Journal of Biological Control, 27(2): 88–94, 2013

Research Article

Decision tree induction model for the population dynamics of mirid bug, *Creontiodes biseratense* (Distant) (Hemiptera: Miridae) and its natural enemies

M. PRATHEEPA*#, K. MEENA[1], K. R. SUBRAMANIAM[1] and H. BHEEMANNA[2]

Department of M.C.A. Shrimathi Indira Gandhi College, Trichirappalli 620 002, Tamil Nadu, India

#Present Address: National Bureau of Agriculturally Important Insects, H. A. Farm Post, Hebbal, Bangalore 560 024, India.

[1]Bharathidasan University, Trichirappalli 620 024, Tamil Nadu, India

2 Department of Entomology, University of Agricultural Sciences, Raichur 584 102

** Corresponding author Email:* mpratheepa.nbaii@gmail.com

ABSTRACT: The mirid bug, *Creontiodes biseratense* (Distant) (Hemiptera: Miridae) is as a serious pest of cotton crop. Forecasting model by linking the pest incidence with season, crop phenology, biotic and abiotic factors enable to understand the dynamics of pest occurrence likely to occur. A data mining technique decision tree induction model is proposed for forecasting the pest incidence and study the population dynamics of mirid bug, *C. biseratense* in relation to its natural enemies *viz.*, spider *Lycosa* sp. and coccinellid *Cheilomenes sexmaculata* Fabricius and abiotic factors. The results of the decision tree agreed well with statistical analysis.

KEY WORDS: *Creontiodes biseratense*, cotton, spiders, coccinellids, decision tree, information theory, abiotic

(Article chronicle: Received: 29-01-2013; Revised: 18-05-2013; Accepted: 02-06-2013)

INTRODUCTION

Cotton enjoys premier status among all cash crops in the country and is the principal raw material for the flourishing textile industry. In Karnataka, the area under cotton cultivation is 3.71 lakh ha with a production capacity of 8.00 lakh bales and an average productivity of 367 Kg/ha (Anon., 2008). In south India (Karnataka), the mirid bug, *Creontiodes biseratense* (Distant) (Hemiptera: Miridae) mirids is gaining importance on Bt cotton, causing heavy shedding of squares and bolls which lead to significant reduction in seed cotton yield (Patil *et al*., 2006; Udikeri, *et al*., 2009). Nymphs and adults of *C. biseratense* damage the squares, flower and tender bolls by sucking the sap resulting into gradually yellowing, shrivelling and premature dropping of squares and young bolls (Venkateshalu *et al*., 2010; Ravi and Patil, 2008). The major concern is the inadequate knowledge about the factors influencing population dynamics. The extent of pest damage varies at different stages of crop growth (Trivedi *et al*., 2005). Abiotic factors *viz.*, temperature, relative humidity, rainfall and wind speed and biotic factors such as natural enemies influence the occurrence and population build up of many insect pests (Sreedevi and Verghese, 2007). There is an urgent need to generate data on the dynamics of pests in a holistic manner and develop models for forecasting the pest incidence in advance to reduce crop loss. Decision tree techniques, a data mining tool, is largely popular in classification application because the models generated closely resemble human reasoning and are easily understood (Zhao and Ram, 2004). The major advantage of a decision tree is its interpretability, the decision can be represented in terms of a set of rules (Basak and Krishnapuram, 2005). In this paper, we propose a decision tree induction model for forecasting *C. biseratense* incidence and its population dynamics on cotton by analyzing both biotic and abiotic factors.

MATERIALS AND METHODS

Data base

Data sets were obtained from the Regional Agricultural Research Station, Raichur, Karnataka, India from unsprayed experimental plots under the All India Coordinated Cotton Improvement Project (AICCIP) on Bt cotton. Sample size was 25 plants/500 m^2 area. Since mirid bug occurs on squares of cotton (Ravi and Patil, 2008; Venkateshalu *et al*., 2010), weekly observations on nymphs of mirid bug count per square were recorded during 2009 to 2012 period. Natural enemies *viz.*, spider (*Lycosa* sp.) (NE1) and coccinellid

(*Cheilomenes sexmaculata* Fabricius) (NE2) populations were recorded per plant. Weekly means of daily weather parameters like maximum temperature (MaxT), minimum temperature (MinT), average humidity (RH) and sum of rainfall (RF) and number of rainy days (RFD) per week were computed. Current week's pest incidence was related to previous week's abiotic and biotic factors for analysis. Different stages of the cotton crop like vegetative (1-5 weeks after germination), square initiation, flowering and boll formation, boll maturity and boll bursting were considered for analysis. Sample data is given in Table 1 along with attributes/variables and class/target variable.

Training and Testing Phase

The overall dataset (72 records) was divided into two parts *viz.*, training phase and testing phase. Two thirds of records (48 records) have been chosen for training phase (from 2009 to 2010) and one third of records (24 records) have been chosen for testing phase (2011). The decision tree was built on testing set based upon the intrinsic nature of the training set.

Assigning class label pattern

Pest incidence was classified into two classes namely, HIGH and LOW based up on mirid bug count per square with incidence $\geq$ 1 mirid bug/square was considered as HIGH and < 1 mirid bug/square was considered as LOW. These class values were assigned in the database for the training set of data.

Data discretization

The discretization technique, equal binning method was used to convert all numerical attributes (continuous variables) NE1, NE2, MaxT, MinT, RH, RF and RFD into the categorical values / labels as A1, A2, A3 and A4 as option of total bins 4 has been given. The labels of A1 to A4 for each numerical attribute, has its own range values stored in the database. The proposed algorithm for discretization of numerical attributes based on the binning method is as follows:

Step i. Get user input as no. of bins required denoted as '*b*'

Step ii. *nb* (No. of bins) = *n* / *b*, where '*n*' denotes total no. of records

Step iii. Sort the numerical attribute values into ascending order

Step iv. For each numerical attribute, assign the label *A*1 into 1 to '*nb*' records and A2 into

Table 1.: Sample records/tuples from the database

Tuples	Attributes / Variables										
Sl. No	Crop Stage	Season	NE1	NE2	MaxT (°C)	MinT (°C)	RH (%)	RF (mm)	RFD	Mirid bug per square (PI)	Class of Target variable (PI)
1	2	Monsoon	0.144	0.096	29.8	21.4	75.5	36.4	4	0	Low
2	2	Monsoon	0.168	0.112	32.5	21.9	74.5	7.6	2	0	Low
3	3	Postmonsoon	0.228	0.152	32.5	21.4	80.5	92.4	4	0.28	Low
4	3	Postmonsoon	0.24	0.16	29.2	20.7	83	362.6	3	1.08	High
5	3	Postmonsoon	0.21	0.14	32.3	18	61.5	0	0	5.22	High
6	4	Postmonsoon	0.264	0.176	30.5	20.1	77	8.2	1	6.35	High
7	4	Postmonsoon	0.624	0.416	33.1	14.6	84	0	0	10.36	High
8	5	Winter	0.168	0.112	28.8	15.7	65.5	0	0	0.06	Low
9	3	Postmonsoon	0.204	0.136	31	20.3	80	18.2	2	0.16	Low
10	5	Winter	0.288	0.192	30.3	12.9	68	0	0	1.12	High

NE1 – spiders; NE2– coccinellids; MaxT: maximum temperature (weekly mean); MinT: minimum temperature (weekly mean); RH:relative humidity (weekly mean); RF: weekly sum of rainfall in mm; RFD: number of rainy days per week; PI–Pest incidence; Crop stage 1– 1–5 weeks from the sowing date; Crop stage 2 – square initiation stage; Crop stage 3 – flowering and boll formation stage; Crop stage 4 – Boll maturity stage; Crop stage 5 – Boll bursting stage '*nb* + 1' to '*nb* + *nb*' and *A*3 into '2*nb*+*1*' to '2*nb* + *nb*' and so on until the counter no. of bins becomes zero.

Step v. The range values of *A*1 to *An*, were stored separately in the database for each numerical attribute.

Step vi. If extra records are available after creation of bins, the last label, An was assigned.

Attribute selection measure

The data classification process is aimed at reducing uncertainty or gaining information about the target (classification) attribute. The Shannon information theory was used for the attribute selection measure. It is a measure of how good an attribute is for predicting the class of each of the training data. The expected information needed to classify a tuple/record in D is given by

$$\text{Info (D)} = \sum_{i=1}^{m} p_i \log_2(p_i)$$

where Info(D) is also known as the entropy of D, that is, total information value (Han and Kamber, 2001). Let D, the data partition, be a training set of class labelled tuples. p_i is the probability that an arbitrary tuple in D belongs to class C_i and is estimated by n(C_i, D) / n(D) where n(C_i, D) denotes total number of tuples or records in class C_i and n(D) denotes total number of tuples or records in D, respectively.

Entropy of each attribute or variable A, with labels $\{A_1,A_2,...,A_v\}$ is used to split D into 'v' subsets where 'v' varies from 1 to 4 (total number of bins) and Info_A(D) is calculated based on the total number of records from A1 to A4 labels with each class values of PI and it can be represented as:

$$\text{Info}_A \text{ (D)} = \sum_{j=1}^{v} \frac{n(D_j)}{n(D)} x Info(D_j)$$

Info_A is denoted as the information value for the attribute A. Info_A(D) is the expected information required to classify a tuple from D based on the partitioning by A.

Information is gained by branching an attribute A, and gain value for an attribute A is calculated as

$$\text{Gain (A)} = \text{Info(D)} - \text{Info}_A\text{(D)}$$

The attribute A with the highest information gain, (Gain(A)), is chosen as the splitting attribute at node N. Here, the attribute A would do the "best classification" (Gupta, 2006).

Gain value was calculated recursively for each subset evaluation (decision space) and the stopping criterion until the gain value reached zero was used. The gain values for all the attributes at iteration 1 (level 1) are given in Table 2.

Construction of decision tree

The root of the decision tree was fixed as PI and it was at level 0. The maximum gain value of the attribute was crop stage (Table 2). Hence, the tree generation started from the crop stage at level-1. The decision tree is a non-backtracking algorithm hence it was constructed in a top-down manner. The tree constructed in the form of binary tree. The tree has decision node as condition and output/result of that condition derived as yes or no options. The 'yes' option always grows as left child and 'no' option always grows as right child in the tree. The node ends when the condition not able to proceed further in the 'no' option as right child. The end of the leaf node denotes the class label of pest incidence. The classes always defined in the left child of the tree *i.e.* 'yes' option of the condition/decision node. The decision tree of mirid bug incidence on cotton, based on biotic and abiotic factors have been given in Fig. 1 and Fig. 2.

Statistical analysis

To understand the role of abiotic and biotic factors on pest incidence, correlation analyses were done by using the Statistical Package for Social Sciences (SPSS V 17.0) and the results are presented in Table 2.

Decision Tree

The binary form of decision tree derived based on Shannon information theory are given in Fig. 1 and Fig. 2. IF-THEN rules could be derived based on the decision tree.

During flowering and boll formation stage of the cotton crop when the spider population is more the mirid bug population is also high. Similarly, when the relative humidity is 72 – 77.5% the pest population is high (Fig. 1).

During boll maturity stage of the crop, when the spider population was more the pest incidence was high. Similarly, when the maximum temperature ranged from 28.8°C to 30.4°C, the pest incidence was high. The pest incidence was high when there was a rainfall coupled with the maximum temperature ranges from 30.4°C to 31.5°C (Fig. 2).

Table 2: Comparison of information gain value and correlation analysis of PI with biotic and abiotic factors

Attribute/ variable name	Information gain value	Correlation analysis (r value)
Cropstage	0.26	0.288*
Season	0.21	0.209
NE1	0.20	0.509**
NE2	0.19	0.509**
MaxT	0.04	0.014
MinT	0.02	–0.092
RH	0.01	0.166
RF	0.01	–0.087
RFD	0.01	–0.118

Note: ** correlation is significant at 0.01 level
* correlation is significant at 0.05 level

Mirid bug, spider and coccinellid incidence was high when the crop was in flowering, boll formation stages and boll maturity stages during the period 2009-10 and 2010-11, respectively (Fig. 3). The spiders and coccinellids population are high when the mirid bug incidences are high (Fig. 3).

The Shannon information gain value for all attributes at level 1 and correlation analysis of PI with biotic and abiotic factors are given in Table 2.

The information theory shows that among all the attributes, crop stage plays major role on pest incidence since the gain value of 'crop stage' is the highest. These results corroborate with the reports by Surulivelu and Dhara Jothi (2007) and Khan *et al*. (2007) that shedding of most of the damaged tender bolls were due to *C. biseratense* infestation and all stages of mirids caused damage to bolls and squares. Among were biotic and abiotic factors, the gain values of NE1 and NE2 was maximum which indicated that spiders and coccinellids were playing a major positive role on mirid bug incidence which indicates that spiers and coceinellids may not be efficient predators of *C. biseratense*. Correlation analysis

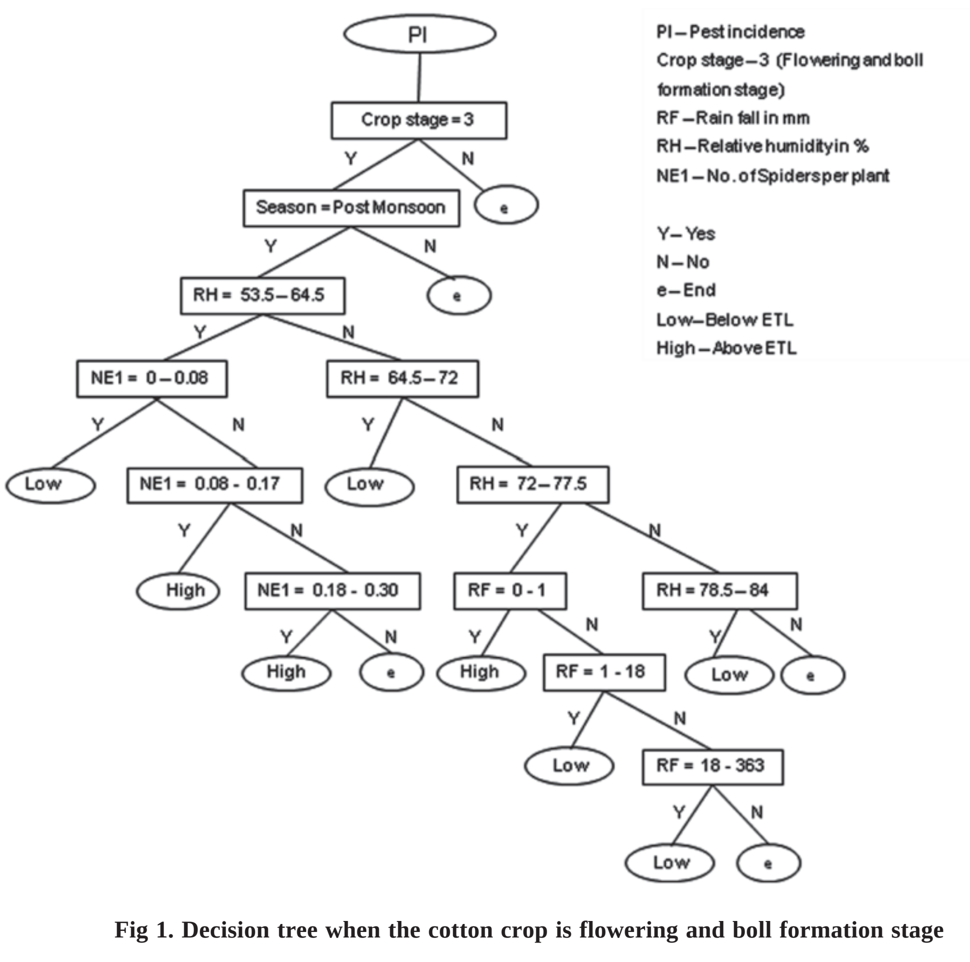

Fig 1. Decision tree when the cotton crop is flowering and boll formation stage

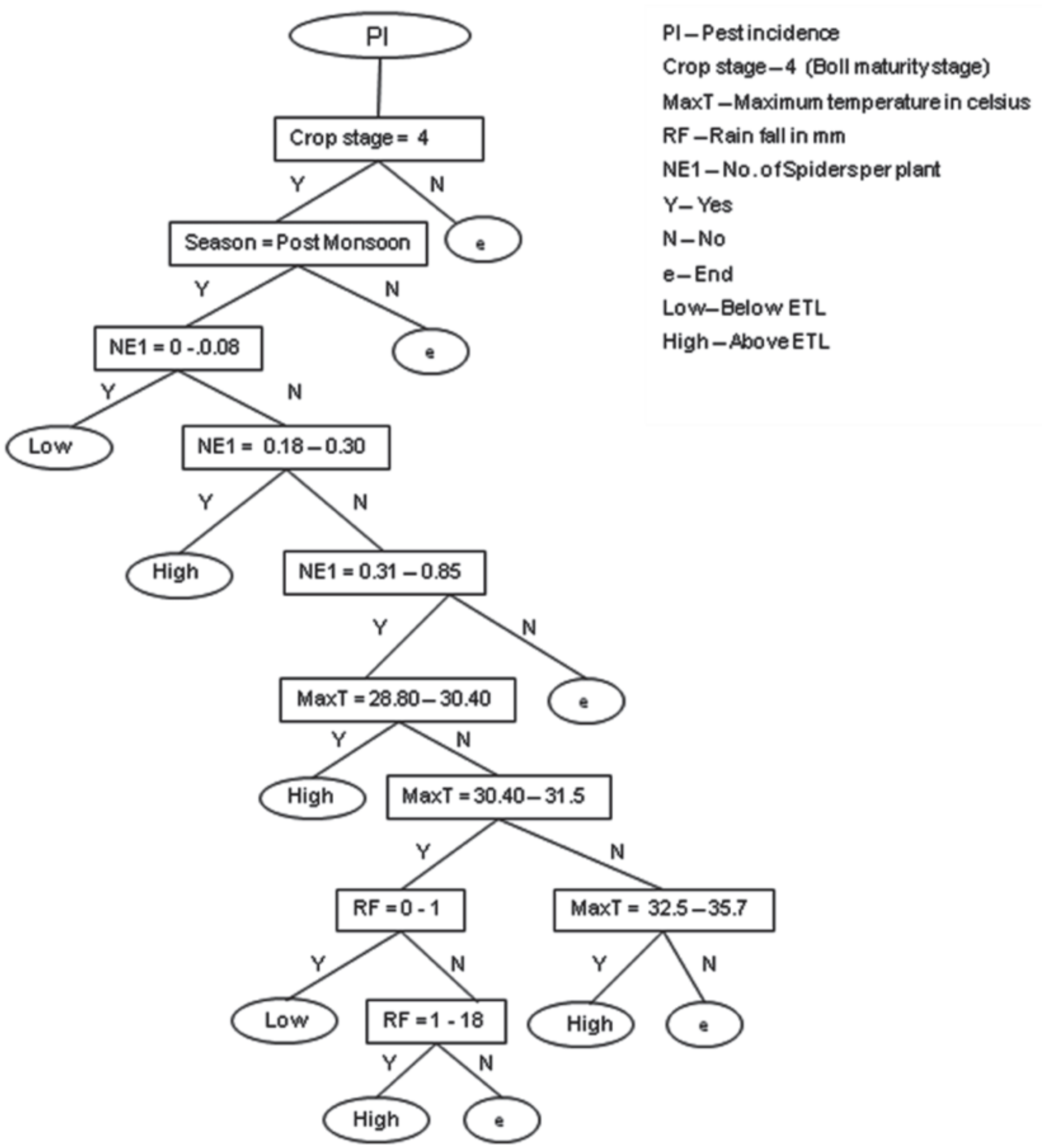

Fig 2.: Decision tree when the cotton crop is boll maturity stage

also revealed that crop stage, NE1 and NE2 had significant impact on mirid bug incidence (Table 2). The decision tree constructed indicated the variable NE1 existed as information gain value computed at each level and also at flowering, boll formation stage and boll maturity stage which indicated that spiders played a major role on mirid bug incidence. IF-THEN rules can be derived based on the decision tree (Fig. 1 and Fig. 2). Decision tree revealed that when the spider population is high, the mirid bug incidence was also high indicating a density dependent relationship. The importance of spiders in preventing crop loss by relating the abundance of spiders feeding on *Heliothis* spp. to fruit damage levels in unsprayed cotton field at Tehran, Iran was mentioned by Ghavami (2008). Similarly, the natural enemy fauna such as coccinellid *Menochilus* sp.; and spiders *Neoseona* spp. were also greater when cowpea, groundnut or mungbean were intercropped with cotton (Venugopal, 1995).

The proposed decison tree induction model predicted the mirid bug incidence as high or low based on user threshold value. Crop stage played a major role on mirid bug incidence based on Shannon information theory. Mirid bug population was more when the spider and coccinellid incidence was high IF-THEN rules can be derived from the decision tree. Decision tree induction approach enabled to study the role of biotic and abiotic factors on pest incidence.

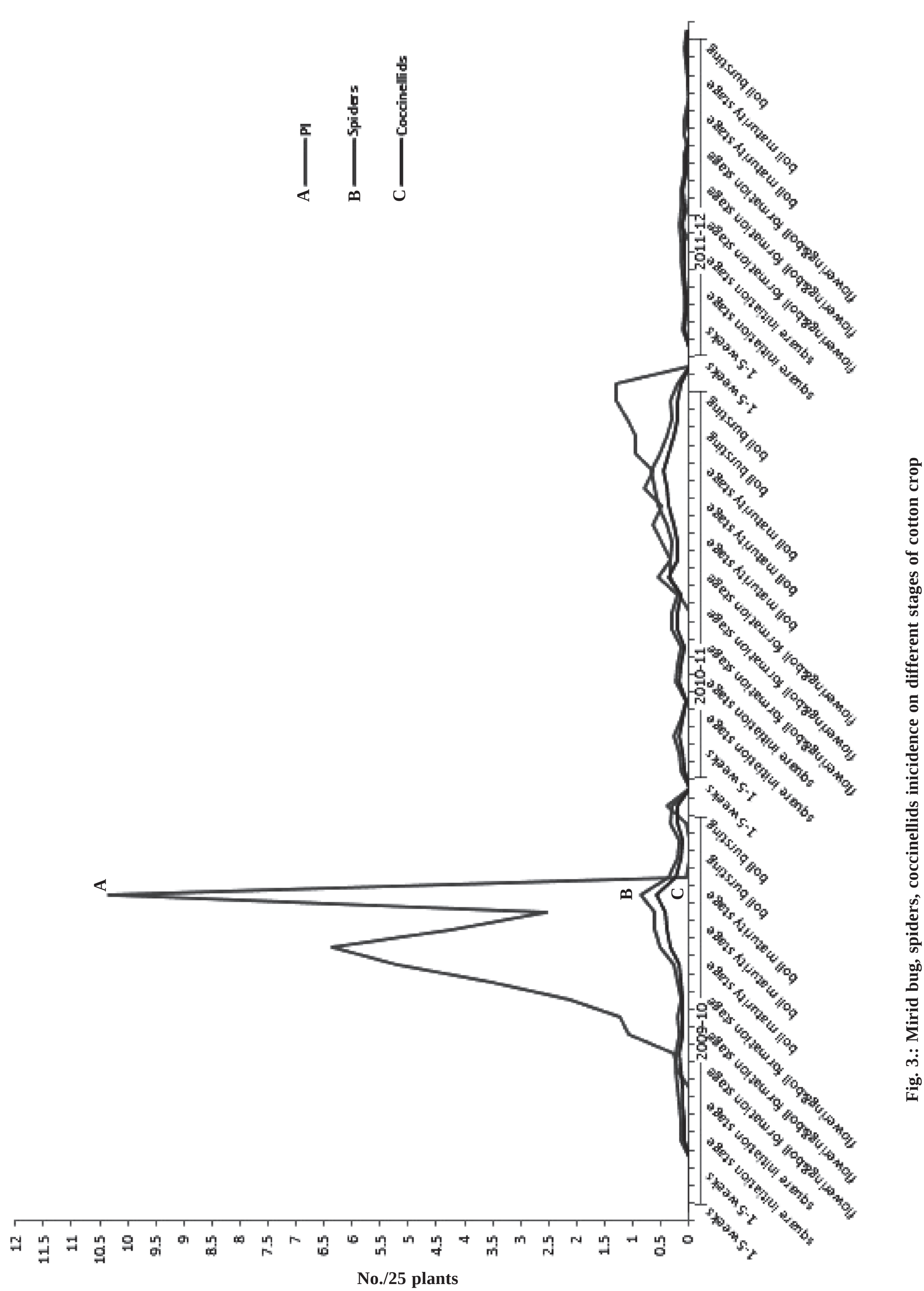

Fig. 3.: Mirid bug, spiders, coccinellids inicidence on different stages of cotton crop

REFERENCES

Anonymous, 2008a. Project coordinator's report (2007–08). All India co-ordinated cotton improvement project, pp. 4.

Basak J, Krishnapuram R. 2005. Interpretable hierarchical clustering by constructing an un-supervised decision tree. *IEEE Trans Knowl Data Engg* **17** (1): 121–132.

Ghavami S. 2008. The potential of predatory spiders as biological control agents of cotton pests in Tehran Provinces of Iran. *Asian J Expl Sci.* **22** (3): 303–306.

Han J, Kamber M. 2001. Classification and prediction. pp. 285–375. In *Data Mining Concepts and Techniques*, 2nd ed. Jim Gray, Indian Reprint. Elsevier. Morgan Kaufmann.

Khan M, Quade A, Murray D. 2007. Damage assessment and action threshold for mirids, *Creontiades* spp. In Bollgard II cotton in Australia. Second International *Lygus* Symposium Asilomar. *J Insect Sci*. **8**: 49, p. 27.

Patil BV, Bheemanna M, Patil SB, Udikery SS, Hosmani A. 2006. Record of mirid bug, *Creontiades biseratense* (Distant) on cotton from Karnataka, India. *Insect Env*. **11**:176–177.

Ravi PR, Patil BV. 2008. Biology of mirid bug, *Creontiades biseratense* (Distant) (Hemiptera: Miridae) on Bt cotton. *Karnataka J Agric Sci.* **21**(2): 234–236.

Sreedevi K, Verghese A. 2007. Ecology of aphidophagous predators in pomegranate ecosystem in India. *Communication Agrl Appl Biol Sci.* **72**(3): 509–516.

Surulivelu T, Dhara Jothi B. 2007. Mirid bug, *Creontiodes biseratense* (Distant) damage on cotton in Coimbatore. http://www.cicr.gov.in

Trivedi TP, Yadav CP, Vishwadhar, Srivastava CP, Dhandapani, Das DK, Singh, J. 2005. Monitoring and forecasting of *Heliothis / Helicoverpa* population, pp. 119-140 In: Sharma HC (Ed.) *Heliothis / Helicoverpa Management – Emerging trends and strategies for future research.* Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi.

Udikeri SS, Patil SB, Shaila HM, Guruprasad GS, Patil SS, Kranthi KR, Khadi BM. 2009. Mirid menace – a potential emerging sucking pest problem in cotton. http://www.icas.org.

Venkateshalu V, Kalmath B, Swamy L, Sushila N, Mallapur CP, Reddy N. 2010. Performance of different Bt cotton hybrids against mirid bug, *Creontiades biseratense* (Distant) (Miridae: Hemiptera). *Kar J Agric Sci*. **23**(1):109–110.

Venugopal Rao N. 1995. Bioecology and management of *Helicoverpa armigera* in the cotton ecosystem of Andhra Pradesh, Hyderabad, India. Andhra Pradesh Agricultural University. Ph.D Thesis.

Zhao H, Ram S. 2004. Constrained cascade generalization of decision trees, *IEEE Trans Knowl Data Engg* **16 (**6): 727–739.