3

Research Achievements

The research targets set by the Institute were implemented by six Divisions of the Institute, viz. Design of Experiments, Statistical Genetics, Forecasting and Agricultural System Modelling, Sample Surveys, Computer Applications and Centre for Agricultural Bioinformatics. The basic, applied, adaptive and strategic research in Agricultural Statistics and Informatics is carried out under six broad programmes that cut across the boundaries of the Divisions and encourage interdisciplinary research. The six programmes are as under:

1. Development and Analysis of Experimental Designs for Agricultural System Research
2. Forecasting, Modelling and Simulation Techniques in Biological and Economic Phenomena
3. Development of Techniques for Planning and Execution of Surveys and Statistical Applications of GIS in Agricultural Systems
4. Development of Statistical Techniques for Genetics/ Computational Biology and Applications of Bioinformatics in Agricultural Research
5. Development of Informatics in Agricultural Research
6. Teaching and Training in Agricultural Statistics and Informatics

Programme 1: DEVELOPMENT AND ANALYSIS OF EXPERIMENTAL DESIGNS FOR AGRICULTURAL SYSTEM RESEARCH

Application of optimization technique based algorithms for construction of incomplete block designs

In order to maintain homogeneity among the experimental units within blocks incomplete block designs are very useful. Small blocks, with number of experimental units smaller than the total number of treatments in the experiment, help in reducing the intra-block variance leading thereby to precise treatment comparisons. Incomplete block deigns have been used in many agricultural experiments. However, the experimenters often face the problem of selecting a suitable design for given number of treatments, v, number of blocks, b and the common block size, k. An efficient incomplete block design may not be always available for given number of treatments, blocks and block sizes. Therefore, the purpose of this study was to address the problem of obtaining highly efficient incomplete block designs using the optimization approaches, particularly the linear integer programming approach.

A constraint satisfaction approach for construction of incomplete block design with specified concurrence matrix has been proposed. A multi-step linear integer programming approach to construct a proper binary incomplete block design with specified parameters and concurrence matrix has also been developed. Nearly balanced concurrence matrix is also generated through the algorithm. Such concurrence matrices are known to lead to efficient designs. Using the two approaches, construction of different classes of binary incomplete block designs viz. balanced incomplete block designs, regular graph designs, semi-regular graph designs etc. is illustrated with examples. Modification of the algorithm for obtaining incomplete block designs for tests vs

control(s) comparisons has also been shown and illustrated with examples. All the proposed methods have been implemented using R and SAS packages. An R package called *'ibd'* has been developed and is available on cran.r-project.org/web/packages/ibd/index.html. SAS macros have also been prepared.

The algorithm is fairly general in nature and can generate an efficient design for given parameters, provided such a design exist. However, for the benefit of the experimenters a catalogue of efficient incomplete block designs in a restricted parametric range $3 \leq v \leq 20$, $b \geq v$, $2 \leq k \leq \min(10, v-1)$ with $vb \leq 1000$ is prepared. The layouts of the designs are available on Design Resources Server at http://iasri.res.in/design/ibd/ibd. A screenshot of the webpage containing the catalogues is given below.

Webpage containing catalogue of efficient incomplete block designs

The user can see the layout of the design for a given parametric condition, by clicking on the view design.
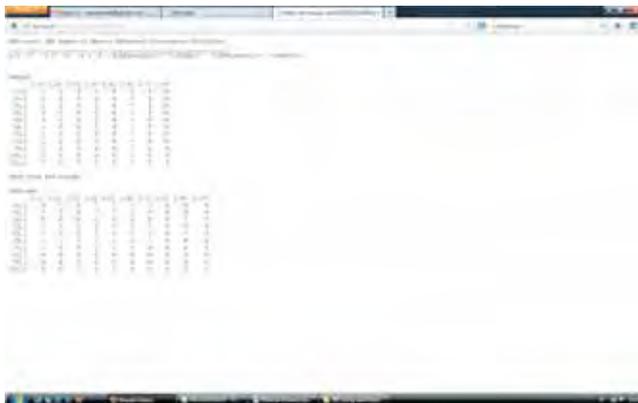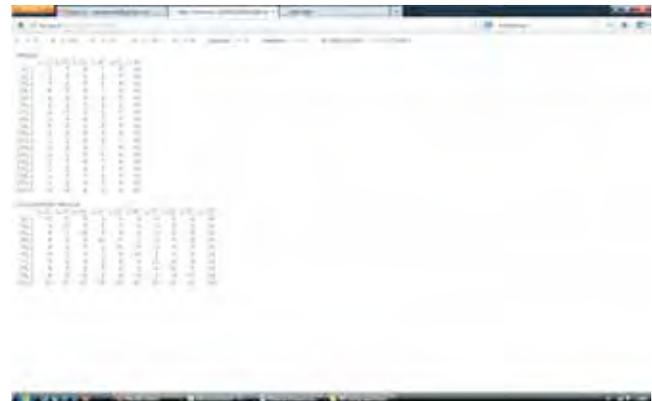

Layout of an efficient incomplete block design

The proposed algorithm has also been utilized to construct balanced treatment incomplete block designs for $2 \leq v \leq 12$, $v \leq b \leq 50$, $2 \leq k \leq v-1$. Balanced treatment incomplete block designs are useful for comparing test treatments with a control. A list of designs obtained in the above range is also presented. The layouts of the designs are available on Design Resources Server at http://iasri.res.in/design/btib/btib. A screenshot of the webpage containing the catalogues is given below.

Webpage containing catalogue of balanced treatment incomplete block designs

By clicking on the view design for a given parametric combination, user can see the layout of the design.

Layout of a balanced treatment incomplete block design

Strengthened Design Resources Server

For dissemination of research in Design of Experiments, Design Resources Server (www.iasri.res.in/design) was further strengthened through adding new link on orthogonal arrays. It has been strengthened by adding the following links:

Row-column designs in two rows

- Row-column designs in two rows are useful for 2-colour microarray experiments. A new link 'Catalogue and Generation of Row-Column Designs' (http://www.iasri.res.in/drs/) has been initiated for generation of Row-Column designs with two rows along with lower bounds to A- and D- efficiencies for parametric range $3 \leq v \leq 10$, $v \leq b \leq v(v-1)/2$, $11 \leq v \leq 35$, b = v, and (v, b) = (11,12), (11,13), (12,13), (12,14), (13,14), (13,15), (13,16) under fixed and mixed effects models. Some screen shots are as.

Block designs with factorial treatment structure with block size 2 for baseline parameterization

- Developed a module for online *generation of block designs with block size 2 for factorial experiments* with baseline parameterization in b = v-1

blocks (where *v* is the number of treatment combinations v = $s_1 x s_2 x \ldots x s_n$, $2 \leq n \leq 10$ factors and v = $s_1 x s_2 x \ldots x s_n$ and for 2 factor mixed level factorial experiments in $v-1 \leq b \leq (v-1)+(s_1-1)(s_2-1)$ arrays and made available at http://www.iasri.res.in/dbp/. Some screen shots are given below.
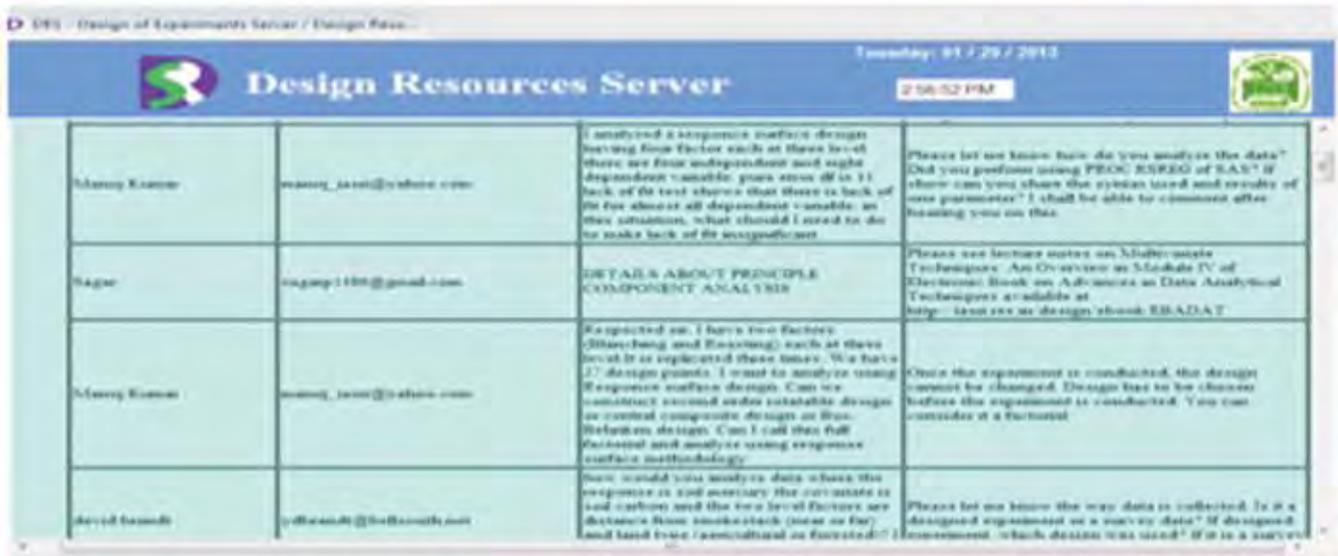

Books on Design of Experiments

- A list of books on Design of Experiments has been provided for the benefit of the visitors of this web resource, the faculty, the researchers in Design of Experiments and the students. No claim is being made for this list to being exhaustive. New additions would be made to it from time to time.

Usage of the Server

- The server has a facility of "Ask a Question" through which a lot of questions are being received and answered. More than 40 questions asked through the link 'Ask a Question' were answered for providing e-advisory services.
- During 01 April, 2012 - 31 March, 2013, Google analytics gave 12342 page views through 458 cities of 91 countries. Average time taken on page was 3.51 minutes.

Efficient designs for drug testing in veterinary trials

- Veterinary trials are generally conducted for drug testing, to solve specific and practical problems like diseases and toxicology testing or safety testing conducted by pharmaceutical companies. Research on living animals is carried out when it will reveal information which cannot be obtained in other ways. For ethical and economic reasons, it is important to design veterinary trials well, to analyze the data correctly and to use the minimum number of animals to achieve the objectives. A method of constructing designs for making comparisons of investigational products with two active controls has been obtained that are suitable for veterinary trials. The two controls can also be taken as one active control and placebo facilitating the experimenter to pursue multiple goals in one trial like establishing superiority to placebo and at the same time allowing comparisons of the investigational products with an active control. The efficiency of these designs under a nested model by considering several observational units within each experimental unit, in comparison to an orthogonal design with same number of treatments has been studied. Further, in a drug-drug interaction study the purpose of the experimenter is to investigate whether co-administration of two or several drugs will alter the absorption profile of each drug. For the class of symmetric factorial (v^3), row-column designs (RCDs) with $3v$ rows and v^2 columns, developed for studying the multi-drug interaction effects, the general form of effects confounded in rows and columns have been identified and are found to be of third order.

Experimental designs in the presence of indirect effects of treatments

- Indirect effects are effects which occur in an experiment due to the units which are adjacent (spatially or temporally) to the unit being observed. A class of complete circular block designs strongly balanced for spatial (neighbour) effects up to distance 2 has been obtained. The parameters of the design so obtained are v, $b = v(v-1)/2$, $r = (v-1)(2v-1)/2$ and $k = 2v-1$. The information matrix for estimating the direct as well as spatial indirect effects from the neighbouring units has been derived and the designs are found to be totally balanced.

- A class of strongly balanced designs balanced for temporal indirect effects up to second residual effects has also been obtained. The parameters of the design so obtained are v (prime) treatments, $p = v(v-1)$ periods and $n = v$ experimental units.

18

- A class of minimally neighbour balanced row-column designs with parameters v (even), $p = q = r = v$ has been obtained which are variance balanced for the estimation of elementary contrasts pertaining to direct effects of treatments.

- Universal optimality of block design with spatial indirect effect from neighbouring unit under a general non-additive model has been established in the presence of interactions among the treatments applied in the adjacent plots as these effects contribute significantly to the response. A class of complete block designs balanced for neighbour effects from left neighbouring unit is shown to be universally optimal for the estimation of direct and neighbour effects of treatments.

- Considering more than one relationship between observations on units over space, two series of linear trend free block (one complete and one incomplete) designs balanced for estimating direct and neighbour indirect effect of treatments have been proved to be trend free for higher order trend effects.

- A large number of Neighbour Balanced Designs (NBDs) and Crossover Designs (CODs) are developed in the literature. For easy accessibility and quick reference of these designs by the experimenters, a software Web Generation of Experimental Designs Balanced for Indirect Effects of Treatments has been developed and deployed at www.iasri.res.in/webdbie. List of neighbour balanced block designs and crossover designs for $v \le 20$ was prepared for developing the online catalogue. The designs can also be generated from the online catalogues developed for the purpose. This software provides freely available solution for the researchers and students working in this area.

Home Page

- The software generates five classes of Neighbour Balanced Block Designs (v treatments, b blocks, r replications and k block size) and eight classes of Crossover Designs (v treatments, p periods and n units/sequences).
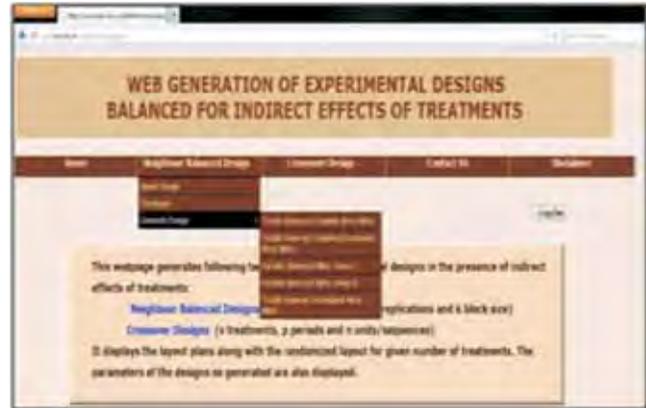
Neighbour Balanced Designs

- The webpage displays the layout plans along with the randomized layout for given number of treatments. The parameters of the designs so generated are also displayed. The details of the designs are also included.

Neighbour Balanced Design for v = 5

Crossover Designs

Crossover (Williams Square) Design for v = 5

- The online catalogue (v ≤ 20) of Neighbour Balanced Designs and Crossover Designs has been developed and is included in the software. Search facility of all designs and designs for some particular value of parameters have been provided showing the layout of the design.

Mating-Environmental designs under two-way blocking setup

Mating-Environmental Row-Column (MERC) designs are suitable for breeding programmes compared to traditional mating designs as it provides designs which serve both the purposes, i.e. mating designs laid out using a row-column design, for the breeders. MERC designs yields more precise comparison among general combining ability (gca) effects as they eliminate two perpendicular source of variations in the field. They facilitate comparison among gca effects free from specific combining ability (sca) effects. Three series of designs which are variance balanced for estimating the elementary contrasts pertaining to gca effects free from sca effects are

Series 1: The parameters are number of crosses (v) = number of rows, (p) = number of columns, (q) = number

of replications, (r) = $\dfrac{t(t-1)}{2}$, where t is the number of

lines. For this class of MERC designs, the information matrix for estimating the contrasts pertaining to crosses is

$$\mathbf{C}_{\ddot{A}} = \frac{t(t-1)}{2}\left[\mathbf{I} - \frac{2\mathbf{J}}{t(t-1)}\right]$$

and the information matrix for estimating the contrasts pertaining to gca effects after eliminating *sca* effect is of

the form $\mathbf{C}_{gca} = \dfrac{(t-1)^2}{2(t-2)}\left[\mathbf{I} - \dfrac{\mathbf{J}}{(t-1)}\right].$

Series 2: The parameters are $v = \dfrac{t(t-1)}{2}$,

$p = \dfrac{(t-3)\ (t-2)}{2}$, $q = \dfrac{t\ (t-1)}{2}$ and $r = \dfrac{(t-3)(t-2)}{2}$,

where t (number of lines) should be a prime number. For this class of designs, the information matrix for estimating the contrasts pertaining to gca effects after eliminating *sca* effect is

$$\mathbf{C}_{gca} = \frac{t(t-3)(t-4)}{2(t-2)^2}\left[\mathbf{I} - \frac{\mathbf{J}}{t}\right].$$

Series 3: The parameters are $v = \dfrac{t(t-1)}{2}$, $p = \dfrac{t(t-1)}{2}$,

q = t and r = t, where t should be a prime number. For this class of MERC designs, the crosses are partially balanced and the information matrix for estimating the contrasts pertaining to *gca* effects after eliminating *sca* effect is of

the form $C_{gca} = \dfrac{t}{(t-2)}\left[I - \dfrac{J}{t}\right].$

- Developed macros using SAS IML (Interactive Matrix Language) for generation and randomization of three series of MERC designs for the above three series of

20

designs which will be of immense use to researchers for constructing MERC designs as it provides ready-made layout plans.

- Methodology has been developed for the orthogonal partition of the information matrix for estimating elementary contrasts pertaining to gca and sca effects from a diallel (or partial diallel) cross experiment laid out under a two-way blocking set up considering F_1's along with Selfing's (Griffing's method II).

Planning, designing and analysis of data relating to experiments conducted under AICRP on Long Term Fertilizer Experiments

The data generated from long term fertilizer experiments on various crop wise characters *viz.* grain and straw yield, plant nutrients concentration/ uptake and available soil nutrients after the completion of each crop cycle received from eight cooperating centres of 2010-11 and two cooperating centres of 2011-12 were analysed.

The mixed model methodology has been used for the repeated measure analysis of two cooperating centres Coimbatore and Bangalore (for both kharif and rabi seaons) by taking year as time variable for the data of grain yield. Treatment, year and the interaction treatment × year are highly significant at both the centres. In general, least square means were highest for the treatment 100% NPK+FYM for both Rabi and Kharif seasons for both the centres. For the Coimbatore cooperative centre of Kharif season, data of 38 years were analyzed by making four groups according to trend in grain yield. An increasing trend was observed during the period 1981-82 to 1994-95. The data of seventeen cooperating centers for all the characters of LTFE is now available at the site http://www.iasri.res.in/isde.

Planning, designing and analysis of "On Farm" Research experiments planned under Project Directorate for Farming Systems Research

Three types of experiments viz; Response of nutrients, Diversification/Intensification of cropping system and Sustainable production system are planned and conducted at 31 On Farm Centres (OFR) in farmers' field under Project Directorate of Farming Systems Research, Modipuram during 2010-11. The data of 117 experiments conducted at 2286 farmers' field at OFR centres are processed for statistical analysis.

- During 2011-12 a new experiment "On-Farm Integrated Farming System Research" has been initiated at 28 OFR Centres replacing the experiment (sustainable production system).The objective of this experiment is to address critical constraints of small and marginal farmers' for overall productivity improvement and to increase the profitability of households and ensure livelihood security of the farmers. The treatments are formed by making interventions in modules of different components of integrated farming system (IFS) such as crops, animals, value addition and processing / subsidiary enterprises and kitchen gardening/poultry/ fisheries etc. By using the input and output costs of these interventions in various modules of IFS, the impact of interventions in terms of productivity and profitability of small and marginal farmers can be evaluated.

- On line data entry of experiment "Response of nutrients" conducted during 2011-12 has been initiated for the first time by OFR Agronomists and On-line data entry and analysis by 24 OFR Agronomists has been carried out. Statistical analysis of the data of 36 experiments (Diversification and/or Intensification of cropping system) conducted at 28 OFR Centres during 2011-12 has been processed for statistical analysis.

- Coefficient of variation (CV) of On-Farm experiments conducted during 2011-12 have been evaluated and presented in the tables. In experiment (Response of nutrients), the analysis have been carried out for crop yields and two characters analysis have been done for experiment "Intensification/Diversification of cropping system" namely caloric value and net return obtained in the cropping sequences.

Distribution of CV(%) in ON-FARM experiments

Type of Experiment	CV				
	0-5	5-10	10-15	15-20	> 20
Experiment-1	11	34	2	-	1
Experiment-2	25	25	14	5	-

It is observed that 34 of the experiment-1 have percent coefficient of variation (C.V.) in the range of 5 to 10 %. The C.V. of experiment-2 has been found in 25 cases in both 0-5 and 5-10% range, whereas it is observed that C.V. in 5 cases is in 15-20% range.

Information System for Designed Experiments (ISDE)

ISDE is a web-enabled information system (available as a link on http://www.iasri.res.in) wherein, presently, information relating to databases on agricultural field experiments (excluding purely varietal trials) conducted in the country, on-farm and on-station experiments conducted under the supervision of Project Directorate of Farming Systems Research and Long Term Fertilizer Experiments are stored and maintained on-line. During the period under report, regular activities like collection, storage, validation and retrieval of experimental data were in progress. Agricultural Field Experiments Database contains more than 33000 experiments on different crops. Data relating to 1317 experiments have been entered on-line between 01 January, 2012 and 31 December, 2012.

As part of integration of distinct databases, a common script has been written to generate report based on databases from AFEIS, On-Farm Experiment-1 (Response of nutrients) and On-Station Experiment-1A (Intensification / Diversification of cropping sequence based on high value crops). The report supplies combined report of:

● Year and Objective of AFEIS experiments
● Year and crop sequences of On-Farm experiments-1 and
● Year and Treatments of On-Station experiments-1A

Some important achievements in this system are:
● On-line entry of eleven (11) experiments by the individual scientists,
● On-line data entry for On-farm Experiment-1 (Response of nutrients) by different centers. Analysis and other reports were generated for 21 centers. Rest 9 centers are being processed and
● On-line data entry for LTFE was successfully tested for Pantnagar and Bhubneshwar centers.

Clicking on ISDE link on IASRI site www.iasri.res.in gives the home page as follows:

Given below is the combined report of index of experiments for the three databases AFEIS, On_Farm Experiment-1 and On-Station Experiment-1A:
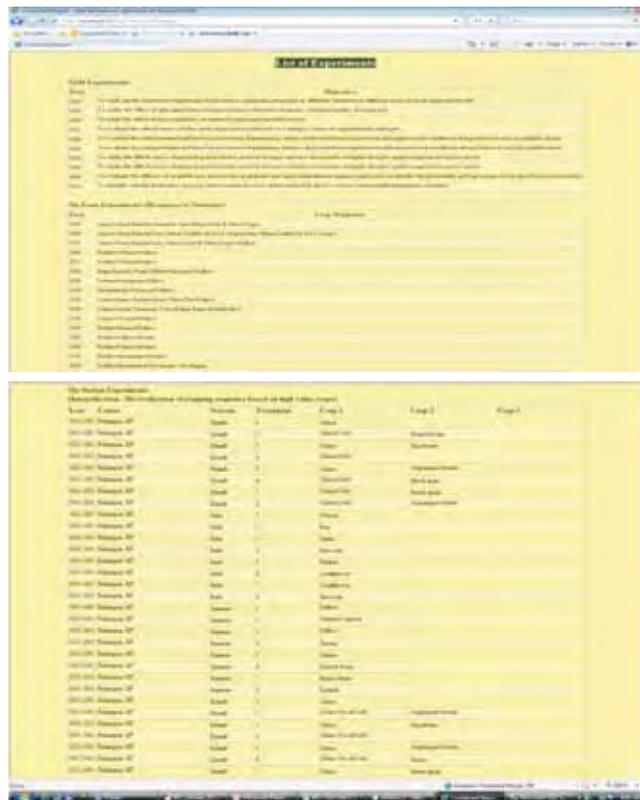
Planning, designing and analysis of experiments planned On Stations under the Project Directorate for Farming Systems Research

For the Project Directorate for Farming Systems Research, the experiments On Stations are planned and conducted under four types of research programmes viz. Development of new cropping systems; Nutrient management in cropping systems; Development of system based management practices and Maximum yield research using randomised complete block (RCB) design, factorial RCB design, split plot designs, strip plot designs and reinforced 32 x 2 balanced confounded factorial experiments.

● Analysis work for the 158 experiments conducted during the year 2010-11 has been completed. Data of 300 experiments conducted during the year 2011-12 have been received and analysis of 56 experiments has been completed. Results have been tabulated in the form of summary tables and have been sent to

the respective scientist- in-charge of the cooperating centres. The final tables of the results of the experiments have been prepared and sent to Project Directorate for Farming Systems Research, Modipuram for inclusion in the Project Report of All India Coordinated Research Project on Integrated Farming Systems.

- Combined analysis of data of the experiment (2a) [Permanent plot experiment on integrated nutrient supply system in a cereal based crop sequence] pertaining to three centres viz., Bhubaneswar, Sabour and R S Pura was performed separately for kharif and rabi seasons and it was found that mean square errors estimated over years were heterogeneous as per Bartlett's chi-square test. Further, the data were subjected to Aitken's transformation as the mean square errors estimated over years were heterogeneous for all the data set. The transformed data were further analysed and it was found that the Year × Treatment interaction was significant. Hence, the treatment effects were tested against Year × Treatment interaction and subjected to Tukey's HSD to identify the best treatment group. At Bhubaneswar centre it was found that for kharif crop T_{11}(75% recommended NPK dose through fertilizers+ 25% N through green organic matter (green leaf manuring or through Azolla)) is the best treatment giving maximum average yield and which is significantly different from the rest. At Sabour centre in rabi crop, both T_6 (75% recommended NPK dose through fertilizers) and T_{10} (100% recommended NPK dose through fertilizers) are found to be at par and are significantly different from the rest. It was also found that for kharif crop T_6 (50% recommended NPK dose through fertilizers + 50% N through Compost/FYM/Gobar gas slurry) and T_7 (75% recommended NPK dose through fertilizers + 25% N through Compost/FYM/Gobar gas slurry) are at par on giving maximum average yield. For rabi crop, T_6 (75% recommended NPK dose through fertilizers) is found to be the best treatment giving maximum average yield and which are significantly different from the rest and at RS Pura centre, It was also found that for kharif crop T_5 (100% recommended NPK dose through fertilizers), T_6 (50% recommended NPK dose through fertilizers + 50% N through Compost/FYM/Gobar gas slurry) and T_7 (75% recommended NPK dose through fertilizers + 25% N through Compost/FYM/Gobar gas slurry) are at par on giving maximum average yield. For rabi crop, T_6

(75% recommended NPK dose through fertilizers) is found to be the best treatment giving maximum average yield and which are significantly different from the rest.

Programme 2: FORECASTING, MODELLING AND SIMULATION TECHNIQUES IN BIOLOGICAL AND ECONOMIC PHENOMENA

Forecasting models using functional data analysis and nonlinear support vector regression techniques

The commodity basis pattern was perdicted for the upcoming time as a mixture of historical commodity basis pattern. It was assumed that Y(t) is distributed as a mixture of Gaussian processes

$$Y(t) \sim \sum_{k=1}^{K} \pi_k f_k(.|t)$$

where K is the number of mixture component and f_k is the density function of the kth Gaussian mixture component with mean function and covariance surface $\mu_k(t)$ and $\Sigma_k(t, t')$ respectively.

As in Functional Clustering approach, the parameters were estimated by maximizing mixture likelihood function. Programs for application of Functional clustering approach have been developed using R software. Cash price and future price data of soybean for Indore have been collected from NCDEX (National Commodity and Derivatives Exchange Limited) website. A total of 2448 data points were used for model development and 300 data points were used for model validation purpose. The methodology has been applied to predict commodity basis which is a function of cash price and future price. Using commodity basis, predicted cash price (Rs/Kg) of Indore market has been obtained. The performance of the method in forecasting cash price of soybean is given below.

Measures	Number of Clusters							
	3	4	5	6	7	8	10	12
Mean Absolute Error (MAE)	2.70	3.05	2.85	**2.47**	2.85	2.83	2.95	2.97
Root Mean Squared Error (RMSE)	3.70	4.55	3.83	**3.30**	3.92	3.88	4.04	4.09
Mean Absolute Percentage Error (MAPE)	7.47	8.26	7.94	**6.99**	7.86	7.83	8.17	8.19
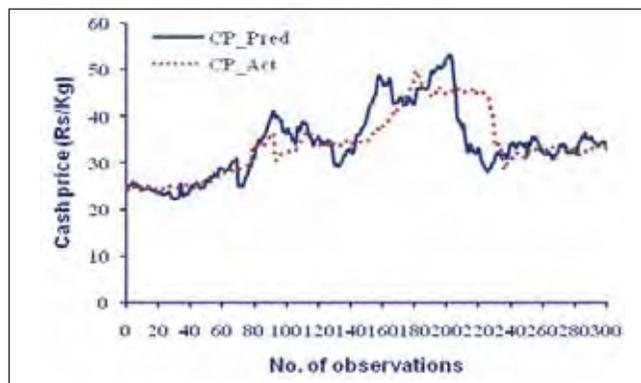
Predicted cash price (Rs./Kg) of Soybean (Indore market) along with actual data.

Programs for application of support vector regression technique have been developed using R software. The methodology has been illustrated to predict maize crop yield (response variable). Four predictor variables considered were total human labour (Rs./ha), farm power (Rs./ha), fertilizer consumption (kg./ha) and pesticide consumption (Rs./ha).

It is clear from the results that the Support vector regression technique perform better than Artificial neural network methodology in modelling and forecasting the data under consideration.

Performance of different methods in modelling

Measures	Support Vector Regression	Artificial Neural Network
MAE	7.53	8.47	7.87
RMSE	9.26	10.32	9.53

Predicted maize crop yield and Goodness-of-fit measures

S. No.	Actual Maize Yield	Predicted Maize Yield	
		Support Vector Regression	Artificial Neural Network
1.	25.00	24.56	26.28
2.	36.14	36.85	37.13
3.	43.67	41.06	40.60
4.	22.32	22.29	23.92
5.	29.94	30.39	30.91
6.	37.31	38.31	38.62
7.	32.93	31.59	31.38
8.	36.32	35.49	34.91
9.	18.75	19.27	20.29
10.	17.75	16.46	16.98
Goodness-of-fit measures			
MAE	0.92	1.45	
RMSE	1.15	1.57	
MAPE	3.04	4.99	

Weather based forewarning models for Onion Thrips (*Thripstabaci Lindeman*)

For this study data have been taken from Directorate of Onion and Garlic Research (DOGR), Pune. The field trials were conducted on different dates at fortnightly intervals (15-Jun, 01-Jul, 15-Jul, 01-Aug, 15-Aug, 01-Sep, 15-Sep, 01-Oct, 15-Oct, 01-Nov, 15-Nov, 01-Dec and 15 Dec.) in different seasons at Pune during 2000 to 2008. Models were developed for each date of planting of crop. Weekly data on weather variables starting from one week before the crop sowing up to two weeks of crop growth were considered for forewarning time (crop age) of first appearance of thrips (Y_1), whereas weather variables starting from one week before the crop sowing up to six weeks of crop growth were considered for forewarning time (crop age) of peak population of thrips (Y_2) and maximum thrips population (Y_3). Weather variables namely maximum temperature, minimum temperature, morning relative humidity, evening relative humidity, bright-sunshine hours (for rabi season only) and rainfall (for kharif season only) were considered. Models have been validated using data on subsequent years not included in developing the models. The forecasts for different characters in various dates of planting were at par with the observed ones. The percentage deviation of forecast for different characters in various dates of planting using weather indices based regression models were low for crop age at first appearance of thrips (Y_1) and crop age at peak population of thrips (Y_2) whereas deviation was higher for maximum population of thrips (Y_3). An attempt has been also made to develop fuzzy regression models, for various characters in different dates of planting. The average widths for linear regression models vis-a-vis their fuzzy counterparts were much higher for all values of fitness criterion (h). Thus, fuzzy regression methodology is more efficient than linear regression technique. The pattern over the crop season for onion thrips has been developed through non-linear models taking time as independent parameter for different dates of planting. The following was considered for this purpose,

$$Y_t = ae^{-bt} (1+de^{-bt})^{-2} + \varepsilon$$

Y_t is counts of thrips at time t. The residuals were analyzed for all dates of planting of onion thrips in different years. Similarly, for each data set Shapiro-Wilk statistic was calculated. The result showed that none of the assumptions of randomness and normality of residual was violated for any data set. The model provided a good fit to all the data sets. Thus, this model captured the fluctuations in thrips population in different years. Neural

networks models were developed for weekly thrips population considering the residuals (obtained through non-linear models taking time as independent parameter for different dates of planting) as output variable and weather indices as input variables. MLP based neural network with different hidden layers (one and two) and different number of neurons in a hidden layer with hyperbolic function as an activation function was obtained for weekly thrips population. Models have been validated using data on subsequent years not included in developing the models. Mean Absolute Percentage Error was mimimum for neural network approach which indicated that reliable forewarnings were possible well in advance using this technique.

The composed error model was used to estimate a damage function for thrips population. Using data from the experiments, the estimated damage function such as Linear, Logistic, Quadratic, Cobb–Douglas, Negative exponential and Hyperbolic were used to estimate expected yield loss due to thrips in onion in Pune. The results revealed that the mean proportional yield losses was 20.3 % for Cobb–Douglas damage function.

Pest and diseases dynamic vis-a-vis climatic change under National Initiative on Climate Resilient Agriculture (NICRA)

Weekly weather data on maximum temperature (MaxT), minimum temperature (MinT), relative humidity in the morning (RHI), and in the afternoon (RHII), rainfall (RF) / bright sunshine hours (BSH) for various locations (Kanpur: 1971-2011; Hyderabad: 1980-2010; Bangalore: 1980-2010; Pusa: 1980-2010; Pantnagar: 1970-2008; Prabhani: 1980-2010; Varanasi: 1980-2008; Pune: 1971-2008; Raipur: 1971-2011; Anantpur: 1985-2010; Mandya: 1985-2011 and Warrangal: 1982-2011) were considered. For each location, 52 weekly series, 3 seasonal series and 12 monthly series were obtained. For each series, trends through non parametric methods were obtained for various meteorological variables for different locations. Besides, the monthly rainfall series data for 141 years (1871–2011) for 30 meteorological sub-divisions in India along with temperature (maximum and minimum) for all-India and seven homogeneous regions, viz., Western Himalaya (WH), Northwest India (NWI), North Central India (NCI), Northeast India (NEI), West Coast (WC), East Coast (EC) and Interior Peninsula (IP) for the period 1901-2007 from Indian Institute of Tropical Meteorology (IITM: http://www.tropmet.res.in) were also obtained. The monthly long-term annual, seasonal and monthly trends in rainfall in different sub-divisional meteorological stations and trends for temperature (maximum and minimum) for various homogenous regions were also investigated. To ascertain the presence of statistically significant trend in climatic variables such as temperature, relative humidity, rainfall and bright sunshine hour, non-parametric Mann–Kendall (M-K) test has been employed. The M-K test checks the null hypothesis (H_o) that the data (x_1, x_2, x_3, x_N) have no trend versus the alternative hypothesis of the existence of increasing or decreasing trend.

Sen's estimator has been used for determining the magnitude of trend in meteorological time series data. Weather indices based models for Pod Borer (% pod damage due to pod borer) have been developed for Gulbarga , Kanpur and Rahuri. Initiated the disease–pest forewarning models in crops for national agro-advisory service using SATMET product at IMD (Agrimet), Pune in collaboration with NCIPM, New Delhi.

Weather based forewarning of mango pests

Weather based models have been developed for forewarning time of first appearance (for Mohanpur and Paria – Kesar and Alphanso varieties) and weekly disease incidence of powdery mildew at Paria for two varieties (Kesar and Alphanso). Weather indices have been obtained using the data on weather variables which have been used as regressors alongwith time of flush in the model for forewarning time of first appearance. For forewarning weekly disease incidence, the model was developed under the assumption that disease incidence was due to two reasons, natural disease growth pattern (non-linear model) and weather. Therefore, the model has been developed in two steps, modeling natural growth pattern and relating the deviations (from natural pattern) to appropriate lagged weather variables. Stepwise regression technique has been used to select the important variables in the model. Using these models, reliable forewarning of time of first appearance of disease (within a difference of one week) could be provided at the earliest at 2nd standard meteorological week (smw) for Mohanpur and 47th smw for Paria. Reliable forewarning of weekly per cent disease incidence (PDI) could be obtained using weather and per cent disease incidence upto preceding week. Forecasts of weekly PDI for subsequent year at Paria for the two varieties are given in the following figures .

Percent Disease Incidence for 2011-12 (Kesar)

Percent Disease Incidence for 2011-12 (Alphanso)
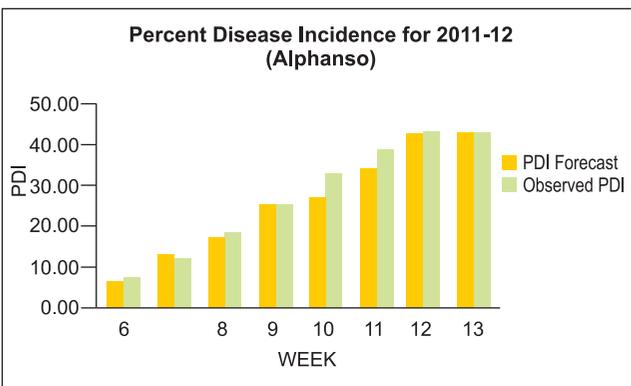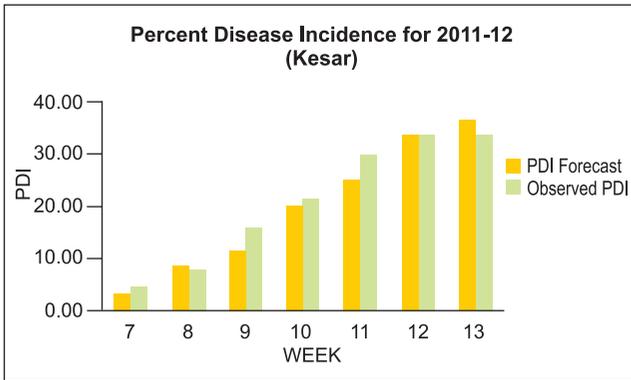
Development of forecasting module for podfly, Melanagromyzaobtusa Malloch in late pigeonpea

Development of models based on qualitative as well as quantitative data has been attempted for forewarning damage due to pod fly in late pigeonpea for Kanpur using historical data. Regression models were developed using weather indices as independent variables while % pod damage was used as dependent variable. Forecasts have been obtained for subsequent years not included in model development. Models for forecasting of % pod damage due to pod fly at different weeks of forecast were obtained. For qualitative forewarning, data in the quantitative form was classified in to two categories by taking epidemic status as 1 for the pod damage (%) more than 15 and 0 otherwise. For development of model for forewarning epidemic status, logistic regression model was used taking weather indices as regressors. The results indicated that for most of the years the approach provided correct epidemic status and forecast for % pod damage closed to the observed ones. Reliable quantitative forecasts for per cent damage due to pod fly in late pigeonpea at Kanpur could be obtained at 4th smw using data on maximum temperature, minimum temperature

and evening relative humidity. Over all epidemic status (qualitative forewarning) could be provided at first smw using data on max. temperature and evening relative humidity. Forecast for subsequent years (not included for model development) captured qualitative status correctly. The quantitative forecasts for % pod damage were 38.18 and 35.1 against observed values 31.7 and 32.5 for the years 2010-11 and 2011-12 respectively.

Study on robustness of sequential testing procedures on some distributions used in agricultural pest control

Sequential testing procedure was developed for testing the hypothesis $H_o: \alpha = \alpha_o$ against $H_1: \alpha = \alpha_1 (\alpha_1 > \alpha_o)$ for the parameter α when the other parameter 'm' is known for size-biased negative binomial distribution with probability mass function

$$P(X=x) = \binom{m+x-1}{x-1} \alpha^{x-1} (1-\alpha)^{m+1}; x = 1, 2, \ldots$$

where $0 < \alpha < 1, m > 0$.

Decision criteria has been developed over the stopping

bounds $A = \dfrac{1-\beta}{\alpha}$ and $B = \dfrac{\beta}{1-\alpha}$ with strength (α, β)

of the test. The estimating equation of

$$E\left(\left\{ \frac{\alpha_1}{\alpha_o} \right\}^{h(X_i - 1)} \left\{ \frac{1-\alpha_1}{1-\alpha_o} \right\}^{(m+1)h} \right) = 1,$$

has been derived for solving non- zero solution of 'h' for the size-biased negative binomial distribution. Sequential probability ratio tests have also been developed to test simple hypothesis for unknown parameter of a family of continuous distributions when another parameter is known.

Development of forecasting methodology for fish production from ponds of upland region

Fish growth data of three different fish species viz., grass carp, silver carp and common carp obtained from polyponds and earthen ponds were thoroughly analyzed. Different growth models were attempted to fit the growth datasets of fish obtained from polyponds and earthen ponds. There was no extreme autocorrelation and the assumption of homoscedastic error structure was not

violated. Also, residual analyses showed that the randomness assumption and normality assumption were fulfilled. However, the high correlation among the estimated parameters and the nonlinear behavior of the estimated parameters were of concern. Consequently, partially reparameterized versions of Gompertz and logistic models with expected value parameters were developed.

Weather based yield forecasts for rice and wheat using non-linear regression techniques

Weather data on temperature (maximum and minimum), relative humidity and total rainfall from the year 1970-71 to 2009-10 have been utilized for model fitting and two years data 2008-09 and 2009-10 used for validation of the model. Crops yield forecast models have been developed for different districts of Uttar Pradesh using weekly weather data. Residuals were obtained from the selected nonlinear models and linear models. Weather Indices (WI) were obtained and WI based regression models were developed using WI as independent variables while character under study such as crop yield as dependent variable for wheat and rice crop. Comparison of forecast models developed through different approaches was done on the basis of RMSE and MAPE. The results indicated that non-linear models based approach provided better models (or at par) for forecasting in the comparison of linear model approach. The performance of these forecasts was judged on the basis of Mean Absolute Percentage Error of forecasts.

Development of crop yield forecasting models using Generalized Autoregressive Conditional Heteroscedastic (GARCH) and Wavelet techniques

Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) time-series model along with its estimation procedure was studied. Five models at five important stages of wheat growth were developed by including the most important weather variables. The weekly maximum temperature at Crown root initiation (CRI) stage, tillering stage, anthesis stage, milk stage and dough stage and evapotranspiration at CRI stage were used for model development. As an illustration, ARIMAX models were employed for forecasting of wheat yield in Kanpur district of Uttar Pradesh. Comparative study of the fitted models was carried out from the viewpoint of Relative mean absolute prediction error (RMAPE). It was demonstrated that the ARIMAX methodology was able to provide pre-harvest forecasts based on weather variables at various stages of wheat crop growth, starting from CRI stage (21 days after sowing) to dough stage (126 days after sowing).

It was observed that, as wheat crop grew towards maturity, pre-harvest forecasts got closer to actual values.

Autoregressive integrated moving average with exogenous variable-Generalized autoregressive conditional heteroscedastic (ARIMAX-GARCH) methodology was employed for describing volatile data by incorporating the exogenous variables in the mean-model. As an illustration, ARIMAX and ARIMAX-GARCH models were employed for modelling and forecasting of wheat yield for Kanpur district of Uttar Pradesh, India. Comparative study of the fitted models was carried out from the viewpoint of dynamic one-step ahead forecast error variance along with Mean square prediction error (MSPE), Mean absolute prediction error (MAPE) and Relative mean absolute prediction error (RMAPE). The formulae for more than one-step ahead out-of-sample forecasts along with forecast error variances for the fitted ARIMAX-GARCH model were derived analytically by recursive use of conditional expectation. Superiority of ARIMAX-GARCH model over ARIMAX approach was demonstrated for the data under consideration. For the selected ARIMAX-GARCH model, Maximum overlap discrete wavelet transform (MODWT) coefficients were computed for the weekly maximum temperature series at CRI stage for wheat yield time-series data of Kanpur district for forecasting of maximum temperature at CRI stage. After obtaining the forecast of maximum temperature by Wavelet methodology, these forecasts were used for forecasting of wheat yield by the model developed.

A study of stochastic volatility (SV) models through particle filtering

Estimation procedure for fitting SV model proposed by Taylor and SVM model proposed by Koopman and Uspensky has been developed through Particle filtering. Code for the same has also been developed in MATLAB, 2007. Formulae for optimal out-of-sample forecasts for SV model have been derived. The developed estimation procedure for fitting SV model has been applied for modelling and forecasting the volatile India's monthly Basmati rice export data. It has been shown, using appropriate statistical measures, that SV model fitted through Particle filter performed relatively better than GARCH model for modelling as well as forecasting.

Development of methodology for estimation of compound growth rate and its web-based solution

Compound growth rates were estimated in respect of non-monotonic situations for all the three possibilities, viz Over-

damped, critically damped, and under-damped. Code was constructed in R language for estimation of compound growth rate for the above three non-monotonic situations. As an illustration, India's nine oilseeds production data (in Million tonnes) during 1986-87 to 2002-03 was considered.

To assess goodness of fit of the model, Mean Square Error (MSE) were computed. Using the Critically-damped model, the compound growth rate for the nine oilseeds production in India during the period 1986-87 to 2002-03 were computed. Codes were constructed in R language for estimation of compound growth rate using nonparametric methodologies applying (a) Moving average technique under Time Domain approach, and (b) Kernel Smoothing technique for error dependent process under Time Domain approach by using modified plug-in technique. The optimum bandwidth was obtained to estimate time varying growth rate.

Code was also constructed in SAS IML for estimation of compound growth rate by (a) applying local linear kernel smoothing using various bandwidth, and (b) obtaining interval estimate of growth rate based on local linear kernel smoothing.

As an illustration, India's total foodgrain production data (in Million tonnes) during 1960-61 to 2010-11 was considered.

It was observed that the error series was long range dependent. Therefore, a novel approach of iterative estimation of the optimal bandwidth under long range dependence was formulated. The steps were as follows:

i) Estimate an "optimal" bandwidth $\hat{h}_{opt}/T$, assuming only short range dependent errors.

ii) Let $h'_0/T = \hat{h}_{opt}/T$.

iii) For j = 1,2,... estimate g(.) using h'_{j-1} and let $\hat{X}_i = r_i - \hat{g}_{h'_{(j-1)}/T}(i/n)$. Estimate long memory parameters to compute AIMSE.

iv) Estimate second derivative of g(.) and denoted by $\hat{g}_2(u, h_2/T)$ where $h_2/T = (h'_{j-1}/T)T^{\alpha/(4p+2\alpha)}$.

v) Use estimated second derivative of g(.) to finally calculate AMISE over various values of h/T and get optimal h'_j.

vi) Repeat process (iii) to (v) until convergence is achieved.

Technology forecasting in "Visioning Policy Analysis and Gender (VPAGe)"

Technology Forecasting (TF) tools have been employed to forecast future technological needs and trends in Indian agriculture. TF and Technology Assessment (TA) have been done with the following tools: Analytical Hierarchy Process, Brainstorming, Cross impact analysis, Fisher Pry/ Pearl, Gompertz and Lotka-Volterra substitution models, Framework Forecasting, Scientometrics and Multi-Dimensional Scaling. The agricultural subdomains/ commodities considered were Plant Breeding and Genetics, Rainfed Agriculture, Fisheries, Cotton and Rice. Implications of frontier sciences viz. Remote Sensing (RS) and Information and Communication Technology (ICT) on agricultural R&D have also been done.

The future technological needs in "Plant Genetics and Breeding" (PG&B) domain for sustainable agriculture have been ascertained by Brainstorming workshop. Scientometric analysis in PG&B of agriculture revealed that India is focusing more not only on subdomains like abiotic and biotic stresses but also on niche areas like bioinformatics, Marker Assisted Selection (MAS), transgenics etc. In rainfed agriculture, Analytic Hierarchy Process (AHP) revealed that priority setting on extension, policy and biophysical sectors came out to be 29%, whereas for socio-economic and technological sectors it was 7%. By Multi-Dimensional Scaling (MDS) approach, it was found that water harvesting and water saving technologies were the best strategies to cope with climate change in the coming years among the different technologies considered. In addition, the study revealed that stability of crops should have highest research priority followed by early maturity, broad adoption, stress resistance and high yield potential in achieving high productivity in rainfed areas. In Fisheries sector, AHP was employed to build a hierarchy consisting of "decision criteria" leading to various "alternative courses of actions/ factors" within each of them for achieving the goal of a well-established fisheries sector and the AHP tree thus obtained showed that both 'technological' (with 'fuel saving technologies' alternative carrying highest priority of above 18%) and 'institutional and policy' criteria (with 'infrastructural facilities in fish landing centres' alternative carrying a highest priority of above 13%) contributed 45% each in achieving the set goal while the 'extension' criteria contributed 10%.

The substitution models viz., Fisher-Pry/Pearl, Gompertz and Lotka Volterra models were fitted for data on area under adoption of Bt Cotton in India. It was found that by 2013, if the same trend continues, all of area under Indian Cotton will be substituted by Bt Cotton. Kane's KSIM cross impact simulation model was utilised for inferring about the future behaviour of variables of Indian cotton viz., Production, Export, Import and Supply over time. It

was inferred from the study that if no curb on imports was done then it may increase over time in the long run. A conceptual cum critical TF technique namely, Framework Forecasting was also attempted to study the cotton scenario in future both in the Indian as well as in the global context. While the baseline future envisioned the dominance of Bt cotton in almost all the major countries, the alternative future to Cotton production were also outlined such as plausible options like water resistant, fire resistant, wrinkle free and drought tolerant cottons. For rice, it was found that, if technological needs are fulfilled in low productivity districts, it would lead to an increase of 7% in production. Priority areas for applications of RS on agricultural research and development have been identified. Agricultural extension was one of the dominating applications of ICT over others. Thus, TF and TA tools have been successfully employed to aid decision making in agriculture.

Enhancing resilience of agriculture to climate change through technologies, institutions and policies

Agro-climatic zone-wise trend in temperature and precipitation have been estimated by both parametric and nonparametric methods. Wavelet analysis in frequency domain has been used for detection of trend in rainfall in above zones. It was found that there is significant trend present in all the zones.

Study of asymmetry in retail wholesale price transmission for selected essential commodities

It has been observed that changes in wholesale prices are neither fully nor partially transmitted to consumer prices via retail price. The study indicated that the retail traders are more active and are not following the price signals coming from wholesale traders even in short run. The fall in wholesale price is partially transmitted whereas the rise in wholesale prices is more than fully transmitted to the consumers. In both the situations the retail traders are earning a huge profit from trading. The results of Error Correction Model for Rice markets indicated that there existed persistence asymmetry in marketing of Rice and the extent of asymmetry was more acute in Hyderabad (Rs.1.27) followed by Cuttack (Rs.1.21) and Delhi (Rs.0.99) markets. The lowest extent of asymmetry was found in Amritsar (Rs.0.77) market. For Wheat markets, the results indicated that asymmetry do exist in marketing of Wheat and the extent of asymmetry was more acute

in Chennai (Rs.1.33) followed by Hyderabad (Rs.1.20) and Delhi (Rs.1.01). The lowest extent of asymmetry was found in Bangaluru market (Rs.0.89). For Gram markets, the extent of asymmetry was more acute in Chittoori (Rs.1.18) followed by Bhopal (Rs.1.08), Delhi (Rs.0.99). The lowest extent of asymmetry was found in Sri Ganganagar market (Rs.0.96). For Moong, the extent of asymmetry was more acute in Delhi (Rs.1.16) followed by Kolkata (Rs.1.08). The lowest extent of asymmetry was found in Chennai market (Rs.0.87). In Rape seed & Mustard Oil, the extent of asymmetry was more acute in Delhi (Rs.1.44) followed by Kolkata (Rs.0.94) with lowest extent of asymmetry found in Kanpur market (Rs.0.91). For Sugar trading, the extent of asymmetry was more acute in Hyderabad (Rs.1.49) followed by Delhi (Rs.1.27) and Sri Gaganagar (0.67). The lowest extent of asymmetry was found in Kolkata market (Rs.0.52). In trading of Apples, the extent of asymmetry was more acute in Delhi (Rs.1.02) followed by Lucknow (Rs.0.90) and Chennai (0.86). In trading of Onions, the extent of asymmetry was more acute in Chennai (Rs.1.06), Kolkata (1.03), Mumbai (0.97) and in Hyderabad (0.90). The lowest extent of asymmetry was found in Lucknow market (Rs.0.82). The varying level of asymmetry practically is an indication of market efficiency attained in different markets of the same commodity.

The value of long run adjustment was almost close to zero in all the markets indicating that most of the changes in wholesale markets were already transmitted to consumers in the short run and very little was left for long run adjustments.

An econometric study of water markets in canal command area of North-Western Rajasthan

Secondary data on volumetric statistics on groundwater resources, groundwater table and source-wise irrigated area were collected and analysed. The structure of water markets in North-western Rajasthan was studied. The cropping pattern and productivity of major crops were also examined under different forms of water markets. It was observed that three-fifths of net sown area in North-Western region were irrigated. The region was dominated by canal irrigation. Although, the annual growth in groundwater irrigated area was impressive (14 per cent) during 2000-01 to 2008-09, there was a further scope for groundwater development in the region as its development was 46 and 80 per cent in Sri Ganganagar and Hanumangarh districts in 2009.

Programme 3: DEVELOPMENT OF TECHNIQUES FOR PLANNING AND EXECUTION OF SURVEYS AND STATISTICAL APPLICATIONS OF GIS IN AGRICULTURAL SYSTEMS

Farm power machinery use protocol and management for sustainable crop production

For a comprehensive assessment of mechanization and agricultural machinery manufacturing/supply scenario in Punjab and Rajasthan, primary data collection on farm power-machinery and their uses in the selected villages of Ludhiana district of Punjab State has been completed and is in progress in Rajasthan State. Secondary data on population of farm power resources, crop production input uses, mechanization status etc. was acquired. Data analysis of the acquired secondary data was done. The linkages between the research/ educational institutions and the manufacturer of agricultural machinery were studied. Expert system for efficient farm machinery selection is being developed. The flow chart for development of expert system has been prepared.

Small area inference using survey weights

In this era of decentralization, the thrust of planning process has shifted from macro to micro level. The thrust of research efforts has shifted to development of precise estimators for small areas. Small area estimation (SAE) techniques are used to produce reliable estimates for small areas. As a consequence, SAE is now very common in survey sampling, with several methods proposed in the literature. However, research continues on the identification of SAE techniques that are efficient and also simple to implement, with estimation of mean squared error (MSE) a particular problem. Unit level linear mixed models are often used in SAE, and the empirical best linear unbiased prediction (EBLUP) based approach is widely used for producing small area estimates under such models and proven to be efficient. However, this approach of SAE does not make use of the unit level survey weights. As a result, small area estimator based on this approach is not design consistent unless the sampling design is self-weighting within areas. The Pseudo empirical best linear unbiased prediction (Pseudo-EBLUP) approach overcomes this limitation by using sample weights and also leads to design consistent small area estimator.

A bias-robust method for estimating the MSE of Pseudo-EBLUP estimator that remain approximately unbiased under failure of assumptions about second order moments has been developed. The proposed estimator is based on conditional approach of MSE estimation and provides area specific MSE estimates for the Pseudo-EBLUP. In addition, the conditional approach of MSE estimation leads to estimator of MSE that is simpler to implement, and potentially more robust. In particular, it performed reasonably well overall in terms of estimating true MSE for the Pseudo-EBLUP.

Spatial nonstationarity in small area estimation under area level model

In many Small area estimation (SAE) problems, it is not possible to use the unit level small area model simply because of the unavailability of the unit level data. In such circumstances, SAE is carried out under area level small area models. The Fay-Herriot model (Fay and Herriot 1979) is widely used area level model in SAE. This model relates small area direct survey estimates to area-specific covariates, often obtained from various administrative and census records etc. The SAE under this model is one of the most popular methods used by private and public agencies because of its flexibility in combining different sources of information and explaining different sources of errors. There are situations (for example agricultural, environmental and economic data), where the relationship between variable of interest and covariates is not constant over the study space, a phenomenon referred to as spatial nonstationarity. This area level model does not account for spatial nonstationarity present in the data. A geographically weighted pseudo empirical best linear unbiased predictor (GWEBLUP) for small area means was introduced under geographically weighted version of area level model. In SAE, the mean squared error (MSE) estimates are required for measuring uncertainty or reliability and producing the confidence interval of small area estimates. The MSE of GWEBLUP for small area means was developed, and then asymptotically unbiased estimator of the MSE with the second-order accuracy was derived based on the Taylor series approximation. Empirical studies were undertaken to examine the empirical performance of the proposed MSE estimation method. The MSE estimator of the GWEBLUP appeared to provide good approximation of true MSE along with desirable level of coverage and stability performance.

Study to develop methodology for crop acreage estimation under cloud cover in the satellite imageries

The techniques of simple kriging, stratified kriging, simple co-kriging, stratified co-kriging were applied to remove

cloud in the satellite images. Spatial imputation techniques were evolved for generation of cloud free images based on row-wise pixels, column -wise pixels, both row-wise and column-wise pixels, neighbouring pixels and by ratio and regression approach. Cloud free images were generated using all these techniques and then these techniques were compared by estimating area under paddy crop from the generated cloud free images.

Study of sample sizes for estimation of area and production of food grain crops

Data of Crop Cutting Experiments (CCEs) for different crops (having smaller sample sizes) pertaining to number of States under Improvement of Crop Statistics (ICS) scheme were obtained from NSSO for the agricultural year 2010-11. Estimates of average yield for wheat and paddy crops at State level were obtained with suitable precision, however, for other food grain crops like bajra, gram, black gram, green gram, horse gram, jowar, maize, barley, ragi etc., these were obtained with very high estimates of percentage standard errors. Sample sizes for different levels of margin of errors were worked out for estimation of average yield of different crops. ICS scheme related crop area data for different districts of some States for the agricultural year 2010-11 was analysed. Estimates of area under different crops were obtained with very high estimates of percentage standard errors.

Study to develop an alternative methodology for estimation of cotton production

The estimates of average yield of cotton along with its percentage standard error has been obtained for all the five districts of Maharashtra namely, Aurangabad, Buldana, Aurangabad, Jalna and Jalgaon using the proposed methodology. In the process of exploring other sampling designs, estimation procedure for estimating average yield of cotton using double sampling under stratified two stage sampling framework was also developed. The survey for primary data collection for validation of the developed alternative methodology was planned. Two districts of Maharashtra State namely, Aurangabad and Amravati and two districts of A.P. State namely, Warangal and Guntur were selected for validation. Training for data collection was imparted in both the districts of Maharashtra and A.P. States. Field data collection in both the states has been completed with the help of respective State Govt. officials. Supervision of data collection was done in both the states at regular intervals. The process of acquisition of data from both the

states is in progress. The developed alternative sampling methodology was presented in the National Workshop on Improvement of Agricultural Statistics organized by Directorate of Economics and Statistics, Ministry of Agriculture, Govt. of India at New Delhi and the methodology is likely to be adopted in all the cotton growing states of the country after validation as announced in the workshop.

A study on calibration estimators of finite population total for two stage sampling design

In sample surveys, auxiliary information on the finite population is often used to increase the precision of estimators of finite population total or mean or distribution function. In the simplest settings, ratio and regression estimators incorporate known finite population parameters of auxiliary variables. The calibration approach proposed by Deville and Sarndal (1992) is one of the other techniques widely used for making efficient use of auxiliary information in survey estimation. However, in many cases, the population could be spread over a wide area entailing very high travel expenses for the personal interviewers. In addition, efficient supervision of the field work can be difficult, which could result in high non-response rates and severe measurement errors. In such situations, two-stage sampling designs are preferably considered. Different calibration estimators of the finite population total have been developed based on the assumption that the population level auxiliary information is available both at the psu and ssu level under two stage sampling design. The variance of these estimators along with their estimators of variance have also been developed under two-stage sampling design. In particular, twelve different situations of data availability have been considered and estimators have been obtained. The empirical evaluations revealed that all the developed calibration approach based estimators under two-stage sampling design were better than the usual estimator under two-stage sampling design with no auxiliary information.

Impact assessment of agroforestry model in Vaishali district of Bihar State

Impact of agroforestry on socio-economic conditions of farmers of Vaishali district of Bihar State was to be assessed due to the ICFRE agroforestry project launched in the district. Sample selection was done as per the proposed sampling design i.e. stratified two stage sampling treating blocks as strata, villages as first stage units and households as ultimate stage units. Planning

for survey and imparting training to field investigators for primary data collection was done. The designing of schedules for primary data collection was done. The designed schedules were finalized. Testing of the designed schedules was done in the field during a visit to Forestry Research and Extension Centre (FREC), Patna and Vaishali district by IASRI officials during 15-22 January, 2013. The schedules were modified after testing in the field and were translated in Hindi language. Training (class room as well as in the field) to field investigators for data collection was imparted using Hindi version of schedules. Supervision of data collection was done and doubts of field investigators in collection of data and filling the schedules were clarified. Second round of supervision was done by IASRI officials during 11-17 March, 2013.

Small area estimation for skewed data

In many surveys, e.g. agricultural, environmental and business surveys, data are typically skewed and linear model assumptions are questionable. Commonly used methods for small area estimation are based on the assumption that a linear mixed model can be used to characterize the relationship between the survey variable Y and an auxiliary variable X in the small areas of interest. In particular, empirical best linear unbiased prediction (EBLUP, Prasad and Rao, 1990) and model-based direct (MBDE) estimation (Chambers and Chandra, 2009) are typically based on linear model assumptions. However, when the data are skewed, the relationship between Y and X may not be linear in the original (raw) scale, but can be linear in a transformed scale, e.g. the logarithmic scale. In such cases estimation based on a linear model for Y is expected to be inefficient, and an appropriate technique for small area estimation should then be based on a linear mixed model for a transformed version of Y. The use of transform variable based estimation was explored when carrying out small area estimation for skewed data, focussing on the widely used log-log transformation. An empirical best predictor for small area means, in the sense that it has minimum mean squared error in the class of unbiased predictors, has been developed. The proposed method is expected to be more efficient than the existing methods of small area estimation for skewed data.

Assessment of quantitative harvest and post harvest losses of major crops/commodities in India

Sampling frame for selection of districts, blocks and villages for the study was prepared using Census 2001 data. Selection of 120 districts for the study, 2 blocks from each selected district and 5 villages from each selected block was done. The schedules and instruction manual for primary data collection in all 120 districts were finalized. An orientation meeting was held at Agricultural Research Station, Durgapura, Jaipur in which training for primary data collection and enquiry based data entry software was imparted to the Research Engineers and Principal Investigators under AICRP on Post Harvest Technology. The updation of observation based data entry software was done. An Orientation meeting was held at CIPHET, Ludhiana in which training for updated observation based data entry software was imparted to the Research Engineers and Principal Investigators under AICRP on Post Harvest Technology.

Calibration based product estimator in single and two phase sampling

Auxiliary information is often used by survey statisticians to increase the precision of estimators of commonly used parameters. Some examples of estimators of population mean or population total, which use auxiliary information, are ratio and regression estimators. The ratio estimator, in particular, is useful when there is positive correlation between the study and the auxiliary variable. However, in many practical situations, the study and the associated auxiliary variable are negatively correlated. In these situations, the product estimator, developed by Murthy (1964), is a viable alternative. A new product estimator has been proposed using the calibration approach under the assumption that a negative correlation exists between the study and the auxiliary variable. In addition, expressions for bias, variance and variance estimator of the proposed calibration approach based estimator have also been developed. The calibration approach was used a second time on the variance estimator of the proposed estimator for further improvement in variance estimator. Empirical evaluations showed that the developed methodology was reliable and stable alternative to traditional product estimator for estimation of population parameters.

Estimation of finite population total using calibration based regression type estimator for inverse relationship between study and auxiliary variable

The ratio estimator is widely used for estimation of finite population mean or total when the study and the auxiliary variable have positive correlation and their regression line passes through origin. However, in real life data, the study

and the associated auxiliary variable are sometime negatively correlated. In this case, the product estimators discussed in Murthy (1964) and Sud *et al.* (2013) can be applied for estimation of finite population total. These estimators are highly biased if the fitted regression line between study and the associated auxiliary variable does not pass through origin. For such cases, a regression type estimator has been developed using the calibration approach under the assumption that a negative correlation exists between the study and the auxiliary variable and the regression line does not pass through the origin. The expressions for bias, variance and variance estimator of the new estimator have also been obtained. The improved performance of the proposed estimator over the usual regression estimator is demonstrated through a simulation study. The double sampling approach based calibrated estimator has also been dealt with for the situation when the auxiliary information is not available. The improved performance of the double sampling based calibration estimator over the usual regression estimator, in terms of the criterion of mean square error of the estimator, is also demonstrated through a simulation study.

Construction of food security index

Sub-indices of Food Security Index (FSI) were constructed for all the three states namely, U.P., Bihar and Punjab. Thematic maps were generated based on constructed FSI and their sub-indices for all the three States using Geographic Information System (GIS).

Sample Survey Resources Server

Sample Survey Resources Server, hosted at www.iasri.res.in, is a web resource created with a goal to disseminate research in theory, application and computational aspects of sample survey among (i) the statisticians in academia, (ii) the practicing statisticians involved in advisory and consultancy services, (iii) scientists in the National Agricultural Research System, and (iv) the statisticians involved in conducting large scale sample surveys, particularly in the National Statistical System with focus on Agricultural Statistical System. This resource focuses on propagating research in sample survey including designing a survey, estimation procedures with support of online software for computing purposes, analysis of survey data, e-learning, etc. This resource is useful to surveyors in agricultural sciences, biological sciences, social sciences, industry and in statistical organizations in the centre and the states in planning and designing surveys and then in analyzing the complex survey data generated.

An important feature of this web resource is that it provides an online calculator for the determination of sample size for estimating the population mean or population proportion for simple random sampling without or with replacement sampling design. An exhaustive list of books on sampling theory is also available on the server. Lectures on glossary of sampling theory, fundamentals of survey sampling and small area estimation under area level model along with the R code for analysis of data serve as E-learning material for the readers. Other links useful for survey statisticians are also available at this resource.

Among the other important features of the web resource is the link "Ask a Question" through which the user can ask questions and seek clarifications through Email. This link is partially in operation and needs to be further strengthened.

It is expected that the material provided at this server would help the survey practitioners in general and in agricultural sciences in particular and those survey practitioners involved in planning, designing and analysis of large scale complex surveys in the national statistical system in improving the quality of research in their respective sciences and making their research globally competitive.

A snap shot of the web resource is given below:

Agricultural Research Data Book (ARDB)

Agricultural research is a vital input for planned growth and sustainable development of agriculture in the country. Indian Council of Agricultural Research, being an apex scientific organization at national level, plays a crucial role in promoting and accelerating the use of science and technology programmes relating to agricultural research and education. It also provides assistance and

support in demonstrating the use of new technologies in agriculture.

Information pertaining to agricultural research, education and related aspects available from different sources is scattered over various types of published and unpublished records. The Agricultural Research Data Book (ARDB) 2012, which is the fifteenth in the series, is an attempt to put together main components/indicators of such information. The Data Book comprising 172 Tables is organized, for the purpose of convenience of the users, into ten sections namely, Natural Resources; Agricultural Inputs; Animal Husbandry, Dairying and Fisheries; Horticulture; Production and Productivity; Agricultural Engineering & Produce Management; Export & Import; India's Position in World Agriculture; Investment in Agricultural Research & Education; and Human Resources under National Agricultural Research System (NARS). This edition contains latest information / data as available in the country by the end of May, 2012. In ARDB 2012, some value editions like predicting the future year production of food grain crops etc., based on previous years data using statistical models, pictorial/graphical representations of data have been done. For depicting state-wise data, thematic maps have been prepared using GIS. Efforts have been made to incorporate the comments and suggestions received from various users. The first ARDB was brought out in the year 1996 and since then, it has been brought out every year.

Visioning, Policy Analysis and Gender (V-PAGe) (Sub-Prog. III): Policy analysis & market intelligence (NAIP Project)

Under agricultural commodity futures, trade, the futures and spot prices of soybean in NCDEX exchange were found to be cointegrated and sharing a long run relationship. There is a causality flow from futures markets towards spot markets indicating information flow from futures to spot markets. At the same time, there is also a reverse information flow happening in case of August-2008 and June-2009 contracts signifying price discovery in both futures and spot markets. This finding to a large extent answers to the apprehensions about the destabilizing impact of commodity futures markets in India. The investigation into potato farmers' participation in future markets revealed that there is a scope for farmers for participating in commodity markets as their marketed surplus is higher and if the cold storage owners can be used as agency for financing, providing reliable market intelligence and quality and quantity certification. The hypothecation of commodities in the warehouse should be treated towards margin to facilitate farmers' participation in futures market.

Programme 4: DEVELOPMENT OF STATISTICAL TECHNIQUES FOR GENETICS/COMPUTATIONAL BIOLOGY AND APPLICATIONS OF BIOINFORMATICS IN AGRICULTURAL RESEARCH

Study of synonymous codon usage and its relation with gene expressivity in genomes of halophilic bacteria

Codon selection pattern was studied in three different organisms which inhabit different habitats. This study helped in exploring the factors which were governing codon selection pattern of these organisms. It also provided an insight into gene expression level of these organisms.

From the findings, it was suggested that in all the three bacterium isolated from moderate, low and high halophilic conditions, there were a large number of genes with high G+C content, and G+C content at the third codon position is higher than that of A+T. Accordingly, it was suggested that the usage frequency of codons ending with G or C bases was higher than that ending with A or T bases. High level of heterogeneity was seen within the genes of various functions in all the organisms. It was observed that in all cases, codon usage was largely determined by compositional constraints. Translational selection also seemed to affect shaping the codon usage variation among the genes. Therefore, the variation in codon usage among the genes might be due to mutational bias at the DNA level and natural selection acting at the level of mRNA translation. Length of the genes also affected the codon usage bias, while aromaticity and hydrophobicity of the encoded proteins played minor role in shaping codon usage bias.

Blastn was used for finding similarity between highly expressed genes of S. ruber and all the gene sequences of C. salexigens. For this, the highly expressed gene sequences of S. ruber were used as a query against database containing all the sequences of moderately halophilic bacterium, C. salexigens and also against sequence database of non-halophilic bacterium, Rhizobium. Blastn ended with no sequence similarity in these organisms. It may be inferred that the functional aspects of all the genes in extremely halophilic, moderately halophilic and non-halophilic organisms are diverse and thus the similarity in their genes could not be recorded.

Agriculturally important microbes such as Rhizobium, Azotobacter etc. are helpful for increasing soil fertility, nitrogen fixation and are used as biofertilizers.

Knowledge of optimal codons might provide more useful information for gene engineering and/or evolution studies related to nitrogen fixation. The findings may facilitate in developing bio-fertilizers for improving soil fertility if expression of these identified genes was induced in case of non-symbiotic bacterium. Further, identification of salt-tolerant genes in halophilic bacteria and transfer of such genes to other agriculturally important bacteria will help them to adapt under conditions of salinity environment. Inoculation of these bacteria with salt-tolerant characteristics in in-vivo condition may facilitate the enhancement of crop productivity.

Algorithm for gene classification based on gene expression data

The gene expression data of chickpea under abiotic stress was obtained and the consensus clustering was applied on it that identified the co-regulated genes. Customization of penalized classifier called Least Absolute Shrinkage and Selection Operator (LASSO) using kernel function was done. Code of customized classifier has been written in MATLAB and applied to gene expression data of Arabidopsis thaliana (Model Plant). Accuracy of the developed model has been tested through Leave One Out Cross Validation technique.

Agriculturally important plants suffer from different abiotic stress conditions like draught, salinity, temperature and biotic stresses like different diseases. Classification of genes according to different abiotic and biotic stress will be of great help in genetic engineering and breeding program to develop the varieties which may be able to survive in stress conditions.

Analysis and determination of antimicrobial peptides: A machine learning approach

A database of the antimicrobial peptides for cattle collected from various source like APD2, CAMP, AMSDb, DAMP etc. and published literature was developed using PHP and MySQL. It is available at http://cabindb.iasri.res.in/amp/database.html. Analysis for N and C terminals were done for AMPs followed by occurrence of all 20 natural amino acids at both termini. It was observed that at N terminal, residues like R, L and C were more preferred while at C terminal, residues like R, K and C were more preferred. Thus, it can be concluded that both termini preferred more of positively charged residues. The dataset was further analyzed and prediction

models using Artificial Neural Network for classification of antimicrobial peptides were developed with accuracy of 94.95.

In silico identification of abiotic stress (salinity) responsive transcription factors and their cis-regulatory elements in grape

In India, grape cultivation is concentrated in the semi arid region and a number of factors like unpredictable rains, poor irrigation, water quality, excessive use of fertilizers etc. adds to the already prevalent saline soil conditions leading to decline in productivity. The understanding of salinity stress will be greatly enhanced by elucidating the structural and regulatory genes responsible for developmental and physiological processes under stress response. With this objective, a total of 16785 salt stress Expressed Sequence Tags (ESTs) of *Vitis vinifera we*re extracted from public domain. 1201 contigs were assembled to be taken for BLAST, mapping, annotation along with the domain identification using BLAST2GO tool. The identified 243 contigs were taken individually to Phytozome v 8.0 for BLAST followed by Fgenesh to predict the genes. 21 genes were found and 7 classes of transcription factors (Dof, GATA, ARF, ERF, MYB, RAV and WRKY) were detected in the identified genes.

Bio-prospecting of genes and allele mining for abiotic stress tolerance

An *in silico* approach was followed to identify the key residues of Mn-SOD that regulate the salt stress tolerance, across species. A total of 22 species including microbes, fishes, animals and plants were selected for this study

and the tertiary structures of the Mn-SODs were predicted through homology modelling whose three dimensional structures were not available in PDB. The predicted structures were submitted at PMDB, which were further accepted. The residues conserved throughout species were then identified from sequence alignment and further their conservancy was studied both at structural and functional level. The residues thus found to be conserved are reported as the key residues that play a significant role in salt stress tolerance mechanism in terms of contribution to the cofactor and substrate specificity, active site gateway formation and protein stability.

Low temperature tolerance is a desired trait for rice grown in North Eastern hill region of India. Efforts to find newer and/or better alleles for this abiotic stress has led to identification of Single Nucleotide Polymorphisms (SNPs) across the ORFs of transcription factor DREB (Dehydration Responsive Element Binding protein) viz. DREB1A and DREB1B induced in response to low temperature in upland genotype UR14. Domain analysis of DREB1A and 1B revealed that the proteins has a DNA binding domain (AP2) but the SNPs do not lie on it, which reflect that the SNPs might not affect the main function of the protein. In order to verify this fact at structural level, tertiary structures of DREB proteins (both normal and with SNP) have been predicted by fold recognition method and validated by standard procedures.

The analysis of phenotypic data received from one of the consortium centres highlights the prevalence of wide spectrum of genetic variability across SSILs and STILs that could be a potential repository for dissecting the molecular basis of salinity stress responses.

Whole genome association analysis in complex diseases: An Indian initiative

Whole genome SNP data for Ulcerative colitis diseases was analysed by Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF) to identify SNPs associated with the trait. The performance of the LASSO, RF vis-a-vis Support Vector Machines (SVMs) for prediction of disease status was assessed through prediction metrics.

Identification and characterization of genomic sequences responsible for salinity-stress in cereal crops-rice, sorghum, maize and wheat

Under this study, 116242 expressed sequence tags (ESTs) were downloaded, clustered and assembled into 11042

contigs after pre-processing the ESTs by removing the polyA/polyT tail. Biological functions were assigned to 11000 contigs through Gene Ontology (GO) and the remaining contigs showed no functional assignment. The remaining contigs were mapped on to maize chromosomes and full length gene sequences were designed. Altogether, 9 such genomic region were obtained, with TSS (transcription start site), PolyA tails at the extremes ends and CDS (coding sequences) in between, as novel candidate genes. These designed candidate genes were further validated by means of promoter analysis.

Computational identification of putative miRNAs and their characterization in *Heliothis virescens*

ESTs of *Heliothis virescens* and miRNA of insect species has been downloaded from miRBase. A BLAST search has been carried out between these. These searches have been filtered by applying various criterions. Sliding windows concept has been applied and then the secondary structures of these sequences have been found using RNAfold program. MiPred software is used for the classification of real precursor from pseudo precursor sequences. Targets of these have been found using MiRanda program. Four novel putative miRNA were identified from H. virescens from ESTs sequences based on homology search. Their targeted proteins were also identified. These findings also strengthened the bioinformatics approach for new miRNAs identification from insect species whose genome was not yet sequenced. The ESTs based identification also confirmed the miRNAs expression. This approach holds great promise for the future as it allows a wide range of potential targets for suppression of gene expression in the insect. Additional genetic /molecular studies will be needed to understand whether miRNAs typically regulate only a handful of key targets or co-ordinately regulate multiple targets which are equally important.

Gene expression analysis using synonymous codon usage analysis for Drosophila

Coding sequences of drosophila having Cytochrome P450 mono-oxygenase had been downloaded from NCBI site. Various Codon Usage Indices have been calculated and multivariate analysis has been done to establish differential level of expressions and the pattern of synonymous codon usage of CYP genes in drosophila. This study was helpful in understanding P450s enzyme system involved in

resistance mechanism. The main objective of this study was to apply synonymous codon usage bias to understand the expressivity level of different categories of resistant genes in model insect. This study also helped to develop new strategies to mitigate the insecticide resistance development in pest insects which was serious concern for agriculture and human health.

In-silico identification of genes responsible for late blight disease in potato

The virulent and susceptible genes were indentified and information was sent to CPRI, Shimla. Based on the findings of this study a new inter-institutional project with NBPGR and CPRI has been initiated.

Programme 5: DEVELOPMENT OF INFORMATICS IN AGRICULTURAL RESEARCH

Strengthening Statistical Computing for NARS

Strengthening Statistical Computing for NARS (www.iasri.res.in/sscnars) targets at providing

– research guidance in statistical computing and creating sound and healthy statistical computing environment and

– providing advanced, versatile, innovative and state-of-the-art high end statistical packages for analysis of data so as to enable drawing meaningful and valid inferences and converting research output into knowledge

The efforts also involve designing intelligent algorithms to implement statistical techniques particularly for analyzing massive data sets, simulation, bootstrap, etc. Capacity building, achievements, usage and impact is summarized in the sequel.

Capacity Building

– 211 researchers have been trained on Data Analysis using SAS through 11 training programmes of one week duration each. With this the number of researchers trained has gone upto 1883 through a total of 91 training programmes. Out of these 11 training programmes in 2012-13, 02 were organized by IASRI, New Delhi and rest 09 by consortium partners. 04 were organized at doorsteps of the users such as RVSKVV, Gwalior; SKRAU, Bikaner, NIRJAFT, Kolkata and VPKAS, Almora. 03 of these 11 training programmes were on specific topics such as Design of Experiments, Sample Surveys and Multivariate Analysis.

– 40 researchers were trained through two training programmes on (i) Analysis of Design of Experiments using SAS and (ii) Biometrical Analysis using SAS organized by Nodal Officer from IGKV, Raipur.

Updates, Upgrades and Installation

– Updates and upgrades were received. To sort out implementation issues and refinement in installation process, handing over of updates and upgrades (SAS EAS 9.3, JMP 10, JMP Genomics 6.1 and all products for 64 bit windows) and to have a face to face interaction with nodal officers, third Workshop-cum-installation training programmes at 09 Statistical Computing Hubs were organized.

– The Workshop-cum-Installation training at IASRI was organized during 25-26 June, 2012. Dr. S Ayyappan, Secretary DARE and Director General, ICAR inaugurated the Workshop. Bulletin on Indian NARS Statistical Computing Portal was released during the occasion.

– The software has been installed on 2095 computers (1623 reported earlier) in all 151 NARS organizations (on an average 13 machines per NARS organization). SAS Genetics successfully installed in Thin Client Environment at Statistical and Computational Genomics Laboratory, IASRI, New Delhi.

Strengthened Indian NARS Statistical Computing Portal

– Indian NARS Statistical Computing Portal (http://stat.iasri.res.in/sscnarsportal) has been strengthened by adding 13 new modules of analysis of data generated from Completely Randomized Designs, Resolvable Block Designs, Row-Column Designs,

Nested Block Designs, Split-Split-Plot Designs, Split Factorial (main A, sub B x C) Designs, Strip Plot Designs, Response Surface Designs, Univariate Distribution Fitting, Test of Significance based on t-test and Chi-square test, Discriminant Analysis, Correlation and Regression Analysis. The data can be analysed by uploading *.xls, *.xlsx, *.csv and *.txt files.

Macros for customized analysis and E-reference manuals

● For customized analysis, macros for analysis of data generated from Strip Plot Designs have been developed and made available on the project website http://www.iasri.res.in/sscnars/StripPlot.aspx.

● Following 04 reference manuals consisting of 52 lectures have been uploaded on project website:
  – Genetics/ Genomics Data Analysis Using SAS: 14 lectures
    http://www.iasri.res.in/sscnars content_Genetics.htm
  – Data Analysis in Social Sciences Research Using SAS: 19 lectures
    http://www.iasri.res.in/sscnars/content_social.htm
  – Data Mining Using SAS: 11 lectures
    http://www.iasri.res.in/sscnars/content_dm.htm
  – Data Analysis Using R: 08 lectures
    http://www.iasri.res.in/sscnars/content_rmanual.htm

Sensitization of researchers

● Website of the project is being maintained and updated regularly. During 15 November, 2010 – 31

March, 2013 Google analytics gave 21900 page views across 367 cities of 71 countries. During 01 April, 2012- 31 March, 2013, there were 11747 page views across 322 cities of 66 countries. Average time on page was 3.18 minutes.

● With the cooperation and support from CRIDA, Hyderabad and NAARM, Hyderabad, first in-house Webinar session was conducted. The participants were given the exposure of Design Resources Server and Indian NARS Statistical Computing Portal. Second in-house WebEx session was conducted on 16 February, 2013.

● WebEX session on JMP Genomics 6.0 and JMP DOE were also arranged.

● To sensitize the researchers about the availability of this high end statistical package, 336 participants were sensitized through 12 sensitization training programme-cum-workshop organized at NBAIM, Mau; Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut; ICAR RC NEHR Regional Station, Gangtok Sikkim; Assam Agricultural University, Khanpara; College of Agriculture, Bapatla, ANGRAU; TANUVAS, Chennai; ICAR RC ER, Patna; NDRI, Karnal; MPUAT, Udaipur; CIFE Mumbai and Directorate of Onion and Garlic Research, Pune.

● 313 scientists have been sensitized on Statistical Computing through FOCARS by NAARM, Hyderabad (a total of 571 scientists were sensitized).

● Several researchers were also sensitized through the sensitizations programmes conducted by Nodal Officers at NBPGR, New Delhi; NCAP, New Delhi; Junagarh Agricultural University, Junagarh and IISR, Lucknow.

● Presentations were made in 13 training programmes/ Workshops/ Conferences/ Special Sessions at different NARS organizations.

Usage and impact in NARS

The capacity building efforts have paved the way for publishing of research papers in the high impact factor journals. Researchers have started making effective use of the software.

● Based on feedback received from NARS organizations, 105 research reports (98 reported earlier), 201 research papers (100 reported earlier) have been published/ accepted for publication by analyzing the data using high end statistical computing facility; 143 students (60 reported earlier) have used

this in their dissertations; 1229 students (984 reported earlier) have used in their course work. The number of data sets analyzed is more than 3420 (1589 reported earlier) across NARS.

- Reference manual cited in Journal of Doctoral Research in Economics, The Bucharest Academy of Economic Studies. Macro for augmented designs cited by Jennifer Kling, Oregon State University in Introduction to Augmented Designs; Ir. Suprayogi from Laboratory of Plant Breeding and Biotechnology, Central Java, Indonesia and Ogbonna Alex C. from National Root Crops Research Institute, Umudike, Nigeria.
- Number of hits at Indian NARS Statistical Computing Portal since 01 April 2011: 41097.
– Nodal officer from CMFRI, Kochi reported saving of 20 man months in compilation of data related to Marine Fish Household Census 2010 consisting of 10 lakh households with 16 attributes.

Implementation of Management Information System (MIS) including Financial Management System (FMS) in ICAR

IASRI is implementing a robust and flexible MIS & FMS System which includes solution for Financial Management, Project Management, Material Management, Human Resource Management and Payroll at ICAR with the funding support from NAIP. A contract agreement was signed on 19 January 2012 between IBM and IASRI on behalf of ICAR to implement the ERP solution based on Oracle Application R12. Business Process Owners and core team members were identified at ICAR and in different institutions in different functional areas.

- Requirement study was carried out in collaboration with ICAR headquarters and partner organizations, and Requirement analysis workshop was organized at IASRI. Based on the requirement study, AS IS documents which cover the current process followed in ICAR institutions have been prepared in the functional areas of Financial Management, Project Management, Material Management, Human Resource Management and Payroll. Six AS IS documents have been prepared in different functional areas.
- Development /System Demonstration Instance of Oracle application was installed and configured for system design and development of TO BE Process

Scenarios. TO BE Design Documents were created in all functional areas based on solution mapping on Oracle Application. Six TO BE documents have been prepared.

- Technical Architecture document which included recommendation for Production and Non-Production hardware along with infrastructure and bandwidth requirements was finalized.
- Web site for MIS/FMS system was created and all the documents related to system are accessible from website.
- System Design and Technical Development (Reports, Customizations) were developed in each functional area of FMS/MIS system. System design along with reports were demonstrated with core business process owners in iterative manner in different meetings. Based on the feedback, the system was strengthened.
- Integrated solution was demonstrated to the Core Team and Business Process Owners in the workshops organised and the feedback received from Business Process Owners was addressed.
- Templates for data collection have been prepared in different functional areas for data digitization related activity. Data digitization awareness workshop was carried out at IASRI. Data digitization teams have been formulated at Phase 1a institution along with some other ICAR institutions. Sample Data (10% for UAT) is being entered in the templates by Phase 1a institutes.

Inauguration of the Requirement Analysis Workshop for ICAR ERP System

AS-IS,TO-BE and Technical Architecture documents related to system

Web site for MIS/FMS system at http://www.iasri.res.in/misfms

Establishment of National Agricultural Bioinformatics Grid in ICAR (NABG)

Genomic Database, Portal for genome sequence submission, Cattle genomic resource information system, Crop stress responsive gene Database, Micro-satellite databases of Pigeonpea, Micro-satellite databases of Buffalo have been developed. Four research papers have been published and two accepted in high impact factor journals from the outcome of this project during the period under report. Numbers of research articles are in pipeline for publications. Three research projects have been initiated on the basis of outcome of the research studies in this project and number of inter-institutional research projects including externally funded projects have been initiated. Also, during this year, scientists were working on eight different research studies.

Genomic Databases Developed

A cattle genomic resource information system (CGRIS)

The system is being updated with the identified 4000 additional SNPs, related to different diseases, growth traits and immunity. The database was updated with a new module on *Epitope Vaccines* that contains 175 predicted epitopes for diseases like FMD, BVD, IBR and Calf Scour. An interactive Jmol viewer was designed to visualize the structures of epitopes.

PIPEMicroDB: Microsatellite database and primer generation tool for pigeonpea genome

(http://cabindb.iasri.res.in/pigeonpea/)

Molecular markers play a significant role for crop improvement in desirable characteristics, such as high yield, resistance to disease and others that will benefit the crop in long term. *PIgeonPEa Microsatellite DataBase (PIPEMicroDB)* is an automated primer designing tool for pigeonpea genome, based on chromosome wise as well as location wise search of primers. This stores 123387 STRs extracted *in silico* from pigeonpea genome. This tool enables researchers to select STRs at desired interval over the chromosome. Further, one can use individual STRs of a targeted region over chromosome to narrow down location of gene of interest or linked QTL. These marker searches based on characteristics and location of STRs is expected to be beneficial for researchers/molecular breeder for varietal improvement.

BuffSatDb: Micro-satellite databases of buffalo

(http://cabindb.iasri.res.in/buffsatdb)

Though India has sequenced water buffalo genome, its draft assembly is based on cattle genome BTau 4.0. Thus de novo chromosome wise assembly is a major pending issue for global community. The existing radiation hybrid of buffalo and these reported STR can be used further in final gap plugging and "finishing" expected in de novo genome assembly. QTL and gene mapping needs mining of putative STR from buffalo genome at equal interval on each and every chromosome. Such markers have potential role in improvement of desirable characteristics, such as high milk yields, resistance to diseases, high growth rate. The STR mining from whole genome and development of user friendly database is yet to be done to reap the benefit of whole genome sequence. By in silico microsatellite mining of whole genome first STR database of water buffalo, BuffSatDb (Buffalo MicroSatellite Database) have been developed which is a web based relational database of 910529 microsatellite markers, developed using PHP and MySQL database. Microsatellite markers have been generated using MIcroSAtellite tool. The search may be customised by limiting location of STR on chromosome as well as number of markers in that range. This was a novel approach and not been implemented in any of the existing marker database. This database has been further appended with Primer3 for primer designing of the selected markers enabling researcher to select markers of choice at desired interval over the chromosome. The unique add-on of degenerate bases further helps in resolving presence of degenerate bases in current buffalo assembly.

Being first buffalo STR database in the world, this would not only pave the way in resolving current assembly problem but shall be of immense use to global community in QTL/gene mapping critically required to increase knowledge in the endeavour to increase buffalo productivity, especially for third world country where rural economy is significantly dependent on buffalo productivity. These markers can be used for parentage testing, breed identification, population structuring and admixture analysis. They can also be used for germplasm

identification especially in germplasm exchange or issues related to trans-border movement of germplasm.

Parallelized workflows for gene prediction, phylogenetic analysis and primer designing

Development of web pages for user profile generation and logging in page has been completed. Program has been developed for uploading the sequences on the server.

Development of script for automatic download of genomic sequences is in progress. However, a script has been developed for automatic downloading the sequence from the NCBI website using "ftputil" program as a standalone version. Script for degenerate bases of the uploaded sequence has been developed. It can also handle multiple sequences. Also, program has been developed for extraction of sequences from the uploaded input sequence file in which multiple sequences are present. A linked hash map has been used for this purpose. Further, development of framework for the pipelines phylogenetic analysis and SSR-primer is completed. Integration of MISA in the SSR-primer workflow has been done. It handles multiple sequences and shows the result on web. Bioinformatics tools, available for parallelized platform, required for development of pipelines are being explored and downloaded for integration. Various tools for phylogenetic analysis have been reviewed and tools for SSR markers generation and primer designing identified. Designing of home page for the workflows has been done. Web pages for user profile generation and log-in page have been developed.

Home page of the Workflow

Program has been developed for uploading the sequences on the server. Development of script for automatic download of genomic sequences is in progress. Integration of MISA in the workflow has been done. It handles multiple sequences and shows the result on web. Bioinformatics tools, available for parallelized platform, required for development of pipelines under the project are being explored and downloaded for integration.

Phylogenetic Analysis Workflow diagram

SSR-Primer Design Workflow diagram

Web based software for codon usage analysis for gene expression identification

This study aims at development of a comprehensive web solution for synonymous codon usage analysis for gene expression identification using client-server architecture. Review of literature has been done. Software has been developed using JSP, NetBean, HTML and JavaScript. Various modules have been developed for the generation of base indices, GC3 Contents, A3, T3, G3 and C3 contents, Codon Bias Index, Codon Adaption Index, Frequency of Optimal Codons (Fop) of codon usage bias data reduction technique like correspondence analysis in codon usage analysis with respect to software development. Web based calculation of codon usage indices would help researchers to calculate these indices using any standard browser.

Project Information & Management System of ICAR (PIMS-ICAR)

Project Information & Management System of ICAR (PIMS-ICAR) designed, developed and implemented at IASRI, New Delhi at http://pimsicar.iasri.res.in/ to help in taking decisions to check duplication in research projects both at divisional as well as inter divisional level of ICAR. PIMS-ICAR has also been integrated with Half Yearly Progress Monitoring of Scientists (HYPM) system developed and implemented for all the ICAR institutes. The integration has facilitated the visibility of Research Projects details of ongoing projects with respective PIs and Co-PIs in HYPM. As per the data entry status available in PIMS-ICAR, the ICAR institutes have initiated project data entry process for more than 5550 ongoing and 5656 completed projects into PIMS-ICAR from their respective institutes. The RPF-III of 4024 projects has already been uploaded by institutes and is available in PIMS-ICAR.

Besides, PIMS-ICAR has been included in the curriculum of the training programmes like FOCARS, MDP, EDP and Refreshers courses organized by NAARM, Hyderabad. For hands-on exercise by the trainees, the training demo version of PIMS-ICAR software has been installed on the LAN server of NAARM, Hyderabad.

National Information System on Agricultural Education Network (NISAGENET)

The NISAGENET web portal is being maintained at the Central Server of IASRI, New Delhi and is accessible at http://www.iasri.res.in/Nisagenet/. The system is operational in all 65 Universities/Organizations' involved in imparting higher agricultural education in the country. The database of this system contains the information on various aspects such as Academic data of the universities, Infrastructural facilities, Budget provision, Manpower employed, Faculty and R&D activities. Moreover, it has an exhaustive Query/Reports system to provide information at Country, State, University and College levels. To maintain the NISAGENET system and to initiate data management activities from all agricultural universities, Regular contact and Technical Support is being maintained with the Nodal Officers for collection and entry of updated data. The data with regard to faculty status during the year 2010-11 and 2011-12 of the constituent/ affiliated colleges is being uploaded by the respective AUs/Colleges. To expedite data management activities from AUs, 3 Appraisal cum Data Validation Workshop, for the Nodal Officers of NISAGENET were organized at SKUAST Jammu, TANUVAS Chennai and Banaras Hindu University, Varanasi respectively. A Workshop on NISAGENET for Associate Nodal Officers of ANGRAU was also organized by ANGRAU at College of Agriculture, Bapatla. As per requirement of the Education Division, ICAR report module has been strengthened to generate the following additional reports:

- Discipline wise reports on faculty members at universities/colleges in the country.
- Reports on experimental farm area at universities/colleges in the country.
- Reports on discipline wise intake capacity, enrolment and passed out students at Masters and Doctoral level in Agricultural Statistics, Statistics, Biostatistics, Bioinformatics and Computer Applications in NARS.
- Reports on Diploma and Certificate courses offered at universities/colleges in the country.
- Reports on State wise distribution of universities/colleges in the country.

- Reports on the universities/colleges that will qualify to celebrate their Silver/Golden/Platinum Jubilee in the 12th Five Year plan.

Half-Yearly Progress Monitoring System of the Scientists in ICAR (HYPM)

For Half-Yearly Progress Monitoring (HYPM) of the Scientists in ICAR http://hypm.iasri.res.in was implemented from 01 April 2012 for online submitting the proposed targets by the scientists for the first half year period (01.04.2012 to 30.09.2012). The system facilitate to enter proposed targets for the coming half year and achievements of the completed half-year with respect to Research, Teaching, Training, Extension and Other Prioritized Activities independently.

The Reporting Officer has access to the Proposed Targets and Achievements details submitted by all concerned scientists to add his/her remarks and give recommendations on the basis of the progress reports/ inputs submitted by the concerned scientists. Reviewing Officer has dual facilities as he/she may be the Reporting Officer for some scientists like Head of Divisions and Reviewing Officer for other scientists. The Reviewing Officers are able to add their own assessment remarks and final overall grading on the Proposed Targets and Achievements of all scientists. For monitoring progress of the scientists at DG/SMD/ICAR level, various reports are generated for the proposed targets status as submitted by the scientists and comments of the Reporting/ Reviewing Officers. HYPM has been included in the curriculum of the training programmes like FOCARS, MDP, EDP and Refreshers courses organized by NAARM, Hyderabad. For hands-on exercise by the trainees, the training demo version of HYPM software has been installed on the LAN server of NAARM, Hyderabad. The access of HYPM (http://hypm.iasri.res.in) is made available to all Institutes/ Bureaus/ Directorates/NRCs for on line submission of the achievements of completed half yearly period – I (01-04-2012 to 30-09-2012) and simultaneously proposed targets of half year period – II (01-10-2012 to 31-03-2013). Provided help and guidance to the Nodal officers for customization & implementation of HYPM from their respective institutes.

Exploration of Central Data Warehouse (CDW) for Knowledge Discovery

The prototype for developing OLAP cubes in SAS OLAP Studio with the help of SAS Enterprise Guide using NSSO data (61st Round) has been developed. Data preprocessing and data preparation for carrying out classification and association task was done. Integration of different data tables to include variables to prepare data for carrying out classification and association rule mining task was done. Data from 3 districts (Faridabad, Sonipat and Rohtak) of Haryana was classified. Association rule mining has been carried out on one district (Faridabad) of Haryana.
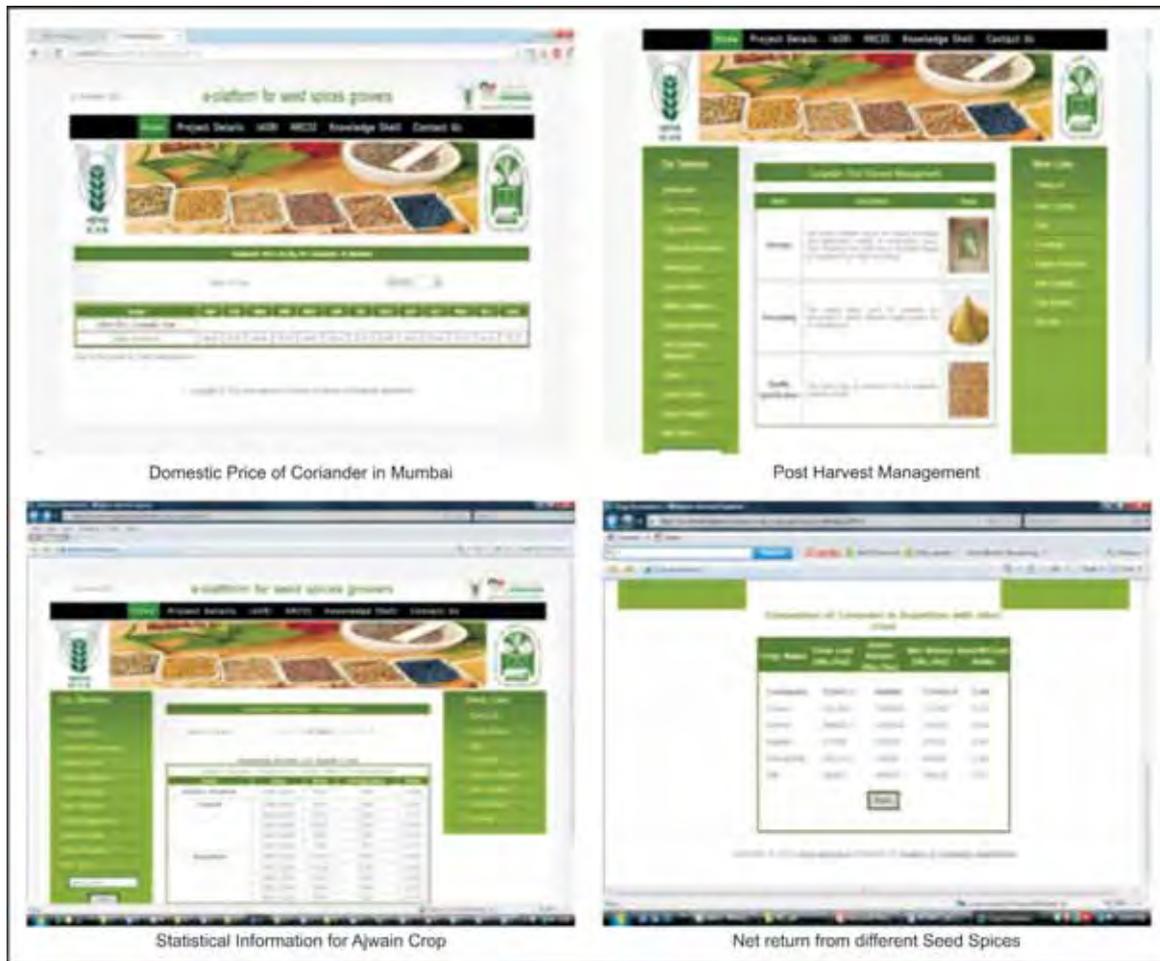
ePlatform for seed spice growers

This system is envisaged to provide guidance and information on economic issues related to seed spices. It will also provide the cost benefit ratio of each seed spice based on the area, agro climatic conditions and the factors relevant to crop production. It will project the spice that brings better return to a farmer and henceforth will help in identifying the right crop to be produced by him.

The data submission status of the achievements of Half Yearly Period (01-04-2012 to 30-09-2012) by the Scientists, Reporting and Reviewing officers

No. of Institutes Registered with HYPM (Password Issued to PME Cell I/Cs)	Scientist Registered with HYPM	Scientist Submitted Target	Scientist Submitted Achievement	Reporting Officer Commented	Reviewing Officer Reviewed
97	4487	3981	3948	3800	3602

The data submission status of the online proposed targets of Half Year Period (01-10-2012 to 31-03-2013) by the Scientists, Reporting and Reviewing officers

No. of Institutes Registered with HYPM (Password Issued to PME Cell I/Cs)	Scientist Registered with HYPM	Scientist Submitted Target	Reporting Officer Commented	Reviewing Officer Reviewed
97	4487	4127	4040	3938

Domestic Price of Coriander in Mumbai

Post Harvest Management

Statistical Information for Ajwain Crop

Net return from different Seed Spices

Modules on Market Intelligence, Crop Economics and Post Harvest Management and Statistical Information have been developed. Market Intelligence module provides price of seed spices in various Mandis. Post Harvest Management module provides information on Storage, Grading and Quality specifications. Crop Economics provides information on production economics of all seed spice crops, cost benefit ratio and suggests the most suitable crop. The statistical information provides information on area, production, yield and export potential of seed spices.

Web Enabled Statistical Package for Factorial Experiments (SPFE 2.0)

The SPFE 2.0, which is a web enabled version of SPFE 1.0 developed earlier at IASRI, gives the designs for symmetrical and asymmetrical factorial experiments and

also performs analysis of the data generated. It generates randomized layout of the designs for factorial experiments with or without confounding. The software requires user input as a list of independent interactions to be confounded.

Different interaction(s) can be assigned for confounding in different replications. It also generates regular fractional factorial plans for symmetrical factorial experiments. The data are analyzed as per procedure of blocked/unblocked designs for single factor experiments. The treatment sum of squares can be partitioned into sum of squares due to main effects and interactions. A null hypothesis on any contrast of interest can also be tested. The package is also useful for illustration purposes in the classroom teaching as well as for the researchers in Statistics with interest in experimental designs particularly in factorial experiments. The package has been developed using C# and ASP.NET using the .NET technology.

The main features of the package are (i) Generation of selected design, (ii) Randomized layout of the design, (iii) Analysis of the data and (iv) Probability calculation.

(i) Generation of the Designs

This module generates designs for the following four situations viz., (i) Complete factorial without confounding (ii) Complete factorial with confounding (iii) Fractional factorial plans, and (iv) Balanced confounded designs.

(ii) Randomization

SPFE has a built-in facility to generate randomized layout of designs for all the options. This randomization includes randomization of the Replications, Blocks within Replications and Treatment combinations within each block. The randomization is achieved by using some standard library functions of C#. These functions generate uniform random variates by taking seed as time of the system clock.

(iii) Analysis of Data Generated

The option analysis in the Menu-Bar consists of the following submenus, which are displayed as follows:
- Single Factor
- Multiple Factor
- Main Effects and Interactions
- Single Degrees of Freedom Contrasts
- User Defined Contrast

(iv) Probability Generation

It generates the probability using the following distribution
- t-distribution
- Chi-square Distribution
- F-Distribution

SPFE Help

This feature includes the SPFE web help with index, content and search facility i.e. Individual Help and Complete Help of each Module of SPFE 2.0. It gives detail theory about each module, how to use each module from SPFE 2.0.

Management system for post graduate education-II

The aim is to strengthen the software "Management System for Post Graduate Education" for management of day to day activities of the University. The software is based on Web technologies and is accessible from the desktops of the students, faculty members and administrative officials in different disciplines under the P.G. School, IARI. Following activities were undertaken for incorporating user feedback by enhancing existing functionalities of the implemented modules and provide support to the users:

- Search functionality was created in the administrator section to search for students, users, courses and thesis. The same will be provided to all the users.
- A functionality was created to remove the left-out faculties and to add a course qualifying exam and thesis evaluation (2 credits each) in PPW.
- A functionality was created and implemented to record every transaction made by the faculty in the administrator module.
- PG school calendar was linked for the year 2012-13 for all users.
- Provisional certificates have been generated for all the M.Sc. students enrolled in 2010.
- Module for Clerk, TOSC, AAO, In-charge AIM sub-modules for PPW, ORW and Result workflows have been implemented.
- Registration of old and newly admitted M.Sc. and Ph.D. students for all the three trimesters on the Management System P.G. School, IARI was done.
- Mark sheets have been prepared through the system for all M.Sc. students enrolled in 2010.

The following reports were developed:

- The report for result submitted by course leaders.
- Report of class schedules uploaded by faculty members for various courses at professor level and dean level.
- Report on Multilingual degree generation and printing.
- The student reports were enhanced by adding their photos.

Technology Assessed and Transferred

Expert System on Seed Spices

- Expert system on seed spices has been developed and implemented to advise farmers on variety selection, field preparation, fertilizer application, schedule of irrigation, plant protection from pests diseases/nematodes. The system was demonstrated in Krishi Vigyan Mela at IARI during 6-8 March 2013. A presentation of the system was made and farmers were apprised about the e-platform during a training programme organized for the farmers by NRCSS, Ajmer. A demonstration of e-platform and formal interaction was also made by the team of e-platform for Seed Spice Growers with the farmers from Jodhpur at the Krishi Vigyan Kendra, Ajmer on 31 October 2012.

Maize AGRIdaksh

- Maize AGRIdaksh is the first system developed using AGRIdaksh tool, which provides ICT based expert advice on maize crop and allows interaction with experts using internet. Farmers can login to the website and can query for different pests and diseases and for their control and prevention. They can also seek help for the varieties recommended for their region for different purposes. Demonstrated the system in Krishi Vigyan Mela during 6-8 March 2013 at IARI, New Delhi. Farmers and other visitors found the system to be very useful. The site was accessed from 16 countries by over 2200 users from April 2012 to March 2013. Mushroom AGRIdaksh is also developed and available online. AGRIdaksh has been enhanced by incorporating multilingual features.

❑