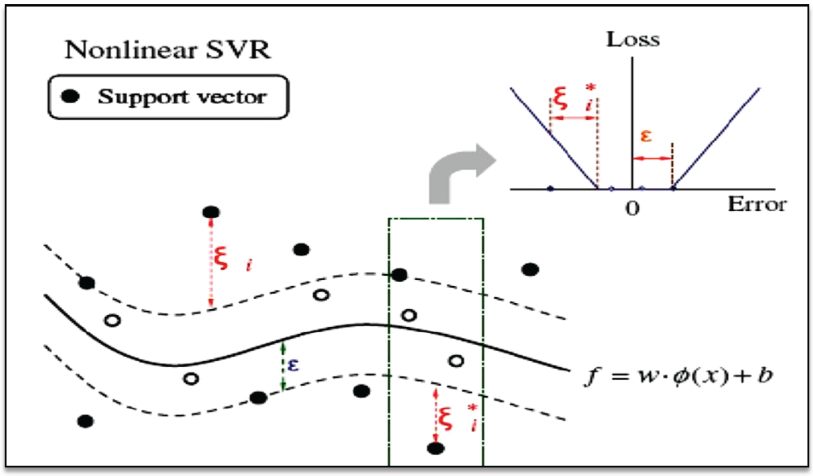

Fig. 3.10.1: A schematic representation of Vapnik ε -insensitive loss function and accuracy tube under nonlinear SVR model setup

In the above loss function the data points situated within the tube have no loss as these are less than ε and for this reason they are not used in any decision-making process. The points on or outside the boundary lines are known as support vectors and these are accountable for the loss. The ε -insensitive loss function is the cause behind this kind of sparseness algorithm which simplifies the calculation of nonlinear SVR. Here ξ_i and ξ_i^* are two positive slack variables which represents the distance between actual values and respective boundary lines of the tube. Using this input equation (21) can be reformulated as;

$$R_P(w,b,\xi_i,\xi_i^*) = \frac{1}{2}\|w\|^2 + C\left[\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right] \tag{22}$$

With the ε -insensitive function the SVR, the minimization of $\frac{1}{2}\|w\|^2 + C\left[\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right]$ is carried out subject to

$$w^T\varphi(x_i) + b - y_i \le \varepsilon + \xi_i^*,$$

$$y_i - w^T\varphi(x_i) - b \le \varepsilon + \xi_i^* \text{ and} \tag{23}$$

$$\xi_i, \xi_i^* \ge 0; \quad i = 1, 2, \ldots, N$$

In SVR kernels can be used to accomplish nonlinear mapping in higher dimensions. Then the regression function becomes:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{24}$$

where α_i and α_i^* are lagrange multipliers and

$k(x_i, x)$ is a kernel function.

3.11 Nonlinearity Test

In a time-series data to check the occurrence of nonlinearity pattern, Brock-Dechert-Scheinkman (BDS) test is used widely. To detect the nonlinear structure in a time series it utilizes the correlation dimension. This test can also be used to check how good the fit estimation model is. Generally, in forecasting the available time-series data is made stationary by differencing as per need and then suitable linear model (e.g., ARIMA (p, d, q), exponential smoothing etc.) is fitted to get the linear component from the series. Then by using the BDS test nonlinearity pattern is checked on the extracted residuals of the formerly fitted linear model. In this test the null hypothesis i.e., the residuals are independent and identically distributed (i.i.d.) is established against the alternate hypothesis that there is presence of hidden nonlinearity, hidden nonstationarity or other structure in the residuals. It is also preferable over conventional nonparametric tests used for residual analysis. So, for our current investigation BDS test is carried out using R software.

3.12 Forecast Evaluation Methods

In many scholarly investigations, the models that perform best for within-sample data are unlikely to predict better out-of-sample results. So, the data is divided into two sections: first one for sample period forecasts to generate confidence in the model and the other part for post-sample period forecasts to use in planning and other purposes. However, it is advisable to choose the most parsimonious model based on the first n observations and evaluate them with m holdout data. For the present investigation following error range indexes have been used to validate the prediction performances of different models and any model fulfilling most of the following criteria is selected.

(i) Root Mean Squared Error (RMSE)

RMSE is the positive square root of Mean squared error (MSE). It measures how much a dependent series differs from its model-predicted level, expressed in the same units as the dependent series. RMSE has the following expression,

$$RMSE = \sqrt{\left(1/m \right)\left[\sum_{h=1}^{m} \left(\hat{y}_{n+h} - y_{n+h} \right)^2 \right]} \tag{25}$$

(ii) Root Median Squared Error (RMdSE)

Outliers are troublesome when it comes to select best model among set of forecasting methods. So RMdSE is used to determine the robustness of a model against outliers. RMdSE has the following expression,

$$RMSE = \sqrt{Median\left(\hat{y}_{n+h} - y_{n+h} \right)^2} \tag{26}$$

(iii) Diebold-Mariano (DM) test

DM test can be used to test if the two models have significantly different Mean Squared Percentage Errors (MSPEs) or MSEs. Let two models A and B have MSPEs as $MSPE_A$ and $MSPE_B$ respectively. For testing, a new variable *dj* (j = 1,2, …, m) is formed, such that *dj* =1, if $MSPE_A > MSPE_B$ and 0 otherwise. Then DM test statistic is given as,

$$DM = \left({}^{m}\!/_{4} \right)^{1/2} \left[\sum_{j=1}^{m} d_j - \left({}^{m}\!/_{2} \right) \right] \sim N(0,1) \tag{27}$$

If the test is significant then it indicates that the MSPE or MSE of the two model are significantly different.

4. Results and Discussion

Chapter 4

RESULTS AND DISCUSSION

Current chapter comprises the highlights of the results after carrying out analyses on the available data set for the present study concerning the previously discussed objectives. For the investigation 36 data points consisting of 4 seasons are considered out of which initial 32 data points are used for model building purpose and rest 4 data points are used for validation purpose. The results have been analyzed statistically by using R-Studio Version 4.1.2 and SAS 9.2.

4.1 Descriptive Statistics

The descriptive statistics of two major pests of Jute i.e., Yellow Mite and Semi Looper are highlighted in Table 4.1.1. A perusal of Table 4.1.1 indicates that average incidence of yellow mite is 4.31 per cm^2 and semi looper is 4.70 % with standard deviation being 5.74 and 5.05 respectively. The variability measured in terms of coefficient of variation (CV) is found to be 133% and 108% for yellow mite and semi looper respectively. So, from this result it can be concluded that there is high variability in terms of both insect pests.

For both yellow mite and semi looper the measure of skewness is found to be positive in nature as it is 1.81 and 1.12 respectively. Also, the former series is leptokurtic and later series is platykurtic in nature as the values are 3.36 and 0.40 respectively.

Further Descriptive Statistics of all the independent variables at current week, lag1 and lag2 are represented in Table 4.1.2, 4.1.3 and 4.1.4 respectively.

Table 4.1.1 Descriptive statistics of Incidence of Major Jute Pests

Parameters	Yellow Mite (no/sq cm)	Semi Looper (% Infestation)
Mean	4.31	4.70
Standard Deviation	5.74	5.05
Coefficient of Variation (CV)	1.33	1.08
Minimum	0.00	0.00
Maximum	25.62	17.87
Skewness	1.81	1.12
Kurtosis	3.36	0.40

Table 4.1.2 Descriptive statistics of Weather Variables at Current Week

Parameters	Rainfall (mm)	MaxT (^{0}C)	MinT (^{0}C)	MaxRH (%)	MinRH (%)
Mean	113.49	31.47	22.33	85.15	74.67
Standard Deviation	126.40	1.97	2.30	10.30	9.79
Coefficient of Variation (CV)	1.11	0.06	0.10	0.12	0.13
Minimum	0.00	28.29	13.00	46.43	37.86
Maximum	533.90	36.86	25.29	99.14	92.57
Skewness	1.77	0.67	-1.80	-1.73	-1.19
Kurtosis	2.45	-0.11	5.13	4.29	3.60

Table 4.1.3 Descriptive statistics of Weather Variables at One week lag

Parameters	Rainfall Lag1 (mm)	MaxT Lag1 (^{0}C)	MinT Lag1 (^{0}C)	MaxRH Lag1 (%)	MinRH Lag1 (%)
Mean	87.11	31.86	21.98	82.12	71.36
Standard Deviation	93.86	1.82	2.18	11.88	11.48
Coefficient of Variation (CV)	1.08	0.06	0.10	0.14	0.16
Minimum	0.00	28.93	17.41	45.00	35.00
Maximum	404.40	36.86	25.29	97.71	92.57
Skewness	1.68	0.67	-0.29	-1.63	-1.19
Kurtosis	2.43	0.16	-0.94	2.68	2.62

Table 4.1.4 Descriptive statistics of Weather Variables at Two weeks lag

Parameters	Rainfall Lag2 (mm)	MaxT Lag2 (^{0}C)	MinT Lag2 (^{0}C)	MaxRH Lag2 (%)	MinRH Lag2 (%)
Mean	78.83	31.49	21.55	81.85	70.63
Standard Deviation	100.60	1.88	2.06	11.16	11.38
Coefficient of Variation (CV)	1.28	0.06	0.10	0.14	0.16
Minimum	0.00	28.76	17.40	45.00	35.00
Maximum	533.90	36.86	25.01	97.00	88.57
Skewness	2.86	0.67	-0.23	-1.75	-1.36
Kurtosis	9.53	0.28	-0.74	3.46	2.31

4.2 Seasonal Plot

In this portion of the current chapter, seasonal aspects like seasonal plots and seasonal indices are illustrated graphically and discussed thoroughly. To test the hypothesis of seasonality, the data points on 35, 45, 55 and 65 DAS have been plotted against the mean incidence of Yellow Mite and Semi Looper respectively from 2013-2021.

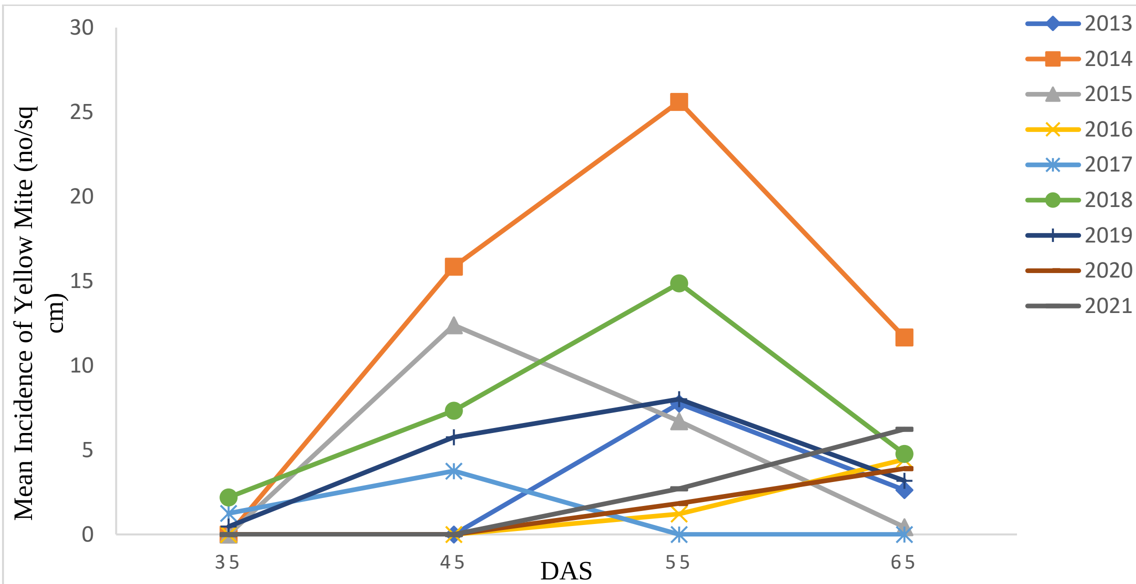

Fig. 4.2.1: Seasonal Plot of Yellow Mite incidence

From Fig. 4.2.1 it can be concluded that the peak incidence in yellow mite is observed during 55 DAS in 2013, 2014, 2018 and 2019. Also, for 2015 and 2017 the peak is on 45 DAS and for 2016, 2020 and 2021 it is on 65 DAS. So, from the above figure presence of seasonality is ambiguous.

. From Fig. 4.2.2 it was evident that the peak incidence in semi looper is detected during 45 DAS in recent years like 2015 to 2020. For earlier years like 2013 and for 2014 and 2021 this is attained slight later i.e., on 65 DAS and 55 DAS respectively. Also, from the figure it can be concluded that seasonality is present in this data set.

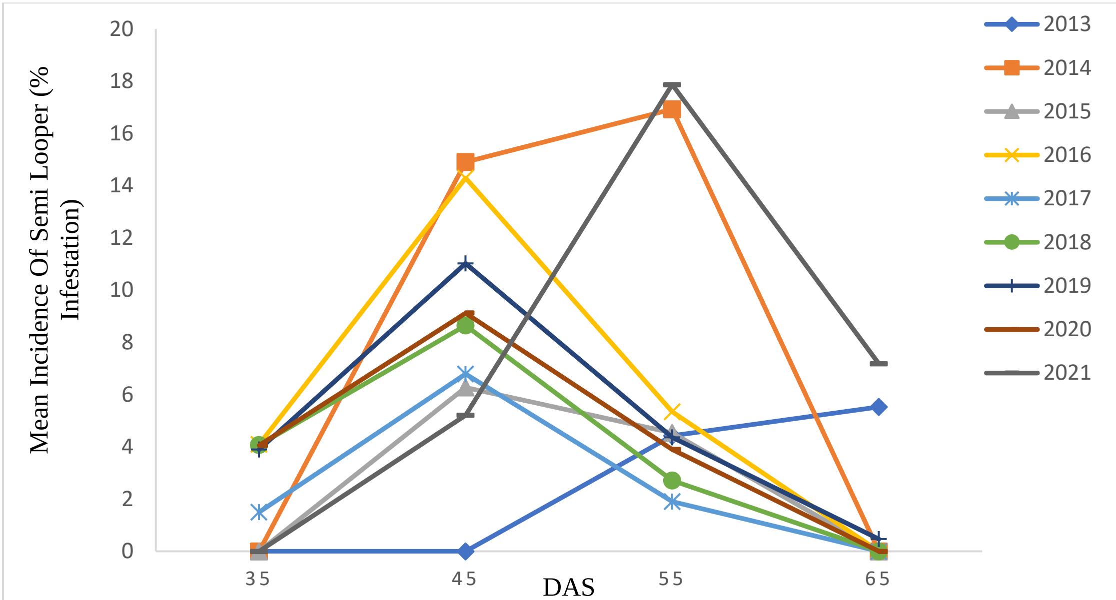

Fig. 4.2.2: Seasonal Plot of Semi Looper incidence

4.3 Seasonality Test

To check the presence of seasonality, WO test has been conducted and the results are noted in Table 4.3.1 and 4.3.2 for Yellow Mite and Semi Looper respectively.

Table 4.3.1 WO test for Yellow Mite

Test Statistic		p-value	
0	0.111	1	0.078

Table 4.3.2 WO test for Semi Looper

Test Statistic		p-value	
1	0.0001	0.001	0.005

From Table 4.3.1 it is clear that seasonality is not present as the p-value of the combined test is above 0.05. Thus, for Yellow Mite seasonality is not present at 5% level of significance, but from Table 4.3.2 it can be observed that p-value of the combined test is below 0.05. Therefore, seasonality is present for Semi Looper at 5% level of significance.

4.4 Two-way ANOVA

To study the variation in pest incidence across seasons (DAS) and years, two-way analysis of variance ANOVA technique has been implemented while seasons and years are considered as the sources of variation.

From Table 4.4.1 it is clearly depictable that annual variation in yellow mite incidence is highly significant i.e., at 1% level and seasonal variation is also significant at 5% level.

Table 4.4.1 Two-way ANOVA using Yellow Mite Incidence as dependent variable

Source of Variation	DF	Sum of Squares	Mean Square	F Value	p-value
DAS	3	239.17	79.72	4.53	0.012*
Year	8	492.35	61.54	3.49	0.008**
Error	24	422.65	17.61		
Total	35	1154.18			

**: Significant at 1%; *: Significant at 5%

Table 4.4.2 Grouping of Seasons based on pair-wise comparison of Yellow Mite incidence

DAS	Mean
55	7.63^{A}
45	5.01^{A}
65	4.14^{AB}
35	0.44^{B}

*: Means with same letter are not significantly different

It can be observed from Table 4.4.2 that mean incidence in 55 DAS is highest but it is at par with 45 and 65 DAS, as the said 3 seasons belong to group A, but 65 DAS belonging to group A is also at par with 35 DAS, while both 65 DAS and 35 DAS belong to group B.

Table 4.4.3 Grouping of Years based on pair-wise comparison of Yellow Mite incidence

Year	Mean
2014	13.29^{A}
2018	7.30^{AB}
2015	4.88^{B}
2019	4.36^{B}
2013	2.60^{B}
2021	2.24^{B}
2020	1.43^{B}
2016	1.42^{B}
2017	1.26^{B}

*: Means with same letter are not significantly different

From Table 4.4.3 it is clearly evident that mean incidence of yellow mite in 2014 is highest among all i.e., 13.29 and it is only at par with 2018. So, both mean incidence in 2014 and 2018 belong to group A. But incidence in 2018 is also at par with other years such as 2015, 2019, 2013, 2021, 2020, 2016 and 2017. Thus, all these years except 2014, belong to group B.

Table 4.4.4 Two-way ANOVA using Semi Looper Incidence as dependent variable

Source of Variation	DF	Sum of Squares	Mean Square	F Value	p-value
DAS	3	333.08	111.03	6.32	0.01**
Year	8	139.34	17.42	0.99	0.47
Error	24	421.90	17.58		
Total	35	894.33			

**: Significant at 1%

Similarly in case of semi looper incidence only seasonal variation is highly significant i.e., at 1% level but annual variation is not significant at all, which can be observed in Table 4.4.4.

Table 4.4.5 Grouping of Seasons based on pair-wise comparison of Semi Looper incidence

DAS	Mean
45	8.47^{A}
55	6.89^{A}
35	1.96^{B}
65	1.46^{B}

*: Means with same letter are not significantly different

From Table 4.4.5 it is evident that 45 DAS has the maximum mean incidence of semi looper and it is also at par with 55 DAS i.e., both 45 and 55 DAS belong to group A. But mean incidence in 35 DAS and 65 DAS belong to group B and it has comparatively less mean incidence than group A.

Table 4.4.6 Grouping of Years based on pair-wise comparison of Semi Looper incidence

Year	Mean
2014	7.96^{A}
2021	7.57^{A}
2016	5.94^{A}
2019	4.94^{A}
2020	4.27^{A}
2018	3.86^{A}
2015	2.70^{A}
2017	2.55^{A}
2013	2.49^{A}

*: Means with same letter are not significantly different

From Table 4.4.6 it is evident that mean incidence of all the years belong to one group, i.e., there is no significant difference between mean incidence of different year and mean incidence of 2014 being highest among all the years, which is 7.96.

4.5 Correlation Analysis

Further to analyze the relationship between pest incidence and weather parameters, Pearson's correlation coefficients are conducted at current, lag1 and lag 2 week and the corresponding results are depicted in Table 4.5.1 and 4.5.2.

Table 4.5.1 Pearson's correlation coefficient between Mean incidence of Yellow Mite and weather parameters

Weather Parameters	Current Week	One Week Lag	Two Weeks Lag
Rainfall	-0.24	-0.24	-0.19
MaxT	-0.15	0.21	0.34
MinT	**-0.37***	-0.34	**-0.37***
MaxRH	-0.28	**-0.43****	**-0.61****
MinRH	-0.11	-0.3	**-0.51****

**: Significant at 1%; *: Significant at 5%

From Table 4.5.1 it can be perceived that mean incidence of yellow mite has a significant negative association with MinT in current week and MaxRH at lag2. While correlation between MaxRH in lag1, MaxRH and MinRH in lag 2 with mean pest incidence is highly significant in a negative direction.

Table 4.5.2 Pearson's correlation coefficient between Mean incidence of Semi Looper and weather parameters

Weather Parameters	Current Week	One Week Lag	Two Weeks Lag
Rainfall	-0.07	-0.12	-0.16
MaxT	-0.12	0.25	**0.39***
MinT	-0.01	-0.07	-0.24
MaxRH	-0.17	-0.33	**-0.49****
MinRH	-0.02	-0.23	**-0.44****

**: Significant at 1%; *: Significant at 5%

From Table 4.5.2 it can be depicted that MaxT at lag 2 is significantly positively correlated with mean incidence of semi looper. But at lag 2 MaxRH and MinRH are negatively correlated with mean incidence of semi looper and the association is highly significant.

4.6 Fitting of Different Models of Yellow Mite

In the present investigation ARIMA, ARIMAX and SVR models are considered for modeling of incidence of yellow mite. Out of 36 data points initial 32 data points have been used for model building purpose as training data set and rest 4 data points have been used as testing data set for validating the proposed models.

The exogenous variables having significant association with incidence of yellow mite are denoted in Table 4.6.1 with respective correlation coefficients and p-values. In Table 4.6.2 the Variance Inflation Factor (VIF) values are noted for the correlated exogenous variables and variables having VIF values more than 5 are eliminated as it indicates presence of moderate multicollinearity. Thus, remaining three variables are considered for analysis purpose and the best pair among different combinations has been chosen for final analysis in ARIMAX and SVR model.

Table 4.6.1 Correlated Exogenous variables with correlation coefficient and p-values for Yellow Mite

Weather Parameters	r-value	p-value
MinT	-0.37	0.03*
MaxRH lag1	-0.43	0.01**
MinT lag2	-0.37	0.04*
MaxRH lag2	-0.61	0.00**
MinRH lag2	-0.51	0.00**

**: Significant at 1%; *: Significant at 5%

Table 4.6.2 VIF values of correlated exogenous variables for Yellow Mite

Variables	MinT	MaxRH lag1	MinT lag2	MaxRH lag2	MinRH lag2
VIF Value	1.55	3.54	3.11	8.67*	5.92*

*: Moderate Multicollinearity

4.6.1 Test for Stationarity

To check the presence of stationarity in the current data series Augmented-Dickey-Fuller (ADF) test and Phillips-Perron (PP) tests have been applied and from Table 4.6.3 it is evident that p value of ADF test is 0.01 i.e., it is highly significant. Thus, the null hypothesis in ADF test is rejected at 1% significance level, thus indicating the data series is stationary. Similarly in PP test also p-value being 0.05, indicates that the data series is stationary at 5% level of significance. It indicates there is no requirement of regular differencing. It is also evident from time plot in Fig. 4.6.1, as there is not much presence of trend in the time series data. Thus, order of d is 0 for ARIMA model.

Table 4.6.3 ADF and PP test for stationarity

ADF Test		PP Test	
Test Statistic	**p-value**	**Test Statistic**	**p-value**
-3.59	0.01**	-3.61	0.05*

**: Significant at 1%; *: Significant at 5%

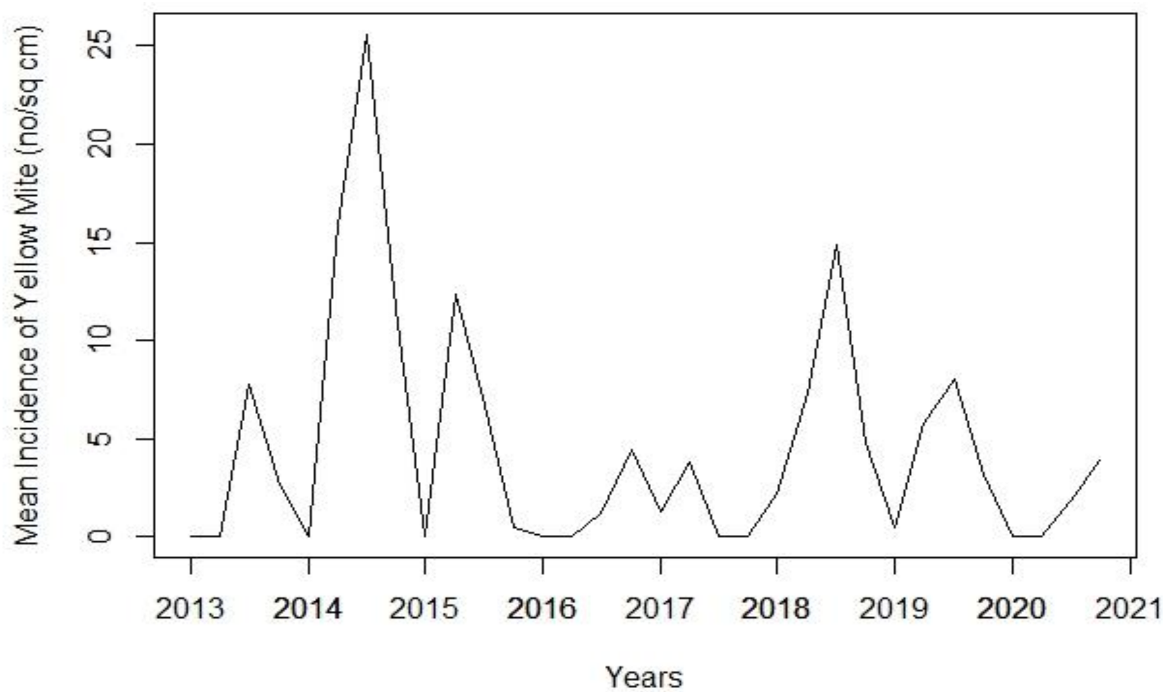

Fig. 4.6.1: Time Series Plot of Yellow Mite incidence

The calculated acf and pacf values are plotted and represented as Fig. 4.6.2 and Fig. 4.6.3 and since acf and pacf values are not significant at seasonal lags i.e., at lag 4 and lag 8 etc., thus, it can be concluded that seasonality is not present. So, there is no need of any seasonal differencing and thus, order of D is 0.

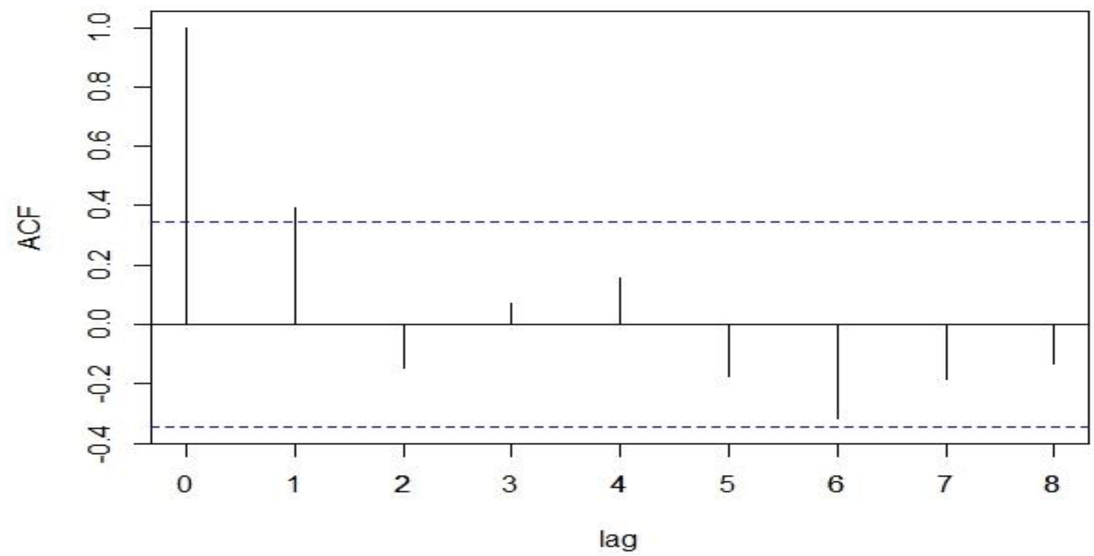

Fig. 4.6.2: Plot showing ACF values of Yellow Mite incidence

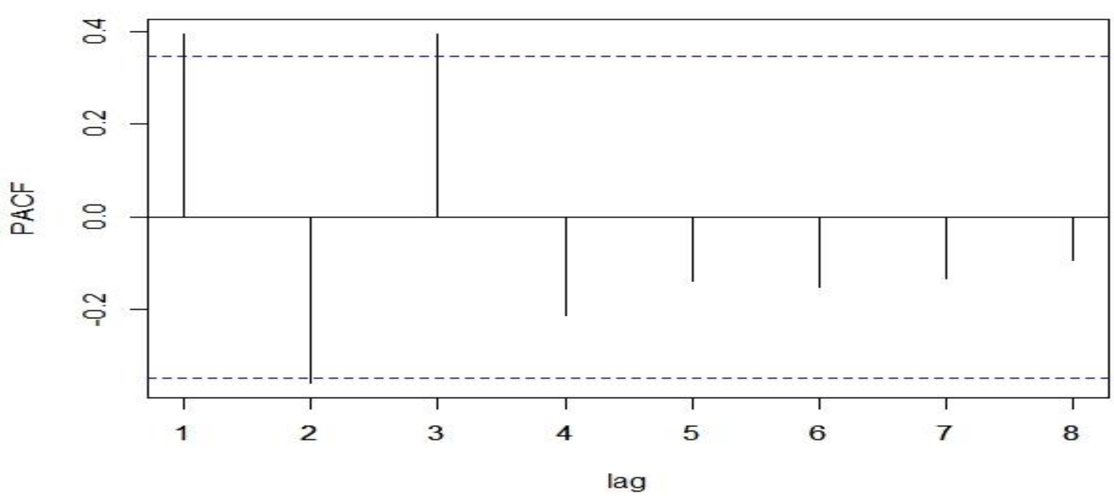

Fig. 4.6.3: Plot showing PACF values of Yellow Mite incidence

Order of regular differencing and seasonal differencing i.e., d and D can also be determined by using “ndiffs” and “nsdiffs” function in R software. The results obtained in R software also indicates that values of d and D are also 0.

4.6.2 ACF and PACF values

To determine the possible order of AR (p) and MA (q) of the ARIMA model, acf and pacf values are calculated based on the original series as the value of d and D is 0. These values are noted in table 4.6.4 and plots can be seen in Fig. 4.6.2 and 4.6.3.

Table 4.6.4 ACF and PACF values at lags up to 8

Lag	ACF	PACF
1	0.40	0.40
2	-0.15	-0.36
3	0.07	0.39
4	0.15	-0.21
5	-0.18	-0.14
6	-0.32	-0.15
7	-0.18	-0.13
8	-0.13	-0.09

From Fig. 4.6.2 it can be observed that acf value is significant at lag 1, so the order of q maybe 1. But from Fig. 4.6.3 pacf values are significant at lag 1, lag 2 and lag 3, so the order of p can’t be determined clearly. So, from the above details tentative order of the ARIMA model may be (0, 0, 1).

4.6.3 Fitting of ARIMA model

After confirmation of stationarity by ADF and PP test earlier, ARIMA model building has been carried out using acf and pacf plots, and suitable model has been selected on the basis of minimum AIC and BIC criteria. Accordingly, ARIMA (0, 0, 1) model is selected using “auto.arima” function in R software. The estimate of parameters, its standard error (S.E.) and respective p-values are presented in Table 4.6.5.

Table 4.6.5 Parameter Estimates of the ARIMA (0, 0, 1) model for Yellow Mite incidence

Model	Parameters	Estimate	S.E.	p-value
ARIMA (0, 0, 1)	**C**	4.55	1.43	0.002**
	MA1	0.72	0.11	0.001***

***: Significant at 0.1%; **: Significant at 1%.

RMSE and RMdSE of the fitted ARIMA (0, 0, 1) model on the training data set found to be 4.78 and 2.62 respectively.

4.6.4 Fitting of ARIMAX model

As an extension of ARIMA model, here in ARIMAX model independent variables are included additionally. In present study maximum RH at lag1 and minimum temperature at lag2 are considered as two exogenous variables as proved to be the best possible pair and then model building process has been carried out. On the basis of minimum AIC and BIC value ARIMAX (0, 0, 1) model is selected as best fitted model using "auto.arima" function in R software and estimated parameters with S.E. and p-values are presented in Table 4.6.6.

Table 4.6.6 Parameter Estimates of the ARIMAX (0, 0, 1) model for Yellow Mite incidence

Model	Parameters	Estimate	S.E.	p-value
ARIMAX (0, 0, 1)	MaxRH lag1	-0.12	0.13	0.350
	MinT lag2	0.64	0.48	0.179
	MA1	0.68	0.13	0.001***

***: Significant at 0.1%.

RMSE and RMdSE of the fitted ARIMAX (0, 0, 1) model on the training data set is 4.77 and 2.83 respectively.

4.6.5 Fitting of SVR model

As it is evident from earlier sections that seasonality is not present in the current data set. Therefore, there is no need of any seasonal adjustment in the data set for application of SVR methodology. Further, from Section 4.6.2 and Section 4.6.4 it is also clear that there is no AR component present in the possible ARIMA model. So, there is no justification for fitting regression models of pest incidence with itself. Thus, the best fitted model is selected on the basis of lowest training and testing error, which is found to be $y \sim x$ and the parameters are represented in Table 4.6.7.

Table 4.6.7 Parameters of the SVR (*y* ~ *x*) model for Yellow Mite incidence

Type	Kernel	Cost (C)	Gamma	Epsilon (ε)	No. of Support Vectors
eps-regression	radial	1	0.5	0.1	30

RMSE and RMdSE of the fitted SVR (y ~ x) model on the training data set is 5.24 and 2.52 respectively.

4.6.6 Model Validation

In order to compare the accuracy of different models undertaken, different criterions can be used. Here, RMSE and RMdSE are used to compare the forecast performance of different models on the basis of analysis carried on testing data set and have been furnished in Table 4.6.8. From the table it is observed that ARIMAX model produces the least RMSE value as compared to others. It is then followed by SVR model which has RMSE value more than ARIMAX model but less than ARIMA model. At last, ARIMA model is there having the highest RMSE values among all the models. The same can also be confirmed by observing the RMdSE values.

Table 4.6.8 Predictive Abilities for ARIMA, ARIMAX and SVR models for Yellow Mite

Model	Parameter	RMSE	RMdSE
ARIMA	(0,0,1)	3.41	3.39
ARIMAX	(0,0,1)	1.86	1.61
SVR (y ~ x)	eps-regression, radial	2.08	2.19

Evident from several literatures published earlier, SVR model should be the best fitted out of all the three models considered in the current investigation. But it has also been seen that SVR model performs better in non-linear data set as compared to linear data set, but results obtained from ARCH-LM test confirm that the residuals are homoscedastic in nature. Thus, the data set is stationary in mean and variance. Therefore, SVR model has not given better result as compared to the ARIMAX model.

4.6.7 Residual Diagnostics

After getting the best model as ARIMAX model, to check the appropriateness of the fitted model, residual diagnostics has been carried out. After building the appropriate model, Box-Ljung test is carried out on the residuals to check whether the residuals are autocorrelated or not and then results are presented in Table 4.6.9. As the p-value is more than 0.05, the null hypothesis is not rejected and thus it can be concluded that the residuals are independent.

Table 4.6.9 Box-Ljung Test for Yellow Mite incidence

Test Statistic	p-value
6.75	0.56

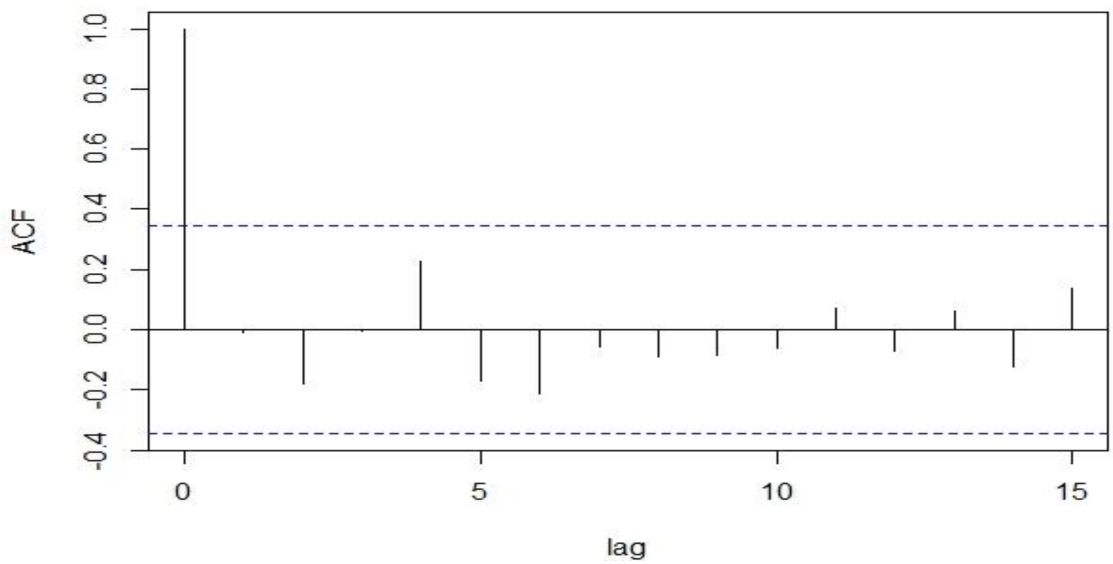

Fig. 4.6.4: Plot showing ACF values of residuals of ARIMAX model of Yellow Mite incidence

The same can also be proved from Fig. 4.6.4 as the acf values of the residuals are non-significant in nature indicating that the residuals are independent.

In order to check the normality of residuals, Shapiro-Wilk test has been implemented and the results are shown in Table 4.6.10. As the p-value is less than 0.01 i.e., highly significant, the null hypothesis is rejected and thus it can be concluded that the residuals are not normally distributed.

Table 4.6.10 Shapiro-Wilk Test for Yellow Mite incidence

Test Statistic	p-value
0.87	0.001**

**: Significant at 1%

To check for the presence of ARCH effect i.e., to test the presence for conditional heteroscedasticity, ARCH LM test has been carried out. From Table 4.6.11 it can be depicted that the null hypothesis could not be rejected as the p-value is greater than 0.05. Thus, it can be concluded that up to lag 8 there is no existing ARCH effect in residuals, i.e., residuals are homoscedastic in nature. The same can also be confirmed from the shape of the plot in Fig. 4.6.5, as there is constant variance.

Table 4.6.11 ARCH LM Test for Yellow Mite incidence

Test Statistic	p-value
9.62	0.29

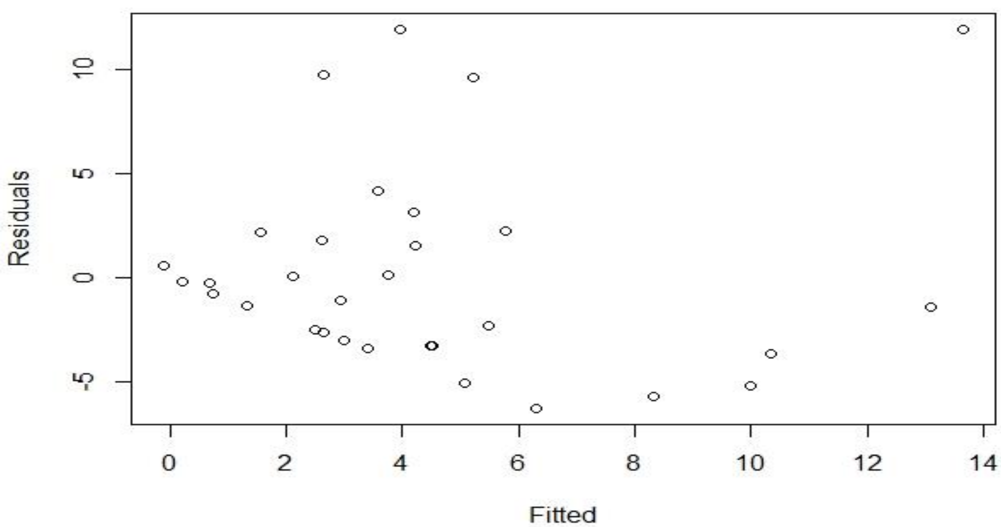

Fig. 4.6.5: Plot showing Fitted Values vs. Residuals of ARIMAX model for Yellow Mite incidence

To test the presence of any remaining component in the residuals after fitting the linear ARIMAX model, i.e., to check the presence of non-linearity in the residuals BDS test has been carried out and the results are represented in Table 4.6.12. As from the results it is clearly evident that most of the epsilons are non-significant in nature, except the first one. So, it can be concluded that there is absence of any non-linear pattern which implies that hybrid models can't be implemented further.

Table 4.6.12 BDS Test for Yellow Mite incidence

Dimension (m)	Epsilon (ε)		Test Statistic	p-value
2	eps (1)	2.42	2.01	0.04*
	eps (2)	4.85	1.45	0.15
	eps (3)	7.27	0.90	0.37
	eps (4)	9.70	1.31	0.19
3	eps (1)	2.42	-0.85	0.40
	eps (2)	4.85	0.27	0.78
	eps (3)	7.27	0.19	0.85
	eps (4)	9.70	1.24	0.21

*: Significant at 5%

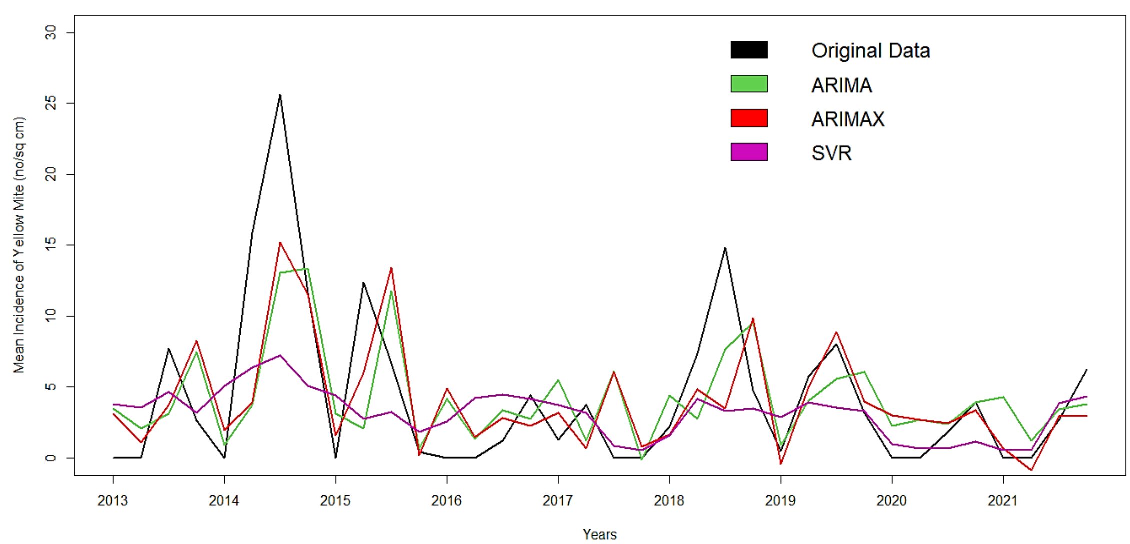

Fig. 4.6.6: Plot showing original vs fitted values by ARIMA, ARIMAX and SVR model for Yellow Mite incidence

Fig 4.6.6 represents the plots of original values versus fitted values by different models like ARIMA, ARIMAX and SVR.

4.7 Fitting of Different Models of Semi Looper

Just like Yellow Mite, in case of Semi Looper also SARIMA, SARIMAX and SVR models have been used for model building and forecasting purpose using the initial 32 data points as training data set and remaining 4 data points as testing data set.

In Table 4.7.1 significant relationship of different exogenous variables with semi looper incidence has been noted down with respective correlation coefficients and p-values. VIF values of different exogenous variables are noted in Table 4.7.2 and as the VIF value of MinRH at lag2 indicates moderate multicollinearity, so it is eliminated and rest two variables have been chosen for final analysis in ARIMAX and SVR model.

Table 4.7.1 Correlated Exogenous variables with correlation coefficient and p values for Semi Looper

Weather Parameters	r Value	p-Value
MaxT lag2	0.39	0.03*
MaxRH lag2	-0.49	0.00**
MinRH lag2	-0.44	0.01**

**: Significant at 1%; *: Significant at 5%

Table 4.7.2 VIF values of correlated exogenous variables for Semi Looper

Variables	MaxT lag2	MaxRH lag2	MinRH lag2
VIF Value	1.72	5.66	6.60*

*: Moderate Multicollinearity

4.7.1 Test for Stationarity

In order to check stationarity of the data series ADF and PP tests have been applied on the data set and the results have been noted down in Table 4.7.3. From the table it can be concluded that for both ADF and PP test the p-values are 0.01 i.e., highly significant. So, the null hypothesis is rejected and it can be noted that the given data set is stationary in nature. It implies that regular differencing is not required here. It is also evident from time plot in Fig. 4.7.1, as there is not much presence of trend in the time series data. Thus, order of d is 0 for ARIMA model. The same can also be confirmed after using "ndiffs" function in R software as the result is 0.

Table 4.7.3 ADF and PP test for stationarity

ADF Test		**PP Test**	
Test Statistic	**p-value**	**Test Statistic**	**p-value**
-4.86	0.01**	-4.76	0.01**

**: Significant at 1%

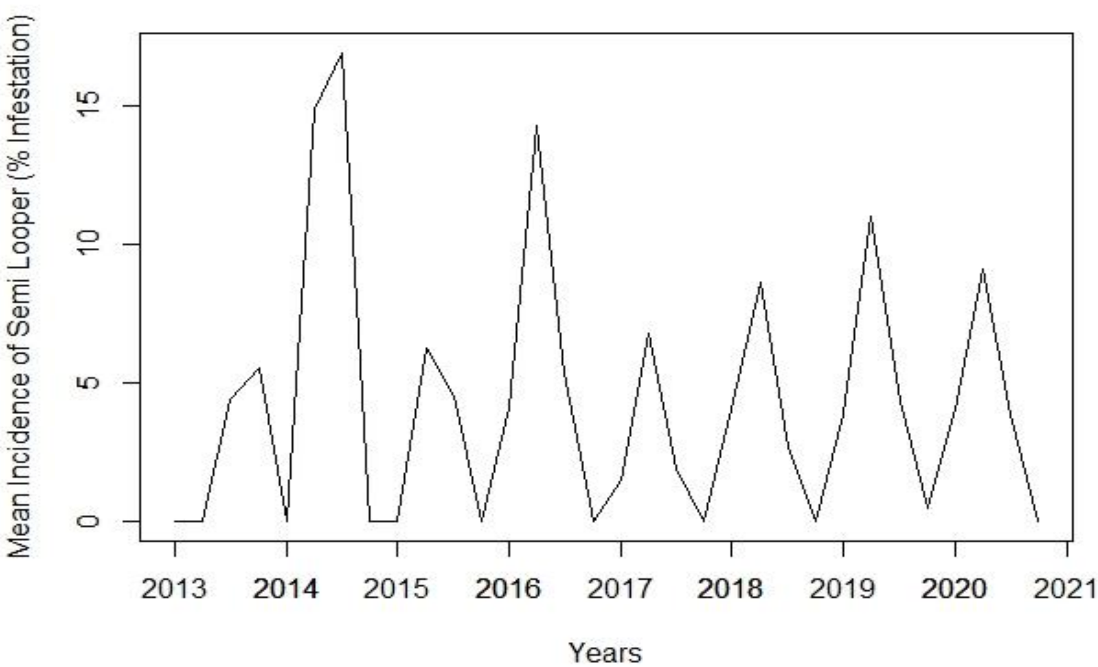

Fig. 4.7.1: Time Series Plot of original data set for Semi Looper incidence

Table 4.7.4 Seasonal Indices of Pest Incidence

Seasons	Semi Looper Incidence
35 DAS (S1)	0.51
45 DAS (S2)	2.05
55 DAS (S3)	1.27
65 DAS (S4)	0.17

Also, by observing the pattern of the plot from Fig. 4.7.1, it can be inferred that there may be presence of seasonality in the time series data set. So, the seasonal indices are also calculated and represented in Table 4.7.4 and by looking at the values it can be concluded that seasonality is present. Also, by seeing the plots in Fig. 4.7.2 and 4.7.3 which represent the acf and pacf values respectively, the acf and pacf values are significant at lag 4 and lag 8. Then using "nsdiffs" function in R software also suggests that there is need of seasonal differencing of 1st order. Thus, order of D is 1. The following Fig. 4.7.4 represents the time plot of Semi Looper incidence after seasonal differencing of 1st order.

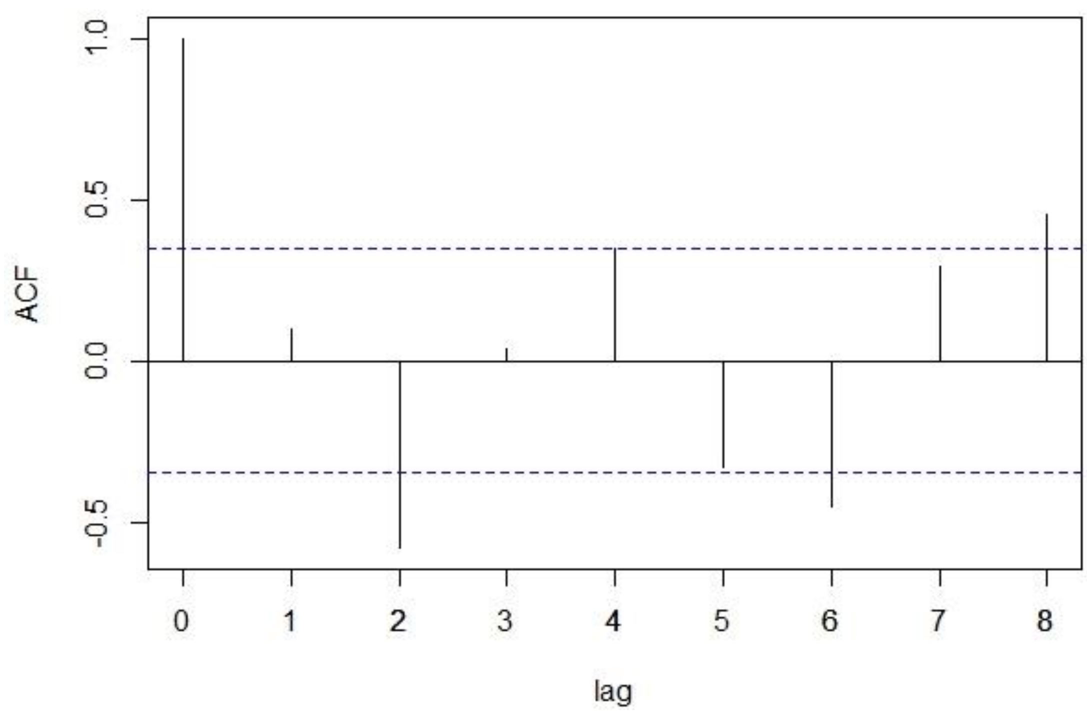

Fig. 4.7.2: Plot showing ACF values of original data set for Semi Looper incidence

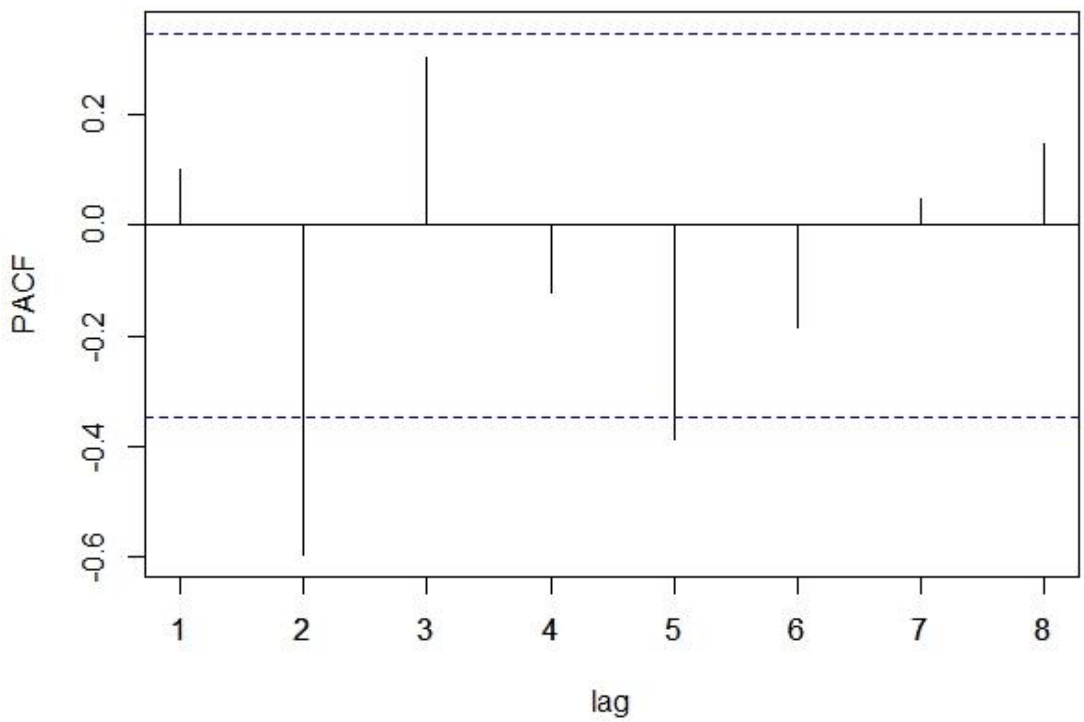

Fig. 4.7.3: Plot showing PACF values of original data set for Semi Looper incidence

Table 4.7.5 ACF and PACF values of original data set at lags up to 8

Lag	ACF	PACF
1	0.10	0.10
2	-0.58	-0.60
3	0.04	0.30
4	0.35	-0.12
5	-0.33	-0.39
6	-0.45	-0.19
7	0.29	0.05
8	0.46	0.15

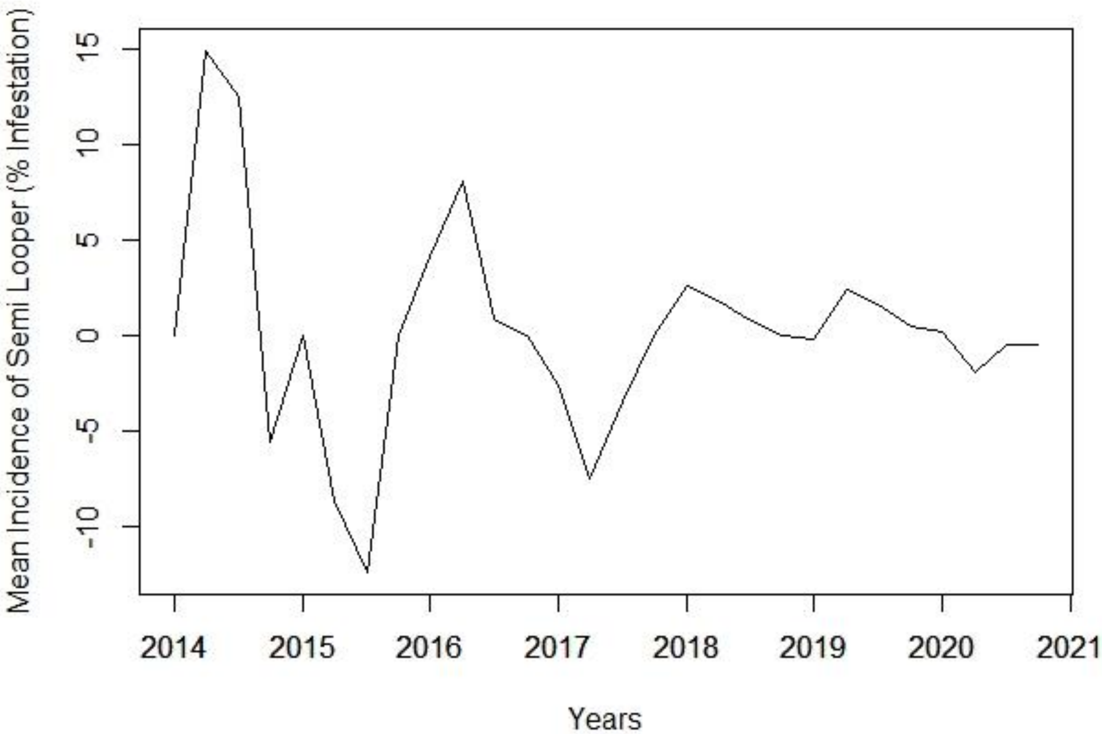

Fig. 4.7.4: Time Series Plot of seasonally differenced data set of Semi Looper incidence

4.7.2 ACF and PACF values

To determine the possible order of AR (p) and MA (q) of the SARIMA model, acf and pacf values are calculated based on the original series as the value of d is 0. But for calculation of SAR (P) and SMA (Q) of the SARIMA model, acf and pacf values are obtained after seasonal differencing as D is equal to 1. These values are noted in Table 4.7.5 and Table 4.7.6 and plots can be seen in Fig. 4.7.2, 4.7.3, 4.7.5 and 4.7.6.

From Fig. 4.7.2 as the acf value is remarkably significant at lag 2, so order of q may be 2. But from Fig. 4.7.3 it is not clearly evident as pacf values are significant at different lags. So, the most possible order may be (0, 0, 2).

Similarly, as the acf values are not significant at initial lags as evident from Fig. 4.7.5, so order of Q may be 0. From Fig. 4.7.6 as pacf value is significant at lag 1, so P maybe of order 1. So, the probable SARIMA model may be of order (0, 0, 2) $(1, 1, 0)_4$.

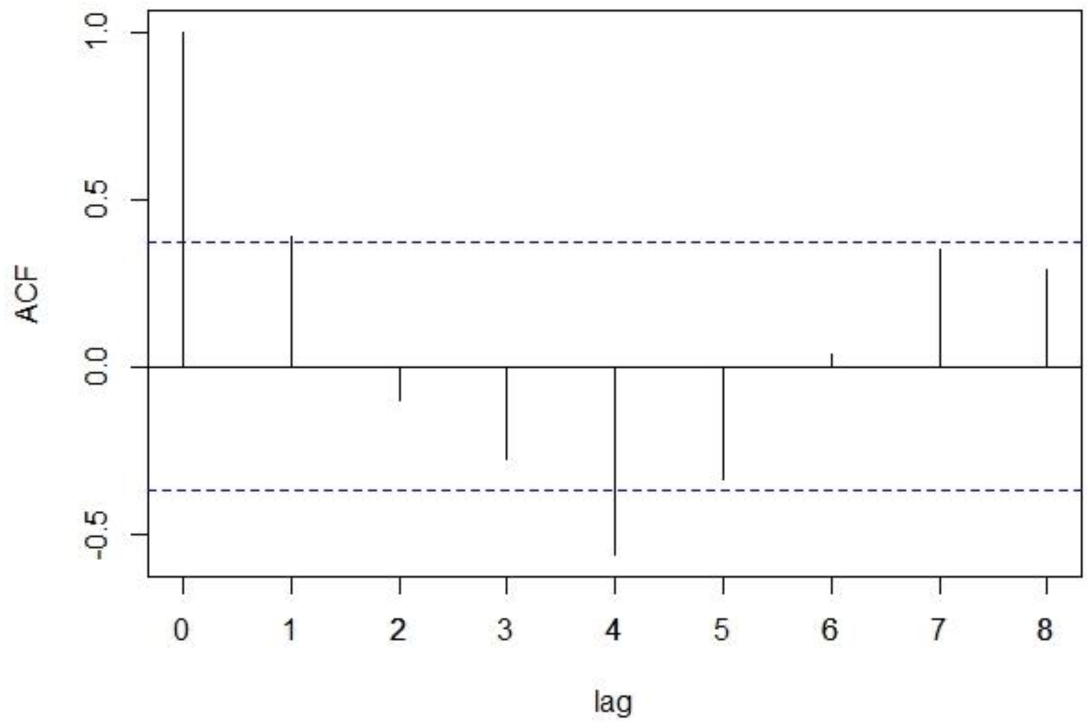

Fig. 4.7.5: Plot showing ACF values of seasonally differenced data set for Semi Looper incidence

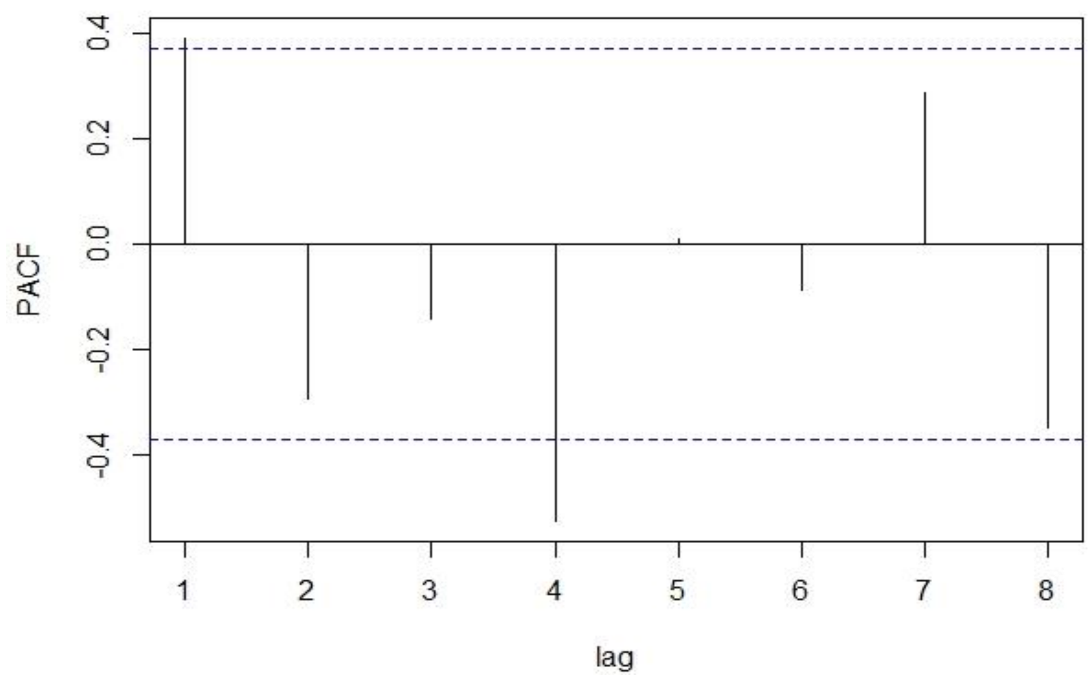

Fig. 4.7.6: Plot showing PACF values of seasonally differenced data set for Semi Looper incidence

Table 4.7.6 ACF and PACF values of seasonally differenced data set at lags up to 8

Lag	ACF	PACF
1	0.39	0.39
2	-0.10	-0.30
3	-0.28	-0.14
4	-0.56	-0.53
5	-0.33	0.01
6	0.04	-0.09
7	0.35	0.29
8	0.29	-0.35

4.7.3 Fitting of SARIMA model

After carrying out the required seasonal differencing, SARIMA models have been fitted using acf and pacf plots. From different models on the basis of the least AIC and BIC value SARIMA (0, 0, 2) (1, 1, 0)$_4$ model is selected and the estimates of parameters, respective s.e and p-values are depicted in Table 4.7.7.

Table 4.7.7 Parameter Estimates of the SARIMA (0, 0, 2) (1, 1, 0)$_4$ model for Semi Looper incidence

Model	Parameters	Estimate	S.E.	p-value
SARIMA (0, 0, 2) (1, 1, 0)$_4$	MA1	0.38	0.20	0.064
	MA2	-0.45	0.21	0.033*
	SAR1	-0.78	0.13	0.001***

***: Significant at 0.1%; *: Significant at 5%.

RMSE and RMdSE of the fitted SARIMA (0, 0, 2) (1, 1, 0)$_4$ model on the training data set is 2.81 and 1.66 respectively.

4.7.4 Fitting of SARIMAX model

In this present investigation maximum temperature at lag2 and maximum RH at lag2 are considered as exogenous variables as the VIF values are not that detrimental and SARIMAX models have been fitted. Considering the lowest AIC and BIC value SARIMAX (0, 0, 0) (0, 1, 0)$_4$ is found to be the best fitted model using “auto.arima” function in R software and estimated parameters with S.E. and p-values are presented in Table 4.7.8.

Table 4.7.8 Parameter Estimates of the SARIMAX (0, 0, 0) (0, 1, 0)$_4$ model for Semi Looper incidence

Model	Parameters	Estimate	S.E.	p-value
SARIMAX (0, 0, 0) (0, 1, 0)$_4$	**MaxT lag2**	0.34	0.26	0.186
	MaxRH lag2	-0.21	0.04	0.001***

***: Significant at 0.1%.

RMSE and RMdSE of the fitted SARIMAX (0, 0, 0) (0, 1, 0)$_4$ model on the training data set is 3.17 and 1.48 respectively.

4.7.5 Fitting of SVR model

As it is clear from section 4.7.4 that there is no AR component present in the possible SARIMAX model, so pest incidence is not fitted with itself for SVR model building purpose. But in the fitted SARIMAX model seasonal differencing is there. Therefore, seasonal adjustment in the data set has been carried out before applying SVR methodology. Thus, *ys* ~ *x* (where, *ys* is seasonally adjusted pest incidence values) is found to be the best fitted model on the basis of lowest training and testing error and the parameters are represented in Table 4.7.9.

Table 4.7.9 Parameters of the SVR (*ys* ~ *x*) model for Semi Looper incidence

Type	Kernel	Cost (C)	Gamma	Epsilon (ε)	No. of Support Vectors
eps-regression	radial	1	0.5	0.1	28

RMSE and RMdSE of the fitted SVR (*ys* ~ *x*) model on the training data set is 2.81 and 1.23 respectively.

4.7.6 Model Validation

To choose the best model out of all the fitted models, different accuracy comparison criterions can be used. For our present investigation RMSE and RMdSE are used for comparison of different models used for forecasting purpose and the values are noted down in Table 4.7.10.

Table 4.7.10 Predictive Abilities for SARIMA, SARIMAX and SVR models for Semi Looper

Model	Parameter	RMSE	RMdSE
SARIMA	$(0,0,2)\ (1,1,0)_4$	8.28	6.18
SARIMAX	$(0,0,0)\ (0,1,0)_4$	6.58	6.07
SVR (*ys* ~ *x*)	eps-regression, radial	6.83	4.88

From Table 4.7.10 it can be clearly perceived that SARIMAX model is found to be the best model having the least RMSE value. It is then followed by SVR model and the last one having highest RMSE value is SARIMA model, but by observing the RMdSE values SVR model is found to be the best model followed by SARIMAX and SARIMA respectively.

4.7.7 Residual Diagnostics

To check the appropriateness of the best fitted model, different diagnostic tests have been carried out on the residuals. To check whether the residuals are autocorrelated or not, Box-Ljung test has been carried out and the results are denoted in Table 4.7.11. From the table it is evident that p-value is more than 0.05 i.e., the null hypothesis is accepted and thus, the residuals are independent.

Table 4.7.11 Box-Ljung Test for Semi Looper incidence

Test Statistic	p-value
7.30	0.50

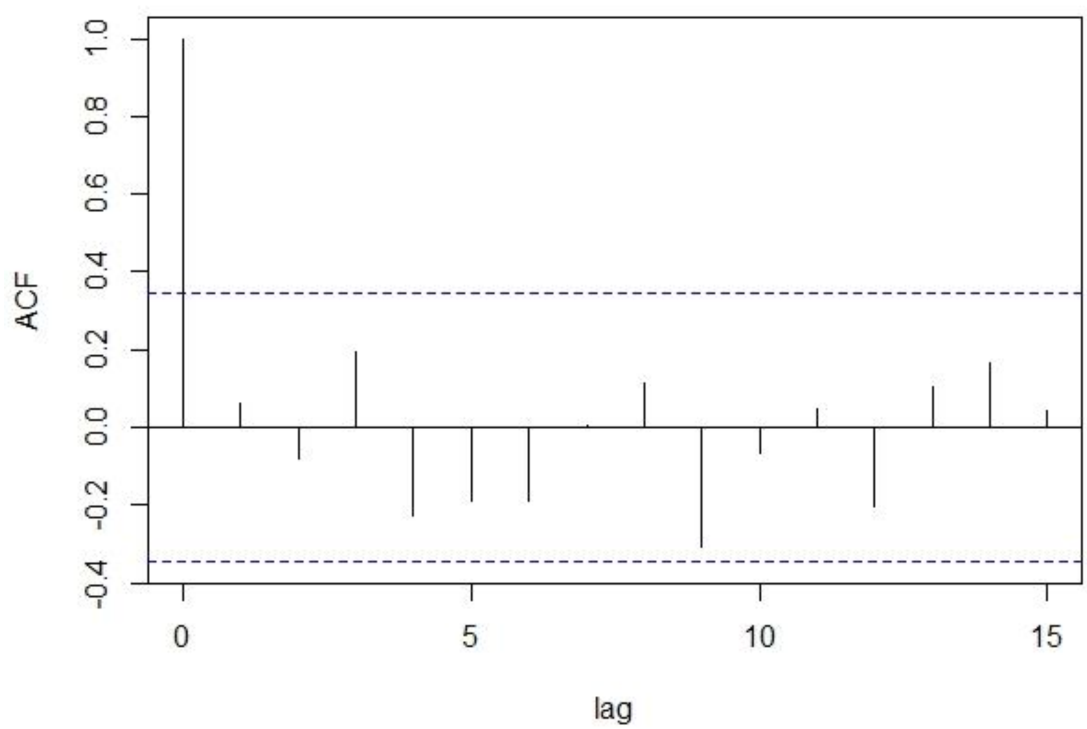

Fig. 4.7.7: Plot showing ACF values of residuals of ARIMAX model of Semi Looper incidence

The same can also be verified from Fig. 4.7.7 as the acf values of the residuals are non-significant in nature indicating that the residuals are independent.

To test the normality, Shapiro-Wilk test has been applied on the residuals and from the results depicted in Table 4.7.12 it can be concluded that it is significant, as the p value is less than 0.05. Thus, the null hypothesis is rejected and it is clear that the residuals are not normally distributed.

Table 4.7.12 Shapiro-Wilk Test for Semi Looper incidence

Test Statistic	p-value
0.93	0.03*

*: Significant at 5%

To test the presence of heteroscedasticity, ARCH LM test is generally used. Here also after applying ARCH LM test the results are denoted in Table 4.7.13. As the p-value is greater than 0.05, the null hypothesis is accepted. Thus, it can be concluded that there is no presence of ARCH effect in the residuals up to lag 8, i.e., the residuals are homoscedastic in nature. The same can also be confirmed from the shape of the plot in Fig. 4.7.8, as there is constant variance.

Table 4.7.13 ARCH LM Test for Semi Looper incidence

Test Statistic	p-value
7.70	0.46

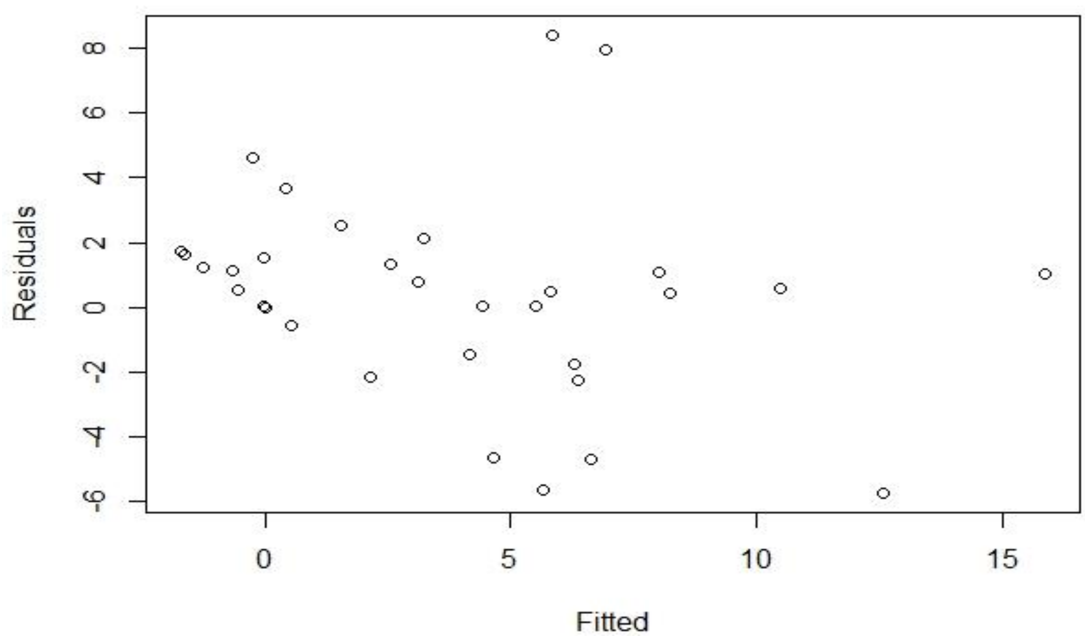

Fig. 4.7.8: Plot showing Fitted Values vs. Residuals of SARIMAX model for Semi Looper incidence

BDS test has been carried out to check the presence of non-linear component in the residuals after fitting different SARIMA models. From the results denoted in Table 4.7.14 it can be depicted that is absence of any non-linear pattern as all the epsilons are non-significant in nature. This implies that further application of hybrid models is not possible in the particular case.

Table 4.7.14 BDS Test for Semi Looper incidence

Dimension (m)	Epsilon (ε)		Test Statistic	p-value
2	eps (1)	1.60	-0.89	0.37
	eps (2)	3.19	0.34	0.73
	eps (3)	4.79	0.12	0.90
	eps (4)	6.38	-0.02	0.99
3	eps (1)	1.60	0.06	0.96
	eps (2)	3.19	0.35	0.73
	eps (3)	4.79	0.25	0.80
	eps (4)	6.38	0.06	0.95

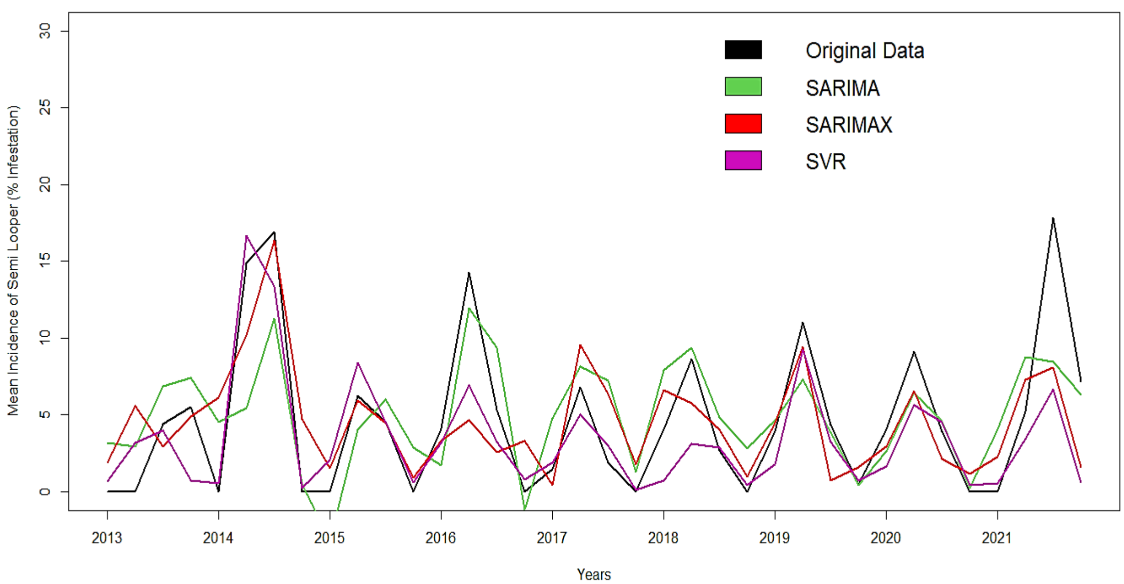

Fig 4.7.9: Plot showing original vs fitted values by SARIMA, SARIMAX and SVR model for Semi Looper incidence

Fig. 4.7.9 represents the plots of original values versus fitted values by different models like SARIMA, SARIMAX and SVR.

4.8 Diebold-Mariano (DM) test

DM test has been carried out to check any significance difference is there between RMSE values of all possible pairs of models and the results obtained are represented in Table 4.8.1 and 4.8.2.

Table 4.8.1 DM test for Yellow Mite

Model	Test Statistic	p-value
ARIMAX vs SVR	-0.83	0.41
ARIMAX vs ARIMA	-0.02	0.99
SVR vs ARIMA	0.83	0.41

Table 4.8.2 DM test for Semi Looper

Model	Test Statistic	p-value
SARIMAX vs SVR	0.64	0.53
SARIMAX vs SARIMA	0.70	0.49
SVR vs SARIMA	0.01	0.99

As it is evident from the Table 4.8.1 and 4.8.2 it is clear that there is no significant difference between the RMSE values of all possible pairs of models for both Yellow Mite and Semi Looper.

4.9 Forecasting of Pest incidence

After getting ARIMAX (0, 0, 1) and SARIMAX (0, 0, 0) $(0, 1, 0)_4$ as the best fitted models for Yellow Mite and Semi Looper respectively on the basis of model validation, out-of-sample forecast has been carried out for the year 2022 at 35, 45, 55 and 65 DAS and the results are represented in Table 4.9.1 and 4.9.2.

Table 4.9.1 Out-of-sample forecast for Mean incidence of Yellow Mite (no/sq cm) for 2022

DAS	Mean Incidence (no/sq cm)	MaxRH lag1 (%)	MinT lag2 (^{0}C)
35 DAS (S1)	4.12	95.51	21.07
45 DAS (S2)	2.38	94.46	21.52
55 DAS (S3)	2.89	90.56	21.52
65 DAS (S4)	3.45	92.36	21.52

Table 4.9.2 Out-of-sample forecast for Mean incidence of Semi Looper (% Infestation) for 2022

DAS	Mean Incidence (% Infestation)	MaxT lag2 (^{0}C)	MaxRH lag2 (%)
35 DAS (S1)	5.05	31.49	86.44
45 DAS (S2)	3.13	31.49	84.56
55 DAS (S3)	3.27	31.49	83.52
65 DAS (S4)	4.92	31.49	82.95

5. Summary and Conclusions

Chapter 5

SUMMARY AND CONCLUSIONS

Brief remarks about the results obtained in chapter 4 are mentioned in this current section. The findings from the analyses and results have been briefly covered in this portion that follows. In the present investigation seasonal plots have been constructed to check seasonality in incidence of Yellow Mite and Semi Looper. Further Pearson correlation analysis has been carried out to check the association between pest incidence and several weather parameters. At last, different forecasting methods have been fitted using initial 32 data points as training data set and rest 4 data points as testing data set and then best model is chosen on the basis of evaluation parameter. Some of the major outcomes are addressed followingly;

- From seasonal plots it is evident that for Yellow Mite and Semi Looper highest mean incidence is on 55 DAS and 45 DAS respectively for most of the years.
- Also, from the same plots it is observed that highest mean incidence is in the year 2014 for both the pests.
- From the results obtained in two-way ANOVA, the above mentioned two points are also verified to be true.
- By observing the seasonal plots and calculated seasonal indices, it is clear that for Yellow Mite seasonality is not present, but for Semi Looper seasonality is there.
- The same can also be confirmed from the results obtained in WO test.
- From the obtained results of correlation analysis, it is found that mean incidence of Yellow Mite has a significant negative correlation with minimum temperature (MinT) in current week and maximum relative humidity (MaxRH) at lag2. While correlation between MaxRH in lag1, minimum and maximum relative humidity (MinRH and MaxRH) in lag 2 with mean pest incidence is highly significant in a negative direction.

- Similarly for Semi Looper is found that maximum temperature (MaxT) at lag 2 is significantly positively correlated with mean incidence. But at lag 2 minimum and maximum relative humidity are negatively correlated with mean incidence of semi looper and the association is highly significant.

- From time series plots for both Yellow Mite and Semi Looper it is observed that any significant upward or downward trend is absent. So, in the time series data set trend is not present.

- For Yellow Mite different forecasting methods like ARIMA, ARIMAX and SVR have been fitted using training data set and then followed by model validation using RMSE and RMdSE values. On the basis of least RMSE value, ARIMAX is found to be the best fitted model followed by SVR and ARIMA. Similar results are also obtained by comparing RMdSE values.

- Similarly for Semi Looper forecasting methods like ARIMA, ARIMAX, SARIMA, SARIMAX and SVR have been fitted and it is observed that SARIMAX model produces the least RMSE value followed by SVR and SARIMA, but on the basis of RMdSE values SVR model is the best fitted followed by SARIMAX and SARIMA.

- From the results obtained in DM test, it is clear that there is no significant difference between the RMSE values of all possible pairs of models for both Yellow Mite and Semi Looper.

- Finally, forecasting of jute pest incidence has been carried out in Cooch Behar district of West Bengal for the year 2022 at 35, 45,55 and 65 DAS using the best fitted models.

Conclusion

Jute being one of the most popular fibre crops, has a tremendous effect on the small and marginal farmers of Cooch Behar district of West Bengal. As a drawback, every year insect pest infestation causes severe physical and economical losses in jute cultivation. So, in order to get insight about the major pest incidence the present investigation has been carried out and the major findings have been discussed in the earlier section. The obtained results in forecasting the pest incidence and the weather parameters can be used in warning the farmers in advance. The insights from the current study can also be used to make them prepared for necessary sustainable pest management methods. Thus, the present study can be useful in different manners for helping the farming community.

6. Future Scope of Work

Chapter 6

FUTURE SCOPE OF WORK

The current chapter comprises of some expected future works as it is evident from the findings attained and covered in earlier chapters that additional work could have been done to produce better results, which are discussed below.

- These forecasting methods can also be extended to other places in Terai zone.
- These methods can also be applied for some other disease data collected under the same project.
- There is also presence of some outlying observations in this time series data. Therefore, effects of these outlying observations may be lessened by using some suitable techniques for handling outliers in time series data.
- As the residuals are not normally distributed, so suitable transformation may be done prior to model fitting.
- In real-life situations, time-series data are rarely pure linear or nonlinear in nature. So, in order to tackle this situation Hybrid model is approached, which decomposes a time-series data into linear and nonlinear components (Zhang, 2001). Some popular examples of hybrid models are ARIMA-ANN, ARIMA-SVR etc. For the present data set after conducting BDS test, it is confirmed that the residuals are linear in nature. Therefore, hybrid model is not suitable for the present data set. But for some other pest data set, if residuals are found to be nonlinear in nature following fitting of linear models like ARIMA or ARIMAX, then use of hybrid model may be considered.
- In future, reasons for high values of out-of-sample forecast in Yellow Mite and Semi Lopper during 35 DAS may be explored using some advanced machine learning or deep learning techniques such as, random forest, wavelet ANN,

Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and other deep learning algorithms.

7. Bibliography

Chapter 7

BIBLIOGRAPHY

Agrawal R and Mehta S C (2007) Weather based forecasting of crop yields, pests and diseases- IASRI models. *Journal of the Indian Society of Agricultural Statistics* **61:** 255-263.

Almeyda E, Paiva J and Ipanaqué W (2020) Pest incidence prediction in organic banana crops with machine learning techniques. *Engineering International Research Conference* **:**1-4.

Arya P, Paul R K, Kumar A, Singh K N, Sivaramne N and Chaudhary P (2015) Predicting pest population using weather variables: an ARIMAX time series framework. *International Journal of Agricultural and Statistical Sciences* **11:** 381-386.

Aswathi V S and Duraisamy M R (2018) Comparison of Prediction Accuracy of Multiple Linear Regression, ARIMA and ARIMAX Model for Pest Incidence of Cotton with Weather Factors. *Madras Agricultural Journal* **105:** 313-316.

Balikai R A and Venkatesh H (2019) Weather based prediction models to forecast major pests of rabi sorghum in Karnataka, India. *Journal of Experimental Zoology, India* **22:** 131-134.

Bhat M N, Ahmad M, Vennila S, Singh G, Sardana H R, Saxena A K, Sridhar V and Yadav S K (2018) Severity, weather influence and prediction of early blight of Tomato for Eastern dry zone of Karnataka. *Annals of Plant Protection Sciences* **26:** 165-169.

Bhosale A A, Hole U B and Aghav S T (2019) Studies on seasonal incidence of hairy caterpillar on jute. *Journal of Entomology and Zoology Studies* **7:** 1159-1161.

Bierens H J (1987) ARMAX model specification testing, with an application to unemployment in the Netherlands. *Journal of Econometrics* **35:** 161-190.

Box G E P and Jenkins G M (1976) *Time Series Analysis: Forecasting and Control*. Revised Edition, San Francisco: Holden Day.

Chapke R R (2013) Role of jute cultivation in farmers' livelihood. *Indian Res J Ext Edu* **13:** 132-135.

Chiu L Y, Rustia D J A, Lu C Y and Lin T T (2019) Modelling and forecasting of greenhouse whitefly incidence using time-series and ARIMAX analysis. *International Federation of Automatic Control* **52:** 196-201.

Damos P T and Savopoulou-Soultani M (2010) Development and statistical evaluation of models in forecasting moth phenology of major lepidopterous peach pest complex for Integrated Pest Management programs. *Crop Protection* **29:** 1190-1199.

Das L K and Singh B (1977) Economic control measures against the major pests of jute. *PANS* **23:** 159-161.

Das S S, Ray S, Sen A, Siva G S and Das S (2019) Statistical study on modeling and forecasting of jute production in West Bengal, India. *International Journal of Current Microbiology and Applied Science* **8:** 1719-1730.

de Oliveira Aparecido L E, de Souza Rolim G, da Silva Cabral De Moraes J, Costa C T S and de Souza P S (2020) Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology* **64:** 671-688.

de Oliveira J F L and Ludermir T B (2014) A distributed PSO-ARIMA-SVR hybrid system for time series forecasting. *International conference on systems, man, and cybernetics* 3867-3872.

Durgabai R P L, Bhargavi P and Jyothi S (2018) Pest management using machine learning algorithms: a review. *International Journal of Computer Science Engineering and Information Technology Research* **8:** 13-22.

Dutta S, Rakshit S and Chatterjee D (2020) Use of Artificial Intelligence in Indian Agriculture. *Food and Scientific Reports* **1:** 65-72.

Gotyal B S, Babu V R, Satpathy S and Selvaraj K (2019) Biological control of key pest of jute, *Spilosoma obliqua* Walker (Lepidoptera: Arctiidae)-a case study. *Journal of Biological Control* **33:** 222-224.

Islam M N, Islam K S, Jahan M and Rahman M S (2020) Incidence of the yellow jute mite *Polyphagotarsonemus latus* (Acari: Tarsonemidae) depending on the growing season, day hours, sowing time, plant canopy and plant age in Bangladesh. *Israel Journal of entomology* **50:** 75-84.

Kamruzzaman A S M, Alam M Z and Miah M R U (2013) Effect of yellow mite, Polyphagotarsonemus latus (banks) density on hosts (*Corchorus olitorius* L.) phenology and assessment of yield loss under net house condition. *African Journal of Agricultural Research* **8:** 380-385.

Katke M K, Balikai R A and Venkatesh H (2009) Seasonal incidence of grape mealy bug, *Maconellicoccus hirsutus* (green) and its relation with weather parameters. *Pest management in horticultural ecosystems* **15:** 9-16.

Kumar A, Tanwar N, Deepankar, Ravinder, Nimbrayan P K, Rahul, Kumar V, Sain V and Malik S (2021) ARIMA vs ARIMAX Modelling for Mustard Yield Forecasting in Haryana. *Agricultural Mechanization in Asia.* **52:** 2553-2562.

Kumari M (2019) A study on the crop damage from bihar hairy caterpillar and available means to reduce it extent of such damage. *International Journal in Physical and Applied Sciences* **6:** 1-5.

Lama A, Singh K N, Singh H, Shekhawat R, Mishra P and Gurung B (2022). Forecasting monthly rainfall of Sub-Himalayan region of India using parametric and non-parametric modelling approaches. *Modeling Earth Systems and Environment* **8:** 837-845.

Lee H, Moon A, Moon K and Lee Y (2017) Disease and pest prediction IoT system in orchard: A preliminary study. *International Conference on Ubiquitous and Future Networks* **9:** 525-527.

Lv J, Wang X, Xie D and Wei C (2009) A hybrid approach of neural network with particle swarm optimization for tobacco pests prediction. *International Conference on Computer and Computing Technologies in Agriculture* **2:** 1251-1260.

Ma G and Ma C S (2022). Potential distribution of invasive crop pests under climate change: incorporating mitigation responses of insects into prediction models. *Current Opinion in Insect Science* **49:** 15-21.

Marković D, Vujičić D, Tanasković S, Đorđević B, Ranđić S and Stamenković Z (2021) Prediction of pest insect appearance using sensors and machine learning. *Sensors* **21:** 1-13.

Mishra R, Sharma V and Singh B P (2014) Population Dynamics of Mite (Polyphagotarsonemus Latus Bank) and Correlation with Weather Parameters on Processing Varieties of Potato. *Voyager* **5:** 122-134.

Mitra D and Paul R K (2017). Hybrid time-series models for forecasting agricultural commodity prices. *Model Assisted Statistics and Applications* **12:** 255-264.

Mokhlis N H M, Burhan N A A, Roeslan N F A, Moin S N A Z and Nur U A M (2021) Forecasting healthcare stock price using ARIMA-GARCH model and its value at risk. *International Journal of Business and Economy* **3:** 127-142.

Narayanasamy M, Kennedy J S and Geethalakshmi V (2017) Weather Based Pest Forewarning Model for Major Insect Pests of Rice–An Effective Way for Insect Pest Prediction. *Annual Research & Review in Biology* **21:** 1-13.

Patil J and Mytri V D (2013) A prediction model for population dynamics of cotton pest (Thrips tabaci Linde) using multilayer-perceptron neural network. *International Journal of Computer Applications* **67:** 19-26.

Paul R K, Prajneshu and Ghosh H (2013) Statistical modelling for forecasting of wheat yield based on weather variables. *Indian Journal of Agricultural Sciences* **83:** 180-183.

Paul R K, Vennila S, Singh N, Chandra P, Yadav S K, Sharma O P, Sharma V K, Nisar S, Bhat M N, Rao M S and Prabhakar M (2018). Seasonal dynamics of sterility mosaic of pigeonpea and its prediction using statistical models for Banaskantha region of Gujarat, India. *Journal of The Indian Society of Agricultural Statistics* **72:** 213-223.

Permanasari A E, Hidayah I and Bustoni I A (2013) SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. *International Conference on Information Technology and Electrical Engineering* 203-207.

Prasad Y G and Prabhakar M (2012) Pest monitoring and forecasting. *Integrated pest management: principles and practice* **:**41-57.

Rahman S and Khan M R (2012a) Incidence of pests in jute (*Corchorus olitorius* L.) ecosystem and pest–weather relationships in West Bengal, India. *Archives of Phytopathology and Plant Protection* **45:** 591-607.

Rahman S and Khan M R (2012b) Effect of plant characteristics of jute varieties on incidence of pests in West Bengal, India. *Archives of Phytopathology and Plant Protection* **45:** 608-619.

Roy T N (2016) Is agriculture diversification for sectoral transformation–An emperical study in Coochbehar district of West Bengal. *Journal of Crop and Weed* **12:** 121-127.

Sadat A and Chakraborty K (2015) Insect pest constrains of jute and its control by biological agents under modern Eco-Friendly sustainable production system. *International Journal of Multidisciplinary Research and Development* **2:** 316-321.

Sarkar S and Majumdar B (2016) Present Status of Jute Production and Technological and Social Interventions Needed for Making Jute Agriculture Sustainable and Remunerative in West Bengal. *Indian Journal of Natural Fibers* **3:** 23-36.

Shamna A, Jha S K, Chakraborty A K and Kumar S (2017) Jute Production: The Need for Linkage among Research, Farmers and Industries in West Bengal, India. *Current Journal of Applied Science and Technology* **25:** 1-6.

Singh A K, Jha S K, Majumdar B, Roy M L, Sarkar S and Ghorai A K (2019) Impacts of climate smart jute farming on resource use efficiency, productivity and economic benefits in rural Eastern India. *Outlook on Agriculture* **48:** 75-82.

Sinha M K, Mitra S, Ramasubramanian T and Mahapatra B S (2009) Crop diversification for profitability in jute and allied fibre crops. *Indian Journal of Agronomy* **54:** 221-225.

Suyal P, Gaur N, Rukesh P K N and Devrani A (2018) Seasonal incidence of insect pests and their natural enemies on soybean crop. *Journal of Entomology and Zoology Studies* **6:** 1237-1240.

Vapnik V (2000) *The nature of statistical learning theory*. SpringerVerlag New York Inc, New York.

Webel K and Ollech D (2021) WO test in seastests: Seasonality Tests. Retrieved September 16, 2021, from https://rdrr.io/cran/seastests/man/wo.html.

Xiao Q, Li W, Kai Y, Chen P, Zhang J and Wang B (2019) Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. *BMC bioinformatics* **20:** 1-15.

Zhang G P (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50:** 159-175.

ouriginal
by Turnitin

Document Information

Analyzed document	Chinmaya Subhrajyoti Panda (A-2020-016-M) Thesis.docx (D143231197)
Submitted	8/24/2022 8:44:00 PM
Submitted by	Prof. Debsankar Gupta
Submitter email	debsankargupta.research@gmail.com
Similarity	1%
Analysis address	debsankargupta.research.ubkv@analysis.urkund.com

Sources included in the report

W	URL: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/ Fetched: 8/20/2020 7:12:51 PM	88	1
SA	**Uppsala Universitet / ERIC AIDOO_Master Thesis.pdf** Document ERIC AIDOO_Master Thesis.pdf (D4210950) Submitted by: johan.lyhagen@statistik.uu.se Receiver: lyhagen.uu@analys.urkund.se	88	1
SA	**Copenhagen Business School / Predictive_Exam.pdf** Document Predictive_Exam.pdf (D108599172) Submitted by: DE_CBSProd_1269332.arccbs@submitters.urkund.com Receiver: arcanic.cbs@analyse.urkund.com	88	1
SA	**12631246.pdf** Document 12631246.pdf (D50365481)	88	1
W	URL: https://otexts.com/fpp2/seasonal-arima.html Fetched: 11/19/2020 11:05:31 PM	88	1
SA	**Babasaheb Bhimrao Ambedkar University, Lucknow / ARIMA project (9).pdf** Document ARIMA project (9).pdf (D125813922) Submitted by: anujkmsmath2016@gmail.com Receiver: anujkmsmath2016.bbau@analysis.urkund.com	88	1
SA	**Report_Chapter 1-5_Wan Nur Afrina_2021175575.pdf** Document Report_Chapter 1-5_Wan Nur Afrina_2021175575.pdf (D142399835)	88	3

Entire Document

Chapter 1
INTRODUCTION

29-08-2022
(Chairman, Advisory Committee)

Chinmaya Subhrajyoti Panda
29-08-2022