Institute Project Code: IASRI P.R.- 01/2020
आई॰ ए॰ एस॰ आर॰ आई॰ /पी॰ आर॰-01/2020
AGEDIASRISIL201700900095

Development of count time-series models for predicting pest dynamics using weather variables

हर कदम, हर डगर
किसानों का हमसफर
भारतीय कृषि अनुसंधान परिषद
Agrisearch with a human touch

Prawin Arya	प्रवीण आर्य
Md. Wasi Alam	मौ. वसी आलम (from 20/07/2018)
Bishal Gurung	बिशाल गुरुंग (19/04/17 to 19/07/2018)

पूर्वानुमान एवं कृ ष प्रणाली प्रतिकथन प्रभाग
Forecasting and Agricultural Systems Modelling

भारतीय कृषि साख्यिकी अनुसंधान संस्थान (भा.कृ.अनु.प.)
लाइब्रेरी एवेन्यु, पूसा, नई दिल्ली – 110 012
Indian Agricultural Statistics Research Institute (I.C.A.R.)
Library Avenue, PUSA, New Delhi – 110 012
www.iasri.res.in

2020

आमुख

कृषि भारत की अर्थव्यवस्था में एक प्रमुख भूमिका निभाती है क्योंकि यह कुल सकल घरेलू उत्पाद (जीडीपी) के लगभग सत्रह प्रतिशत का योगदान देती है और लगभग 60 प्रतिशत आबादी प्रत्यक्ष या अप्रत्यक्ष रूप से कृषि पर निर्भर करती है। खेती एक अत्यधिक जोखिम भरा व्यवसाय है क्योंकि यह विभिन्न प्राकृतिक आपदाओं जैसे मौसम, आग, बाढ़, ठंढ, ओलावृष्टि, कीट और बीमारियों आदि के संपर्क में होता है। भारत में, अधिकांश किसान गरीब हैं और उनके पास बेहद सीमित साधन और संसाधन हैं, यहाँ तक कि प्रकृति की मार स्वरूप एक फसल की विफलता उन्हें गरीबी के दुष्चक्र में धकेल देती है। कृषि अत्यधिक लागत वाली और अनिश्चितताओं से भरी हुई है और समय पर उपाय किसान के जोखिम को कम कर सकते हैं। कीटों के हमलों से खतरे अक्सर स्थानीय होते हैं, लेकिन मानसून की विफलता या फसल की कीमतों में गिरावट जोखिमों को और ज्यादा कठोर बना देते है । फसलों में कीट और रोगों की इन घटनाओं ने कृषि को बहुत जोखिम भरा उद्यम बना दिया है और महंगे बीज और खेती की ज्यादा लागत के कारण किसान नई तकनीकों को अपनाने में बहुत आशंकित रहते हैं। कीटों के हमलों के कारण कुल फसलों की उपज का एक चौथाई हिस्सा हर साल नष्ट हो जाता है। अतः इन समस्याओं को कम करने के लिए, समय पर पूर्वानुमान विधियाँ एक महत्वपूर्ण और अत्यंत उपयोगी सूचना प्रदान करता है। भारत में कीटनाशक का उपयोग शाकनाशकों और फफूंदनाशकों की तुलना में अधिक किया जाता है। भारत में कीटनाशकों का मुख्य उपयोग धान के बाद कपास की फसल के लिए होता है। कपास की फसल को इन कीटों से होने वाले नुकसान को दूर करने के लिए कीटनाशकों के गहन उपयोग की आवश्यकता होती है, जिसके कारण कीटनाशकों की ज्यादा लागत और कम पैदावार के कारण देश के कई हिस्सों में इसकी खेती को गैर-आर्थिक बना दिया है। लेकिन किसानों के लिए उत्पादकता और आय बढ़ाने के लिए, कीट और बीमारी प्रबंधन महत्वपूर्ण है। ये प्रबंधन विभिन्न फसलों की उपज में वृद्धि करते हैं।

Count time series में इवेंट्स समय के लगातार क्रम में होते हैं जैसा कि कीट डेटा में होता है। पूर्णांक समय श्रृंखला असतत समय श्रृंखला का एक महत्वपूर्ण वर्ग है और INAR विधि इस प्रकार की समय श्रृंखला के लिए अनुकूल है जो पॉइसन, नकारात्मक द्विपद, सामान्यीकृत पॉइसन वितरण का अनुसरण करती है। फसलों में कीट / रोगों के संक्रमण के लिए मौसम संबंधी कारक भी अत्यधिक जिम्मेदार होते हैं। इस अध्ययन में GLMX, INARX और पूर्णांक ANN मॉडेल्स को मौसम के मापदंडो के साथ अनुमानित किया गया है तथा कीट और रोग की प्रारंभिक चेतावनी किसानों के लिए प्रस्तुत की गयी है जिसके फलस्वरूप वह समय रहते आगे की कार्यवाही करने में सक्षम हो जाता है ।

परियोजना अन्वेषक

PREFACE

Agriculture plays a predominant role in India's economy as it contributes about seventeen percent of total gross domestic product (GDP) and about 60 percent of population depends on agriculture sector directly or indirectly. However farming is highly risked prone as it is exposed to various natural calamities such as weather, fire, flood, frost and hailstorm, pest and diseases, etc. In India, most of the farmers are poor and have extremely limited means and resources, even a single crop failure of a disastrous nature pushes them in the vicious cycle of poverty. Agriculture being highly cost intensive and full of uncertainties and timely measures can minimize the farmer's risk. Threats from pest attacks are often localized but underlines the multitude of risks apart from those related to monsoon failure or a crash in crop prices. These incidences of pest and diseases in crops have made agriculture very risky venture and due to high seed cost and cost of cultivation farmers are very apprehensive in adopting new technologies. About one fourth of total crops yield is lost each year due to pest attacks. To mitigate these problems, reliable and timely forecast provides an important and extremely useful input in formulation of policies. In India, insecticide is used more compared to herbicides and fungicides. Main use of pesticides in India is for cotton crop followed by paddy. The cotton crop requires an intensive use of pesticides to overcome the incidence of damages from pests, thereby making its cultivation uneconomic in many parts of the country due to the high cost of pesticides and low yields. But, to enhance the productivity and income to the farmers, forewarning of pest and disease and pest management is crucial. These management practices potentially increase the yield of different crops.

In count time series the events occur in the consecutive points of time as it occurs in pest count data. Integer-valued time series is an important class of discrete-valued time series models and INAR process is well-suited for such type of time series which follows poisson, negative binomial, generalized poisson distributions. Meteorological factors are also highly responsible for pest/diseases infestation in crops. In this study advanced models like GLMX, INARX and integer valued ANN models with weather parameters as exogenous variables were developed for modeling and predicting pest dynamics to address appropriate solutions for early warning of pest and disease infestation.

Project Investigators

Contents

Chapter I

INTRODUCTION

1.1. Count Time Series Modelling for pest dynamics

Agriculture being highly cost intensive and full of uncertainties have great impact on the livelihood of farmers, if timely measures are not taken to minimize the risk, they may fall in the trap of vicious cycle. Not only this, farmers in several states are battling with growing incidence of pest attacks on a variety of crops. The threats farmers face from pest attacks are often localised but underlines the multitude of risks apart from those related to monsoon failure or a crash in crop prices Therefore, incidence of pest and diseases in crops have made agriculture very risky venture. Due to high cost of cultivation, farmers are very apprehensive in adopting new technologies. About 15-25 per cent of crops yields is lost each year due to pest attacks. To mitigate these problems, reliable and timely forecast provides an important and extremely useful input in formulation of policies. In India, merely 70% of the pesticide used is insecticide and the use of herbicides and fungicides is correspondingly less compared to insecticides. The main use of pesticides in India is for cotton crops (36%), followed by paddy (20%). Andhra Pradesh is the highest pesticides consuming state (23%) followed by Punjab and Maharashtra (Bhardwaj and Sharma 2013). In India Cotton is a major commercial crop for sustainable economy of India and livelihood of the Indian farming community. It is cultivated in 11.0 Million hectares in the country. India accounts for about 32% of the global cotton area and contributes to 21% of the global cotton produce, currently ranked second after China, but the productivity is found to be very low because it is prone to pest attacks and damag largely by many pests. The main pests of cotton crops are American bollworm (Helicoverpa armigera), Whitefly (Bemisia tabaci), Jassids (Amrasca bigutella bigutella), and Pink bollworm (Pectinophora gossypiella) etc. The cotton crop requires an intensive use of pesticides to overcome the incidence of damage from these pests. The major cotton producing states include Maharashtra, Gujarat, Andhra Pradesh, Punjab, Karnataka, and Madhya Pradesh. But to enhance the productivity and income to the farmers, forewarning of pest and disease is crucial. In agriculture, disease and pest management is very much important. Because these management practices potentially increase the yield of different crops.

In count time series, the events occur in the consecutive points of time, which is commonly occurs in many situations, for example, the number of road accidents in a week, number of seeds germinated in a week etc. Integer-valued time series is an important class of discrete-valued time series models. The INAR process is well-suited for many time series which follows poisson, negative binomial, generalized poisson distributions etc. As a nonlinear and nonparametric class of model integer based neural network is very potential to capture the count time series trend and it have wide application in many areas like image classification, pattern recognition etc.

Over the years, different methodologies were introduced from time to time. Since meteorological factors are highly responsible for pest/diseases infestation in crops, therefore, advanced models like INARX and ANN along with weather parameters may address appropriate solutions for early warning of pest/disease infestation for investigating and predicting pest/disease status. With these background the INAR and integer based neural network models by considering information on exogenous variables will be developed for modelling and predicting pest dynamics in cotton crop. It is generally agreed that forecasting methods should be assessed for accuracy by using out-of-sample forecasts rather than goodness of fit to past data. In order to understand the probabilistic behaviour of future data, out of-sample forecasts are required. Formulae for optimal out-of-sample forecasts were derived in this study.

1.2. Review of Literature

Crop pests are evolving to spread to new area by adapting to the climate change. It is very difficult to predict the attacks of pest and diseases. Different researchers, over the time have given various methodologies for forecasting of pest attacks. The review of the available literature relevant to the proposed study has been furnished in this section with a perspective to overview the various methodologies and procedures employed in this study.

McKenzie (1985) developed Binomial autoregressive model for binomial count observations and the structure of model is well-interpretable. For stationary sequence of count observations.

McKenzie, E. (1985b) Contribution to the Discussion of ‘Modelling and Residual Analysis of Nonlinear Autoregressive Time-Series’ by A.J. Lawrance and P.A.W. Lewis, J.R. Statist. Soc.(B) 47,187-188.

Al-Osh and Alzaid (1987) described integer-valued autoregressive model of stationary sequence with lag-one dependence and is known as INAR (1) model or Poisson INAR (1) model. They showed that this model is most suitable for discrete observations.

Alzaid (1987) first introduced integer-valued random variables for lag-one is known as INAR (1) process or Poisson INAR(1) process. They showed that this model is most suitable for count observations. They showed that the distributional properties and correlation structure of the model are similar to the continuous valued autoregressive or AR(1) process. Different estimation procedures such as maximum likelihood estimation (MLE), conditional least squares (CLS) and Yule-Walker (YW) method were also described.

Alzaid and Al-Osh (1990) extended INAR (1) process up to the pth order which is useful for modelling discrete-time dependent counting process. They showed the difference with the Gaussian AR (p) process in terms of correlation, Markovian Property and regression.

Bockenholt (1999) developed INAR model with Poisson regression for study the regularity and predictability of purchase behaviour over time. The process facilitates the analysis of heterogeneity and autocorrelation.

Agrawal et al. (2001) developed forecasting model for wheat in Vindhyanchal Plateau zone of Madhya Pradesh. It was reported that reliable forecasting yield could be obtained when both the crops were 12 weeks old i.e. about 2 months before harvest.

Hellstrom (2002) described the modelling of count on tourism demand. He used the basic INAR (1) model and used it to realistic empirical economic applications. He also extended the INAR (1) model up to different lags.

Agrawal and Mehta (2007) developed several weather based forecasting models for crop yield of rice, wheat, sorghum, maize and sugarcane at selected districts/agro climatic zones/states of India using regression analysis, discriminant function analysis and water balance technique.

Bu and McCabe (2008) developed estimation and model selection procedure for a class of integer valued autoregressive models for any number of lags.

Pavlopoulos and Karlis (2008) developed INAR (1) model which discusses about the non-linear structure of auto-regressive Markov Chain on total time length of the series, where error follows a finite mixture distribution of Poisson laws.

Weib (2008) discussed the count data analysis in time series using AR(p) model. Some marginal distributions of the discrete self-decomposing distributions family were outlined.

Silva et al. (2009) proposed Bayesian methodology for forecasting integer-valued time series, modelled by the INAR (1) process. Point predictions as well as confidence intervals for the predicted values are obtained. The predicted values are compared with their classic counterparts.

Sang (2010) discussed the design of Multilayer Perceptron (MLP) especially for pattern classification problems. This discussion included how to decide the number of nodes in each layer, how to initialize the weights of MLPs, how to train MLPs among various error functions, the imbalanced data problems, and deep architecture.

Karlaftis and Vlahogianni (2011) discussed differences and similarities between these two approaches; Statistical methods and neural networks. They reviewed relevant literature and attempt to provide a set of insights for selecting the appropriate approach.

Pedeli and Karlis (2011) discussed a bivariate integer valued autoregressive (INAR) process of order 1. They have given emphasis on bivariate poison and bivariate negative binomial innovations.

Rozman et al. (2012) developed a hybrid model based on image analysis and neural network. From the end of fruit thinning in June till harvesting digital images of 120 trees of yellow-skin 'Golden Delicious' (four times) and 120 trees of red-skin 'Braeburn' (five times) were captured from intensive orchards.

Sang (2012) proposed a new error function, in order to improve the error back-propagation algorithm for the classification of imbalanced data sets. This method was compared with the two-phase, threshold-moving, and target node methods through simulations in a mammography data set and the proposed method attained the best results.

Sharma et al. (2012) elaborated Artificial Neural Network or ANN, its various characteristics and business applications. They also showed that "what are neural networks" and "Why they are so important in today's Artificial intelligence?" Because numerous advances have been made in developing intelligent system, some inspired by biological neural networks.

Bhardwaj and Sharma (2013). Studied impact of pesticides application in agricultural Industry in India.

Weib and Pollett (2012) introduced chain binomial population model and also established a relationship with ecology and epidemiology. The connection of chain-binomial models with binomial autoregressive (AR) processes was also developed.

Kumar et al., (2013) used Multilayer perceptron (MLP) and Radial basis function (RBF) neural network to predict the outbreak of disease and pest of mustard crop. MLP neural network was found better in terms of mean absolute percentage error (MAPE).

Kumari et al. (2013) developed a model to forecast the productivity and pod damage by Helicoverpa armigera using artificial neural network model in pigeonpea (Cajanus Cajan). Sigmoid and linear functions were used as activation function hidden and output nodes respectively.

Rudra (2013) presented an application of Artificial Neural Network (ANN) to forecast inflation in India during the period 1994-2009. The paper finally concluded that multivariate models were better forecasting performance over the univariate model.

Enciso-Mora et al. (2009) developed INAR processes which are perfectly suited for modelling count data including the explanatory variables into the model. An efficient MCMC algorithm was constructed to analyze the model and incorporates both explanatory variables and order selection.

Karale and Sharma (2014) investigated probability models for explaining population dynamics of major insect pests under rice-potato-okra cropping system.

Kumari et al. (2014) presented time series forecasting of losses due to pod borer, pod fly and productivity of pigeonpea (Cajanus cajan) for North West Plain Zone (NWPZ) by using artificial neural network (ANN). The ANN model performed better as compared to ARIMA model.

Arya et al. (20015) developed ARIMAX time series model for modelling and forecasting the pest population using count data with climatic information as exogenous variable. This model was found to be an appropriate model for forecasting pest population.

1.3. Motivation

Over the last few years, the class of models particularly applicable to the analysis of time series count data have been studied. Count outcome variables are sometimes log-transformed and analysed using OLS regression. Many issues arise with this approach, including loss of data due to undefined values generated by taking the log of zero (which is undefined), as well as the lack of capacity to model the dispersion. Integer-valued autoregressive (INAR) models, Poisson models and negative Binomial models have also been studied by many researcher’s models take the autocorrelation and discrete nature of the data into account. INAR and generalized linear models (GLM) have many applications, not only to the analysis of counts of events, but also in other field like in the analysis of survival data.

An attempt is made to incorporate exogenous variables in GLM and INAR model for their improvement. Integer based Neural network which is generally applied in image processing task, has been attempted for developing integer based neural network using exogenous variables for predicting pest dynamics.

1.4. Objectives

With above discussed motivations and research gaps following objectives were framed;

i. To predict the pest dynamics using generalized Linear Models (GLM) with exogenous variables
ii. To develop Integer-Valued Autoregressive model with exogenous variable (INARX) for pest dynamic prediction
iii. To develop integer based neural network model with exogenous variable for pest dynamic prediction
iv. To compare the proposed models with conventional models

1.5. Expected output

This study is an attempt in the direction for relief to consumer and opportunity to farmers for crop planning and to enhance crop production. The outcome from this research will be helpful to policy makers in formulating polices for enhancement of social and economic development.

1.6. Data Description

In this study, the variable under study is pest and disease data of Bt. cotton crop (average number of pest on 3 leaves selected randomly on cotton plants) along with standard meteorological weekly (SMW) weather data viz., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were used. The duration of the collected data are from 2008-09 to 2012-13 for Bt cotton crop from different centres. The pest chosen were Aphids at two centres (Akola and Vadodra) and Jassids at six centres viz. Akola, Banswara, Faridkot, Guntur, Perambalur and Vadodra. The data from different centres were divided in to two sets, the first one were used for model building as training data set and data from the last 12 observations were used for validation of model as testing data set.

1.7. Statistical Software packages used

Data analysis and programming codes for proposed methodologies were developed using different R packages.

1) tscount
2) forecast
3) lmtest
4) tseries

1.8. Significance of Research

This study gives an insight in the direction for relief and opportunity to farmers for minimizing their cost of cultivation by optimizing use of insecticides and enhancing their income from the produce. This will also act as an instrument in enhancing social welfare, economic development and providing opportunity to farmers in crop planning.

Chapter II

Prediction of pest dynamics using generalized Linear Models (INGARCH) with exogenous variables

2.1 Introduction

Count time series appear naturally in various areas and are the result of a process of measuring the number of discrete events over some period of time. Typically, these models assume that the process that generates the events is independent of time (t). This means that they are memory less. The time between events are assumed independent and exponentially distributed and most of the practical data violates this assumption. Examples showing the wide range of applications are the daily number of hospital admissions from public health, the number of stock market transactions per minute, the hourly number of defective items from industrial quality control, daily Insect/ Pest attack etc. Models for count time series should take into account that the observations are non-negative integers and they should capture suitably the dependence among observations.

A convenient and flexible approach is to employ the generalized linear model (GLM) methodology (Nelder and Wedderburn 1972) for modeling the observations conditionally on the past information. The time series of count data follows either poisson or negative binomial process. The count time series with dependent variable follows poisson process is termed as integer valued generalized conditional autoregressive (INGARCH) model of order p q. These models are also known as autoregressive conditional Poisson (ACP) models. These models were discussed by Heinen (2003), Ferland *et al*. (2006) and Fokianos *et al*. (2009).

2.2 INGARCH Model

GLM estimators are maximum likelihood estimators that are based on a density in the linear exponential family (LEF). These include the normal (Gaussian) and inverse Gaussian for continuous data, Poisson and Negative binomial for count data, Bernoulli for binary data (including logit and probit) and Gamma for duration data. GLM models follows the distributions which are other than Normal distributions. Let us denote the count time series by $\{Y_t : t \in N\}$ and time varying r-dimensional covariate vector say $\{X_t : t \in N\}$ i.e. $X_t = (X_{t,1,} \ldots, X_{t,r,})^T$. The conditional mean becomes $E\left(Y_t \,|F_{t-1}\right) = \lambda_t$ and F_t is historical data. The generalized model form is expressed as follows;

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^{p} \beta_k \tilde{g}\ (Y_{t-i_k}) + \sum_{l=1}^{q} \alpha_l g\ (\lambda_{t-jl}) + \eta^T$$

Where, g is link function, $\tilde{g}$ is transformation function, $g(\lambda_t)$ is linear predictor and η^T is parameter vector. To allow for regression on arbitrary past observations of the response, $P = \left\{i_{1,}\, i_{2,}\, \ldots, i_{p,}\right\}$ and 0<$0 < i_1 < i_2 < \cdots < i_p < \infty$ for leads to lagged observations $Y_{t-i_1}, \ldots, Y_{t-i_p}$Set $Q = \{j_{1,} j_{2,}\, \ldots, j_{q,}\}$ and 0 < $0 < j_1 < j_2 < \cdots < j_p < \infty$. The set Q lagged in parameter mean i.e. $\lambda_{t-i_1}, \ldots, \lambda_{t-i_p}$. Specification of the model order, i.e., of the sets P and Q, are guided by considering the empirical autocorrelation functions of the observed data. This approach is described for ARMA models in many time series analysis literatures. General class of linear models that are made up of 3 components: Random, Systematic, and Link Function. Random component identifies dependent variable (Y) and its probability distribution. Systematic Component identifies the set of explanatory variables (X_1,…, X_k). Link Function identifies a function of the mean that is a linear function of the explanatory variables and describes how the mean, depends on the linear predictor.

Cases of GLM:

Case 1: Consider the situation where g and $\tilde{g}$ are equal to identity i.e. $g(x)$= $\tilde{g}(x) = x$, further P={1,…,p}, Q={1,…,q} and $\eta = 0$ then the GLM model becomes poisson model as follows;

$$\lambda_t = \beta_0 + \sum_{k=1}^{p} \beta_k\ Y_{t-i_k} + \sum_{l=1}^{q} \alpha_l \lambda_{t-jl}$$

Assuming further that $Y_t|Y_{t-1}$ is Poisson distributed, then we obtain an INGARCH model of order p and q, abbreviated as INGARCH (p, q). These models are also known as autoregressive conditional Poisson (ACP) models (Heinen 2003, Ferland *et al.* 2006 and Fokianos, *et al.* 2009).

Case 2: The Negative Binomial distribution allows for a conditional variance to be larger than the mean $\lambda_t$which is often referred to as over-dispersion (with over dispersion parameter $\emptyset$) (Christou and Fokianos 2014). It is assumed that $Y_t|F_{t-1} \sim NegBionom(\lambda_t, \emptyset)$. When $\emptyset \rightarrow \infty$. The Poisson distribution is a limiting case of the Negative Binomial

$$\mathrm{P}\left(Y_t = y|\mathcal{F}_{t-1}\right) = \frac{\Gamma(\phi + y)}{\Gamma(y+1)\Gamma(\phi)}\left(\frac{\phi}{\phi + \lambda_t}\right)^{\phi}\left(\frac{\lambda_t}{\phi + \lambda_t}\right)^{y}, \quad y = 0, 1, \ldots.$$

2.3. INGARCH-X model

The standard INGARCH model allows to make forecasts based only on the past values of the forecast variable. The model assumes that future values of a variable linearly depend on its past values, as well as on the values of past exogenous variables. The INGARCHX model is an extended version of the INGARCH model. It also includes other independent (predictor) variables. The model is also referred to as the vector INGARCH or the dynamic regression model. The INGARCHX model is like a multivariate regression model but allows to take advantage of autocorrelation that may be present in residuals of the regression to improve the accuracy of a forecast.

2.4. Data description

In this study the variable under study is pest and disease data of Bt. cotton crop (average number of pest on 3 leaves selected randomly on cotton plants) along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were used. The duration of the collected data are from 2008-09 to 2012-13 for Bt cotton crop from different centers. The pest chosen were Aphids at two centers (Akola and Vadodra) and Jassids at six centers viz. Akola, Banswara, Faridkot, Guntur, Perambluru and Vadodra**.** The data from different centers were divided in to two sets, the first one were used for model building as training data set and data from the last 12 observations were used for validation of model as testing data set.

2.5. Results and Discussion

2.5.1. Results of Aphids of Akola centre

Aphids counts of cotton data (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Aphids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 1). Maximum number of pests is 57 and minimum is zero, and coefficient of variation (CV %) is 102 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.1.1).

Table 2.5.1.1: Summary statistics of No. of Aphid and weather variables of Akola centre

	No. of Aphids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	14.68	32.87	21.07	30.00	84.82	62.04
Standard Error	1.38	0.36	0.30	4.44	0.92	1.00
Kurtosis	0.03	37.15	-0.45	4.56	27.62	0.18
Skewness	0.94	4.62	-0.48	2.17	-3.87	0.29
Minimum	0.00	24.05	12.50	0.00	8.71	35.80
Maximum	57.43	64.40	27.31	218.00	99.14	98.50
CV (%)	102.21	11.78	15.65	160.64	11.80	17.43

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.1.2 depicts the parameter estimation of INGARCH model for Aphids pest as the coefficient is significant (P=0.009). After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated as p-value is found to be 0.024. As explained in methodology section, we developed INGARCH-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 2.5.1.3 depicts the parameter estimation of INGARCH-X model. Parameter for exogenous variables are all non-significant as probability of significance is >0.05 except model coefficient which is significant (P=0.010). Residuals of fitted model are also significant.

Table 2.5.1.2: INGARCH model specifications for No. of Aphid of Akola centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	4.615	1.54	3.004	0.003	56.85	<0.0001	5.120	0.024
β	0.647	.025	2.588	0.009				

Table 2.5.1.3: INGARCH-X model specifications for Aphid of Akola centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.481	3.316	0.145	0.885	56.85	<0.0001	11.063	<0.001
β	0.494	0.192	2.572	0.010				
MAXT	0.007	0.060	0.114	0.909				
MINT	0.064	0.094	0.684	0.494				
RAIN	0.002	0.006	0.261	0.794				
MAX_RH	-0.009	0.025	-0.341	0.733				
MIN_RH	-0.001	0.030	-0.028	0.978				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.1.4 and 2.5.1.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model in both training and testing data set. Possible reasons for this performance could be non-significance of exogenous variables it means exogenous variables have no linear relationship with Aphids count.

Table 2.5.1.4: Model performance in training data set for Aphid of Akola centre

Criteria’s	INGARCH	INGARCH-X
MAE	6.65	7.09
MSE	106.29	117.75
RMSE	10.31	10.85

Table 2.5.1.5: Model performance in testing data set for Aphid of Akola centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
41	34	23.37	8.68
42	38	19.73	4.97
43	26	17.38	3.91
44	30	15.86	3.55
45	29	14.87	3.42
46	25	14.24	3.37
47	26	13.82	3.35
48	26	13.56	3.34
49	25	13.38	3.34
50	21	13.27	3.34
MAE		12.11	23.93
MSE		155.02	587.93
RMSE		12.45	24.25

2.5.2. Results of Aphids of Vadodra centre

Aphids counts of cotton data for Vadodra centre (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Aphids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.2.1). Maximum number of pests is 66 and minimum is zero, and coefficient of variation (CV %) is 119.80 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.2.1).

Table 2.5.2.1: Summary statistics of No. of Aphid and weather variables of Vadodara centre

	No. of Aphid	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	15.69	32.27	18.80	10.44	76.18	45.70
Standard Error	1.56	0.25	0.46	2.77	0.95	1.47
Kurtosis	-0.34	-0.50	-1.36	14.83	-0.56	0.20
Skewness	0.94	-0.36	-0.09	3.91	-0.25	0.92
Minimum	0.00	23.10	7.50	0.00	44.66	18.66
Maximum	65.75	37.32	27.00	173.80	95.46	97.30
CV (%)	119.80	9.25	29.88	320.09	15.06	38.91

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.2.2 depicts the parameter estimation of INGARCH model for Aphids pest as the coefficient is significant (P=0.006). After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated as p-value is found to be 0.643. As explained in methodology section, we developed INGARCH-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 2.5.2.3 depicts the parameter estimation of INGARCH-X model. Parameter for exogenous variables are all non-significant as probability of significance is >0.05 except model coefficient which is significant (P=0.001). Residual test of fitted model is also found to be insignificant, hence residuals are white noise.

Table 2.5.2.2: INGARCH model specifications for No. of Aphid of Vadodara centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	1.919	0.664	2.890	0.004	105.92	<0.0001	0.214	0.643
β	0.873	0.319	2.733	0.006				

Table 2.5.2.3: INGARCH-X model specifications for Aphid of Vadodara centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.357	6.174	0.058	0.954	105.92	<0.0001	105.92	0.184
β	0.801	0.240	3.336	0.001				

MAXT	0.039	0.162	0.240	0.810				
MINT	-0.018	0.109	-0.169	0.866				
RAIN	0.002	0.010	0.214	0.830				
MAX_RH	-0.012	0.047	-0.264	0.792				
MIN_RH	0.005	0.039	0.117	0.907				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.2.4 and 2.5.2.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model in both training and testing data set. Possible reasons for this performance could be non-significance of exogenous variables it means exogenous variables have no linear relationship with Aphids count.

Table 2.5.2.4: Model performance in training data set for Aphid of Vadodara centre

	INGARCH	INGARCH-X
MAE	5.32	24.43
MSE	64.89	786.15
RMSE	8.06	28.04

Table 2.5.2.5: Model performance in testing data set for Aphid of Vadodara centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
46	27	32	25
47	27	30	20
48	31	28	16
49	29	27	14
50	44	25	12
51	45	24	11
52	52	23	11
1	60	22	10
2	51	21	10
3	36	20	10
MAE		16.77	26.28
MSE		433.94	919.52
RMSE		20.83	30.32

2.5.3. Results of Jassids of Akola centre

Jassids count of cotton data for Akola centre (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.3.1). Maximum number of pests is 1 and minimum is zero, and coefficient of variation (CV %) is 92.61%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.3.1).

Table 2.5.3.1: Summary statistics of No. of Jassids and weather variables of Akola centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.80	32.87	21.07	30.00	84.82	62.04
Standard Error	0.15	0.36	0.30	4.44	0.92	1.00
Kurtosis	2.79	15.00	10.87	2321.77	100.30	116.98
Skewness	1.54	37.15	-0.45	4.56	27.62	0.18
Minimum	0.00	4.62	-0.48	2.17	-3.87	0.29
Maximum	1.23	24.05	12.50	0.00	8.71	35.80
CV (%)	92.61	11.78	15.65	160.63	11.80	17.43

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.3.2 depicts the parameter estimation of INGARCH model for Jassids pest as the coefficient is significant (P < 0.0001). After model fitting, diagnostic checking of residuals is done and residuals are found to be uncorrelated as p-value is found to be 0.042. As explained in methodology section, we developed INGARCH-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 2.5.3.3 depicts the parameter estimation of

INGARCH-X model. Parameter for exogenous variables are all non-significant as probability of significance is >0.05 except model coefficient which is significant (P=0.001) and MINT (P=0.035). Residual test of fitted model is also found to be significant, hence residuals are not white noise.

Table 2.5.3.2: INGARCH model specifications for No. of Jassids of Akola centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.274	0.093	2.961	0.003	77.799	<0.0001	4.119	0.042
β	0.838	0.076	11.072	<0.0001				

Table 2.5.3.3: INGARCH-X model specifications for Jassids of Akola centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	-2.165	1.308	-1.656	0.098	77.799	<0.0001	28.423	<0.0001
β	0.929	0.127	7.292	<0.0001				
MAXT	0.018	0.015	1.185	0.236				
MINT	0.063	0.030	2.112	0.035				
RAIN	-0.001	0.002	-0.590	0.555				
MAX_RH	-0.001	0.009	-0.151	0.880				
MIN_RH	0.002	0.010	0.195	0.846				
SSH	-0.024	0.050	-0.486	0.627				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.3.4 and 2.5.3.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model on training data set, but not performed well on testing data set. Possible reasons for this performance could be non-significance of exogenous variables it means exogenous variables have no linear relationship with Jassids count.

Table 2.5.3.4: Model performance in training data set for Jassids of Akola centre

	INGARCH	INGARCH-X
MAE	0.54	0.85
MSE	0.80	1.50
RMSE	0.90	1.22

Table 2.5.3.5: Model performance in testing data set for Jassids of Akola centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
41	1	1.11	0.32
42	1	1.21	0.22
43	1	1.29	0.15
44	1	1.35	0.14
45	1	1.41	0.13
46	0	1.45	0.13
47	0	1.49	0.13
48	0	1.53	0.13
49	0	1.55	0.13
50	0	1.58	0.13
MAE		0.73	0.55
MSE		0.70	0.40
RMSE		0.84	0.63

2.5.4. Results of Jassids of Banswara centre

Jassids count of cotton data for Banswara (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), minimum relative humidity (MIN_RH) and SSH were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.4.1). Maximum number of pests is 9 and minimum is zero, and coefficient of variation (CV %) is 75.33%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.4.1).

Table 2.5.4.1: Summary statistics of No. of Jassids and weather variables of Banswara centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH	SSH
Mean	3.38	32.62	21.85	34.26	81.47	53.38	6.01
Standard Error	0.25	0.23	0.42	5.80	0.74	1.87	0.29
Kurtosis	-0.95	-0.42	-0.30	10.42	1.07	-1.23	-1.47
Skewness	0.24	-0.10	-0.95	2.86	-1.19	-0.34	-0.15
Minimum	0.00	26.80	10.70	0.00	58.00	16.00	0.10
Maximum	9.10	38.80	28.40	368.20	91.00	85.00	9.90
CV (%)	75.33	7.18	19.54	173.40	9.32	35.88	49.22

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.4.2 depicts the parameter estimation of INGARCH model for Jassids pest as the coefficient is significant (P < 0.0001). After model fitting, diagnostic checking of residuals is done and residuals are found to be white noise as p-value is found to be 0.969. As explained in methodology section, we developed INGARCH-X model for Jassids pest dynamic prediction by incorporating exogenous variables. Table 2.5.4.3 depicts the parameter estimation of INGARCH-X model. Parameter for exogenous variables are all non-significant as probability of significance is >0.05 except model coefficient which is significant (P=0.001). Residual test of fitted model is also found to be significant, hence residuals are not white noise.

Table 2.5.4.2: INGARCH model specifications for No. of Jassids of Banswara centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.463	0.146	3.169	0.002	64.91	<0.0001	0.002	0.969
β	0.837	0.080	10.495	<0.0001				

Table 2.5.4.3: INGARCH-X model specifications for Jassids of Banswara centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	-3.950	3.039	-1.300	0.194	64.91	<0.0001	8.95	0.003
β	0.854	0.118	7.211	<0.0001				
MAXT	-0.027	0.063	-0.431	0.667				
MINT	0.054	0.046	1.175	0.240				
RAIN	-0.001	0.001	-0.439	0.660				
MAX_RH	0.040	0.027	1.460	0.144				
MIN_RH	0.002	0.010	0.204	0.838				
SSH	0.034	0.051	0.674	0.500				

Table 2.5.4.4: Model performance in training data set for Jassids of Banswara centre

	INGARCH	INGARCH-X
MAE	1.04	1.16
MSE	1.95	2.35
RMSE	1.40	1.53

Table 2.5.4.5: Model performance in testing data set for Jassids of Banswara centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
36	4	3.8	0.1
37	4	3.7	0.1
38	5	3.5	0.0
39	5	3.4	0.0
40	4	3.3	0.0
41	4	3.2	0.0
42	3	3.2	0.0
43	3	3.1	0.0
44	2	3.1	0.0
45	2	3.0	0.0
MAE		0.73	3.34
MSE		0.83	12.23
RMSE		0.91	3.50

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.4.4 and 2.5.4.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model on training as well as testing data sets. Possible reasons for this performance could be non-significance of exogenous variables it means exogenous variables have no linear relationship with Jassids count.

2.5.5. Results of Jassids of Faridkot centre

Jassids count of cotton data for Faridkot Centre (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz.*, maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.5.1). Maximum number of pests is 5 and minimum is zero, and coefficient of variation (CV %) is 77.43%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.5.1).

Table 2.5.5.1: Summary statistics of No. of Jassids and weather variables of Faridkot centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.60	32.68	19.13	12.07	92.01	44.63
Standard Error	0.13	0.40	0.75	3.14	1.14	1.96
Kurtosis	-0.81	-0.17	-0.88	18.27	5.27	-1.11
Skewness	0.22	-0.57	-0.62	3.97	-2.23	0.23
Minimum	0.00	22.40	4.50	0.00	52.00	16.00
Maximum	4.90	39.60	28.20	183.10	100.00	85.00
CV (%)	77.43	11.16	36.11	236.58	11.40	40.44

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.5.2 depicts the parameter estimation of INGARCH model for Jassids pest as the coefficient is significant (P < 0.0001). After model fitting, diagnostic checking of residuals is done and residuals are found to be white noise as p-value is estimated to be 0.59. As explained in methodology section, we developed INGARCH-X model for Jassids pest dynamic prediction by incorporating exogenous variables. Table 2.5.5.3 depicts the parameter estimation of INGARCH-X model. Parameters for exogenous variables like rain, Max_RH and MIN_RH are found to be significant as p-value is < 0.05, however, remaining variables are insignificant. Residual test of fitted model is also found to be insignificant (P value=0.211), hence residuals are white noise.

Table 2.5.5.2: INGARCH model specifications for No. of Jassids of Faridkot centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.456	0.142	3.204	0.001	37.595	<0.0001	0.27871	0.59
β	0.735	0.104	7.079	<0.0001				

Table 2.5.5.3: INGARCH-X model specifications for Jassids of Faridkot centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	-0.556	1.955	-0.284	-0.556	37.595	<0.0001	1.563	0.211
β	0.616	0.158	3.906	0.616				
MAXT	-0.067	0.060	-1.123	-0.067				
MINT	0.120	0.045	2.696	0.120				
RAIN	0.000	0.003	-0.140	0.000				
MAX_RH	0.007	0.012	0.546	0.007				
MIN_RH	-0.008	0.010	-0.807	-0.008				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.5.4 and 2.5.5.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model on training data set, but, on testing data set, performance of INGARCH-X model is found to be better.

Table 2.5.5.4: Model performance in training data set for Jassids of Faridkot centre

	INGARCH	INGARCH-X
MAE	0.596	0.715
MSE	0.670	0.882
RMSE	0.818	0.939

Table 2.5.5.5: Model performance in testing data set for Jassids of Faridkot centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
41	2	1.93	1.35
42	2	1.87	1.13
43	2	1.83	0.97
44	1	1.80	0.91
45	1	1.78	0.87
46	1	1.76	0.86
47	1	1.75	0.84
48	1	1.74	0.84
49	1	1.74	0.84
50	0	1.73	0.83
MAE		0.73	0.45
MSE		0.71	0.34
RMSE		0.84	0.58

2.5.6. Results of Jassids of Guntur centre

Jassids count of cotton data for Guntur Centre (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.6.1). Maximum number of pests is 5 and minimum is zero, and coefficient of variation (CV %) is 63.19%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.6.1).

Table 2.5.6.1: Summary statistics of No. of Jassids and weather variables of Guntur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.50	32.88	21.06	28.51	85.69	62.13
Standard Error	0.09	0.36	0.32	4.52	0.69	1.04
Kurtosis	0.27	41.85	-0.45	5.55	-0.29	0.18
Skewness	0.65	5.19	-0.47	2.35	-0.43	0.31
Minimum	0.00	26.70	12.50	0.00	66.85	35.80
Maximum	4.58	64.40	27.31	218.00	99.14	98.50
CV (%)	63.19	11.64	15.77	166.24	8.41	17.54

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.6.2 depicts the parameter estimation of INGARCH model for Jassids pest as the coefficient is significant (P < 0.0001). After model fitting, diagnostic checking of residuals is done and residuals are found to be white noise as p-value is estimated to be 0.299. As explained in methodology section, we developed INGARCH-X model for Jassids pest dynamic prediction by incorporating exogenous variables. Table 2.5.6.3 depicts the parameter estimation of INGARCH-X model. Parameters for all exogenous variables are found to be insignificant as p-value is >0.05, except model coefficient which is significant (P=0.001). Residual test of fitted model is also found to be significant (P value=0.0002), hence residuals are not white noise.

Table 2.5.6.2: INGARCH model specifications for No. of Jassids of Guntur centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.465	0.162	2.858	0.004	40.537	<0.0001	1.076	0.299
β	0.677	0.121	5.581	<0.0001				

Table 2.5.6.3: INGARCH-X model specifications for Jassids of Guntur centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	-0.576	2.016	-0.286	0.775	40.537	<0.0001	13.667	0.0002
β	0.697	0.202	3.445	0.001				
MAXT	-0.008	0.028	-0.288	0.774				
MINT	0.039	0.036	1.073	0.283				
RAIN	-0.003	0.003	-1.060	0.289				
MAX_RH	-0.011	0.017	-0.616	0.538				
MIN_RH	0.012	0.011	1.126	0.260				

Table 2.5.6.4: Model performance in training data set for Jassids of Guntur centre

	INGARCH	INGARCH-X
MAE	0.565	0.528
MSE	0.518	0.556
RMSE	0.720	0.746

Table 2.5.6.5: Model performance in testing data set for Jassids of Guntur centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
42	1	1.14	0.91
43	1	1.24	0.91
44	3	1.30	0.88
45	3	1.35	0.88
46	3	1.38	0.87
47	1	1.40	0.87
48	1	1.41	0.87
49	2	1.42	0.87
50	2	1.43	0.87
1	3	1.43	0.87
MAE		0.65	1.04
MSE		0.42	1.07
RMSE		0.65	1.04

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table

2.5.6.4 and 2.5.6.5. Based on reported MAE, MSE and RMSE, one can interpret that INGARCH model performed better as compare to INGARCH-X model on testing data set, but, on training data set, performance of INGARCH model is found to be slightly better.

2.5.7. Results of Jassids of Perambalur centre

Jassids count of cotton data for Perambalur Centre (average number of pest in 3 leaves selected randomly of Aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz.*, maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), and maximum relative humidity (MAX_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 2.5.7.1). Maximum number of pests is 4 and minimum is zero, and coefficient of variation (CV %) is 66.69%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 2.5.7.1).

Table 2.5.7.1: Summary statistics of No. of Jassids and weather variables of Perambalur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH
Mean	1.30	30.64	22.29	22.67	73.65
Standard Error	0.08	0.26	0.16	3.82	1.57
Kurtosis	0.06	0.90	0.66	5.58	21.70
Skewness	0.69	0.15	-0.34	2.41	2.30
Minimum	0.00	22.70	16.50	0.00	8.26
Maximum	3.90	39.00	26.00	188.00	184.80
CV (%)	66.69	8.83	7.46	174.55	22.12

As explained in methodology section, INGARCH and INGARCH-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INGARCH modeling. Table 2.5.7.2 depicts the parameter estimation of INGARCH model for Jassids pest as the coefficient is significant (P

< 0.0001). After model fitting, diagnostic checking of residuals is done and residuals are found to be white noise as p-value is estimated to be 0.385. As explained in methodology section, we developed INGARCH-X model for Jassids pest dynamic prediction by incorporating exogenous variables. Table 2.5.7.3 depicts the parameter estimation of INGARCH-X model. Parameters for all exogenous variables are found to be insignificant as p-value is >0.05, except MINT which is border line significant (P=0.049). Residual test of fitted model is also found to be significant (P value=0.002), hence residuals are not white noise.

Table 2.5.7.2: INGARCH model specifications for No. of Jassids of Perambalur centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	0.295	0.144	2.051	0.040	47.758	<0.0001	0.754	0.385
β	0.767	0.124	6.175	<0.0001				

Table 2.5.7.3: INGARCH-X model specifications for Jassids of Perambalur centre

Parameters	Estimate	S.E.	Z	Prob.	Box-Pierce Non-Correlation Test			
					For original series		For residuals	
					χ^2	Prob.	χ^2	Prob.
Intercept	-0.763	1.585	-0.482	0.630	47.758	<0.0001	9.587	0.002
β	0.948	0.262	3.619	<0.0001				
MAXT	-0.073	0.043	-1.699	0.089				
MINT	0.123	0.062	1.968	0.049				
RAIN	0.001	0.003	0.240	0.810				
MAX_RH	-0.004	0.007	-0.543	0.587				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 2.5.7.4 and 2.5.7.5. Based on reported MAE, MSE and RMSE, one can interpret that

INGARCH model performed better as compare to INGARCH-X model on training as well as testing data sets.

Table 2.5.7.4: Model performance in training data set for Jassids of Perambalur centre

	INGARCH	INGARCH-X
MAE	0.434	0.476
MSE	0.319	0.404
RMSE	0.565	0.636

Table 2.5.7.5: Model performance in testing data set for Jassids of Perambalur centre

SMW (2012-13)	Actual	Forecast	
		INGARCH	INGARCH-X
41	0	0.30	0.47
42	0	0.52	0.47
43	0	0.70	0.67
44	1	0.83	0.67
45	1	0.93	0.76
46	1	1.01	0.76
47	2	1.07	0.80
48	1	1.12	0.80
49	1	1.15	0.81
50	1	1.18	0.81
MAE		0.25	0.23
MSE		0.09	0.12
RMSE		0.30	0.35

2.6. Conclusion

Based on reported MAE, MSE and RMSE, one can interpret that INGARCHX model performed better as compare to INGARCH model in training data set for some center, but, INGRACH performed better under testing data set and for some centers it was vice versa. In this section, we obtained mixed results.

Chapter III

Development of Integer-Valued Auto Regressive model with exogenous variables for pest dynamics prediction

3.1. Introduction

Integer valued autoregressive model for first order, INAR (1) was first introduced by McKenzie (1985, 1988) and Al-Osh and Alzaid (1987) independently for modelling and forecasting the sequences of dependent counting process.

Integer valued ARMA (INAR) models are discrete analogues of the (standard, real valued) ARMA model. The INAR model is given by

$$X_t = \sum_{i=1}^{p} \alpha_i \, o \, X_{t-i} + \mathcal{E}_t$$

Where, $\mathcal{E}_t$is independent and identically distributed random variables with $\mathrm{E}[\mathcal{E}_t] = 0 \; and \; Var \; [\mathcal{E}_t] = \sigma^2$ which follows Gaussian distribution and 'o' denotes thinning operator.

The binomial thinning operator implies $\alpha \, o \, X$ arises from X by Binomial thinning operation. The main properties of binomial thinning operator are

$$(\alpha \, \mathrm{o} X \mid X = x) \sim B(x, \alpha)$$

$$\alpha \, \mathrm{o} X \leq X$$

McKenzie (1985) developed integer-valued auto-regressive model of first order *i.e.* INAR(1) model independently and the model is formally given by the following equation

$$X_t = \alpha * X_{t-1} + \varepsilon_t \quad t = 0, 1, 2, \ldots$$

Here $0 < \alpha < 1$ and $\{\varepsilon_t\}$ is independently and identically distributed integer-valued random variables with $E(\varepsilon_t) = \mu_\varepsilon$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$.

The conditional mean and variance of the INAR (1) model is given by

$$E(X_t \mid X_{t-1}) = E\{Bin(X_{t-1}, \alpha) + Po(\lambda)\} = \alpha X_{t-k} + \lambda$$

$$Var(X_t \mid X_{t-1}) = \alpha(1-\alpha) X_{t-1} + \lambda$$

Where λ is a Poisson parameter. The exogenous variables will be incorporated to develop INARX model and the parameters will be estimated by maximizing the likelihood function. For this different thinning operators (o) will be tried to implement in INAR model inclusion of explanatory variables into the INAR model may extend the applicability of INAR model which greatly extends the range of time series data sets for which INAR can be applied. Roy *et al*. (2016) have applied the INAR model in pest population dynamics studies in agriculture.

3.2. INAR-X model

Our aim for INARX modelling is to extend the analysis of INAR models to incorporate explanatory variables viz. temperature, humidity and rainfall which are correlated with the infestation of pest attacks. That is, we look to increase the flexibility of INAR models whilst maintaining the AR structure of the model. In particular, explanatory variables can be used to model a (linear) trend or periodicity as well as other covariates which may affect the outcome of the time series data. Unlike standard AR processes for INAR processes trends and periodicity cannot easily be removed by transforming the original time series since any transformation would need to preserve the integer nature of the data. Therefore if there are trends and periodicity in the data these have to be incorporated in the modelling of the data with explanatory variables being a natural way of including such information. The work by Branna¨s (1995), incorporated explanatory variables into an INAR model where he considers an INAR (1) model only.

Suppose that for each time point there are r explanatory variables. For t ∈ Z and i = 1, 2, . . . r, let wt,i denote the value of the ith explanatory variable at time t and let $w_t = (w_{t,0}, w_{t,1}, \ldots, w_{t,r})$ where $w_{t,0}$ = 1 for all t ∈ Z. Let p denote the maximum AR order of the model and for j = 1, 2, . . . p let $\delta_t = (\delta_{j,0}, \delta_{j,1}, \ldots, \delta_{j,r})$. Let $\gamma_t = (\gamma_0, \gamma_1, \ldots, \gamma_r)$. Then for t ∈ Z, the INAR(p) model with explanatory variables is given by

$$X_t = \sum_{j=1}^{p} \alpha_{t,j} \circ X_{t-j} + Z_t$$

where Zt ∼ P o(λt), $\alpha_{t,j} = \{1 + \exp(w_t^T \delta_j)\}$-1 and $\lambda_t = \exp(w_t^T \gamma)\}$. The special case where p = 1 was considered in Brannas (1995). In Brannas (1995), separate explanatory variables were used for $\alpha_{t,1}$ and λ_t and is covered by the current set up by fixing some of the components of δ_j and γ equal to 0. The model defined above is the full model. It will often be the case that a simpler model which does not include all the explanatory variables or all the AR terms will suffice. Therefore we assume that there exists $R \subseteq \{1, 2, \ldots, r\}$ and $A \subseteq \{1, 2, \ldots, p\}$ such that for $i \notin A$, $\alpha_{t,i} = 0$ and for $j \notin R$, $\delta_{i,j} = 0$ and $\gamma_j = 0$, where A and R are unknown and are parameters in the model to be estimated.

Finally, it is important to note that the explanatory variables can be used to model a linear trend or periodicity. Unlike standard AR processes for INAR processes trends and periodicity cannot easily be removed by transforming the original time series since any transformation would need to preserve the integer nature of the data. Therefore if there are trends and periodicity in the data these can be incorporated through explanatory variables.

3.3. Data description

Suppose $\{X_t\}$ $t = 0, 1, 2, \ldots, n$ be a count-data time series with a finite range $\{0, \ldots, n\}$ of counts, where $n \in N = \{1, 2, \ldots\}$ is known and the series has serial dependence similar to the Gaussian autoregressive (AR) process. If the marginal distribution follows binomial distribution i.e. B(n, p) where $p \in (0; 1)$ is called Binomial AR (1) model, was first proposed by McKenzie (1985). Particular case of the binomial AR(1) model for describing binomial counts with a first-order autoregressive serially correlated structure was described by Weiß and Kim (2013). Asymptotic distribution of the conditional least-squares estimators of the parameters of the binomial AR(1) model were also discussed.

3.4. Results and Discussion

3.4.1. Results of Aphids of Akola center

In this illustration, aphids counts of cotton data (no. of aphids on cotton plants) along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 from ICAR- National Research Centre for Integrated pest management (ICAR-NCIPM) under NICRA scheme. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Regardless of the study, descriptive statistics for no. of aphids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.1.1). Considering the values of skewness and kurtosis, one can decipher that, the data under consideration follows positively skewed with symmetrical kurtosis, maximum number of pests are 57 and minimum are zero, and coefficient of variation (CV %) is 102 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.1.1).

Table 3.4.1.1: Summary statistics of No. of Aphid and weather variables of Akola center

	No. of Aphids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	14.68	32.87	21.07	30.00	84.82	62.04
Standard Error	1.38	0.36	0.30	4.44	0.92	1.00
Kurtosis	0.03	37.15	-0.45	4.56	27.62	0.18
Skewness	0.94	4.62	-0.48	2.17	-3.87	0.29
Minimum	0.00	24.05	12.50	0.00	8.71	35.80
Maximum	57.43	64.40	27.31	218.00	99.14	98.50
CV (%)	102.21	11.78	15.65	160.64	11.80	17.43

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.1.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.1.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.1.2: Parameter specifications of INAR for Aphids in Akola center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.763	56.85	<0.0001	0.85	0.356
lambda	14.66				

Table 3.4.1.3: Parameter specifications of INAR-X for Aphids in Akola centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.315	56.85	<0.0001	27.31	<0.0001
lambda	8.989				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.1.4 and 3.4.1.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.1.4: Model performance in training data set for Aphid of Akola center

	INAR	INAR-X
MAE	4.723	7.98
MSE	92.416	136.95
RMSE	9.613	11.70

Table 3.4.1.5: Model performance in testing data set for Aphid of Akola center

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
41	34	43	18
42	38	47	20
43	26	20	21
44	30	37	17
45	29	32	18
46	25	25	14
47	26	30	12
48	26	29	17
49	25	27	17
50	21	20	17
MAE		9.92	10.25
MSE		99.39	123.68
RMSE		9.96	11.12

3.4.2. Results of Aphid of Vadodra centre

Aphids counts of cotton data for Vadodra centre (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data viz., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of aphids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.2.1). Maximum number of pests is 66 and minimum is zero, and coefficient of variation (CV %) is 119.80 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.2.1).

Table 3.4.2.1: Summary statistics of No. of Aphid and weather variables of Vadodara centre

	No. of Aphid	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	15.69	32.27	18.80	10.44	76.18	45.70
Standard Error	1.56	0.25	0.46	2.77	0.95	1.47
Kurtosis	-0.34	-0.50	-1.36	14.83	-0.56	0.20
Skewness	0.94	-0.36	-0.09	3.91	-0.25	0.92
Minimum	0.00	23.10	7.50	0.00	44.66	18.66
Maximum	65.75	37.32	27.00	173.80	95.46	97.30
CV (%)	119.80	9.25	29.88	320.09	15.06	38.91

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.2.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.2.3

depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.2.2: Parameter specifications of INAR for Aphids in Vadodara center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.888	105.92	<0.0001	2.03	0.154
Lambda	1.755				

Table 3.4.2.3: Parameter specifications of INAR-X for Aphid in Vadodara centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.272	105.92	<0.0001	94.80	<0.0001
Lambda	5.625				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.2.4 and 3.4.2.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.2.4: Model performance in training data set for Aphid of Vadodara centre

	INAR	INARX
MAE	3.77	12.55
MSE	47.20	236.75
RMSE	6.87	15.38

Table 3.4.2.5: Model performance in testing data set for Aphid of Vadodara centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
46	27	21	15
47	27	28	13
48	31	36	13
49	29	29	20
50	44	60	14
51	45	49	18
52	52	62	18
1	60	72	20
2	51	47	22
3	36	25	20
MAE		5.89	21.66
MSE		61.85	563.42
RMSE		7.86	23.74

3.4.3. Results of Jassids of Akola centre

Jassids count of cotton data for Akola centre (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data viz., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.3.1). Maximum number of pests is 1 and minimum is zero, and coefficient of variation (CV %) is 92.61%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.3.1).

Table 3.4.3.1: Summary statistics of No. of Jassids and weather variables of Akola centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.80	32.87	21.07	30.00	84.82	62.04
Standard Error	0.15	0.36	0.30	4.44	0.92	1.00
Kurtosis	2.79	15.00	10.87	2321.77	100.30	116.98
Skewness	1.54	37.15	-0.45	4.56	27.62	0.18
Minimum	0.00	4.62	-0.48	2.17	-3.87	0.29
Maximum	1.23	24.05	12.50	0.00	8.71	35.80
CV (%)	92.61	11.78	15.65	160.63	11.80	17.43

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.3.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.3.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.3.2: Parameter specifications of INAR for Jassids in Akola center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.864	77.799	<0.0001	4.624	0.032
Lambda	0.240				

Table 3.4.3.3: Parameter specifications of INAR-X for Jassid in Akola centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.110	77.799	<0.0001	77.89	<0.0001
Lambda	2.094				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.3.4 and 3.4.3.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set.

Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.3.4: Model performance in training data set for Jassids of Akola centre

	INAR	INARX
MAE	0.44	1.235
MSE	0.74	2.395
RMSE	0.86	1.547

Table 3.4.3.5: Model performance in testing data set for Jassids of Akola centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
41	1	0.9	2.2
42	1	0.9	2.2
43	1	0.9	2.2
44	1	0.9	2.2
45	1	0.9	2.2
46	0	0	1.9
47	0	0	1.6
48	0	0	2.1
49	0	0	2.1
50	0	0	2.1
MAE		0.41	1.62
MSE		0.49	2.78
RMSE		0.07	1.67

3.4.4. Results of Jassids of Banswara centre

Jassids count of cotton data for Banswara (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), minimum relative humidity (MIN_RH) and SSH were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.4.1). Maximum number of pests is 9 and minimum is zero, and coefficient of variation (CV %) is 75.33%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.4.1).

Table 3.4.4.1: Summary statistics of No. of Jassids and weather variables of Banswara centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH	SSH
Mean	3.38	32.62	21.85	34.26	81.47	53.38	6.01
Standard Error	0.25	0.23	0.42	5.80	0.74	1.87	0.29
Kurtosis	-0.95	-0.42	-0.30	10.42	1.07	-1.23	-1.47
Skewness	0.24	-0.10	-0.95	2.86	-1.19	-0.34	-0.15
Minimum	0.00	26.80	10.70	0.00	58.00	16.00	0.10
Maximum	9.10	38.80	28.40	368.20	91.00	85.00	9.90
CV (%)	75.33	7.18	19.54	173.40	9.32	35.88	49.22

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.4.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.4.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.4.2: Parameter specifications of INAR for Jassids in Banswara center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.833	64.91	<0.0001	0.589	0.442
Lambda	0.572				

Table 3.4.4.3: Parameter specifications of INAR-X for Jassid in Banswara centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.073	64.91	<0.0001	53.70	<0.0001
Lambda	2.693				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.4.4 and 3.4.4.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.4.4: Model performance in training data set for Jassids of Banswara centre

	INAR	INARX
MAE	0.832	1.907
MSE	1.477	5.402
RMSE	1.216	2.324

Table 3.4.4.5: Model performance in testing data set for Jassids of Banswara centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
36	4	4.1	3.0
37	4	4.1	3.0
38	5	6.1	3.0
39	5	5.3	3.1
40	4	3.3	3.1
41	4	4.1	3.0
42	3	2.1	2.5
43	3	2.9	2.2
44	2	0.9	2.2
45	2	1.8	1.8
MAE		2.75	0.98
MSE		0.42	1.35
RMSE		0.65	1.16

3.4.5. Results of Jassids of Faridkot centre

Jassids count of cotton data for Faridkot Centre (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data viz., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.5.1). Maximum number of pests is 5 and minimum is zero, and coefficient of variation (CV %) is 77.43%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.5.1).

Table 3.4.5.1: Summary statistics of No. of Jassids and weather variables of Faridkot centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.60	32.68	19.13	12.07	92.01	44.63
Standard Error	0.13	0.40	0.75	3.14	1.14	1.96
Kurtosis	-0.81	-0.17	-0.88	18.27	5.27	-1.11
Skewness	0.22	-0.57	-0.62	3.97	-2.23	0.23
Minimum	0.00	22.40	4.50	0.00	52.00	16.00
Maximum	4.90	39.60	28.20	183.10	100.00	85.00
CV (%)	77.43	11.16	36.11	236.58	11.40	40.44

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.5.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.5.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.5.2: Parameter specifications of INAR for Jassids in Faridkot center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.729	37.595	<0.0001	1.745	0.187
Lambda	0.439				

Table 3.4.5.3: Parameter specifications of INAR-X for Jassid in Faridkot centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.034	37.595	<0.0001	21.81	<0.0001
lambda	2.173				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.5.4 and 3.4.5.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.5.4: Model performance in training data set for Jassids of Faridkot centre

	INAR	INARX
MAE	0.498	0.843
MSE	0.662	1.163
RMSE	0.814	1.078

Table 3.4.5.5: Model performance in testing data set for Jassids of Faridkot centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
41	2	2.1	2.2
42	2	2.1	2.2
43	2	2.1	2.2
44	1	0.1	2.2
45	1	0.8	2.2
46	1	0.8	2.2
47	1	0.8	2.2
48	1	0.8	2.2
49	1	0.8	2.2
50	0	0.0	2.2
MAE		0.98	1.08
MSE		0.11	1.58
RMSE		0.33	1.25

3.4.6. Results of Jassids of Guntur centre

Jassids count of cotton data for Guntur Centre (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH), and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.6.1). Maximum number of pests is 5 and minimum is zero, and coefficient of variation (CV %) is 63.19%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.6.1).

Table 3.4.6.1: Summary statistics of No. of Jassids and weather variables of Guntur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.50	32.88	21.06	28.51	85.69	62.13
Standard Error	0.09	0.36	0.32	4.52	0.69	1.04
Kurtosis	0.27	41.85	-0.45	5.55	-0.29	0.18
Skewness	0.65	5.19	-0.47	2.35	-0.43	0.31
Minimum	0.00	26.70	12.50	0.00	66.85	35.80
Maximum	4.58	64.40	27.31	218.00	99.14	98.50
CV (%)	63.19	11.64	15.77	166.24	8.41	17.54

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.6.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.6.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.6.2: Parameter specifications of INAR for Jassids in Guntur center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.623	40.537	<0.0001	0.096	0.757
Lambda	0.573				

Table 3.4.6.3: Parameter specifications of INAR-X for Jassid in Guntur centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.101	40.537	<0.0001	25.92	<0.0001
Lambda	2.262				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.6.4 and 3.4.6.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.6.4: Model performance in training data set for Jassids of Guntur centre

	INAR	INARX
MAE	0.475	0.877
MSE	0.433	1.161
RMSE	0.659	1.078

Table 3.4.6.5: Model performance in testing data set for Jassids of Guntur centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
42	1	0.8	2.4
43	1	0.8	2.4
44	3	4.8	2.4
45	3	3.6	2.6
46	3	3.6	1.8
47	1	0.0	1.8
48	1	0.8	2.4
49	2	2.8	2.4
50	2	2.2	2.5

1	3	4.2	2.5
MAE		1.30	0.77
MSE		0.66	0.77
RMSE		0.81	0.87

3.4.7. Results of Jassids of Permbluru centre

Jassids count of cotton data for Permbluru Centre (average number of pest in 3 leaves selected randomly of aphids on cotton plants) per three leaves along with standard meteorological weekly (SMW) weather data viz., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), and maximum relative humidity (MAX_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 3.4.7.1). Maximum number of pests is 4 and minimum is zero, and coefficient of variation (CV %) is 66.69%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 3.4.7.1).

Table 3.4.7.1: Summary statistics of No. of Jassids and weather variables of Perambalur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH
Mean	1.30	30.64	22.29	22.67	73.65
Standard Error	0.08	0.26	0.16	3.82	1.57
Kurtosis	0.06	0.90	0.66	5.58	21.70
Skewness	0.69	0.15	-0.34	2.41	2.30
Minimum	0.00	22.70	16.50	0.00	8.26
Maximum	3.90	39.00	26.00	188.00	184.80
CV (%)	66.69	8.83	7.46	174.55	22.12

As explained in methodology section INAR and INAR-X models were fitted to aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for INAR modeling. Table 3.4.7.2 depicts the parameter estimation of INAR model for Aphids pest. After model fitting, diagnostic checking of residuals is done and residuals are found to be autocorrelated, it means some information is present in the residuals. As explained in methodology section, we developed INAR-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 3.4.7.3 depicts the parameter estimation of INAR-X model. Residuals of fitted INAR-X model are also significant.

Table 3.4.7.2: Parameter specifications of INAR for Jassids in Perambalur center

	INAR	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.726	47.758	<0.0001	0.19	0.66
Lambda	0.353				

Table 3.4.7.3: Parameter specifications of INAR-X for Jassids in Perambalur centre

	INARX	Box-Pierce Non-Correlation Test			
		For original series		For residuals	
		χ^2	Prob.	χ^2	Prob.
Parameter	0.354	47.758	<0.0001	27.61	<0.0001
lambda	0.939				

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 3.4.7.4 and 3.4.7.5. Based on reported MAE, MSE and RMSE once can interpret that INAR model performed better compared INAR-X model in training data set and in testing data set. Based on the result obtained, one can say that INAR model has better forecasting efficiency for out of sample forecast in this data set.

Table 3.4.7.4: Model performance in training data set for Jassids of Perambalur centre

	INAR	INARX
MAE	0.329	0.676
MSE	0.295	0.588
RMSE	0.543	0.767

Table 3.4.7.5: Model performance in testing data set for Jassids of Perambalur centre

SMW (2012-13)	Actual	Forecast	
		INAR	INARX
41	0	0.0	0.9
42	0	0.0	1.5
43	0	0.0	1.5
44	1	1.6	1.5
45	1	0.9	1.7
46	1	0.9	1.7
47	2	2.9	1.7
48	1	0.2	1.9
49	1	0.9	1.7
50	1	0.9	1.3
MAE		0.67	0.77
MSE		0.16	0.73
RMSE		0.40	0.86

3.5. Comparison of forecasting performance

The Root Mean Square Error (RMSE) mean absolute percentage error (MAE) has been computed to compare the forecasting performance of all the models under considerations in both training and validation data set for both pests in different centers separately.

Conclusion

The Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAE) has been computed to compare the forecasting performance of all the models under considerations in both training and validation data set for both pests in different centers separately. Based on the results obtained one can interpret that, INAR model outperformed the INARX model in both training and testing data set.

Chapter IV

Development of Integer based neural network with exogenous variables for pest dynamic prediction

4.1. Introduction

Artificial neural networks (ANNs) are nonlinear model that are able to capture various nonlinear structures present in the data set. One significant advantage of the ANN models over other classes of nonlinear models is that ANNs are universal approximators that can approximate a large class of functions with a high degree of accuracy. ANN model specification does not require prior assumption of the data generating process, instead it is largely depending on characteristics of the data. Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting. The model is characterized by a network of three layers of simple processing units connected by a cyclic links. The relationship between the output (y_t) and the inputs (y_{t-1}, …, y_{t-p}) is expressed as follows;

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right) + \varepsilon_t$$

Where, $\alpha_j (j = 0,1,2, \ldots\ldots, q)$ and $\beta_{ij} (i = 0,1,2, \ldots\ldots, p,\ j = 0,1,2, \ldots\ldots, q)$ are the model parameters often called the connection weights, p is the number of input nodes and q is the number of hidden nodes.

The logistic function $g(x) = \frac{1}{1+\exp(-x)}$ is often used as the hidden layer activation function. Along with logistic function, we will try other activation function like bipolar logistic, tanh, wavelet activation function etc. Data normalization is often performed before the training process begins. When nonlinear transfer functions are used at the output nodes, the desired output values must be transformed to the range of the actual outputs of the network. Even if a linear output transfer function is used, it may still advantageous to standardize the outputs as well as the inputs to avoid computational problems, to meet algorithm requirement and to facilitate network learning. In general data normalization is beneficial in terms of classification rate and mean squared errors, but the benefit diminishes as network and sample size increase. In addition, data normalization usually slows down the training process. Normalization of the output values (targets) is usually independent of the normalization of the inputs. For time series modelling problems, however, the normalization of targets is typically performed together with the inputs. The choice of range to which inputs and targets are normalized depends largely on the activation function of output nodes, with typically [0, 1] for logistic function and [-1, 1] for hyperbolic tangent function. It should be noted that, as a result of normalizing the target values, the observed output of the network should be corresponding to the normalized range. Thus, to interpret the results obtained from the network, the outputs must be rescaled to the original range. From the user's point of view, the accuracy obtained by ANNs should be based on the rescaled data sets. Performance measures should also be calculated on the rescaled outputs.

Training and test sample are typically required for building an ANN model. The training sample is used for ANN model development and test sample is adopted for evaluating the prediction ability of the model. Sometimes a third one called the validation sample is also utilized to avoid the overfitting problem or to determine the stopping point of the training process. It is common to use one test set for both validation and testing purposes particularly with small data sets. In view of the fact, the selection of the training and test sample may affect the performance of ANNs.

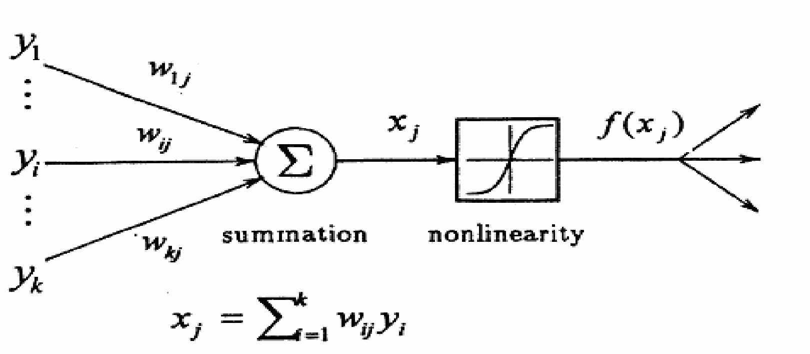

Fig.1: Architecture of Neural Network

The main issue here is to divide the data into the training and test sets. Although there is no solution to this problem, several factors such as the problem characteristics, the data type and the size of the available data should be considered in making the decision. Most researchers select the training and test sets based on the rule of 90% vs. 10%, 80% vs. 20% or 70% vs. 30%. The amount of data for the network training depends on the network structure, the training method and the complexity of the particular problem or the amount of noise in the data on hand. The ANN modeling efficiency increases as the training sample size increases. Using artificial neural network approach, one of the weather variable i.e. rainfall will be forecasted and will be utilized in conjunction with transfer function model for time series crop yield forecast.

Kumari *et al.* (2013) forecasted the pigeonpea productivity and pod damage by Helicoverpa armigera using artificial neural network model and it has been inferred that Levenberg- Marquardt algorithm gave the best performance in the prediction of damage and productivity of long duration pigeonpea for NEPZ in India for the year. Liu *et al*. (2013) carried out research on prediction about Fruit Tree Diseases and Insect Pests Based on Neural Network. Huang *et al.* (2010) explained development of soft computing and applications in agricultural and biological engineering. Draghici (2002) studied the capabilities of neural networks using limited precision weights. Bhagawati *et al*. (2015) gave weather based plant disease forecasting system using artificial neural network. Yang *et al*. (2009) developed prediction model for population occurrence of paddy stem borer based on back propagation, artificial neural network and principal components analysis.

4.2. IANN-X model

The standard IANN model allows to make forecasts based only on the past values of the forecast variable. The model assumes that future values of a variable linearly depend on its past values, as well as on the values of past exogenous variables. The IANNX model is an extended version of the IANN model. It also includes other independent (predictor) variables. The model is also referred to as the vector IANN model.

4.3. Data description

In this study the variable under study is pest and disease data of Bt. cotton crop (average number of pest on 3 leaves selected randomly on cotton plants) along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were used. The duration of the collected data are from 2008-09 to 2012-13 for Bt cotton crop from different centers. The pest chosen were Aphids at two centers (Akola and Vadodra) and Jassids at six centers viz. Akola, Banswara, Faridkot, Guntur, Perambalur and Vadodra**.** The data from different centers were divided in to two sets, the first one were used for model building as training data set and data from the last 12 observations were used for validation of model as testing data set.

4.4. Results and Discussion

4.4.1. Results of Aphids of Akola center

In this illustration, Aphids counts of cotton data (average number of pest in 3 leaves selected randomly) of Aphids on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Aphids count and weather variables are ascertained to comprehend the nature of data under consideration and the results are presented in Table 4.4.1.1. Considering the values of skewness and kurtosis, one can decipher that the data under consideration follows positively skewed with symmetrical kurtosis, maximum number of pests are 57 and minimum are zero, and coefficient of variation (CV %) is 102 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.1.1.

Table 4.4.1.1: Summary statistics of No. of Aphid and weather variables of Akola center

	No. of Aphids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	14.68	32.87	21.07	30.00	84.82	62.04
Standard Error	1.38	0.36	0.30	4.44	0.92	1.00
Kurtosis	0.03	37.15	-0.45	4.56	27.62	0.18
Skewness	0.94	4.62	-0.48	2.17	-3.87	0.29
Minimum	0.00	24.05	12.50	0.00	8.71	35.80
Maximum	57.43	64.40	27.31	218.00	99.14	98.50
CV (%)	102.21	11.78	15.65	160.64	11.80	17.43

As explained in methodology section IANN and IANN-X models were fitted to Aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Table 4.4.1.2 depicts the parameter estimation of IANN model for Aphids pest. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.1.2), based on the lowest RMSE values. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.1.3 depicts the parameter estimation of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.1.3) based on lowest RMSE values. Residuals of fitted IANN-X model are also non-significant.

Table 4.4.1.2: IANN (2,2,1) model specifications for Aphid of Akola center

Particulars	Specifications
Input lags	2
Hidden nodes	2
Output nodes	1
No. of weights	9
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.488, Prob=0.485

Table 4.4.1.3: IANN-X (2,4,1) model parameters specifications for Aphid of Akola center

Particulars	Specifications
Input lags	2[8]
Hidden nodes	4
Output nodes	1
No. of weights	41
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.894, Prob=0.344

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4.4.1.4 and 4.4.1.5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-

X model performed better compared IANN model in both training and testing data set. Possible reasons for this performance could be inclusion of exogenous variables, it means exogenous variables have non-linear relationship with Aphids count.

Table 4.4.1.4: Model performance in training data set for Aphid of Akola center

	IANN	IANN-X
MAE	5.82	1.61
MSE	88.30	5.32
RMSE	9.40	2.31

Table 4.4.1.5: Model performance in testing data set for Aphid of Akola center

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
41	34	24	27
42	38	19	28
43	26	14	25
44	30	11	24
45	29	10	14
46	25	9	9
47	26	9	6
48	26	9	5
49	25	8	9
50	21	8	5
MAE		15.96	12.86
MSE		264.13	203.58
RMSE		16.25	14.27

4.4.2. Results of Aphids of Vadodra centre

In this illustration, Aphids counts of cotton data (average number of pest in 3 leaves selected randomly) of Aphids on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Regardless of the study, descriptive statistics for no. of Aphids count and weather variables are ascertained to comprehend the nature of data under consideration (Table 4.4.2.1). Maximum number of pests is 66 and minimum is zero, and coefficient of variation (CV %) is 119.80 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory (Table 4.4.2.1).

Table 4.4.2.1: Summary statistics of No. of Aphid and weather variables of Vadodara centre

	No. of Aphid	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	15.69	32.27	18.80	10.44	76.18	45.70
Standard Error	1.56	0.25	0.46	2.77	0.95	1.47
Kurtosis	-0.34	-0.50	-1.36	14.83	-0.56	0.20
Skewness	0.94	-0.36	-0.09	3.91	-0.25	0.92
Minimum	0.00	23.10	7.50	0.00	44.66	18.66
Maximum	65.75	37.32	27.00	173.80	95.46	97.30
CV (%)	119.80	9.25	29.88	320.09	15.06	38.91

As explained in methodology section IANN and IANN-X models were fitted to Aphids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Table 4.4.2.2 depicts the parameter estimation of IANN model for Aphids pest. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.2.2), based on the lowest RMSE values. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.896. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.2.3 depicts the parameter estimation of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.2.3) based on lowest RMSE values. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.485.

Table 4.4.2.2: IANN (4,2,1)model specifications for Aphid of Vadodara centre

Particulars	Specifications
Input lags	4
Hidden nodes	2
Output nodes	1
No. of weights	13
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.017, Prob=0.896

Table 4.4.2.3: IANN-X (2,6,1) model parameters specifications for Aphid of Vadodara centre

Particulars	Specifications
Input lags	2 [7]
Hidden nodes	6
Output nodes	1
No. of weights	55
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.488, Prob=0.485

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4 and 5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed slightly better as compare to IANN model in both training and testing data set. Possible reasons for this performance could be inclusion of exogenous variables.

Table 4.4.2.4: Model performance in training data set for Aphid of Vadodara centre

	IANN	IANN-X
MAE	24.16	24.06
MSE	792.77	798.30
RMSE	28.16	28.25

Table 4.4.2.5: Model performance in testing data set for Aphid of Vadodara centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
46	27	46	43
47	27	51	46
48	31	52	51
49	29	53	34
50	44	52	27
51	45	51	23
52	52	49	19
1	60	48	22
2	51	47	24
3	36	46	20
MAE		13.15	21.23
MSE		231.57	533.74
RMSE		15.22	23.10

4.4.3. Results of Jassids of Akola centre

Jassids count data (average number of pest in 3 leaves selected randomly) on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids and weather variables are presented in Table 4.4.3.1. Maximum number of Jassids is 1 and minimum is zero, and coefficient of variation (CV %) is 92.61 %, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.3.1.

Table 4.4.3.1: Summary statistics of No. of Jassids and weather variables of Akola centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.80	32.87	21.07	30.00	84.82	62.04
Standard Error	0.15	0.36	0.30	4.44	0.92	1.00
Kurtosis	2.79	15.00	10.87	2321.77	100.30	116.98
Skewness	1.54	37.15	-0.45	4.56	27.62	0.18
Minimum	0.00	4.62	-0.48	2.17	-3.87	0.29
Maximum	1.23	24.05	12.50	0.00	8.71	35.80
CV (%)	92.61	11.78	15.65	160.63	11.80	17.43

As explained in methodology section IANN and IANN-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Table 4.4.3.2 depicts the parameter estimation of IANN model for Jassids pest. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.3.2), based on the lowest RMSE values given in Table 4. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.878. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.3.3 depicts the parameter estimation of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.3.3) based on lowest RMSE values given in Table 4. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.251.

Table 4.4.3.2: IANN (2,2,1) model specifications for Jassids of Akola centre

Particulars	Specifications
Input lags	2
Hidden nodes	2
Output nodes	1
No. of weights	9
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.0.023, Prob=0.878

Table 4.4.3.3: IANN-X (2,4,1) model specifications for Jassids of Akola centre

Particulars	Specifications
Input lags	2[8]
Hidden nodes	4
Output nodes	1
No. of weights	41
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=1.319, Prob=0.251

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4 and 5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed better as compare to IANN model in training but in testing data set, performance of IANN is found to be better.

Table 4.4.3.4: Model performance in training data set for Jassids of Akola centre

	IANN	IANN-X
MAE	0.55	0.34
MSE	0.74	0.20
RMSE	0.86	0.45

Table 4.4.3.5: Model performance in testing data set for Jassids of Akola centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
41	1	1	1
42	1	1	1
43	1	1	1
44	1	1	2
45	1	1	3
46	0	1	2
47	0	1	2
48	0	1	2
49	0	1	2
50	0	1	2
MAE		0.39	1.19
MSE		0.21	1.93
RMSE		0.45	1.39

4.4.4. Results of Jassids of Banswara centre

Jassids count data (average number of pest in 3 leaves selected randomly) on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz.*, maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids and weather variables are presented in Table 4.4.4.1. Maximum number of Jassids is 9 and minimum is zero, and coefficient of variation (CV %) is 75.33%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.4.1.

Table 4.4.4.1: Summary statistics of No. of Jassids and weather variables of Banswara centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	3.38	32.62	21.85	34.26	81.47	53.38
Standard Error	0.25	0.23	0.42	5.80	0.74	1.87
Kurtosis	-0.95	-0.42	-0.30	10.42	1.07	-1.23
Skewness	0.24	-0.10	-0.95	2.86	-1.19	-0.34
Minimum	0.00	26.80	10.70	0.00	58.00	16.00
Maximum	9.10	38.80	28.40	368.20	91.00	85.00
CV (%)	75.33	7.18	19.54	173.40	9.32	35.88

As explained in methodology section IANN and IANN-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Table 4.4.4.2 depicts the parameter estimation of IANN model for Jassids pest. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.4.2), based on the lowest RMSE values given in Table 4. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.882. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.4.3 depicts the parameter estimation of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.4.3) based on lowest RMSE values given in Table 4. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.480.

Table 4.4.4.2: IANN (2,2,1) model specifications for Jassids of Banswara centre

Particulars	Specifications
Input lags	7
Hidden nodes	4
Output nodes	1
No. of weights	37
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.022, Prob=0.882

Table 4.4.4.3: IANN-X (2,4,1) model parameters specifications for Jassids of Banswara centre

Particulars	Specifications
Input lags	2[8]
Hidden nodes	7
Output nodes	1
No. of weights	71
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.498, Prob=0.480

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4.4.4.4 and 4.4.4.5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed better as compare to IANN model in training but in testing data set, performance of IANN is found to be better.

Table 4.4.4.4: Model performance in training data set for Jassids of Banswara centre

	IANN	IANN-X
MAE	0.46	0.24
MSE	0.39	0.10
RMSE	0.63	0.31

Table 4.4.4.5: Model performance in testing data set for Jassids of Banswara centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
36	4	3	5
37	4	4	4
38	5	3	6
39	5	3	7
40	4	4	8
41	4	2	5
42	3	3	5
43	3	3	4
44	2	1	5
45	2	3	3
MAE		0.97	1.83
MSE		1.43	4.53
RMSE		1.20	2.13

4.4.5. Results of Jassids of Faridkot centre

Jassids count data (average number of pest in 3 leaves selected randomly) on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids and weather variables are presented in Table 4.4.5.1. Maximum number of Jassids is 5 and minimum is zero, and coefficient of variation (CV %) is 77.43%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.5.1.

Table 4.4.5.1: Summary statistics of No. of Jassids and weather variables of Faridkot centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.60	32.68	19.13	12.07	92.01	44.63
Standard Error	0.13	0.40	0.75	3.14	1.14	1.96
Kurtosis	-0.81	-0.17	-0.88	18.27	5.27	-1.11
Skewness	0.22	-0.57	-0.62	3.97	-2.23	0.23
Minimum	0.00	22.40	4.50	0.00	52.00	16.00
Maximum	4.90	39.60	28.20	183.10	100.00	85.00
CV (%)	77.43	11.16	36.11	236.58	11.40	40.44

As explained in methodology section, IANN and IANN-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Table 4.4.5.2 depicts the number of parameters under consideration of IANN model for Jassids pest. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.5.2), based on the lowest RMSE values given in Table 4.4.5.4. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.485. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.5.3 depicts the parameter estimation of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.5.3) based on lowest RMSE values given in Table 4. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.960.

Table 4.4.5.2: IANN (2,5,1) model specifications for Jassids of Faridkot centre

Particulars	Specifications
Input lags	2
Hidden nodes	5
Output nodes	1
No. of weights	21
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.488, Prob=0.485

Table 4.4.5.3: IANN-X (2,5,1) model parameters specifications for Jassids of Faridkot centre

Particulars	Specifications
Input lags	2 [7]
Hidden nodes	5
Output nodes	1
No. of weights	46
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.002, Prob=0.960

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4 and 5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed better as compare to IANN model in training as well as testing data set.

Table 4.4.5.4: Model performance in training data set for Jassids of Faridkot centre

	IANN	IANN-X
MAE	0.499	0.196
MSE	0.504	0.074
RMSE	0.710	0.271

Table 4.4.5.5: Model performance in testing data set for Jassids of Faridkot centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
41	2	2	2
42	2	2	2
43	2	2	2
44	1	2	2
45	1	2	2
46	1	2	1
47	1	2	1
48	1	2	1
49	1	2	1
50	0	2	1
MAE		0.89	0.39
MSE		1.06	0.21
RMSE		1.03	0.46

4.4.6. Results of Jassids of Guntur centre

Jassids count data of Guntur center (average number of pest in 3 leaves selected randomly) on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids and weather variables are presented in Table 4.4.6.1. Maximum number of Jassids is 5 and minimum is zero, and coefficient of variation (CV %) is 63.19%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.6.1.

Table 4.4.6.1: Summary statistics of No. of Jassids and weather variables of Guntur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH	MIN_RH
Mean	1.50	32.88	21.06	28.51	85.69	62.13
Standard Error	0.09	0.36	0.32	4.52	0.69	1.04
Kurtosis	0.27	41.85	-0.45	5.55	-0.29	0.18
Skewness	0.65	5.19	-0.47	2.35	-0.43	0.31
Minimum	0.00	26.70	12.50	0.00	66.85	35.80
Maximum	4.58	64.40	27.31	218.00	99.14	98.50
CV (%)	63.19	11.64	15.77	166.24	8.41	17.54

As explained in methodology section, IANN and IANN-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.6.2), based on the lowest RMSE values given in Table 4.4.6.4. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.457. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.6.3 depicts the parametric specifications of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.6.3) based on lowest RMSE values given in Table 4.4.6.4. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.485.

Table 4.4.6.2: IANN (1,1,1) model specifications for Jassids of Guntur centre

Particulars	Specifications
Input lags	1
Hidden nodes	1
Output nodes	1
No. of weights	4
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.552, Prob=0.457

Table 4.4.6.3: IANN-X (1,4,1) model parameters specifications for Jassids of Guntur centre

Particulars	Specifications
Input lags	1 [6]
Hidden nodes	4
Output nodes	1
No. of weights	33
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.488, Prob=0.485

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4 and 5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed better as compare to IANN model in training as well as testing data sets.

Table 4.4.6.4: Model performance in training data set for Jassids of Guntur centre

	IANN	IANN-X
MAE	0.539	0.281
MSE	0.487	0.139
RMSE	0.698	0.373

Table 4.4.6.5: Model performance in testing data set for Jassids of Guntur centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
42	1	1	2
43	1	1	2
44	3	1	2
45	3	1	2
46	3	1	1
47	1	1	1
48	1	1	1
49	2	1	1
50	2	1	1
1	3	1	1
MAE		0.93	0.89
MSE		0.86	0.80
RMSE		0.93	0.89

4.4.7. Results of Jassids of Perambalur centre

Jassids count data of Permbluru centre (average number of pest in 3 leaves selected randomly) on cotton plants per three leaves along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), and maximum relative humidity (MAX_RH) were collected from 2008-09 to 2012-13 for Bt cotton crop from different centers. The data from 31st SMW 2008-09 to 40th SMW 2012-13 were used for model building as training data set and data from 41st SMW 2012-13 to 50th SMW were used for validation of model as testing data set.

Descriptive statistics for no. of Jassids and weather variables are presented in Table 4.4.7.1. Maximum number of Jassids is 4 and minimum is zero, and coefficient of variation (CV %) is 66.69%, it means data under consideration is highly heterogeneous. The summary statistics for weather variables are self-explanatory and are presented in Table 4.4.7.1.

Table 4.4.7.1: Summary statistics of No. of Jassids and weather variables of Perambalur centre

	No. of Jassids	MAXT	MINT	RF	MAX_RH
Mean	1.30	30.64	22.29	22.67	73.65
Standard Error	0.08	0.26	0.16	3.82	1.57
Kurtosis	0.06	0.90	0.66	5.58	21.70
Skewness	0.69	0.15	-0.34	2.41	2.30
Minimum	0.00	22.70	16.50	0.00	8.26
Maximum	3.90	39.00	26.00	188.00	184.80
CV (%)	66.69	8.83	7.46	174.55	22.12

As explained in methodology section, IANN and IANN-X models were fitted to Jassids data, before model estimation one has to ensure that data under consideration is autocorrelated, as probability of significance for original series is <0.0001, the data is autocorrelated, so one can proceed for IANN and IANN-X modeling. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.7.2), based on the lowest RMSE values given in Table 4.4.7.4. After model fitting, diagnostic checking of residuals has been done and residuals are found to be non-autocorrelated on the basis of insignificant p-values i.e. 0.527. As explained in methodology section, we developed IANN-X model for Aphids pest dynamic prediction by incorporating exogenous variables. Table 4.4.7.3 depicts the parametric specifications of IANN-X model. Different combination of input lags and hidden nodes were tried, and optimum parameters were selected (Table 4.4.7.3) based on lowest RMSE values given in Table 4.4.7.4. Residuals of fitted IANN-X model are also non-significant i.e. p-value is 0.832.

Table 4.4.7.2: IANN (1,1,1) model specifications for Jassids of Perambalur centre

Particulars	Specifications
Input lags	1
Hidden nodes	1
Output nodes	1
No. of weights	4
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.400, Prob=0.527

Table 4.4.7.3: IANN-X (2,4,1) model parameters specifications for Jassids of Perambalur centre

Particulars	Specifications
Input lags	1[5]
Hidden nodes	3
Output nodes	1
No. of weights	22
I:H activation function	Sigmoidal
H:O activation function	Identity
Box-Pierce Non-Correlation Test for residuals	X-Sq.=0.045, Prob=0.832

Table 4.4.7.4: Model performance in training data set for Jassids of Perambalur centre

	IANN	IANN-X
MAE	0.429	0.293
MSE	0.308	0.150
RMSE	0.555	0.387

Table 4.4.7.5: Model performance in testing data set for Jassids of Perambalur centre

SMW (2012-13)	Actual	Forecast	
		IANN	IANN-X
41	0	0	1
42	0	1	3
43	0	1	2
44	1	1	2
45	1	1	2
46	1	1	1
47	2	1	1
48	1	1	1
49	1	1	2
50	1	1	1
MAE		0.33	0.89
MSE		0.14	1.30
RMSE		0.38	1.14

Further, performance of selected models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE has been depicted in table 4.4.7.4 and 4.4.7.5. Based on reported MAE, MSE and RMSE, one can interpret that IANN-

X model performed better as compare to IANN model in training data set, but, IANN performed better under training data set.

Conclusion:

Based on reported MAE, MSE and RMSE, one can interpret that IANN-X model performed better as compare to IANN model in training data set, but, IANN performed better under testing data set.

Chapter V

Comparative study of different models for pest dynamics predictions

5.1 Introduction

In this chapter, different models developed in the study are compared and discussed. Certain criteria's that are used to make comparison of modeling and forecasting ability among different models are as follows;

Mean squared error:

The mean square error (MSE) is the average of sum of squared error values and written as $MSE = \frac{\sum_{i=1}^{N}(Y_i - \widehat{Y_i})^2}{N}$

Root Mean squared error (RMSE):

The Square root of mean squared error which is also known as standard error of estimate in regression analysis or the estimated white noise standard deviation in time series model's analysis, which is expressed as follows;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \widehat{Y_i})^2}{N}}$$

Where, Y_i is the Actual value, $\hat{Y}_i$ is the predicted value and N is the number of observations.

Mean Absolute error (MAPE):

Mean absolute percentage error is another criterion to measure the performance of forecasting model and is written as:

$$MAE = \frac{1}{N}\left|Y_i - \hat{Y}_i\right|$$

Where, Y_i is the Actual value, $\hat{Y}_i$ is the predicted value and N is the number of observations.

5.2 Comparison of forecasting performance in training and testing data set

Results of different pests in different centers are discussed as under different sections for both training and testing data set.

5.2.1 Results of Aphid of Akola center

Performance of developed models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE have been depicted in Table 5.2.1.1 and 5.2.1.2. Based on reported MAE, MSE and RMSE one can interpret that integer based artificial neural network models performed better compared to other model in both training data set and INAR model outperformed all the models in testing data set. As, discussed in chapter II and III residuals of INGARCH, INGARCH-X, and INAR-X are significant and at the same time, residuals of IANN and IANN-X models are non-significant, it means artificial neural network models are good fit for Aphids of Akola center. Further, INAR model performed better in testing data set. Possible reasons for the better fit and better forecasting performance of IANN and IANN-X models could be generalization ability of integer based artificial neural network model to capture the complex and non-linear relationship present in the data.

Table 5.2.1.1: Model performance in training data set for Aphid of Akola centre

	INGARCH	INGARCH-X	IANN	IANN-X	INAR	INAR-X
MAE	6.65	7.09	5.82	1.61	4.723	7.98
MSE	106.29	117.75	88.30	5.32	92.416	136.95
RMSE	10.31	10.85	9.40	2.31	9.613	11.70

Table 5.2.1.2: Model performance in testing data set for Aphid of Akola centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCH-X	IANN	IANN-X	INAR	INAR-X
41	34	23.37	8.68	24	27	43	18
42	38	19.73	4.97	19	28	47	20
43	26	17.38	3.91	14	25	20	21
44	30	15.86	3.55	11	24	37	17
45	29	14.87	3.42	10	14	32	18
46	25	14.24	3.37	9	9	25	14
47	26	13.82	3.35	9	6	30	12
48	26	13.56	3.34	9	5	29	17
49	25	13.38	3.34	8	9	27	17
50	21	13.27	3.34	8	5	20	17
MAE		12.11	23.93	15.96	12.86	9.92	10.25
MSE		155.02	587.93	264.13	203.58	99.39	123.68
RMSE		12.45	24.25	16.25	14.27	9.96	11.12

5.2.2 Results of Aphid of Vadodra centre

Based on reported MAE, MSE and RMSE (Table 5.2.2.1 and 5.2.2.2) one can interpret that model without exogenous variables performed better as compared to model without exogenous variables such as INAR, INGARCH and INAR respectively. INAR model outperformed all the models in both training and testing data set. Possible reasons for the better fit and better forecasting performance of INAR model could be generalization ability of INAR model to relationship present in this data set. In this data set IANN and IANNX models not performed well, the reasons could be, for linear data set parametric models like INAR and INGARCHX are better fit as compared to integer based artificial neural network models.

Table 5.2.2.1: Model performance in training data set for Aphid of Vadodara centre

	INGARCH	INGARCH-X	IANN	IANN-X	INAR	INAR-X
MAE	5.32	24.43	24.16	24.06	3.77	12.55
MSE	64.89	786.15	792.77	798.30	47.20	236.75
RMSE	8.06	28.04	28.16	28.25	6.87	15.38

Table 5.2.2.2: Model performance in testing data set for Aphid of Vadodara centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCH-X	IANN	IANN-X	INAR	INAR-X
46	27	32	25	46	43	21	15
47	27	30	20	51	46	28	13
48	31	28	16	52	51	36	13
49	29	27	14	53	34	29	20
50	44	25	12	52	27	60	14
51	45	24	11	51	23	49	18
52	52	23	11	49	19	62	18
1	60	22	10	48	22	72	20
2	51	21	10	47	24	47	22
3	36	20	10	46	20	25	20
MAE		16.77	26.28	13.15	21.23	5.89	21.66
MSE		433.94	919.52	231.57	533.74	61.85	563.42
RMSE		20.83	30.32	15.22	23.10	7.86	23.74

5.2.3 Results of Jassids of Akola centre

Based on reported MAE, MSE and RMSE (Table 5.2.3.1 and 5.2.3.2) one can interpret that IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.3.1: Model performance in training data set for Jassids of Akola centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	0.54	0.85	0.55	0.34	0.44	1.235
MSE	0.80	1.50	0.74	0.20	0.74	2.395
RMSE	0.90	1.22	0.86	0.45	0.86	1.547

Table 5.2.3.2: Model performance in testing data set for Jassids of Akola centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
41	1	1.11	0.32	1	1	0.9	2.2
42	1	1.21	0.22	1	1	0.9	2.2
43	1	1.29	0.15	1	1	0.9	2.2
44	1	1.35	0.14	1	2	0.9	2.2
45	1	1.41	0.13	1	3	0.9	2.2
46	0	1.45	0.13	1	2	0	1.9
47	0	1.49	0.13	1	2	0	1.6
48	0	1.53	0.13	1	2	0	2.1
49	0	1.55	0.13	1	2	0	2.1
50	0	1.58	0.13	1	2	0	2.1
MAE		0.73	0.55	0.39	1.19	0.41	1.62
MSE		0.70	0.40	0.21	1.93	0.49	2.78
RMSE		0.84	0.63	0.45	1.39	0.07	1.67

5.2.4 Results of Jassids of Banswara centre

Based on reported MAE, MSE and RMSE (Table 5.2.4.1 and 5.2.4.2) one can interpret that IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.4.1: Model performance in training data set for Jassids of Banswara centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	1.04	1.16	0.46	0.24	0.832	1.907
MSE	1.95	2.35	0.39	0.10	1.477	5.402
RMSE	1.40	1.53	0.63	0.31	1.216	2.324

Table 5.2.4.2: Model performance in testing data set for Jassids of Banswara centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
36	4	3.8	0.1	3	5	4.1	3.0
37	4	3.7	0.1	4	4	4.1	3.0
38	5	3.5	0.0	3	6	6.1	3.0
39	5	3.4	0.0	3	7	5.3	3.1
40	4	3.3	0.0	4	8	3.3	3.1
41	4	3.2	0.0	2	5	4.1	3.0
42	3	3.2	0.0	3	5	2.1	2.5
43	3	3.1	0.0	3	4	2.9	2.2
44	2	3.1	0.0	1	5	0.9	2.2
45	2	3.0	0.0	3	3	1.8	1.8
MAE		0.73	3.34	0.97	1.83	2.75	0.98
MSE		0.83	12.23	1.43	4.53	0.42	1.35
RMSE		0.91	3.50	1.20	2.13	0.65	1.16

5.2.5 Results of Jassids of Faridkot centre

Based on reported MAE, MSE and RMSE (Table 5.2.5.1 and 5.2.5.2) one can interpret that IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.5.1: Model performance in training data set for Jassids of Faridkot centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	0.596	0.715	0.499	0.196	0.498	0.843
MSE	0.670	0.882	0.504	0.074	0.662	1.163
RMSE	0.818	0.939	0.710	0.271	0.814	1.078

Table 5.2.5.2: Model performance in testing data set for Jassids of Faridkot centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
41	2	1.93	1.35	2	2	2.1	2.2
42	2	1.87	1.13	2	2	2.1	2.2
43	2	1.83	0.97	2	2	2.1	2.2
44	1	1.80	0.91	2	2	0.1	2.2
45	1	1.78	0.87	2	2	0.8	2.2
46	1	1.76	0.86	2	1	0.8	2.2
47	1	1.75	0.84	2	1	0.8	2.2
48	1	1.74	0.84	2	1	0.8	2.2
49	1	1.74	0.84	2	1	0.8	2.2
50	0	1.73	0.83	2	1	0.0	2.2
MAE		0.73	0.45	0.89	0.39	0.98	1.08
MSE		0.71	0.34	1.06	0.21	0.11	1.58
RMSE		0.84	0.58	1.03	0.46	0.33	1.25

5.2.6 Results of Jassids of Guntur centre

Based on reported MAE, MSE and RMSE (Table 5.2.6.1 and 5.2.6.2) one can interpret that IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.6.1: Model performance in training data set for Jassids of Guntur centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	0.565	0.528	0.539	0.281	0.475	0.877
MSE	0.518	0.556	0.487	0.139	0.433	1.161
RMSE	0.720	0.746	0.698	0.373	0.659	1.078

Table 5.2.6.2: Model performance in testing data set for Jassids of Guntur centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
42	1	1.14	0.91	1	2	0.8	2.4
43	1	1.24	0.91	1	2	0.8	2.4
44	3	1.30	0.88	1	2	4.8	2.4
45	3	1.35	0.88	1	2	3.6	2.6

46	3	1.38	0.87	1	1	3.6	1.8
47	1	1.40	0.87	1	1	0.0	1.8
48	1	1.41	0.87	1	1	0.8	2.4
49	2	1.42	0.87	1	1	2.8	2.4
50	2	1.43	0.87	1	1	2.2	2.5
1	3	1.43	0.87	1	1	4.2	2.5
MAE		0.65	1.04	0.93	0.89	1.30	0.77
MSE		0.42	1.07	0.86	0.80	0.66	0.77
RMSE		0.65	1.04	0.93	0.89	0.81	0.87

5.2.7 Results of Jassids of Perambalur centre

Based on reported MAE, MSE and RMSE (Table 5.2.7.1 and 5.2.7..2) one can interpret that IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.7.1: Model performance in training data set for Jassids of Perambalur centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	0.434	0.476	0.429	0.293	0.329	0.676
MSE	0.319	0.404	0.308	0.150	0.295	0.588
RMSE	0.565	0.636	0.555	0.387	0.543	0.767

Table 5.2.7.2: Model performance in testing data set for Jassids of Perambalur centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
41	0	0.30	0.47	0	1	0.0	0.9
42	0	0.52	0.47	1	3	0.0	1.5
43	0	0.70	0.67	1	2	0.0	1.5
44	1	0.83	0.67	1	2	1.6	1.5
45	1	0.93	0.76	1	2	0.9	1.7
46	1	1.01	0.76	1	1	0.9	1.7
47	2	1.07	0.80	1	1	2.9	1.7
48	1	1.12	0.80	1	1	0.2	1.9
49	1	1.15	0.81	1	2	0.9	1.7
50	1	1.18	0.81	1	1	0.9	1.3
MAE		0.25	0.23	0.33	0.89	0.67	0.77
MSE		0.09	0.12	0.14	1.30	0.16	0.73
RMSE		0.30	0.35	0.38	1.14	0.40	0.86

5.2.8 Results of Jassids of ANGRAU centre

Based on reported MAE, MSE and RMSE (Table 5.2.8.1 and 5.2.8.2) one can interpret that IANNX model outperformed in training data set compared to all other models and INGARCH model outperformed all the model in testing data set. We can generalize that model performance is totally a data driven concept, different models performs differently in different data set.

Table 5.2.8.1: Model performance in training data set for Jassids of ANGRAU centre

	INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
MAE	0.56	0.53	0.54	0.29	3.707	8.024
MSE	0.52	0.56	0.49	0.15	20.949	135.10
RMSE	0.72	0.75	0.70	0.38	4.577	11.623

Table 5.2.8.2: Model performance in testing data set for Jassids of ANGRAU centre

SMW (2012-13)	Actual	Forecast					
		INGARCH	INGARCHX	IANN	IANNX	INAR	INARX
42	1	1.14	0.91	1	1	43.7	39.9
43	1	1.24	0.91	1	1	45.5	40.3
44	3	1.30	0.88	1	2	47.2	40.8
45	3	1.35	0.88	1	2	49	41.2
46	3	1.38	0.87	1	1	50.8	32.3
47	1	1.40	0.87	1	1	52.6	32.8
48	1	1.41	0.87	1	1	54.4	42.6
49	2	1.42	0.87	1	1	56.1	43
50	2	1.43	0.87	1	1	57.9	43.4
51	3	1.43	0.87	1	1	59.7	43.9
MAE		0.71	1.11	1.01	0.81	21.87	9.69
MSE		0.86	1.75	1.49	1.03	34.22	173.44
RMSE		0.93	1.32	1.22	1.01	5.85	13.16

Conclusion:

The study has been conducted to develop count time series models for modeling and forecasting pest dynamic prediction. As an illustration, the developed models have been developed in aphids and jassids of cotton pests at different centers of India. For, Aphid of Akola center Performance of developed models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE. Based on reported MAE, MSE and RMSE one can interpret that integer based artificial neural network models performed better compared to other model in both training data set and INAR model outperformed all the models in testing data set. For Aphid of Vadodra center INAR model outperformed all the models in both training and testing data set. For Jassids of Akola, Banswara, Faridkot, Guntur and Perambalur centers IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. For Jassids of ANGRAU center IANNX model outperformed in training data set compared to all other models and INGARCH model outperformed all the model in testing data set. Finally, we conclude that no models are performing better in all the training data set and in testing data set. Based on the results obtained in this study one can conclude that IANNX model outperformed all the models in training data set and INAR model outperformed all the models in testing data set.

सारांश

अत्यधिक लागत वाली और अनिश्चितताओं से भरी कृषि का किसानों की आजीविका पर बहुत गहरा प्रभाव पड़ता है । यदि समय पर जोखिम को कम करने के लिए उपाय नहीं किए जाते, तो वे दुष्चक्र के जाल में गिर सकते हैं। यही नहीं, कई राज्यों में किसान कई तरह की फसलों पर कीटों के हमले की बढ़ती घटनाओं से जूझ रहे हैं। फसलों पर कीटों के हमलों का खतरा अक्सर स्थानीय होता है, लेकिन मानसून की विफलता या फसल की कीमतों के उतार चड़ाव के कारण फसलों में कीटों और बीमारियों की घटनाओं ने कृषि को बहुत जोखिम भरा उद्यम बना दिया है। इतना ही नहीं अधिक खेती की लागत के कारण नई तकनीकों को अपनाने में किसान बहुत आशंकित रहते हैं। लगभग 15-प्रतिशत फसलों की पैदावार हर साल कीटों के हमलों के 25 कारण नष्ट हो जाती है। इन समस्याओं को कम करने के लिए, विश्वसनीय और समय पर पूर्वानुमान तकनीकिया एक महत्वपूर्ण और अत्यंत उपयोगी इनपुट प्रदान करती है। काउंट टाइम सीरीज़ में समय के लगातार बिंदुओं में घटनाएं घटित होती हैं, जो आमतौर पर कई स्थितियों में होती है, उदाहरण के लिए, एक सप्ताह में सड़क दुर्घटनाओं की संख्या, एक सप्ताह में अंकुरित होने वाले बीज की संख्या आदि। पूर्णांकमूल्यवान समय श्रृंखला - मूल्यवान समय श्रृंखला मॉडल का महत्वपूर्ण वर्ग।-है। असततINAR प्रक्रिया कई समय श्रृंखला के लिए अच्छी तरह से अनुकूल है जो पॉइसन, नकारात्मक द्विपद, सामान्यीकृत पॉइसन वितरण इत्यादि का अनुसरण करती है। नूयरल नेटवर्क मॉडल nonlinear और nonparametric होने के कारण गिनती समय श्रृंखला को अच्छी तरह से एस्टिमेट करत है।

समयसमय पर पेश किया गया। चूंकि मौसम संबंधी कारक -समय पर विभिन्न पद्धतियों को समय- बीमारियों के संक्रमण के लिए अत्यधिक / फसलों में कीटजिम्मेदार होते हैं, इसलिए, मौसम के मापदंडों के साथ INARX और IANN जैसे उन्नत मॉडल कीट बीमारी की स्थिति की जांच और भविष्यवाणी करने के / रोग के संक्रमण की प्रारंभिक चेतावनी के लिए उपयुक्त समाधान का पता लगा सकते हैं। इन / लिए कीट पृष्ठभूमि के साथINAR और पूर्णांक आधारित ANN मॉडल को बहिर्जात चर की जानकारी पर विचार करके कपास की फसल में मॉडलिंग और कीट गतिशीलता की भविष्यवाणी के लिए विकसित किया गया। यह आम तौर पर धारणा है कि पिछले आंकड़ों के लिए goodness of fit के बजाय out of sample पूर्वानुमानों का उपयोग करके आकलन किया जाना चाहिए। भविष्य के डेटा के संभावित व्यवहार को समझने के लिए, out of sample पूर्वानुमानों की आवश्यकता होती है। इस अध्ययन में इष्टतम आउट सैंपल फोरकास्ट के सूत्र-ऑफ- दिये गए हैं।

पिछले कुछ वर्षों में, विशेष रूप से समय श्रृंखला गणना डेटा के विश्लेषण के लिए लागू मॉडल के वर्ग का अध्ययन किया गया है। गणना परिणाम चर कभी ट्रांसफ़ॉर्म किए जाते हैं और-कभी लॉग-log प्रतिगमन का उपयोग करके विश्लेषण किया जाता है। इस दृष्टिकोण के साथ कई समस्याएं उत्पन्न होती हैं, जिसमें शून्य का (जो अपरिभाषित है)log लेने से उत्पन्न अपरिभाषित मूल्यों के कारण डेटा की हानि, साथ ही साथ फैलाव को मॉडल करने की क्षमता की कमी भी शामिल है। कई शोधकर्ता के मॉडल द्वारा डेटा के ऑटोकॉर्लेशन और असतत प्रकृति को ध्यान में रखते हुए पूर्णांक) मूल्यवान ऑटोरिएरेटिव-INAR) मॉडल, poisson मॉडल और negative binomial मॉडल का भी अध्ययन किया गया है। INAR और INGARCH मॉडेल्स न केवल घटनाओं

की गणना के विश्लेषण के लिए, बल्कि अन्य क्षेत्रों में भी जैसे डेटा के अस्तित्व के विश्लेषण में उपयोग किया जाता है। उनके सुधार के लिए INGARCH और INAR मॉडल में बहिर्जात चर को शामिल करने का प्रयास किया जाता है। पूर्णांक आधारित न्यूरल नेटवर्क जो आमतौर पर छवि प्रसंस्करण कार्य में लगाया जाता है, कीट डायनेमिक्स की भविष्यवाणी के लिए बहिर्जात चर का उपयोग करके पूर्णांक आधारित तंत्रिका नेटवर्क को विकसित करने का प्रयास किया गया है।

इस अध्ययन में अध्ययन के तहत चर बीटी के कीट और रोग डेटा हैं। कपास की फसल कपास के) साथ मानक मौसम संबंधी -के साथ (पत्तों पर कीट की औसत संख्या 3 पौधों पर बेतरतीब ढंग से चुने गए) साप्ताहिकSMW) मौसम डेटा अर्थात, अधिकतम तापमान)MAXT), न्यूनतम तापमान)MINT), वर्षा)RF), अधिकतम आर्द्रता)MAX_RH) और न्यूनतम सापेक्ष आर्द्रता)MIN_RH) का उपयोग किया गया। विभिन्न केंद्रों से बीटी कपास की फसल के लिए एकत्रित आंकड़ों की अवधि 2008-2012 से 09-तक है। चुने गए कीट 13 और जसिड में छह केंद्रों पर एफिड्स थे। अ (अकोला और वडोदरा) दो केंद्रोंकोला, बांसवाड़ा, फरीदकोट, गुंटूर, पेरम्बलुर और वडोदरा। विभिन्न केंद्रों के डेटा को दो सेटों में विभाजित किया गया था, पहले मॉडल का उपयोग प्रशिक्षण डेटा सेट के रूप में किया गया था और पिछले अवलोकनों के डेटा का उपयोग परीक्षण 12 डेटा सेट के रूप में मॉडल के सत्य◌ापन के लिए किया गया था। प्रस्तावित तरीकों के लिए डेटा विश्लेषण और प्रोग्रामिंग कोड अलग अलग आर पैकेजों का उपयोग करके-tscount, पूर्वानुमान, lmtest और tseries विधयों विकसित किए गए थे। ।

मॉडलिंग और कीटों के गतिशील पूर्वानुमान के पूर्वानुमान के लिए गणना समय श्रृंखला मॉडल विकसित करने के लिए अध्ययन आयोजित किया गया है। एक उदाहरण के रूप में, विकसित मॉडल भारत के विभिन्न केंद्रों में कपास कीटों के एफिड्स और जसिड्स में विकसित किए गए हैं। के लिए, प्रशिक्षण डेटा सेट के (मॉडल सत्यापन) और परीक्षण डेटा सेट (मॉडल निर्माण) तहत विकसित मॉडल के अकोला केंद्र प्रदर्शन के एफिड, एमएई, एमएसई और आरएमएसई का उपयोग कर। रिपोर्ट किए गए MAE, MSE और RMSE के आधार पर, कोई भी व्याख्या कर सकता है कि पूर्णांक आधारित कृत्रिम तंत्रिका नेटवर्क मॉडल ने प्रशिक्षण डेटा सेट और INAR मॉडल दोनों में अन्य मॉडल की तुलना में बेहतर प्रदर्शन किया है और डेटा सेट के परीक्षण में सभी मॉडलों को बेहतर नहीं बनाया है। वड़ोदरा केंद्र के एफिड के लिए आईएनएआर मॉडल ने प्रशिक्षण और परीक्षण डेटा सेट दोनों में सभी मॉडलों को बेहतर बनाया। अकोला, बांसवाड़ा, फरीदकोट, गुंटूर और पेरम्बलुर केंद्रों के जसिड्स के लिए IANNX मॉडल अन्य सभी मॉडलों की तुलना में प्रशिक्षण डेटा सेट में बेहतर प्रदर्शन किया और INAR मॉडल ने परीक्षण डेटा सेट में सभी मॉडल को पीछे छोड़ दिया। ANGRAU केंद्र IANNX मॉडल के जसिड्स के लिए अन्य सभी मॉडलों की तुलना में प्रशिक्षण डेटा सेट में बेहतर प्रदर्शन किया गया और INGARCH मॉडल ने डेटा सेट के परीक्षण में सभी मॉडल को पीछे छोड़ दिया। अंत में, हम यह निष्कर्ष निकालते हैं कि कोई भी मॉडल सभी प्रशिक्षण डेटा सेट और परीक्षण डेटा सेट में बेहतर प्रदर्शन नहीं कर रहा है। इस अध्ययन में प्राप्त परिणामों के आधार पर, कोई यह निष्कर्ष निकाल सकता है कि IANNX मॉडल ने प्रशिक्षण डेटा सेट में सभी मॉडलों को बेहतर बनाया और INAR मॉडल ने डेटा सेट के परीक्षण में सभी मॉडलों को बेहतर बना दिया।

Summary

Agriculture being highly cost intensive and full of uncertainties have great impact on the livelihood of farmers, if timely measures are not taken to minimize the risk, they may fall in the trap of vicious cycle. Not only this, farmers in several states are battling with growing incidence of pest attacks on a variety of crops. The threats farmers face from pest attacks are often localised but underlines the multitude of risks apart from those related to monsoon failure or a crash in crop prices Therefore, incidence of pest and diseases in crops have made agriculture very risky venture and due to high seed cost and cost of cultivation farmers are very apprehensive in adopting new technologies. About 15-25 per cent of crops yields is lost each year due to pest attacks. To mitigate these problems, reliable and timely forecast provides an important and extremely useful input in formulation of policies. In count time series the events occur in the consecutive points of time, which is commonly occurs in many situations, for example, the number of road accidents in a week, number of seeds germinated in a week etc. Integer-valued time series is an important class of discrete-valued time series models. The INAR process is well-suited for many time series which follows poisson, negative binomial, generalized poisson distributions etc. As a nonlinear and nonparametric class of model integer based neural network is very potential to capture the count time series trend and it have wide application in many areas like image classification, pattern recognition etc.

Over the year's different methodologies were introduced from time to time. Since meteorological factors are highly responsible for pest/diseases infestation in crops, therefore, advanced models like INARX and ANN along with weather parameters may address appropriate solutions for early warning of pest/disease infestation for investigating and predicting pest/disease status. With these backgrounds the INAR and integer based neural network models by considering information on exogenous variables will be developed for modelling and predicting pest dynamics in cotton crop. It is generally agreed that forecasting methods should be assessed for accuracy by using out-of-sample forecasts rather than goodness of fit to past data. To understand the probabilistic behaviour of future data, out of-sample forecasts are required. Formulae for optimal out-of-sample forecasts were derived in this study.

Over the last few years, the class of models particularly applicable to the analysis of time series count data have been studied. Count outcome variables are sometimes log-transformed and analyzed using OLS regression. Many issues arise with this approach, including loss of data due to undefined values generated by taking the log of zero (which is undefined), as well as the lack of capacity to model the dispersion. Integer-valued autoregressive (INAR) models, Poisson models and negative Binomial models have also been studied by many researcher's models take the autocorrelation and discrete nature of the data into account. INAR and INGARCH have many applications, not only to the analysis of counts of events, but also in other field like in the analysis of survival data. An attempt is made to incorporate exogenous variables in INGARCH and INAR model for their improvement. Integer based Neural network which is generally applied in image processing task, has been attempted for developing integer based neural network using exogenous variables for predicting pest dynamics.

In this study the variable under study is pest and disease data of Bt. cotton crop (average number of pest on 3 leaves selected randomly on cotton plants) along with standard meteorological weekly (SMW) weather data *viz*., maximum temperature (MAXT), minimum temperature (MINT), rainfall (RF), maximum relative humidity (MAX_RH) and minimum relative humidity (MIN_RH) were used. The duration of the collected data are from 2008-09 to 2012-13 for Bt cotton crop from different centre s. The pest chosen were Aphids at two centre s (Akola and Vadodra) and Jassids at six centre s viz. Akola, Banswara, Faridkot, Guntur, Perambalur and Vadodra**.** The data from different centre s were divided in to two sets, the first one were used for model building as training data set and data from the last 12 observations were used for validation of model as testing data set. Data analysis and programming codes for proposed methodologies were developed using different R packages viz., tscount, forecast, lmtest and tseries.

The study has been conducted to develop count time series models for modeling and forecasting pest dynamic prediction. As an illustration, the developed models have been developed in aphids and Jassids of cotton pests at different centres of India. For, Aphid of Akola centre Performance of developed models under training data set (model building) and testing data set (model validation) using MAE, MSE and RMSE. Based on reported MAE, MSE and RMSE one can interpret that integer based artificial neural network models performed better compared to other model in both training data set and INAR model outperformed all the models in testing data set. For Aphid of Vadodra centre INAR model outperformed all the models in both training and testing data set. For Jassids of Akola, Banswara, Faridkot, Guntur and Perambalur centre s IANNX model outperformed in training data set compared to all other models and INAR model outperformed all the model in testing data set. For Jassids of ANGRAU centre IANNX model outperformed in training data set compared to all other models and INGARCH model outperformed all the model in testing data set. Finally, we conclude that no models are performing better in all the training data set and in testing data set. Based on the results obtained in this study one can conclude that IANNX model outperformed all the models in training data set and INAR model outperformed all the models in testing data set.

References

1. Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pest and diseases – IASRI Models. *J. Ind. Soc. of Agril. Stat.* **61 (2):** 255-263.
2. Agrawal, R. Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis. *Ind. J. Agri. Sci.,* **71 (7):** 487-490.
3. Bhardwaj, T. and Sharma, J.P. (2013). Impact of Pesticides Application in Agricultural Industry: An Indian Scenario. International Journal of Agriculture and Food Science Technology. 4(8): 817-822.
4. Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (inar(1)) process. Journal of Time Series Analysis.8.261-275
5. Alzaid, A. A. and Al-Osh, M. (1990). An Integer-Valued pth-order autoregressive structure (INAR (p)) Process. J. Appl. Prob. 27, 314-324
6. Bockenholt, U. (1999). An INAR (1) negative multinomial regression model for longitudinal count data. Psychometrika,64, 53–67
7. Bu, R. McCabe, B. (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov Chain approach. International journal of forecasting. 24 ,151–162
8. Hellstrom (2002). Count data modelling and tourism demand. Umea Economic Studies No. 584,100 pages
9. Karlaftis, M.G.and Vlahogianni, E. I.(2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies,* 19(3): 387-399.
10. Kumar, A., Ranjana Agrawal, R. and Chattopadhyay, C. (2013). Weather based forecast models for diseases in mustard crop, *Mausam.* 64(4):663-670.
11. Kumari, Prity, Mishra, G. C. and Srivastava, C. P. (2013). Forecasting of productivity and pod damage by Helicoverpaarmigera using artificial neural network model in pigeonpea (*Cajanus cajan*). *International Journal of Agriculture, Environment and Biotechnology* 6(2): 335-340.
12. Kumari, Prity, Mishra, G.C. and Srivastava, C.P. (2014). Time series forecasting of losses due to pod borer, pod fly and productivity of pigeonpea (Cajanus cajan) for North West Plain Zone (NWPZ) by using artificial neural network (ANN). International Journal of Agricultural and Statistical Science 10(1):15-21.

13. McKenzie , E. (1985). Some simple models for discrete variate time series . Water Resour. Bull. 21 (4): 645 – 650.
14. Pavlopoulos H and Karlis D (2008) INAR (1) modelling of over dispersed count series with an environmental application. Environmetrics, 19, 369–93.
15. Arya Prawin, Ranjit Kumar Paul, Anil Kumar, K. N. Singh, N. Sivaramne and Pradeep Chaudhary (2015). Predicting pest population using weather variables: An ARIMAX time series framework. Int. J. Agricult. Stat. Sci. 11 (2), pp. 381-386.
16. Pedeli X. and Karlis D. (2011). A bivariate INAR (1) process with application. Statistical Modelling. 11(4), 325–349
17. Roy, H.S., Paul, R.K., Bhar, L.M. and Arya, P. (2016). Application of INAR model on the pest population dynamics in Agriculture. Journal of Crop and Weed, 12(2):96-101.
18. Rozman, Crtomir; Cvelbar, Urska; Tojnko, Stanislav; Stajnko, Denis; Pazek, Karmen; Martin Pavlovic and Vracko, Marjan. (2012). Application of Neural Networks and Image Visualization for Early Forecast of Apple Yield. Erwerbs-Obstbau, 54(2): 69-76.
19. Rudra, P. Pradhan. (2013). Forecasting Inflation in India. International Journal of Asian Business and Information Management 2(2): 64-73.
20. Sang, Hoon Oh (2010). Design of Multilayer Perceptrons for Pattern Classifications. The Journal of the Korea Contents Association,10(5): 99-106.
21. Sang, Hoon Oh (2012). Improving the Error Back-Propagation Algorithm for Imbalanced DataSets, International Journal of Contents, 8(2):7-12.
22. Sharma, Vidushi; Rai, Sachin and Dev, Anurag (2012). A Comprehensive Study of Artificial Neural Networks, International Journal of Advanced Research in Computer Science and Software Engineering, 2: 10, 278-284.
23. Silva N., Pereira, I. and Silva M. E. (2009). Forecasting in inar (1) model. *REVSTAT*, 7,119-134
24. Weib, C. H. and Pollett, P. K. (2012). Chain Binomial Models and Binomial Autoregressive Processes.Biometrics.68.815-824.
25. Weib, C. H., (2008). Serial dependence and regression of Poisson INARMA models. Journal of Statistical Planning and Inference ,138,2975 – 2990.