IASRI...

स्वर्ण जयन्ती

GOLDEN JUBILEE

भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE

1959-2009

... an era of excellence

GOLDEN JUBILEE

भाकृअनुप
ICAR


भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
(भा.कृ.अनु.प.)
लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली–110012

INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE
(ICAR)
LIBRARY AVENUE, PUSA, NEW DELHI- 110 012
www.iasri.res.in

*. . . towards excellence in research and education in Statistics and Informatics for enriching the quality of agricultural research*

# Messages

## M.S. SWAMINATHAN RESEARCH FOUNDATION

**M. S. Swaminathan**
*Chairman*

# MESSAGE

The Indian Agricultural Statistics Research Institute (IASRI) has been a pioneer in introducing statistical methods in agricultural research and experimentation. Although in its present form the institute is only fifty years old, its history dates back to 1930. Giants in the field of agricultural statistics like Drs P V Sukhatme and V G Panse laid the foundation for promoting agricultural research on statistically sound lines. Thanks to IASRI, India has become the world leader in this field. IASRI is an important partner of the IARI PG School. In this capacity the institute has helped to train large numbers of Ph.D. and M.Sc. scholars. Thus, IASRI is the mother institution in our country for agricultural statistics, since its alumni are now holding important positions in agricultural universities, ICAR institutes and government departments.

I wish the institute continued success in its efforts to ensure the needed statistical rigour in all our experimental work. This will help in obtaining results which are scientifically sound and practically effective. I am confident that during the next fifty years, IASRI will become the flagship of the statistics for sustainable agriculture movement.

M S Swaminathan

3rd Cross Street, Taramani Institutional Area, Chennai (Madras)-600 113, INDIA
Phone: +91-44-2254 2790, 2254 1698 Fax: +91-44-2254 1319
E-mail: chairman@mssrf.res.in, swami@mssrf.res.in

सत्यमेव जयते

डा. मंगला राय
सचिव एवं महानिदेशक

**DR. MANGALA RAI**
SECRETARY & DIRECTOR-GENERAL

भारत सरकार
कृषि अनुसंधान और शिक्षा विभाग एवं
भारतीय कृषि अनुसंधान परिषद्
कृषि मंत्रालय, कृषि भवन, नई दिल्ली 110 001

GOVERNMENT OF INDIA
DEPARTMENT OF AGRICULTURAL RESEARCH & EDUCATION
AND
INDIAN COUNCIL OF AGRICULTURAL RESEARCH
MINISTRY OF AGRICULTURE, KRISHI BHAWAN, NEW DELHI 110 001
TEL.: 23382629; FAX: 91-11-23387293; E-MAIL: mrai.icar@nic.in

# MESSAGE

It is a matter of great satisfaction for us to note that Indian Agricultural Statistics Research Institute (IASRI), which began as a small statistical section in 1930 in the then Imperial Council of Agricultural Research (ICAR), is now celebrating Golden Jubilee of its foundation on 2 July, 2009.

The Institute has made significant academic and research contribution in the fields of design of experiments, statistical genetics, forecasting techniques, sample surveys, econometrics and computer applications in agriculture and software development. The institute continues to provide specialized support to the National Agricultural Research system by way of planning and designing agricultural experiments, analyzing voluminous data using advanced, appropriate and complicated statistical techniques so as to draw valid conclusions and also interpreting the results and converting them into knowledge. It has now become an internationally recognized institution in agricultural statistics with its alumni occupying prestigious positions in India and abroad.

The world is changing fast and so are agriculture and the agricultural research. Newer technologies like remote sensing, geo-informatics, artificial neural networks, data warehouses and data mining, newer emerging areas of research like gene expressions, bioinformatics, genomics and proteomics, nanotechnology etc. pose tough challenges to the statisticians of today. Present day realities of Climate change, bio-safety, agricultural marketing and modeling of bio-systems have to be dealt with determination. Statistical theories, methodologies and the Information Communication Technologies have to converge so that the knowledge generated through agricultural research gets transformed, translated and transferred to the farming community for their progress and prosperity. These challenges require additional resources and new path ways. I am confident that the scientists of the institute will endeavour to meet challenges of today and tomorrow.

On this historic occasion I convey my heartiest congratulations to the Director, Scientists and all the present and past members of the IASRI family. Our younger colleagues have a bigger challenge of maintaining the past glory and taking the Institute to newer heights in the area of agricultural statistics and computer applications. I wish the Institute a very bright future.

**(Mangala Rai)**

# MESSAGE

I most warmly felicitate IASRI on its attainment of the age of 50 years. In association with the late Dr Frank Yates, I spent a year in New Delhi (1952-1953) as FAO consultant to ICAR with the intention that he and I should advise on the development of statistical science in the service of Indian agricultural research. This was an adventurous year for my family, because my wife and I brought with us our first child, our 6-month old infant daughter. We have been left with very happy memories of the kindness and friendship shown to us at all times by the people among whom we lived and with whom I worked. The report that Yates and I submitted played its part in ICAR's decision to establish IARS as separate institute with the consequences that we see today. Fortunately, under a succession of outstanding Directors (notably my dear friends VG Panse and Prem Narain) IASRI has grown to its present eminence in the national life of a great country. I count myself fortunate to have been invited several times to return on short consultancy engagements, so keeping in touch with developments and possibly contributing ideas from experience elsewhere.

My love for India and interest in her scientific welfare makes me wish that I could visit you again, but my advanced age will prevent this. Though I am no longer in touch with current IASRI activities nor sadly do I know any of the leading scientists employed there, I take this opportunity of suggesting one topic for an important project. As I read the Institute's history, I am struck by the recurrent theme of research into forecasting of yields and forewarning of major outbreaks of disease and pests. Is not the time ripe for a thorough survey of this aspect of agricultural statistics? How successful has IASRI's contribution been? Has its methodological advice been widely used and have quantitative predictions been close enough to the eventual truth for timely deployment of effective corrective actions or countermeasures? Make of this a major statistical study that can serve as an example for forecasters worldwide.

**David J. Finney**
**13 Oswald Court, South Oswald Road**
**Edinburgh EH9 2HY, UK**

डा. प्रणव सेन
DR. PRONAB SEN

भारत के मुख्य सांख्यिकीविद्
Chief Statistician of India

सत्यमेव जयते

राष्ट्रीय सांख्यिकीय आयोग
National Statistical Commission
भारत सरकार /Government of India
सरदार पटेल भवन, संसद मार्ग, नई दिल्ली-110001
Sardar Patel Bhavan, Sansad Marg, New Delhi-110 001
फोन / Tel.: 23742150 फैक्स / Fax: 23742067
E-mail: pronab@nic.in

# MESSAGE

The Golden Jubilee of the Indian Agricultural Statistics Research Institute (IASRI) is a landmark occasion in the history of the development of applied statistics in India. Over this long period, IASRI has not only contributed substantially to the development of agriculture statistics in the country, but has also left an indelible impression on the international statistical system. Despite its primary focus being on agricultural statistics, IASRI has contributed to the development of applied statistics methodologies across a wide range of subjects and sectors. The institute can rightly take pride in its achievements, and I am sure that it will scale greater heights in the coming years. I congratulate IASRI on completing its 50th year of existence and I look forward to its continued excellence in all areas of applied statistics.

**(Dr. Pronab Sen)**

**Indian Statistical Institute**

**Professor Bikas K Sinha**
Bayesian and Interdisciplinary
Research Unit [BIRU], ISI
AND
Acting Chairman,
National Statistical Commission
GoI

# MESSAGE

Dear Dr. Bhatia,

I am happy to learn that the Indian Agricultural Statistics Research Institute [IASRI] has been celebrating its Year-long Golden Jubilee [July 3, 2008 to July 2, 2009] in a befitting manner and that you will observe the Institute's Foundation Day on July 2, 2009 in a special function.

My first visit to IASRI dates back to December 1972 and the last visit was only a few months back ! I have closely observed with utmost interest your institute's multi-dimensional growth in activities related to all statistical aspects of agricultural research.

I have personally been closely associated with a few major academic programs in the Division of Design of Experiments and have thoroughly enjoyed interacting with the scientists located therein.

In my estimation, research, training and consultancy activities of your colleagues in various scientific divisions have attained a quality level of international standard. My interaction with some of them [in the Divisions of Design of Experiments, Sample Surveys and Biometrics] has been a matter of great personal satisfaction.

I must congratulate you and your esteemed colleagues on this occasion for carrying forward the mission and vision of this immensely important centre for statistical research in agriculture.

I extend my hearty wishes to you all on the eve of the special function.

Bikas Kumar Sinha

**(Bikas K Sinha)**

डा. एस. अय्यप्पन
उप महानिदेशक (मत्स्य)

**Dr. S. AYYAPPAN**
Deputy Director General (Fisheries)

भारतीय कृषि अनुसंधान परिषद्
कृषि अनुसंधान भवन - II
पूसा, नई दिल्ली 110 012

**INDIAN COUNCIL OF AGRICULTURAL RESEARCH**
KRISHI ANUSANDHAN BHAVAN-II
PUSA, NEW DELHI 110 012

# MESSAGE

I am happy to learn that the Indian Agricultural Statistics Research Institute, New Delhi is celebrating its Golden Jubilee year during July, 2008-July, 2009.

Over the last 50 years, the institute has ably served the Indian agriculture in general, and Agricultural Statistics in particular. Its contributions to the R&D as well as human resource development in the area are unparallel, that have been recognized internationally. The IASRI has stood tall as an academic institution, with further plans for advanced research and capacity building in the coming years. While complimenting the staff members of the Institute for their contributions, I wish all success in their endeavours.

**[S. Ayyappan]**

Ph: (O) 2584 6738 (R) 2584 3190 FAX: 00-91-11-2584 1955
e-mail: ayyappans@icar.org.in & ayyapans@yahoo.co.uk

भारतीय कृषि अनुसंधान परिषद्
कृषि अनुसंधान भवन-II, पूसा, नई दिल्ली 110 012
**Indian Council of Agricultural Research**
Krishi Anusandhan Bhavan-II, Pusa, New Delhi 110 012

**डॉ. सत्य प्रकाश तिवारी**
उप महानिदेशक (शिक्षा)
*Dr. S. Prakash Tiwari*
Deputy Director General (Edn.)

# MESSAGE

I congratulate the Indian Agricultural Statistics Research Institute on the Golden Jubilee of its foundation on 02 July 2009 for having grown in stature and established itself as a leader in research and education in Agricultural Statistics and Computer Applications.

Significant and original contributions have been made by the institute in the areas of design of experiments, sampling techniques, statistical genetics, modelling and forecasting techniques. More recently the institute has taken a lead in development of information systems, software, decision support systems and the use of technologies like Remote Sensing and GIS in agriculture. The institute has already become a champion in disseminating research, capacity building of scientists, human resource development, etc. by way of its teaching and training programmes.

I congratulate the entire family of IASRI, the past and present ones, on this occasion of the Golden Jubilee of the Institute.

**(S.P. Tiwari)**
**Deputy Director General (Education)**

*Tel (Office): 011-25841760 Fax: 011-25843932; E-mail: sptiwari@icar.org.in*

डॉ. एच. पी. सिंह
उप महानिदेशक
(बागवानी)
**Dr. H.P. SINGH**
Deputy Director General
(Horticulture)

भारतीय कृषि अनुसंधान परिषद्
कृषि अनुसंधान भवन-II
पूसा, नई दिल्ली 110 012
**INDIAN COUNCIL OF AGRICULTURAL RESEARCH**
KRISHI ANUSANDHAN BHAVAN-II
PUSA, NEW DELHI 110 012

# MESSAGE

I am extremely happy to know that IASRI is celebrating its Golden Jubilee this year. My heartiest congratulations to each and every member of the IASRI family.

IASRI, over the past 50 years, has contributed significantly in the area of design of experiments, sample surveys, statistical genetics, statistical modeling, forecasting and computer application for the benefit of agricultural research workers. The Institute is recognized world-wide in these fields. The work done by the scientists have appeared in reputed national/international journals. In addition to this, the Institute has developed a number of statistical software packages and information systems for agricultural scientist and agricultural research managers. The Institute has been playing a key role in the development and strengthening of the National Agricultural Research System and National Agricultural Statistics System. The Institute has been contributing towards Human Resource development programme in the form of M.Sc. and Ph.D. courses and has trained numerous scientists on application of statistical techniques in agricultural research.

I am happy to see that the research programme of IASRI is also changing as per the currents demands in agricultural research. The Institute has initiated work on Geo-informatics, Bioinformatics, Web resources on Design of Experiments and other important areas. I am sure that IASRI would continue to make its impact on Indian Agriculture and would continue to enjoy its national and international reputation.

At this occasion, I convey my best wishes to IASRI for future success and also wish the students, the staff and the scientists a very happy Golden Jubilee Celebration.

H.P. Singh

**(H.P. Singh)**

Ph: (Off.) 2584 2068, (Resi.) 25841189, (Mob.) 9871450730 Fax: 91-11-2584 1976 E-mail: hpsingh@icar.org.in

भारतीय कृषि अनुसंधान परिषद्
कृषि भवन, डा. राजेन्द्र प्रसाद मार्ग, नई दिल्ली–110001
**Indian Council of Agricultural Research**
Krishi Bhavan, Dr. Rajendra Prasad Road, New Delhi-110001

**डा0 कमल मल्ल बुजरबरूवा**
उप महानिदेशक (पशु विज्ञान)
**Dr. K.M. Bujarbaruah**
Deputy Director General
(Animal Sciences)

# MESSAGE

I am very happy to know that Indian Agricultural Statistics Research Institute, one of the premier research institute, under the aegis of ICAR is celebrating its Golden Jubilee year from 3rd July, 2008 to 2nd July, 2009. The Institute has been the nerve centre for statistical research in agriculture in the country and has not only been a nodal centre providing the much needed support in data analysis for the benefit of researchers and students but also a hub for human resource development. Statistics has been a powerful tool for data compilation, interpretation and data projection and with the advent of statistical models and packages data presentation has been made simpler with greater accuracy. Today when biotechnological tools are being used in all facets of biological sciences, bio-informatics has emerged as a single tool for large data management, particularly in areas of bioprospecting, molecular genetics and breeding and gene sequence analysis. Keeping the track record of IASRI in view, I am sure the institute will gear up itself to be a meaningful contributor to agricultural scenario analysis and developing fore-warning models.

I compliment the Director and the staff, both past and present, of the institute for their yeomen service to agricultural research.

I also wish the Institute all success in their future endeavor and growth.

**(K M Bujarbaruah)**

Phone : (O) 23381119, 23388991, Extn.: 200, Resi: 24674386; Fax: 0091-11-23097001, 23387293
E-mail: kmbujarbaruah@rediffmail.com, ddgas.icar@nic.in

भारतीय कृषि अनुसंधान परिषद्
कृषि अनुसंधान भवन-II, पूसा, नई दिल्ली 110 012

**INDIAN COUNCIL OF AGRICULTURAL RESEARCH**
KRISHI ANUSANDHAN BHAVAN-II, PUSA, NEW DELHI 110 012

Ph : 91-11-25848364 (O), 25843496, 25849786 (R)
Fax : 91-11-25848366
E-mail : aksingh@icar.org.in; aks_wtc@yahoo.com

**डॉ. अनिल कुमार सिंह**
उप महानिदेशक (प्रा सं प्रा)

**Dr. Anil Kumar Singh**
Deputy Director General (NRM)

# MESSAGE

On this historic occasion of the Golden Jubilee Year Celebrations of Indian Agricultural Statistics Research Institute (IASRI), I take this opportunity to congratulate each and every staff member, currently on rolls of the institute and those who have retired after giving yeoman service to the institute. Right from its very humble beginning in 1930 as a Statistical Section of the then Imperial Council of Agricultural Research to its new *avtaar* in the shape of Institute of Agricultural Research Statistics in 1959 and the current IASRI, this institute has provided exemplary leadership in the field of Agricultural Statistics both nationally and internationally. In addition to research, it has played a stellar role in developing human resources through client-oriented customized as well as regular training courses in addition to the degree programmes. Over the years, the institute has had the good fortune of having many legendary scientists associated with it starting from Shri M. Vaidyanathan, Dr. P.V. Sukhatme, Dr. V.G. Panse, Dr. G.R. Seth to name a few and many stalwarts that followed. The institute has kept up with the pace of developments of the modern times and is now being looked upon to provide solutions to the emerging challenges faced by the nation through newer techniques of data mining and analysis, design of experiments, use of ICT, modelling and simulation techniques in agriculture.

I understand that the institute is organizing a series of seminars, workshops and a host of other academic activities during the Golden Jubilee Year Celebrations. My best wishes for grand successes in all the activities and for a sparkling future of the institute.

***(A.K. Singh)***

डॉ. मदन मोहन पाण्डेय
उप महानिदेशक (अभियांत्रिकी)
**Dr. M. M. Pandey**
Deputy Director General (Engineering)

भारतीय कृषि अनुसंधान परिषद्
कृषि अनुसंधान भवन-II
पूसा, नई दिल्ली 110 012

INDIAN COUNCIL OF AGRICULTURAL RESEARCH
KRISHI ANUSANDHAN BHAVAN-II
PUSA, NEW DELHI 110 012

# MESSAGE

It gives me immense pleasure to note that Indian Agricultural Statistics Research Institute (IASRI) is celebrating Golden Jubilee of its establishment. The first thought that comes to mind is the journey that this unique institution has made so far. It has not only facilitated the country in its quest for food security, it has also created world class human resource to sustain the agricultural progress in future.

An institution is born with a purpose. The institution takes shape to fulfill the intended purpose. As the time advances, the purpose may get modified. It is then incumbent on the institution to transform itself to fulfill the modified purpose. With this analogy, IASRI has effectively fulfilled its professional obligations so far. The institute needs to now take stock of its strengths and weaknesses in the context of contemporary and future expectations so as to perceive the opportunities and threats. There are more and tougher challenges that require different and sharper skills. While conventional statistical tools may continue to be useful, we must work with our stakeholders to break new grounds. Newer opportunities lie in precision farming, environmental statistics, bio-informatics, bio-safety and global trade forecasting.

I congratulate the IASRI family on the occasion of the Golden Jubilee celebrations and hope that no stone would be left unturned in its resolve to emerge as an institute of great relevance.

**(M.M. Pandey)**

Ph: 91-11-2584 3415 Fax: 91-11-2584 2660 E-mail: mmpandey@icar.org.in

भाकृअनुप
ICAR

**Tele.: 91-11-25843277 (O)**
**E-mail: kdkokate@rediffmail.com**
**kdkokate@icar.org.in**

**Indian Council of Agricultural Research**
**(Division of Agricultural Extension)**
**Krishi Anusandhan Bhavan, Pusa, New Delhi-110012**

**Dr. K.D. Kokate**
**Deputy Director General (Agril. Extn.)**

# MESSAGE

I am immensely delighted to know that Indian Agricultural Statistics Research Institute (IASRI) New Delhi is celebrating its Golden Jubilee Year from July 03, 2008-July 02, 2009 and in this connection series of Seminars and Workshops on different thematic areas in Agricultural Statistics and Computer Applications have been organized by the Institute. It is also heartening to note that on its annual day i.e. 02 July 2009, the Institute is bringing out a publication highlighting its success stories, the impact of its research, teaching and training activities and also experiences of its alumni. The publication will be useful to researchers and teachers in bridging the gaps in the existing knowledge and meeting the challenges of newer emerging areas.

The institute has been playing a crucial role in research and education in Agricultural Statistics in the country and it is my sincere wish that IASRI continues its glorious traditions of carrying out research, teaching and training in the areas of Agricultural Statistics and Computer Applications to make these globally competitive.

**(K.D. Kokate)**

भारतीय कृषि अनुसंधान परिषद्
कृषि भवन, नई दिल्ली–110 114
**Indian Council of Agricultural Research**
Krishi Bhavan, New Delhi 110 114

**प्रो० स्वपन कुमार दत्ता**
उपमहानिदेशक (फसल विज्ञान)
**Prof. Swapan Kumar Datta**
Deputy Director General (Crop Science)

# MESSAGE

It gives me immense pleasure to know that Indian Agricultural Statistics Research Institute is celebrating its Golden Jubilee year during July 03, 2008-July 02, 2009. IASRI is a premier institute of Indian Council of Agricultural Research with glorious tradition of carrying out research, teaching and training in areas of Agricultural Statistics and Computer Applications. The contributions of the Institute in research, teaching and training have been monumental. Through innovative applications of statistical techniques in crop improvement programmes, the Institute has played a significant role in improving the quality of agricultural research and making them globally competitive.

On this historic occasion, I take this opportunity to congratulate each and every staff member, present and past, of the Institute and all its Alumni. I wish a grand success in all the present and future activities of the Institute.

**(Swapan Kumar Datta)**

Phone (Off.) 23382545, Fax 23097003 E-mail: ddgcs.icar@nic.in

# MESSAGE

It is a great pleasure for me to have the opportunity to write this message. As an employee, I had worked in the then 'ICAR Statistical Wing' from 1 January 1958 onwards, and it was during those days that there came the proposal to change the name of the institution. Eventually, on 2 July 1959, the name was changed to 'Institute of Agricultural Research Statistics'. The recognition by the Indian Government of the importance of Agricultural Statistics was certainly a milestone in the progress of India, with its large population and low crop yields per acre as was the case those days. So, we the employees were much delighted.

Even though I stayed in the institute only one and a half months after it became an Institute (because of my departure to USA to work on my Ph.D.), I have always relished my association with it. The Statistics Wing and later the Institute (and the 'Indian Society of Agricultural Statistics', a scientific academic association that has been behind its creation) has had world class leaders (like PV Sukhatme, and VG Panse) associated with it from the very beginning. In the mid-fifties, it was widely considered as one of the two main centers of Statistics in India. As it turned out, many of the students and workers from there later on flourished as leading scientists in the world.

A decade or so later, the name was changed to its current name. However, the good work continued. Over the years, numerous people from there have been considered to be world class scientists. This did not go unnoticed in India itself. For example, right now, there is one person there (Dr. VK Gupta) who is a National Professor.

The Institute has many important wings, all doing good work. There are plans to move also in new related fields such as Bio-informatics. Perhaps it would be a good idea to also explicitly add one further wing, namely one that studies in depth 'stochastic modeling of biological growth', including the growth of plants, and of diseases and pests on them. Combined with Bio-informatics, this will bring the field to its deepest levels.

I have had the privilege of visiting the Institute and interacting with the people there during this long period. The Society has also been a great inspiration, and I had the privilege of serving as its Sessional President in 1977. My very first paper, on sugarcane diseases, was published, in 1957, in its journal.

I wish the very best to happen to the institute; I would like to see it continue having a leading role among world bodies of its kind.

I am sure fifty years later some one will write an even stronger message than I have.

**J N Srivastava**
[Editor-in-Chief, Journal of Statistical Planning and Interference;
CNS Research Professor Emeritus, Colorado State University]
Insitute for the Exploration of Higher Reality

# MESSAGE

It gives me a great pleasure and personal pride in conveying my heart felt congratulations and best wishes to the Indian Agricultural Statistics Research Institute, New Delhi for its golden jubilee. The Institute has made a significant contribution to the Indian and world agriculture during the past half century and I pray for its successful mission in the future.

I have been privileged to have my association with the institute and its early form for the past sixty five years. I started as a Computer Assistant in the Office of Statistician of the Imperial Council of Agricultural Research in the summer of 1944 with Dr. P.V. Sukhatme as the Statistician. We provided statistical computations using manual computing machines, for crop cutting experiments to estimate crop yield. The application of statistical theory to practical national problems as well as performing crop cutting experiments in the field, created an interest in statistics in me. Consequently, after finishing my master's degree in Mathematics and some teaching, I went abroad to study Statistics and received a doctoral degree from Stanford University.

One could hardly imagine in 1944 that this office, hardly occupying three rooms in Connaught Place, New Delhi would flower into a world famous Indian Agricultural Statistics Research Institute. I personally feel very proud of my association with the Institute. This association has been further nurtured by my frequent visits to the Institute, some extending over months. I also have been Food and Agricultural Organization Consultant to the Institute in 1987 and 1990. In 1980, I was awarded a project by the U.S. National Science Foundation and reported on the scientific work done by the Institute.

During my tenure at The Ohio State University, Columbus, Ohio, several statisticians of the Institute have visited the campus and have given seminars. Many directors of the Institute like Dr. P.V. Sukhatme, Dr. G.R. Seth, Dr. Daroga Singh and Dr. Prem Narain have made extended visits to the Department of Statistics at The Ohio State University. Their seminars attracted many agricultural researchers. The impact of research and training conducted at the Institute on world agriculture is hard to measure but it must be considerable.

I salute the Indian Agricultural Statistics Research Institute for attaining this milestone and for doing pioneering work for the welfare of the society at large. We are confident that it will continue to serve the humanity in future by its work.

**Jagdish Rustagi**

Professor and Chairman Emeritus
Department of Statistics
The Ohio State University
Columbus, Ohio U.S.A.

भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
(भा.कृ.अ.प.) लाईब्रेरी एवेन्यू, पूसा, नई दिल्ली–110012 (भारत)
Indian Agricultural Statistics Research Institute
(ICAR) Library Avenue, New Delhi-110012 (INDIA)

डा. विनोद कुमार गुप्ता
राष्ट्रीय प्रोफेसर
Dr. V.K. Gupta
National Professor

# MESSAGE

I am immeasurably enchanted with an enormity of satisfaction to know that Indian Agricultural Statistics Research Institute (IASRI), which made a humble beginning as a statistical section of the then Imperial Council of Agricultural Research in 1930, has grown in stature and established itself as a pioneer in research and education in Agricultural Statistics and Computer Applications, both nationally and internationally. The institute has used the power of Statistics as a science, blended judiciously with Information Communication Technology, to meet the challenges of agricultural research in newer emerging areas and also to help enhance the quality of agricultural research. The journey of the institute towards wonderful excellence in its research, teaching and training programmes has now completed 50 years and the institute is celebrating Golden Jubilee of its foundation with a sense of fulfilment and satisfaction on 02 July 2009 and this historic event fills my heart with infinite pleasure. I feel proud and privileged when I say that I have been a witness of an era of this great institute for almost 39 years, first as a student in 1970 and then as a faculty from 1975. The Institute has made outstanding, innovative and original contributions in Statistics and Computer Applications by an intelligent and thoughtful combination of basic, applied and strategic research. Its contributions to the field of design of experiments, sample surveys, statistical genetics and statistical modelling have made the institute known internationally. The institute has also achieved brilliance in human resource development, capacity building of scientists and research dissemination among peers. The contributions of the institute towards development of software, information systems, expert systems, data warehouse and e-advisory to provide support to agricultural research are monumental.

The Institute would now take a step forward into another era with more challenges than ever, particularly on account of the depleting natural resources, climate change, uncertainties of economy and marketing and with newer emerging areas of research like bio-safety, biotechnology, nano-technology, genomics and proteomics, integrated farming system, etc. The challenges are stunning because of dwindling human resource. But I have full faith and confidence that the scientists of the Institute will be able to meet the challenges of today and future with success. I am sure after 50 years from now, some people would be writing messages describing the research and teaching achievements of the institute over 100 years and I am abundantly confident that there would be infinitely many success stories to describe.

I congratulate all the past and the present Directors, scientists, technical personnel and the administrative staff on the occasion of the Golden Jubilee celebrations of the Institute. The unflinching and untiring efforts of the stalwarts like Late Dr. VG Panse, Late Dr. PV Sukhatme, Late Dr. GR Seth, Late Dr. Daroga Singh and all the past and present incumbents of the Institute have brought dignity to this Institute. The institute has a bigger challenge of maintaining the past glory and taking the institute to newer heights so as to remain the champion in the area of agricultural statistics and computer applications. May the graph of success and magnificence of the institute be ever rising is my wish and my prayer too!

V. K. Gupta

**V.K. Gupta**
ICAR National Professor

Tel.: Off.: 011-25843508 Res.: 011-23857058 Mob.: 9899105601 Fax: 011-25841564
E-mail:vkgupta@iasri.res.in Web site: http://www.iasri.res.in

# MESSAGE

I am very happy to learn that the Institute is celebrating the Golden Jubilee this year. The Institute celebrated the Silver Jubilee in 1984 when I was Director. During the intervening period the Institute has made tremendous progress and has maintained, as in previous decades, the tempo of research, training and computer applications in the field of agricultural statistics. It has achieved international recognition for its high quality of research contributions. A number of statistical techniques developed by the Institute has been adopted by agricultural scientists, development workers, planners and administrators for deriving maximum information helpful in policy decisions. The Institute has developed a large number of computer application software and has provided with web based solutions to meet the requirements of agricultural research workers in the country. Its training programmes in the science of statistics and computer application has benefited a large number of students who are currently occupying senior positions in research institutions both in India and abroad.

Agricultural Statistics is regarded as an applied science in which the statistical methodology is investigated in a given field of agriculture covering crops, livestock, poultry, fisheries and forestry. This inevitably involves continued efforts for interaction with subject-matter specialists that may lead to newer problems in the theory of statistics for which necessary methodology may not be available. The Institute has therefore been undertaking researches both in the theory of statistics as well as in its practical application to agricultural problems. The Institute is credited with the development of several useful methodologies over the years from this angle. Just to mention one of them that had wide impact not only in India but also world over is the objective method for collecting yield statistics of principal food crops involving crop cutting survey techniques.

I am sure under the able leadership of the Director and with the support of scientists, staff and students, the institute is poised to make further significant contributions in times to come.

I am glad to extend my sincere greetings on this happy occasion and wish the Institute great success.

**(PREM NARAIN)**

# World Buffalo Trust

Flat No.205, Block No.F-64,C/9, Sector 40, Noida-201303, Uttar Pradesh. Telefax : 91-120-2579627
E-mail:wbtnoida@vsnl.net, pnbhat@airtelmail.in

***Prof. (Dr) P.N. Bhat***
*Chairman*


## MESSAGE

I am extremely pleased that the Indian Agricultural Statistics Research Institute (IASRI) is celebrating Golden Jubilee year of achievement completing on 2nd July 2009. This institute has been responsible for introducing statistical technologies to Agricultural Research and production systems. It has been responsible for the development of tools for Plant and Animal Breeding Research, which has put the country on world map of leadership in the field. The institute has major computer network and Agris information system which is used for planning and monitoring of research programmes.

I wish the staff and the Director of the Institute to lead this institute as an institute of global excellence in future.

**Prof. (Dr) P.N. Bhat**

Phones:011-25843635 (O), Tele Fax:011-25840851. e-mail:adghrd@icar.org.in

भारतीय कृषि अनुसंधान परिषद्

**INDIAN COUNCIL OF AGRICULTURAL RESEARCH**

कृषि अनुसंधान भवन- II, पूसा, नई दिल्ली 110 012

*Krishi Anusandhan Bhavan - II, Pusa, New Delhi – 110012*

***Dr. S.D. SHARMA***
***Assistant Director General (HRD)***
***& Controller of Examination(Education Division)***

# MESSAGE

Dear Dr. Bhatia,

I am happy to note that having completed 50 years of its existence as a unique premier institution in agricultural statistics, the Indian Agricultural Statistics Research Institute, is celebrating the Golden Jubilee Year during July 3, 2008 to July 2, 2009. I am aware that during the intervening period the Institute has achieved several notable milestones with significant impact on the national/international scenario through addressing emerging challenges in general and the research contributions to agricultural statistics in particular. It is really heartening to note that you would be bringing out a publication highlighting the success stories, the impact of its research, teaching and training activities along with the experiences of its alumni to serve as documentation for the posterity.

I feel great pride in having been associated with this great institution, as a student in my formative years and also, later as a Director, for giving directions for the growth of the Institute to the best of my abilities and understanding. I also take special pride in closely working with the colleagues at IASRI during all these years. I am sure the Institute has many more miles to go and shall continue to create an indelible impact on the national and international scenes. Please convey my blessings to all the colleagues for very successful growth.

With best of wishes,

Sincerely yours,

**(S.D. Sharma)**

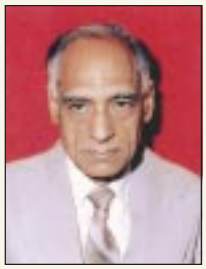
**Dr. S.K. Raheja**
**M.A., D.A.S,. Ph.D.**

Ex-Director
Indian Agricultural Statistics Research Institute
(ICAR), New Delhi
Consultant :
World Bank, 70 Lodi Estate, New Delhi

B-21, NARAIN VIHAR
NEW DELHI-110028

Phones: 25779637
9899079089
E-mail: raheja.sk@gmail.com

# MESSAGE



It is with a sense of great pride and fulfillment that I look upon the completion of 50 years of setting up of the Institute on July 2, 1959. The Institute has done pioneering and path breaking research in the field of Statistics as applied to agriculture and allied sciences right from its inception. This was in a large measure due to the strong foundation the Institute had in the form of a Statistical Branch and Statistical Wing which worked under the able guidance of internationally renowned statisticians like Dr. P.V. Sukhatme and Dr. V.G. Panse. To name a few of the outstanding contributions of the Institute to the development of statistical methods, the Technique of Crop Cutting Experiments for Estimation of Crop Yield and Production which is now preferred method of crop yield estimation in a number of countries in Asia, Africa and Latin America, Methodology for study of Impact of Intensive Agricultural District Programme followed by High Yielding Varieties Programme after the advent of Green Revolution in the late 60's, Intensive Livestock Development Programme after the advent of White Revolution in the 70's, development of new designs for Agricultural and Animal Experimentation to minimize and conserve the use of scarce resource material, development of improved Models for Estimation of Crop Yield Forecasts and development of Computer Software to provide a fast and reliable analytical framework are but a few examples of achievements of the Institute in the last 5 decades. The institute faced a number of challenges in this period which only served to give a further impetus to the grit and resolve of the scientists to treat them as small impediments towards achieving greater heights.

The Institute has a strong broad based training programme to cater to the needs not only of professional statisticians but also scientists in other fields of agriculture and allied fields. The institute is alma mater to a number of scientists occupying senior positions in the global scientist fraternity. Today, the Institute serves as a beacon light to all the budding scientists and research workers who crave for an opportunity to participate and work in the multi faceted activities of the Institute in the field of both basic and applied research and the training programme.

My Best Wishes to the Institute for emerging as Centre of Highest Learning in Statistical Methods recognized at the national and international level not only in agriculture and allied fields but also in all other areas of biological sciences.

**S.K. Raheja**

# MESSAGE

I am very happy to learn that the Indian Agricultural Statistics Research Institute (IASRI) is celebrating its Golden Jubilee this year. The Institute, formerly known as Institute of Agricultural Research Statistics (IARS), evolved from a small wing of the Indian Council of Agricultural Research (ICAR) to a full fledged research and training centre. Initially, IARS was involved mostly in project work and in the training of officers of state agricultural departments. Two advanced level courses were also run to cater to the need of trained professional statisticians. During the initial period, a few scientists were engaged in fundamental research of good quality and, this kind of activity increased the visibility of the Institute. In 1964, the Institute started M.Sc. and Ph.D. programmes in agricultural statistics and, I happened to be one of the 9 students who were admitted to the M.Sc. programme. My association with the Institute has thus been very long, five years (1964-1969) as a student and more than 18 years (1970-1989) as a faculty. I left IASRI in mid-1989 to join the Indian Statistical Institute but continue to have close association with the Institute.

Over the years that I have been associated with the Institute, I have seen it grow as a major centre of research in India. The Institute scientists have contributed significantly in several areas of statistics, notably in Design of experiments and Sample surveys. It is imperative that the younger generation of scientists continue to sustain the good work done in earlier years, so that the future is at least as glorious as the past. I convey my best wishes to all concerned in achieving this goal.

**Aloke Dey**
INSA Senior Scientist
Indian Statistical Institute

# MESSAGE

It is heartening to note that IASRI is completing fifty years of its existence. It grew from a mere Cell to a full fledged Institution during these years and has earned good reputation world over as a premier Institution in Agricultural Statistics.

I have my association with IASRI first as a Student of PSCC and Diploma during 1963-1964, then as a Research Apprentice from August 1964 to December, 1965, and finally as a faculty member in various capacities from August 1972 to April, 1987.

It was in April, 1987 that I took voluntary retirement from ICAR, and moved to Lucknow. However, I never lost my contacts with IASRI, the latest interaction being working as Chairman QRT, IASRI and a Member of Joint RAC, for IASRI and NCAP.

I must confess that I have special feelings for IASRI. This is the Institution, which has made me a Statistician in true spirits. Not only my association with IASRI as a student was most rewarding, I became truly competitive during my second phase at IASRI as a faculty. It was this later tenure of about fifteen years which made me appreciate both the strengths and weaknesses of the Institute. All these years, some of us always strived to raise the level in all the spheres – teaching, research, and applications.

It is indeed a great pleasure to get associated on this great occasion of celebrations. I very heartily wish that the Institute attains new heights and is looked upon by others as a savior in solving challenging problems of estimation and analysis.

**(A. K. Nigam)**
Executive President
Institute of Applied Statistics and
Development Studies, Lucknow

**Dr. V. K. Sharma**
Emeritus Scientist, ICAR
Former Professor (Ag. Statistics)
PG School, IARI & Head
Division of Design of Experiments, IASRI

# MESSAGE

I am indeed very happy to learn that the Institute is bringing out a souvenir as a part of its Golden Jubilee Celebration. I am fortunate to be one of the alumni of the Institute. Historically, the Institute made its beginning way back in 1930 as a small Statistical Section in the then Imperial Council of Agricultural Research that grew over time and was named as the Institute of Agricultural Research Statistics (ICAR) in 1959. The Institute earned world-wide recognition for its contribution to agricultural research and evolving techniques for development of agricultural statistics needed for planning. In addition to in-service training programmes at national and international levels, the Institute has been contributing towards Human Resource Development programme by way of conducting M.Sc. and Ph.D. programme in Agricultural Statistics since 1964-65 and M.Sc. in Computer Application since 1985-86 in collaboration with Indian Agricultural Research Institute. Many foreign nationals have also been benefited from these programmes. The Institute got its present name, in 1978 and has been recognized as the Centre of Advanced Studies in Agricultural Statistics and Computer Application. It is a matter of great pride that the alumni of the Institute have been occupying prestigious positions in reputed organizations in India and abroad. The Institute is one of its kind in the whole world and is known for its many important and original contributions in Design of Experiments, Sample Survey, Statistical Genetics and Modeling of biological and economic phenomena. The Institute has developed a number of statistical software packages and information systems for the benefit of agricultural scientists and agricultural research managers. The Institute has also been playing a key role in the development and strengthening of the National Agricultural Statistics System. It is repository of information on agricultural research and allied statistics. Besides, the Institute is engaged in developing a data warehouse on national agricultural resources.

The credit of all these achievements go to the dynamic leadership and foresightedness of the present and past Directors and well wishers of the Institute and also to the Heads of Schemes/ Divisions, scientists, technical, administrative and other staff who worked whole heartedly and coherently. At this occasion I send my best wishes and wish new higher heights for the Institute in future under the leadership of the present Director, Dr. V.K. Bhatia.

**(V.K. Sharma)**

# MESSAGE

I am delighted that IASRI is celebrating 50 years of very successful research, training and outreach programs. Congratulations!

IASRI has played a pivotal role in Indian agricultural research by developing important designs for conducting experiments and surveys and developing methods for analyzing the resulting data. Over the past 50 years, statisticians at the institute made fundamental contributions to *Design of Experiments*, *Sample Surveys*, *Statistical Genetics* and *Statistical Modeling* and are recognized world-wide as leaders in these fields. In addition to publishing important papers, they have written several important books and developed user friendly softwares. The institute has trained numerous students who went on to enjoy very successful research careers.

I am delighted to see that the research program at IASRI is keeping up with the changing times and technologies. I understand that IASRI is in the process of establishing a center for genomics. Such a center would be a great resource for scientists at IARI and other institutions generating high dimensional genomic data. Through such a center, IASRI would continue to make its impact on Indian agriculture and hence continue to enjoy its national and international reputation for fundamental contributions to statistical science and agricultural research.

I wish the students, the staff and the scientists of IASRI a very happy golden jubilee celebration.

**Shyamal D. Peddada**
Senior Investigator
National Institute of
Environmental Health Sciences (NIH)
Research Triangle Park, NC, USA.

# MESSAGE

I joined the Institute of Agricultural Research Statistics (IARS) in 1970 after earning a B.Sc. degree from the University of Delhi. Joining the Institute was an interesting experience – after one of the professors at the Institute said, "It's not a good decision to join IARS rather than pursuing an M.Sc. degree in Statistics at the University of Delhi." Well, contrary to a not so welcome reception received, I learned a lot about applications of statistics and earned the M.Sc. degree in Agricultural Statistics. I also earned a research diploma in experimental design under the guidance of Professor Aloke Dey and it gave me a foundation to conduct research which has helped me throughout my career. Thereafter I was fortunate to go to Iowa State to earn my doctorate degree in Statistics. Just as I was fortunate to work under Professor Dey at IARS, I had the privilege of working under the guidance of the late Professors B.V. Sukhatme and Oscar Kempthorne as well as Professor H.A. David at Iowa State University. I have visited my alma mater (IARS) regularly for the last three decades and have enjoyed meeting several directors of the institute, some of whom were my professors when I was pursuing my master's degree and others who were my seniors. I am very happy to know that at the present time, my classmate is the Director of the Institute and I am looking forward to my next visit to New Delhi. Earlier training and experience at IARS has been very valuable for my career path as well as my success. In 2004, I was honored to be selected a Fellow of the American Statistical Association, have served as Chair of the Department of Statistics here at The University of Akron since 1998 and recently also became interim Dean of the College of Arts and Sciences.

I congratulate the staff of the institute for successfully completing 50 years of the foundation of the institute. I expect the institute to take greater strides in the coming years.

**Chand Midha**
Acting Dean, College of Arts & Sciences
Associate Provost
The University of Akron
Akron, OH 44325-1913

# MESSAGE

I joined IASRI, the former IARS, in September 1974 session as a Ph. D. student registered at IARI with major in Design of Experiment. Having graduated from Banaras Hindu University and with a keen interest in pursuing research in Design of Experiments, I kept my mind and soul geared to seek a study opportunity at IASRI. I was fortunate to get Professor Aloke Dey as my supervisor in the area of my interest. Due to excellent research and teaching resources including dedicated faculty members, IASRI was seen as the premier institution in the country to develop skills and knowledge on applied statistics to support agro-biological and environmental research. IASRI was seen as an international seat of learning as many of my contemporary students came from other countries.

One feature I noticed was the periodical review of the institute's activities and achievements to keep its mission and mandate consistently evolving in tune with the changing needs of the society and environment it serves. Due changes have taken place in the infra-structure, the organizational structure and the personnel over the past 35 years since my student life at IASRI. I do wish IASRI to keep the dynamic nature of its planning, monitoring and evaluations of its programmes on. The technological advancement in ICT, biotechnology and exponential growth in the volume of databases on almost every aspect of human activities including exploitation of plant genetic resources, temporal-spatial information on environmental parameters, and other natural resources including water and land, are posing challenge to the statisticians on designing ways to collect reliable data, develop methods to summarise them, draw inferences on the processes contributing to the data and explore ways for generalization and implications for future. I look forward towards IASRI to take lead in suitable statistical theory and methods including the Bayesian approaches as well in building the design and decision support system in the service of society and the environment.

**Murari Singh**
Department of Mathematics and Statistics
Concordia University
1455 De Maisonneuve Blvd. West
Montreal, QC, H3G1M8
Canada

# CONTENTS

# From Director's Desk...

ALTHOUGH it has happened by destiny and not by design, yet it gives me unbounded pleasure to be the Director of this prestigious institute at a historic moment when it is celebrating the Golden Jubilee of its foundation. The institute has infinitely many contributions and success stories spread over a period of almost 79 years when it started as a small statistical section in 1930. The name Institute of Agricultural Research Statistics (IARS) was given to it on 02 July 1959, though it was given the status of a full fledged institute in the ICAR in April 1970. Since 1959, 02 July every year has been celebrated as foundation day of the institute. The year 2008 – 2009 has been the year of celebrations for the institute, enjoying the glory of its past as also doing an introspection to produce a roadmap to navigate towards brilliance while facing and meeting successfully new challenges of research while marching towards platinum jubilee. On 02 July 2009, the institute would become 50 years old and it is time to take a stock of its march towards excellence through this publication "*IASRI . . . an era of Excellence*". I would like to express my deepest sense of admiration towards the founder Directors, all the past Directors, Joint Directors, to all the past and the present Heads of Divisions, Scientists, Technical personnel, Administrative, Supporting and other staff members who have built this institute and brought it to its present form by way of sheer vision, wisdom, dedication, hard work, support, help and team work. It would be my pleasant duty to express my heartfelt gratitude towards the galaxy of eminent scientists who have served on various committees like Institute Research Council, Institute Management Committee, Research Advisory Committee, Quinquennial Review Team, and other committees for their guidance in deciding the research agenda, their critical evaluation of the research programmes to make it internationally competitive and acceptable, fresh and fragrant ideas, support and help, during this journey of 50 years as a consequence of which the research achievements have become radiant and vibrant.

The schema for the celebrations of the Golden Jubilee year was drawn by a team of young scientists led by Dr. Rajender Parsad. This team of youth was augmented with an experienced group comprising of Heads of Divisions and led by Dr. V.K. Gupta so as to give beautiful shape to the diagram drawn by this young group. Based on the suggestions, a calendar of events was drawn punctuated with several events, major ones being (a) the organization of Seminars by guest faculty made up of distinguished Statisticians and Agricultural Scientists from India and abroad, and (b) conducting one-day workshops on specialized topics of current interest.

The seminars were delivered by internationally renowned scientists like Dr. JN Srivastava, Colorado State University, USA; Dr. Bikas K. Sinha, Member, National Statistical Commission, Government of India; Dr. Ravindra Khattree, Professor of Statistics, Oakland University, Rochester, Michigan, USA; Dr. Aloke Dey, INSA Senior Scientist, ISI New Delhi; Dr. Pitam Chandra, Assistant Director General, ICAR, New Delhi; Dr. Prem Narain, Former Director, IASRI, New Delhi; Dr Murari Singh, Department of Mathematics and Statistics, Concordia University, Canada; Dr. V.K. Sharma, Former HD(DE) and Professor, IASRI, New Delhi.

To disseminate the research work carried out by the institute and to identify newer statistical and informatics issues that need attention, five workshops were conducted. The themes of these workshops were (a) Design of Experiments in Agricultural Research; (b) Applications of Small Area Estimation Techniques; (c) Expert Systems in Agriculture; (d) Remote Sensing and GIS Techniques for Decision Support in Agriculture; and (v) Statistical and Computational Issues in Genomics. Scientists from different parts of the country participated in these workshops. These workshops received high appreciation. I take this opportunity to thank profusely the guest faculty, all the participants, the conveners/co-conveners and others who helped in conducting successfully these workshops.

An important breakthrough in the capacity building and sensitization of the scientists of NARS was the initiation of a new concept of organizing "*Travel Training Programmes*" in different areas of statistics and computer applications. Two such travel training programmes on Advances in Design of Experiments were organized for the scientists of Acharya NG Ranga Agricultural University, Rajendra Nagar, Hyderabad and its centers. These were organized at Guntur and Hyderabad. These are now becoming popular and requests from other research Institutes have also started pouring in.

Another important item in the framework of the Golden Jubilee celebrations was to bring out a publication entitled '*IASRI . . . an Era of Excellence*' to be released on the foundation day of the institute. This publication consists of messages from the leaders in the ICAR and all the well wishers, the reminiscences of alumni and the brilliance achieved during these 50 years in the research and teaching programmes of the institute. Bringing out such a publication is an arduous task but I am immensely pleased to mention that Dr. V.K. Gupta, ICAR National Professor, gladly accepted to shoulder this demanding task as Chairman, Publication Committee. Not only did he conceive the ideas, the contents and the format of this publication, but he also took personal interest in giving it the final shape. He got full support from all the members of the editorial committee, viz., Dr. P.K. Malhotra, Dr. Rajender Parsad and Dr. Seema Jaggi who together toiled with Dr. Gupta in realizing the dream of bringing out this publication in time. The whole team deserves laurels for their dedication and untiring efforts in shaping this publication.

With a deep sense of gratitude I would like to express my gratefulness to all those who have sent their good wishes and blessings in the form of message for the accelerated growth of the institute. I would also like to place on record my appreciation towards all the contributors who have helped in making this publication a possibility by way of submitting their articles in time. I would also wish to express my sincere thanks to one and all who have helped directly or indirectly in bringing out this publication.

On this occasion, I congratulate the whole family of IASRI and wish them best and only best in their future research endeavors. Let us all march together in pursuit of higher goals, elevated targets, supremacy in research, . . . to help agricultural research grow exponentially so as bring solace to the farming community.

VK Bhatia

**(VK Bhatia)**

# Indian Agricultural Statistics Research Institute: A Profile

**V.K. Gupta, V.K. Bhatia and Rajender Parsad**

## 1. Historical Development

IASRI has been and continues to be a premier Institute of the ICAR with glorious tradition of carrying out research, teaching and training in the areas of *Agricultural Statistics* and *Computer Application*. Recognizing the importance of research and education in Agricultural Statistics way back in 1930, the then Imperial Council of Agricultural Research established a small *Statistical Section* to assist the State Departments of Agriculture and Animal Husbandry in planning and designing their experiments, analysing the experimental data, interpreting the results, rendering advice on the formulation of the technical programmes and examining the progress reports of the schemes funded by the Council. The statistical section was headed by a statistician and Late Shri M. Vaidyanathan was the first statistician of the statistical section.

The research activities of the statistical section took a big leap with the appointment of Late Dr. PV Sukhatme as statistician in 1940. Basic research in the theory of statistics was undertaken to bridge the gaps in the available methodologies for solving practical problems in agricultural research. With an expansion in its scope and research activities being undertaken, the statistical section was re-organised as the *Statistical Branch* in 1945 with associated expansion in its strength. The statistical branch was headed by a Statistical Advisor. The advisory work relating to agriculture and animal husbandry was separated into two units, each under the charge of a statistician.

A new dimension was added to the activities of the statistical section towards the end of 1943 when the enquiry committee set up by the Government of India to enquire into the causes of the Bengal famine came to the conclusion that one of the main factors responsible for the famine was the inappropriate statistics of crop production available at that time. As desired by Government of India and the ICAR, the statistical section undertook research in the methods of collection of yield statistics of crops by developing techniques for yield estimation based on the methods of random sampling. In the course of this work the statistical section also undertook research in sampling theory and imparted practical training to the field staff in different States. The work involved in scrutiny, compilation and analysis of the voluminous data collected in the crop cutting surveys was immense and the activities of the statistical branch were ever mounting.

An important development that took place in 1945 was the initiation of two regular certificate courses, one course of six months duration, called Junior Certificate Course (JCC) and the other course of one year duration called Senior Certificate Course (SCC). Besides, there was another course of one year duration that was introduced to train professional statisticians. Subsequently, a Diploma course involving a research project of one year duration, in addition to the Professional Statisticians' Certificate Course (PSCC) consisting of one year course work in advanced statistics, was also introduced. This necessitated the augmentation of the staff and accordingly, the strength of the branch was increased to two Professors of Statistics and other staff. These certificate courses helped in strengthening the linkages of the institute

with the state departments of agriculture and animal husbandry. The certificate courses started in 1945 were discontinued by the Indian Council of Agricultural Research (ICAR) in 1985-86. However, during 1997, the Senior Certificate Course in 'Agricultural Statistics and Computing' was revived. This course is now of six months duration and lays more emphasis on statistical computing using statistical software. The course was divided into two modules viz. (i) Statistical Methods and Official Agricultural Statistics, and (ii) Use of Computers in Agricultural Research, of three months duration each.

The period 1945-49 was one of intensive activities in the statistical branch when under its technical guidance, yield estimation surveys were carried out in almost all the States covering Wheat and Rice crops. By 1949, the results of sample surveys were also being used for official forecasts. Gradually these surveys were extended to other crops such as Jowar, Bajra, Maize and Cotton.

The statistical branch was renamed as *Statistical Wing* in 1949. In 1951, Late Dr. PV Sukhatme moved to FAO as Chief, Statistics Branch, Rome. Late Dr. VG Panse took over as Statistical Advisor of the Council in 1951. The Statistical Wing soon acquired the international recognition as a Centre of Research and Training in the field of Agricultural Statistics. Subsequently, in recognition of the important role as a training and research institution and based on the recommendations of two FAO experts Dr. Frank Yates and Dr. DJ Finney, the Statistical Wing was re-designated as the *Institute of Agricultural Research Statistics* (IARS) on 02 July 1959. In 1966 on retirement of Late Dr. VG Panse, Late Dr. GR Seth took over as statistical advisor. In April 1970, the IARS was given the status of a full fledged institute of the ICAR headed by a Director with Late Dr. GR Seth as the first Director. On 01 January 1978 the name of the institute was changed to Indian Agricultural Statistics Research Institute (IASRI) emphasizing the role of Agricultural Statistics as a full fledged discipline by itself.

In 1955, the statistical wing moved to its present campus. The building provided adequate space for library, reading rooms, lecture halls, auditorium and office for its technical and ministerial staff. A hostel with modern amenities was constructed for the trainees. Collaboration with the All India Co-ordinated Agronomic Experiments Research Projects of the ICAR started in 1956. This project consisted of two components viz., (a) complex experiments conducted at research centres, and (b) simple experiments conducted on cultivators' fields. A mechanical data processing unit was added to the institute in 1957.

The year 1964 witnessed tremendous changes in the activities of the Institute. An MOU was signed with IARI, New Delhi to start new degree courses leading to M.Sc. and Ph.D. in Agricultural Statistics. In 1981, a two years Diploma Course in Advanced Computer Programming was introduced. On the recommendations of UNDP, this course was soon discontinued and in 1985 another new course leading to an M.Sc. degree in Computer Applications in Agriculture was initiated in collaboration with IARI, New Delhi. This course was redesignated as M.Sc. degree in Computer Application during 1993-94. The institute has so far produced 173 Ph.D. and 287 M.Sc. students in Agricultural Statistics and 81 M.Sc. students in Computer Application. The alumni of the Institute are at present occupying high positions in Universities and other academic research institutions of USA, Canada and other countries.

During 1964, yet another mile stone was the installation of an IBM 1620 Model II electronic computer. In 1977, the IBM 1620 computer was replaced with a third generation computer Burroughs B-4700. The Institute has always kept pace with the latest state-of-the-art-technologies. In 1991 the Burroughs B-4700 was replaced by a Super Mini COSMOS LAN SERVER. The LAN and Intranet were further strengthened with Fibre optics and UTP cabling and at present LAN is supporting all the three buildings

of the Institute and almost every scientist has a PC on the work table.

With the advent of LAN Server and PCs, Divisional computing Labs were created. A Remote Sensing and GIS Lab with latest software facilities was established. A statistics Lab has been created very recently in which almost all the software are available on the PCs. Recently, with the broadening of the horizon of research activities in newer emerging areas, an Agricultural Bioinformatics Laboratory (ABL) has also been established at the Institute.

In 1970, a Staff Research Council was constituted at the Institute to carefully examine and finalize the research programmes of the Institute and to monitor the progress of the projects running at the Institute. The name of the Staff Research Council has recently been changed to Institute Research Council.

In October 1975, a Management Committee of the Institute was constituted in pursuance of the decision of the ICAR. The main purpose of setting up of the Management Committee was to ensure that the working of the Institute is smooth and the research programmes are being run without any problem. The administrative and the financial aspects of the institute are also governed by the Management Committee after approval from the ICAR. The research programmes of the Institute are also broadly guided by the Management Committee.

The functioning of the Institute as a Centre of Advanced Studies in Agricultural Statistics and Computer Application during October 1983 to March 1992 under the aegis of United Nations Development Programme was another landmark in the history of the Institute. The purpose of this programme was to develop the Institute as a centre of excellence with adequate infrastructure and facilities to undertake advanced training programmes and to carry out research in various emerging areas of Agricultural Statistics and Computer Application. Under this programme, a number of illustrious statisticians and computer scientists from abroad visited the institute with a view to interacting with the scientists, giving seminars / lectures and suggesting gaps in the research programmes of the institute and helping initiate research to bridge the gaps. Under the programme some scientists of the Institute received training for capacity building from abroad.

Another singular development in the growth of the Institute was the Centre of Advanced Studies programme in Agricultural Statistics and Computer Application established during the VIII five year plan in 1995. Under this program the institute organizes training programmes on various topics of current interest for the benefit of scientists of National Agricultural Research System (NARS). These training programmes cover specialized topics of current interest in statistics and agricultural sciences. There is yet another form of training courses, which are tailor made courses and are demand driven. The coverage in these courses is need based and the courses are organized for specific organizations from where the demand is received. The Institute has conducted such programmes for Indian Council of Forestry Research, Indian Statistical Services probationers and Senior officers of Central Statistical Organization, and many other organizations. The Institute has also conducted several international training programmes on request from FAO, particularly for African, Asian and Latin American countries. The Institute has broadened the horizon of capacity building by opening its doors to the agro-based private sector. One such training programme was organized for research personnel of E.I. DuPont Pvt. Ltd. The Institute has also conducted training programmes for the scientists / research personnel of CGIAR organizations such as ICARDA and Rice-Wheat Consortium for Indo-Gangetic plains.

On the basis of a resolution passed in 65th meeting of the ICAR Society, a Research Advisory Committee was constituted by the council on July 14, 1994 to perform the following functions:

1. To suggest research programmes based on national and global context of research in the thrust areas.
2. To review the research achievements of the Institute to see that these are consistent with the mandate of the institute.
3. Any other function that may be specifically assigned by the Director General, ICAR.

In any research and teaching institute, library is its biggest and most powerful resource, which provides support to its research and teaching programmes. The library of the Institute has witnessed many developments over the years. Presently, the Institute library enjoys the status of a regional library. It plays a vital role in meeting the information needs of the in-house users as well as users from the NARS. Library Information System is fully automated and bar-coded. Now an E-search and retrieval facility is available through LAN.

## 2. Mandate

Having made a humble beginning as a statistical section of the then Imperial Council of Agricultural Research in 1930, the Indian Agricultural Statistics Research Institute has come a big way and established itself as a pioneer in research and education in Agricultural Statistics and Computer Applications in the country. The vision of the institute is to use the power of Statistics as a science blended judiciously with Information Communication Technology to enhance the quality of agricultural research. To convert this vision into a reality, the institute has set for itself a mission to undertake research, teaching and training in Agricultural Statistics and Computer Applications so that these efforts culminate into improved quality of agricultural research and also meet the challenges of agricultural research in newer emerging areas. The functions and activities of the institute have been re-defined from time to time in the past. The present mandate of the Institute is

a. to undertake basic, applied, adaptive, strategic and anticipatory research in Agricultural Statistics and related fields and use these researches in meeting challenges and improving quality of agricultural research

b. to conduct post graduate teaching and in service, customized and sponsored training courses in Agricultural Statistics and Computer Application at National and International level so as to be a leading centre of excellence in Human Resource Development

c. to provide methodological support in strengthening National Agricultural Statistics System by establishing linkages with State Departments of Agriculture and allied fields, other research institutions, industry, etc.

d. to lead in development of Agricultural Knowledge Management Systems for NARS

e. to provide advisory and consultancy services for strengthening the NARS and undertaking sponsored research and consultancy for National and International organizations.

The research, teaching and training activities are carried out under the following six broad programmes cutting across the six Divisions:

1. Development and analysis of experimental designs for agricultural systems research

2. Forecasting and remote sensing techniques and statistical applications of GIS in agricultural systems
3. Development of techniques for planning and analysis of survey data including economic problems of current interest
4. Modelling and simulation techniques in biological systems
5. Development of information technology in agricultural research
6. Teaching and training in agricultural statistics and computer application.

**3. Organizational Set Up**

Consequent upon becoming a full fledged Institute of the ICAR in April 1970, the research, teaching and training activities of the Institute were organized in a number of Divisions with the main objective of formulation and implementation of research projects in specific fields. During April 1970, the following seven Divisions were created:

1. Statistical Research (Crop Sciences)
2. Statistical Research (Animal Sciences)
3. Sample Survey Methodology
4. Crop Forecasting Methodology
5. Econometric Analysis
6. Computer Science and Numerical Analysis
7. Training and Basic Research

Over time, the organizational structure also underwent changes as a result of the recommendations of the Quinquennial Review Teams (QRT). In 1985, the QRT suggested dismantling the Division of Training and Basic Research and adding a new division of Bio-statistics and Statistical Genetics. It also suggested restructuring the seven existing divisions into the following six divisions:

1. Sample Survey Methodology and Analysis of Survey Data
2. Design of Experiments and Analysis of Experimental Data
3. Bio-statistics and Statistical Genetics
4. Statistical Economics
5. Forecasting Techniques for Crops, Diseases and Pests
6. Computing Sciences

In 1998-99 once again on the recommendations of the QRT, the names of some of the Divisions were changed although the constitution remained the same. The new names were:

1. Sample Surveys
2. Design of Experiments
3. Biometrics
4. Statistical Economics
5. Forecasting Techniques for Crops, Diseases and Pests
6. Computer Applications

During 1999-2000 the ICAR suggested changing the names of some Divisions without affecting the mandate of the Divisions. The structure of the Divisions is the following:

1. Sample Surveys
2. Design of Experiments
3. Biometrics
4. Econometrics
5. Forecasting Techniques
6. Computer Applications

**4. Achievements**

The success stories of the Institute have been many and fairly wide spread. The contributions towards research, teaching and training have been monumental. It is a matter of great pride for the Institute that two of its scientists have received the most prestigious National Award in Statistics in memory of Late Dr. PV Sukhatme, for outstanding life time achievements in Statistics. One scientist is presently occupying the prestigious ICAR National Professor Chair, two scientists have been the National Fellow of the ICAR, one scientist received the GP Chatterjee Memorial Lecture award from Indian National Science Academy (INSA) and one scientist received the Shri Om Prakash Bhasin award for science and technology in the field of agriculture and allied sciences. Six scientists have been adjudged as the 'Best Teacher' of the PG School of IARI, New Delhi. Several of its scientists have received 'Young Scientist Award' from National Academy of Agricultural Sciences, Indian Council of Agricultural Research and many other scientific societies / associations. Several scientists have been the elected members and one scientist had been the Council member of the International Statistical Institute. Scientists are also Editors, Associate Editors and Members of the Editorial Board of many National and International Journals. The scientists of the institute have published eight text books, four handbooks and many monographs. Agricultural Research Data Book consisting of collation of information on various aspects of agriculture is also brought out. In the sequel, an attempt has been made to provide a glimpse of the achievements of the institute over the past 50 years.

**4.1 Research achievements**

The Institute has made some outstanding and useful contributions to the research in Agricultural Statistics in the fields like Design of Experiments, Statistical Genetics, Forecasting techniques, Statistical Modelling, Sample Surveys, Econometrics, Computer Applications in Agriculture, Software development, etc. IASRI has conducted basic and original research on many topics in design of experiments, sample surveys, biometrics, etc. By way of its publications in International Journals, the Institute has come to be known globally. IASRI has been providing and continues to provide support to the NARS by way of analyzing voluminous data using advanced and appropriate analytical techniques and providing efficient designs for experimentation. IASRI has also been very actively pursuing advisory service that has enabled the institute to enrich the quality of agricultural research in the NARS. Through its advisory, IASRI has made its presence visibly felt in NARS and now experimenters look to IASRI for designing experiments and analysis of experimental data. The efficient designs like balanced incomplete block designs, group divisible and extended group divisible designs, reinforced extended group divisible designs, square and rectangular lattice designs, alpha designs, reinforced alpha designs, augmented designs, designs for fitting

response surfaces, etc. and advanced analytical techniques including, contrast analysis, linear models with nested structures, experiments with mixtures methodology, mixed effects models, biplot, etc. have been adopted by the experimenters. The analytical techniques for estimating/projecting the Energy Requirement in the Agricultural Sector has been exploited for the analysis of countrywide data. The analytical techniques for the analysis of data from the experiments conducted to study the post harvest storage behaviour of the perishable commodities like fruits and vegetables are being widely used in NARS. The Institute works in close collaboration with NARS organizations and has many projects being run at the institute in collaboration with All India Co-ordinated Research Projects and ICAR Institutes. Institute has developed linkages with the CGIAR organizations such as CIMMYT, IRRI and ICARDA. The status of experimentation is now changing and with the support provided in terms of suggesting efficient designs and analyzing the data using modern complicated statistical tools, the research publications of the agricultural scientists are finding a place in high impact factor international journals.

The methodology for General Crop Estimation Surveys (GCES), cost of cultivation studies, Integrated Sample Surveys (ISS) for livestock product estimation, fruits and vegetable survey, which are being adopted throughout the country are research efforts of IASRI. Methodology based on small area estimation technique for National Agricultural Insurance Scheme suggested by IASRI has been pilot tested in the country. A status paper on chronological development and present status of information support system for management of agriculture was prepared as a part of State of Indian Farmer: A Millennium Study of Ministry of Agriculture. The sample survey methodology for imported fertilizer quality assessment, fish resources estimation, flower production estimation, horticultural crops estimations, etc. have been developed and passed on to the user agencies. Integrated methodology for estimation of multiple crop area of different crops in North Eastern Hilly Regions using Remote Sensing data has been developed.

The Institute has also made very significant contributions in developing the analytical techniques for the estimation of genetic parameters, models for pre-harvest forecasting of crop yields, models for forewarning of incidence of pests and diseases and econometrics and statistical modeling of biological phenomena using structural time series and machine learning approaches. The methodology developed for forecasting based on weather variables and agricultural inputs was used by Space Application Centre, Ahmedabad, to obtain forecast of Wheat yield at national level. Models developed for forewarning of aphids in mustard crop were used by National Research Centre for Rapeseed and Mustard to provide forewarning to farmers, which enabled them to optimize plant protection measures and save resources on unnecessary sprays consecutively for three years. The modification in the procedure of estimation of genetic parameters has been suggested for incorporating the effect of unbalancedness, presence of outliers, aberrant observations and non-normality of data sets. Procedures for studying genotype × environment interactions have been developed and used for the analysis of data generated from Crop Improvement Programmes.

### 4.2 Achievements in information communication technology

Today information is accessible to anyone and anywhere. It is increasingly becoming a basic economic resource and a structuring factor in society. The Institute has the capability of development of Information Systems, Decision Support Systems and Expert Systems. Realizing the need of integration of databases to prepare a comprehensive knowledge warehouse that can provide desired information in time to the planners, decision-makers and developmental agencies, Integrated National Agricultural Resources

Information System (INARIS) with the active support of 13 sister institutes as partners has been developed. The data warehouse comprises of 59 databases on agricultural technologies of different sectors of agriculture and related agricultural statistics at districts/state/national levels, population census including village level population data as well as tehshil level household assets and livestock census. Subject-wise data marts have been designed, multi-dimensional data cubes developed and published in the form of on-line decision support system. The Institute has also developed information systems for agricultural field experiments, animal experiments and long term fertilizer experiments conducted in NARS. Besides, a comprehensive Personnel Information Management Information System network has been implemented for the ICAR for man power planning, administrative decision making, and monitoring. For National Agricultural Technology Project, a Project Information and Management System Network (PIMSNET) was developed and implemented for concurrent monitoring and evaluation of 845 projects. A National Information System on Agricultural Education Network in India (NISAGENET) has been designed, developed and implemented on the recommendations of the National Statistical Commission (NSC-2001) so as to maintain and update the data regularly on parameters related to agricultural education in India. An Expert System on Wheat crop Management has also been developed and implemented.

A milestone in the research programmes of the Institute was created when it started developing indigenous statistical software packages mainly for analysis of agricultural research and animal breeding data and then generation of experimental designs for various experimental situations, both unstructured and factorial structure of treatments, catalogues of designs, randomized layout of design and analysis of data generated. Statistical package for analysis of survey data is perhaps one of its kinds of indigenous packages.

The creation of Design Resources Server, an e-learning and e-advisory resource for the experimenters, has been another revolution in the growth of the institute. The server provides a platform to popularize and disseminate research and also to further strengthen research in newer emerging areas in design of experiments among peers over the globe in general and among the agricultural scientists in particular so as to meet the emerging challenges of agricultural research. This is hosted at www.iasri.res.in/design.

## 5. Future Agenda

In an effort to maintain the past glory, the Institute plans to maintain a judicious balance of basic research and innovative applied research and also to expand the horizon of the teaching and training programmes so as to remain the leader in human resource development. The Institute would delve on developing roadmaps on combating the major problems of depleting natural resources, climate change, uncertainties of economy and marketing by addressing the statistical issues involved in it. The fusion of Statistics and Informatics would help addressing this alarming problem. The Institute would undertake research in newer emerging areas with fresh vigor and zeal. Block designs having orthogonal factorial structure with balance have many applications in crop sequence experiments and in cDNA micro array experiments. Generation of designs with full efficiency on main effects and controlled efficiency on interactions is one area of research. Fractional factorial designs, permitting estimation of main effects and some selected, important two factor interactions, with large number of factors and with scarce resources is another important area of research with many applications. A special class of fractional factorials with number of design runs smaller than the total degrees of freedom for all the main effects and the intercept

term, called as supersaturated designs, is also an important area of research that would need attention. Generation of orthogonal arrays with varying symbols, particularly resolvable or nested arrays assumes significance in view of their applications in many fields. A special class of orthogonal arrays of strength one is very useful in computer experiments. Integrated farming systems research is becoming the order of the day. Data designing for such systems will be of considerable importance and the institute would take lead in this area. Small area estimation is another important area of research that would be addressed to and its applications in policy planning at block and village levels, in crop yield estimation, demographic studies and health sciences, poverty and malnutrition, etc. will be made. Bayesian analysis of survey data and experimental data would be another area of active research. The institute needs to take a deep plunge into the internationally adapted areas of research in statistical genomics, proteomics, bio-informatics and nanotechnology. The creation of a bioinformatics Lab is the first step in this direction. The institute would also like to initiate action on starting an M.Sc. degree course in Bio-informatics jointly with IARI. Remote sensing, GIS, ANN and other machine learning algorithms would be used in developing models for biological and economic phenomena. These techniques would also be used for developing models for forecasting crop yields, forewarning and early warning systems for incidence of pest and disease. Data mining activities would be taken up rigorously to add value to the various information systems and the data ware house available at IASRI.

The important activity of developing indigenous statistical software packages would be rejuvenated with more vigor and zeal. A blend of statistics and computer applications would help in developing agricultural knowledge information system and decision support systems. Such systems would be beneficial in building a bridge between the farm scientists and the farming community so as to bring prosperity to the farmers.

The institute has also been playing a leadership role in developing methodologies for the National Agricultural Statistics System (NASS). Several methodologies developed have been passed on to the state departments for replicating them in the states. This activity needs to be further reinvigorated with new zeal and vigour to provide support to the NASS to help in proper implementation of policy planning and developmental programmes. The small area estimation techniques, the categorical data analysis of survey data, regression analysis of survey data, etc. would be judiciously and innovatively used for developing methodologies.

IASRI has also been very actively pursuing advisory service that has enabled the institute to enrich the quality of agricultural research in the NARS. This activity has made a strong impact in providing help to improve the quality of agricultural research. This activity needs to be further strengthened to become more effective in helping make agricultural research globally competitive.

The developmental activity is a continuing process. It is hoped that the institute would keep growing with an accelerated pace. It is also wished that the institute would continue providing leadership in the disciplines of agricultural statistics and computer applications. The institute would also revisit its role in agricultural research and would keep helping improve the quality of agricultural research by undertaking research in other newer emerging areas.

# Reminiscences of IASRI

**A.K. Nigam***

- It is heartening that Indian Agricultural Statistics Research Institute (IASRI) is celebrating Golden Jubilee of its foundation on 02 July 2009. Starting as a small statistical section in the then Imperial Council of Agricultural Research and in no time taking the shape of an institute speaks volumes of its journey towards excellence. The institute is known for its research and teaching programmes both nationally and internationally.
- I have been associated with IASRI for several decades. This includes two periods of stay at IASRI, first as a student and Research Apprentice during 1962-1965, and then as a faculty in various capacities from August 1972 to April 1987. I had an occasion to work under several Directors – Late Dr. V. G. Panse, Late Dr. Daroga Singh, Dr. M. N. Das and Dr. Prem Narain. While being away at Banaras Hindu University (BHU) during 1966 to 1972, I still had my roots at IASRI as I was pursuing my Ph.D. under Dr. Das. It is indeed a great pleasure for me to write my experiences / impressions about this great institution.
- My association with IASRI as a student was most rewarding and I spent my most memorable days as a hosteller. The competition in Professional Statisticians Certificate Course (PSCC) was fierce as there was a historical divide between north and south – a fellow from Lucknow University had been the topper for the last several years. The teaching here was entirely different from traditional universities as more emphasis was put on concepts and real world live data examples rather than derivations and proofs of theorems. I must confess that the best teacher that I came across was in our practical classes. Mr. M. B. Jain always laid more emphasis on concepts, and was open for discussions. The Practical Register was my most prized possession; it helped me a lot in imparting the concepts of Statistics to agricultural graduates when I moved to BHU as a lecturer in 1966. It was also used by many of my colleagues as a reference material for teaching. It also served as resource material for the IASRI Handbook on *Analysis of Agricultural Experiments*, compiled and edited by me and Dr. V.K. Gupta.
- The life at the hostel, which is now the Panse Guest House, changed me from a timid shy boy to a flamboyant, ready to take the world on its stride, a complete youth. I quickly learnt and in no time became champion of games like Table tennis and Bridge. I still recall that we (V.K. Bhargava was my partner) defeated the duo – V.G. Panse and B.V. Sukhatme in the final of open bridge tournament, which was played at Dr. Panse's residence. I also won many other titles – doubles in badminton and carom, courtesy V. K. Bhargava, an outstanding sports person, who happened to be my partner.
- After completing PSCC and Diploma in Agricultural Statistics, I joined as a Research Apprentice at IASRI and worked under Shri V.N. Amble. It was at this stage that I had my first real exposure to data analysis and computer programming. I vividly remember how happy we were when we jointly wrote our first *8 in 1* computer programmes in Fortran IV for fitting 8 production functions in one go.

*Executive President, Institute of Applied Statistics and Development Studies, Lucknow.

- My second stint with IASRI started in August 1972 as a faculty member. A strong group led by Dr. Aloke Dey was already very active in research in combinatorics related to design of experiments. They had already published good quality papers in International Journals. I had no option but to join this race for survival and self esteem. It took almost five years to develop my own team of collaborators, mostly new faces, being just fresh post Graduates and Doctorate students. Within a few years, we started competing with the other team.
- Our work was not only restricted to combinatorics. We never lost sight of importance of analysis component in design of experiments and of applications in different areas. We even took up to sample surveys and our work on interplay of design of experiments and survey sampling received good acclaims internationally. The potential of the work impressed Professor J. N. K. Rao who invited me for a visiting assignment to Carleton University immediately after I quit IASRI in 1987. I followed up my first visit by consecutive four more visits in subsequent years and then I discontinued these because of family pressures back at home. During these years, I worked on the famous problem of optimum controlled sampling, and on bootstraps.
- Somehow, during these years it so happened that the institute started laying more emphasis on applied research, and at times to trivial applications, rather than basic research. Most of the applied research was still along the same lines and themes initiated by the duo – Dr. P. V. Sukhatme and Dr. V. G. Panse. Most of the work of these stalwarts was innovative using appropriate techniques and had earned worldwide acclaims. But, over the years there was a declining trend in the quality of applied work as same ideas and techniques could no more be classified as innovative after a gap of 2-3 decades. Some of us were eager that the quality of applied research became innovative again, but somehow, the idea had very few takers. Some of us felt that this change in attitude was perhaps not good for the healthy growth of a research institute. It cannot be denied that applied research is important, but it has to be innovative. For the existence of a research organization, it is also equally important to conduct basic research which has a strong potential of immediate application or in near future to solve problems of agricultural sciences.
- Late Dr. V. G. Panse has always been a legend because of his balanced approach to research. While he himself was basically not a theoretician, he always appreciated good theoretical work. I am reminded of the afternoon when I was giving a seminar talk on my Diploma thesis work related to construction of rotatable designs – pure theory based upon trial and error. The first question after my talk came from a very senior faculty member – *what is the application of your work in agriculture?* Almost simultaneously Dr. Panse expressed his anguish and intervened with his strong but motivating remark, '*this is an outstanding theoretical contribution'*.
- I really felt honored when I was invited to deliver Panse Memorial lecture at the International Conference on Statistics and Informatics in Agricultural Research held at Delhi in 2006 to mark the Diamond Jubilee Celebrations of the foundation of Indian Society of Agricultural Statistics. It may not be out of place to record here some part of the tributes I paid to this great person – "*His contributions to agricultural statistics are monumental and it would do no justice if one tries to enumerate them. While Dr. Panse was a strict administrator, he was also a very kind hearted person. He was a terror to incompetence, but an impetus to the conscientious work.*"

- Although I left IASRI in 1987, after taking voluntary retirement, my connections with the Institute remained intact. From outside, it was a helpless feeling to note that ICAR was struggling to appoint a competent Director for IASRI. It was somewhat satisfying when I was approached by the then DDG (Engg), ICAR to take over the responsibility of leading the institute as Director. This unfortunately came a little too late and because of my other pre-occupations I had to decline the offer.
- I continued my association with the Institute through different Committees. I happened to become the Chairman of QRT for the period 2000-2005, and Member, RAC jointly for IASRI and NCAP. This provided me an opportunity to have a closer look at the activities of the Institute. It was really heartening to see that the Institute has made a valiant attempt to improve the status of agricultural experimentation in National Agricultural Research System (NARS) by way of rigorous advisory services being pursued by the scientists of the Institute. It was a pleasant surprise for me to find that alpha designs, extended group divisible designs, second order rotatable designs, experiments with mixtures, etc. have been used by the experimenters in NARS. The applications of non-linear models, state-space models, models for forewarning of pests and diseases in crops, small area estimation have been made by the scientists of the institute for explaining the behaviour of complicated and multi-dimensional agricultural research. Development of indigenous software and information systems is another area in which the institute has taken a big leap.
- In spite of the achievements as above in applied research, it is disturbing to note that fewer scientists are involved now in basic research. This has resulted into very few publications in International Journals with high impact factors. Basic research is the input to applied research. Therefore, this trend needs to be reversed and a balanced approach in basic and innovative applied research needs to be maintained. I, therefore, strongly feel that reviving the Division of Training and Basic Research at the institute should be top priority in the agenda of progress of the institute. This has also been one of the recommendations of QRT.
- For doing good quality basic and innovative applied research work, it is essential to have a blend of young new talent and experienced competent scientists. However, it is disheartening to note that the average age of the scientists at IASRI is very high. Since last many years, no new scientists have been posted at the Institute. Therefore, immediate efforts should be made to infuse new blood into the system so as to undertake research in newer emerging challenging areas.
- As stated earlier, the Institute has also developed outstanding methodologies that have helped improve the National Agricultural Statistical System (NASS) of the country. But some studies have been trivial in nature and seem repetitive. Perhaps there has been a seemingly trivial imbalance in the efforts of the institute towards applied research for NARS and NASS. This needs to be addressed for the sound health of the institute.
- The institute has some brilliant workers but unfortunately they are few in numbers. These workers have to be given all impetus and encouragement, and in turn, they are to provide leadership to motivate and energize the young workers. A spirit of competitiveness is to be induced to derive the best from the scientists. Then alone, it would attain the heights and recognition which it is aiming for.

# Indian Agricultural Statistics Research Institute (IASRI): A Student's Experience

**Shyamal D. Peddada***

I congratulate IASRI on its golden jubilee celebrations! Throughout its history, the statisticians at IASRI played a crucial role in Indian agricultural research by developing important designs for conducting experiments and surveys. Some of the pioneering works in the areas of *Design of Experiments, Sample Surveys and Statistical Genetics* were conducted by the statisticians at IASRI. I consider myself to be very fortunate that I was trained by some of these outstanding researchers and teachers. It is indeed an honor for me to write this short note recounting my academic experiences at IASRI while I was a student there.

## Training at IASRI

I never had a course in Probability or Statistics prior to joining the M.Sc. program and hence everything was new and challenging during the first trimester. It took me a while to appreciate the inductive logic of statistics - drawing inferences about a population using a random sample from the population. However, I was very fortunate that during first trimester I was taught by Professors Aloke Dey, V. K. Gupta, S. C. Rai and U. G. Nadkarni, who laid the foundations in matrix algebra, statistical theory and data analysis. After the first trimester I started to enjoy statistics very much and things started to click and make sense. A unique feature of the statistics program at IASRI was that almost every *theory course* was accompanied by a *practical class (or lab)* where we analyzed real data from agriculture. Not only did these practical classes reinforce the relevance of statistical theory and methodology for solving practical problems, they also taught us some of the important computational algorithms. In those days we used mechanical computing machines and not electronic calculators or computers, although IASRI had two "state of the art" computers (IBM 1620 and Burroughs 4700). These punch-card generation mainframe computers were typically used for research purposes only and not for class-room projects. Since the mechanical computers were physically demanding, we had to learn clever ways of using them, which made us to think deeply about the algorithms we were using.

Although I did not take a course taught by every member of the training unit, the general consensus was that members of the training unit were dedicated researchers and committed teachers. They had a serious "no-nonsense" approach to teaching with very high expectations from their students. Several of my professors were also very prolific in publishing papers, and that served as a motivation for me to do Ph.D. in statistics and follow in their footsteps. In fact I discovered research problem for my Ph.D. dissertation while taking a course on statistical genetics taught by Professor V. K. Bhatia. While analyzing a genetics data in the practical class taught by Professor Bhatia, my classmates and I independently obtained a negative estimate for the variance of a random effect term in a linear mixed model. Our initial reaction was that perhaps we were all using a wrong formula or we were making an error in our calculations

*Senior Investigator, Biostatistics Branch, National Institute of Environmental Health Sciences (NIH), Research Triangle Park, NC 27709.

(recall that we were not using electronic computers or even an electronic calculator but simple mechanical computing machines). We did not recognize that this could happen with a positive probability when using an Analysis of Variance (ANOVA) based estimator. During the next lecture Professor Bhatia explained to us that an unbiased estimator of the variance component can take a negative value with a positive probability. I found it to be very intriguing and decided that for my Ph.D. dissertation I would work on the problem of non-negative estimation of variance components. That is exactly what I did three years later in my Ph.D. under Professor C. R. Rao at University of Pittsburgh.

In my view, the M.Sc. program at IASRI (in 1970's) was already years ahead of many graduate programs around the world. It was a very inter-disciplinary program, a popular theme these days at many universities in the US. In our M.Sc. program we were required to take a broad spectrum of courses in statistics and agriculture. In all I took about 70 credit hours of course work, which included 50 credit hours of statistics courses (a few of which were math courses) and the remaining were courses in various disciplines of agriculture, such as genetics, agronomy, plant pathology, entomology, soil science, etc.

In addition to the course work, we were required to do a thesis project in our major area, which was Design of Experiments for me. My thesis advisor was Professor P. N. Soni, a remarkably kind man who was very nice to work with. For my thesis, I worked on response curves for fertilizer trials. Although it was an applied project, I learnt a lot about basic data analysis. Through the course work and my thesis work I appreciated the importance of statistics and mathematics for solving real world problems. The training I received at IASRI continues to have a major impact on my research program and my passion for developing appropriate statistical methods for solving real world scientific problems in a wide range of topics.

The M.Sc. program at IASRI was a well rounded program. In addition to taking a broad range of courses and doing a thesis project, students were required to give seminar lectures on a regular basis. These seminars were well attended by all students and faculty. Furthermore, on a regular basis, we had speakers from all over the world visiting the institute to give seminars. By giving seminars and listening to seminars on a regular basis, I learnt a good deal about how to communicate. It made me a good communicator and helped me as a teacher here in the US. In 1985, I was nominated for teaching excellence award at Central Michigan University and I largely attribute this nomination to the training I received at IASRI.

I am grateful for all the wonderful experiences I had at IASRI. In particular, I am very indebted to all my professors for laying the foundations of my academic career.

## Present and the Future of IASRI

### Faculty

Since leaving the institute in 1980, I have tried to keep in contact with my Professors at IASRI, and got updates on the changes and progress made at the institute. Several faculty members I used to know when I was a student have now either retired or moved to other places. The remaining faculty provided important leadership and direction, and consequently the institute continues to enjoy its key position on the international stage. Furthermore, they hired several very impressive statisticians who are making their mark and keeping the tradition of IASRI alive. Not only that IASRI continues to enjoy a great international reputation in the traditional fields of *Design of Experiments, Sample Surveys and Statistical*

*Genetics*, but they are also developing strong programs in *Biometrics and Statistical Modeling*, *Forecasting*, *Statistical Computing*. I congratulate each member of these divisions for conducting outstanding research consistent with IASRI tradition.

**Statistical computing and software**

I am very impressed to see that IASRI has developed useful statistical packages such as, Statistical Package for Block Designs (SPBD 1.0), Statistical Package for Agricultural Research (SPAR1 and SPAR2), Statistical Package for Factorial Experiments (SPFE 1.0), Statistical Package for Augmented Designs (SPAD), etc. To a large extent these packages reflect the research work conducted by the researchers at IASRI. I also visited the Design Resources Server (www.iasri.res.in/design/) developed by the scientists at the institute. It is a very impressive website with all kinds of important tools and resources for a researcher. The layout of the page is very nice and I congratulate the staff maintaining the website. By developing these software packages and web tools, the scientists of the institute are demonstrating that their research work is not only of theoretical value, but has important practical relevance.

**Genomics and environmental health**

During my last visit to IASRI in 2007, I was delighted to see that the institute is actually positioning itself very well for the age of genomics by hiring brilliant young scientists such as Dr. Rajender Parsad, and Dr. Rao. The Indian agriculture would greatly benefit by expanding IASRI's research program in the area of genomics by adding more faculty and developing strong collaborations with IARI scientists and others within India and outside India. Perhaps they could conduct workshops and conferences on various topics related to bioinformatics by bringing statisticians, bioinformaticians, computer scientists and biologists together. On the methodological front, this institute has the unique expertise to develop new experimental designs that are suitable for experiments relating to the study of plant genome. I am not sure if theory of optimal designs for high throughput experiments such as microarrays, has been developed yet. Perhaps IASRI may want to take the lead research in this area, if they are not already doing so.

**Students of IASRI**

Since the inception of its training program about five decades ago, IASRI enjoyed a great tradition of producing strong students who had (or having) successful careers all over the world. During my last visit I had the opportunity to meet with the next generation of scientists who are being trained at IASRI. I am sure these students are having the same positive experiences as I did when I was a student. I am sure each of them will succeed in their careers. I wish them all the best.

**On a personal note**

This summary of my experiences at IASRI would not be complete if I did not pay tribute to a very important person in my life, namely, my father (late) Professor P. Prabhakara Rao, who served ICAR for about 30 years and was a member of the training unit at IASRI. Both my parents were regarded as outstanding teachers at their respective institutions, so a career in teaching and research was the obvious one for me. My father was the main source of inspiration for me to consider a career in statistical science. I never regretted that I chose the field and profession he loved so much. I hope some of his passion for the field rubbed off of me as well. I am eternally grateful to him.

# A Profile of Design of Experiments at IASRI

**Rajender Parsad and V.K. Gupta**

## 1. Importance and Historical Development

The genesis of the Institute dates back to 1930 when a small Statistical Section came into existence in the then Imperial Council of Agricultural Research to assist the State Departments of Agriculture and Animal Husbandry in planning their experiments, analysis of experimental data, interpretation of results and also rendering advice on the formulation of the technical programmes and examining the progress reports of the schemes funded by the Council. Ever since then, the division of Design of Experiments has been the pride division of the Indian Agricultural Statistics Research Institute (IASRI) enjoying recognition all over the globe.

Designing experiments forms the backbone of any research endeavor in the discipline of agriculture and allied sciences. Dr. MS Swaminathan in his message to International Conference on Statistics and Informatics in Agricultural Research said, "*It is the effective use of the tools of statistical design of experiments that paved the way for the green revolution*." This statement spells out the importance of the subject of Design of Experiments for improving the quality of agricultural research.

The advisory work relating to agriculture and animal husbandry of the statistical section was separated into two units in 1945. The two units *viz.*, crops and animal husbandry were separately under the charge of a statistician. After getting the status of a full fledged Institute in 1970, the concept of Divisions was introduced and the Division of Statistical Research in Crop Sciences was created. On the recommendations of Quinquennial Review Teams (QRT) in 1985, the Division of Statistical Research in Crop Sciences was renamed as Division of Design of Experiments and Analysis of Experimental Data. Again on the recommendations of QRT in 1998-99, the Division was rechristened as Division of Design of Experiments.

In the early years of development of the Institute, the emphasis of research was more on the analysis of designed experiments rather than developing new designs. At IASRI of those days, one of the major activities was 'advisory' work related to agricultural and animal husbandry experiments. During the course of such advisory work, it was often found necessary to develop methods of analysis appropriate to the experiment and to that end, several contributions were made. It was during the decade of 1950 that the Institute started making important theoretical contributions in developing new methods of construction and analysis of experiments. In late 1950's with the efforts of Dr JN Srivastava suitable confounded designs for asymmetrical (or mixed) factorial experiments were obtained. The decade of 1960 saw many major contributions, particularly from Dr MN Das in designs for factorial experiments, rotatable and other designs for fitting response surfaces and designs for experiments with mixtures. During early1970's till late 1980's the research in Design of Experiments got a very good impetus under the guidance of two leading statisticians Dr. Aloke Dey and Dr. AK Nigam. Among other statisticians who helped in building collaborations and analysis of voluminous data were Shri MG Sardana, Shri. KS Krishnan, Dr. SK Raheja, Ms. CR Leelawathi, Shri SR Bapat, Dr. PN Bhargava, Shri RC Khanna, Shri PN Soni, to mention a few. The present day generation has given a new shape to the theory and applications of design of experiments by maintaining an ideal balance between theory and applications. The scientists of the Division participate

actively in planning and designing of experiments in the National Agricultural Research System (NARS) and also involve themselves in the analysis of experimental data.

The present agenda of research in the Division is the following:

- Cropping system research
- Experimental designs for agricultural, animal, agro-forestry and fisheries research
- Development of information systems for agricultural and animal experiments

## 2. Designs in National Agricultural Research System (NARS)

Generally speaking, there are two broad categories of experiments, one comprising of designs with unstructured treatments also popularly known as varietal designs. These designs are single factor experiments with many levels and all possible pair wise treatment comparisons or a subset of such comparisons are of interest to the experimenter. The other type of experiments is the factorial experiments with several factors each having many levels. All the possible combinations of the levels of all the factors, or a subset of all possible level combinations, form the treatments. The interest in such experiments is in the factorial effects like main effects of all the factors and their interactions. Not only is the choice of design important, but the choice of treatments in the experiment is also important. And while selecting a design one should not lose sight of the inference problem and the analysis to be performed. A design good for one inference problem may become worse for another inference problem. In the sequel we describe salient achievements in the field of design of experiments.

## 3. Research Achievements

The achievements of the division of design of experiments are many and widely spread. Several important, useful and original contributions have been made in both basic research and innovative applications of the theory of statistical designs and analysis of experimental data. The scientists of the division have published 5 text books *viz*., (i) Handbook of Design and Analysis of Agricultural Experiments; (ii) Design and Analysis of Experiments; (iii) Theory of Block Designs; (iv) Characterization and Analysis of Block Designs; and (v) Fractional Factorial Plans. Five research monographs *viz.* (i) Supplemented Block Designs; (ii) Orthogonal Main Effect Plans; (iii) Optimality of Multi-dimensional Designs; (iv) $\alpha$-Designs; and (v) Hadamard Matrices have been published and have received wide appreciation because of their usefulness. The Division has established closer linkages and collaborations with All India Co-Ordinated Research Projects (AICRP) of the Indian Council of Agricultural Research (ICAR). Collaborations have also been established with CGIAR organizations such as Rice–Wheat Consortium for Indo-Gangetic plains; International Centre for Agricultural Research in Dry Areas, etc. The Division has also taken lead in getting outside funded projects, particularly from the AP-Cess Fund of ICAR, Department of Science and Technology and National Agricultural Innovation Project.

The research work done in the division can be categorized into following broad headings.

### 3.1 Basic research

The intense research activity in the area of design and analysis of experiments has continued rigorously throughout. Many notable contributions have been made in basic research that has been accepted at an international level. To mention a few: designs for symmetrical and asymmetrical factorial experiments;

designs for fitting response surfaces; designs for experiments with mixtures; weighing designs; fractional factorial and orthogonal resolution plans, orthogonal arrays; supersaturated designs for two-level, multi-level and mixed level factorial experiments; block designs with factorial treatment structure with emphasis on extended group divisible designs; variance balanced, efficiency balanced, partially efficiency balanced, simple partially efficiency balanced and generalized efficiency balanced block designs; optimality and combinatorial aspects of designs (designs for making test treatments-control treatments comparisons; row-column designs and structurally incomplete row-column designs; resolvable block designs; designs for bioassays, cross over designs; designs for estimation of competition effects; designs with correlated observations; designs with nested blocking structure; designs for microarray experiments; designs for bio-equivalence trials; designs for multi-stage experiments with more than one set of non-interacting treatments applied in succession; designs for diallel and double cross experiments); designs for on-Farm trials; design and analysis of multi-response experiments; detection and handling of outliers, etc.

Robustness aspects of designs against loss of data (both single factor and multi-factor experiments); presence of outliers, interchange and exchange of treatments, model inadequacy, presence of systematic trends, etc. have been studied.

More recently, a breakthrough in research has been the computer aided search of efficient designs for various experimental settings. This has been possible by making use of computer algorithms based on exchange and interchange procedures.

Use of combinatorial designs in obtaining efficient survey sampling plans with unequal probabilities of selection leading thereby to controlled selection was another major contribution. The use of mixed orthogonal arrays as Balanced Repeated Replications in the variance estimation of a non-linear statistic from a large scale complex survey data was an original contribution of the Division. Balanced sampling plans have also been obtained using the applications of polygonal designs. A detailed description of these achievements is given in separate article on “Glimpses of Basic Research in Design of Experiments at IASRI”.

### 3.2 Applied research

The contributions made in the applied research in Design of Experiments have been monumental. Significant research contributions in applied research are described in the sequel.

#### Shape and size of plots and blocks

It is well known that the statistical designing of experiments requires proper blocking of experimental units. The past experience in the NARS shows that in majority of the experiments, the block mean square is small as compared to the error mean square. As a consequence, the precision for treatment comparisons is small. The analysis of several uniformity trials data sets has revealed that long rectangular or square blocks are not homogeneous because several fertility patches are nested within these blocks. So either incomplete block designs or nested block designs with smaller sub blocks nested within bigger blocks would be more beneficial in controlling experimental error. In the absence of uniformity trials, strategy for preparing contour maps and for their periodic updation has also been devised. The procedure of preparing yield and fertility contour maps based on moving averages and cluster analysis developed for determining the shape of plots and blocks has been used by Allahabad Agricultural Institute, Allahabad in

preparing yield and fertility contour maps for Development of Site Specific Nutrient Management.

**Diagnostics in designed field experiments**

In NARS, a large number of experiments are conducted. The data from these experiments is generally analyzed without giving any importance to the validity of assumptions. The departures from these assumptions make the interpretation based on these statistical techniques invalid. Division has taken a lead in carrying out a diagnostic study of design and analysis of field experiments using the data from a large number of experiments. The main emphasis was given on normality and homogeneity of error variances. The departure from these assumptions was observed in large number of cases. Remedial measures like Box-Cox transformations and non-parametric measures have been suggested. A SAS code that is capable of carrying out the diagnostics and remedial measures at one go has been developed. This will encourage the use of diagnostic measures in the experimental data and arriving at statistically valid conclusions. The experimental data with non-normal and /or heterogeneous errors is then tested for presence of outliers. It was found that in most of these experiments, at least one outlier was present. Nearest neighbour methodology has been used for improving the precision of treatment comparisons where the randomized layout was available. Dr. R. Chidambaram, Principal Scientific Advisor to Government of India in his inaugural address delivered during the 57$^{th}$ Annual Conference of Indian Society of Agricultural Statistics held at GBPUA & T, Pantnagar during February 05 – 07, 2004 emphasized the importance of the work done on diagnostics in designed field experiments.

**Plot sampling in designed field experiments**

The plot-sampled data from designed field experiments is generally analyzed as per analytical procedure of the design adopted on the plot means. The plot variances, however, are different from plot to plot and may violate the assumptions of constancy of variances and normality of observations. Through some empirical investigations, plot-sampled data from field experiments has been used to evolve variance-stabilizing transformations so as to satisfy the assumptions of analysis of variance and it is recommended that this data may be used for obtaining appropriate transformations. Fuzzy regression theory was modified and used in the analysis of plot sampled data from field experiments.

**On-farm research**

On-Farm research helps in testing the technologies developed at the research stations by taking into account the realistic environment. Therefore, more and more emphasis is being laid by different organizations on On-Farm research. Based on the degree of the control of the researchers and farmers, the On-Farm trials can be classified into three categories *viz.* (i) trials designed and managed by researchers, (ii) trials designed by researcher and managed by farmers, and (iii) trials designed and managed by farmers. The on-farm research suffers primarily from lack of control on variability due to (a) variation in farmers' managerial skills and resources, (b) plot to plot variation, etc. It has been shown that the resolvable block designs are quite helpful in On-Farm research to take care of variation due to management skills of the farmers or plot to plot variation. An application of augmented designs has been shown in the trials designed and managed by farmers, with some intervention of the researcher. A linear, nested mixed effects model has been suggested for the analysis of on-farm trial data currently being generated by Project Directorate of Cropping Systems Research (PDCSR), Modipuram. This will be helpful in identifying the specific development blocks that seemed to favour one treatment over the other. This will

also help in identification of the recommendation zones that is not possible through comparing the treatment effects averaged over all the development blocks alone.

**Fertilizer-response studies**

The data from experiments on cultivators' fields have been used extensively for obtaining fertilizer-response ratios. These ratios have also been used by the researchers in studying the trends of fertilizer response ratios over years. The fertilizer response ratios have also been obtained for micronutrients. The fertilize-response ratios developed have been used by the planners and Task Force on Balanced Use of Fertilizers constituted by Ministry of Agriculture, Government of India.

The Institute has been engaged in formulating and updating yardsticks of additional production from time to time of various crops from the use of several agricultural inputs besides fertilizers. Yardsticks of rice were worked out for high yielding, locally improved and locally indigenous varieties for nitrogen, phosphorus, potassium, zinc, lime, FYM, green manure, bio-fertilizers, herbicides, etc. Yardsticks due to irrigation were worked out for various crops utilizing the limited available data.

Statistical investigations on the fertilizer use efficiency in relation to cultural practices have also been carried out. It has been shown that through the use of recommended cultural practices, fertilizer efficiency and fertilizer response ratios can be improved.

A study entitled 'Statistical evaluation of fertilizer requirements according to dates of sowing' was carried out with the objective of estimating fertilizer requirement corresponding to the normal and delayed sowing for different crops.

Some studies on the estimation of direct and residual effects of nitrogenous fertilizers applied alone or in combination with organic and biofertilizers *viz*., farmyard manure (FYM), Azolla in crop sequences has been made in the division. Application of slow release nitrogenous fertilizers to Kharif-rice was quite useful as compared to urea both from the point of view of productivity and stability considerations. Similarly nitrogen-requirement can also be met through Azolla alone or in combination with organic source.

A study on the interactions with reference to resource constraints of agronomic factors was carried out with the broad objective of identifying certain interactions in relation to crop production, of different crops at reduced level of certain agronomic factors and to obtain optimum number of replications, locations and years required to test the performance of such interactions.

**Long term fertilizer experiments**

Under All India Coordinated Research Project (AICRP) on Long Term Fertilizer Experiments (LTFE) continuous cropping and manuring for more than two decades has resulted in sizeable build up of certain nutrients like phosphorus or depletion of some plant nutrients like zinc in the soil at several locations. To examine the reduction or even temporary suspension of phosphatic fertilizers or addition of any other deficient nutrient element over the prescribed minimum, a nested two way design has been suggested using mid course bifurcation of plots in one of the replications. These bifurcated plots were used for superimposition of suitable treatments.

In long-term fertilizer experiments the unit plot sizes range between 150 to 300 square meters and it

becomes unmanageable, especially for harvesting, due to shortage of labour, time and budgetary provisions. The researchers are forced to go in for sample harvests purely based on operational conveniences without any scientific reasoning. It results in large errors. A scientific method of locating the sample plot for harvest in the main experimental plot considering the practical feasibility and operational convenience has been suggested. The sampling technique consists of locating a sample plot within the whole plot by selecting a row at random from the South-West corner of the main plot and a plant within the selected row at random along the length of the whole plot. This randomly selected plant would serve as the South-West corner for demarcating a sample plot of specified size. The dimensions of the sample plot are determined by first fixing its size, which is generally 10-15 percent of the whole plot, and then obtaining the number of consecutive rows to be included in the sample which is equal to one less than the integral part of the ratio of the square root of the sample plot size divided by row spacing. The number of rows multiplied with row spacing would provide the breadth of the plot, which is then used for computing the length of the sample plot. Starting from the South-West corner of the sample plot and moving along the length and breadth of the whole plot, the sample plot of calculated dimensions is demarcated. This method was tried on wheat crop of long-term fertilizer experiment that is in progress at IARI, New Delhi. Analysis of data collected from sample plots didn't reveal any significant difference between the sampled and whole plot yields.

From the analysis of data from AICCRP on LTFE, it has been observed that incorporation of FYM at the rate of 10-15 t/ha/year with recommended NPK fertilizer doses to the Kharif crop in the sequence has a pronounced effect in enhancing the efficiency of chemical fertilizers.

Long term fertilizer experiments are generally conducted on a fixed crop sequence with treatments as graded levels of fertilizers. Same design with the same randomization is used over years. The observations are collected from same experimental plots over years. As a consequence, the observations may be correlated. To account for the correlation among the observations over years, multivariate analysis of variance (MANOVA) is suggested for the analysis of data from these experiments. The comparison of treatments after MANOVA is a problem. To tackle this problem, a multivariate treatment contrast analysis procedure based on Wilk's Lambda criterion has been developed. This procedure will be of immense use in the analysis of data from AICCRP on LTFE.

**Crop sequence experiments**

A study on crop sequence in terms of their agronomic productivity, monetary returns and energy equivalents has been made in different agro-climatic zones of various states in the country. Data on each crop were analyzed by an appropriate analysis of variance technique after their conversion into protein, carbohydrate and calorie equivalents. For examining the consistency of the performance of a sequence over years, it was assumed that the sequence that has retained by and large, same rank in the individual year as well as combined over the years is considered consistent in its performance over years.

A statistical investigation on production, economic and energy potential of crop sequences in different agro-ecosystems has also been carried out. Statistical investigations have also been carried out to investigate the long-term effect of fertilizers on productivity of cereal crop sequences.

Some experiments have been conducted by PDCSR on Research Stations to assess the relative performance of the crop rotations of different periods in terms of crop productivity, yield stability and

soil health. Analytical procedures for such experiments have been developed and are illustrated with the help of examples. Statistical analysis of experiments on determining level and frequency of phosphorus application in different cropping systems has been carried out.

**Intercropping experiments**

Interpretation of intercropping data and its analysis presents considerable problems where the magnitude or even the existence of yield advantage over sole cropping is not immediately apparent. There are several analyses to be carried out on the basis of separate component crop-wise yields and combined yields. Usually a homoscedastic model is assumed for the analysis of such data. But this assumption may not be valid in intercropping experiments because the factors such as spacing, crop geometry and intercropping proportions, etc. may be responsible for the heteroscedasticity. So the presence of heteroscedasticity in intercropping experiments and its effects on drawing inferences on the performance of component crops both individually and collectively have been studied on the basis of eleven sets of data on Intercropping experiments, collected in the past, drawn from different places and/or years. It was observed that the heteroscedasticity is more prevalent in gram crop than in wheat crop. Also the effect of such unequal variances on the comparison of treatment effects separately on each component crop was observed. In case of presence of heteroscedasticity the relative values of the unequal variances associated with the different levels of a factor were estimated with the help of componentization of error mean squares. These ratios were utilized for drawing conclusions with the help of approximated distributions of the usual F-ratio under the homoscedastic model.

A strategy for analysis of data from replacement series experiments using experiments with mixtures methodologies has been suggested. A method of analysis of intercropping experiments involving sole crop treatments has been developed using Wilk's lamda criterion. A procedure of bivariate stability analysis of data has also been developed.

**Agroforestry experiments**

Methodological studies relating to agroforestry experiments were taken up. Various limitations in the data collection of agroforestry experiments have been identified. A review of designs and analytical techniques useful for agroforestry experiments has also been prepared. Non-linear models have been developed for the decline in crop yield over years in agroforestry experiments. A procedure based on land equivalent ratios has also been developed for analysis of data from agroforestry experiments. Analytical techniques based on linear nested models have also been suggested. The analytical techniques developed have been used for the analysis of data generated from agroforestry experiments on evaluation of fodder trees with and without crops under rainfed arable farming for semi-arid conditions conducted at Indian Grassland and Forage Research Institute, Jhansi.

**Variance components estimation**

The use of incomplete block designs has been restrictive because the experimenters did not have analysis of variance based method to estimate genetic parameters like genotypic and phenotypic variance and correlation from unbalanced data. A procedure of estimation of these parameters from a non-orthogonal design has been developed based on the analysis of variance of a dual design under a mixed effects model. The expressions of the estimates of the variances of the estimated genotypic and phenotypic variances have also been obtained. This technique has been used successfully in the research work on

Garden Pea of the Division of Vegetable Crops, IARI, New Delhi; Sunflower in Panjabrao Krishi Vidyapeeth, Akola and Pearlmillet in CCS Haryana Agricultural University, Hisar.

**Experiments with fixed quantity of inputs applied in splits**

Experiments with mixtures methodology is very useful strategy in designing and analyzing experiments involving fixed quantity of inputs applied in splits at different crop growth stages. Such experiments could involve a fixed quantity of nitrogen application at different crop growth stages; fixed quantity of nitrogen from different sources applied in splits; split doses of fixed quantity of irrigation, etc. This strategy has enabled the experimenters to conduct such experiments using more efficient designs and establish a relationship between various splits and response that can be used for interpolation of the response at the design points not tried in the mixtures (combinations). The experiments with mixtures methodology has also been used in the development of an analytical procedure for obtaining optimum proportion of area allocated to different crops in replacement series intercropping experiments. This methodology has also been used in experiments where more than one factors are involved, like fixed doses of nitrogen and irrigation.

**Food processing experiments**

Experiment with mixture methodology has been used for obtaining the optimum combination of different fruit juices in the quality evaluation of Ready to Serve fruit beverages. This technique has been used in the research work of the Division of Fruits and Horticultural Technology, IARI, New Delhi. Principal component analysis has been applied in developing an index of the quality of the processed food based on the hedonic scores on taste, colour, flavour, texture, overall acceptability, etc. This index has further been used for identification of the best treatment. This technique has been used in the research work of the Division of Agricultural Engineering and Division of Post Harvest Technology, IARI, New Delhi.

**Analysis of groups of experiments**

The method of combined analysis of data from experiments conducted in different environments (locations or years) using the concept of nested models has also been developed. The location/years have been assumed as bigger blocks and experiments *i.e.* replications are nested within the environments. A method of combined analysis of groups of experiments with some treatments common to all the experiments and some treatments common to some experiments has been developed. Analytical procedure for the combined analysis of experiments with factorial treatment structure with some treatment combinations common has also been developed.The procedure of combined analysis of data when the environment effect is random, has also been developed and used in the analysis of data from initial varietal trials conducted under the aegis of National Research Centre for Rapeseed and Mustard, Bharatpur. In case of cross over interactions, site regression biplot technique has been suggested for identification of entries for a subset of locations.

**Experiments with artificially created environments**

There are many multi-factor experiments wherein it is difficult to change the levels of one factor. Consider an experiment in which three levels of irrigation are complete submergence of the field, 3 days submergence and one day submergence. The other factors are the fertilizers with four doses of nitrogen and three doses of phosphorous. The experiment is conducted in three replications. In this experiment in

one block all the three levels of irrigation can not be experimented. However, with each level of irrigation the 12 combinations of nitrogen and phosphorous can be tried. The levels of irrigation can be treated as artificially created environments (with environment effects taken as fixed) and groups of experiments methodology can be applied for analysis of data. This concept led to valid statistical analysis of several experiments dealing with water management conducted at Water Technology Centre, IARI, New Delhi.

The procedure has also been used for the analysis of experimental data pertaining to pollination studies under different artificially environments (sowing dates) and population geometry (different planting ratios for male and female rows) in sunflower hybrid seed production. In case of artificially created environments it is suggested that the data on some auxiliary variables may also be collected. These auxiliary variables may be taken as covariates in the analysis.

To study the storage behaviour of vegetables / fruits, many post harvest experiments are conducted under different storage conditions. The data generated from these experiments have been analyzed using the storage temperatures as artificially created environments. Here, the artificially created environment effects have been assumed as fixed. This analytical technique has helped many M.Sc. and Ph.D. students of the Division of Vegetable Crops and Division of Agricultural Engineering, IARI, New Delhi to draw statistically valid inferences from their data. This concept of artificially created environments has applications in silviculture experiments as well.

**Energy requirement in agricultural sector**

A linear programming approach has been developed for estimating/projecting the energy requirement in agricultural sector. The approach uses the maximization of yield subject to the constraints on the availability of energy from different sources like Human Labour, Animal labour, Diesel, Electricity, Seed Rate, Farmyard Manure (FYM), Fertilizer, Chemicals, Machinery, Total Energy, etc. The procedure has also been used for minimization of total energy requirement for obtaining a given level of yield. The concept of energy use efficiency has also been introduced. This technique has been extensively used by the All India Co-ordinated Research Project on Energy Requirement in Agricultural Sector, Central Institute of Agricultural Engineering, Bhopal at all its 12 Centres around the country.

**Development of fatigue score on camels**

A fatigue score card based on qualitative and quantitative physiological characteristics of camel has been developed in collaboration with Krishi Vigyan Kendra, Rewari under the aegis of All India Co-ordinated Research Project on Utilization of Animal Energy. This fatigue score card is very useful for the farmers, military farms, etc. in deciding the tiredness of the camel. Suitable work cycles have also been obtained for various drafts.

**Farmers' participatory trials for resource conservation technology**

The Rice-Wheat Consortium (RWC) for the Indo-Gangetic plains generally evaluates different resource-conserving technologies (RCTs) in a farmers' participatory mode. It is a key challenge to analyze unbalanced data (due to variations in the number of replications of different RCTs, the use of different varieties, farmers' preferences for testing different options, etc.) generated from these trials using proper statistical tools so as to draw meaningful and valid conclusions. To achieve this, data collection and data preparation are of paramount importance. An additional problem with the analysis of data generated from these trials is that different researchers use different terms for similar practices. IASRI has reiterated the use of

uniform terminology for RCTs and experimental variables and a common data entry format, including specified units. The usefulness of the linear mixed effects model has also been emphasized to analyze the data generated from these trials by taking farmer effects or field effects as random and RCT effects (henceforth called treatments) as fixed. To identify the best performance of any treatment in a given environment, one can make all possible pair-wise treatment comparisons using adjusted means / best linear unbiased predictors of treatments. The procedure of analyzing groups of experiments has been used to study the interaction of treatments with crop varieties, years, soil types, and land leveling, etc. In the case of crossover interactions, the site regression biplot technique has been suggested to identify subsets of treatments to be recommended for specific environments. All the techniques developed / suggested have been illustrated with examples. The whole data generated from farmers' participatory trials for conservation agriculture conducted under the aegis of Asian Development Bank Project on Enhancing Farmers Income and Livelihoods through Integrated Crop and Resource Management in the Rice-Wheat System in South Asia at Bangladesh, Pakistan, Nepal, India (Modipuram and Balia) centres have been analyzed by Rice-Wheat Consortium for Indo-Gangetic plains using the techniques developed at IASRI in collaboration with CIMMYT and IRRI.

**Covariance analysis**

Analysis of covariance has been used in various experimental situations by taking residuals from previous season / year data, soil status, plant stand, etc. as covariates. Taking initial body weights of animals as covariate in grazing systems analysis, the analysis of covariance has been performed.

**Bayesian analytical techniques for experimental data**

Bayesian analytical technique for the analysis of experimental data generated through randomized complete block design has been developed both for conjugate and non-conjugate family or prior distributions. The procedures developed improved the precision on treatment comparisons in the analysis of experimental data from long term fertilizer experiments.

**Crop-weather studies**

A methodology was developed to study the behaviour of crop response to long-term fertilizer treatment with reference to weather and to examine the association between the responses of different crops of successive seasons. Investigations revealed that most of the variation in responses of rainfed (sorghum) as well as irrigated (wheat) crop, over a period of 13 years was due to weather.

Statistical aspects for characterization of drought in relation to crop with the main objective of working out drought threshold values related to a crop and to know the possibilities of occurrence of drought under various durations were studied. A method of quantifying drought threshold rainfall value for droughts of various durations such as 1$^{st}$ week through seventeen weeks for bajra, cotton and groundnut crops for one district each of Karnataka and Maharashtra was developed. Chances of occurrence of droughts of various durations at different stages of crop growth were also obtained. Out of these crops drought threshold values were the lowest for cotton crop in both the districts.

To study the effect of moisture stress on yields, an investigation was carried out with the objectives of examining reduction in yield and preparing stress index along with testing of the efficiency of the model when unit is changed from a day through fortnight.

**Spatial analysis of data**

A software for analyzing Spatial Variability and Interpolation (SVI) has been developed. It enables the user to visualize the spatial variability in graphical form (both 2D and 3D). It includes Variogram Specification Wizard and Graphical Visualization Tool. Variogram Specification Wizard guides the user about the steps necessary for specifying a variogram type viz. isotropic or anisotropic. Graphical Tool visualizes the calculated results in 2D (for isotropic) and 3D (for anisotropic) graphical form. Computer programs for obtaining variograms in different directions and at different lags have been incorporated in SVI. Programs for kriging in regular grids, with four inbuilt fitting models viz. spherical, Gaussian, exponential, and logarithmic have been developed.

Spatial variations in the analysis field trial data have been investigated using the procedure of variograms and REML. The procedures have been illustrated with the help of data pertaining to experiments conducted on Aonla, Kinno and Tuber crops.

The data pertaining to different parameters of soil (organic carbon, pH, electrical conductivity, etc.) from different locations and over years has been analysed using principal component analysis and variograms have been created to study the spatial and temporal variability in rice-wheat cropping system for assessment of spatial and temporal variation of soil microbial diversity. The methodology for spatial analysis of experimental data with interference effects has been developed and illustrated.

**Analysis of growth data**

The repeated measures data obtained from Jabalpur and Tirupati pertaining to growth of body weights of pigs for the period 1986 to 1990 have been analyzed using multivariate analysis of variance and profile analysis. The technique of profile analysis gives a detailed analysis of growth data. From profile analysis it is observed that there is no interaction between groups (male and female) and time points. Significant differences were observed among the time points but there is no difference among groups. The growth trend of pigs is studied by using the non-linear growth models.

**3.3 Collaborative research**

One of the major activities of Design of Experiments at IASRI since its inception has been working in close collaboration with its NARS partners. The collaborations with NARS partners dates back to 1956 when the Institute started the research work on planning, designing and analysis of experiments conducted under the All India Co-ordinated Agronomic Experiments Research Project (AICRP) of the ICAR. This project consisted of two components *viz.*, (a) complex experiments conducted at research centres (model agronomic experiments) and (b) simple fertiliser experiments conducted as experiments conducted on cultivators' fields. Today we work in close collaboration with large number of sister Institutes, non-governmental organizations, CGIAR organizations, etc. These are given as under:

**Planning, designing and analysis of data relating to**

- On Station and On Farm experiments under the Project Directorate for Cropping Systems Research {All India Co-ordinated Agronomic Experiments Research Project (1956-57 to 1982-83); AICRP on Cropping Systems Research and Experiments on Cultivators' Fields (1982-83 to 1988-89); Project Directorate of Cropping Systems Research (since 1988-89).
- All India Co-ordinated Research Project on Long-Term Fertilizer Experiments (since July 1985);

- All India Co-ordinated Research Project on Soil Test Crop Response Correlations (since March 2000)
- All India Co-ordinated Research Project on Rapeseed and Mustard (since April 2004)
- All India Co-ordinated Research Project on Wheat and Barley (since 2006)
- AICRP on Utilization of Animal Energy in collaboration with KVK, Rewari (01.12.1998 - 31.03.2000)
- AICRP on Energy Requirement in Agricultural Sector (07.04.2000-31.07.2001)
- Some investigations on design and analysis of agro-forestry experiments (IGFRI, Jhansi: 01.03.2000 – 31.08.2006)
- Assessment of spatial and temporal variation of soil microbial diversity in Rice-Wheat Cropping System with different management practices (IARI, New Delhi during June 2001 – May 2004)
- Institute of Applied Statistics and Development Studies, Lucknow (15.09.2001-14.12.2003)
- Design and analysis of farmer participatory research trials for conservation agriculture (Rice-Wheat Consortium for Indo-Gangetic Plains, CIMMYT and IRRI (since January 2005)
- Long term manurial and fertilizer experiment on potato based cropping systems (CPRS, Modipuram since January 2007).

**4. Advisory Services**

Despite the developments in design of experiments at global level in general and in NARS in particular, the status of experimentation is not at expected level. Most of the experiments are conducted in Randomnized Complete Block (RCB) designs / split plot designs. In most of these experiments, percent coefficient of variation is high and as a consequence precision on treatment comparison is low. Concerted efforts are being made to improve status of experimentation in the NARS through training programmes-cum-workshops; dissemination workshops; participation in annual workshops of AICRPs and rigorous advisory services. This has witnessed a revolution in the form of adoption of newer, efficient designs and sophisticated analytical techniques for statistical data analysis by agricultural researchers in their research endeavors. The efforts made have enabled the Institute to enrich the quality of agricultural research in the NARS. Through its advisory, IASRI has made its presence visibly felt in NARS and now experimenters look to IASRI for designing experiments and analysis of experimental data. The adoption of sophisticated analytical techniques for statistical data analysis has helped to improve the quality of agricultural research. Some of the adaptations are given below:

- Balanced incomplete block designs for identification of appropriate crop sequence with high productivity to suit specific needs of different agro-eco systems.
- Group divisible and extended group divisible designs, reinforced group divisible designs, reinforced balanced confounded factorial designs for crop sequence experiments and experiments for rainfed agriculture.
- Second order rotatable response surface designs with equispaced doses and experiments with mixtures for food processing and value addition experiments.
- Square and rectangular lattice designs, $\alpha$-designs, augmented designs, etc. for varietal trials and crop improvement programmes.
- Modern sophisticated analytical techniques such as experiments with mixture methodology, principal component analysis, cluster analysis, canonical correlation analysis, analysis of nested structures,

covariance analysis, etc. have also been adopted. The analytical techniques for the analysis of data from the experiments conducted to study the post harvest storage behaviour of the perishable commodities like fruits and vegetables are being widely used in NARS.

- A design for fitting response surface for the AICRP on Soil Test Crop Response Correlations that incorporates the effect of both the inorganic and organic fertilizers into the model has been developed. The experiment is a symmetrical factorial of the type 3×3×3 in 21 runs. QRT team of AICRP on STCR has approved the design. This design is being adopted by all the centres.
- $\alpha$-designs are being used at 22 co-operating centres of AICRP on Rapeseed and Mustard, National Bureau of Plant Genetic Resources, New Delhi and CSK HPKV, Palampur.

**5. Information Systems**

In the early 1950's an effort was initiated to develop a system of maintaining at a central place the information on all the experiments conducted in NARS. The objective was that for future experimentation it can provide useful information in choice of size and shape of plots, selection of treatments, etc. This effort was initiated in the form of a project on National Index of Agricultural Field Experiments. Ever since its initiation, this project has been continued with modifications from time to time. In 1960's and 1970's for the benefit of research workers, the data on experiments collected from different regional research centres were compiled and published in the form of compendia volumes so that the experimenters can use the information in their future experimentation. With the advent of fast speed computers in 1970's the data was stored on Magentic tapes. In 1980s this was converted into Information System. With the availability of network facilities, now we have following online information systems

- Agricultural Field Experiments Information System
- National Information System on Long Term Experiments
- National Information System on Animal Experiments

Agricultural Field Experiments Information System contains data on more than 27,000 experiments conducted in NARS. This system is available online. Data entry can be made online. User can analyze the data generated through basic designs online once the data is entered. This information system can be further strengthened by sensitizing the policy managers, researchers so as to make this more effective and useful. Additional data on experiments pertaining to cropping systems research, agricultural, animal, horticultural, fishery sciences, and wasteland experiments conducted in the NARS, along with value addition, will further enhance the usage of this information system. To ensure the continuous data flow, it should be made mandatory for the experimenters to put the data into the information system. A part of this can be achieved through ensuring the availability of general purpose software for all and having the customized modules for the analysis. Further, to utilize the information systems developed at IASRI, research projects are undertaken to convert the information into knowledge that can be fruitfully utilized by the subject matter divisions.

**6. Statistical Software Packages**

Without the availability of reliable, easy to use and cost effective software, it would be difficult to use the best statistical analytical techniques in research. Hence, development of indigenous, need based, user-friendly Graphic User Interface (GUI) based statistical software packages for experimental designs,

useful for NARS, requires urgent attention. A lead has been taken in the development and commercialization of indigenous statistical software packages like

- Statistical Package for Balanced Incomplete Block Designs (SPBD Release 1.0)
- Statistical Package for Factorial Experiments (SPFE 1.0)
- Statistical Package for Augmented Designs (SPAD)

for cataloguing and generating randomized layout of designs and analysis of experimental data. These packages have been developed and have been widely appreciated by both the statisticians and agricultural experimenters. A lot of requests over the globe have been received for SPAD. These indigenous softwares can be used for class room teaching and helps in improving the quality of teaching.

Some other indigenous statistical packages that have been developed or are being developed include Statistical Package for Repeated Measurement Designs; Statistical Package for Multivariate Analysis of variance; Statistical Package for Partially Balanced Incomplete Block Designs.

**7. Web Resources on Designed Experiments**

Although important and useful contributions have been made in Agricultural Statistics so as to help in improving the quality of agricultural research, yet it is impossible to reach every field of agricultural sciences in NARS. In view of this it is desirable to initiate an e-advisory service.

A beginning in this direction has already been made through initiating a Design Resources Server (www.iasri.res.in/design) with an objective to popularize and disseminate research in design of experiments among experimenters in agricultural sciences, biological sciences, animal sciences, social sciences and industry in planning and designing their experiments for making precise and valid inferences on the problems of their interest, generally treatment contrasts. It also provides support for analysis of data generated so as to meet the objectives of the study. This server also aims at providing a platform to the researchers in Design of Experiments for disseminating research and also strengthening research in newer emerging areas so as to meet the challenges of agricultural research. This server attempts to spread advances in theoretical, computational and statistical aspects of Design of Experiments among the mathematicians and statisticians in academia and among the practicing statisticians involved in advisory and consultancy services.

The Design Resources Server contains a lot of useful information for scientists of NARS. The material available on the server has been partitioned into four main components:

- **Useful for experimenters:** Electronic Books, online generation of randomized layout of designs, online analysis of data, analysis of data using various softwares
- **Useful for research statisticians:** Literature and catalogues of BBB designs, designs for making test treatments-control treatment comparisons, supersaturated designs, block designs with factorial treatment structure, online generation of Hadamard matrices, MOLS and orthogonal arrays
- **Other useful links:** Discussion Board, Ask a Question, Who-is-where, Important links
- **Site information:** Feedback, How to Quote Design Resources Server, Copyright, Disclaimer, Contact us and Site map

The Design Resources Server is like a mobile Library on Design of Experiments in particular and Statistics in general. It provides useful information both for active researchers in statistics as well as stake holders like scientists in NARS and others all over the globe. The server is dynamic in nature and new links on various topics are added to it regularly.

It is a copyright of IASRI(ICAR). For the period May 26, 2008 – May 25, 2009, Google Analytics gave 5786 page views, 4698 unique page views and usage through 566 cities across 80 countries in 6 continents.

Some other similar type of activities are: Web solutions for PBIB designs (http://www.iasri.res.in/webPBIB3/); Circular Designs (http://iasri.res.in/Circular_Designs/) and Lattice Designs (http://www.iasri.res.in/Lattic_designs/main.htm).

**8. Human Resource Development**

The Division has played a key role in human resource development. Besides teaching the M.Sc./Ph.D. Students of P.G. School, I.A.R.I., New Delhi, Division takes active part in organization of summer/winter schools on Advances in Design of Experiments and related topics for Statisticians and experimenters; Training programmes under Centre of Advanced Studies in Agricultural Statistics and Computer Applications. In fact the training programmes that are being organized to bring agricultural scientists and statisticians together so as to derive the maximum academic advantage through interaction with the faculty and among the fellow participants are being widely appreciated in NARS. The Division is also conducting sponsored training programmes for scientific personnel of Indian Council for Forestry Research and Education (ICFRE); Central Statistical organization, E.I. DuPont India Pvt. Ltd. Customized and tailor made training programmes for research personnel of AICRP on Energy Requirement in Agricultural Sector; AICRP on Rapeseed and Mustard and Rice-Wheat Consortium for Indo-gangetic plains have also been conducted.

**9. Future Perspective**

The research in the area of Design of Experiments has been of high quality and maintaining the tradition, the research programmes would be further strengthened to meet the challenging requirements in the newer emerging areas of agricultural research. Some important areas on which the attention would be focused are described in the sequel.

Research in newer emerging areas would be the priority. Research on designs for factorial experiments with scarce resources, mixed orthogonal arrays and especially mixed orthogonal arrays of strength one would be undertaken. Block designs having orthogonal factorial structure with balance having full efficiency on main effects and controlled efficiency on interactions would also receive attention. With proper modifications these would have applications in micro array experiments, computer experiments, designs for cropping systems research, etc. Efficient designs for multi-response experiments useful in food processing, intercropping and agroforestry experiments would be obtained. Integrated farming systems research is another emerging area that would be taken up. Data requirements for such experiments would be studied and the problems of data designing, which includes the experimental design, the choice of treatments, the inference problem and the analysis of data to be performed, etc. would be studied. Designing and analysis of experiments for identification of varieties resistant to biotic and/ or abiotic stresses would

also be the research priority. Designing experiments and analytical techniques for experiments involving very few treatments (may be only two) but very large plots (may be one hectare or more) would be developed. Efficient and robust designs for animal experiments, horticultural experiments and fisheries experiments would be generated. Bayesian designing of experiments and Bayesian analysis of data would be another priority area of research. Efficient designs for bio-equivalence trials and for biological assays would be obtained. Designing geo-spatial experiments and analysis of geo-spatial data would become another priority of research. Exploiting the information technology, efforts would be made to generate computer aided efficient designs for different experimental settings. The activity of developing indigenous software for generation of randomized layout of the design and then subsequent analysis of data including contrast analysis, groups of experiments, etc. would be reinvigorated.

Concerted efforts would be made to strengthen linkages with the All India Coordinated Research Projects for planning and designing of experiments and analysis of experimental data. These linkages would also be extended to subject matter divisions, State Agricultural Universities (SAUs) and National Research Centres (NRCs). The possibility of undertaking collaborative projects with CGIAR organizations in Consortium Mode would be explored.

For dissemination of research and for capacity building of scientists, need based, customized, tailor made training programmes would be organized. These would be supplemented with travel seminars, workshops, etc. The consultancy and advisory services would be strengthened so as to provide leadership in the planning and designing of experiments and analysis of experimental data to the students and scientists in NARS. This would be possible through rigorous use of Design Resources Server and by further strengthening it for providing e-learning and e-advisory services.

The above steps will definitely help in strengthening the basic, applied and adaptive research as well as teaching and training in Design of Experiments.

# Glimpses of Basic Research in Design of Experiments at IASRI

**Rajender Parsad, Seema Jaggi, V.K. Gupta, Cini Varghese and Krishan Lal**

## 1. Introduction

The basic premise of the creation of a statistical section was to provide advisory on designing agricultural experiments to the state departments of agriculture and animal husbandry, to help them in analysis of data and interpretation of results. In the process of advisory, a need was felt to develop theories and designs and this laid the foundation of basic research in Design of Experiments with applications in solving complicated problems of agricultural research to bridge the gaps in the existing knowledge. Ever since then there has been a focused attention on basic research in design of experiments. The Institute has maintained pace with the developments at international level and for this reason the school of Design of Experiments at the Institute occupies a place of pride nationally and internationally. The research work carried out in Design of Experiments has been published in the International and National Journals of repute with high impact factors. As a consequence of important contributions made in Design of Experiments, several text books and monographs have been brought out which have a widespread citation.

The research work on theoretical developments in Design of Experiments was initiated during the decade of 1950. Many important contributions have been made in developing new methods of construction and analysis of experiments; in particular in finding suitable confounded designs for asymmetrical (or, mixed) factorials. The decade of 1960 saw many major contributions in designs for factorial experiments, rotatable and other response surface designs and designs for experiments with mixtures. The intense research activity in the area of design and analysis of experiments has continued rigorously throughout. The major emphasis has been obtaining optimal / efficient and robust designs for agricultural and allied sciences.

Some of the areas which received special attention included block designs (efficiency- and partially efficiency- balanced designs); weighing designs, crossover designs, orthogonal arrays and their application in obtaining fractional factorial plans, optimal block and row-column designs for making all the possible pair wise treatment comparisons, robust designs against the disturbances like loss of data, interchange and exchange of treatments, outliers, etc. Use of combinatorial designs in efficient survey sampling plans with unequal probabilities leading thereby to controlled selection was another major contribution. The use of mixed orthogonal arrays as Balanced Repeated Replications in the variance estimation of a non-linear statistic from a large scale complex survey data was an original contribution of the Institute.

More recently, the institute has undertaken basic research in optimality aspects of designs for test treatments versus control treatments comparisons. The comparisons are made with differential precision. The optimality aspects of block designs with nested blocks and nested rows and columns have been extensively studied. Similarly, the optimality aspects of designs with correlated observations, presence of trend in the observations within a block, designs for diallel and double cross experiments, designs for bioassays, designs for micro-array experiments, designs for crop sequence experiments, designs for estimation of competition effects from neighboring units, designs for two-way heterogeneity setting and

a connection between these designs and block designs with two non-interacting sets of treatments, etc. are some of the major basic research contributions that have been accepted at an international level. We begin with the concepts of optimality and robustness in designed experiments. In Section 3 we give the developments in the theory of design of experiments concerning single factor experiments. Section 4 is devoted to multi-factor experiments and other important considerations are given in Section 5.

## 2. Optimality and Robustness of Designs

In the initial stages of development of experimental designs, emphasis was laid on constructing designs that were in some sense as symmetric as possible in their treatment of the statistical parameters of interest (for example, randomized complete block designs, balanced incomplete block designs, Latin square designs, Youden square designs, etc.). Such designs yielded "information matrices" (coefficient matrices of the reduced normal equations for estimation of treatment effects) that, especially in the pre-computer age, made statistical calculations easy; the designs had aesthetic appeal to mathematicians and often had algebraic or geometric representations that helped one to understand and construct them; and the symmetric treatment of parameters of interest seemed a reasonable property that made such designs yield statistical estimators that looked intuitively as accurate as possible for the given number of observations.

The advent of high speed computers; the realization of the importance and need to choose and adopt an experimental design that is best according to some well defined statistical criterion, led to the development of a subject like optimality of designs. The theory of optimal designs was almost non-existent till about the end of the Second World War, except for a remarkable early paper by Smith and the important paper of Wald. Professor Jack K. Kiefer initiated the serious and rigorous work on optimality aspects of designs. For the given experimental situation and an inference problem, there are a number of designs available, called a class of designs, which can be used to achieve some specified set of objectives. The choice of an appropriate design for a particular setting depends upon (a) the inference problem whose answer is sought for; (b) the class of designs in which choice is to be made (the class should have more than one design); and (c) the criterion or criteria to be used for the selection of the design. To be specific,

*A design d* is said to be $\psi$-optimal for a given inference problem on $\eta$, a set of parametric function, in a class of competing designs* ***D*** *with some specified parameters, if d* minimizes $\psi$ [information matrix of best linear unbiased estimators (BLUE) of $\eta$] over* ***D****. The function $\psi$ satisfies some pre-defined properties. It could be average variance of BLUE of $\eta$, generalized variance of BLUE of $\eta$, maximum variance of the elements of BLUE of $\eta$, etc.*

Some commonly used optimality criteria in the literature are A-optimality, D-optimality, E-optimality, MV-optimality, S-optimality, MS optimality, $\phi_p$ - optimality, universal optimality, DS-optimality, G-optimality, etc. The significant results obtained on the optimality aspects of designs are given in the concerned sections.

Optimal design theory has been developed and studied under what may be called as ideal conditions. Accidents and disturbances do occur even in a well-planned experiment and they render even an optimal design perform poorly and results in loss in efficiency. Various disturbances that commonly occur are – (i) missing observations(s), (ii) exchange of a treatment, (iii) interchange of a pair of treatments, (iv) presence of outlier(s), (v) presence of a systematic trend in the experimental units, (vi) model

inadequacies, etc. In the presence of one or more of these disturbances even an optimal design may be poor. This realization is the motivation to the development of designs that are robust against one or more of the above disturbances. A design *d* is said to be robust against one or more disturbances if it remains insensitive to the presence of one or more disturbances in terms of design properties. To be specific, *consider a design d having some property "A" is subjected to some accident or disturbance. If the resulting design* $d^*$ *obtained after the accident also possesses property "A", then the design is said to be robust against the accident.* The property *A* could be connectedness; variance balance; optimality; orthogonal factorial structure; efficiency, etc. The results obtained on the robustness aspects of designs are given in section 5.6.

## 3. Single Factor Experiments

In this section we describe the results on optimality aspects of designs for unstructured treatments also termed as varietal designs or single factor designs.

### 3.1 Block designs

We begin with the simplest experimental setting where there is only one blocking system. The inference problem being addressed is that of complete set of orthonormalized treatment contrasts or all possible pair wise treatment comparisons.

**Incomplete block designs for all pair wise treatment comparisons**

Balance in incomplete block (BIB) designs is a desirable statistical property. In literature, the term balance has been used in several senses viz., variance balance, efficiency balance, generalized efficiency balance and pair wise balance. While variance balance, efficiency balance and generalized efficiency balance are based on statistical properties, pair wise balance is a combinatorial property. A general method of construction of binary variance balanced block designs with unequal block sizes under a heteroscedastic model has been obtained. The method begins with a pair wise balanced design which is very easy to obtain and then converts it into a variance balanced design. It has been shown that the existence of a pair wise balanced design implies the existence of a binary variance-balanced design. All the designs hitherto known in the literature can be obtained from this general method of construction. An exhaustive catalogue of binary balanced block designs has been prepared for average replication smaller than 31 along with their resolvability status. These designs are universally optimal over a wide class of designs. The catalogue also includes the binary balanced block designs under the heteroscedastic model where intra block variances are assumed to be proportional to non-negative real power of block sizes. The resolvability status of these designs is also indicated. Several new methods of construction of *n*-ary variance balanced block designs have been given.

Partially Balanced Incomplete Block (PBIB) designs are an important class of incomplete block designs that can be used to economize on the use of experimental material. Some results on non-existence of PBIB designs with *m*-associate classes have been obtained by exploiting the complementary property of block designs. Some new group divisible and nested group divisible designs have been identified. Some new 3-class association schemes along with method of construction of PBIB designs based on these schemes have been obtained. A catalogue containing parameters of PBIB (3) designs has been prepared and made available on web.

The concept of partially efficiency balanced designs and simple partially efficiency balanced designs were introduced. In simple partially efficiency balanced designs, some contrasts are estimated with full efficiency.

Optimality aspects of block designs (universal optimality, E-optimality, etc.) have been studied under the usual fixed effects and mixed effects models.

**Resolvable block designs**

From practical considerations of the experimenters, the use of incomplete block designs becomes restrictive because all the treatments do not appear in adjacent piece of land and, therefore, the demonstration of variety effect / treatment effect in the field cannot be made. To overcome this problem, it is recommended that resolvable block designs with smaller block sizes may be used. A resolvable block design is a design in which the blocks can be grouped in such a way that every treatment appears once in each group; in other words, each group is a complete replication. Resolvable block designs have been studied by various authors in the literature. Resolvable solutions of BIB designs and PBIB designs are available. A simple class of resolvable block designs is the Lattice designs (square lattice, rectangular lattice, cubic lattice and circular lattice), but these solutions are available for a limited number of varieties and block sizes. For the benefit of the experimenters, a comprehensive catalogue of Lattice designs has been prepared. These designs have been compiled and brought out in the form of a technical bulletin for a ready reference of experimenters. The online generation of these designs is possible from IASRI web page. Catalogues of resolvable BIB designs and binary balanced block designs have also been prepared.

In agricultural experiments, the number of varieties is not fixed and the experimenter has to choose number of replications as per experimental resources. BIB designs, PBIB designs, cyclic designs, etc. may not meet the requirements of the experimenters as these designs would require many replications. Similarly the use of Lattice designs also becomes restrictive because of the restriction on number of treatments. α-designs are another class of resolvable incomplete block designs with number of treatments being a composite number. These designs can be generated in two and three replications also. The block size of these designs is a factor of number of treatments. For the benefit of the experimenters, a comprehensive catalogue of efficient α-designs for $6 \le v(=sk) \le 150$, $2 \le r \le 5$, $3 \le k \le 10$ and $2 \le s \le 15$ has been prepared along with lower bounds to A- and D- efficiencies and generating arrays. Here $v$ is the number of treatments, $r$ is the number of replications and $k$ is the block size. Several designs obtained are more efficient than the corresponding designs available in the literature. A module for computer aided generation of α-designs has also been prepared. A randomized layout of these designs can be generated through Design Resources Server (http://www.iasri.res.in/design/Alpha/Home.htm).

**Minimally connected designs with extra observations**

For cost effective utilization of experimental resources, it is desired that the experiment be run in minimum possible number of experimental units. To ensure that all possible pair wise treatment comparisons are possible in a block design, the minimum number of experimental units required is equal to one less than the sum of the number of blocks and number of treatments. A design in minimal number of experimental units that permits the estimation of all possible pair wise treatment comparisons is called a minimally connected design. The basic problem with the use of minimally connected designs in agricultural experimentation is that they do not provide an estimate of error. Therefore, to get an estimate

of error, some modifications in these designs need to be made by adding some more experimental units. Keeping this in view, a catalogue of block designs with $n = v + b - 1 + i$, $i = 1, 2, 3, 4, 5, 6, 7, 8$ has been prepared, where $n$ is the total number of experimental units. Block contents along with lower bounds to A- and D-efficiencies are also given in the catalogue.

**Design and analysis of multi-response experiments**

Experiments in which data on several responses are measured from an experimental unit corresponding to the application of a treatment are known as multi-response experiments. Multi-response experiments are of two types viz. complete multi-response experiments (all the response variables are recorded from each experimental unit) and incomplete multi-response experiments (recording of all the response variables from each experimental unit is not feasible). For complete multi-response experiments, it has been shown that the designs that are efficient for single response experiments are also efficient for complete multi-response experiments provided the dimensionality of the response vector is smaller than the error degrees of freedom. A stepwise procedure for the multivariate analysis of variance of data generated through block designs for complete multi-response experiments has been developed. To identify the sets of treatments that are significantly different from others, expressions of sum of squares and cross products matrix due to desired treatment contrasts for complete multi-response experiments conducted using block designs have been derived. If two treatments are found to be significantly different in the pair wise treatment comparisons, then we need to identify the better treatment from the two significantly different treatments. A procedure based on J-plot and Euclidean distance of the treatment means has been developed for identification of best treatment in complete multi-response situations.

A method of construction of designs for incomplete multi-response experiments using combination of randomized complete block (RCB) design as treatment-wise design and BIB design as response-wise design has been obtained. The designs obtainable from this method are economical from resource point of view. Stepwise analytical procedure for analysis of data generated from incomplete multi-response designs obtained through above method has also been developed.

**Outliers in block designs**

An outlier in a set of data is an observation (or an observation vector) that appears to be inconsistent with the remainder of the observations in that data set. Occurrence of outlier(s) is common in every field in which data collection is involved. In agricultural experiments, outlier(s) is/are likely to appear in the experimental data due to disease and/or insect-pest attack on some plots in the field, or due to unintentional heavy irrigation on some particular block(s) or plot(s) of the experiment. Outlier(s) may creep in due to transcription errors. Presence of such abnormally high or low observations may cause a deviation from the assumptions particularly those of normality and homogeneity of observations. It is, therefore, important to detect the presence of outlier(s) along with deviations from these assumptions and suggest remedial measures.

Test statistics have been developed for identification of outlier(s) that takes care of the masking effect and also enables one to detect outliers. Robust methods of analysis based on M-estimation and Least Median Squares (LMS) method have been developed. Designs that are robust in presence of one or two outliers have been identified. Graphic user interface based software has been developed for analyzing experimental data in the presence of outlying observations. The software has the following features: can

identify outliers in experimental data; can directly apply the robust methods of estimation viz. M-estimation (Huber's function) or LMS method for analyzing the data and has option to analyze the data after deleting the outlying observations.

A test statistic for detection of a single outlier vector in complete multi-response experiments run in a block design has been developed on the lines of Cook-statistic for single response variable in a block design.

**Block designs for test treatments - control(s) comparisons**

In practice there may arise experimental situations where it is desired to compare several treatments called test treatments to standard treatments called controls. For example, in screening experiments or in the beginning of a long-term experimental investigation, it is initially desired to determine the relative performance of new test treatments with respect to the existing treatments. Interest may be to study the performance of new types of hybrid varieties, method of cultivation, pesticides, herbicides, etc. and an existing (old) one is to be replaced by one of these newer kinds. The main interest is to compare the new (test) treatments with the old (control) treatment(s) and thus a higher precision is desired for these estimates. For these experimental situations, the usual designs for making all possible pair wise comparisons among treatments may not be efficient. Series of block designs with equal/unequal block sizes for test treatments versus control(s) comparisons have been developed. The problem of making test treatments versus control comparisons when the block effects are random has been investigated. Efficient designs for the experimental settings, (i) where the test versus test comparisons are made with less precision but test versus control comparisons are made with high precision and (ii) when there are several controls and each control has a different importance have been developed. A catalogue of efficient designs is also prepared.

In many experimental situations, for example, in germplasm evaluation trials the genetic material collected from the exploration trips is very limited and the quantity of the seed is not sufficient to have more than one replication of the test treatments. Such experiments are generally conducted using augmented randomized complete block designs. Here, the experimenters would often like to know how many times the control treatments be replicated in each of the blocks so as to maximize the efficiency per observation for making test treatments versus control treatments(s) comparisons. An expression has been obtained for optimum number of replications of the control treatments in an augmented randomized complete block design that maximizes the efficiency per observation. An algorithm for generation of these designs that maximize the efficiency per observation has been developed.

**3.2 Row-column designs**

When the heterogeneity in the experimental material is present in two directions then row-column designs such as Latin square designs, Youden square designs, generalized Youden designs, Pseudo Youden designs, Youden type designs, etc. are useful. It has been shown that a generalized Youden design under non-regular setting can never be universally optimal. Universal Optimality, E-optimality and Type 1 and Type 2 optimality aspects of row-column designs have been studied. Optimality aspects of structurally incomplete row-column designs with unequal row and column sizes have been investigated. Sufficient conditions for the designs to be simultaneously universally optimal for treatment and column classification have been obtained. General methods for construction of universally optimal structurally incomplete row- column designs have been obtained. A correspondence between structurally incomplete row-column designs, and designs for two stage experiments has been established. This correspondence will go a long

way in unifying the research efforts that were made in the literature in two different directions viz. row-column designs and block designs for two disjoint sets of treatments applied in stages or succession. The designs identified are applicable to experimental situations where treatments are to be applied in sequences, particularly in crop sequence, horticultural and animal science experiments.

Semi-Latin squares are useful for situations when the number of levels of both the nuisance factors (factors of heterogeneity) is same and is a factor of number of treatments. Such experimental situations are common in consumer testing, glass house crops, residual effect experiments, sugar beet trials, food industry, etc. A new method of construction of (n×n)/2 semi-Latin squares has been obtained.

For comparing several test treatments with control treatment(s), efficient row-column designs and structurally incomplete row column designs have been obtained.

### 3.3 Block designs with nested factors

#### Nested block designs

A nested block design is defined as a design with two systems of blocks such that the second system of blocks is nested within the first system of blocks. These designs are quite useful in many experimental situations. For example, consider a field experiment conducted using a block design and harvesting is done block wise. Harvested samples are to be analyzed for their contents either by different technicians at same time or by a technician over different periods of time. The variation due to technicians or time periods may be controlled by another blocking system. Technicians or time periods form a system of blocks that are nested within blocks. Such experimental situations are also common in post harvest value addition of horticultural and vegetable crops. Nested block designs are also quite useful in agricultural field experiments where the plots with similar fertility occur in patches rather than in a uniform direction. Nested balanced incomplete block (NBIB) designs are useful for these situations. In a NBIB design, block classification ignoring sub-blocks is a BIB design and sub-block classification ignoring blocks is also a BIB design. A complete catalogue of NBIB designs with number of replications $r \leq 30$ has been prepared. The catalogue contains a total of 299 designs. Out of 299 designs, 8 designs are non-existent. A new method of construction of NBIB designs has been obtained. Using this method and trial and error solutions, block layouts of 22 new NBIB designs have been obtained. The layout of 199 designs with block contents has been prepared. The solution for the block layout of remaining 92 designs is unknown and the statisticians need to develop methods of construction of these NBIB designs. The designs catalogued have also been identified for 1-resolvable and 2-resolvable sets.

A NBIB design may not exist for all parametric combinations or even if it exists may require a large number of replications, which the experimenter may not be able to afford. To deal with such situations, nested partially balanced incomplete block (NPBIB) designs have been introduced in the literature. Some new methods of construction of NPBIB designs based on group divisible association scheme have been obtaned. Catalogues of NPBIB designs based on 2- associate and 3- associate class association schemes have been prepared. Universal optimality aspects of non-proper block designs with nested rows and columns have been studied. Some general methods of construction of optimal designs under this setup have also been given.

A β-version of the software for generation of efficient nested block designs has been developed. It consists of 3 modules viz. (i) nested block designs with independent observations; (ii) nested block designs

when the observations within a sub-block have a nearest neighbour correlation structure and (iii) nested block designs when the observations within a sub-block have a nearest autoregressive correlation structure. Once the number of treatments ($v$), number of bigger blocks ($b$), bigger block size ($k$), number of sub-blocks nested within a block ($q$) and the value of correlation coefficient ($r$) is entered (in case of correlated observation), the design gets generated.

**Nested block designs for making test treatments-control treatment comparisons**

Nested block (NBIB and NPBIB) designs are useful for experimental situations where the experimenter is interested in making all possible pair wise treatment comparisons with as high a precision as possible. However, there do occur experimental situations where the experimenter is interested in comparing several new treatments with existing practice with high precision and the comparisons among the test treatments are not of much importance. To deal with such situations, nested balanced treatment incomplete block (NBTIB) designs have been introduced. Some new methods of construction of NBTIB designs making use of NBIB designs, initial block solutions, etc. have been developed. A method of construction of nested block designs for making test treatments-control treatment comparisons has been developed which yields minimally connected designs with respect to sub-blocks. The design with respect to bigger blocks is a group divisible treatment design.

**Doubly nested block designs**

Nested block designs are useful for experimental situations with two sources of variation in the experimental units, one nested within another. There may, however, arise a situation when there exists another source of variation among the units in sub-blocks of an incomplete block design. Consider a field experiment which is conducted at several locations using a nested block design where locations are taken as blocks and harvesting time is taken as sub-blocks. But harvested samples from each sub-block are to be analyzed for their content in laboratory by different technicians. To control variation due to technicians, this may be taken as another blocking factor. Hence, nesting of units within sub-blocks may be required. Doubly nested incomplete block designs are useful for such situations. An exhaustive catalogue of doubly nested balanced incomplete block (DNBIB) designs for number of treatments $v \leq 20$ and $r \leq 20$ is available in the literature. A DNBIB design may not exist for all parametric combinations or even if it exists, may require large number of replications which the experimenter may not be able to afford. To deal with such situations, a new class of designs called doubly nested partially balanced incomplete block (DNPBIB) designs has been introduced. Some general methods for construction of DNPBIB designs are obtained using DNBIB designs, NBIB designs and PBIB designs.

**Block designs with nested rows and columns**

For the experimental situations where there are two cross-classified factors causing heterogeneity in the experimental material and are nested within the blocking factor, block designs with nested rows and columns have been introduced. Several methods of construction of designs with nested rows-columns (BIB-RC designs) have been developed and a comprehensive catalogue of BIB-RC designs has also been prepared. Universal optimality aspects of non-proper block designs with nested rows and columns have been investigated. Some general methods of construction of optimal block designs with nested rows and columns have been obtained and a catalogue of universally optimal designs has been prepared. Some methods of construction of block designs with nested rows and columns that are optimal according to Type 1 criterion have also been obtained.

Efficient block designs with nested rows and columns for making test treatment-control treatment(s) comparisons have also been obtained.

### 3.4 Computer aided search of efficient designs

In the literature on design of experiments, methods of constructing optimal designs are given for various experimental settings. However, these methods do not give flexibility to obtain an optimal/efficient design for any desired parametric combination in a given experimental situation. This limitation can be overcome by using intelligent computer algorithm. Computer algorithms for generation of efficient designs based on exchange and interchange procedures have been developed. The procedure of computing lower bounds to A- and D-efficiencies has been incorporated in the algorithms. Several new efficient designs for making all possible pairwise treatment comparisons have been obtained. Efficient designs for the parametric combinations for which BIB designs are non-existent or solutions are unknown have also been obtained. Efficient designs for making test treatments-control treatment(s) comparisons with differential precision have been obtained through the computer aided search. The computer algorithms have also been modified and implemented for obtaining efficient designs under correlated error structures for (i) block designs for making all possible pairwise treatment comparisons; (ii) block designs for making test treatments-control treatment comparisons; (iii) nested block designs and (iv) cross over designs. Algorithms for construction of efficient $\alpha$-designs have also been developed using exchange and interchange procedures.

### 3.5 Block designs for correlated observations

One of the requirements in the analysis of data from designed experiments is that the observations should be independent. But, there may arise situation when the observations are correlated. In agricultural field experiments, usually the fertility levels are changing over space. Plots occurring close together within a field area are more similar than plots occurring far away from each other. Thus, the observations from the neighbouring plots are likely to be correlated. The observations may be correlated because of the layout of the plots, pest infections from neighbouring plots or some local factors which blocking cannot remove. In industrial experiments, suppose only a certain number of experimental runs can be made during a particular time period such as a plant shift or a day. Here, time contiguity is used as criterion of blocking and there may exist correlation within the time block caused by the equipment ageing and wear out. In this situation, it is desirable to obtain designs and perform analysis in which the correlations are taken into account. Catalogues of efficient block designs and nested block designs for making all possible pair-wise treatment camparisons under nearest neigbour and auto correlated error structures have been prepared. A catologue of efficient block designs for making test treatments-control treatment comparisons and change over designs under autoregressive correlated error stuructures has been prepared. Series of pair-wise uniform incomplete block designs under this situation have been obtained.

### 3.6 Block designs with neighbour effects

In agricultural field experiments, it may so happen that the treatment applied to one plot in a block may affect the response on the neighbouring plots if the blocks are formed using adjacent plots with no gaps. Neighbour balanced block (NBB) designs have been defined to ensure that no treatment is unduly disadvantaged by its neighbour. Some series of balanced/ partially balanced and complete/

incomplete NBB designs have been obtained. Optimality aspects of block designs with neighbour effects under fixed and mixed effects model have been studied. Some series of NBB designs for comparing a set of test treatments to a control have been developed. The designs obtained are totally balanced in the sense that all the contrasts among test treatments for direct, left and right effects are estimated with same variance and all the contrasts pertaining to test versus control for direct, left and right effects are estimated with the same variance. The concept of NBB designs has been defined for the experimental situation where the treatments are the combinations of levels of two factors and only one of the factors exhibits neighbour effect. Some methods of constructing complete NBB designs for two factors (useful in agroforestry experimets) neighbour balanced for one factor have been obtained. These designs are variance balanced for estimating the direct effects of contrasts pertaining to combinations of levels of both the factors. An incomplete NBB design for two factors has also been given which is found to be partially variance balanced with three associate classes. Several methods of construction of serial designs have been obtained and the analytical procedures have been developed using the linear model accommodating the left and right neighbour effects on each treated plot. Several methods of construction of complete as well as incomplete block designs have been obtained by making use of the directed graphs and factorial structure. An analytical procedure for the analysis of these designs under the assumption of (i) equal left and right neighbour effects and (ii) no competition effect of a treatment with itself has also been developed. The performance of a series of incomplete NBB design for autoregressive [AR(1)] and nearest neighbour (NN) error correlation structure is studied when generalized least squares estimation is used. The designs with AR(1) structure turns out to be more efficient and the efficiency of direct effects of treatments is more as compared to neighbour effects under both the structures.

## 4. Multi-factor Experiments

The theoretical developments in experimental designs at IASRI began with obtaining confounded designs for asymmetrical factorial experiments in 1950's. Since then a lot of developments have taken place in multi-factor experiments. Some of these are described in the sequel.

### 4.1 Confounded designs for asymmetrical factorial experiments

Several efficient confounded designs for asymmetrical factorial experiments have been obtained using the technique of collapsing and replacement and a catalogue of efficient designs have been prepared.

Balanced confounded factorial experiments for symmetrical as well as asymmetrical factorial experiments, providing flexibility in the choice of factors and their levels have been constructed. Procedures of analyzing these designs have also been suggested.

### 4.2 Designs for crop sequence experiments

Equivalence between extended group divisible (EGD) designs and designs for crop sequence experiments has been established. This equivalence has encouraged the agricultural experimenters in the use of EGD designs for their experimentation. This equivalence is also useful for the experimental situations where the blocking criteria are implemented in succession and there is likelihood of the interaction effect between the direct effects and residual effects of the treatments applied at two successive stages. Some new methods of construction of EGD designs have been obtained and a catalogue of these designs with a maximum of 5 replications has also been prepared.

### 4.3 Fractional factorial plans

For screening experiments or when the higher order interactions are negligible or the experimental resources are not enough for one complete replication, fractional factorial plans and particularly orthogonal resolution plans have been obtained. A new method of construction of confounded designs for fractional asymmetrical factorial experiments has been developed. This method is quite general and is applicable to the symmetrical factorial experiments as well. Some series of resolution III plans for asymmetrical factorial exprements have been constructed. Orthogonal resolution IV plans for asymmetrical factorials have been obtained. Two series of designs for $3\times2^n$ and $5\times2^n$ experiments are constructed. General class of main-effect plans for $3^m\times2^n$ experiments has been proposed. These plans are irregular fractions, and have been constructed with the help of BIB designs. They are generally non-orthogonal in nature. A complete catalogue of orthogonal main effect plans has been prepared. A new exchange algorithm (EA) for construction of D-optimal designs was developed. This EA uses D-optimality criterion and can be applied in the settings of fractional factorial plans). This EA is used to construct 54 D-optimal $2^m$ fractional factorial plans of resolution V for $m$ = 4, 5 and 6. D-optimal saturated main effect plans for three factors when the first factor is at three levels have been obtained using a computer algorithm.

Several methods of construction of symmetric and asymmetric orthogonal arrays have been obtained. Resolvable orthogonal arrays have been used in construction of fractional factorial plans. Some methods of obtaining resolvable (symmetric) orthogonal arrays and resolvable mixed orthogonal arrays have been given. The methods are derived from the use of Kronecker Product and Kronecker sum of matrices.

### 4.4 Supersaturated designs

In industrial, biological and agricultural experiments, there occur experimental situations where a large number of factors are to be tested but only few of the factors are active. In such experiments, the experimenter's endeavour is to minimize the number of runs to identify the active factors for efficient utilization of resources and minimization of cost and time. Supersaturated design (SSD) is a fractional factorial plan whose run size is not large enough for estimating the mean and all the main effects. The main attraction for using SSDs is their run size economy. SSDs for two-level, multi-level and mixed level factorial experiments have been obtained using resolvable orthogonal arrays; Hadamard matrices and proportional frequency plans. Some criteria for comparing SSDs for asymmetrical factorial experiments and/or unbalanced designs are given. A column-wise co-ordinate exchange algorithm for generation of SSDs for two-level, multi-level and mixed level factorials has also been developed.

### 4.5 Response surface designs

The above discussion relates to the experiments in which the levels or level combinations of one or more factors are treatments and the data generated from these experiments are normally analyzed to compare the level effects of the factors and also their interactions. Though such investigations are useful to have objective assessment of the effects of the levels actually tried in the experiment, this seems to be inadequate, especially when the factors are quantitative in nature and cannot throw much light on the possible effect(s) of the intervening levels or their combinations. In such situations, it is more realistic and informative to carry out investigations with the twin purposes:

a) to determine and to quantify the relationship between the response and the settings of a set of experimental factors, and

b) to find the settings of the experimental factor(s) that produces the best value or the best set of values of the response(s).

If all the factors are quantitative in nature, it is natural to think the response as a function of the factor levels and data from quantitative factorial experiments can be used to fit the response surface over the region of interest. The special class of designed experiments for fitting response surfaces is called response surface designs. Second order rotatable response surface designs (SORD) have been obtained through BIB designs. The concept of group divisible rotatable response surface designs has been introduced. Second order rotatable and group divisible rotatable response surface designs have been obtained for 4 level and 6-levels factorial experiments. The rotatable designs for asymmetric factorial experiments have been introduced using the transformation of levels of rotatable designs for symmetrical response surface designs.

Method of construction of SORDs with blocking has been obtained. The blocking of the SORD satisfying the condition of orthogonal blocking usually requires a large number of experimental points. However, if one of the requirements of orthogonal blocking is waived, the problem of blocking becomes somewhat easier and the size of the experiment in such a case can be greatly reduced. Though in such designs, surface parameters are not all free from block effects.

Some modified and/or rotatable response surface designs useful for response optimization for symmetric as well as asymmetric factorials with equispaced doses have been obtained. Some efficient response surface designs for slope estimation for symmetrical factorials with equispaced doses have also been obtained using MINIMAX criterion. A catalogue of modified and/or rotatable response surface designs with orthogonal blocking is also prepared. Modified and rotatable second order response surface designs developed have been used by the Division of Agricultural Engineering and Division of Post Harvest Technology, IARI, New Delhi and Division of Biotechnology, IVRI, Izatnagar.

The problem of selecting explanatory variables for the case of non-linear models has also been investigated. The design criterion heavily depends on initial guess of the parameters in non-linear set up. The criterion of D-optimality is used for selecting any design. Replication structure of the experiments under a non-linear set up has been examined. Replicating all the basic design points equally gives increased valued of the criterion. A sufficient condition for a replicated design to be optimum is provided.

First and second order response surface models with neighbour effects from immediate left and right neighbouring units have been considered and the conditions for the estimation and rotatability of these models have been derived. Response surface methodology in the presence of neighbour effects as well as correlated error structure has been developed. Designs for fitting response surface models with quantitative-cum-qualitative factors have been studied under neighbour effects/ correlated error structure.

In several response surface designs, data on several response variables is recorded. Such experimental situations are common in food processing experiments. The experimenter is interested in obtaining a combination that simultaneously optimizes all the response variables. A procedure of estimation of parameters from linear multi-response models for incomplete multi-response experiments has been developed. Procedure of simultaneous optimization of complete and incomplete multi-response experiments has been developed and SAS codes are developed for its implementation.

### 4.6 Experiments with mixtures

There do occur experimental situations where a fixed quantity of inputs, may be same dose of fertilizer, same quantity of irrigation water or same dose of insecticide or pesticide etc. are applied. The fixed quantity of input is a combination of two or more ingredients. For example, fixed quantity of water may be a combination of different qualities of water sources or fixed quantity of nitrogen may be obtained from different sources. In a pesticide trial, a fixed quantity of pesticide may be obtained from four different chemicals. In these experiments the response is a function of the proportion of the ingredient in the mixture rather than the actual amount of the mixture. The experiments with mixture methodology are quite useful for these experimental situations. Experiments with mixtures methodology will give a functional relationship between proportions of inputs applied and response and is helpful to interpolate responses at points that have not been tried in the experiment. Optimum combination of proportions can also be obtained through this approach.

In some experimental situations it may not be possible to explore the total range of all the components because one may require that at least a certain amount of a particular component may be present in all blends or one may insist that the proportion of component may not exceed a certain proportion. Such constraints give lower and/or upper bound on different components of mixtures. In certain experimental situations there may be one or more factors that are not component(s) of the mixture but influence the response. This type of factor(s) is(are) called process variable(s). A method of construction of restricted region simplex design in the presence of process variables when upper bound is imposed on one of the components of the mixture has been obtained using response surface designs with equispaced doses. Extreme vertices algorithm has also been developed for obtaining designs with restricted regions. The designs for experiments with mixtures with orthogonal blocking have also been obtained.

There do occur many experimental situations, wherein designs for multi-factor mixture experiments are required. Designs for multifactor mixture experiments available in literature are useful for fitting only the first order model for two factor mixture experiments only. In agricultural experiments, behaviour of different ingredients may be quadratic in nature, therefore, multifactor designs for experiments with mixtures need to be obtained so as to fit the second order response surface for multifactor mixture experiments. Designs for multi-factor mixture experiments have been obtained using response surface designs and projection matrices.

## 5. Some Other Considerations

### 5.1 Designs for biological assays

In a biological assay, a test (new) preparation and a standard one are compared by means of reactions that follow on their applications to some subjects. Characterizations of designs for symmetrical and asymmetrical parallel line assays have been obtained and methods of construction of designs satisfying these characterizations have been generated. These designs are generally simple partially efficiency balanced designs because of the nature of the contrasts of interest. The contrasts on interest need to be estimated with full precision.

Efficient designs for symmetrical parallel line assays that estimate all the three contrasts of interest with as high a precision as possible have been obtained and catalogued. Some methods of construction of symmetric multiple parallel line assays and slope ratio assays have also been obtained.

### 5.2 Designs for microarray experiments

Microarray experiments are conducted to study the relative expression levels of thousands of genes simultaneously and identification of differentially expressed genes. Work on design and analysis of microarray experiments was initiated and optimality aspects of designs by taking array, dye and variety effects in the model has been studied. Lower bounds to A- and D-Efficiencies have been obtained under mixed effects model taking array effects as random. Algorithm has been developed for obtaining efficient block designs for micro array experiments. Designs that are efficient for both the situations when array effect is taken fixed and as well as random effects have been identified. Analytical procedures based on mixed effects model considering array effects as random has been developed for identification of differentially expressed genes.

### 5.3 Designs for two-line and four line cross experiments

Optimal proper and non-proper incomplete block designs for diallel cross experiments for the estimation of contrasts among the general combining ability effects have been obtained and catalogued. Optimal balanced augmented block designs for diallel crosses have also been obtained and catalogued. Efficient partial diallel cross plans have been obtained from the three-class association schemes and a catalogue of these plans has been prepared. Optimal block designs for diallel crosses for estimation of heritability and ratio of variance components have been obtained to deal with the experimental situations in which the parental lines are randomly selected from a population of lines.

### 5.4 Block designs for multistage experiments

In experiments where the experimental unit is long lived, for example, horticultural experiments with trees as experimental units, different sets of treatments may be applied at different stages of the growth of tree. Other experimental situations like this could be experiments using animals as units. Such designs are known as multi-stage designs with two (or more) sets of non-interacting (or may be interacting) treatments. To deal with such experimental situations, block designs for two interacting / non-interacting sets of treatments applied in succession have been developed. Through these designs, the comparison among the direct effects of the treatments applied at second stage and residual effects of the treatments applied at the first stage can be made. A correspondence has been established between structurally incomplete row-column designs, block designs with nested rows and columns and designs for two stage experiments.

### 5.5 Cross over designs

Experiments with biological entities, like trials with animals, often involve application of a sequence of treatments to each experimental unit over varying periods of time using crossover designs (CODs). When experimental units exhibit larger variability or when the number of units is limited due to ethical or cost reasons, a COD allows treatments to be adequately replicated. In the literature, these designs are known by many other names *viz.*, repeated measurements designs, change over trials, switchover trials, designs involving sequences of treatments, time-series designs, before-after designs, *etc*. These designs are used advantageously in nutrition experiments with dairy cattle, clinical trials in medical research, psychological experiments, long-term agricultural field experiments, bioequivalence trials, *etc*. The peculiarity of a COD is that any treatment applied to a unit in a certain period influences the responses of the unit not only in the period of its application but also leaves residual effects in the following periods.

The residual effects may persist for one or more periods. Generally a blank period is left for washing away the effect of the treatment in the previous period.

CODs can be classified on the basis of the properties they possess. We present a brief account of such classifications along with the contributions made by division of Design of Experiments.

A COD is said to be uniform over periods (experimental units), if each treatment occurs equally often in each period (experimental unit). A COD is said to be uniform if it is uniform over both periods as well as units.

Pre-period or $0^{th}$ period is a period preceding the first experimental period in which appropriate treatments are applied to the experimental unit but either the observations are not recorded or if recorded they are not taken into consideration while carrying out the analysis of the data. By adding a pre-period to the design, the first period observations also receive the residual effects of the treatments and hence the data become homogeneous. A COD allowing the estimation of direct and first residual effects with one pre-period, in which the treatments in the period are exactly the same as those in the last period, is called a circular COD.

CODs in which every treatment is preceded equally often by every other treatment in the sequences have been called combinatorial balanced with respect to first order residual effects. A design is said to be strongly balanced if each treatment is preceded by every treatment including itself equally often. A COD permitting the estimation of first order residual effects is called variance balanced if the variance of any elementary contrast among the direct effects is constant, say $\alpha$, and the variance of any estimated elementary contrast among the residual effects is also constant, say $\beta$. Here $\alpha$ and $\beta$ may not be equal. If $\alpha = \beta$, then these designs are known as totally balanced COD.

The work at IASRI on CODs was started with the problem of construction of CODs having incomplete periods, balanced for first order residual effects of treatments. A simple method of construction of William square designs balanced for first order residual effects of treatments has been developed at the Institute. A catalogue of CODs requiring only 50 or less experimental units have been prepared that serves as a ready recknoer for the experimenters.

Universal optimality aspects of CODs when number of periods is less than number of treatments has been investigated and sufficient conditions have been obtained for a COD to be universally optimal in a pre-defined class of designs. Optimality aspects of some series of CODs in the presence of first and second order residual effects have also been investigated. Series of balanced CODs, optimal in a given class of designs has been obtained.

Series of CODs balanced for first as well as second order residual effects have been obtained. Two classes of totally balanced designs have been obtained considering the presence of first order residual effects of treatments.

A minimal balanced design contains the minimum possible number of experimental units. A general method of construction of minimal balanced CODs for odd number of treatments along with an outline of their analysis has been given. Further, two new classes of minimal strongly balanced CODs assuming the presence of first order residual effects has been obtained. These designs are partially variance balanced with *m*-associate classes based on a circular association scheme. A class of cyclic CODs has been obtained.

Circular CODs have been obtained by associating circular designs with a complete set of mutually orthogonal Latin squares.

A method of constructing completely balanced CODs that permit the estimation of direct effects orthogonal to all other effects, when the residual effects of treatments last for two consecutive periods, has been developed. Some methods of construction of partially balanced CODs have also been given. A series of ternary CODs along with a method of analysis has been proposed. The assumption in these designs was that in a given sequence a treatment should not occur more than two times and the same treatment should not appear in two consecutive periods.

Designs that give independent estimates of direct and residual effects, of approximately equal precision, are obtained by adding an extra period to the original design. In the new final period, the treatments that were applied in the previous final period are repeated. Some series of extra-period CODs have been obtained.

The performance of two-treatment CODs in the presence of time trends has been studied. The efficiency of estimation of various treatment effects of two-treatment CODs has been improved by adding a pre-period with appropriate treatments. Some efficient two-treatment CODs when the errors are auto-correlated has been obtained.

The combinatorial aspects of CODs for making test treatments- control treatment comparisons have been studied. Schur optimality aspects of CODs for comparing several treatments with a control have been studied. It has been shown that the Schur-optimal control CODs are necessarily balanced. Some methods of construction of Schur-optimal balanced control CODs have been given.

In many agricultural experiments and veterinary medical trials, it is often required to measure the effect of response from simultaneous application of levels of two or more factors over varying periods where levels of at least one factor exhibit residual effects. An amalgamation of CODs and factorial experiments may be required in such situations. Two-factor symmetrical split-type CODs have been developed. Two classes of uniform strongly balanced CODs with two factors balanced for residual effects of levels of one factor have been obtained. For situations where the two factors are unrelated but existence of first order residual effects of both the factors are suspected, CODs with two non-interacting factors assuming residual effects to be present have been obtained using mutually orthogonal Latin squares and William squares.

Almost all the studies have been carried out under the assumption that observations are independent but two successive observations can have a correlation structure. The necessary and sufficient conditions for designs to be balanced in presence of auto-correlated errors have been obtained. In a COD, a sequence of treatments is applied to an experimental unit and observations are recorded over the periods and hence it is very much possible to observe interaction between treatments and experimental units. A non-additive model with treatment $\times$ unit interaction effects is recommended for such situations, as these effects also contribute significantly to the response measured. Under a non-additive model, a uniform balanced non-circular class of CODs with a pre-period has been shown as universally optimal for the estimation of direct effects.

In general, the effects in the COD model are considered as fixed and observations are uncorrelated and distributed normally. Many a time the assumptions made in the model are violated and have an effect

on the estimates of treatment contrasts. Hence, variance of treatment contrasts, when the unit effects are considered as random and the observations are correlated and/or non-normally distributed, have been estimated through simulation studies. An attempt has been made to study the efficiency of treatment comparisons using Bayesian analysis, empirically. It was seen that the estimate obtained under Bayesian approach using GIBBS Sampling (BUGS) is much smaller than that obtained under classical approach for the same data set.

CODs have immense applications in bioequivalence trials. Bioequivalence refers to the degree to which clinically important outcomes subjected to a new preparation resemble those of a previously established preparation. Standard CODs are being used in bioequivalence trials, but they treat all pair-wise comparisons of formulation effects with equal importance. The experimenters are more interested in the comparison of several test formulations to an established standard or control or reference formulation rather than in all pair-wise comparisons. Some series of reference balanced CODs for bioequivalence trials with or without considering residual effects have been obtained.

### 5.6 Robust designs

It has been described in section 2 that robustness of designs may be investigated against disturbances or accidents like loss of data, interchange or exchange of treatments, presence of a systematic trend among observations in a block, presence of outliers, etc. The criterion of robustness against any of these disturbances could be in terms of connectedness of design, variance balance, optimality, orthogonal factorical structure, trend free, efficiency of the residual design after the accident, etc. The robustness of designed experiments has been studied under a general linear model for inferring on a set of treatments as per the connectedness criterion as well as the A-efficiency criterion. It has been shown that a design robust under a homoscedastic set up is also robust under the general heteroscedastic set up with correlated observations. The robustness of block designs has been studied in particular against the loss of data, in general against loss of any $t$ observations. Designs robust against the loss of one, two and three observations have been identified. Robustness of block designs against the loss of any two blocks, not necessarily disjoint, has also been studied. Similarly, the robustness of row-column designs has also been investigated against the loss of data. The resistance of variance balanced block designs with unequal block sizes that are universally optimal has been investigated against the loss of all the observations pertaining to a treatment. Some methods of construction of resistant designs have also been obtained. Binary variance balanced block designs have been shown to be robust against (i) missing observations; (ii) exchange of a treatment; (iii) interchange of a pair of treatments and (iv) presence of one or two outlier(s). Several diagnostic criteria for detection of outliers in experimental data have been developed. Robustness aspects of designs against the presence of one outlier have been investigated. It has been shown that all BIB designs and binary balanced block designs are robust in presence of two outliers. All other E-optimal designs are robust in presence of one outlier. Several 2-associate class PBIB block designs and variance balanced row-column designs that satisfy the property of adjusted orthogonality, nested balanced incomplete block designs have been shown to be robust in presence of a single outlier.

Balanced treatment incomplete block designs have been identified that are robust against the presence of any one of the following disturbances (i) one observation missing, (ii) one block missing, (iii) exchange of a treatment, (iv) interchange of a pair of test treatments, and (v) presence of a single outlier. Balanced bipartite block designs that are resistant against the loss of all observations pertaining to a test treatment

have been introduced.

Robustness of neighbour balanced designs has been investigated against loss of data and also for different correlation structures and against different values of correlation coefficient.

Binary balanced block designs for diallel crosses have been shown to be robust against the presence of any one of the following disturbances (i) exchange of a cross, (ii) interchange of a pair of crosses, (iii) single observation missing, (iv) all the observations pertaining to a block missing, and (v) presence of a single outlier.

It is well known that an EGD design, if exists, has the property of Orthogonal Factorial Structure with balance. During the course of the investigation, a particular level of some factor turns out to be lethal resulting in the loss of all the observations pertaining to treatment combinations with this level of factor. High cost or non-availability of inputs may sometimes force the experimenter to discard such observations and continue with the experimentation. Such kind of exigencies may disturb the original structure of the design. Thus the concept of structure resistant designs has been introduced at IASRI for the first time. The designs retain the property of orthogonal factorial structure with balance when all the observations pertaining to the treatment combinations containing one level of a particular factor are lost. Robustness aspects of designs for such situations have been studied and structure resistant designs have been obtained.

For multi factor experiments, sometimes level of a particular factor may be lethal or experimental material pertaining to this level of the factor could not be procured in time and as such all the observations pertaining to the treatment combination containing that level of the particular factor are lost. Robustness aspects of designs for such situations have been studied and structure resistant designs have been obtained.

Robustness aspects of response surface designs against single missing observation have been investigated using efficiency per obsevation criterion.

Robustness of designs for bioassays against the loss of a dose has been investigated. The loss of all the observations corresponding to a dose (standard or test) gives rise to designs for asymmetric parallel line assays. This is another example of structure resistant designs. Using the structure resistance property, several designs for asymmetric assays have been obtained.

Robustness aspects of CODs have also been investigated against missing observation(s) using efficiency criterion. CODs balanced with respect to first order residual effects were found to be robust against the loss of observation(s) from one experimental unit.

In some experimental situations, there may exist a systematic trend in the experimental units. Designs that allow the estimation of contrasts of interest free from trend effects are known as trend free designs. Conditions for a BIB design, NBIB design and proper binary variance balanced block design for diallel crosses to be trend free have been obtained and catalogues of trend free BIB designs, NBIB designs and proper binary variance balanced block designs for diallel crosses have been prepared for the benefit of the experimenters.

Systematic trend may also be present in the experimental units while conducting a factorial experiment. If there is trend in the experimental units, the interest of experimenter is to eliminate the trend effect and obtain the estimates of contrasts of interest (main effects and lower order interactions) free from trend effects. The resulting designs are called as trend-free designs for factorial experiments for the effects of

interest and ordered application of treatments to experimental units is called run order. Computer aided search has been made for linear trend-free designs for 2-level factorial experiments both with and without confounding. Computer aided search has also been made to identify the two-factor and three factor interactions that can be estimated free from trend effects along with main effects. Computer aided search has also been made to obtain fractional factorial plans that provide the estimates of main effects free from block effects.

In a CODs also, observations are taken over periods from experimental units and hence it is quite possible that the units exhibit trend over the periods. Validity of least square technique for the analysis of CODs balanced for first and second order residual effects of treatments when the experimental units exhibit time trends over periods has been investigated. A class of two-period totally balanced trend-free CODs in the presence of first residual effects that allow the estimation of treatment effects orthogonal to trend effects has been developed.

### 5.7 Sampling vis-à-vis design of experiments

The use of interesting combinatorial properties of binary, proper block designs like balanced and doubly balanced incomplete block designs, partially balanced incomplete block designs, cyclic designs, etc. enables one to generate inclusion probability proportional to size sampling schemes for sample size more than two. The sampling schemes are easy in operation. The ease in sample selection is achieved by building an association between sampling design and the experimental design. The experimental design is mapped onto the sampling design. The treatments (v) are the population units (N), the totality of the blocks is the sample space (S containing b samples), the sample size (n) is the block size (k) and the contents of a block are the sampled units. In this respect the sampling scheme is very simple in operation. A probability measure is defined on the sample space. The selection of one block with a pre-assigned probability gives the entire sample. The sampling schemes so generated are in fact controlled selections in the sense that only a subset of all the possible samples of size n has a non-zero probability of selection while all the remaining samples have a zero probability of selection. As the choice of design is not unique, the desired precision of estimation can also be achieved by controlling the second order inclusion probabilities. Many sampling schemes have been generated by this technique.

For estimating the variance of nonlinear statistics like regression and correlation coefficients, birth and death rates, from a complex survey data, the method of balanced repeated replication is more useful than the other methods like linearization, bootstrap and jackknife repeated replications. Once again a relationship is established between mixed orthogonal arrays of strength two and the balanced repeated replications. This relationship has enabled to draw samples with unequal number of selections per stratum. The use of proportional frequency plans as balanced repeated replication has also been exploited by adjusting for bias. A bias corrected non-linear statistic is also obtained. Based on simulation studies, it has been shown that the proposed method is better than the grouped method.

Design combinatorics and linear programming approach have been exploited in generating balanced sampling plans for sampling from populations in which adjacent units provide similar observations due to natural ordering of units in time or space. An algorithm for generation of balanced sampling plans excluding adjacent units (BSA plans) has been developed. A new family of distance balanced sampling plans (DBSP) with the property that the second order inclusion probabilities are non-decreasing function

of distance between the two concerned units is developed. Inclusion probability propertional to size sampling plans excluding adjacent units have been developed by making use of binary, proper and unquireplicated block designs.

# Pre-harvest Crop Production Forecast Methodologies: IASRI Studies

**Chandrahas and Ranjana Agrawal**

## 1. Introduction

Reliable forecasts of crop production before the harvest constitute a problem of topical interest. Such forecasts are needed by the Government, agro-based industries, traders and agriculturists alike. The Government needs these for use as a basis for its policy decisions in regard to procurement, distribution, buffer-stocking, import-export, price fixation and marketing of agricultural commodities while agro-based industries, traders and the agriculturists need forecasts for planning their operations properly. To meet such needs, crop forecasts under the prevalent system in India are being issued for principal crops at four points of time during a year by the Directorate of Economics and Statistics, Ministry of Agriculture (DESMOA), New Delhi. The first official forecast is issued in the middle of September, second is made in the month of January, third is prepared towards the end of March / beginning of April and fourth in the month of June. These forecasts are, however, of a subjective nature since these are based on eye-estimates and personal judgement of agricultural officials and the final crop production estimates, though based on objective crop-cutting experiments, are of limited utility as these become available quite later after the harvest.

In view of this, there is a need for an objective methodology for pre-harvest crop forecasting. This involves building up suitable forecast model(s) which has certain merits over the traditional forecasting method. These merits include the objectivity of the forecast and its ability to provide a measure of reliability which a traditional forecast method can not provide. This, as such, calls for the necessity of objective methods for pre-harvest forecast of crop yields. In this article, various studies undertaken at Indian Agricultural Statistics Research Institute (IASRI), New Delhi on crop forecasting have been presented.

## 2. Forecasting Studies at IASRI

As crop acreage and yield rates constitute the two components of crop production of which the former is available before harvest and the latter only after harvest, the work on forecasting at the Institute was restricted to developing yield forecast models. The main factors affecting crop yield are inputs and weather experienced by the crops during growth period. Use of data on these factors forms one approach for forecasting crop yields. The other approach uses plant vigour measured either through plant characters or through remotely sensed data. The second approach is based on the fact that various factors affect crop growth through plant processes. These effects are manifested through crop stand, number of tillers, root length, leaf area, number of ear heads etc., which ultimately determine crop yield. A number of techniques based on different types of data have been developed at the Institute for various crops. These are discussed in the following sections.

### 2.1 Weather based models

Crop yield is affected by technological change and weather variability. It can be assumed that the technological factors will increase yield smoothly through time and, therefore, years or some other

parameter of time can be used to study the overall effect of technology on yield. The weather variability both within and between seasons is another and the only uncontrollable source of variability in yields. The weather variables affect the crop differently during various stages of development. Thus extent of weather influence on crop yield depends not only on the magnitude but also on the distribution pattern of weather variables over the crop season which, as such, calls for the necessity of dividing the whole crop season into fine intervals and studying relationships of weather variables with crop yield in these intervals. This will increase number of variables in the model and in turn a large number of model parameters will have to be evaluated from the data. This will require a long series of data for precise estimation of parameters which may not be available in practice. Thus, a technique based on relatively smaller number of manageable variables and at the same time taking care of entire weather distribution may solve the problem.

**Models using composite weather indices**

Fisher (1924) and Hendricks & Scholl (1943) have suggested models which require small number of parameters to be estimated while taking care of distribution pattern of weather over the crop season. Fisher divided the whole crop season into 5/7 days intervals. He assumed that the effect of change in weather variable in successive periods would not be abrupt or erratic but an orderly one that follows some mathematical law. He assumed that these effects as well as magnitude of the weather variable in successive periods are composed of the terms of polynomial functions of time. Substituting these in usual regression model,

$$Y = A_0 + A_1X_1 + A_2X_2 + ... + A_nX_n + e$$

(here Y denoted yield and $X_w$ weather variable in $w^{th}$ period w = 1,2,....,n and e error term) and utilising the properties of orthogonal and normalised functions, he obtained

$$Y = A_0 + a_0\rho_0 + a_1\rho_1 + a_2\rho_2 + ... + a_k\rho_k + e$$

where $A_0$, $a_0$, $a_1$, $a_2$,....$a_k$ denoted parameters to be determined, $\rho_i$ ( i =1,…, k) distribution constants of $X_w$. Fisher has suggested to use k = 5 for most of the practical situations. In fitting this equation for k = 5, the number of parameters to be evaluated will remain 7, no matter how finely growing season is divided. This model was used by Fisher for studying the influence of rainfall on the yield of wheat.

Hendricks and Scholl modified Fisher's technique. They assumed that a second degree polynomial in period number would be sufficiently flexible to express the effects in successive periods. Under this assumption, the model suggested by Hendricks and Scholl was

$$Y = A_0 + a_0 \sum_{w=1}^{n} X_w + a_1 \sum_{w=1}^{n} w X_w + a_2 \sum_{w=1}^{n} w^2X_w + e$$

where $A_0$, $a_0$, $a_1$ & $a_2$ represented model parameters.

In this model number of parameters to be determined reduced to 4, irrespective of n. This model was extended to study joint effects of weather variables and an additional variate T representing the year was included to make allowance for time trend.

At IASRI, the model suggested by Hendricks and Scholl has been modified by expressing effect of

changes in weather variables on yield in the $w^{th}$ period as second degree polynomial in respective correlation coefficients between yield and weather variables [Agrawal *et al.* (1980), Agrawal and Jain (1982), Agrawal *et al.*(1983), Jain *et al.* (1980)]. This is expected to explain the relationship in a better way as it gives appropriate weightage to weather in different periods. Under this assumption, the models were developed for studying the effects of weather variables on yield using complete crop season data whereas forecast model utilised partial crop season data. These models were found to be better than the one suggested by Hendricks and Scholl.

These models were further modified [Agrawal *et al.* (1986)] by expressing the effects of changes in weather variables on yield in $w^{th}$ period as a linear function of respective correlation coefficients between yield and weather variables. As trend effect on yield was found to be significant, its effect was removed from yield while calculating correlation coefficients of yield with weather variables to be used as weights. Effects of second degree terms of weather variables were also studied. The results indicated that (i) the models using correlation coefficients based on yield adjusted for trend effect were better than the ones using simple correlations, (ii) inclusion of quadratic terms of weather variables and also the second power of correlation coefficients did not improve the model. This suggested the use of following type of model to study effects of weather on yield.

$$Y = A_0 + a_0 Z_0 + a_1 Z_1 + c\,T + e \quad \text{where} \quad Z_j = \sum_{w=1}^{n} r_w^j X_w \quad ; \quad j = 0, 1$$

here $r_w$ denoted correlation coefficient of weather variable in $w^{th}$ period with yield (adjusted for trend effect, if present). The model was further extended for studying joint effects.

The forecast model has been developed using partial crop season data considering all weather variables simultaneously. The model finally recommended was of the form

$$Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'=1}^{p} \sum_{j=0}^{1} a_{ii'j} Z_{ii'j} + cT + e$$

where

$$Z_{ij} = \sum_{w=1}^{m} r_{iw}^j X_{iw} \quad \text{and} \quad Z_{ii'j} = \sum_{w=1}^{m} r_{ii'w}^j X_{iw} X_{i'w}$$

$r_{iw}/r_{ii'w}$ denoted correlation coefficient of yield (adjusted for trend effect, if present) with $i^{th}$ weather variable/ product of $i^{th}$ and $i`^{th}$ weather variables in $w^{th}$ period, m period of forecast and p number of weather variables used.

In this model, for each weather variable, two weather indices were developed, one as simple accumulation of weather variable and the other one as weighted accumulation of weather variable over different periods, weights being correlation coefficients of weather variable in respective periods with yield (adjusted for trend effect, if present). Similarly, for interaction of weather variables, indices were generated using period-wise products of weather variables taking two at a time. Stepwise regression technique was used to select the important weather indices.

The above approach was used to forecast yield of rice and wheat at district level in different situations, viz (i) rainfed district having deficient rainfall (rice), (ii) rainfed district having adequate rainfall (rice) and (iii) irrigated district (wheat). Data starting a fortnight before sowing were considered as this period

is expected to have effect on establishment of the crop. The results revealed that reliable forecasts can be obtained when the crops are 10-12 weeks old i.e. about 2 to 2½ months before harvest. This approach was also used to develop forecast model for sugarcane for Kolhapur district using fortnightly weather data [Mehta *et al.*(2000)]. Deviations of forecasts for subsequent years (not included in model development) from observed ones ranged between 5-10 %.

These studies, carried out at district level, revealed data requirement of about 25 years for reliable forecasts. Such a long series may not be available for most of the locations. Therefore, model development was attempted at agro climatic zone level. The models were developed by pooling the data of various districts within the zone so that a long series could be obtained in a relatively shorter period. Models were developed for wheat in Vindhya Plateau zone and for rice in then Chattisgarh Plain & Bastar Plateau zone taken together (as a portion of Bastar district falls under Chattisgarh Plain zone whereas remaining under Bastar Plateau zone and yield figures are available at district level only). Agricultural inputs, previous year's yield and moving averages of yield were taken as the variables taking care of variation between districts within the zone. Year was included to take care of technological changes. Different strategies for pooling district level data for the zone were adopted. Results revealed that reliable forecasts can be obtained using this methodology at 12 weeks after sowing i.e. about 2 months before harvest at zone level also. The data requirement reduced to 10-15 years as against 25 years (approx.) for district level models. The study also revealed that forecast model will be appropriate to forecast the yield of a zone even if data for some districts within the zone are not available at model development stage or at forecasting stage [Agrawal *et al.* (2001)].

This approach was further studied in detail for various districts and agro climatic zones of Uttar Pradesh for one major kharif crop (rice), one major rabi crop (wheat) and one long duration crop (sugarcane) so as to come out with a suitable methodology for forecasting crop yields at state level. It was observed that performance of the models based on only weighted indices was almost at par with those developed earlier. Therefore, the simplified form of the model based on only weighted weather indices has been recommended. With this approach, reliable forecasts for rice and wheat could be obtained when crop was 11 weeks old i.e. approx. 2½ months before harvest. Sugarcane forecast could be obtained in the middle of September using data of 14 fortnights (starting from March first fortnight). The forecasts for subsequent years were compared with two types of observed yields - one based on the districts for which forecasts were obtained while the other one based on all districts of the state. The results are presented below:

**Forecasts for rice, wheat and sugarcane in Uttar Pradesh**

| Observed yield based on | Percent deviation of forecasts from observed yield | | |
|---|---|---|---|
| | Rice | Wheat | Sugarcane |
| districts used in forecasts | 4.2 | 0.7 | 0.8 |
| all districts of U.P. | 3.3 | 4.3 | 4.5 |

The methodology is simple, adoptable, does not involve use of very detailed data collection/ sophisticated statistical tools and at the same time provides reasonably good forecasts. It is suitable at district, agro climatic zone as well as state level. This approach has been used by Space Application

Centre, Ahmedabad to obtain forecast for wheat yield at national level. The forecast thus obtained was with only 3% deviation from observed yield.

**Model using discriminant function analysis**

At district level, model based on time series data of weather variables was also developed using discriminant function analysis. The long series data of 25-30 years were classified into three groups - congenial, normal and adverse on the basis of crop yield. Using weather data of these groups, linear and quadratic discriminant functions were fitted. These functions were used to find weather scores for each year at different phases of crop growth which were used as regressors in forecast model along with agricultural inputs and time trend. The study was carried out for rice in Raipur district. The performance of quadratic discriminant function was found to be better than linear discriminant function. This may be due to heterogeneity in dispersion matrices of weather variables in different groups. This approach could provide reliable yield forecast, with approx. 1% deviation from observed yield, about two months before harvest [Rai and Chandrahas (2000)].

In another study by Aditya (2008), use of discriminant function analysis has been studied for wheat crop in Kanpur district of Uttar Pradesh. Discriminant scores derived through different procedures were used as regressors in forecast models. The recommended model involved working out discriminant scores using weather data of first week, then obtaining discriminant scores using weather data of second week alongwith discriminant scores based on first week's weather variables and so on. Finally, two discriminant scores obtained using weather data of the last week and discriminant scores based on data upto previous week were used as regressors in the model alongwith the time trend. This model provided forecast of yield about one and a half months before harvest with deviation from the observed one around 5%.

**Models using water balance technique**

Using water balance technique, models have been developed for rainfed crops, rice (Raipur), sorghum (Delhi & Parbhani) and maize (Delhi). In this approach, water deficit/surplus has been worked out at weekly intervals. Weighted stress indices were prepared phase-wise by applying appropriate weights to surplus as well as deficit depending upon stage at which it occurred. These stress indices for each year have been used as regressors alongwith time trend to develop the forecast model.

The estimated soil moisture in the root zone at the end of $i^{th}$ week was obtained from the relation

$$S_i = S_{i-1} + R_i - WR_i$$

where, $R_i$ = rainfall in the $i^{th}$ week

$WR_i$ = water requirement of the crop in $i^{th}$ week

= $k_i \times evap_i$ ($k_i$ crop coefficient and $evap_i$ pan evaporation in the $i^{th}$ week)

Stress due to deficit, $St_i = 1 - AE_i / WR_i$, where $AE_i$ Actual Evapotranspiration.

Stress due to water surplus was taken as '1' if $S_i$ > water holding capacity.

Using this technique the forecasts were obtained for sorghum, maize and rice respectively 6, 4 and 5 weeks before harvest. Deviations of forecasts from observed yields for subsequent years ranged between 1–11% [Saksena *et al.* (2001)].

### 2.2 Plant characters based models

Effects of various causal factors (weather, inputs, etc.) are manifested through crop stand, number of tillers, leaf area, number of earheads, etc. which ultimately determine crop yield. As such, plant characters can be taken as the integrated effects of all the causal factors and may serve as good scientific indicators of plant condition. Considerable work has been done at the Institute for developing crop yield forecast models based on plant characters for paddy (Sambalpur, West Godavari, Delhi & Ludhiana), jute (24 Parganas & Purnea), wheat (Ludhiana, Delhi & Aligarh), jowar (Sangli), cotton (Baroda, Jalgaon & Aligarh), tobacco (Prakasam), sugarcane (Meerut & Kolhapur), apple (Shimla) and groundnut (Rajkot). Two types of approaches for modeling have been attempted - Between year model and Within year model.

#### 2.2.1 Between year models

These models are developed taking previous year(s) data. Objective yield forecasts are obtained by substituting the current year plant data into a model developed from the previous year(s). An assumption is made that the present year is a part of the composite population of these (previous) years. The data utilized was on plant characters collected at different periodic intervals from farmers' fields using stratified multistage sampling design for 3-4 years taking Village Level Workers' circles/Community Development Blocks/Taluks in the selected district as strata. The samples were selected by selecting two villages as first stage units from each stratum, 2-4 fields from each selected village as second stage units and two plots per field as third stage units. In all, around 200-250 fields from a district have been selected for study in each crop. In case of apple crop, orchards and trees formed the second and third stage units respectively. The sampling units at each stage were selected by simple random sampling without replacement and were kept fixed for the entire crop season. Various models studied under this approach are as follows :

**Linear regression model**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + e$$

where Y represented crop yield, $X_i$'s plant characters and $\beta_0$ & $\beta_i$ 's regression coefficients. These models utilized data at one point of time only during the crop growth [Sardana *et al.* (1972), Singh *et al*. (1976), Jha *et al*. (1981), Singh and Bapat (1988), Rai, *et al.* (1998)]. In these models, the regressand Y and the regressors $X_i$'s were used in the original scale and in some models, these were transformed to log, square root and reciprocal scales. As all the models were found roughly at par, for simplicity and ease in interpretation, the models using variables in original scale were recommended. This approach provided forecasts atleast one month before harvest for various crops at different locations. For cotton crop, first picking yield was also used alongwith plant characters in the model.

**Model based on growth indices**

The models using data at one point of time do not provide the idea of growth pattern of plant characters. Therefore, models using growth indices were attempted to improve upon usual linear regression models. In these models regressors were growth indices based on plant characters observed on two or more points

of time during the crop growth which were obtained as weighted accumulations of observations on plant characters in different periods, weights being respective correlation coefficients between yield and plant characters. The form of the model was

$$Y = \beta_0 + \Sigma \beta_i G_i + e$$

where $G_i = \sum_{w=n_1}^{n_2} r_{iw} X_{iw}$ represented the index of the $i^{th}$ character, w period identification, $n_1$ & $n_2$ the initial & final periods considered in developing the index of the character, $r_{iw}$ simple/partial correlation coefficient between yield and $i^{th}$ character in $w^{th}$ period, $X_{iw}$ the value of the $i^{th}$ character in $w^{th}$ period.

This model was attempted for hybrid jowar in Sangli district and was found better than conventional linear regression model [Jain *et al*. (1985)] .

**Principal component regression**

Forecast models were attempted using principal components of biometrical characters alongwith crop inputs as regressors. [Jain *et al.* (1984), Chandrahas *et al.* (1989)]. Principal components were obtained using data on biometrical characters at one or more points of time.

**Probability model based on markov chain**

In multiple regression approach least squares technique is used for estimating the parameters of the model. The optimality properties of these estimates are described in an ideal setting which is not often realised in practice. It has been observed that regression based on different subsets of data produces very different results, raising questions of model stability. To overcome some of the drawbacks of regression model, Markov Chain theory has been used for developing probability model for forecasting crop yield. This method, being completely model free, does not require any assumption about independent/dependent variables. Markov Chain method has the advantage of providing non parametric interval estimates and is robust against outliers/extreme values.

In this method, growth process of the crop is divided into phenological stages. A markov chain model is constructed by defining a set of states, which describe the average condition of a group of plants at specified time within the phenological stages. Individual states are defined on the basis of available qualitative and quantitative information to describe the plant condition. Transition matrices are worked out which give the transition probabilities of a plant (or a group of plants) moving from any possible state of any stage in the growth process to any state of next stage in the model. Predicted yield distributions are obtained by using transition matrices which are used to provide the yield forecast. This method was applied to forecast yield of sugarcane [Jain and Agrawal (1992(a)), Agrawal and Jain (1996)]. Models using higher order markov chain using data on plant characters as such / principal components/ growth indices of plant characters were also attempted [Jain and Ramasubramanian (1998), Ramasubramanian and Jain (1999), Ramasubramanian *et al.* (2004)].

**2.2.2 Within year models**

The between year models while performing satisfactorily in typical years may falter in atypical years. A model which uses data from the current growing season only may be beneficial in improving forecasts during a year with atypical growing conditions. Such approach will also be useful for forecasting crop

yield in areas where past data are not available. These models had been developed to provide forecasts of pertinent components of crop yield relying entirely on growth data collected from plant observations during the current growing season. This approach has been attempted for rice and wheat using data collected from IARI Research farm, New Delhi.

Logistic model was used to model growth pattern of dry matter accumulation and head / panicle weight. Partial crop season data were utilised to fit the curve and the value of variables under study at harvest was predicted through this curve which in turn was used to forecast yield. As over-estimation was observed when the models were fitted using partial crop season data based on the data points falling on the lower side of the curve where growth has steep rise, the usual logistic model was modified so as to capture the value of variables under study at maturity from partial crop season data. The proposed modified logistic model was

$$Y_t = \frac{\alpha^{\sqrt{\frac{t_m}{t_f}}}}{\left(1+\beta\rho^t\right)} + e_t$$

where $Y_t$ represented dependent growth variable, t independent time variable, $t_m$ time of maturity and $t_f$ time of forecast, a > 0, $\beta > 0$, $0 < r < 1$.

The modified growth model worked well to adjust for over-estimation. By this approach, the forecasts of total dry matter and head / panicle weight at maturity were obtained about one month before harvest and 15 days before harvest respectively. The forecasts were found close to the observed ones [Jain et al. 1992(b)].

### 2.3 Use of farmers' appraisal in the model

Farmers are the best judge of the likely production in the field as the farmers engaged in the cultivation of the crop have their own inbuilt concept in their mind which takes into account the totality of the effects of input variables, climatic effects, a view of biometrical characters and also the soil characteristics to arrive at the expected yield. Farmers' appraisal, therefore, could serve as a good input for forecasting the yield and replace some of the characters requiring expertise or use of sophisticated instruments for their measurements and thus reducing the cost on data collection. A study has been carried out to study the feasibility of using farmers' appraisal in the forecast model for sugarcane [Agrawal and Jain (1996)]. The results revealed that a reliable forecast could be obtained using plant population and farmers' appraisal. Farmers' appraisal in this model has replaced plant height used in the model based on plant characters only.

Another methodology based on farmers' appraisal data has been developed using Bayesian approach. This approach for forecasting is basically based on the judgement / opinions / views of the target group in which the information is obtained directly about the concerned characters to arrive at a forecast. Chandrahas and Rai (2001) developed methodology for obtaining wheat yield forecast in Muzaffarnagar district using farmers' appraisal data through Bayesian approach. Expert opinion data were collected in a number of rounds by interviewing the selected farmers regarding their assessment about the likely crop production and chance of occurrences in various yield classes. From these responses average prior probabilities were computed. Actual harvest yield and farmers' appraisal data on yield for previous year(s) were taken into account to obtain posterior probabilities which were then used for obtaining Bayesian forecast of

crop yield for current year. Forecast could be obtained about two months before harvest. The significant advantage of the Bayesian approach is that it provides forecasts representing the composite thinking of a number of farmers actually engaged in cultivation of the crop. This approach provides quick and less expensive forecasts and the accuracy is expected to be more than purely eye-estimate based forecasts.

### 2.4 Integrated model / composite forecast

The crop yield depends on many types of variables viz. weather factors, performance of plants during crop growth stages, agricultural inputs etc., therefore, it will be better if various aspects are considered in the forecast model. In this reference, a model has been tried using data on plant characters alongwith agricultural inputs for jowar (Jain *et al.* 1985). The results revealed that inclusion of date of sowing and fertilizer in the model increased the coefficient of determination by 8%. For apple and groundnut crops, integrated yield forecast models were attempted using data on biometrical characters, crop inputs and weather variables [Chandrahas and Narain (1992), Singh *et al.* (1991)].

Sometimes it is not possible to include all the variables in a single model. Therefore, a study has been carried out to obtain composite forecast as a suitable combination of forecasts obtained from different models. Various strategies for combining forecasts have been suggested under different situations [Mehta *et al.* (2000)].

### 2.5 Some aspects on sampling and error of forecast

#### Use of successive sampling

The forecast of crop yield based on biometrical characters involves the use of at least two years' data. The data of first year are used to set up the forecast model and the second year data on regressors to forecast the mean yield. Thus, the yield forecast estimator is a function of observations recorded in the two successive years. Accordingly, an attempt was made to provide a minimum variance linear unbiased estimator of yield forecast employing the theory of successive sampling. The study on optimum allocation of sample showed that independent sampling in the forecast year leads to more precise forecast [Chandrahas (1984)].

#### Error of forecast

Usually sampling design is ignored while fitting the models. The effect of sampling for regressors on the variance of forecast error was studied [Chandrahas (1984)]. The expression for error of forecast was derived by considering sampling and model components simultaneously [Jain *et al*. (1985)].

#### Effect of measurement errors on forecast

In fitting the models the regressors are normally assumed to be free from measurement errors which in practice may not be true. When the variables are subject to measurement errors, then as discussed by Cochran (1970), coefficient of determination $R^2$ gets reduced to $R^2.g_y.g_w$ where $g_y$ is the coefficient of reliability of dependent variable and $g_w$ is the weighted mean of the coefficients of reliabilities of regressors. Chandrahas and Rai (1998) worked on this aspect for developing forecast models for sugarcane crop. This study showed that coefficient of determination increases when the measurement errors were taken into consideration while analyzing the observed data. Use of replicated measurements for a number of units has been suggested to enable an estimation of the extent of measurement errors involved in the data.

### Application of two-phase sampling

Models based on plant characters require periodical data collection from farmers' fields. Therefore, some characters involving high cost of data collection or use of sophisticated instruments, though important, may not find place in the models or else the sample size will be required to be reduced drastically with a view to keep cost on data collection within a reasonable limit. Goyal (2003) suggested the use of two-phase sampling wherein data on less costlier characters are collected from a larger sample and data on characters involving high cost or labour are collected from a sub-sample. The appropriate methodology for developing forecast model based on entire data (i.e. data on some characters on smaller sample and some characters on whole sample) has been suggested. The approach has been successfully demonstrated by forecasting sugarcane yield in Meerut district.

## 3. Conclusion

Models for forecasting yield of various crops at various locations have been developed at IASRI using various types of data / approaches. The weather (indices) based models have been studied in detail for different crops at various locations. The developed methodology has been demonstrated at district, agro-climatic zone as well as state level. The methodology has been successfully used by various research workers.

The approaches for which data collection from farmers' fields is required, can be integrated with crop cutting surveys where sample plots may be selected in advance using sampling design adopted under these surveys. The data required for forecasting purposes can be collected at required points of time from these plots and used for providing forecasts before harvest. The methodologies developed at IASRI can be more widely used throughout the country after pilot testing at other locations.

## References

Aditya, Kaustav (2008). Forecasting of crop yield using discriminant function technique. M.Sc. thesis, PG School, IARI, New Delhi.

Agrawal, Ranjana, Jain, R.C. and Singh, D.(1980). Forecasting of rice yield using climatic variables. *Ind. J. Agric. Sci.,* **50 (9)**, 680-684.

Agrawal, Ranjana and Jain, R.C. (1982). Composite model for forecasting rice yield. *Ind. J. Agric.Sci*., **52 (3)**, 189-194.

Agrawal, Ranjana, Jain, R.C. and Jha, M.P. (1983). Joint effects of weather variables on rice yields. *Mausam*, **34 (2)**, 177-181.

Agrawal, Ranjana, Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop weather relationship. *Mausam*, **37 (1)**, 67-70.

Agrawal, Ranjana and Jain, R.C. (1996). Forecast of sugarcane yield using eye estimate alongwith plant characters. *Biometrical Journal,* **38 (5)**, 731-739.

Agrawal, Ranjana, Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural input on agro-climatic zone basis. *Ind. J. Agric. Sci.,* **71 (7)**, 487-490.

Chandrahas, Bapat, S.R. and Singh, B.H. (1983). On pre-harvest forecasting of yield of sugarcane. *Agricultural Situation in India*, **38(5)**, 291-292.

Chandrahas (1984). Some investigations on sampling and prediction models in crop forecasting methodology. Ph.D. thesis, PG School, IARI, New Delhi.

Chandrahas, Aneja, K.G. and Singh, B.H. (1989). Tobacco yield forecast models based on principal components of biometrical characters and crop inputs. Project Report, IASRI, New Delhi.

Chandrahas and Narain, Prem (1992). Pre-harvest forecasting of yield of apple. Project Report, IASRI, New Delhi.

Chandrahas and Rai, T. (1998). Effect of errors of measurement on estimates of parameters of forecast model. *Ann. Agric. Res.*, **19(2)**, 139-143.

Chandrahas and Rai, T. (2001). Pilot study for developing Bayesian probability forecast model based on farmers' appraisal data on wheat crop. Project Report, IASRI, New Delhi.

Cochran, W.G. (1970). Some effects of errors of measurements on multiple correlation. *J. Amer. Stat. Assoc.*, **65**, 22-34.

Fisher, R.A. (1924). The influence of rainfall on the yield of wheat at Rothamsted. *Roy. Soc. London, Phil. Trans.*, **B**, **213**, 89-142.

Goyal, Gaurav (2003). Crop yield forecast using double sampling. M.Sc. Thesis, P.G. School, IARI, New Delhi.

Hendrick, W.A. and Scholl. J.C. (1943). Technique in measuring joint relationship - The joint effects of temperature and precipitation on crop yield. *N. Carolina Agric. Exp. Sta. Tech. Bull.*, 74.

Jain, R.C., Agrawal, Ranjana and Jha, M.P. (1980). Effect of climatic variables on rice yield and its forecast. *Mausam*, **31(4)**, 591-596.

Jain, R.C., Sridharan, H. and Agrawal, R. (1984). Principal component technique for forecasting of sorghum yield. *Indian J. Agric. Sci.*, **54(6)**, 467-470.

Jain, R.C., Jha, M.P. and Agrawal, R. (1985). Use of growth indices in yield forecast. *Biometrical Journal*, **27(4)**, 435-439.

Jain, R.C. and Agrawal, R. (1992 (a)). Probability model for crop yield forecasting. *Biometrical Journal*, **34(4)**, 501-511.

Jain, R.C., Agrawal, Ranjana and Singh, K.N. (1992(b)). A within year growth model for crop yield forecasting. *Biometrical Journal*, **34(7)**, 789-799.

Jain, R.C. and Ramasubramanian, V. (1998). Forecasting of crop yields using second order Markov Chains. *Jour. Ind. Soc. of Agril. Stat.*, **51(1)**, 61-72.

Jha, M.P., Jain, R.C. and Singh, D. (1981). Pre-harvest forecasting of sugarcane yield. *Ind. J. Agril. Sci.*, **51(11)**, 757-761.

Mehta, S.C., Agrawal, Ranjana and Singh, V.P.N. (2000). Strategies for composite forecast. *Jour. Ind. Soc. Agril. Stat.*, **53(3)**, 262-272.

Rai, T., Swarup, Onkar and Chandrahas (1998). Yield forecast model for wheat (Triticum aestivum) crop. *Ind. Jour. of Agric. Sci.,* **68 (4)**, 206-208.

Rai, T. and Chandrahas (2000). Use of discriminant function of weather parameters for developing forecast model of rice crop. Project Report, IASRI, New Delhi.

Ramasubramanian, V. and Jain, R.C. (1999). Use of growth indices in Markov Chains models for crop yield forecasting. *Biometrical Journal,* **41(1)**, 99-109.

Ramasubramanian, V., Agrawal, Ranjana and Bhar, Lal Mohan (2004). Forecasting sugarcane yield using multiple markov chains. Project Report, IASRI, New Delhi.

Saksena, Asha, Jain, R.C. and Yadav, R.L.(2001). Development of early warning and yield assessment models for rainfed crops based on agro-meteorological indices. Project Report, IASRI, New Delhi.

Sardana, M.G., Khosla, R.K., Ohri, N.K. and Mitra, P.C. (1972). Pre-harvest forecasting of yield rate of jute. *Jute Bull.* **35 (7** and **8**).

Singh, B.H. and Bapat, S.R. (1988). Pre-harvest forecast models for prediction of sugarcane yield. *Ind. J. Agril. Sci.,* **58(6)**, 465-469.

Singh, B.H., Jain, R.C., Madan Mohan and Bapat, S.R. (1991). Pre-harvest forecasting of yield of groundnut. Project Report, IASRI, New Delhi.

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

# Forewarning Crop Pests and Diseases: IASRI Methodologies

**S.C. Mehta, Ranjana Agrawal and Amrender Kumar**

## 1. Introduction

Agriculture has always played and will continue to play a dominant role in the growth of Indian economy in the foreseeable future. It represents the largest sector producing around 28 per cent of the GDP. Achieveing self sufficiency in food grains production, has ensured a high priority for agricultural sector in the successive development plans of the country. Agricultural production is grossly affected by attack of pests/diseases on various crops and the losses in food grains only are to the tune of Rs 90,000 cr each year in India. The prevention of such losses needs substantial consideration and accordingly forewarning of pests and diseases is essential for taking timely control measures. These are also required for assessing losses.

Pest/disease infestation in crops is highly influenced by meteorological factors. The weather based modeling for early warning of pest/disease infestation may provide appropriate tool for investigating and predicting pest/disease status.

Indian Agricultural Statistics Research Institute (IASRI) has played a prominent role in developing methodologies for forewarning of different aspects relating to Alternaria Blight, White Rust, Powdery Mildew and Aphid (Mustard), Aphid (potato), American boll worm, Pink boll worm, Spotted boll worm & Whitefly (Cotton), *Spodoptera litura*, Late leaf blast & Rust (Groundnut), Pyrilla, Early shoot borer & Top borer (Sugarcane), Pod fly, Pod borer, Sterility Mosaic & Phytophthora Blight (Pigeon pea), Fruit fly, Hopper & Powdery Mildew (Mango) and Gall midge (Rice) for selected centers. In this paper, various models developed at the institute, mostly weather based, using quantitative as well as qualitative data, are discussed.

## 2. Models Based on Quantitative Data

Models were developed for forewarning time of first appearance of disease/pest, time of maximum disease severity/pest population, maximum disease severity/pest population and age-wise/standard meteorological (smw) week-wise or year-wise pest population/ disease severity. Various types of models developed are as follows:

### 2.1 Between year models and

These models were developed using previous years' data. An assumption was made that the present year is a part of the composite population of the previous years and accordingly the relationships developed on the basis of previous years' data will be applicable for the present year. The forecast for pests and diseases could be obtained by substituting the current year data into a model developed upon the previous years.

#### 2.1.1 Regression models

The approach was attempted for forewarning aphid population in potato in various weeks.

The form of the model was

$$Y = f(X) + e$$

where Y **:** Aphid population of the current week,

f(X) **:** Functions of X (weather variables with appropriate lags)

e **:** Error term

Different functions were tried and the appropriate were cos, log and exponential. As an example, model for third week of December (Pantnagar) was

$$Y = 80.25 + 35.78\cos(6.81X_{22} + 8.03) + 40.25\cos(2.79X_{12} - 14.82)$$

where $X_{i2}$ (i =1, 2) **:** Maximum and minimum temperature lagged by 2 weeks

[Trivedi *et al.* (1999)]

The observed and predicted values are given in Fig. 1.

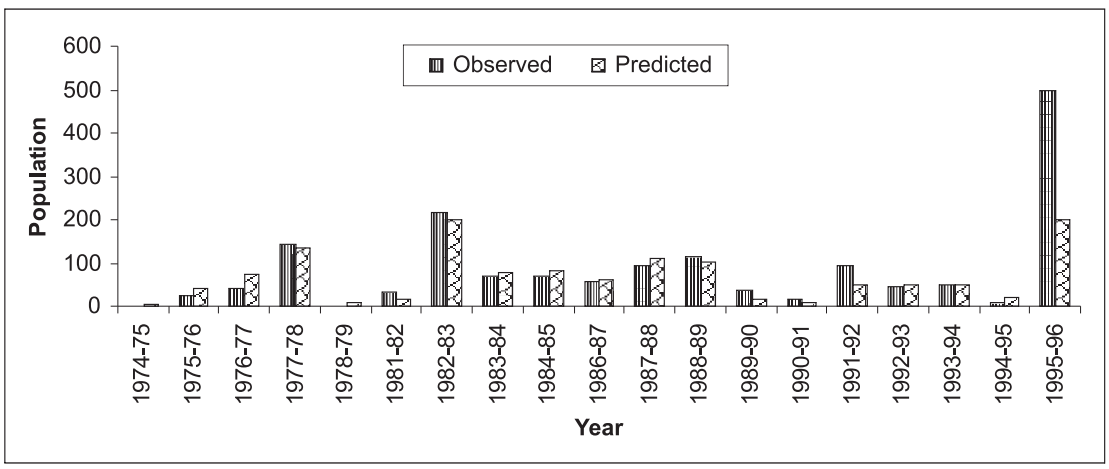



**Fig. 1:** Aphid population in third week of December at Pantnagar

### 2.1.2 Complex polynomial

For forewarning aphid population in potato, complex polynomials using group method of data handling (GMDH) technique were also fitted. The main feature of this technique is that it selects the structure of the model without using a prior information about relationship of dependent variable with independent variables.

The form of the model was

$$Y = a + \sum_{i=1}^{m} b_i X_i + \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} X_i X_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} d_{ijk} X_i X_j X_k + ....$$

where Y **:** Aphid population of the current week

X's **:** Weather variables with appropriate lags.

The polynomial for third week of December (Pantnagar) was

$$Y = 82.86 - 0.72\,G_1 - 0.35\,G_2 + 0.004\,G_1^2 - 0.002\,G_2^2 + 0.007\,G_1G_2$$

$$(R^2 = 0.95)$$

where

$$G_1 = 496.0 + 1123.30\,X_{21} - 1172.12\,X_{22} - 85.59\,X_{21}^2 + 46.25\,X_{22}^2 + 40.27\,X_{21}X_{22}$$

$$G_2 = -6359.0 + 477.80\,X_{11} + 42.93\,X_{32} - 12.65\,X_{11}^2 - 1.32\,X_{32}^2 + 2.87\,X_{11}X_{32}$$

$X_{ij}$ **:** $i^{th}$ weather variable (i=1,2,3 corresponded to maximum temperature, minimum temperature and relative humidity) in $j^{th}$ lag week. [Trivedi *et al.* (1999)]

Observed and predicted values from this model are given in Fig. 2.

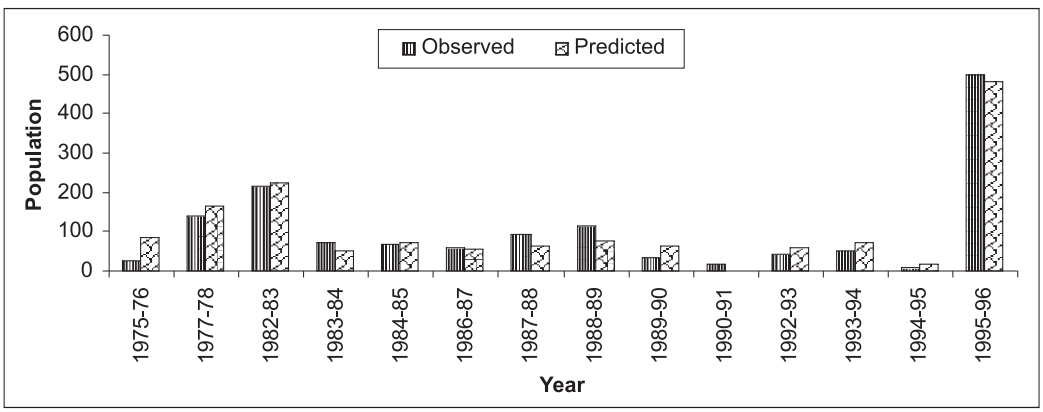



**Fig. 2 :** Aphid population in third week of December at Pantnagar

### 2.1.3 Weather indices based model

The model was used for forecasting various aspects viz. maximum pest population/disease severity, time of first appearance, time of maximum pest population/disease severity and weekly pest population/ disease severity for important pests and diseases of rice, mustard, pigeon pea, sugarcane, groundnut and cotton at various locations.

In this type of model, for each weather variable two indices have been developed, one as simple total of values of weather variable in different weeks and the other one as weighted total, weights being correlation coefficients between variable to forecast and weather variable in respective weeks. The first index represents the total amount of weather variable during the period under consideration while the other one takes care of distribution of weather variable with special reference to its importance in different weeks in relation to the variable to forecast. On similar lines, indices were computed with products of weather variables (taken two at a time) for joint effects.

The form of the model was

$$Y = a_0 + \sum_{i=1}^{p}\sum_{j=0}^{1} a_{ij}Z_{ij} + \sum_{i \neq i'}^{p}\sum_{j=0}^{1} b_{ii'j}Z_{ii'j} + e$$

where

$$Z_{ij} = \sum_{w=n_1}^{n_2} r_{iw}^{j} X_{iw} \quad \text{and} \quad Z_{ii'j} = \sum_{w=n_1}^{n_2} r_{ii'w}^{j} X_{iw} X_{i'w}$$

| | | |
|---|---|---|
| Y | : | variable to forecast |
| $X_{iw}$ | : | value of $i^{th}$ weather variable in $w^{th}$ week |
| $r_{iw}$ | : | correlation coefficient between Y and $i^{th}$ weather variable in $w^{th}$ week |
| $r_{ii'w}$ | : | correlation coefficient between Y and product of $X_i$ and $X_{i'}$ in $w^{th}$ week |
| p | : | number of weather variables considered |
| $n_1$ | : | initial week for which weather data were included in the model |
| $n_2$ | : | final week for which weather data were included in the model |

In some cases previous disease incidence / pest population (or their indices) and /or previous year's last population have also been included in the model. Stepwise regression technique was used for selecting important variables to be included in the model. The reliable forewarnings through this approach are possible at least one week in advance [Agrawal *et al.* (2004), Chattopadhyay *et al.* (2005-a), Chattopadhyay *et al.* (2005-b), Desai *et al.* (2004) and Dhar *et al.* (2007)].

For example, in mustard, models were developed utilising weekly data starting from standard week of sowing up to $50^{th}$ standard meteorological week i.e. second week of December for several years. The weather variables considered were maximum temperature, minimum temperature, maximum relative humidity, minimum relative humidity and bright sunshine hours - [$X_1$ to $X_5$] except for Morena & Bharatpur where data on bright sunshine was not available and therefore the models were based on only first four variables. For Kanpur, the variables considered were maximum temperature, minimum temperature, mean relative humidity, wind speed & evaporation - [$X_1$ to $X_5$]. The models for forecasting different aspects along with coefficient of determination ($R^2$) are presented in table 1.

Models have been validated using data on subsequent years not included in developing the models. In most of the cases, there was good agreement between forecasts and observed values. Results of validation are given in Fig. 3 to 5. [Chattopadhyay *et al.* (2005-a) and Chattopadhyay *et al.* (2005-b)].

**Table 1 :** Models for forecasting different aspects of aphid (mustard)

| Aspect | Center | Model | $R^2$ |
|---|---|---|---|
| crop-age at first appearance | Behrampur | $Y = -32.75 + 0.009\, Z_{121} - 0.73\, Z_{20}$ | 0.99 |
| | Pantnagar | $Y = 431.04 + 2.68\, Z_{31}$ | 0.56 |
| | Hisar | $Y = 56.72 + 0.029\, Z_{241} + 2.74\, Z_{11}$ | 0.60 |
| | Ludhiana | $Y = 198.12 + 0.14\, Z_{351} + 0.08\, Z_{451} - 0.011\, Z_{130}$ | 0.92 |
| | Morena and Bharatpur | $Y = -112.91 + 0.02\, Z_{131} + 4.98\, Z_{11}$ | 0.71 |
| | Kanpur | $Y = 7.96 + 0.42\, Z_{241}$ | 0.86 |

| | | | |
|---|---|---|---|
| crop-age at maximum population | Behrampur | $Y = 52.48 + 0.02\ Z_{121} + 1.58\ Z_{51}$ | 0.84 |
| | Pantnagar | $Y = 53.75 + 0.02\ Z_{120}$ | 0.56 |
| | Hisar | $Y = 102.19 + 0.30\ Z_{121} + 1.20\ Z_{50}$ | 0.85 |
| | Ludhiana | $Y = -133.56 + 0.09\ Z_{241}$ | 0.67 |
| | Morena and Bharatpur | $Y = 77.1 + 0.03\ Z_{121}$ | 0.63 |
| | Kanpur | $Y = 74.07 + 0.04\ Z_{231}$ | 0.92 |
| maximum population | Behrampur | $Y = 327.34 - 0.039\ Z_{230} + 0.06\ Z_{141}$ | 0.99 |
| | Pantnagar | $Y = 168.81 + 3.40\ Z_{241} + 101.22\ Z_{21}$ | 0.82 |
| | Hisar | $Y = 88.47 + 5.03\ Z_{121} - 0.33\ Z_{450}$ | 0.98 |
| | Ludhiana | $Y = -6114.29 + 9.46\ Z_{30} + 3.55\ Z_{451} + 113.07\ Z_{11} + 2.56\ Z_{131}$ | 0.95 |
| | Morena and Bharatpur | $Y = 690.77 + 79.29\ Z_{21} - 0.019\ Z_{340}$ | 0.65 |
| | Kanpur | $Y = 7531.96 + 882.83\ Z_{41} - 0.65\ Z_{130}$ | 0.75 |

N. B.: Morena & Bharatpur were taken together because these fall under same agro climatic zone.

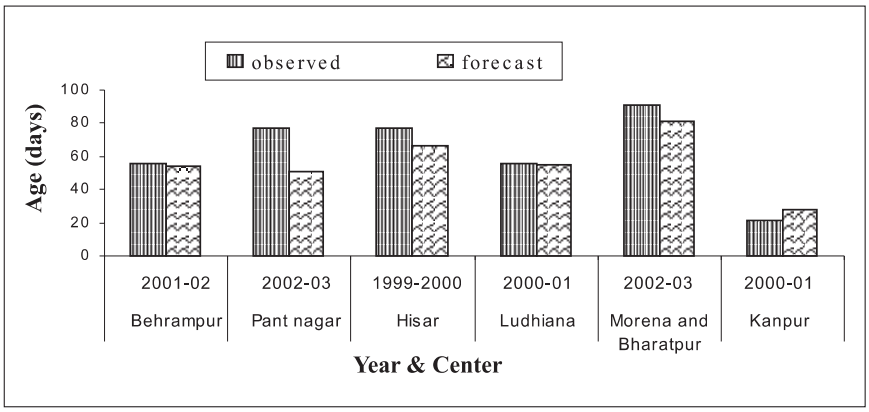



**Fig. 3 :** Crop age at first appearance of aphid

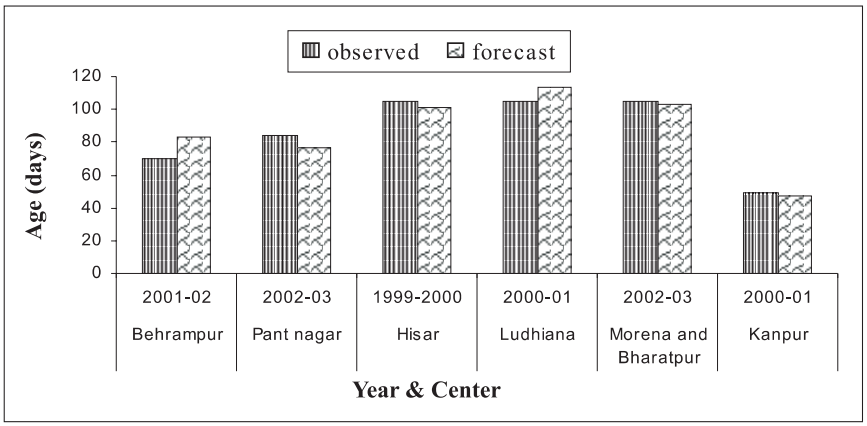



**Fig. 4 :** Crop age at maximum population

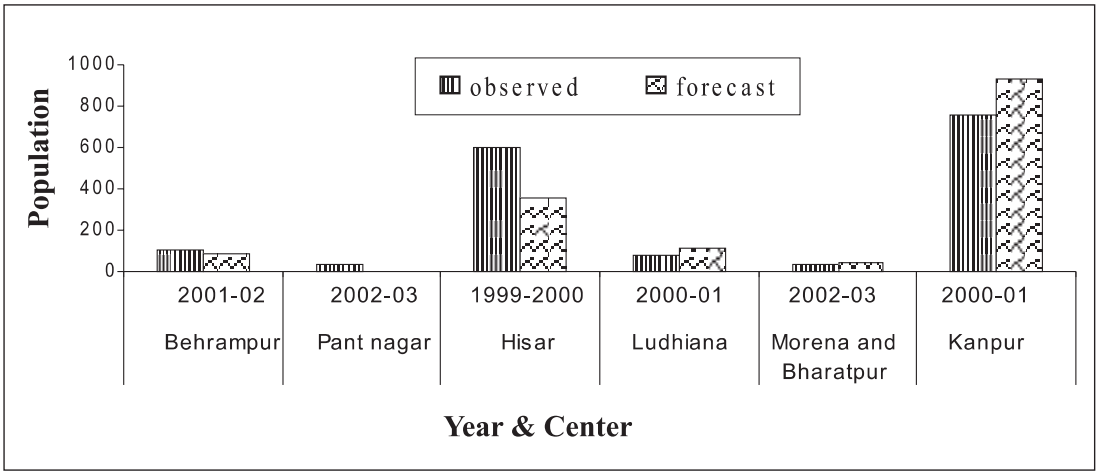



**Fig. 5 :** Maximum aphid population

The above models have been used to provide forewarning to the concerned farmers by National Research Center for Rapeseed Mustard, Bharatpur consecutively for three years which enabled them to optimize plant protection measures.

### 2.1.4 Model by deviation method

In situations, when data are available for few years (5-7 years) at different time intervals (say, weekly) which are inadequate for development of usual models, a methodology (Deviation Method) has been proposed. It has been assumed that the pest population/disease severity at any time is due to natural cycle of the pest/disease and the prevailing weather. To identify the natural pattern, data at different intervals can be averaged over years and a suitable model identified. Then a model can be fitted using deviations from natural pattern as a dependent and weather as independent variables.

The methodology has been illustrated using six years' available data for weekly fruit fly population in mango at Rehman Khera Farm, Central Institute for Subtropical Horticulture, Lucknow. [Mehta *et al.*(2001)]

The model for natural pattern was

$$Y_t = \frac{33.64 - 1.79t}{1 - 0.16\,t + 0.0067\,t^2}, \quad t : \text{Week}, \quad Y_t : \text{Fruit fly population at week } t.$$

The observed and predicted values are given in Fig. 6.

The final forecast model was

$$Y_d = -125.77 + 0.67\,Y_2 + 0.12\,(1/X_{22}^2) + 10.66\,X_{12}^2 + 0.0013\,Y_3^2 + 31.79\,(1/Y_3) - 21.32\,X_{12} - 2.15\,(1/X_{33}^2) - 1.75\,(1/X_{34}^2)$$

where $Y_d$ : Deviation of fruit fly population from natural cycle

$Y_i$ : Fruit fly population in $i^{th}$ lag week

$X_{ij}$ : Deviation of $i^{th}$ weather variable (from average), i = 1,2,3 corresponded to maximum temperature, minimum temperature and relative humidity in $j^{th}$ lag week.

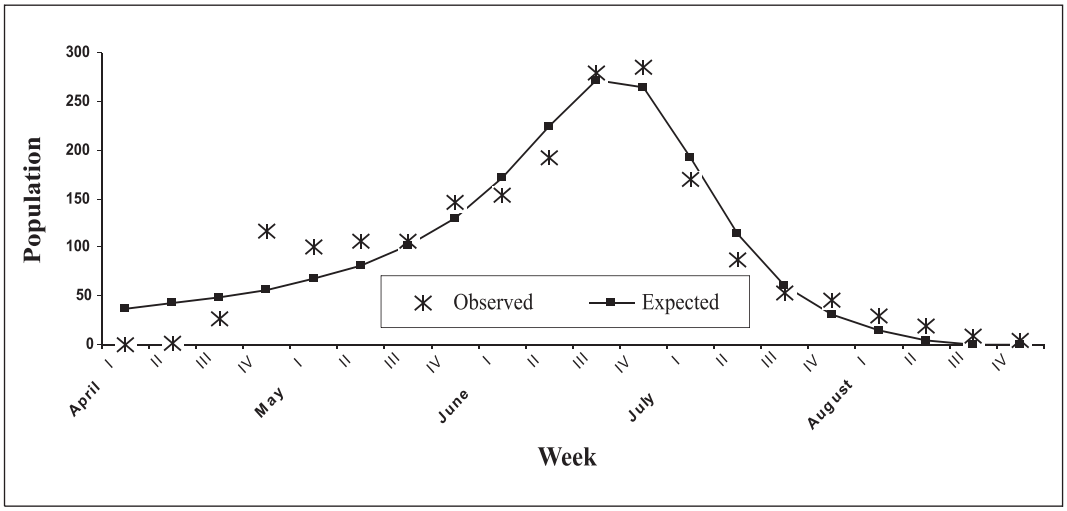



**Fig. 6 :** Natural pattern of fruit fly

Using this model, forewarning can be issued one week in advance. Forewarnings for subsequent year are given in Fig. 7

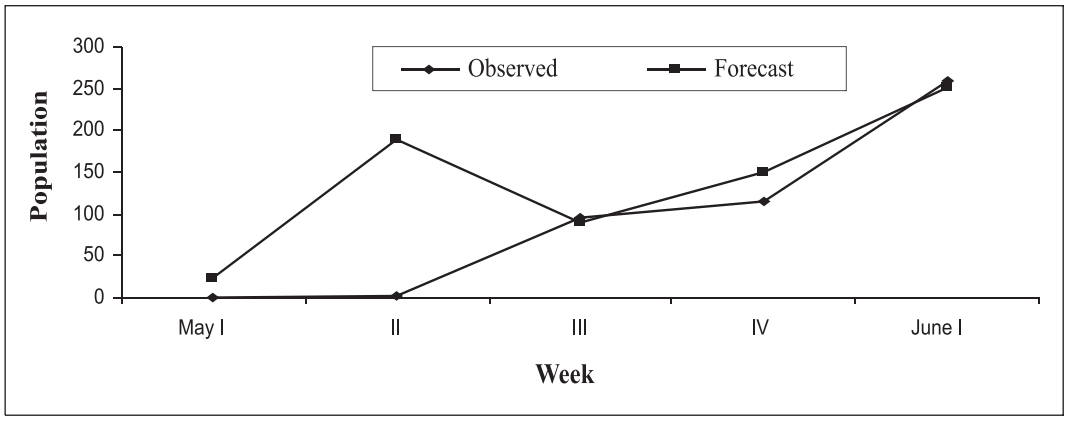



**Fig. 7 :** Forecast of fruit fly population

It can be seen from the figure that all the forecasts (except May II week) are in good agreement with the observed populations.

### 2.1.5 Use of artificial neural network (ANN) technique

ANN provides an efficient alternative tool for forecasting purposes. ANNs are data driven self-adaptive methods in that there are few apriori assumptions about the models for problems under study.

These learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe. After learning from the data, ANNs can often correctly infer the unseen part of a population even if data contain noisy information. As forecasting is performed via prediction of unseen part from examples of past behaviour, it is an ideal application area for ANNs, at least in principle. [Dewolf and Francl (1997 and 2000)]

This technique was applied for forewarning various aspects relating to different pests / diseases such as Alternaria Blight and Powdery Mildew (Mustard), American boll worm, Pink boll worm, Bacterial blight & Whitefly (Cotton), Leaf minor (Groundnut), Helicoverpa armigera (Pigeon pea) and Yellow stem borer (Rice) at selected centers. For instance, the model has been developed for forecasting time (crop-age) at first appearance of Powdery mildew in Mustard (S.K.Nagar), by using weather data (maximum temperature, minimum temperature, bright sunshine hour, relative humidity (morning and evening)), for the period 1999-2000 to 2005-06 with different dates of sowing. The model was validated on data for the subsequent year 2006-07. The observed and forecast values were in good agreement (Fig. 8).

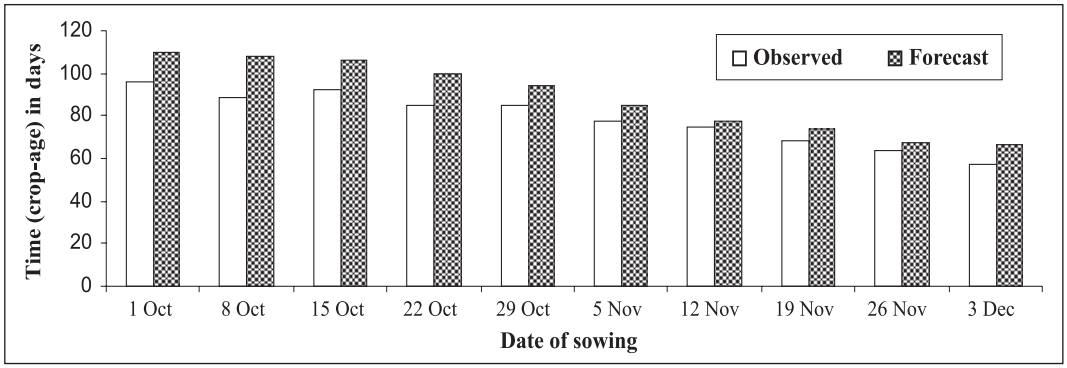



**Fig. 8 :** Forecast for time (crop-age) at first appearance of Powdery mildew (Mustard)

## 2.2 Within year growth model

In case historical data are not available and about 10-12 data points are available between time of first appearance of disease/pest and maximum disease severity/pest population, then forecast of maximum disease severity / pest population can be obtained on the basis of current season data using within year growth model. The technique comprises of fitting an appropriate model to the pattern of disease development / pest population using partial crop season data and forecasting the maximum value on the basis of this model.

This technique was used for forecasting percent disease severity (pds) of Alternaria Blight in Varuna variety of mustard at Kumarganj in the year 1999-2000 for different dates of sowing. The model was of the form

$$Y_t = A \exp (B/t) + e$$

where t **:** week after sowing, $Y_t$ **:** percent disease severity at week t,

A & B **:** model parameters.

Using this model reliable forecast could be obtained two weeks in advance. The observed, predicted and forecasts of maximum per cent disease severity are presented in table 2. [Agrawal *et al.* (2004)]

**Table 2 :** Observed, predicted and forecast for Alternaria Blight in Mustard

| Date of sowing | Observed max. pds | Predicted max. pds. (Full Model) | Forecast at lag | | |
|---|---|---|---|---|---|
| | | | 1 week | 2 week | 3 week |
| 27–09–99 | 73.88 | 75.15 | 69.69 | 69.07 | 65.02 |
| 04–10–99 | 75.60 | 75.60 | 75.66 | 76.68 | 79.28 |
| 12–10–99 | 70.62 | 66.83 | 63.98 | 73.47 | 79.57 |

## 3. Model for Qualitative Data

The timely control measures to prevent pest/ disease outbreak can be taken even if the information on the extent of severity is not available but merely the epidemic status is accessible. This information could be obtained through modeling qualitative data. Such models have added advantage that these could be obtained even if the detailed and exact information on pest count/disease severity is not available but only the qualitative status such as epidemic or no epidemic / low, medium or high is known. Such a situation arises quite often in pest/disease data.

The technique was applied for forecasting epidemic status of Alternaria blight & White rust (Mustard), Whitefly (Cotton), Pyrilla (Sugarcane) and Powdery mildew & Fruit fly (Mango).

Ordinal logistic models were developed to forecast probability of occurrence (Y=1) / non-occurrence (Y=0) of the pest/disease. In cases where the data were in quantitative form, the same were converted to dichotomous form using threshold values.

The form of the model was

$$P(Y=1) = \frac{1}{1+\exp(-L)} + e$$

where $L = \sum \beta_i X_i$

$X_i$ **:** weather variables/weather indices

$P < 0.5$ indicates that the probability of epidemic occurrence will be minimal

$P \geq 0.5$ indicates that there is more chance of occurrence of epidemic.

Different combinations of weather variables (maximum & minimum temperature, relative humidity (morning & evening) and mean relative humidity), along with their interactions, were tried in construction of function L. The combination that provided most of the prediction probabilities matching with the observed ones was identified. The results of fitting above model in different cases are presented in table 3. [Agrawal *et al.* (2004), Mehta *et al.* (2001) and Misra *et al.* (2004)].

**Table 3 :** Forecasting outbreak of pest and disease using Ordinal Logistic model

| Crop (Location) | Pest / Disease | Time of | |
|---|---|---|---|
| | | Damage | Forecast |
| Mustard (Hisar) | Alternaria blight | Feb. | Mid Dec. |
| | White rust | Feb. | Mid Dec. |
| Cotton (Lam) | Whitefly | Mid Dec. | Mid Nov. |
| Sugarcane (Muzaffarnagar) | Pyrilla | Oct.-Nov. | May |
| Mango (Lucknow) | Powdery mildew | May- June | 3rd week of March |
| | Fruit fly | May-June | 2nd week of March |

## 4. Conclusion

To sum up, different types of models were developed at IASRI for forewarning pests and diseases. The performance of the models was found to be good. The models provided timely forewarning of various aspects of pests / diseases.

The methodologies were used successfully by various workers and organizations. The models developed for mustard aphid have been used by National Research Center for Rapeseed Mustard, Bharatpur to provide forewarning to the farmers consecutively for three years which enabled them to optimize plant protection measures and save expenditure on unnecessary spray of chemicals.

## References

Agrawal, Ranjana, Mehta, S.C., Kumar, Amrender and Bhar, L.M. (2004). Development of weather based forewarning system for crop pests and diseases. Project Report, IASRI, New Delhi Mission mode project under NATP, PI, Dr. Y.S. Ramakrishna, CRIDA, Hyderabad.

Chattopadhyay, C., Agrawal, R., Kumar, A., Bhar, L.M., Meena, P.D., Meena, R.L., Khan, S.A., Chattopadhyay, A.K., Awasthi, R.P., Singh, S.N., Chakravarthy, N.V.K., Kumar, A., Singh, R.B. and Bhunia, C.K. (2005-a). Epidemiology and forecasting of Alternaria blight of oilseed Brassica in India – a case study. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz (Journal of Plant Diseases and Protection)*, 112(4), 351-365.

Chattopadhyay, C., Agrawal, R., Kumar, Amrender, Singh, Y.P., Roy, S.K., Khan, S.A., Bhar, L.M., Chakravarthy, N.V.K., Srivastava, A., Patel, B.S., Srivastava, B., Singh, C.P. and Mehta S.C. (2005-b). Forecasting of *Lipaphis erysimi* on oilseed Brassicas in India – a case study. *Crop Protection*, 24, 1042-1053.

Desai, A. G., Chattopadhyay, C., Agrawal, Ranjana, Kumar, A., Meena, R.L., Meena, P.D., Sharma, K.C., Srinivasa Rao, M., Prasad, Y.G. and Ramakrishna, Y.S. (2004). *Brassica juncea* powdery mildew epidemiology and weather- based forecasting models for India – a case study. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz (Journal of plant Diseases and Protection)* 111(5), 429-438.

Dewolf, E.D. and Francl, L.J., (1997). Neural network that distinguish in period of wheat tan spot in an outdoor environment. *Phytopathalogy*, 87(1), 83-87.

Dewolf, E.D. and Francl, L.J. (2000). Neural network classification of tan spot and stagonespore blotch infection period in wheat field environment. *Phytopathalogy*, 20(2), 108-113.

Dhar, Vishwa, Singh, S.K., Kumar, M., Agrawal, R. and Kumar, Amrender (2007). Prediction of pod borer (Helicoverpa armigera) infestation in short duration pigeonpea (Cajanus cajan) in central Uttar Pradesh. *Indian Journal of Agricultural Sciences*, 77(10), 701-704.

Mehta, S.C., Agrawal, Ranjana, Shukla, R.P. and Sharma, Shashi (2001). A Statistical model for prediction of mango fruit fly outbreak. Presentation in National Seminar on Agro-meteorological research for Sustainable Agricultural Production organized by Gujarat Agricultural University, Anand.

Misra, A.K., Om Prakash and Ramasubramanian V. (2004). Forecasting Powdery mildew caused by Oidium mangiferae in Mango using logistic regression models. *Indian Journal of Agricultural Sciences*, 74(2). 84-87.

Trivedi, T.P., Jain R.C., Mehta, S.C. and Bhar, L.M. (1999). Development of Forewarning System for Aphid, Myzus Persicae (Sulzer) on Potato. Annual Report, NCIPM, New Delhi, 43-49.

# Geo-informatics in Agricultural Research and Development: An IASRI Perspective

**Anil Rai, Prachi Misra Sahoo and Tauqueer Ahmad**

## 1. Introduction

Geo-informatics is integrated technology for collection, transformation and generation of information from integrated spatial and non-spatial data bases. Remote sensing, Geographical Information Sciences (GIS), Global Positioning Systems (GPS), Relational Data Base Management Systems (RDBMS) are some of its important ingredients. It is a powerful tool for assessment, monitoring, planning and management of agricultural research and development. Management of agricultural resources is a myriad activity of conservation practices and land/water resources aimed at increasing the food production. Substantial increase in crop production could be achieved by bringing additional land under cultivation, improved crop management technology through use of high yielding, input responsive and stress tolerant crop varieties, improved pest control as well as by increasing irrigation and fertilizer inputs. These inputs together with reliable information on i) existing land use and acreage under various crops, ii) soil types and extent of problem soils, iii) monitoring of surface water bodies (to determine water availability in irrigation systems) for ground water development and (iv) management of natural calamities etc. will enable formulation of appropriate strategies to sustain the pace of agricultural development. This in turn calls for a holistic approach, which must combine short-term management of agricultural resources at micro-level with long term global perspectives, keeping in view of socio-economic and cultural environment of the people. The role of space geo-informatics in finding new resources for agriculture development for optimally managing the already available resources in order to maximize agriculture production is recognized world wide and is found to be highly potential.

Agricultural remote sensing involving crops and soils are quite complex. These complexities are due to dynamic nature and inherent complexity of biological materials. In order to handle these complex problems, remote sensing technology offers numerous advantages over traditional methods of conducting agricultural and other resource surveys. Advantages include, the potential for accelerated surveys, capability to achieve a synoptic view under relatively uniform lighting conditions, availability of multi-spectral data for providing intense information, capability of repetitive coverage to depict seasonal and long-term changes and availability of imagery with minimum distortion etc. Therefore, it permits direct measurement of important agro-physical parameters. Remote sensing of earth resources utilizes electromagnetic waves, which ranges from short wave length ultra violet through visible near infrared and thermal infrared in the longer wave length, active radar and passive microwave systems. A great advancement in applications of computers to this science is the development of capability of storing vast and varied information, ranging from historical information and aerial photography to spacecraft data, ground reference, and other forms of ancillary data. All these information is stored in the form of highly useful database/information system. Thus remotely sensed data and its derived information have become an integral component of agricultural management system in the country.

Applications of space borne remote sensing data for large area crop survey was explored in USA under Corn Blight Watch Experiment (CBWE) in 1971 which was followed by large number of

experiments/large-scale remote sensing program. In a country like India, with vast geographic spread and great diversity in its set up, the need to apply remote sensing technology for national development was recognized during early 70's. The pioneering experiment was of coconut root–wilt disease using colour-infrared aerial photography. Numbers of studies were conducted for methodological development in this area. Remote sensing activities in India received a tremendous boost with the launch of Indian Remote Sensing Satellite-1A (IRS 1A) in March 1988. India is moving fast in development of new satellite systems and recently launched number of satellites dedicated to specific area of applications such as OceanSat, CartoSat, ResourceSat etc. Radar Imaging Satellite (RISAT), a microwave remote sensing mission with Synthetic Aperture Radar (SAR) operating in C-band and having a 6 x 2 meter planar active array antenna based on trans-receiver module architecture has been launched recently. Many technological developments, which occurred in 20th century contributed to the development of the concept of precision farming which includes GPS, GIS and high resolution remote sensing satellite data. In following sections, contributions of this institute are presented in this emerging field.

## 2. Contributions of IASRI

Indian Agricultural Statistics Research Institute (IASRI) was established on July 02, 1959 as an Institute of Agricultural Research Statistics. The mandate of the Institute is to undertake basic, applied and adaptive research in Agricultural Statistics, to conduct post graduate and in-service training courses in Agricultural Statistics and Computer Applications, to provide consultancy services, to act as a repository of information on Agricultural Statistics for research. Institute has been identified as an Advanced Centre of Excellence in education and training in Agricultural Statistics and Computer Applications. Apart from this institute also liaise with institutions of National Agricultural Research System (NARS), National Agricultural Statistical System (NASS), Department of Space (DOS) etc., to assist in the development and strengthening of quality of agricultural research and agricultural statistics through undertaking research and consultancy projects. Institute is celebrating the Golden Jubilee year (July 03, 2008 – July 02, 2009) of its foundation. This institute also recognized potential of geo-informatics technologies and initiated work in the direction of generation of crop production statistics since early nineties. During these years institute has undertaken number of research projects / studies and made significant contributions broadly in the areas of (i) Crop yield estimation (ii) Spatial stratification techniques (iii) Small area estimation (iii) Spatial sampling (iv) Spatial modeling (v) Classification techniques (vi) Integrated surveys for hilly regions and (vii) Web GIS. Some of important contributions of the institute are briefly described in following sub-sections

### 2.1 Crop yield estimation

Research on crop yield estimation has been taken up by the institute since beginning of nineties. Goyal (1990) demonstrated that remotely sensed satellite spectral data in the form of vegetation indices has been used to post- stratify the cropped area into area of homogeneous crop vigor and consequently improved estimators are proposed to estimate the crop yield using remote sensing data along with the ground enumerated yield estimation survey data. The usefulness of the suggested procedure has been demonstrated by using the Landsat (TM) satellite data and the crop yield data from yield estimation survey based on crop cutting experiments from Sultanpur district of Uttar Pradesh. Further, it has been observed that Normalized Difference Vegetation Index (NDVI) as compared to Ratio Vegetation Index (RVI) has higher potential to discriminate vegetation vigor and hence has the higher potential to be used

in crop yield estimation surveys. An attempt was also made to quantify the effect of misclassification of units on the size of post- strata and the post- stratified estimator of the crop yield. The expressions for the bias and variance of the post-stratified estimator have been derived in terms of the extent of misclassification. Since, spectral reflectance is a manifestation of integrated effects of all inputs like weather, soil and agricultural practices, therefore it is expected that the spectral data can be used with advantage for crop yield forecasting. Therefore, attempt was also made to study the relationship between wheat yield and the spectral parameters obtained through the hand- held spectral radiometers to explore the usefulness of spectral data in crop yield forecasting.

Singh and Goyal (1993), Singh *et al.* (2000) and Singh and Goyal (2000) extended the earlier results for estimation of wheat crop yield for district Rohtak, Haryana using crop cutting experiments data for the year 1995-96 and satellite spectral data from the Indian Remote Sensing Satellite IRS-1B LISS II data for February 17, 1996. Post stratified estimator of crop yield using spectral data in the form of vegetation indices NDVI and RVI for stratification have been obtained for the district. The efficiency of the post stratified estimator based on NDVI and RVI compared to the usual estimator comes to 1.42 and 1.28 respectively. This study thus almost confirm the findings of the earlier study that the district level estimator of crop yield may be obtained by reducing the number of crop cutting experiments to about 2/3$^{rd}$ without loosing the precision thus resulting in great savings of cost. Further, two small area estimators of crop yield, namely the direct estimator and the synthetic estimator have been developed at tehsil level using post stratification based on NDVI. The standard error of both the direct estimator and the synthetic estimator at tehsil level is within 5 per cent and as expected the synthetic estimator is more efficient as compared to the direct estimator.

Ibrahim (1992) investigated the utility of the multi-date spectral data taken at selected intermediate times in the growing season, in a Markov chain model to forecast crop yield. A hand-held spectral radiometer has been used for collecting the spectral responses from the experimental plots of wheat crop at fortnightly intervals during the growth of the crop. The spectral parameters observed at the different growth stages as well as the observed yield have been utilized to simulate a spectral population along with the corresponding yield based on a stochastic model.

Das (2004) proposed alternative approach of crop yield estimation using multi resolution satellite data. The attempts were made to make use of satellite spectral data and spatial sampling technique for crop acreage estimation, crop yield estimation and crop yield forecasting, which involves use of satellite data of coarse spatial resolution, which is cheaper with larger aerial coverage. It was shown that remotely sensed satellite data can be used effectively as area frame for conducting crop yield estimation surveys. Fine resolution data is costly and aerial coverage is less, whereas poor resolution data has larger aerial coverage with lesser cost. Since, spectral reflectance is a manifestation of integrated effects of all inputs like weather, soil and agricultural practices, it is expected to have a very high correlation with crop vigor and hence the crop yield. Since, estimation of yield using different spectral indices provides different estimates attempts were made to combine the estimators from different vegetation indices using multiple frames sampling estimation technique. It has been found that considerable gain in efficiency is obtained in multiple frame sampling estimates as compared to usual estimator using single index. A common approach for classification of satellite data is supervised maximum likelihood classification. In supervised maximum likelihood classification, each pixel is allotted with a particular theme depending on its digital

number. The presence of the mixed pixels is a nuisance with performing classification, because in the conventional classification procedures, a pixel is considered as an elementary unit for the analysis. An alternative method of satellite data classification is proposed, which involves the use of spectral signature curve for training the satellite images to estimate the underlying class signatures through the use of fuzzy classification developed by indicator Kriging. Attempt was also made to develop forecasting model based on spectral data and agro-meteorology.

### 2.2 Spatial sampling

In agricultural surveys often the parameter of interest is geographical in nature i.e. the observations are dependent through space thus classical statistics cannot be applied as such. Dependence implies correlation and spatial dependence implies the presence of spatial autocorrelation. Since, in geographical data 'adjacent units are often more alike than units that are far apart', it is desirable to exploit this information in the sampling designs. In this way duplicate information partly contained in areas already sampled can be avoided. Another advantage is that the sampling cost can be economized without loosing the reliability of the estimates. An attempt by Misra (2001) is being made to improve the conventional survey methodology for agricultural surveys with the help of spatial sampling procedures. The potential of GIS to handle various kinds of information through their geographic coordinates and Remote Sensing with its advantage of wide area coverage, repetitive coverage and synoptic view have been exploited for the study. An improved spatial sampling technique known as *Contiguous Unit Based Spatial Sampling (CUBSS)* Technique is proposed in this study (Sahoo *et al.*, 2006). The technique incorporates size measure along with spatial contiguity of the units in the population. The spatial correlation is estimated for auxiliary character which is used along with size measure in assigning weights for selection of the sampling units. The probability of selection of any unit is governed by these weights. The principle of sample selection is that the probability of selection of any unit increases as the distance from the units (area) already selected increases. The sample selection criterion is based on the weights, accounting for spatial variability and the size measure accounting for areal extent. Further, a suitable unbiased estimator which takes into account the order of the draw is suggested for this situation. The study is carried for regular lattice i.e. assuming the area to consist of regular units. In order to tackle the problem of irregularity of the sampling units, distance based neighbor are suggested. Based on these neighbors the modified formula for spatial correlation is also suggested in this study. For defining these neighbors, the concept of lagged variable and lagged series is being used. A spatial sampling technique termed as *Distance Unit Based Spatial Sampling (DUBSS)* is also proposed in this study and its efficiency is compared with the existing ones and CUBBS technique by carrying out a suitable simulation study. The proposed technique performs considerably better than all the other techniques.

Ranked Set Sampling (RSS) is found to provide better results than Simple Random Sampling (SRS) when ranking is easy and cost effective, especially in ecological and agricultural surveys. After critical review of literature it has been found that there is a hardly any technique of RSS, which takes into account spatial dependence of observation generated by spatial variables. Kankure (2007) made an attempt to develop spatial ranked set sampling methodology for the estimation of finite population mean. Four sampling designs were proposed which takes into account the spatial relationships of the areal sampling units in the population while selecting a sample. The proposed Spatial Ranked Set Sampling (SRSS) procedures involve the selection of ultimate sampling units in two stages. In the first stage Random

Spatial Clusters (RSC) of sampling units in the population are formed and in the second stage ranked set sample of specified size is being selected. Sample selection at the first stage is done by applying Dependent Areal-Unit Sequential Technique (DUST). This technique is based on giving different probabilities of selection to the sampling units in such a way that nearer units, or already selected units in the sample, get lesser probability of selection, while farther units get higher probability of selection. It is one of the desirable characteristics of a sampling design in spatially correlated population. The spatial component of the data is incorporated at this stage by dividing the entire population into Random Spatial Clusters (RSCs) by considering first phase units as the key units. The spatial clusters are formed on the basis of nearest neighborhood approach with respect to randomly selected units. Any particular unit will fall exclusively in a single spatial cluster, which has been formed by the key unit nearest to it on the basis of Cartesian distances calculated using latitudes (La) and longitudes (Lo) of the locations. Having selected the RSC, RSS was carried out using two approaches i.e. (i) the entire ranked set sample was selected independently from each RSC and (ii) different sets of the RSS are selected independently from different RSC. Thus, in this study some new and more efficient sampling techniques have been proposed which take into account the spatial correlation present in the geographical units. The results of the study point out that, in spatial surveys, a considerable gain in efficiency of the estimators could be achieved by using distance based sample selection strategies even when applying these for complex sampling schemes such as RSS. The complex algorithms involved in the selection procedure of distance based sampling strategies could be solved with the use of advanced computing and software.

It is well known fact that there is no objective methodology for estimation of area under different crops in North-Eastern states due to typical problems existing in these regions. The north-eastern states particularly Meghalaya, mainly consists of hilly region with thick forest cover. Besides this, the main problem is its undulating topography and non-accessibility of vast area. Further, the relative percentage area under the crops is very less. Mostly terraced farming and Jhum cultivation is practiced in these regions. Moreover, these areas particularly Meghalaya, are covered by clouds most of the time in a year. Thus it is difficult to get cloud free images of these areas. Therefore, use of remote sensing satellite data alone may not be able to provide reliable information. Further, there are no cadastral maps and village boundary maps existing for these regions. In contrary to other states, in north-eastern regions reliable information regarding total number of villages in each district/block is not available. Further, within a villages total number of farmers, number of fields owned by each farmers, crops grown by the framers etc. are also not available in village records. Thus, the traditional methodology of area estimation is not applicable in these regions. Keeping all this in view, a study (Sahoo *et al.*, 2008) was taken up by Indian Agricultural Statistics Research Institute (IASRI), New Delhi in collaboration with North- Eastern Space Applications Center (NESAC) Shillong and Space Applications Center (SAC) Ahmedabad in which methodology was developed using integrated approach of remote sensing, GIS and ground survey for estimation of area under winter paddy crop in Meghalaya. The satellite data of IRS 1D, LISS III sensor has been used for this study. Under this approach the area under paddy has been obtained by usual classification method. There are two major factors affecting the accuracy of the crop area as obtained from the classified satellite image in the hilly regions: (i) Due to undulating topography of the region, misclassification and topographic geometry, there may be large differences of area under crop in the image and actual area under crop on the ground which may also result in larger extent of misclassification errors (ii) The area under paddy crop falling under hill shades or valleys may not be exposed to the

satellite sensor, as satellite sensors are sun-synchronous. Further, small paddy fields are not detectable due to lower spatial resolution of the LISS-III sensors. In order to rectify the area under paddy crop due to undulating topography and misclassification errors, relationship between area under paddy in the classified image and actual area under paddy crop on the ground has been established. The area under paddy which has not been captured by satellite sensor due to hill shades and limitations of spatial resolution of the sensor has been rectified by a suitable sample survey in the buffer created along the National Highway/State roads in GIS environment. Suitable estimators were developed to estimate the area under paddy in this buffer zone. The vector layer of this buffer was overlaid on the satellite-classified image and the corresponding area from the image was extracted. Using these estimates, the area under paddy in the entire district was estimated.

A study was undertaken in Yamunanagar district of Haryana State (Ahmad *et al.*, 2003), to develop a GIS based technique for identification of potential agro-forestry areas. In this study, important factors responsible for growth of agro-forestry were identified, suitability index using Spatial-Analytic Hierarchy Process was constructed and was compared with the Composite development index. A new Objective Analytic Hierarchy Process (OAHP) procedure was used to develop suitability index for agro-forestry of each village. This study strives to find out the potential agro-forestry areas using Geographic Information System (GIS) and proposed OAHP leading to Objective Spatial-AHP to identify and rank areas that are suitable for agro-forestry, using the statistical techniques involved in the proposed OAHP and data contained in GIS maps. This study is useful for the farmers as well as planners. The study is helpful for the social, economic and environmental development of the study area. The study will also be helpful in future research in the area of agro-forestry for other regions.

### 2.3 Spatial modeling

Agricultural fields are Spatial in nature. If we consider productivity of the field with respect to certain crop provided other factors as constant. It does not change abruptly from one field to another. The change is very gradual and the neighboring fields have more or less same structure. In general crop-cutting experiments (CCE) are carried out for yield estimation in selected villages. It may be noted that from crop cutting experiment the estimation of production at lower level like Tehsil's, villages etc is difficult and prone to large error due to small sample size. The application of spatial statistics in agriculture to improve the prediction and estimation may be a useful attempt for small area levels. Spatial characteristics and CCE will help us in giving the better estimate and at lower level also. With the help of available literature in the field of spatial statistics, it is possible to apply suitable spatial models to predict the production surfaces i.e. values of production at each point of the map, of the target region. Field sizes of our country are fairly small, therefore it is more appropriate to utilize the spectral data as auxiliary information. A study was undertaken by Gupta (2002) to develop an integrated methodology for wheat crop yield estimation using the survey data on wheat crop yield from CCE along with the satellite spectral data in the form of vegetation indices i.e. Normalized Difference Vegetation index (NDVI). The use of remote sensing satellite data along with the crop yield data based on CCE can greatly improve the efficiency of crop yield estimators at small area level.

Spatial statistics is based on the assumptions that nearby units are in some way associated and tend to share similar attribute values. The classical statistical theory is based on the assumption of independence of observations. Thus, classical statistical method when applied to geo-referenced data, fail to capture the

spatial dependence of the data. It is unrealistic to assume that the information is location independent while analyzing the data of a character, which is geographical in nature such as productivity of crop, soil parameters and availability of ground water for irrigation. Spatial data is data related to the location of features along with information related to variables of interest. It is always desirable to utilize the knowledge contained in the sample to improve the estimation of spatial statistics. Bayesian approach takes into account any prior knowledge of parameter/variable of interest.

Gupta (2007) made an attempt to develop spatial prediction model under four different situations i.e. (i) using prior information about parameters which is non-informative for known and unknown variance (ii) using prior information about parameters which is informative for known and unknown variance, (iii) using prior information about parameters as natural conjugate prior for known and unknown variance and (iv) using fuzzy approach for linear interval model for vague characters under study. It was shown in this study through simulation that Bayesian regression analysis is always better then simple regression analysis. This may be due to the fact that information contained in the sample as well as about the parameter of the model has been utilized in the estimation procedure. Further, there is significant gain in the precision in case of geographical variables when spatial effects were taken in to account in the estimation procedure under Bayesian framework. It can be seen that variogram models plays significant role in capturing the spatial effect. Spatial Bayesian regression model performs better when spatial effects are incorporated through variogram models. The results obtained through exponential and spherical variogram models were found to be encouraging as compared to other models.

Measurement Errors (ME) in explanatory variables of classical regression model makes the estimators of regression coefficients biased and inconsistent. In case, variable of interest is geographical in nature, regression coefficients do not remain fixed over space and usual regression analysis takes no account of spatial location in its analysis. Therefore, a new technique called Geographically Weighted Regression (GWR) is used in which estimates of regression coefficients are based on local relation instead of global relations among spatial variables of interest. Estimation of regression coefficients when spatial explanatory variables with ME are fixed or random in GWR model is expected to provide efficient estimates as compared to corresponding usual regression model. An attempt has been made in this institute to propose a Functional Spatial Regression (FSR) model under ME and a Structural Spatial Regression (SSR) model under ME for estimation of regression coefficients in case of spatial variables. Explanatory variables under FSR model are assumed to be fixed while it is random in case of SSR model. Modified estimates of spatial regression coefficients were proposed following Ordinary Least Squares (OLS), Generalized Least Squares (GLS), Maximum Likelihood Estimation (MLE) and Method of Moment Estimation (MME) approaches in both model structures. It has been shown through spatial simulation, that proposed estimators are unbiased, consistent and comparatively more efficient than corresponding usual estimators. Four different approaches were followed to incorporate spatial effects in the proposed models (Jha, 2009).

A study was undertaken by Rai *et al.* (2004) in district of Lalitpur in UP due to the fact that this district has been observed to have considerable area under most of the land use classification categories. It has been observed in this study that quality of revenue records in the study area i.e. Lalitpur district is quite reliable for most of the usual nine fold classified land use classes. The statistics of land use classes were restricted to five broader classes, which can be identified by using single time digital data of Remote Sensing out of the above nine-fold classification can be easily obtained using RS. These statistics of land

use classes obtained through RS could be used as auxiliary information in spatial /non-spatial models to get reliable statistics of different classes. The above models can be used to predict the statistics related to these classes for non-surveyed area/villages of the districts. Hence, it is possible to develop reliable land use statistics at any smaller level i.e. panchayat/block/tehsils using above models. In order to take into account of spatial dependence of the neighboring units, the classical sampling technique approach is being modified such that the probabilities of selecting neighboring units, once a particular unit is selected in the sample, becomes less as compared to distant areal units. The best fitted spatial model for each class of land use was found to be different, depending on the spatial distribution of the land use class patches of land in the district. The prediction of area under different land use categories covered under nine fold classification based on satellite data using spatial model seems to be quite satisfactory.

Apart from these studies numbers of research papers were published in national and international journals by the researchers of the institute. Some of the important contributions in this regard are Singh *et al.* (1992), Singh and Ibrahim (1996), Singh *et al.* (2002), Sahoo *et al.* (2005), Sahoo *et al*. (2006), Rai *et al.* (2007), Ahmad *et al*. (2007), etc.

**References**

Ahmad, T., Singh R., Rai, A. and Kant, A. (2007). Model for prediction of area under Yamunanager district of Haryana. *Indian. J. Ag. Sci.,* **77(1)**, 43-45.

Ahmad, T., Singh, R.and Rai, A. (2003). Development of GIS based technique for Identification of Potential Agro-forestry area. Project Report, IASRI, New Delhi.

Das, S.K. (2004). Application of multiple frame sampling technique for crop surveys using remote sensing satellite data. A Ph.D. thesis submitted to P.G. School, IARI, New Delhi.

Goyal, R.C. (1990). Use of Remote Sensing Planning of Agricultural Surveys. Project Report, IASRI, New Delhi.

Ibrahim, A.E.I. (1992.). Use of remote sensing data in a markov chain model for crop yield forecasting. Project report IASRI, New Delhi.

Jha, D.N. (2009). A study an spatial regression model under measurement area framework. Unpublished Ph.D thesis submitted to PG. School. IARI, New Delhi.

Kankure, A.K. (2007). Small area estimation of crop yield at block level using satellite data. Unpublished Ph.D. thesis submitted to P.G. School, IARI, New Delhi.

Misra, P. (2001). Applications of spatial statistics in agricultural surveys. Unpublished Ph.D. thesis, P.G.School, IARI, New Delhi.

Sahoo, P.M., Rai, A., Singh, R., Handique, B.K. and Rao, C.S. (2005). Integrated approach based on remote sensing and GIS for estimation of area under paddy crop in north-eastern hilly region' *J. Ind. Soc. Agric. Stat.,* **59(2)**, 151-160.

Sahoo, P.M., Rai, A., Singh, R., Handique, B.K. and Rao, C. S. (2008). Development of Remote Sensing based Methodology for collection of Agricultural Statistics in Megahalaya. Project Report IASRI, New Delhi.

Sahoo, P.M., Singh, R. and Rai, A. (2006) Spatial sampling proceduse for agricultural surveys using geogeaphical infoemation system. *J. Ind. soc. Agril. Stat.* **60 (2)** 134-143.

Gupta, N.K. (2007). On spatial prediction modeling. Unpublished Ph.D. thesis submitted to P.G.School IARI, New Delhi.

Gupta, N.K. (2002). Applications of spatial models in estimation of wheat production in rohtak distrct of haryana. Unpublished M.Sc. thesis submitted to P.G. School IARI, New Delhi.

Rai, A., Gupta, N.K. and Singh, R. (2007). Small area estimation of crop production using spatial models. Mod. *Asst. Stat. Apl.*, **2(2)**, 89-98.

Rai, A., Srivastava, A.K., Singh, R., and Jain, V.K. (2004). A study of land use statistics through integrated modelling using geographic information system. IASRI, New Delhi Publication.

Singh, R., Goyal, R.C., Saha, S.K. and Chhikara, R.S. (1992). Use of satellite spectral data in crop yield estimation surveys. *Int. J. Rem. Sens.*, **13(14)**, 2583-2592.

Singh, R. and Goyal, R.C. (1993). Use of remote sensing technology in crop yield estimation surveys. Project Report, IASRI, New Delhi.

Singh, R. and Ibrahim, A.E.I. (1996). Use of spectral data in markov chain model for crop yield forecasting. *J. Ind. Soc. Rem Sens.*, **24(3)**, 145-152.

Singh R., Goyal R.C., Pandey L.M. and Shah S.K. (2000) Use of remote sensing technology in crop yield estimation survey-ii. Project report IASRI, New Delhi.

Singh, R. Semwal, D.P., Rai, A. and Chhikara, R.S. (2002). Small area estimation of crop yield using remote sensing satellite data. *Int. J. Rem. Sens.*, **23(1)**, 49-56.

Singh, R. and Goyal, R.C. (2000) Use of remote sensing technology in crop yield estimation surveys. Project Report, IASRI, New Delhi.

# Glimpses of Basic Research in Sample Survey at IASRI

**U.C. Sud, Prachi Misra Sahoo, Hukum Chandra,**
**Tauqueer Ahmad and Vijay Bindal**

## 1. Introduction

The Indian Agricultural Statistics Research Institute made a modest beginning in 1930 as Statistical Section of the Indian (then Imperial) Council of Agricultural Research. It subsequently evolved into a premier Institution in 1959. Since the very beginning one of the mandates of the Institute is to conduct research in agricultural statistics. The Institute has made a significant contribution in the field of experimental design, sampling methods, statistical genetics, bio-statistics, forecasting techniques, statistical economics and computer applications. Large numbers of research papers have been published in National and International journals of repute. Research papers published by the scientists/students of the Institute have been both basic and applied in nature. On the occasion of Golden Jubilee of the Institute it is appropriate to summarize the achievements made by the scientists/students from time to time in various areas. In what follows, we give an account of basic research carried out, in a chronological manner, in the Sample Survey at Indian Agricultural Statistics Research Institute.

The basic research in the Division of sample survey was carried out in all areas of sample surveys i.e. equal and unequal probability sampling, stratified sampling, cluster and multi-stage sampling, systematic sampling, ratio and regression methods of estimation, controlled selection and non-sampling errors. Besides, the scientists/students did basic research in newer areas of sample surveys like analysis of complex surveys which included regression and categorical analysis from survey data, variance estimation techniques, small area estimation, combinatorial aspects of sample surveys etc., conditional design based approach and model assisted approach of estimation. In many papers which appeared in earlier years the theory was illustrated with real data. Similarly, simulated as well as real data was used in large number of papers which appeared in later years.

## 2. Details of Basic Research in Chronological Order

The statistical efficiency of the interpenetrating sampling design in relation to the precision of the estimate and the cost involved was examined.

A general formula appropriate for the estimation of gains in precision due to stratification in a sub-sampling design from finite population was developed and illustrated on the yield data relating to sample surveys carried out in Delhi province.

The gain in efficiency achieved by avoiding the repetition of sub-units in a two-stage sampling plan was explicitly worked out. The percentage reduction in the within unit component of the variance was found to be very nearly equal to the over-all sampling fraction.

An attempt was made in the direction of measurement of non-sampling errors by the method of replicated samples. Necessary formulae were developed to give the estimates of the various components of non-sampling errors. Simple special cases of the replicated samples were discussed with illustrations from the surveys carried out under the auspices of Indian Council of Agricultural Research. Some of the

merits and limitations of method of replicate samples were examined.

It has been shown that for n>2, sampling with varying probability without replacement (PPSWOR) may not be more efficient than sampling with varying probability with replacement (PPSWR). The efficiency will depend on sample size, population size and size measure with which the units are selected.

Necessary theory of sampling with probability proportional to size when 2n psu's were selected from a stratum in two groups of n each, such that either the elements within a group were selected with pps and the rest with equal probability and the groups were selected with replacement. The theory was generalized to m groups. Efficient sampling systems were derived for two-stage sampling when the groups had overlapping psu's.

Efficiency of cluster sampling was examined in relation to two -stage sampling. Specific situations were discussed enumerating the merits of one method over the other.

The method of analysis of variance was used for studying the variance components of two/three-stage sampling designs.

The double sampling scheme was extended to multi-stage sampling design in the context of ratio method of estimation. Numerical illustrations were given in support of the theory.

A suitable sampling methodology was developed whereby the risk of drawing non-preferred samples was minimum but at the same time the sampling scheme confirmed to the principles of random sampling procedures.

The problem of estimation of population mean using ratio method of estimation when information on auxiliary character was unknown involving multi-stage sampling design was extended to the case of several auxiliary variables for a three stage sampling design. Optimum weights were determined. An estimate of variance was also developed. An empirical study was carried out using survey data on guava.

A more general form of an estimate of the variance of customary stratified simple random sample based on one sample unit per stratum was suggested.

It was shown that the usual ratio and regression estimators in controlled simple random sampling were at least as efficient as those in simple random sampling.

A more general result regarding non-negativity of Yates and Grundy's estimate of variance of Horvitz-Thompson estimate of the total was obtained.

A sampling scheme was suggested which not only minimized the selection probability of non-preferred units but also provided efficient estimators than the usual PPS sampling scheme

A slight modification of the usual systematic sampling procedure in the presence of a linear trend in the population was found to be very effective in reducing the error variance of the estimator for the population mean. The efficiency of the modified design in the presence of random, quadratic and periodic trends was evaluated in comparison with usual systematic methods. The adaptability and suitability of the modified systematic sampling procedure was illustrated by application to a survey for estimating the milk yield.

The problem of estimation of population mean in successive sampling under a multi-stage sampling design was considered. The optimum replacement problem was studied.

A selection mechanism was given by means of which one could easily construct a controlled simple random sampling design to draw a sample of any size from a population of any size.

Different estimators of annual milk production were developed. These were used to build estimates of milk production out of a survey carried out in Punjab and Eastern Uttar Pradesh.

A general method of collapsing any number of strata and the use of an ancillary character for collapsing was suggested. The same ancillary character was used in improving the estimate of the parameter under study. The case when the population mean of the ancillary character was unknown was also considered. Double sampling approach was proposed in such a situation

A comparison of estimates of population mean under a two-stage sampling design for two different sampling patterns i.e. when a fraction 'p' of the primary sampling units (psu) with their samples of second stage units (ssu) were retained from previous occasion to the current occasions and a fraction 'q' of the psu's were selected afresh and when all the psu's in the sample on the previous occasion were retained on current occasion but only a fraction 'p' of the sample of ssu's within each psu was retained and a fraction q of the ssu's were selected afresh. Estimators were developed for the current occasion, change over two occasions.

Relative efficiencies of the estimates obtained under different sampling patterns were studied in two-stage sampling on successive occasions. Relative efficiencies were also studied of some alternative replacement procedures when sampling was carried out on three occasions.

Reduction in bias of the ratio method of estimation of population mean was achieved by transformation of origin and scale of the measurement of the auxiliary variate.

It was shown that the Horvitz-Thompson estimator under Midzuno-Sen sampling was always inferior to that of Rao, Hartley and Cochran in all practical situations where PPS sampling was to be preferred. It was found to be worse than even the PPS sampling with replacement estimator in many situations. Some numerical examples were considered for illustration purpose.

The results obtained on controlled sampling with equal probabilities and without replacement were consolidated. A practical and convenient procedure for controlled simple random sampling was proposed.

Minimum variance linear unbiased estimates have been developed for population mean on the most recent occasion, change in the population mean from one occasion to another and overall estimate of population mean overall occasions under a two stage sampling design in which primary stage units were partially replaced and the second stage units in the retained primaries were kept fixed. The entire study was made under a general correlation pattern.

A PPSWOR technique was suggested which ensured sampling with probability exactly proportional to size, and facilitated the calculation of joint inclusion probability of two units for any sample size from a finite population.

A new approach to controlled selection, which ultimately lead to the method of controlled simple random sampling, was suggested which simplified procedures of controlled selection.

An estimation procedure was developed in two-stage successive sampling on the h-th occasion where primary as well as secondary units were partially replaced. Some corrections were made in a similar paper published earlier by different authors.

Necessary theory of non-response in successive sampling on two occasions was developed.

An estimation procedure was suggested which made use of multi-auxiliary information when information on some of the auxiliary variables was not available for the whole population, instead it was known for some parts of the population.

A new systematic sampling procedure was suggested which provides the unbiased estimator of sampling variance, besides maintaining simplicity. On comparing the efficiency of the suggested procedure with the usual systematic sampling and simple random sampling it was observed that in situations where usual systematic sampling performs better than simple random sampling the suggested procedure also leads to the similar results and for some situations it may provide better results than even usual systematic sampling.

A method based on PPS systematic schemes was proposed whereby 'm' unit were selected with PPS systematic sampling with revised probabilities and 'n-m' units with SRSWOR.

The general problem of regression analysis was studied when the population was finite. Expressions for the variance of the regression predictor were obtained depending on how the value of the auxiliary variate was chosen. It was observed that there was a component in variance of prediction which depended upon the sampling procedure used for the selection of the sample.

A method based on PPS systematic sampling, was proposed which selected large number of units through a specific procedure and then from these large number of units the ultimate sample of n units was selected by SRSWOR.

Necessary theory for estimation of population mean was developed using unbiased two-auxiliary variate ratio estimators. Two cases were considered i) when the population means of auxiliary variables were known ii) when the population means were unknown. The theoretical results were supported with empirical results.

Role of Sampling in prediction was examined. The expressions of variance of prediction were obtained for various situations and relative efficiencies were compared.

A procedure was suggested so as to make cluster totals almost equal. With the help of empirical study it was shown that the suggested procedure was more efficient than SRS. An added advantage was that the suggested procedure was simple, objective and convenient.

The limits on probability under which usual sampling model is to be prepared over binomial sampling model have been worked out

Using one auxiliary variable for selection and the other for estimation, the PPS and ratio estimator were suitably combined for estimation of population mean. The proposed estimator was shown to be more efficient than either PPS estimator or ratio estimator under PPS sampling scheme.

A modified ratio estimator was considered by utilizing the relationship between the study variable and the auxiliary variable. The efficiency gains of modified ratio estimator vis-à-vis the regression estimators were demonstrated theoretically as well as empirically.

Using multi-auxiliary information an efficient estimator was developed which utilized cluster sampling in conjunction with ratio/regression estimators. Besides theoretical results, some empirical results were obtained.

Almost unbiased ratio type estimators were considered for different sampling schemes. Efficiency comparisons were made. Usefulness of the sampling schemes was illustrated with the help of real data on price of vegetables.

A criteria based on information on several auxiliary characters was suggested for the purpose of selecting units with unequal probability sampling with replacement. Efficiency advantage of an estimator of total based on this criterion was demonstrated over the usual PPSWR based estimator. An empirical study was carried out for illustration purpose.

It is well known that the ratio estimator of finite population mean/total is unbiased under Lahiri scheme of sampling. A sampling scheme was developed in the institute under which the regression estimator of finite population mean is unbiased. A sampling procedure was given which provided an unbiased regression type estimator.

The problem of estimation of frequency distribution for the current occasion using different sampling designs under successive sampling approach was attempted. Cluster sampling design with PPS sampling was found advantageous for such situations.

Using modified form of sequential sampling, a class of acceptable estimators were developed with varying sample size.

A method was proposed which involved splitting the sizes of the population units arranged in ascending order. The splitting formed columns. The splitting was carried out in such a manner that there were at least n non-zero values in each column and each pair of units had a non-zero value in at least one column. A column was selected with probability proportional to column totals. From the selected columns n units were selected with SRSWOR.

Adopting size stratification when the auxiliary character was approximated by a continuous uniform distribution, a double sampling ratio strategy was suggested and compared with some of the known strategies. The applicability of the strategy to sampling on two successive occasions was also investigated.

A general class of unequal probability sampling schemes for selecting two units without replacement was given. Many known sampling schemes were shown to be members of this class. It was shown that for any member of this class, the Yates-Grundy estimator for variance of the Horvitz-Thompson estimator was always positive.

Two sampling schemes using auxiliary variables for cluster sampling were proposed. In one of the schemes, the auxiliary variable was used in selecting the clusters whereas in the other the elements within the clusters were selected using the auxiliary variables. The efficiencies of the proposed sampling schemes as compared to the conventional cluster sampling with unequal probabilities of selection as well as SRSWOR were worked out empirically for natural as well as constructed populations.

A sufficient condition that the variance of Horvitz-Thompson estimator for inclusion probability proportional to sizes sampling scheme of selecting two units was uniformly smaller than that of Rao, Hartley and Cochran estimator was obtained.

A simple method of inclusion probability proportional to sizes was proposed for samples of size three units. It was shown that the variance of the Horvitz-Thompson estimator based on the proposed sampling scheme was uniformally smaller than that of the customary estimator used in probability proportional to

sizes with replacement sampling. Further, the performance over Rao-Hartley_Cochran and Sampford sampling schemes was studied empirically for some natural populations.

An alternative to the usual regression estimator for a population mean in double sampling was suggested on the basis of a preliminary test of a simple hypothesis about the auxiliary variate-mean. Two-phase sampling was assumed from a bivariate normal population. The gain in efficiency was investigated theoretically and empirically.

A method of forming clusters of two units and their selection in such a way that the units in a cluster did not fall beyond a pre-assigned distance 'd' was suggested. The selection procedure ensured sampling without replacement. The results were empirically illustrated.

Sampling with partial enumeration was defined in the context of multi-character surveys. Efficiency of sampling with partial enumeration was examined vis-à-vis the usual two phase sampling for the study of two characters.

An attempt was made towards optimal use of several auxiliary variables in the form of a single auxiliary variable obtained as a linear function of the several auxiliary variables. The performance of the condensed auxiliary variable was studied in selecting the sample.

A sampling procedure in unequal cluster sampling for fixed sample size, where the number of units in the initial sample of selected clusters exceeds the planned size of units was proposed. A scheme for discarding the excess number of clusters from the initial sample of clusters was also presented. It was shown that the suggested procedure, taking into account simplicity and practical feasibility, could be used in practice in uni-stage unequal cluster sampling designs.

An inclusion probability proportional to size sampling scheme for selecting sample of two units was suggested. The sample selection ensured the non-negativity of Yates Grundy's variance estimator of the Horvitz –Thompson variance of the estimator . For a number of natural populations, the proposed sampling scheme performed better, in terms of smaller variance, than many of the existing unequal probability sampling schemes.

The theory of multiple frame surveys was developed for multi-stage sampling designs in the context of estimation of finite population mean. Efficiency gains vis-à-vis single frame situation were studied under different cases i.e. multiple frame at first stage and single frame at second stage, single frame at first stage and multiple frame at second stage and multiple frames at both the stages.

A ratio type estimator involving two auxiliary variates was suggested. The suggested estimator was found to be more practicable then the already available estimators in the sense that it was always positive.

A sampling scheme was proposed which involved creating a sampling design to ensure inclusion probability proportional to size selection. The sampling design was created through constructing a sample space in which the population units occurred in proportion to their sizes and every pair of units occurred in such a way as to provide a non-negative and stable variance estimator. The selection procedure was extremely simple and the values of the inclusion probabilities could be computed easily. Some examples were considered.

A modified systematic sampling scheme was suggested by combining the concept of random interval with the use of unequal selection probabilities. The suggested scheme had the merit of simplicity and it

was also possible to estimate the variance of the estimator. Empirical comparison with SRS and usual circular systematic sampling indicated that the suggested scheme could be used as an alternative to systematic sampling.

A procedure was given in the context of post-stratification in uni-stage unequal cluster sampling on the basis of the elements of the selected clusters. The results obtained were empirically studied.

An approach was developed to estimate population and sample sizes when units were selected sequentially. Both with and without replacement procedures were considered.

The efficiency of ratio estimator under size stratification was worked out depending upon the size character when it was discrete. Efficiency comparison was made with some of the well known sampling strategies. It was established that the stratified ratio sampling strategy performed satisfactorily.

A stratified ratio sampling strategy using discrete ancillary information was suggested. Relative efficiency comparisons revealed that the stratified ratio sampling strategy performed satisfactorily.

A $\pi PS$ sampling was proposed by modifying the usual PPS systematic sampling. The proposed sampling scheme was compared empirically with the existing sampling schemes. The empirical results were promising.

The problem of estimation of population mean was considered for the case of sampling on two occasions when a fixed proportion of clusters of units were drawn on the first occasion and retained on the second occasion. Optimum allocation was considered under a suitable cost function. The efficiency of matching of clusters of units was examined in relation to the matching of an equivalent simple random sample of units. The theoretical results were empirically illustrated.

The effect of departure of optimum values of sample size and weight was studied in the context of estimation of finite population total under a multiple frame set up. The estimator was found to be robust with regard to departure from optimum values of the weights but not with regard to departure from optimum values of sample sizes.

The method of Balanced Repeated Replication (BRR) was extended to general sample design with arbitrary number of selections from each stratum. It was shown that mixed orthogonal arrays of strength two, or equivalently equal frequency orthogonal main-effect plans for asymmetrical factorials gave a set of balanced sub-samples useful in variance estimation.

The problem of estimation of finite population total in the context of sampling from two-dimensional populations spread over space and time was examined. Combinations of equal probability sampling and systematic sampling over time dimensions were considered. A scheme with double sampling and systematic sampling was also considered. Efficiency gains of the proposed procedure were studied empirically.

Narain's necessary condition for without replacement sampling to have smaller variance than with replacement sampling was modified.

The problem of estimation of finite population mean was attempted when the underlying population followed a simple linear model involving single auxiliary variable Two situations were considered i.e. i) the line of regression passes through origin and the variance of error terms increases with the increase in size of the auxiliary variable ii) there was an intercept term in the model and the variance of error terms

was constant. The errors terms were assumed to be uncorrelated. The method of double sampling was proposed when the information on the auxiliary character was assumed unknown. The efficiency gain of double sampling based predictor over a predictor which did not utilize auxiliary information were demonstrated under a suitably chosen cost function.

Suitable estimators were proposed for estimation of population mean for study of two characters from a two-stage sampling design when data were missing for one or both the characters from some preliminary stage units as well as some second stage units. The theoretical results were illustrated empirically.

A suitable method of estimation was proposed in the context of sampling from imperfect frame and a geographical ordering of units could be established.

An efficient method of estimation was proposed for the following two situations i) when the planned sample size was fixed but the realized sample size was a random variable ii) when the planned sample size itself was random variable. Theoretical results were supported with empirical results.

A sampling scheme was proposed under which the usual regression estimator was unbiased. The proposed scheme had not only the advantage of simplicity but it was found to be efficient than the conventional strategies under use.

Estimation theory for sampling from two dimensional populations with various sampling procedures along space and time dimensions were considered. Both aligned and unaligned sampling for various situations were investigated. While simple random sampling and two-stage sampling were considered along space dimension, systematic sampling was considered along time dimension.

A supplementary randomised response strategy was proposed. Rules for selection of design parameters were obtained. The relative efficiency of the supplementary information quantitative randomised response model over the optimized model was worked out.

Estimators of lactation yield were developed when the lactation records were incomplete. The performance of the proposed estimators were compared empirically using the data collected in Madras in 1980 under a survey entitled “A study of impact of milk supply schemes on rural economy in milk collection areas of Madhavaram milk supply schemes”.

A supplementary randomised response strategy was proposed. Rules for selection of design parameters were obtained. The relative efficiency of the supplementary information quantitative randomised response model over the optimized model was worked out.

A regression model which takes into account the clustering effect was considered. The effect of misspecification of model when the clustering effect was ignored was examined under some suitable cost function.

A sampling scheme for post stratification in two stage sampling was developed. Efficiency gains of an estimator of population mean based on the proposed scheme vis-à-vis an estimator based on the conventional unstratified two-stage procedure were demonstrated empirically.

Using Projective Geometry Approach, Minimum Variance Linear Unbiased Estimators of population parameters were proposed in sampling on two occasions for simultaneous estimation of two characters for levels, change and average over time.

The effect of correlated measurement errors on Ordinary Least Squares estimator was studied when both the variables were subject to measurement errors. The expressions for relative bias and relative mean-square error of regression estimates were derived.

Three variance estimation techniques namely Taylor Series Linearization, Jacknife Repeated Replication and Balanced Repeated Replication were compared in the context of cell proportions through the use of combined ratio estimator. The bias of Taylor Series Linearised estimator was found to be minimum.

Estimators were developed using randomized response models for binary and discrete data. It was shown that the unrelated question randomized response model under SRSWOR performed better than open interview with nominal untruthful reporting of order 5%.

Using available small area models composite estimators were developed. The performance of the proposed estimators were studied empirically.

It was shown that mixed orthogonal arrays of strength two are balanced subsamples needed for variance estimation.

A variance estimator for linear as well as non-linear estimator based on proportional frequency plan was proposed. The proposed variance estimator was shown to be asymptotically consistent.

An unbiased multiplicity estimator was considered in the context of overlapping clusters. The merit of the proposed estimator was studied under a specific situation.

Minimum Variance Linear Unbiased Estimator of any parametric function was developed using Hilbert space method for the case of sampling on h-occasions for univariate as well as multivariate populations.

A preliminary test estimator was proposed to test the significance of Ordinary Least Squares estimator of regression coefficient and the probability weighted estimator. The model based properties of the estimator were studied theoretically while the design based properties were studied through simulation approach. The proposed estimator was found to be better compromise between model based and randomization based inferential framework.

The various modified Chi-square test statistics were evaluated with respect to their actual size of the critical region through empirical investigation. It was observed that the second order corrected statistic based on Satterthwaite approximation performed better in terms of its actual size of the critical region as compared to the nominal level of the critical region at 5%.

Robust estimators of population mean were developed when there were outliers and the assumed model was wrong. The advantages of the proposed robust estimators vis-à-vis already available robust estimators were studied through a simulation study.

The problem of estimation of finite population mean for sampling on two occasions has been considered when the variables were subject to measurement errors. Efficient estimators have been developed for the current, previous and sum over the two occasions.

Empirical investigations on jackknife variance estimation under two-phase sampling when the first phase sample was used for stratification were carried out through simulation study. Under certain conditions, the jackknife variance estimator of reweighted expansion estimator performed satisfactorily

The problem of estimation of finite population mean square for the current occasion was considered for the situation of sampling on two occasions. Estimator of finite population mean square has been developed utilizing data of the two occasions. This estimator was shown to be better than an estimator which utilized only the current occasion data. However, the efficiency gains were less as compared to the case when the parameter was finite population mean.

A combined estimator of population mean under successive sampling which allowed one to take account of available auxiliary information in the matched portion at the estimation stage by means of conditional bias adjustment. The merit of the proposed estimator was studied through a simulation approach.

A new bootstrap technique of variance estimation for dealing with missing observations was developed. An optimum choice of bootstrap sample size in case of missing observations was obtained. The relative efficiency of the proposed technique was studied through simulation approach.

The problem of estimation of population mean when there was partial matching among first stage as well as second stage units was considered. The performance of both linear and non-linear estimators were studied empirically.

The problem of estimation of finite population regression coefficient was considered when the information on design variable was not available. The double sampling approach was suggested in this case. The efficiency of the double sampling based estimator of regression coefficient was examined vis-à-vis a probability weighted Ordinary Least Square (OLS) estimator.

A general result has been obtained whereby it was possible to obtain Minimum Mean Square Linear estimator which belonged to a particular class of biased estimator. The general result has been extended to the case of sampling on more than two occasions.

Necessary theory for estimation of finite population total in case of dual frame surveys has been developed when the domain sizes were assumed unknown.

An integrated approach for estimation of area under paddy has been developed using remote sensing and GIS.

The necessary theory for estimation of finite population regression coefficient in the context of sampling on two occasions was developed.

The bootstrap techniques were compared through simulation approach when the missing observations were imputed through different techniques

The Hansen and Hurwitz technique of obtaining response from the selected non-respondents has been extended to the case of sampling on two occasions for estimation of finite population mean on the current occasion. Three different situations have been considered viz. non-response on the first occasion, non-response on the second occasion and non-response on both first and second occasions. Efficiency gains of the proposed estimator were considered under a suitable cost function.

A finite population framework has been developed in the Institute for the method of Rank Set Sampling (RSS). This framework was further used to develop the method of RSS for two-stage sampling designs with the objective of estimating the finite population mean. Different situations were considered i.e. selection of first stage units by RSS and second stage selection using Simple Random sampling (SRS),

SRS at the first stage and RSS at the second stage and RSS at both the stages. Efficiency gains of RSS based estimators over SRS based estimators were demonstrated empirically with the help of real data.

The mixed model approach of estimation of population mean for the small area was used to develop district level estimates of the parameter amount of loan outstanding using 59$^{th}$ round of National Sample Survey Organization (NSSO) data for the rural areas of Uttar Pradesh. The performance of the Empirical Best Linear Unbiased Predictor (EBLUP) was studied vis-à-vis the direct district level estimates on the basis of the criterion of mean square error. The covariates used in this study were obtained from the 1996 Agriculture Census data.

A framework for estimation of finite population mean when the units were selected using the method of Double Rank Set Sampling (DRSS) was developed. Efficiency gains of the estimator based on DRSS procedure were demonstrated over estimators based on RSS and SRS with the help of real data.

The problem of estimation of finite population mean square has been attempted when the underlying population followed a simple linear model involving single auxiliary variable and the line of regression passed through origin. Further, the errors terms were assumed to be uncorrelated and the variance of error terms increased with the increase in size of the auxiliary variable. The method of double sampling was proposed when the information on the auxiliary character was assumed unknown. The efficiency gain of double sampling based predictor over a predictor which did not utilize auxiliary information was demonstrated under a suitably chosen cost function.

A Spatial Ranked Set Sampling (SRSS) scheme which took into account the spatial correlation between units was proposed. The merit of the proposed scheme vis-à-vis both Simple Random Sampling and Rank Set Sampling was demonstrated through a simulation study.

Two approaches to computer aided construction of Balanced sampling plans excluding contiguous units (BSEC)/polygonal designs were described. In the first approach polygonal designs were obtained using method of symmetrically repeated differences. In the second approach, BSECs were obtained using linear programming to minimize the probability of selecting samples which contained contiguous units subject to the constraints of specified first and second order inclusion probabilities of BSECs. The methods were illustrated with numerical examples.

A new family of distance balanced sampling plans was proposed where second-order inclusion probabilities were a non-decreasing function of the distance between the population units. Some properties of the proposed family of sampling plans were studied using Horvitz-Thompson estimator. Relative efficiency of the proposed family was compared with existing sampling plans. Some results on existence and constructions of the proposed family were discussed.

The problem of estimation of finite population total under nested stratified sampling plan was investigated. Allocation of sample sizes to different primary and secondary strata were obtained for fixed cost and fixed precision. The proposed plan under different allocation was compared with usual stratification.

Projective Geometry Approach was used to develop Minimum Variance Linear Unbiased Estimators of population parameters in sampling on two occasions with the simultaneous involvement of two correlated characters for the current occasion as well as change and average over time for both the characters.

Minimum Variance Linear Unbiased Estimator developed using Projective Geometry Approach was applied to estimate the foodgrain losses at harvest and threshing stage.

# Achievements in Applied Research in Sample Survey at IASRI

**Tauqueer Ahmad, U.C. Sud, Prachi Misra Sahoo,**
**Hukum Chandra and Vijay Bindal**

## 1. Introduction

The Indian Agricultural Statistics Research Institute is a pioneering Institute in the field of research in Agricultural Statistics in India which made a modest beginning in 1930 as Statistical Section of the Indian (then Imperial) Council of Agricultural Research. As per the mandates of the Institute to promote and conduct research in Agricultural Statistics, the Institute has contributed significantly over the years in research in the field of sample surveys, design of experiments, statistical genetics, bio-statistics, forecasting techniques, statistical economics and computer applications. The development of suitable sample survey methodologies for estimation of various parameters pertaining to crops, livestock, fisheries production and allied fields have always been one of the most important activities of the Institute from the very beginning. The Institute was instrumental in the development of sample survey methodology for estimation of crop production through crop cutting approach which formed a sound objective method of estimating crop production in the country. The sample survey methodologies for estimation of livestock numbers, products and attendant practices, for estimation of extent of cultivation and production of fruits and vegetables, for estimation of fish catch both from marine and inland resources, for estimation of cost of production of crops as well as livestock products and for evaluation studies such as assessment of development programmes like IADP, HYVP, dairy improvement programmes etc. are other important contributions of the Institute. Most of the methodologies evolved are being adopted by the concerned departments in various states of the country and this work is being coordinated by the Directorate of Economics and Statistics and other concerned wings of the Ministry of Agriculture, Government of India. Some of these methodologies are being internationally used.

In this article, the research highlights in sample survey at the Institute in different areas of applied research have been presented.

## 2. Sampling Methodologies Developed - At a Glance

One of the most striking features of recent development in statistics is the rapid growth of interest in development of sampling methodologies and their applications as well. The sampling approach being a most cost prohibited way of obtaining an estimate of a character of a population has compelled the survey statisticians to probe into problems of basic research. Survey sampling has considerably helped the concerned organizations in providing estimation of important population parameters for scientific planning in various developmental programmes. The Division of Sample Survey of the Institute undertook research in sample survey techniques and their application in agriculture, animal husbandry, fisheries, forestry and allied fields. By now, several methodological investigations have been carried out by the Division for evolving suitable sampling methodologies. Some of the methodologies have been adopted by State Departments and other related agencies and some are in the process of implementation. The broad categories under which these methodologies can be classified are:

(i) Crops

(ii) Horticulture and Plantation crops

(iii) Livestock

(iv) Fisheries

(v) Evaluation studies, Demographic studies, Socio-economic studies and all such studies which do not fall under above stated four categories.

**2.1 Surveys on crops**

One of the essential features for viability of any survey methodology is that it must be in conformity with the available infrastructure. For estimation of production of crops, one could concentrate on obtaining information on production of crops from the ultimate units. However, in the context of our country, estimation of area and yield is important which ultimately gives the production of crops.

Before 1919, method of determining crop yield was being done by traditional method. This method however, was subjective and unreliable. In 1919, objective method of crop cutting experiments by random selection of villages, fields and plots was recommended by Board of Agriculture. Initial attempt in practice goes to Hubback (1923-25) for carrying out surveys for estimation of yield of paddy in Bihar and Orissa. Hubback's idea of random sampling was pursued further for estimation of area and yield of jute crops. It may be mentioned that before independence, cash crops were considered more important as a source of revenue.

Before 1943-44, the official procedure of annawari estimation was defective and unreliable. This resulted in grave lacuna in the official estimates of yield of food grains. A serious food crisis occurred in the country in which Bengal famine was worst. This crisis could have been averted if reliable estimates of yield of food crops could be available.

Estimation of production of food crops was taken up seriously by Government of India. The Indian Council of Agricultural Research (ICAR) initiated pilot studies for developing methodologies for estimation of area and production of important food crops under the leadership of Sukhatme (1943-44).

The significant research in crop surveys in Sample Survey at the Institute are as under:

The method of crop cutting experiments developed in mid-forties for estimation of yield of a crop was standardized for different field crops which is widely adopted for estimation of production of principal crops in India and other developing countries. The standardized method of crop cutting experiments is a sound objective method of estimation of crop production.

In order to estimate the cost of cultivation of important food and cash crops, sampling techniques were developed. The developed methodology was adopted by the Ministry of Agriculture in the comprehensive survey on cost of cultivation of principal crops which forms the basis of fixing remunerative prices of agricultural commodities.

A suitable sampling and measurement techniques were evolved for estimation of incidence of pests and diseases and assessment of consequent loss in yield of paddy, wheat and maize crops.

In view of the importance given to the production of fodder in the Intensive Cattle Development Programme, pilot surveys were undertaken for evolving sampling methodology for estimation of area, yield and production of principal cultivated fodder crops.

Based on two surveys carried out, a suitable sampling methodology for estimation of area of grazing land and average yield per unit of grazing area in different seasons was evolved and chemical composition and botanical classification of grasses in each season were studied.

With the emphasis on white revolution, studies on cost of production of fodder crops assumed special importance, as cultivated fodders are the major inputs in dairy farming. The methodology available for estimating the cost of cultivation of cereal and cash crops did not seem to fit into studies on cost of production of fodder crops. With this in view, a pilot study for developing a suitable methodology for estimation of cost of cultivation of fodder crops was undertaken in Jalandhar tehsil of Punjab State.

A study was conducted to estimate the cost of cultivation as well as cost of production of potato, to estimate the area and yield rates of different varieties of potato and to study the extent of adoption of improved agronomic practices under cultivator's conditions.

In view of the difficulties to capture all the pickings of cotton during its long period of harvesting, a study was undertaken to develop methodology for estimation of yield of cotton on the basis of data of past few pickings. The results showed that it was possible to estimate the yield of cotton with good precision by adopting double sampling approach. In order to forecast the yield of cotton based on data of first one or two pickings, number of bolls, number of plants and plant height, multiple linear regression approach was tried successfully.

In view of unavailability of standard methodology for estimation of foodgrain losses, a survey was conducted in order to develop a suitable statistical methodology for estimation of post-harvest foodgrain losses specially, wheat.

A sampling methodology was developed to estimate area and production of cultivated fodder crops other than Jowar and Berseem and to review regular production on the lines of food-grains.

With the advent of remote sensing technology during 1970s, it's great potential in the field of agriculture opened new vistas of improving the agricultural statistics system all over the world. An attempt was made to develop a suitable methodology for estimation of crop acreage, crop yield and crop yield forecasting using remote sensing techniques.

The methodology developed under the previous study for crop yield estimation using remote sensing techniques was tested and an improved estimator of crop yield was obtained using post stratification based on satellite data in the form of vegetation indices.

A study was conducted to explore various possibilities of obtaining reliable land utilization statistics with minimum time lag at any smaller geographical level i.e. panchayats/blocks/tehsils based on sample survey using spatial sampling and remote sensing techniques. Under this study, suitable models were developed for integration of land use statistics obtained through different sources i.e. the complete enumeration, spatial survey and remote sensing.

A methodology was developed to estimate the various post-harvest food grain losses due to operational and casual factors and also due to periodic/seasonal fluctuations.

In view of the importance of fertilizers for increasing productivity in agriculture, a sampling procedure for selection of representative samples of fertilizer from ship was developed.

A methodology based on Crop Cutting Experiments (CCE) data as well as farmer appraisal data of crop produce was developed to estimate precise yield rates of different crops at Community Development Block (CDB) level.

A methodology was developed to estimate the cost of production of coconut. The cost of production of coconut was obtained by taking into account the establishment cost of orchard as well as the annual maintenance cost.

The Small Area Crop Estimation Approach (SACEA) developed by the IASRI for assessment of yield rates of different crops at the Gram Panchayat level was tested for large scale adoption in one district each of the State of Uttar Pradesh, Punjab, Rajasthan, Andhra Pradesh and Karnataka. The standard errors of the estimates were in the acceptable range. The report of the study was submitted to the Department of Agriculture & Co-operation, Ministry of Agriculture.

A study to estimate the crop yield at small area level for principal crops namely, wheat and paddy using small area estimation techniques was carried out to examine its applicability for getting reliable estimates of yield at lower levels such as block/tehsil which are very much needed for the purpose of micro level planning.

An appropriate sampling methodology for estimation of area under paddy in Meghalaya state was developed using remote sensing, GIS and ground survey. The developed methodology has been implemented in Meghalaya and an attempt is being made to adopt the methodology for estimation of area under paddy in other states of North-East region. The report of the study was submitted to Space Application Centre, Ahmedabad.

A study is being conducted in two States namely, Maharashtra and Andhra Pradesh in order to investigate the causes of variation between official and trade estimates of cotton production and to suggest ways and means for improving the reliability of estimates of Cotton production from the two sources. The study so far has revealed that official estimate is based on scientific methodology. The methodology is objective, scientific and verifiable and sampling design is proper but various problems have been observed in actual implementation of the Crop Cutting Experiment (CCE) scheme for Cotton. There is no scientific methodology for obtaining trade estimates. The interim report of the study was submitted to the DES, Department of Agriculture & Co-operation, Ministry of Agriculture.

In view of the successful adoption of methodology for estimation of area under paddy in Meghalaya state, a study is being conducted in North-East hilly region in order to develop an integrated methodology for area estimation of major crops i.e. paddy, maize, potato, ginger, pineapple, cashewnut and vegetable as a group using remote sensing, GIS and ground survey.

### 2.2 Surveys on horticulture and plantation crops

Horticulture and plantation crops contribute in various ways towards the prosperity of a nation. Cultivation of fruits and vegetables, if taken as organized industry, has scope for employment and upliftment of rural population. Some of the plantation crops like coconut, arecanut, cashewnut etc. and spices like pepper etc. are important from internal consumption angle and export point of view. Lac is another

commodity of similar value. Lac and cashew are grown in economically backward rural areas such as hilly and tribal areas. The status of statistics regarding these crops is not very satisfactory. The sampling technique developed for estimation of area and yield of various field crops can not be directly applied to horticultural and plantation crops due to inherent differences in various aspects like sowing, growth period, cultivation practices, harvesting etc. Accordingly, IASRI bore the brunt of developing methodology for area and yield statistics of these crops. The significant surveys by the Division of Sample Survey of the Institute are as under:

A suitable methodology for estimation of area and production of important plantation crops like arecanut, coconut and cashewnut was developed and is being used in major growing areas.

A survey was conducted for estimation of yield of lime and to study the cultivation practices of lime in Nellore district of Andhra Pradesh.

A methodology was developed for simultaneous estimation of area and production of fruits and vegetables and to study their marketing practices.

For estimation of cost of production of major fruits and perennial crops, sampling techniques were developed.

In order to develop a methodology for estimation of production of lac and extent of adoption of improved practices for its cultivation and remuneration received by the tribals, a survey was conducted in Bihar. The yield as well as production of lac were found to be greatly influenced by the price received by the tribals and, therefore, for maintaining production at reasonable level, adequate marketing facilities and support price would have to be ensured.

A pilot sample survey was conducted in Valsad district of Gujarat state for determining the cost of production of Chikoo and for studying its marketing practices.

An appropriate methodology for estimation of area and production of major fruits was standardized based on a series of pilot sample surveys conducted for the purpose in important fruit growing regions in the country.

A series of pilot sample surveys were conducted for developing sampling methodology for estimation of area and production of different vegetables in important areas like Delhi, Bangalore, Pune and Nasik. A suitable methodology for the purpose was standardized. Surveys on marketing of vegetables and their price spread were conducted in Delhi to investigate the share of growers vis-à-vis various intermediaries like wholesalers, retailers etc. on the price paid by the consumers. It was found that the share of the intermediaries formed a substantial component of the final price of vegetable.

Surveys were undertaken on cost of cultivation of crops like orange in Maharashtra, banana and mango in Gujarat, vegetables in Delhi and Gujarat and oilseeds and pulses in important growing regions of the country. These surveys provided valuable information on the various cost components of these commodities as well as return to the cultivators with known degree of reliability.

A survey was conducted in Bihar, Uttar Pradesh and Maharashtra states in order to evolve an appropriate methodology for estimation of Lac production.

A survey was conducted to develop a suitable sampling technique for estimation of price spread of vegetables at different stages of marketing and cost of cultivation of important vegetable crops

grown in the area. The different marketing and cultivation practices followed by the growers were also studied.

An attempt was made to study the variability of various components of cost of cultivation of vegetable crops at different stages of sampling and to determine the sample sizes for given levels of precision.

A sample survey was conducted to study the extent of production and cultivation of Cashew practices in Goa.

A suitable methodology was developed for estimation of yield of pepper in the states lacking in statistical procedure for estimation of yield of the crop.

In order to develop a suitable methodology for estimating the production of important vegetable crops and their yield rates on the basis of partial harvest, a suitable theoretical framework was developed for sampling from two dimensional populations spread over space and time with particular reference to vegetable crops and the theory was tested using secondary data collected under earlier vegetable surveys at IASRI. The total production of important vegetable crops and their yield rates were estimated using the developed methodology on the basis of primary data.

A pilot sample survey was conducted for developing a suitable sampling methodology for estimation of area and yield rates of ginger and potato in hilly areas.

A pilot study was undertaken to develop sampling methodology for estimation of area, production and productivity of important flowers on the basis of market arrivals.

Using secondary data, the effect of various inputs was studied on the yield of important vegetable crops. It was found that fertilizer, irrigation, human labour contributed significantly towards yield.

In view of National Statistical Commission (NSC) recommendation, a pilot study was conducted in two States namely, Maharashtra and Himachal Pradesh covering important fruits and vegetables. Under this study, an alternative methodology for estimating area and production of horticultural crops was developed which is cost effective and less time consuming and in which the survey procedures have been simplified. The study has revealed very encouraging results and demonstrated the feasibility of estimating the production of fruits and vegetables with much smaller sample size. As per recommendation made by the NSC, there is a need to test alternative methodology in few States before actually implementing it on a large scale. The report of the study was submitted to the CSO, Ministry of Statistics and Programme Implementation.

### 2.3 Surveys on livestock

The Institute has made a significant headway over the years in research in sampling methods in estimation of livestock and its products, rearing and maintenance cost of livestock etc. The estimates of livestock products such as milk, wool, eggs and meat etc. are needed for policy making and better planning for development of livestock. Division of Sample Survey of the Institute bore the brunt of developing sampling methodology for estimation of livestock and its products. The panorama of development of sampling methodologies is as under:

Appropriate sampling techniques were developed for estimation of production of principal livestock products like milk, wool and meat using the relevant information on livestock practices adopted.

Suitable sampling methodologies were developed for ascertaining the cost of production of important livestock products such as milk, wool, poultry and eggs based on a series of investigations carried out in different animal husbandry regions of the country. Suitable methodology was also evolved for studying the economics of grazing cattle and buffaloes.

The methodologies were standardized on the basis of a number of diversified investigations into alternative designs, procedure for recording and sampling techniques for measurement of (i) wool quality of fleeces of principal breeds of sheep in India in terms of qualitative values of fibre diameter, staple length, crimps per cm., and medulation percentage; (ii) chemical and bacteriological quality of milk supplied by the procedures; (iii) solids-not-fat in milk in terms of Richmond's modified formula and (iv) amount of milk sucked by calves under farm conditions.

Sampling methodologies were developed by the Institute for estimation of production and costing of principal livestock products, when the products were to be covered individually. The developed methodologies were quite suitable for adoption. The cost patterns of these products and changes in the annual output could not be estimated unless sample surveys are repeated every year on each of the product which would not be cost effective. Therefore, an integrated sampling methodology was developed for simultaneous estimation and costing of principal livestock products every year through one single survey.

A systematic sampling technique of milk recording for obtaining reliable estimate of milk production in key village areas was developed. The developed technique was operationally feasible.

A survey was carried out to develop a suitable methodology for estimating the specific rates of fertility & mortality in non-migratory flocks of sheep & goat with respect to breed, sex and age.

A suitable methodology was developed for estimation of vital statistics in a comprehensive way for bovines. The estimates of age-specific mortality rates among cattle and buffaloes obtained from the studies in different parts of the country were utilized for construction of life tables.

A suitable methodology was evolved for estimation of incidence of diseases and occurrence of deaths on account of these diseases, losses in production due to disability and deaths and extent of losses in production and reproduction which could be avoided through protection measures.

For studying the availability of broilers and culled birds for canning purposes, a sampling methodology was developed to estimate month-wise/season-wise age specific vital characteristics affecting the growth and structure of poultry population and to estimate the production of broilers in terms of number and weight and culling of layers by size of farm at regular intervals of time**.** Under this study, appropriate models characterizing the production and culling patterns in poultry farms utilizing the estimated vital characteristics were also developed.

Using the (1989-90) data on Integrated Sample Survey for estimation of major livestock products of Department of Animal Husbandry, Himachal Pradesh and the livestock census (1982) data the synthetic method of estimation for small area was used to estimate milk production for the different districts of Himachal Pradesh.

A sampling methodology was developed for estimation of wool production in India and for collecting information on sheep keeping practices based on a pilot survey in different sheep breeding tracts of the country.

A suitable sampling methodology was developed for estimation of cost of rearing and maintenance of goats, attendant practices of goat keeping, procedures for evaluation of cost components and costs and returns in goat farming.

Studies were conducted on cost of rearing and maintenance of rabbits and cost of production of Angora rabbit wool.

To assess the quantum of herbage intake by animals through grazing, a study was conducted to estimate simultaneously the intake by animals through stall feeding as well as grazing in Dharmapuri district of Tamil Nadu.

An attempt was made to identify the factors responsible for gap in milk yield of buffalo and also to estimate the contribution of each factor through statistical tools viz. path co-efficient analysis, multiple regression technique and principal component analysis.

### 2.4 Surveys on fisheries

Fish is a rich source of protein and its industry provides livelihood to millions of people. Apart from its contributions to national income, fish industry accounts for appreciably export trade. Indian fish industry depends mainly on marine fish. IASRI in collaboration with the fisheries departments of states initiated pilot sample survey for developing sampling technique for estimation of marine fish catch during 1950-51. The other source of fish is from inland water. Inland fisheries comprise of two types of water viz. fresh & brackish. Keeping in view the need for developing a sampling technique for estimation of inland fish catch, IASRI jointly with Central Inland Fisheries Research Institute (CIFRI), Barrackpore undertook a survey in 1979-80 in West Bengal. A brief account of development of sampling methodologies in the field of Fisheries is as under:

A methodology pertaining to marine fisheries involving sampling over space as well as time was developed to estimate the fish catch along the coastal area of the country. The developed methodology was suitably modified by Central Marine Fisheries Research Institute of the I.C.A.R. depending upon the situation and is since been adopted by all maritime states for estimating the fish catch from coastal resources.

Based on secondary data pertaining to Chilka lake in Orissa state, a suitable sampling methodology was developed for estimation of fish catch from a lake and the feasibility of estimating fish catch by using partial data was explored.

Suitable sampling methodologies for estimation of inland fishery resources and catch of inland fish in regions of West Bengal and Orissa were developed.

In view of unavailability of reliable statistics of fish catch from ponds, tanks and reservoirs, a suitable sampling methodology was evolved for estimation of inland fisheries resources and total catch of inland fish. Besides, the prevailing practices of pisciculture were studied.

A survey was conducted to evolve a suitable sampling methodology for estimation of inland brackish water resources and to study the prevailing practices of brackish water pisciculture.

A sampling methodology was developed for estimation of fish catch from rivers and streams of the hilly areas.

## 2.5 Other studies

The Division of Sample Survey of the Institute has been in the forefront in taking up Evaluation studies, Demographic studies and Socio-economic studies and all such studies which do not fall under above stated four categories. The significant highlights of these studies are as under:

In order to ensure timely availability of inputs and credit facilities and to study the institutional changes required, the Intensive Agricultural District Programme was launched by the government in selected districts of the country. The Institute conducted bench mark and assessment surveys to evaluate the success of the programme.

A survey was conducted for studying the impact of new technology on crop production, its disposal and employment in agriculture in Delhi State.

A pilot sample survey was conducted in order to investigate a sampling procedure for assessing the losses in agricultural production as well as livestock caused by flood and to study the impact of floods on crops in the subsequent season.

A survey was undertaken with a view to determine the design and response parameters for studying the comparative performance in terms of production, investment return, etc. of different systems of farming singly or in combination and to compare different systems of farming in respect of labour intensification.

A study was conducted to estimate the energy requirement for different levels of adoption of modern technology in terms of labour and inputs like irrigation, fertilizers, etc.

Appropriate sample survey methodology for evaluation study such as assessment of high yielding varieties programme was developed. In order to develop the methodology, the assessment surveys were conducted on important cereals viz. rice, wheat, maize, jowar and bajra in 88 districts spread over 15 states of the country. The surveys under the study involved yield estimation surveys for determining the yield rates of high yielding varieties and comparative estimates of the traditional varieties and agronomic and agro-economic enquiry for spread of the high yielding varieties and extent of adoption of improved agricultural practices.

A suitable sampling methodology was developed in order to make continual quantitative appraisal of progress in milk collection areas of organized milk supply schemes and Intensive Cattle Development areas. This methodology would help in assessing the impact of milk supply schemes on rural economy in milk collection areas.

A sample survey was conducted to study constraints in transfer of new agricultural technology under field condition. Based on this survey, an attempt was made to develop a suitable sampling methodology for studying the effect of new agricultural technology including high yielding/improved varieties, fertilizers, plant protection chemicals and cultural and management practices for higher productivity of land. The extent to which the potential of high yielding/improved varieties were achieved under field conditions was determined and the constraints in the transfer of new agricultural technology to cultivators' fields were identified and investigated.

A methodology was developed for determining yield rates of crops in command and non-command area of irrigation project with reasonable precision and the impact of command area irrigation project on agronomic and management practices and other development measures was studied.

A study was conducted to develop a sampling procedure for estimating the extent of employment and income of small and marginal farmers and agricultural labourers.

A pilot study was conducted to assess the harvest/production and post harvest/post production losses of wool, meat, eggs and poultry meat, marine fishery, inland fishery, milk and oilseeds at different levels viz. producer, consumer and market**.** Suitable sampling methodologies for estimation of harvest and post-harvest losses of all the commodities and for all the levels under study were developed. The causes of losses at different levels were also identified for all the commodities.

A pilot study on agroforestry in Chhachhrauli block of Yamunanagar district of Haryana state was carried out to study the impact and constraints of agroforestry program in relation to socio-economic structure of the district, spatial distribution of natural resources as well as available infrastructure and to develop integrated approach for planning and development for optimizing the resources in the field of agroforestry.

A document on information support for management of agriculture was prepared. Focusing specifically on 'farmer-the man', the document attempts to establish a link between 'information system for agriculture' and 'the farmer'.

A new GIS based Objective Analytic Hierarchy Process (AHP) namely, Objective Spatial Analytic Hierarchy Process (OSAHP) was developed for identification of potential agroforestry areas. The study was conducted in Yamunanagar district of Haryana state and may be helpful for the social, economic and environmental development of the study area as well as in future research in the area of agroforestry.

To assess the survey capabilities in the private sector a questionnaire was designed and the private survey agencies in the country were asked to fill the questionnaire. On the basis of responses obtained, the different survey agencies were graded in terms of survey capabilities. A report containing survey capabilities of the private agencies was submitted to the CSO, Ministry of Statistics and Programme Implementation (MOS&PI).

A study was initiated with the specific purpose of formulating long-term mechanization strategy. On the basis of a survey carried out in 120 districts in the country, long term farm mechanization strategies were formed for each of the agro climatic zone/state.

Six manuals on (i) Area and crop production Statistics, (ii) Animal husbandry Statistics, (iii)'Agricultural prices and marketing, (iv) Cost of cultivation surveys, (v) Horticulture and spices Statistics and (vi) Fishery Statistics were prepared and submitted to the CSO, Ministry of Statistics and Programme Implementation.

# Assessment of Food and Nutritional Security and Impact of Technologies

**Ashok Kumar, D.R. Singh, A.K. Vasisht, Sivaramane N., Prawin Arya and Sushila Kaul**

## 1. Introduction

To secure adequate food is a primary concern of any country. The cutting edge technological breakthrough has converted the Indian agriculture from dependency into self sufficiency in food production, however, poverty, undernourishment and malnourishment still persists specially in rural areas. The study of dietary pattern and nutritional status of the mass population is one of the indicators to assess the economic development of a country. To bridg the gaps between demand and supply of food for the growing population, efforts are being made to develop new technologies. Furhter, the benefits of new technologies in agriculture are not realized in all the regions especially in underprivileged areas. It is, therefore, important to undertake studies to assess the food and nutritional status of the rural population and to assess the impact of new technologies/research/ tecnnologacal interventions in the different areas of the country. The salient findings of important studies are presented under assessment of food and nutritional security and impact assessment of technologies.

## 2. Aims and Objectives

The change in dietary pattern and nutritional status of different socio-economic groups of rural households due to change in economic environment were studied under dietary pattern and nutritional status of rural households. Further, the socio-economic factors influencing the dietary pattern and nutritional status of rural households and the target groups for prioritization for improving the nutritional status were identified. Fruits and vegetables are important for balanced diet and food security and therefore, there should be a balance between demand and supply of fruits and vegetables in the future. Most of the studies pertaining to the demand estimation of food commodities were confined to estimation of demand for aggregate commodities like cereals, pulses, fats and oils, fruits and vegetables. The few studies were also estimated the demand of individual cereals. Therefore, in this direction a study on estimation of elasticities of demand and supply of major fruits and vegetables was undertaken to study the production and consumption pattern of major fruits and vegetables, to estimate the elasticities of fruits and vegetables at disaggregate level using Almost Ideal Demand System (AIDS) with Linear Approximation (LA) and to project short and long term demand and supply of major fruits and vegetables using the household level consumer expenditure survey data from three quinquennial rounds of National Sample Survey pertaining to the period 1987-88, 1992-93 and 1999-2000.

Earlier study on dietary pattern and nutritional status of rural households suggests that the landless, sub-marginal and marginal category of households in most of the states should be treated as target groups to raise their income to maintain their nutritional status. On this backdrop, the performance of microfinance was studied to examine the effectiveness of Self Helf Groups (SHGs) in reaching vulnerable sections in different agro-climatic zones of selected states and the factors for non-participation of households were also identified under the study on determinants of performance of Self-help Groups (SHGs) in rural

micro-finance (mF). The study was conducted in Uttar Pradesh and Andhra Pradesh in two districts of each state using primary as well as secondary data. The irrigation water plays an important role in achiving food security of a country. Therefore a study was undertaken on water food security scenario analysis for 2025 with an overall goal to strengthen the knowledge base and develop exploratory models as a tool for policy makers, scientists and others to interact and address sustainable water food security related issues in an integrated framework.

Regarding the impact assessment of technologies, a study was undertaken to develop an appropriate model for estimating the economic gains from technological advance in rice production and to estimate the economic gains across different rice growing states due to new technology in rice production. In this study consumers' surplus approach was used to estimate the economic gain. The consumer surplus is the "excess of the price which the consumer would be willing to pay rather than go without the thing, over that what he actually does pay or the difference between what the consumer is willing to pay and what does he actually pays" and the producer surplus is analogous to that of consumer surplus and refers to "a difference between that is actually received from sale of a commodity and the minimum amount to induce a producer to part with it". A model was developed which can estimate year-wise gains from yield enhancing technology in rice cultivation. The historical change in the yield of rice in different states was studied by fitting spline function. Micro irrigation technology was emerged as water and other resource saving technology. In India, however, most of the studies on impact of micro irrigation are conducted on research stations. Therefore, to assess the impact of micro irrigation systems on yield and saving of resources, a study was undertaken on the farmers' fields in Gurgaon and Mahendergarh districts of Haryana.

The fisheries sector generates income and employment and provides livelihood to a large section of economically backward population and stimulates growth for a number of subsidiary industries in the country. Besides, the sector is a source of valuable foreign exchange earnings. Therefore, a study on impact Assessment of fisheries research in India was undertaken in collaboration with other ICAR institutes in order to justify allocation of public funding to fisheries research, to evaluate return to investment of fisheries research in capture and culture fisheries and to assess the socio-economic and environmental impacts of fisheries technologies in different systems. For the measurement of farm specific technical efficiency, one of the most popular methods, the Stochastic Frontier Production method has been used. In the stochastic frontier production function which is popularly known as 'composed error model', the error term • is composed of two independent error components as: • = V-U. The error term V ~ N(0, $\sigma^2_y$) is a two-sided error term symmetrically distributed (- ∞<$V_i$<∞) and it captures the effects of random shocks outside the firm (farm) control, observation and measurement error on independent variables, and usual statistical 'noise' generally found in an empirical relationship. The other independent error component (U) is assumed to be non-negative and represents technical efficiency. This error term is one-sided and is a truncation of the N(0, $^2_u$) distribution (i.e; half normal distribution). The economic interpretation of this error term U is that each farm production must be either on or below the production frontier.

The migratory sheep production system plays an important in the rural economy of the country especially in the economically weaker sections of society in tribal, backward and hilly areas. However, the productivity of this system is very poor. Therefore, improved technologies were intervened in this production system to enhance the productivity under NATP Project on Jai Vigyan National Science and Technology Mission on Household Food and Nutritional Security in Tribal, Backward and Hilly areas

and the impact of technological interventions were assessed. A benchmark survey was conducted during the year 2001, in order to get a comprehensive view of the socio-economic aspects of migratory sheep rearing and food and nutritional security status of the migratory sheep farmers. Three consecutive repeat surveys were conducted for the same migratory sheep households during the years 2002, 2003 and 2004 to assess the impact of technological interventions.

## 3. Assessment of Food and Nutritional Security

### 3.1 Dietary pattern and nutritional status of rural households

The dietary pattern of rural households in almost all states showed a general trend of reduction in consumption of cereal in favour of non-cereal foods during 1999-2000 over 1987-88. The consumption of other non-cereal foods like milk and milk products, egg, meat and fish, fruits and vegetables groups tend to increase in most of the states. However, the shift from cereal based to non-cereal based diet was not visible in all categories of socio-economic groups uniformly. It was also observed that in most of the states, the proportion of deficient rural households was higher in landless, sub-marginal and marginal class and it decreased with the size of holdings. It was observed that Assam and Orissa were the most deficient state in most of the nutrients. Punjab and Haryana were at a better situation in terms of nutrients intake in the diet. In most of the states and land holding categories of rural households, the key factor influencing nutritional status was found to be per capita expenditure on food. In all the states, the coefficients of unit calorie cost and non-food expenditure were found negative and significant, indicating that higher prices of food items and excess expenditure on non-food items affected the nutritional status of rural households adversely. In most of the cases, the coefficient of household size was found to be negative and significant which indicates that bigger the family size and lower the nutritional status of the household. The analysis suggests that the landless, sub-marginal and marginal category of households in most of the states should be treated as target groups to raise their income to maintain their nutritional status.

### 3.2 Demand and supply of fruits and vegetables

The demand for total vegetables was projected to be in the range of 93-106 million tonnes and 105-129 million tonnes for the year 2010 and 2015, respectively under different growth scenarios of the economy. Similarly, the projected demand for total fresh fruits was in the range of 19-22 million tonnes and 22-28 million tonnes for the year 2010 and 2015, respectively. The Box-Jenkins model was used for forecasting supply of individual fruits and vegetables and projections were made for the year 2010 and 2015 after validating the models. The study projected huge demand-supply gap (excess demand scenario) of fruits and vegetables if the same trend in consumption and production continues.

### 3.3 Assessment of performance of self help groups

SHG progress was very diverse in Andhra Pradesh and Uttar Pradesh States. In spite of higher growth rate of population, higher share of cultivators, lower female literacy and higher female population in the selected districts of Uttar Pradesh, the SHGs progress was slow in comparison to A.P. This scenario rejects the hypothesis that “there should be a higher positive correlation between female population and number of SHGs and SHGs are gender and poverty focused but not sector focused”. The probable reason may be that the policies adopted in Andhra Pradesh are seemed to be more effective as location and socioeconomic specific policies and programmes were implemented in the state. Regarding determinants of repayment of loans, the analysis showed that higher amount of loan and socially more heterogeneous

groups are the main factors responsible for default in repayment of loan. The over all position of SHG programme in Andhra Pradesh was better as compared to Uttar Pradesh. It seems that the policies introduced in Andhra Pradesh are area specific and suited to the socio-economic conditions of Andhra Pradesh. Farm based activities like animal husbandry are receiving attention in economic empowerment of women. A member of the SHG also has good access to the banks. It was observed that people intended to take consumption loan from the group. This showed that they felt economically strong, as they are more confident that they can get financial support as and when they require for the upliftment of livelihood. It was also observed that family relations have improved after becoming members of the group and thus status of women in the family have also improved. Domestic violence is reduced to some extent as financial dependence of women on male members is reduced. In case of women control over loan, it was found that it depends on the requirement and purpose of loan.

## 4. Water-Food Security Scenarios

The analysis showed that the cropping pattern in agro-ecological subregions, namely; Ganga Yamuna Doab, Rohilkhand and Avadah Plains, hot, moist semi-arid eco-subregion (AESR 4.3) and Madhya Bharat Plateau and Bundelkhand upland, hot, moist semi-arid eco-subregion (AESR 4.4) was dominated by foodgrain crops. In AESR 4.3, the cereals occupied nearly three-fourth of the gross cropped area and in AESR 4.4, the cereals (43 per cent) and pulses (37 per cent) occupied four-fifth of the cross cropped area. The AESR 4.3 had considerably good proportion of irrigated area (68 per cent) along with low degree of variability among the districts in comparison to AESR 4.4 (20 per cent). On the other hand, development of groundwater was low in both the sub-regions implying the scope for further expansion of irrigated agriculture. However, such scope are quite location specific based on the status of groundwater development. In both the sub-regions, the daily average calorie intake was lower than Recommended Daily Allowances (RDA) in rural as well as urban areas. However, it is to be noted that in terms of nutrition and calorie intake both sub-regions were relatively better than many other parts of the country. WATER-FOOD 2025 model was developed in collebration with NCAP for simulating alternative scenarios of water-food security for 2025 in an integrated framework. Two alternative scenarios; namely, the Business As Usual Scenario (which assumes a continuation of current trends in water and food demand-supply drivers) and the Food Secure Sustainable Groundwater Use Scenario (which postulates no ground water mining and sustaining food security in future) were analyzed for these subregions for 2025 with specific set of interventions. 'Business as usual scenario' analysis showed that the sub-region 4.3 has food deficit and declining groundwater balance, whereas sub-region 4.4 has a very large food deficit but stable groundwater balance. On the other hand, 'Food Secure Sustainable Groundwater Use Scenario' analysis emphasized on yield growth in subregion 4.3 to eliminate the food deficit but groundwater depletion would continue unabated. In subregion 4.4, area expansion, intensity enhancement and yield growth while maintaining stable groundwater balance were explored and quantified to convert from a food-deficit to a food-surplus sub-region.

## 5. Impact Assessment of Technologies

### 5.1 Economic gains from advances in rice cultivation

The effects of technological advance in rice cultivation across states vary from high quality of land-dependent case to low quality to land-invariant case. The results showed that the technological advances

had added another factor responsible for variation in the yields of rice among the states. This indicated that the effect of technological advances in rice production depends on both input use as well as on land quality. The inherent advantages of land quality and higher input use in Punjab, Haryana, Tamil Nadu, Uttar Pradesh, Andhra Pradesh, West Bengal and Maharashtra showed better response to technological advances in rice production as compared to other rice growing states. In general, producers were the major beneficiaries of technological advance in rice production in all the states while the consumers' gains were found to be low as compared to producers gain in all the states. Grain yields from technological advances are mainly dependent on shift in supply curve and the movement of price. Consumers' gain mainly depends on price movement while producers' gains depend on the extent of shift in supply curve. Therefore price policy plays an important role in creating the economic surplus.

### 5.2 Impact of micro-irrigation on productivity and resource saving

With the advent of micro-irrigation in Gurgaon and Mahendergarh districts of Haryana, the benefits of drip irrigation in orchards are taken by only large farmers while sprinkler irrigation is being used by farmers irrespective of size of holdings. The benefit/cost ratios were computed for Guava, Ber, Aonla, Kinnow, Mausami and Anar orchards which are popular in the area. The maximum increase in productivity i.e. 173 percent was observed in Ber while the minimum i.e. 34 percent was obsereved in Aonla. An increase of 27 percent in productivity was observed in wheat and mustared, 20 percent in cotton and 14 percent in bajra. In Gurgaon some of the farms have introduced strawberry and Gladiolus and it was found that both new crops are more profitable under drip irrigation. The benefit/cost ratios were higher in case of subsidy on irrigation systems. The resource saving under drip irrigation varies from crop to crop; however, 40 to 60 percent of waer, 17 to 55 percent of fertilizers, 34 to 50 percent of electricity and 7 to 20 percent of labour can be saved. Under sprinkler irrigation, wheat and mustard in *rabi* and cotton and bajra in *kharif* seasons are grown. About 20 to 30 percent of electricity and labour and 30 percent of water and fertilizers can be saved in wheat and mustard crops while 35 percent of water, 20 percent of fertilizers, 50 percent of labour and 36 percent of electicity can be saved in cotton and about 30, 20 and 20 percent of water, labour and electricity, respectively, can be saved in bajra cultivation under sprinkles. The share of credit and subsidy was 9 and 50 percent of the total investment, respectively.

### 5.3 Return to investment in fisheries research and technical efficiency of fishery farmers

The analysis of scheme-wise outlay for fishery development under different plans revealed that the central government has played an important role in the outlay for fisheries development as there is a direct relationship between central sector schemes outlay and share of fisheries in agricultural outlay. The analysis of expenditure on Fisheries Research showed that the proportionate share of outlay for fisheries research in total fisheries outlay had more than doubled from fourth plan to ninth plan. The incremental GDP in fisheries was impressive in early nineties resulting thereby decrease of incremental capital output ratio from 2.34 in late eighties to 2.02 in early nineties. The socio-economic analysis of the fish farmers from Punjab revealed that the majority of the farmers were having the pond size of less than 2.5 hectares with an average size being 1.4 ha. The net benefit-cost ratio was 0.62 on large farms indicating that the return per rupee of investment was impressive on large farms whereas it was 0.31 for small category of farms. It was found that among all the variables included in Frontier Production Function viz, stocking density, labour, fertilizers and feed, only stocking density and feed were significant, indicating

that one percent increase in stocking density would result in a change in output by 0.24 percent. The average technical efficiency of the fishery farmers in Punjab was 78 percent compared to the best frontier. About 39 % (47 farms) had shown less than 70 per cent efficiency and about 20% (24 farms) had shown efficiency between 70- 80 percent. On the other hand, 41 percent of the farms had shown more than 80 percent efficiency when compared with the best frontier. It is also observed that the small fish farmers are managing their resources better than the other category of farmers. Further, about 96 percent of the difference between the observed and the frontier output was mainly due to the inefficient use of resources which are under the control of fish farmers. Although, the farms are technically sound to some extent, the economic efficiency is low. There is a tremendous chance to increase technical, allocative and economic efficiencies of Punjab fish farmers. The aquaculture in Punjab is profitable, but it is wrought with certain constraints. Hence there is still some scope to increase the profitability in general and of small and medium category of farmers in particular by improving the technical efficiency and mitigating the constraints to the extent possible.

### 5.4 Impact of technological interventions in migratory sheep production system

The results of the study revealed that there was an impressive increase in productivity and income realization due to technological interventions (breeding, health aspects and nutritional supplementation) in migratory sheep production system. There was an increase in the consumption of deficit food items in both the districts, which indicates an improvement in food and nutritional security of the participating households. Further, spillover impacts of technological interventions especially in breeding component were impressive especially in Ajmer District. The findings of the study gave some important policy options to enhance the profitability and sustainability of the migratory sheep production system. Firstly, the efforts should be made to replicate the technological intervention programmes in other areas and secondly, the migratory sheep owners may be educated to adopt better package of practices to realize sustainable returns from this production system.

# Modelling for Risk and Uncertainty, Resource Use Efficiency and Agricultural Marketing

**S.P. Bhardwaj, A.K. Vasisht and Prawin Arya**

## 1. Introduction

The nature of challenges in Indian Agriculture has changed in the recent years. It is no more a question of substituting imports of cereals through rapid growth in their production. The additional production of cereals of course needed in a cost-effective manner to facilitate employment oriented growth. Farmers often prefer farm plans that provide a satisfactory level of security. It is therefore becomes essential for the farm economist to incorporate risk in the farm planning models. Emerging scenario of input utilization of the factors of production under farmer's control has been assessed in a manner that a scientific insight for resolving the problem of measuring management input component present in economically defined concept of Residual due to Risk and Management which has direct bearing for determining rational minimum support price policy with the help of heuristically deduced results of actual and frontier figures relating to levels of output, gross revenue and net returns**.** Crop insurance may be considered as a viable alternative to mitigate the risk & uncertainty. Presently, crop insurance in India is following as area based approach and insures for yield losses only. However, in general, farmers are prone to the price risk along with yield risk. Whenever there is glut in the market due to good yield, there is more than proportional drop in the prices of commodities leading to heavy losses to the farmers. A new risk management tool known as crop revenue insurance (CRI) can address this problem. Crop insurance may help us to minimize the extent of uncertainty and level of risk factor. The next step is how to improve the resource use efficiency in our farming system as most of our inputs bears low marginal productivity which indicates that inputs are underutilized or not properly used. Therefore increasing efficiency in resource use is another approach to enhance farm profitability. The history of economic development of various advanced countries shows that technological change has been the most powerful factor of economic growth. Technological change has helped not only in raising the factor productivity but also in generating surplus for overall development. The adoption of modern technology might disturb the traditional resource equilibrium. Adjustment in resource use in response to technological change is based largely on trial and error method. The study on resource use efficiency at different levels of technology adoption assumes considerable importance in resource allocation pattern. Technological change may affect the relative economic position of the factors of production in terms of their share in total output and thus has implication on income distribution. The marketing is the next area of research to increase profitability in the farm sector. Agricultural marketing plays an important role not only in stimulating production and consumption but in accelerating the pace of economic development. Market information is needed by farmers in planning production and marketing, and is equally required by other market participants in arriving at optimal trading decisions. The existence and dissemination of complete and accurate marketing information is the key to achieving both operational and pricing efficiency in the marketing system. There are several areas of agricultural marketing with which farmers need to be fully familiarized in order to improve price realization. Agricultural marketing is witnessing major changes owing to liberalization and globalization of markets. In this context agriculture has to be market driven, more cost effective, competitive, innovative

and responsive to high tech and I.T applications. In marketing, the Technical Efficiency is said to have increased when operational cost is reduced for performing a function for each unit of output. This can be achieved by reducing physical losses and improvement in technology to carry out a particular function viz storage, transportation, handling and processing. A change in the technique may result in the reduction of per unit cost. Allocative efficiency of the market system refers to efficient allocation of farm products either over time or across the space among the traders, processors and consumers so as to protect the economic interests of producers and consumers. Market integration refers to those markets where prices are determined interdependently. It is assumed that price changes in one market will be fully transmitted to other markets (Engle and Granger, 1987. The Price Spreads represent the difference between retail and farm prices of a specific product at a given point in time. The length of price spread depends upon the number of channels involved in the marketing of a particular product. If the number of marketing channel or middlemen is large consumer has to pay more price because every channel will add its own cost &margin to the price of the commodity. Therefore it is argued that efforts should be to minimise the number of channel and especially unproductive channels should be eliminated from the marketing chain through market reforms.

## 2. Aims and Objectives

To achieve this thrust area followings are the broad objectives. In the study on, the farmer's behaviour towards risk and its effect on resource allocation and farm income the main objectives were, to study farmers behaviour towards risk with respect to purchased inputs, to identify the determinants of attitude towards risk for sampled farmers and to examine the impact of risk on cropping pattern, level of resource use and farm income. This study was based on primary data of the sampled cultivators spread over Umrain and Ram Garh, development blocks in District of Alwar (Rajasthan) and Loni and Bhojpur development blocks in district of Ghaziabad (U.P.). In the study on, Testing relative economic efficiency and determination of factor demand and output supply functions for wheat. The objectives were, to test the relative economic efficiency between large and small farms of wheat crop, to determine the demand for labour and fertilizer in the production, and to determine the output supply function. The farm level wheat data of Punjab and Haryana state for the year 1985-86 collected from Directorate of Economics and Statistics, New Delhi were utilized in this study. In the study on, An Econometric Approach for Measurement of Indemnity and Premium Rates under Crop Revenue Insurance with the objectives to examine the sources of instability for identifying critical variables for Crop Revenue Insurance, to estimate indemnity and premium rate of selected crops in different states of India for Crop Revenue Insurance and to explore the feasibility of Revenue Insurance approach vis-à-vis Yield Insurance approach in Indian agriculture. To measure the instability of economic variables (area, yield and price), Cuddy-Della Valle index (corrected coefficient of variation) is used which takes in to consideration the long-term trend. Premium rates have been estimated with the help of existing yield approach methodologies for different crops of Uttar Pradesh and Karnataka state. The specific objectives of the study on, "A Study on Production Efficiency and Resource Use in Poultry Production" were to study the pattern of poultry production, level of investment and capital requirement in poultry enterprises, to study the pattern of resource use and its efficiency in poultry production, to examine the profitability of poultry enterprises and to study the constraints faced by poultry producers. The data utilized in the study pertains to 41 poultry farms of varying size and spread over different poultry pockets of Ambala district of Haryana State. The data was

obtained from the Department of Animal Husbandry Statistics (AHS), Haryana Government, Chandigarh. The data was collected by Department of AHS under the Integrated Livestock Survey Scheme. Data used in the study was for the period of 1991-92 to 1994-95. In the study on, Econometric Study of Technological Dualism in Egg Production the specific objectives were to study extent of technological dualism and technological change in egg production, to study the effect of technological dualism and technological change on functional income distribution, and to examine economic efficiency under different levels of technology. The project is based on primary survey data of selected layer poultry farms spread over two districts of Punjab namely Ludhiana and Mansa. In Ludhiana district, 30 farms (including deep litter and cage system farms) were randomly selected for canvassing the specified data and in Mansa district, 50 farms were randomly selected. From the selected farms, primary data was collected at fortnightly interval. The primary data were collected for a period of 18 months, from the month of June 2002 to December 2003, as the egg layer bird has a productive life of 18 months. In the study on, Policy Analysis & Market Intelligence (NAIP Project), the objectives are to analyze the supply and demand of major agricultural commodities, to monitor price trends of major commodities in selected domestic and global markets, to study market integration and futures market, and to perform action research on smallholders' participation in futures market. In the, Study of Lac Marketing in India, the objectives were, to study existing marketing channels and price spread in Lac Marketing, to examine extent of market integration in Lac Markets and to suggest remedial measures for improving marketing environment of Lac in India. The three aspects namely Lac cultivation, Lac marketing and Lac processing were examined in the major producing states of Jharkhand, W.Bengal, Chhattisgarh, Madhya Pradesh and Maharashtra by primary Survey data.

### 3. Minimizing Risk and Uncertainty

The farmers are much concerned about risk in their decision making whether they are risk taker or risk averse. The two major source of crop income risk are prices of inputs, crop output and the resultant yield of the crop. More secure farm plan may involve producing less of risky enterprises, diversifying into large number of enterprises to spread risk, using established technology rather than adopting a new one. In our farm planning models a little progress has been made to incorporate risk in the production models. In this direction a study of Farmers behaviour towards risk and its impact on cropping pattern, level of resource use and farm income was undertaken in the division. The study was based on primary survey data of the sampled cultivators spread over Umrain and RamGarh development blocks in the district Alwar: Loni and Bhojpur, development blocks in the district of Ghaziabad. Stochastic linear programming model was used to incorporate risk in the optimal crop plans. The risk-aversion coefficient 'K' was measured through a formulation developedusing the following formulae.

$$K = \frac{1}{\theta}\left(1 - \frac{P_i}{Pf_i}\frac{X_i}{\mu_y}\right)$$

where $\theta = \frac{\sigma_y}{\mu_y} \times 100$, given coefticient of variation of yield, P = given factor prices, $P_i$ = given product prices of $i^{th}$ product, $X_i$ = ith input vector, $f_i$ = elasticity of production function of the input and y = yield.

The specific results of the study indicated that use of chemical fertilizers was largely confined to the Rabi season crops like wheat, mustard and Barley. The rates of fertilizers application were significantly

higher in the district Ghaziabad as compared to district Alwar. The values obtained for risk-aversion coefficient 'K' indicates the extent of risk-aversion for a particular farmer in a particular crop on the purchased input. The observed value of 'K' was higher in wheat crop as compared to mustard, which indicates that returns were more attractive in mustard crop as compared to wheat crop. In the district of Ghaziabad the observed value of 'K' was higher as compared to district Alwar, which again indicates that district Alwar has much higher potential of net returns. Among different categories of farmers the observed value of 'K' reveals no set pattern as it depends mainly on the performance of individual farmer. The main determinants of the attitude towards risk as identified in the study were, age and education of the decision maker in the family, holding size and number of earners and helpers in the family. The optimum crop plans were obtained for the different categories of farmers in the four development block under study. The studies mentioned above help us to minimize the extent of uncertainty and level of risk factor. The next step is how to improve the resource use efficiency in our farming system as most of our inputs bears low marginal productivity which indicates that inputs are underutilized or not properly used. Therefore increasing efficiency in resource use is another approach to enhance farm profitability.

## 4. Testing Relative Economic Efficiency and Determination of Factor Demand and Output Supply Functions

The study aimed to test the relative economic efficiency between large and small farms of wheat crop, determine the demand for labour and fertilizer in the production, and determine the output supply function. The farm level wheat data of Punjab and Haryana state for the year 1985-86 collected from Directorate of Economics and Statistics, New Delhi were utilized in this study. By Unit Output Profit (UOP) function approach and applying the Lau-Yotopoulos test the study inferred that there is no significant difference between the large and small farmers in their relative economic efficiency in both the states. The same set of conclusions are also drawn when the data of both the categories of farmers are applied to Chow-test indicating the fact that the two groups of sample farms do not give different relationship. Therefore, there are reasons to believe that the Punjab and Haryana agriculture has been modernized, relying on new varieties of seeds, fertilizers, irrigation and other chemical inputs for which both the large and small farms have similar access. Further, the salient futures of the present study reveals that in all the estimating equations, the 1985-86 wheat data of Punjab and Haryana provides goodness of fit as witnessed by the values of coefficient of determination and F-statistic. In almost all the cases normalized price elasticities of variable inputs are less than one indicating inelastic response. However, the price elasticities of output obtained from relevant estimating equations are greater than one indicating elastic response. The estimated coefficient of land input is highly significant. Almost in all the cases unconstrained joint GLS coincides with the LS estimates. The reduction in the standard error and substantial improvement in the estimates of the coefficient of normalized prices of variable inputs is obviously a clear cut indication that constrained joint GLS should be preferred over unconstrained joint GLS of SURE methodology under the framework of UOP project function.

## 5. Measurement of Indemnity and Premium Rates under Crop Revenue Insurance

The study aimed to examine the sources of instability for identifying critical variables for Crop Revenue Insurance, to estimate indemnity and premium rate of selected crops in different states of India for Crop

Revenue Insurance, to explore the feasibility of Revenue Insurance approach vis-à-vis Yield Insurance approach in Indian agriculture. To measure the instability of economic variables (area, yield and price), Cuddy-Della Valle index (corrected coefficient of variation) is used which takes in to consideration the long-term trend. Premium rates have been estimated with the help of existing yield approach methodologies for different crops of Uttar Pradesh and Karnataka state. Data sets have been subjected to Normality test wherever required. Premium rates were calculated by the normal curve technique with some exceptions indicating for the instability in the yield for that particular crop. For the nonparametric kernel approach, the window width has been estimated which are found in the range of 120 to 230. For premium estimation under crop revenue insurance, gross return is taken into consideration. In Karnataka state revenue and yield showed low instability compared to area. But in case of Uttar Pradesh, instability was low for all the parameters. Under crop revenue approach, the proportion of almost all the crop strata was having high revenue premium rates than yield premium in Karnataka State. However, in Uttar Pradesh, the revenue premium rates were low in comparison to the yield premium rates. A comparison of estimated revenue premium rates for Karnataka and Uttar Pradesh state showed that the majority of crop strata were having low premium rates in Uttar Pradesh in comparison to Karnataka state.

## 6. Production Efficiency and Resource Use

The study *was* undertaken to identify the factors responsible for inefficient resource use in production and imperfections in marketing of poultry products. The central points of interests that emerged out of the study are the role of size of farms and season on the economic viability and thus profitability of the poultry enterprises. The important features of the study are as follows:

1. Estimates of costs and returns on egg layer farms indicate that feed alone constitutes about 70 percent of the total cost of egg production. Small farms earning about Rs. 2 per bird in summer and rainy seasons and up to Rs.10 in winter season. Large farms on the other hand earning up to Rs.30 per bird. The seasonal effect on returns entered through price of eggs which are higher in winter season.
2. Returns on broiler farms indicate that feed constitutes about 55 percent of the total cost. In summer season the net returns on small farms are least i.e. Rs. 339 per 100 birds. While in other cases the net returns from broiler production remained around Rs. 800 per 100 birds.
3. The small farms are incurring more expenditure on inputs which ranges between Rs. 109 to 141 per 100 birds which is more as compared to medium and large farms. This high input cost lowers the profitability of small farms. However, the average bird's productivity in small farms is not affected by the prevailing inefficiency on the farms. These trends are almost identical in both cases i.e. egg layer farms and broiler farms.
4. Large and medium farms also enjoy the economies of scales and thus making the output-input ratio more favourable to them.
5. Marginal productivity of feed on egg layer farms remained around 3.0 on small and large farms in three seasons. The Marginal Productivity of labour remained high on medium and large farms and ranges 3.5 to 8.0. On small farms it remained less than 3.5.
6. In case of broiler farms marginal productivity of feed is very low on different farms and it ranges between 0.004 to 0.073 farms. Where as labour productivity farms remained less than 0.5 on all farms.

The existing and optimum levels of resource use reveal that funds have to be curtailed on feed consumption on all most all farms in three seasons. The optimum level of resource use indicates that labour input may be enhanced on poultry farms i.e. funds may be diverted to labour as the existing level of labour use is significantly low. The study indicates that poultry showed a high and positive marginal product of labour, this can provide more opportunities of employment. The role of technology is the recent development in enhancing farm profitability. Technological change may affect the relative economic position of the factors of production in terms of their share in total output and thus has implication on income distribution.

**7. Technological Dualism/Technological Change**

The technological change in poultry farming has been defined by the system of keeping poultry birds. The cage system of keeping poultry birds termed as modern technology farms. It is called so for they adopt modern practices such as provision of water nipples, dehumidifiers, quality feed mix, proper medication, etc. The poultry farms operating on deep litter system of keeping birds have been referred as traditional farms. These farms also adopt the above mentioned modern technologies but to a limited extent. The results revealed that the initial capital investment for Cage System farms was more than two folds as that of Deep Litter system farms in Mansa and nearly three folds in Ludhiana District. However, the average investment for a particular technology was higher for farms set up at Ludhiana compared to those at Mansa probably because of urbanization. The estimates of Cost and Returns on Layer Farms indicate that Cage System farms were earning more profit compared to the Deep Litter System farms in both the selected districts. The major components of cost were cost of feed, cost of chicks, fixed cost and vaccination & medication charges. The result of regression analysis indicates that major factors influencing egg production are feed cost, labour cost, medicines and electricity cost. The study of regression analysis showed that on both types of farms most of the input variables except for feed cost are not properly utilized. The analysis showed that input factors are not being utilized efficiently on Deep Litter farms in both the district. However it was observed that if the poultry farms using Deep litter system shift over to Cage system of technology there will be a substantial saving in the input resources. The Chow test Statistics conducted to test the structural relationship between the two technologies proved the superiority of modern technology over the traditional one in terms of efficiency of inputs at both the Districts. The existence of Technological Dualism in egg production reveals that inputs are not being efficiently used on Deep Litter farms .The Chow test further confirms the fact that a shift to the modern Cage technology can save the inputs substantially. Factor share analysis in districts Mansa reveals that the share of labour factor remained about 4 percent, the share of poultry feed which is a proxy variable for capital, was maximum of about 62 percent on both types of farms. The proportionate change of ith factor share (Di) as a result of technological change reveals that there will be saving of inputs in shifting from Deep Litter to Cage Technology. The saving is observed in feed 1.00 percent, labour 36 percent, medicines 23 percent and in electricity cost by 20 percent on the poultry farms in the district. In District Ludhiana the share of labour factor remained about 4 percent on Cage farms and about 5 percent on Deep Litter farms, the share of poultry feed was about 76 percent on Cage farms and 55 percent on Deep Litter farm. The proportionate change of ith factor share(Di) as a result of technological change reveals that by shifting from Deep Litter to Cage system of Technology there will be a saving in Labour (33 percent), Electricity cost (40 percent), Feed (38 percent) and in Medicine cost (11 percent) on the poultry farms in the district. The available viable production technologies have resolved most of the problems related production sector. The marketing is

the next area of research to increase profitability in the farm sector.

## 8. Marketing Research

Agricultural marketing plays an important role not only in stimulating production and consumption but in accelerating the pace of economic development. In pursuance of the suggestions made at various levels, a manual on agricultural prices and marketing was prepared by IASRI, for planning and policy formulation relating to agricultural prices and marketing.

### 8.1 Benefits of efficient marketing

1. Any increase in the efficiency of the marketing process, which results in lower costs of distribution and lower prices to consumers, really brings about an increase in the National Income.
2. A reduction in the cost of marketing is a direct benefit to the society.
3. Marketing process brings new varieties, quality and beneficial goods to consumers. Approximately one third of all persons gainfully employed in the country are engaged in the field of marketing and about one forth of National Income is earned by marketing profession.
4. Scientific marketing has a stabilizing effect on the price level. If producers produce what consumers want and consumers have a wide choice of products there are no frequent ups and downs in price.
5. Marketing is a catalyst for the transmutation of latent resources in to actual resources, of desires into accomplishments and development of responsible economic leaders and informed economic citizens.
6. Marketing brings to the farmers useful implements, tools and fertilizers etc. and the benefits of the use of machines and free after-sales service, and make them modem farmers.

There are several areas in which farmers will be required to remain proactive to derive benefits from emerging scenario of technological changes and marketing system.

1. Direct Marketing: it encourages farmers to undertake grading of farm produce at the farm-gate and obviates the necessity to haul produce to regulated markets for sale. Direct marketing enables farmers and processors and other bulk buyers to economize on transportation costs and to considerably improve price realization. In South Korea, for instance, as a consequence of expansion of direct marketing of agricultural products, consumer prices declined by 20 to 30 per cent and producer-received prices rose by 10 to 20 per cent.
2. Contract Farming: A farming arrangements of different types have existed in various parts of the country for centuries for both subsistence and commercial crops. The commercial crops like sugarcane, cotton, tea, coffee etc. have always involved some forms of contract farming. Even in the case of some fruit crops and fisheries, contract farming arrangements, involving mainly the forward trading of commodities have been observed. However, in the wake of economic liberalization, the concept of contract farming in which national or multinational companies enter into contracts for marketing of the horticultural produce and also provide technologies and capital to contract farmers and has gained importance.

### 8.2 Market intelligence

Market intelligence is an important area for marketing research and a study has been undertaken

under Visioning, Policy Analysis and Gender (V-PAGe) (Sub-Prog. III): (NAIP Project). The study aimed to analyze the supply and demand of major agricultural commodities, monitor price trends of major commodities in selected domestic and global markets, market integration and futures market, and perform action research on smallholders' participation in futures market. The per capita monthly consumption expenditure and proportion of total food consumption expenditure incurred on major food groups (cereals, pulses, oils, milk & milk products, egg, meat & fish, vegetables, fresh fruits, sugar and others) for rural and urban households for different regions of the country were analyzed for different NSS rounds (i.e. 43$^{rd}$, 50$^{th}$, 55$^{th}$ and 61$^{st}$) pertaining to consumer expenditure.

### 8.3 Price spread and market integration

"Study of Lac Marketing in India *was* undertaken to study existing marketing channels and price spread in Lac marketing, extent of integration in Lac markets and to suggest remedial measures for improving marketing environment of Lac in India. The study of Lac cultivation showed that there are three mainly identified Lac host in the country. Palas tree was the largely cultivated host tree in all the Lac growing areas in terms of number of host and scrap Lac production. The Level of net returns from Lac cultivation of selected cultivators from Lac cultivation was comparatively much higher in Maharashtra Rs. 2,14,848 ( in this state cultivators have leased in large number of Lac host from Forest department), Rs. 62,114 in M.P and Rs.1217 in Chhattisgarh states respectively. Rs. 6467 in Ranchi district and Rs. 4229 in W.Singhbhum district of Jharkhand state respectively. Despite such a lucrative returns from Lac cultivation the *extent of Lac host exploitation* is low. Extent of Palas host exploitation in Chhattisgarh about 72% in M.P and Maharashtra about 50 with a little error of about 3%. The average Lac host exploitation rate in the three states was about 51 percent for Palas host respectively. In Jharkhand state the average rate of host tree exploitation in Ranchi district was 46.74%. The low extent of host tree exploitation was the prime factor for low crop production. The study of Price Spreads showed that cultivator receive better price in those markets where Primary Purchaser sold scrap Lac to Processing units directly. This implies that existence of Lac Processing units in Lac growing areas facilitates better price to the cultivators. The study of market integration suggests all the important Lac markets situated within the same state and among different states are not properly integrated and there exit a significant price difference in scrap Lac. The production of scrap Lac in the country can be increased by making timely availability of brood Lac or small funds to the Lac cultivator for the purchase of brood Lac (Brood Lac farms needs to be set up by forest department) , the extent of host exploitation will certainly improved which will result in higher crop production. Installation of Lac Processing units especially in cultivating zone will attract cultivators to produce more of Lac crop in anticipation of better crop price.

### 8.4 Price volatility

Another important aspect is the price volatility which influences marketing strategy and income of the farmers and the empirical research has produced mixed results. Thus a study on Price Volatility and Integration in Spot and Futures Market of Gram was undertaken to empirically evaluate integration of futures and spot market of gram. The analysis of cross correlations of error terms of both spot and future prices provides information on mutual feedback relationship between the two prices. The price series of Gram crop in spot and futures market shows that level data was non-stationary but at their first differences it became stationary (i.e. implying the presence of unit roots in the series). The occurrence of unit roots in

the price data gives a preliminary indication of shocks having permanent or long lasting effect. The same order of integration for both spot and futures prices reveals that there exists a long run price equilibrium relationship between these prices. The study of price series of Gram crop witnessed a persistence and asymmetry in volatility of prices both in spot and future markets. Spot price of gram crop shows a value of 0.57, which is relatively smaller than the value obtained in price series of future prices which is close to one. The results of Granger test detect unidirectional Granger causality from futures to Spot markets. The means of the spot market are greater than those for the futures market. The futures volatility (standard deviation) is greater than that of the spot market volatility. The distribution of the spot prices is comparatively more skewed than futures prices. It was observed that there was mutual feedback between the prices of spot and future markets. The statistically significant coefficient for Zero lag indicates that there is instantaneous transmission of price signals between the markets. The result also indicated that futures prices influence spot prices.

# Five Decades of Research in Statistical Modelling: An Overview

**Prajneshu and Himadri Ghosh**

## 1. Introduction

Statistical modelling plays a very important role in comprehending underlying relationships among crucial variables in an agricultural system. Some examples are: Length-weight relationship, Input-output relationship, Computation of compound growth rate, and Path of adoption of high yielding varieties. A statistical model can be either 'Linear' or 'Nonlinear'. In a linear model, all the parameters appear linearly, whereas in a nonlinear model, at least one of the parameters appears nonlinearly. Those nonlinear models, which can be transformed to a linear model by means of some transformation, are called 'Intrinsically linear'. Before 1963, only either linear models were used or the use of nonlinear models was confined to only using intrinsically linear models. The nonlinear estimation procedure due to Levenberg and Marquardt, which is currently widely used, was developed in 1963. However, as it was extremely complicated, it was hardly ever used for the next two decades or so. In 1980s, software packages for fitting Nonlinear statistical models through nonlinear estimation procedures were developed. In a few years' time, i.e. around early 1990s, most of the standard software packages, like SPSS, SAS, GENSTAT and SPLUS contained such programs. In the subsequent sections, we discuss briefly the salient features of research work carried out at I.A.S.R.I. during last five decades in various sub-areas of Statistical modelling, viz. Nonlinear growth models, Contagious distributions, Structural time-series analysis, Parametric nonlinear time-series modelling, Nonparametric regression methodology, Soft computing, and Growth models in random environment.

## 2. Nonlinear Growth Models

In any fishery, it is of paramount importance to have a knowledge of sustainable yield that can be harvested in perpetuity without altering the fish stock level. Scientific fishery management is generally based on the concept of Maximum sustainable yield (MSY), according to which, at any given population level less than the carrying capacity, a surplus production exists. To this end, several models, like Schaefer and Fox models have been developed for assessing MSY and the optimum fishing effort to achieve the same. These models are widely used for efficient fishery management as these require only the information regarding time-series data on catch and effort, which is readily available in our country for a large number of fish species. As the surplus production models, like Schaefer and Fox models are 'intrinsically linear', the usual practice followed for fitting of these models had been to get rid of the nonlinearity by converting these to a linear model by applying a suitable transformation, like logarithmic, and reciprocal. A disturbing feature of this approach is that it is incorrect and leads to erroneous conclusions. Prajneshu (1991) emphasized that Nonlinear estimation procedures, like Levenberg-Marquardt procedure should invariably be employed to fit these models with additive errors. The author demonstrated that the best model identified for some fisheries data was not only different from the one obtained through the conventional approach but was also biologically more meaningful. Further, the allometric model is widely used for determining the length-weight relationship in animal sciences and fisheries. Generally, the error terms

are assumed to be identically and independently distributed. However, this assumption is quite often violated, particularly for time-series data. Venugopalan and Prajneshu (1997) studied a generalization of allometric model and developed the methodology for fitting the same when the errors follow autocorrelated AR(1) structure.

Prajneshu and Das (1998) carried out a detailed study dealing with modelling of wheat production data at State level in post-Green revolution era. Specifically, several mechanistic nonlinear growth models, viz. monomolecular, logistic, Gompertz, mixed-influence and Richards were applied using Levenberg-Marquardt procedure. The heartening feature of a mechanistic model is that the parameters have specific agricultural/ biological interpretation and provide insight into the underlying mechanism. Six major wheat-growing States, viz. Punjab, Haryana, Uttar Pradesh, Madhya Pradesh, Rajasthan and Bihar were considered. For each of these States, it was found that logistic model has performed the best. It was also concluded that, for the data under consideration, Haryana's performance was the best in view of high intrinsic growth rate of wheat production in that State coupled with the fact that wheat production in that State was still in the fast phase of growth. The utility of these modelling efforts is that it is possible to evaluate the comparative performance of various States as far as wheat production is concerned, This, in turn, may be of immense help to the planners in making important policy decisions. Further, the proportion of area time-series data under high yielding varieties of wheat in India during post-Green revolution era was modelled by Prajneshu and Kandala (2002) through mixed-influence nonlinear growth model. The methodology of replacing one parameter by an 'Expected-value parameter' was employed for fitting of the model. Subsequently, hypothesis testing was carried out by using Wald's test. For the data under consideration, it was concluded that the process of innovation had played a significantly greater role vis-a-vis that of imitation so far as diffusion of high yielding varieties of wheat in India is concerned. This type of result is very satisfying to the research efforts made by Agricultural Scientists. It may be noted that the methodology developed was successful in segregating the effects of the two processes, viz. Innovation and imitation, which otherwise was not possible.

It is well-known that the success rate of fitting the four-parameter nonlinear Richards growth model by using nonlinear estimation procedure is extremely low. Fortunately, a very powerful and versatile optimization technique of Genetic algorithm (GA), motivated by the principles of Genetics and natural selection, has recently been developed. In this methodology, some fundamental ideas of Genetics are borrowed and used artificially to construct search algorithms that are robust and require minimal problem information. The three operators, viz. Selection, crossover, and mutation make GA an important tool for optimization. When a string (parameter solution) is created by GA, it is evaluated in terms of its fitness, which is taken to be the Residual sum of squares. A detailed description of the entire methodology is given by Iquebal *et al.* (2009). The authors have also demonstrated that, for India's total foodgrain time-series data, the GA methodology is successful in fitting the Richards model. The importance of this work is that the GA methodology is applicable even in those cases in which Nonlinear estimation procedures fail to converge.

If there is one concept that has been used the maximum number of times during the past four decades or so in research papers published, particularly in the discipline of Agricultural Economics, it is undoubtedly the 'Computation of compound growth rate'. In a seminal work, the methodology for the same was proposed by Panse (1964). Here, the value of variable under study, like Agricultural production,

productivity, or area at various time-epochs is assumed to be related to the value of that variable at some previous time-epoch through Malthus model. Next, a multiplicative error is generally assumed to describe the random fluctuations. As the underlying model is intrinsically linear, it is transformed to a linear model by means of a logarithmic transformation. The resulting model is fitted to data by the 'Method of least squares' and goodness of fit is assessed by the coefficient of determination $R^2$. Finally, the compound growth rate is estimated as some specific function of one of the parameters. Although, the above methodology has been very widely employed, it has several drawbacks. Firstly, the Malthusian model envisages that the response variable tends to infinity as time tends to infinity, which can not happen in reality. Therefore, continuing to sticking to only this model does not make much sense, particularly when several other more realistic models, like logistic and Gompertz models exist. Secondly, the assumption of multiplicative error tends to be valid only when variability of response variable increases as the variable increases, which rarely holds. The third drawback is that the goodness of fit of the original intrinsically linear model is assessed by reporting the same value of $R^2$ as that obtained for the corresponding linearized model, which is incorrect. A detailed discussion of all these aspects is given by Prajneshu and Chandran (2005). The authors also suggested the correct procedure that needs to be followed for computation of compound growth rates and illustrated the same on India's total foodgrain production time-series data.

The input-output relationship between a response variable and a set of explanatory variables is quite often expressed through Cobb-Douglas production function approach. However, the underlying model is again intrinsically linear. The usual practice is to get rid of the nonlinearity by transforming it into a linear model by means of logarithmic transformation. Prajneshu (2008) highlighted the deficiencies and emphasized that Nonlinear estimation procedures should be employed for fitting the original model. Even when this approach is adopted, the parameter estimates generally exhibit extremely high correlations, implying thereby that the parameters are not estimated independently. To this end, Prajneshu (2008) advocated the use of 'Expected-value parameters' and derived formulae when there are one or two explanatory variables. The methodology developed was demonstrated on wheat yield time-series data of Punjab.

Aphid is a small insect which infests almost every plant. In India, aphids are recognized as serious pests of cereals, oilseeds, pulses and vegetable crops. They cause damage directly to host plants by sucking the plant sap and arrest their growth and development. Accordingly, it is highly desirable to investigate optimal control policies for controlling this pest. To this end, as a first step, Prajneshu (1998) developed a Nonlinear statistical model for describing the dynamics of aphid population growth. The model, in terms of an integro-differential equation, was solved analytically. The formula for obtaining optimal time for insecticidal spray was also derived. Subsequently, the model was successfully applied to ten data sets. Further, it is well established that a plant disease generally progresses at a slow rate in the beginning and subsequently growth rate picks up and reaches the maximum, beyond that again it starts decreasing. This type of behaviour can be captured by several nonlinear growth models, like logistic and Gompertz models. However, one limitation of these models is that the plant parts, viz. number of leaves or number of panicles are assumed to be constant. In reality, as disease grows, plant parts also have same pattern of growth as the disease, though at a slower rate. Accordingly, Chandran and Prajneshu (2005a) theoretically developed generalizations of monomolecular, logistic and Gompertz models and applied these to powdery mildew disease data of mango.

A nonlinear model involving one predator and two prey species was studied by Prajneshu and Holgate (1987). The predators can feed on either species of prey. However, instead of choosing individuals at random, so that the chance that a predator would catch a member of one or other prey species is proportional to their abundances, the predators are assumed to feed preferentially on the most numerous species. This implies a kind of switching from one source of food to another as the prey species alternate in numerical superiority. In population ecology literature, it was believed that switching mechanism invariably leads to stability. However, the authors showed analytically that this is not so and switching mechanism can even have a destabilizing effect.

## 3. Contagious Distributions

Knowledge of Spatial spread is useful in several ways, like Designing efficient sampling programmes for population estimation, development of population models and efficient pest management. To describe the aggregation or clustering behaviour, negative binomial distribution has generally been used. However, there are many instances in which this distribution does not provide a good fit to the data. To this end, Prajneshu and Sarada (1999) advocated the use of some other more advanced contagious distributions, viz. Neyman type A and Thomas distributions. The relevant computer programs for fitting these distributions were also appended using 'Method of moments' as well as 'Method of maximum likelihood' and the methodology was applied to some aphid count data. Similar type of work dealing with the Lagrangian-Poisson distribution (LPD) was also carried out by Sarada and Prajneshu (2000). Although LPD has only two parameters and a simple form, yet it is extremely versatile due to its capability of incorporating greater changes in variance than mean. The variance of LPD can be greater than, or equal to, or less than the mean according as one specific parameter is positive, zero, or negative. This distribution is generally unimodal except in one particular case, where it is bimodal. Further, Sarada *et al.* (2001) thoroughly studied yet another contagious distribution, viz. Polya-Aeppli distribution and applied it to onion thrips' data. An explanation as to why the negative binomial distribution was not appropriate for this data was also provided. The utility of the work described in this Section is that the research workers should start applying these contagious distributions to describe Spatial spread, particularly in those cases where negative binomial distribution does not provide a good fit.

## 4. Structural Time-series Analysis

Statistical modelling of time-series data in Agriculture is usually carried out by employing the well-known Box-Jenkins Autoregressive integrated moving average (ARIMA) methodology. One disadvantage of this methodology is that the data series under consideration is assumed to be either stationary, or can be made so by differencing or some other means. However, this is not always possible. A very promising approach, which does not suffer from the above drawback, is 'Structural time-series modeling (STSM)'. The distinguishing feature of this methodology is that the observations are regarded as made up of distinct components, such as trend, and cyclical fluctuations and each of which is modelled separately. The techniques that emerge from this approach are extremely flexible and are capable of handling a much wider range of problems than is possible through ARIMA approach. Prajneshu *et al*. (2002) thoroughly studied the STSM approach when there are prominent cyclical fluctuations. Specifically, three models of this type, viz. Cycle plus noise model, Trend plus cycle model, and Cyclical trend model were considered. Estimation of parameters was carried out by putting the models in State space form and then applying

Kalman filter. As an illustration, more than sixty years' data on India's lac production exhibiting prominent cycles, was utilized. It is hoped that Agricultural Statisticians would employ STSM methodology for modelling and forecasting purposes.

Achieving food security is one of the foremost aims of our country's agricultural policy. The total foodgrain production witnessed a three-fold increase in nearly three decades during post-Green revolution era. However, this quantum jump barely keeps pace with the burgeoning population growth. Further, in view of setting up of World Trade Organization, farmers are being asked to produce those crops that are suitable for export in the international market. Therefore, it is highly desirable to model and forecast India's total foodgrain production time-series data. Narain *et al*. (1985) developed India's total foodgrain production projections under two alternative scenarios: (i) Constant trend of input factors used for the utilized period, and (ii) 10% increased level of input factors. The authors first projected the explanatory variables themselves on the basis of their linear annual growth rates and then substituted these in the production functions. Ravichandran and Prajneshu (2002) employed two promising dynamical modelling techniques, viz. Bayesian analysis of time-series and Structural time-series modelling. Utilizing the data from 1966 to 1998, both these procedures yielded the forecast for 2020 between 280 and 285 million tonnes, which was much less than similar forecasts made by some other authors. Therefore, a wake-up call was also issued that India's total foodgrain production in future may not even suffice for domestic consumption unless some drastic steps are taken urgently. Evidently, this type of study would help policy makers in formulating appropriate strategies to face the challenges ahead.

## 5. Parametric Nonlinear Time-series Modelling

The area of parametric nonlinear time-series modelling has been rapidly growing during the last three decades or so. The main feature of these models is that these are 'Nonlinear', unlike ARIMA and STSM methodologies discussed in Section 3. There are some aspects in real data sets, like existence of limit cycles, and chaos, which can not be described through linear approaches. The most promising parametric nonlinear time-series model has been the Autoregressive conditional heteroscedastic (ARCH) model, which was introduced by Robert F. Engle in 1982, and for which he was awarded the most prestigious Nobel Prize in Economics in 2003. The main characteristic of this model is that it allows the conditional variance to change over time as a function of squared past errors, leaving the unconditional variance constant. Accordingly, this model is capable of describing data sets exhibiting volatility. In our country, onion prices have been in the headlines, particularly in 1998-99, for their tear jerking effect on consumers, farmers, and Government alike. The temporal variation in onion prices has an impact on consumers if the prices are high, on farmers if these are low, and on Government on both the occasions. Ghosh and Prajneshu (2003) thoroughly studied the modelling and forecasting of volatile onion price data through AR(p)-ARCH(q)-in-mean methodology. It is satisfying to note that the model was successful in capturing the sudden burst in onion prices during the second half of the year 1998.

One limitation of ARCH model is that the unconditional autocorrelation function of squared residuals, if it exists, generally decays very rapidly compared to what is observed in real data. Accordingly, T. Bollerslev in 1986 proposed the Generalized ARCH (GARCH) model in which conditional variance is also a linear function of its own lags. The main objective of GARCH model is not to give better point forecasts but rather to give better estimates of the variance which, in turn, allows more reliable forecast

intervals leading to a better assessment of risk. Recently, Paul *et al*. (2009) thoroughly studied this model and applied it for modelling and forecasting of India's volatile monthly spices export data using EViews software package. Superiority of GARCH model over ARIMA approach was also demonstrated. This study would be of help to the planners to take appropriate policy decisions well in advance in order to meet the targets set for Indian spices export. This, in turn, would help the farmers engaged in spices production in getting timely advice based on sound statistical methodologies. Further, another family of Mixture distributions, viz. Gaussian mixture transition distribution, Mixture autoregressive (MAR) and MAR-Autoregressive conditional heteroscedastic (MAR-ARCH) nonlinear time-series models may be employed to describe those data sets that depict sudden bursts, outliers and flat stretches at irregular time-epochs. Ghosh *et al*. (2006a) derived the formulae for carrying out out-of-sample forecasting up to three-steps ahead through recursive use of conditional expectation and conditional variance. These results enable us to compute best predictor, prediction error variance, and predictive density. The theory developed was successfully applied for modelling and forecasting of volatile weekly wholesale onion price data.

Another important family of parametric nonlinear time-series modelling is the Self-exciting threshold autoregressive (SETAR) models. A heartening feature of this family is that it is capable of describing cyclical data. Some examples of such behaviour are: India's annual lac production/export data, India's summer monsoon rainfall data and Population-sizes of several fish species having prey-predator type of interactions. Formulae for carrying out multi-step ahead out-of-forecasting were derived by Ghosh *et al*. (2006b). The methodology developed was successfully demonstrated on India's lac export time-series data. A mechanistic interpretation for occurrence of cyclicity in the data was also provided.

## 6. Nonparametric Regression Methodology

In Section 1, the parametric approach that needs to be adopted for Computation of compound growth rates was discussed. Quite often it is noticed that there is no appropriate parametric form to describe the time-series data under consideration in a satisfactory manner. In such cases, the powerful 'Nonparametric regression methodology (NRM)' may be employed. This technique imposes very few assumptions about shape of the underlying functions and is, therefore, extremely flexible. Chandran and Prajneshu (2004a) thoroughly discussed the methodology and developed relevant computer programs. The methodology was successfully demonstrated for computing compound growth rates of India's total foodgrain production time-series data. Further, it is well-known that India is among the largest oil economies in the World occupying a distinct position in terms of diversity in annual oilseed crops. Main oilseed crops are: Groundnut, Rapeseed, Mustard, Soybean, Sunflower, Safflower, Sesame, Niger, Linseed, and Castor. Annual production of oilseed crops was virtually stagnating at around 10 million tonnes over a span of more than 15 years despite considerable increase in area under these crops. Generally, the supply lagged far behind the demand, thus forcing the Government to import large quantities of edible oils. Turning point came in 1986 with the setting up of 'Technology Mission on Oilseeds (TMO)'. Chandran and Prajneshu (2005b) employed the nonparametric regression with jump points methodology for describing India's oilseed yield time-series data. It was satisfying to note that the model was able to capture the quantum jump in oilseed productivity due to the efforts of TMO. The 'Additive nonparametric regression approach', which is an extension of NRM to the situation of two explanatory variables was described by Chandran and Prajneshu (2004b). Superiority of proposed methodology over the well-known Multiple

linear regression methodology for explaining Rice productivity data in terms of Fertilizer consumption and Area under irrigation data was also demonstrated.

During last few years, an extremely powerful nonparametric methodology of 'Wavelet analysis' has been rapidly emerging. Novel idea of wavelets is that these are localized in both time and space, whereas traditional Fourier bases are localized only in frequency but not in time. The theory of wavelets permits decomposition of functions into localized oscillating components and so is an ideal tool for modelling and forecasting purposes. Sunilkumar and Prajneshu (2004) discussed various aspects of this methodology with thresholding and applied the same to model and forecast meteorological subdivisions rainfall time-series data using SPLUS wavelets toolkit. Superiority of the methodology over other competing methodologies, like ARIMA and NRM was also demonstrated. Extension of the above methodology, when the errors are not independent but are autocorrelated was carried out by Sunilkumar and Prajneshu (2008). The methodology was successfully applied for modelling and forecasting of India's marine fish production time-series data.

## 7. Soft Computing

It is, by now, well recognized that Agriculture is a 'Soft science', unlike Physics or Chemistry, which are 'Hard sciences'. In the former, there is always some amount of impreciseness or vagueness or fuzziness in the underlying phenomenon, and/or explanatory variables, and/or response variable(s). So, by assuming a crisp relationship, some vital information is lost. Therefore, for a more realistic modelling, there is a need to incorporate this aspect in traditional models, like Multiple linear regression (MLR) model. It may be emphasized that the traditional statistical methodologies are not capable of handling data in which explanatory and/or response variables are expressed in intervals. Kandala and Prajneshu (2002) studied Fuzzy linear regression methodology for crop yield forecasting using remotely sensed data. The underlying phenomenon was considered as fuzzy and so the value of response variable was expressed in intervals. It was shown that the methodology is not only superior to MLR methodology but is also applicable when explanatory variables are highly correlated.

Presently, more than five lakhs Crop-cutting experiments (CCE) in respect of principal crops of foodgrains, oilseeds and horticultural crops, etc. are conducted every year in our country by the Directorate of Economics & Statistics, Ministry of Agriculture to arrive at crop yield estimates at district level. With the growing demand for micro-level planning, the need for building reliable estimates at small area level, say block or even gram panchayat is imperative. To achieve this, the number of CCE has to be increased many folds, which is not practicable. Additional information about farmers' estimates of crop yields at block level, which are crisp values, could be made use of provided these can explain the actual crop yields, which are fuzzy. To this end, Ghosh *et al.* (2008) studied three methods of fuzzy linear regression analysis, viz. Minimization, Maximization, and Conjunction. Subsequently, the methodology was applied to employ farmers' estimates at block level for modelling cotton crop yield at block levels of Sirsa district, Haryana. The Conjunction method performed the best. It was concluded that the farmers' estimates were able to explain satisfactorily the actual crop yields. It is hoped that these types of studies would go a long way in arriving at efficient crop yield estimates at small area level.

The main limitation of Multiple linear regression (MLR) methodology is that it is useful only when the underlying relationship between response and explanatory variables is assumed to be 'Linear'.

However, in a realistic situation, this assumption is rarely satisfied. Also, if there are several explanatory variables, it is well nigh impossible to have an idea of underlying nonlinear functional relationship. Fortunately, to handle such a situation, an extremely versatile approach of 'Artificial neural network (ANN)' is rapidly developing. A distinguishing feature of ANN that makes it valuable and attractive for a statistical task is that, as opposed to traditional model-based methods, ANN is a data-driven self-adaptive method. This modelling approach with ability to learn from experience is extremely useful in many practical problems since it is often easier to have data than to have good theoretical guesses about the underlying laws governing the systems from which data are generated. Singh and Prajneshu (2008a) investigated a particular type of ANN, viz. Multilayered feedforward ANN. To train such a network, two types of learning algorithms, namely Gradient descent algorithm and Conjugate gradient descent algorithm were employed. The methodology was illustrated by taking Maize crop yield data as response variable and Total human labour, Farm power, Fertilizer consumption, and Pesticide consumption as explanatory variables. Superiority of this approach over MLR was also demonstrated.

Incorporating the aspect of 'Fuzzy logic' in ANN, a rapidly developing area of current interest is 'Neuro-fuzzy', which involves a judicious integration of merits of fuzzy and ANN approaches. Singh and Prajneshu (2008b) thoroughly studied an important model, viz. Adaptive neuro-fuzzy inference system (ANFIS). The model was implemented on fuzzy logic toolbox of MATLAB software package using ANFIS. As an illustration, the methodology was applied for development of a forecasting model for secondary data of yield of banana plants on the basis of data at six different stages of growth using several biometrical characters, like plant height, plant girth, and leaf length as predictors.

## 8. Growth Models in Random Environment

The models considered in Section 1 were generally 'Nonlinear statistical models' obtained by adding an error term to the corresponding deterministic model. However, this approach may not be totally satisfactory, particularly for longitudinal data. Thus for a more realistic modelling, a stochastic term describing fluctuations should be added to the deterministic model and then the resulting model should be investigated. Prajneshu (1983a) considered the stochastic version of the logistic model with continuously distributed time delay. The explicit expression for the stationary distribution was worked out and the effect of time delay on various statistics was discussed. Further, Prajneshu (1983b) studied the stochastic Gompertz model with continuously distributed time delay when the parameters are described by correlated Gaussian white noise stochastic processes. Exact expressions for the first and second order moments of logarithm of population size were derived and stability in mean as well as in mean-square was discussed. Prajneshu *et al.* (1986a) considered a nonlinear stochastic model for the spread of an epidemic when there are seasonal variations in infection rate. The resulting model was analyzed by employing the Diffusion approximation technique. It was shown that, for a large population, the process, on suitable scaling and normalization converges to a nonstationary Ornstein-Uhlenbeck stochastic process. Consequently, the number of infectives has, in the steady state, a Gaussian distribution. Prajneshu *et al.* (1986b) also studied the general two-compartmental system with environmental stochasticity. The transfer rates and outputs were assumed to be described by random telegraph processes. The Laplace transforms of the mean-value functions of the amount of substances present in the two compartments were derived. The Gaussian white noise limit case was also discussed and the stability of the system was examined. It was shown that, while the deterministic model was always stable, environmental stochasticity may induce

in the mean-value function all sorts of behaviours, viz. stable, unstable, and oscillatory.

The von Bertalanffy growth model (VBGM) for describing age-length relationship plays a very important role in fisheries research. Prajneshu and Venugopalan (1999) thoroughly studied the VBGM in a random environment. The fluctuations in the system were assumed to be described by a Gaussian white noise stochastic process. The resulting model, in terms of a stochastic differential equation, was solved analytically it was shown that the probability density function of length of a fish follows a Gaussian stochastic process. As an illustration, the methodology was applied to a set of pearl oyster age-length data.

## 9. Concluding Remark

In the above, salient research work done at IASRI, New Delhi during the last fifty years in seven sub-areas of Statistical modelling has been highlighted. However, there is a need to consolidate this area by continuing to conduct basic, applied, as well as adaptive type of research work in the methodologies described above. This type of endeavour would go a long way in arriving at optimal policies for efficient agricultural management.

## References

Chandran, K. P. and Prajneshu (2004a). Computation of growth rates in agriculture: Nonparametric approach. *J. Ind. Soc. Ag. Stat.*, **57**, 382-92.

Chandran, K. P. and Prajneshu (2004b). Modelling rice productivity: Additive nonparametric regression approach. *Natl. Acad. Sci. Lett., India*, **27**, 261-67.

Chandran, K. P. and Prajneshu (2005a). Nonlinear statistical models for plant disease. *Ind. J. Ag. Sci.*, **75**, 361-63.

Chandran, K. P. and Prajneshu (2005b). Nonparametric regression with jump points methodology for describing country's oilseed yield data. *J. Ind. Soc. Ag. Stat.*, **59,** 126-30.

Ghosh, H., Iquebal, M. A. and Prajneshu (2006). On mixture nonlinear time-series modelling and forecasting for ARCH effects. *Sankhya*, **68**, 111-29.

Ghosh, H. and Prajneshu (2003). Nonlinear time-series modelling of volatile onion price data using AR(p)-ARCH(q)-in-mean. *Cal. Stat. Assn. Bull.*, **54**, 231- 47.

Ghosh, H., Singh, R. K. and Prajneshu (2008). Possibilistic linear regression analysis with fuzzy response variable for crop yield estimation. *Cal. Stat. Assn. Bull.*, **60**, 123-44.

Ghosh, H., Sunilkumar, G. and Prajneshu (2006). Self exciting threshold autoregressive models for describing cyclical data. *Cal. Stat. Assn. Bull.*, **58,** 115-32.

Iquebal, M. A., Prajneshu and Ghosh, H. (2009). Application of genetic algorithm optimization technique for fitting nonlinear Richards model. *Ind. J. Ag. Sci.*,**79**, 399-401.

Kandala, V. M. and Prajneshu (2002). Fuzzy regression methodology for crop yield forecasting using remotely sensed data. *J. Ind. Soc. Rem. Sen.*, **30**, 191-95.

Narain, P., Pandey, R. K. and Sarup, S. (1985). On foodgrain production projections. *Ag. Sit. Ind.*, **40,** 383-88.

Panse, V. G. (1964). Yield trends of rice and wheat in first two Five-Year Plans in India. *J. Ind. Soc. Ag. Stat.*, **16**, 1-50.

Paul, R. K., Prajneshu and Ghosh, H. (2009). GARCH nonlinear time-series analysis for modelling and forecasting of India's volatile spices export data. To appear in *J. Ind. Soc. Ag. Stat.*, **63**.

Prajneshu (1983a). A stochastic logistic model with time delay. *Aust. J. Stat.*, **25**, 130-35.

Prajneshu (1983b). Environmental stochasticity of Gompertz model with time delay. *Comm. Stat.*, **A12**, 1359-69.

Prajneshu (1991). Cautionary note about nonlinear models in fisheries. *Ind. J. Fish.*, **38**, 231-33.

Prajneshu (1998). A nonlinear statistical model for aphid population growth. *J. Ind. Soc. Ag. Stat.*, **51**, 73-80.

Prajneshu (2008). Fitting of Cobb-Douglas production functions: Revisited. *Ag. Econ. Res. Rev.*, **21**, 289-92.

Prajneshu and Chandran, K. P. (2005). Computation of compound growth rates in agriculture: Revisited. *Ag. Econ. Res. Rev*., **18**, 317-24.

Prajneshu and Das, P. K. (1998). Modelling wheat (Triticum aestivum) production at state level in post green revolution era. *Ind. J. Ag. Sci.*, **68**, 23-26.

Prajneshu, Gupta, C.K. and Sharma, U. (1986a). A stochastic epidemic model with seasonal variations in infection rate. *Biom. J.*, **28**, 889-95.

Prajneshu, Gupta, C.K. and Sharma, U. (1986b). Stochastic analysis of environmental fluctuations in a compartmental system. *Biol. Cybernetics*, **53**, 343-46.

Prajneshu and Holgate, P. (1987). A prey-predator model with switching effect. *J. Theor. Biol*., **125**, 61-66.

Prajneshu and Kandala, V. M. (2003). Mixed-influence nonlinear growth model. *J. Ind. Soc. Ag. Stat.*, **56**, 19-24.

Prajneshu, Ravichandran, S. and Wadhwa, S. (2002). Structural time series models for describing cyclical fluctuations. *J. Ind. Soc. Ag. Stat.*, **55**, 70-79.

Prajneshu and Sarada, C. (1999). Some contagious distributions for describing spatial spread of aphids *(Homoptera. Aphididae*) when negative binomial distribution fails. *J. Aphidol.*, **13**, 17-26.

Prajneshu and Venugopalan, R. (1999). Von Bertalanffy growth model in a random environment. *Can. J. Fish. Aqu. Sci.*, **56**, 1026-30.

Ravichandran, S. and Prajneshu (2002). Dynamical modelling and forecasting of India's foodgrain production. *Proc. Natl. Aca. Sci., India*, **B 72**, 37-46.

Sarada, C. and Prajneshu (2000). Lagrangian-Poisson distribution for describing spatial spread of aphids. *J. Aphidol.*, **14**, 33-40.

Sarada, C., Prajneshu and Chander, S. (2001) A contagious distribution for describing spatial spread. *Shashpa*, **8,** 131-37.

Singh, R. K. and Prajneshu (2008a). Artificial neural network methodology for modelling and forecasting maize crop yield. *Ag. Econ. Res. Rev.*, **21**, 5-10.

Singh, R. K. and Prajneshu (2008b). Neuro-fuzzy approach for modelling and forecasting: An application. *J. Ind. Soc. Ag. Stat.*, **62**, 214-20.

Sunilkumar, G. and Prajneshu (2004). Modelling and forecasting meteorological subdivisions rainfall data using wavelet thresholding approach. *Cal. Stat. Assn. Bull.*, **55**, 253-65.

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

Venugopalan, R. and Prajneshu (1997). A generalized allometric model for determining length-weight relationship. *Biom. J.*, **39**, 733-39.

# Statistical Applications in Breeding and Genetics

**A.R. Rao, S.D. Wahi and V.K. Bhatia**

## 1. Introduction

The initiation of the work in statistical genetics involving analysis of breeding data on Beetal goats dates back to 1940 when this Institute in its formative stages was working as a statistical section of ICAR. This work under the illustrious leadership of Professor P. V. Sukhatme led, for the first time, to the appreciation of the power of Statistics in drawing inferences on issues in animal sciences and other fields. With the reorganization of research activities under the Indian Council of Agricultural Research during the 70s the stage was set for making the impact of this discipline felt in the research arena of the country. After the recognition of this Institute as a full-fledged Institute of ICAR the work on statistical genetics, which involved research initiatives, both of theoretical and applied nature in plant and animal breeding was carried out in the Division of Animal Sciences. The activity of research in statistical genetics was, further strengthened by carving out a "Statistical Genetics Cell" from the Division of Animal Sciences in 1978. Realizing the importance of this area of research and the amount of work done in the past, the 'Quinquennial Review Team (1971-1981)' as well as the some of the UNDP experts who visited this Institute from time to time recommended that research in the field of statistical genetics and other areas like biometry, bio-assay and bio-statistics must be carried out on a much larger scale rather than by a small cell. Consequent on these recommendations the Division of Bio-statistics and Statistical Genetics came into existence in March 1985. Subsequently, the name of the Division was changed to 'Biometrics', in 1998. New theoretical developments were made from time to time and numerous methodologies for application in plant and animal breeding and related areas were developed. The major research contributions made during last 25 years by the scientists and students in the area of statistical genetics are reviewed. Prior to this period the work has been exhaustively reviewed by Narain *et al.* (1987).

In the next few sections we have given a brief account of the important research and academic accomplishments in the area of statistical applications in genetics and breeding. In the discussion of research achievements, the material has been organized under different heads. At the end a complete list of publications of the scientists, that formed the main basis of the highlights of research achievements, have been provided in this article.

## 2. Estimation of Genetic Parameters

Most of the economic characters in agricultural crops and dairy animals are quantitative in nature. There is a wide range of variability in these characters which depends on the genetic make up of the individuals and the environment in which they are grown. Breeders use this variability for getting improvement in economic characters through efficient selection strategies. The information on genetic parameters, such as heritability, repeatability and genetic correlation is a pre-requisite for making efficient selection strategies by the geneticists and breeders. Keeping in view the importance of these parameters a number of studies have been conducted at the institute, which are as follows.

## 2.1 Estimation of heritability

There are two main approaches to the estimation of heritability ($h^2$) in farm animals, one based on intra-sire regression of offspring on dam and the other based on variance components from full-sib or half-sib analysis. Heritability being a fraction of additive genetic variance to the phenotypic variance should normally lie between 0 and 1. However, when estimated from sample, it frequently turns out to be either negative or takes values exceeding unity due to sampling fluctuations. Such estimates are termed inadmissible estimates. All that has been tried at the Institute is to obtain admissible estimate of heritability and a precise estimate of its variance.

### The problem of inadmissible estimates

Extensive investigations on this problem led to the working out of optimum sample sizes and structures (Prabhakaran and Jain, 1987, a, b, c, 1988, Prabhakaran and Sharma, 1994a, 1995). However, as pointed out by Prabhakaran and Jain (1988) all such probabilities can be reduced to F cumulative probabilities through Satterthwaite(1946) approximation and then computed using Hastings (1955) procedure, without losing much in accuracy. Jogendera Singh (1992) extended the study to cover the heritability under the practical definition in which heritability is the fraction of total fixable genetic variance to the total phenotypic variance. He also computed the probabilities for the unbalanced hierarchical model through Monte Carlo simulation. His results were largely in agreement with those of Prabhakaran and Jain(1988). He also observed that the data imbalance did not pose much of a problem when the observations are fairly good in number and are taken in an optimum fashion. Walke (1993) computed the probabilities of inadmissible estimates of intra-sire regression empirically by resorting to the simulation procedure of Ronningen (1974) and found that the probabilities were in close agreement with the theoretical probabilities as reported in Prabhakaran and Jain (1987a). Subsequently Shukla (1993) considered heritability estimation under finite population assumptions, its effect on the probability of inadmissible estimates and the related saving in sample size requirement.

### Improved estimators of heritability

Garg *et al.* (1984) estimated the information on genetic parameters such as heritability of economic characters as well as on vital characteristics such as rate of mortality, infertility, abortion, sex-ratio involuntary culling etc. in respect of the Indian herds of cattle and buffaloes. When non-availability of suitable data becomes a major constraint, the only way to obtain meaningful estimates of heritability is to go in for restricted estimators, which can shrink the estimates so as to make them fall within the permissible range. Accordingly, a number of restricted estimators were proposed (Verma and Jain, 1989, Prabhakaran and Jain 1990) and illustrated (Prabhakaran and Seema Jaggi, 1996) for improved estimation of heritability. Kaur and Bhatia (1993) conducted an empirical study of different restricted estimation procedures and found that the estimator with minimum average squared bias is superior to other estimators including the MINQUE and MINIMAX estimators of Verma and Jain (1989).

### Variance estimation

Prabhakaran and Jain(1987d) showed that the traditional formula for variance of intra-sire regression heritability is not tenable in view of the random sampling of parents involved and suggested a new expression for the unconditional variance by considering both the offspring and dam observations as random variables. Prabhakaran *et al*. (1990) proposed two new estimators of this variance and compared

for their efficiency. A better approach to the estimation of variance of broad sense heritability was suggested by Prabhakaran and Jain (1986). Prabhakaran *et al*. (1991) based on probabilistic arguments established an identity useful in finding the moments of sample moments of bivariate normal random variables, with applications in many practical situations, established. The variance of intra-sire regression heritability was derived to illustrate its use. The bootstrap procedure was also used (Jugnu *et al.* 1999, Singh and Wahi, 2001) for evaluation of different mating designs, from the point of view of getting precise estimates of heritability. More specifically, the investigation is focused on two aspects, the optimum sample size and structure for half-sib and full-sib mating designs and the design giving least variances under optimum structures, depending upon the level of population heritability. Wahi and Rao (2004, 2005) made an empirical investigation on the estimation of heritability in presence of non-genetic fixed effects under half-sib mating design and it was found that the method of fitting constants commonly used for estimation and adjustment for all fixed effects taken in the model irrespective of their significance, prior to estimation of heritability, has been found comparable in terms of percent bias to that of estimates based on mixed model techniques. Prabhakaran and Rao (2008) developed a new approach for estimation of variance of sample heritability from full-sib analysis and they found that the variance estimates obtained by the new approach are comparable to that of the bootstrap estimates and can be used for estimation of variance of heritability.

**Estimation using bootstrap technique**

Application of bootstrap technique for the variance estimation of genetic parameters was initiated in 1992 in the form of a research project entitled Application of bootstrap technique for studying the statistical properties of genetic parameters. Subsequently this technique was used for the estimation of variance and confidence intervals of heritability and genetic correlation [Bhatia *et al.* 1994; Bhatia and Jayashankar, 1996 a,b, 1998; Wahi *et al.* 1998]. Rao (1997) and Rao and Prabhakaran (2001) proposed an approach of obtaining robust estimates of heritability using bootstrap technique on components of variances.

No formulae give the exact variance of heritability and genetic correlation estimators; the expressions available are quite complex. Therefore, a number of studies (Bhatia *et al.* 1994; Bhatia and Jayasankar, 1996 a,b,1998) were conducted on the suitability of bootstrap technique for obtaining a reasonably good estimate of standard error and in computing interval estimates for these parameters. Bhatia *et al.* (1994) discussed its use for variance estimation. Bhatia and Jayasankar (1996a) computed confidence intervals as well as bias corrected confidence intervals for half-sib correlation heritability adopting the non-parametric bootstrap procedure and demonstrated the workability of the procedure. The procedure was extended to the unbalanced data situation by Bhatia and Jayasankar (1996b). They compared the non-parametric and parametric bootstrapping procedures for estimation of precision and determination of confidence intervals for the half-sib and offspring-parent regression heritabilities and showed that the adequacy of non-parametric approach for any kind of sample data. Subsequently, Wahi *et al.* (1998) applied the non-parametric bootstrap technique to obtain the standard errors and percentile as well as bias-corrected percentile intervals for heritability and genetic correlation estimated from sib analysis and parent-offspring regression. They found that the bootstrap estimates of variance of heritabilitity were always very much lower than those computed from approximations and nearly 200 and 1500 bootstrap replications were needed for obtaining a reliable estimate of standard error and confidence interval respectively.

**Bayesian estimation of heritability**

Rao and Kumar (2001) used Bayesian Using GIBSS Sampling (BUGS) to obtain numerical estimates of parameters of posterior distribution and variance components along with heritability under half-sib mating design with small sample data. A comparison of this approach with traditional approaches, *viz.*, ANOVA, Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) establishes the suitability of the former over later for heritability estimation. Kumar *et al*. (2004) further used BUGS approach to estimate variance components and heritability under two-way nested random effects model. Here too Bayesian technique out performs the other traditional approaches for heritability estimation.

**Robust estimation of heritability**

Kiran *et al*. (2003) studied the empirical distributions of the estimates of heritability from half-sib and full-sib data using analysis of variance estimates of variance components. They found that under half-sib mating design, the heritability estimates always follow normal distribution irrespective of the population parametric values. However this is not true for all the situations of full-sib mating design, particularly at low parametric values of heritability. Under sire component estimation of full-sib heritability, the form of distribution is gamma. Kiran (2003) developed a multivariate approach to estimate the variance components and thereby heritability. Estimates obtained under multivariate approach are compared with that obtained under traditional procedures like ANOVA, ML and REML through simulation studies. The estimates obtained through multivariate approach have lower bias and also have lower mean square error than ANOVA and REML estimates. Kiran *et al*. (2004) studied the effects of the aberrant values, outliers and non normality on heritability estimation and it has been found that the robust method of estimation of heritability provides precise estimates as compared to all other traditional methods of heritability estimation.

**Heritability of herd-life in dairy cattle**

For dairy cows length of productive life or herd-life is a trait of major economic importance and is dependent in the culling decision of individual milk producers. Malhotra (1988) studied the retention time and its relationship with production and reproduction traits of dairy cattle and buffaloes. From this study it was concluded that 3/8 breds and 7/8 breds respectively have lowest and highest culling rates among the cross bred animals when the animals have completed their first lactation. Bhatia *et al*. (1992) studied the culling patterns of different categories of dairy animals through a non-parametric approach. Bhatia and Malhotra (1995 a,b) discussed some aspects of estimation and comparison of retention times in dairy cattle as well as the relationships between the retention time and explanatory variables. Using the path analysis (Paul 1997, Paul and Bhatia 2000a) derived formulae for heritability of herd-life from the correlation between half-sib for the traits. The reliability of the formulae has been demonstrated through simulation.

Stayability is a threshold trait in dairy cattle, which needs a special kind of statistical treatment. Herd-life, a measure of stayability depends on a number of traits including milk production. A realistic measure of heritability of stayabilty can be obtained only if the herd-life is adjusted for production and other auxiliary traits. Kaur (1990) developed an index for assessing the worth of a male on the basis of phenotypic values of his male and female progeny for all-or-none traits. It was shown that maximum gain in accuracy was obtained when the auxiliary traits are included in the index and the phenotypic and additive genetic

correlation between main and auxiliary traits are of opposite signs. Magnussen and Kremer (1995) considered the beta-binomial model for estimating heritability of binary traits using the concepts of selection response and realized heritability. Making use of this idea Paul and Bhatia (2000b) first converted the herd-life adjusted for production to a binary trait using threshold probability and then used the resultant trait for the determination of heritability. The performance of beta-binomial method was further compared, empirically, with that of Dempster and Lerner (1950) method and other procedures of computing narrow sense and family-mean heritability. The comparison was done for different family as well as herd sizes in both adjusted and unadjusted herd-life situations using the relative root mean square error as the basis of comparison. From the results, it is concluded that the family size and herd size have an important role in the estimation of heritability of stayability. While the procedures was based on real data, beta-binomial and Dempster-Lerner showed encouraging results and those based on family mean exhibited were quite unreliable estimates of heritability. The work was later extended to cover unbalanced data situations (Paul and Bhatia, 2001,2002a, b). The important message from the studies is that in situations where assigned values of stayability are unavailable the beta-binomial method is preferable to all other procedures. Further, in case of threshold models the adjustment for auxiliary traits is crucial in the estimation of heritability of stayability (Paul and Bhatia 2002b).

Iqbal *et al.* (2004) obtained the estimates of heritability of stayability by different methods and that too for non-normal, unbalanced and more than one auxiliary trait situations. A modified beta-binomial approach has been proposed to estimate the heritability of stayability. Whenever prior information on the relationship between stayability, production and other reproductive traits are available, one should go for adjustment of two characters for arriving at the true estimate of heritability. Behera (2007) studied the heritability of threshold character (mastitis disease) in sahiwal breed of cows by estimating heritability by different methods and found that the estimate based on beta-binomial method is more precise than other methods.

**Yield survival relationships and culling patterns**

Yield survival relationships in dairy cattle and buffaloes were studied in detail by Bhatia (1984) and Narain and Bhatia (1984b) with the help of survivorship and hazard functions, which again brought out the unsuitability of animals, with higher exotic inheritance level, for Indian conditions. The concept of censoring in dairy cattle was introduced (Bhatia, 1984; Malhotra,1988) for the first time and the culling patterns of different categories of crossbred animals were studied(Bhatia *et al*. 1992a) using the culling and retention functions (Bhatia and Malhotra, 1995a,b). Bhatia *et al*. (1992b) studied the genetic aspects of stayability(a term related to longevity of dairy animals), of animals with different proportions of exotic inheritance, through heritability and genetic correlations. In the identification of factors responsible for culling of a cow, the above studies have also made use of the various statistical properties and the form of the distribution of retention times. In another study Bhatia *et al*. (1987) used Bayes discriminatory analysis in identifying important traits influencing culling in dairy cattle. Bhatia and Muzumdar (1990a, b) conducted an interesting investigation on the association between various traits with their continuous discrete distribution mixture.

**Estimation of generalized heritability**

Some work on the estimation of generalized heritability was also undertaken (Lal Chand and Narain, 1991, Narain and Lal Chand, 1994). Similar studies, on growth performance of crossbred

goats were undertaken by Lal Chand *et al*. (1997, 1998b). The relationship among the various genetic groups is ascertained by plotting the second dominant root against the first root of the determinantal equation.

### 2.2 Repeatability estimation in biennial bearing crops

Biennial rhythm or alternate bearing tendency is quite common in perennial crops. The statistical treatment of experimental data is not straightforward when biennial trend is present, because the variance of biennial effect gets confounded with error variance. The estimation of repeatability in perennial crops is one area, which had needed intervention. Two new procedures (Wahi,1994) based on two-year moving averages were proposed and shown to be superior to conventional approaches based on ANOVA and principal components in terms of robustness to biennial disturbances. Wahi and Rao (2001) obtained the expressions for the estimates of variance of these two estimators of repeatability based on two-year moving averages in perennial crops.

### 2.3 Estimation of genetic correlation

Genetic correlation is one of the important genetic parameters widely used by the breeders and geneticists in the selection and improvement programs. For obtaining admissible estimates of heritability and genetic correlation from half-sib analysis Amemya (1985) proposed a procedure based on partitioning the difference of mean sum of squares and product matrices in respect of 'between sires' and 'within sires' into two components, one of which is positive definite, and using the positive definite component for getting the required estimates. Following this, Lal Chand *et al.* (1998a) determined admissible estimates of heritability and genetic correlation utilizing breeding data of Sindhi Cows on three dairy traits and further demonstrated the workability of the procedure.

Evidence in the literature shows that the estimation of genetic correlation is based on the assumption of normality of the data. Deviation from normality assumption, presence of outliers may have influence on the estimates of genetic correlation. The estimates of genetic correlation under the normal, non-normal distributions, in presence of outliers and probability of inadmissible estimates of genetic correlation have been obtained under half sib mating design [Sarika (2004); Sarika *et al.* (2006a,b)]. The bias in the estimates increases considerably in presence of outliers as compared to normal data. There was considerable decrease in standard errors and mean square errors with increase in sample size. Wahi *et al.* (2006) studied statistical properties of genetic correlation using bootstrap technique and obtained the optimum number of bootstrap replications required for estimating the variance of the genetic correlation.

## 3. Gene-flow Technique for Optimum Selection Strategies

In a population with overlapping generations, the gene flow approach which involves a series of matrix operations based on a matrix specifying the transfer of genes between different sex-age groups provides an exact description of selection response. This technique has been modified to include certain more realistic situations involving non-random mating and stage-structured populations (Muralidharan, 1990, Muralidharan and Jain 1992a, b, 1993, 1995). For developing the theory a 'transmittable genetic value' for an individual has been defined as the average value of its expected progeny resulting from any system of mating which is analogous to the concept of 'breeding value' defined for a random mating population. Also, the genotypic value is characterized in terms of transmittable and residual genetic values and the corresponding components of variance are derived accordingly. These variances can be

estimated by the conventional procedure based on resemblance between relatives. Sreekumar (1994) studied the genetic gain on deleting and restricting various traits in different selection strategies.

## 4. Crossbreeding Studies

A number of studies based on military dairy farm data have been undertaken. These studies have helped in working out the optimum level of exotic inheritance for better performance, inferring about the nature of gene action as also in determining the heterosis present. The studies also have helped in suggesting plans for evolving new dairy breeds of cattle.

### Gene action and heterosis

Sharma and Narain (1986) who considered the gene action in crossbred cattle found that the poly-genes controlling the milk yield traits and age at first calving have probably significant additive and dominance effects, while dry period and calving interval have significant dominance effect only. They also introduced a simple procedure for obtaining the generation means of various crossbred generations, given the proportion of exotic inheritance for that generation. In another interesting study Narain and Sharma (1986) by defining the heterozygosity as the proportion of individuals heterozygous at one more loci gave a formula to determine heterozygosity for a given proportion of exotic inheritance. By fitting regression of various economic traits on proportion of heterozygosity / exotic inheritance, they concluded that the half-bred, which is strictly heterozygous, is optimum for stabilizing the breed resulting from Friesian x Sahiwal crosses. In a subsequent investigation (Sharma and Narain,1988) gene interaction in dairy cattle was studied. Later, Sharma and Pirchner (1991) considered the heterosis in Friesian x Sahiwal crosses. For this they derived the formulae for genetic parameters in various generation mean models in terms of the proportions of parental exotic inheritance and fitted these models to military dairy farm data.

Sharma *et al.* (2000) studied the gene action and heterosis in seven lifetime traits of Holstein-Friesian x Sahiwal crosses by fitting various genetic models to military dairy farm data. The additive-dominance model was found to be adequate for all the traits considered except 'average milk yield per day of productive life'. The heterosis measured from mid-parent varied from 15% in 'total life' to 108% in 'milk yield from all available lactations'. Heterosis measured from superior parent ranged from 6% in 'total milk yield of first three lactations' to 75% in milk yield from all available lactations'.

### Explanation for curvilinear response

Although most of the crossbreeding experiments, in dairy cattle, have revealed a curvilinear relation between production and level of exotic inheritance, very little was known about the reasons behind this curvilinear response. In a couple of investigations (Prabhakaran and Sharma, 1994b; Prabhakaran, 2001) the 'fitness reversal' and Genotype x Environment interaction hypotheses of curvilinearity were tested empirically, using military dairy farm data. It was inferred that the curvilinearity is due to the fitness reversal effect of the production genes and GE interaction has nothing to do with this phenomenon.

## 5. Genotype by Environment Interaction

Genotype-environment interaction (GEI) and yield stability is an area of current interest. The success of crop improvement activities largely depends on the identification of superior varieties for mass propagation. A variety can be considered superior if it has potential for high yield under favorable

environment, and at the same time a great deal of phenotypic stability. A number of statistics, parametric as well as non-parametric have been proposed for the measurement of yield stability. Narain and Bhatia (1984a) have discussed various statistical methods for the analysis of GEI. They also discussed the usual regression approach for explaining GEI by using a non-additive model and the use of external variables to assess the environment along with the basic concepts of stability.

**Inter-relationships among stability parameters**

Rao (1993) and Rao and Prabhakaran(2000) established certain theoretical inter-relationships among common stability parameters. Knowledge of these relationships is useful from computational point of view as well as in knowing the theoretical basis of the observed similarity in the behaviour of various stability parameters.

**Non-parametric stability measures**

There is ample justification for the use of non-parametric measures in the assessment of yield stability of crop varieties. Chief advantages are: (i) No assumptions about the phenotypic observations are needed, (ii) Sensitivity to measurement errors or to outliers is much less compared to parametric measures, (iii) Additions or deletions of one or a few genotypes do not cause distortions to non-parametric measures. (iv) Most of the time, the breeder, is concerned with crossover interaction, an estimate of stability based on rank-information, therefore, seems more relevant, (v) These measures are particularly useful in situations where parametric measures fail due to the presence of large non-linear GEI. For these reasons, non-parametric measures are widely employed in the selection of crop varieties especially when the interest lies in cross over interaction [Thennarasu,1995; Raiger, 1997, Raiger and Prabhakaran, 2000, 2001].

It is a known fact that the non-parametric methods are less powerful than their parametric counterparts. The studies conducted against this background by Raiger (1997) and Raiger and Prabhakaran(2000) have shown that when the number of genotypes in the trial is fairly large, the power efficiency of the non-parametric measures will be quite close to those of the parametric measures. So in situations, which are commonly encountered, i.e. those involving a good number of genotypes being performance-tested in a set of environments whose number is neither too small nor too large, the risk of selecting inferior genotypes from the use of non-parametric measures is minimal.

**Simultaneous selection measures**

Integration of stability with performance through suitable measures will go a long way in selecting high yielding, stable cultivars. Bajpai and Prabhakaran (2000), therefore, developed three new indices for selection of genotypes simultaneously for yield and stability and showed empirically that these indices were superior to Kang's (1993) rank-sum method, which has an inherent weakness that it is weighing heavily towards better yield performers, apart from the arbitrariness in the scoring procedure. Rao *et al.* (2003) and Rao and Prabhakaran (2005) developed statistical procedures based on AMMI model for selecting genotypes simultaneously for high yield and stability. They found that their indices are superior to Bajpai and Prabhakaran (2000) indices. Rao *et al.* (2004) developed two computer programmes, *viz.*, SISGYS1 and SISGYS2 for selecting genotypes simultaneously for yield and stability under genotype x environment x year and genotype x environment situations.

**Stability of several traits**

Another important question that has to be answered by a varietal testing programme is regarding the assessment of stability simultaneously for several traits. Since economic value of a variety is dependent on a number of traits, it is necessary that stability analysis be performed simultaneously for all the traits. The multi-trait stability analysis for these traits becomes all the more important because the response pattern of these traits might be quite dissimilar in nature and so the usual single trait analysis will not help to reconcile these contrasting assessments into a unified conclusion. Stability for several traits, inter-relationships among common stability measures and non-parametric measures, has received special attention (Balakrishnan and Jain, 1988 a, 1989b). Bajpai (1998) proposed a MANOVA procedure suitable for carrying out stability analysis for several traits simultaneously and illustrated it on Sugarcane data. The results indicated that the technique would be quite useful whenever the GE interaction is largely linear.

**Study of crop yield stability**

Detection of GE interaction and analysis of yield stability are essential steps in selecting promising crop varieties for large-scale propagation. Various aspects of GE interaction in vegetable crops were considered by Balakrishnan (1986), Balakrishnan and Jain (1988a,b), Balkrishnan *et al.* (1989a,b) and Rao (1993). Broad sense heritability, useful in the selection of promising parents in vegetatively propagated crops, is estimated as a function of variance components in the analysis of variance of multi-location trial data. Rao and Prabhakaran (2001) proposed a bootstrap procedure for obtaining robust estimates of these variance components and illustrated this by computing heritability for different combinations of variety and location numbers. In an empirical investigation on non-linear genotype x environment interactions (Rao and Prabhakaran 2002), the procedures of two-phase regression, refined two-phase regression and non-linear regression were tried on live data on vegetable crops and the strength and weakness of the procedures were brought out. They also suggested a new procedure to overcome the defects in the two-phase regression approach and the procedure was shown to be comparable with Pooni and Jinks (1980) refined two-phase regression procedure. Based on some simulation studies the need for a revisit to the stability concept of crop varieties was highlighted and in-depth study of the GE analysis for the incomplete and unbalanced data sets conducted (Bhatia, 1991a,b, 1992). Kumar (1997) proposed modified BLUP for studying genotype x environment interaction, where genotypes and g x e interaction were considered as random, under heterogeneous variance structure. It was found that accuracy of modified BLUP is more than the conventional BLUP in predicting yield of genotypes in multi-locational trials.

During the last decade new methods have been developed for studying GEI in economically important crops. The performance of different models like Additive Main effects and Multiplicative Interaction Model (AMMI) and Best Linear Unbiased Prediction Model (BLUP) is evaluated based on the "post-dictive' and predictive assessment criteria and it has been found that BLUP is at par with AMMI [Raju (2002 a,b) and Raju *et al.* (2003, 2006, 2009)]. Further he found that varietal assessment by joint regression analysis is highly unsuitable for the case involving large non-linear interactions as its performance is poor as compared to AMMI or BLUP. The ranking abilities of different stability measures have been found to be better under Expectation-Maximization-AMMI (EM-AMMI) with random environmental effects as compared to EM-AMMI and Modified EM-AMMI with fixed environmental effects. A stability

measure $W_{i(AMMI)}$ using EM-AMMI with random environments methodology is recommended to derive stability conclusions from AMMI model when some cells in two-way GEI table are missing. Also, robust and reliable measures of stability have been identified when the GEI effects are not normal (non-normal and contaminated normal distributions).

Incomplete data are primarily the result of a few genotypes having been not tested in all the environments due to various constraints like, insufficient seed, non-germination and pest and disease attack. To tackle such situations Choudhary (2006) proposed several modified AMMI procedures and compared their performance with the traditional methods for selecting genotypes for high yield and stability under incomplete genotype x environment data. Several new simultaneous selection indices under incomplete GEI data have been developed and tested for their performance in selecting genotypes for high yield and stability. Robustness of simultaneous selection indices against different levels of missing observations has been studied. Few simultaneous selection indices have been identified as robust indices even with the occurrence of 20% missing values.

## 6. Studies on Spatial Patterns

In field experiments the newer concept of spatial relationship among adjacent observations was also brought to the notice of researchers by using geo-statistical techniques such as variography and kriging. Using the notion of co-kriging, the influence of soil characteristics on performance of plant growth traits was illustrated (Samra *et al*. 1987, 1988a,b and 1989). Bhatia and Prasad (2005) studied data processing techniques for statistical analysis of large field variability in hilly and salt affected soil regions and developed software package *Spatial Variability and Interpolation*, which visualize the spatial variability in graphical form was developed. Programs for kriging in regular grids, with four inbuilt fitting models viz., Spherical, Gaussian, Exponential and Logarithmic were also developed.

## 7. Modelling of Scientific Phenomena in Plant and Animal Breeding

### Discriminant and classificatory analyses

Garg *et al*. (1984) using the linear discriminant function showed that infusing exotic inheritance beyond 75% level may result in to significantly lower lactation yield coupled with an enhanced calving interval. For studying the divergence between genetic groups in crossbred cattle Jaiswal and Jain (1989, 1990) proposed genetic group indices on multiple traits under both homogeneous and heterogeneous covariance structures. Comparing the performance of different genetic groups of dairy animals based on multiple traits is likely to yield better information on the economic value of these animals as well as on the optimum level of exotic inheritance at which the crossbred have to be stabilized. Keeping this view, a number of performance indices based on repeated records of animals and linear discriminant function were developed (Lal Chand ,1988, 1990). On the basis of these indices it was concluded that 50% exotic blood is optimum from the point of overall merit of the animal.

In classificatory analysis, the use of Fisher's linear discriminant function rests on the assumption that the relevant data follow multivariate normal distribution and variance- covariance matrices of different populations are equal. These assumptions are seldom fulfilled in the case of performance data generated from crossbreeding experiments. Accordingly, in the comparison of different grades of sheep, Narain *et al*. (1991) used the linear discriminant function based on minimax procedure for classification and found

that this procedure was superior to the conventional method. This was confirmed, by Wahi and Bhatia (1995), in their study based bootstrap technique. Wahi and Kher (1991) made a comparison of different multivariate methods namely, Principal Components Analysis (PCA), Tocher's method and an iterative method of clustering based on successive reallocation of elements based on Mahalanobis $D^2$ statistic using multiple trait data of Gerbera and Dahlia flowers. It was observed that the clusters formed by iterative method were uniformly more homogeneous and unique as compared to the cluster obtained by the other two methods.

Dash (2007) classified seventy-seven maize (*Zea mays* L.) genotypes by 6 different clustering methods including Artificial Neural Networks (ANN) and compared them based on probability of misclassification. It was found that the performance of ANN method is the best among the six methods of clustering irrespective of the sample size and dissimilarity measures used. Wahi *et al.* (2009) compared five different classical clustering methods and found that the Ward's method performed best with least average percentage probability of misclassification followed by non-hierarchical k-means method irrespective of the sample size. Among the different distance measures used under hierarchical clustering methods, the squared Euclidian distance showed least average percentage probability of misclassification followed by city block distance.

**Modelling in crop-insurance**

Crop insurance is a system, which ensures the farmer that his return from the crop will not fall below a certain level in times of partial or complete failure. Although the comprehensive crop insurance scheme was introduced in the country in 1985, there was no sound methodology for working out the insurance premium as well as indemnity payable to the farmer when he suffers a loss due to crop failure. Mittal (1989) studied the statistical aspects of crop insurance on normal and transformed yields of paddy and wheat crops in Uttar Pradesh. The premium rates calculated by kernel method of estimation, which is a non-parametric method and independent of any sort of assumptions, were much lower than the normal curve technique. Further, investigations were taken up (Garg *et al.*, 1991, 1994) which were mainly concerned with density estimation for determination of indemnity and insurance premium based on the observed distribution of crop yields.

**Compartmental models**

Compartmental models are of immense utility in rumen fermentation studies. Keeping this in mind a few stochastic compartmental models for physiological kinetics in animal nutrition were developed by Rana and Narain (1992).

**Growth models**

Sarkar (1998) compared different mathematical models, viz., Wood's model, Morant-Ganasakthy model and Mitscherlich exponential model, for describing lactation yield of cross bred cattle and found that Morant-Ganasakthy model was found to be best among the three and explained maximum variation in lactation-yield. Thus, it can be used by the dairy farm manager and animal breeders. Kolluru (2000) studied growth patterns of cross bred cattle using five different non-linear growth models. They found that Richard's and Logistic models are suitable and equally good to explain the growth patterns in cross bred cattle. Lal Chand *et al.* (2002) and Wahi *et al.* (2004) applied non-linear models to study the growth pattern in Indian breeds of goats and it was found that Gompertz model showed best fit among different

non-linear models. Growth studies serve as an aid in assessing the maximum production potential of livestock. Kundu (2005) has estimated heritability and genetic correlation for the body-weight data of pig species and estimated growth curve parameters (logistic) using full sib mixed model and half sib mixed model considering sex as fixed effect. Based on the estimate of inflection parameter the logistic model was found to be best fitted for growth curves.

## 8. Statistical Genomics and Bioinformatics

### Sire evaluation under MOET

In the past the statistical analysis of breeding data was handicapped due to non-availability of sufficient amount of data. This constraint having been removed under Molecular Ovulation and Embryo Transfer (MOET) technology, the current need is of efficient statistical analysis of data generated under these programmes. As a preliminary initiative in this direction, Sethi and Jain (1993a, b, 1994) examined the merits of sire evaluation based on multiple traits and partial records and proposed procedures suitable under MOET technology.

### Classification of genotypes using RAPD and DNA finger prints

Random Amplified Polymorphic DNA (RAPD) analyses and Amplified Fragment Length Polymorphic DNA (AFLP) analyses, in particular, has been shown to reveal a significant level of DNA polymorphisms in different plant species. These methods are useful for effective germplasm management with respect to estimating diversity, monitoring genetic erosion and removing duplicates from germplasm collections. Notwithstanding the problem of non-reproducibility, these techniques have greater resolving power than the morphological characteristics. Thomas *et al.* (2006) studied the classification of Indian wheat varieties using RAPD based DNA finger prints as well as morphological characters. Kolluru (2006) and Kolluru *et al.* (2007) developed a method to identify a suitable clustering procedure, which could accurately classify sugarcane cultivars, when the AFLP marker data contain missing observations. Among different clustering methods available, fuzzy based clustering technique was identified as the best method to classify AFLP marker based data. Determination of optimum number of markers and their identification by retaining the maximum variability present in the data is a challenging issue. Nearly 35 STMS markers (optimum number) are required for studying the genetic diversity in rice, while retaining the maximum precision in Genetic Distance. Estimation of the standard error of correlation of molecular diversity ($\Phi_{ST}$) in AMOVA (Analysis of molecular variance) has been dealt by bootstrap technique and percentage significance of $\Phi_{ST}$. The estimation of $\Phi_{ST}$ is also done by DNA marker site sampling technique.

### Detection of quantitative trait loci

Quantitative traits are the traits controlled by many genes and each of the genes has a small effect on the trait. The loci controlling quantitative traits are referred to as QTLs (Quantitative Trait Loci) and the procedure of finding and locating the QTLs are called QTL mapping. During past two decades, molecular approaches have advanced in the identification, mapping and isolation of genes for various crop species. Jambhulkar (2007) examined the inheritance of traits by using molecular genetics and gene mapping techniques. Further, quantitative traits are influenced by the environment and tend to show varied degree of QTL×Environment Interaction (QEI). The results on QTL detection have been obtained by considering the main-effect markers and interaction between markers in the model. The results reveal that Jackknife

procedure and Bayesian methodology have been performed better over traditional procedures of QTL detection even in the presence of QEI.

**Bioinformatics**

World agriculture achieved a major boost with the completion of the rice genome sequencing in December 2004. Rice is the most important food source for half the world's population. It is the first crop plant to be sequenced and will, therefore, have a great impact in agriculture. The accurate, map based sequence has already led to the identification of genes responsible for agronomically important traits such as genes that affect growth habit to promote yield and photoperiod genes to extend the range of elite cultivars. This could probably provide the key in improving yield to feed an expanding world population at a time of increasing constraints on agriculture. The entire rice genome sequence is available in public domain. Recently, a project on "Computational analysis of SNPs at functional elements of rice genome" has been taken up at the institute in collaboration with National Research Centre for Plant Biotechnology (NRCPB), New Delhi with the objectives (i) to develop a web based information system on functional elements of rice genome (ii) to design and develop a database on SNPs at functional elements of rice genome (iii) to provide online computational facilities to the users engaged in genomic data analysis and (iv) to annotate SNPs in different features of genome through visual graphic display.

The distribution and function of SNPs is an important area of current research in rice genomics. SNPs are useful for genome wide mapping and study of genes responsible for diseases. As the number of SNPs in public databases continues to grow, identifying functional variants has become an important goal of rice genomics. SNPs can alter the function of DNA, RNA and Proteins, and are generally classed by genome location. Efforts are being made at the institute to locate / predict variations that are likely to have effects on functions of different features of genome. Also, it is being tried to develop a local database and information system on functional elements of rice genome

The sequenced genomes of a wide range of organisms, *viz*., *C. elegans*, *D. melanogaster*, *A. thaliana*, *M. musculus*, *H. sapiens* allow global, comparative analyses of regulatory sequences. The genomic set of splice-site sequences corresponds to a large-scale splicing experiment performed by nature under evolutionary constraints. Roca *et al.* (2008) focused on 5' splice-site (5_ss) sequences of the U2-type GT-AG class, which comprise over 98% of all splice sites, and use disease-causing mutations, human single nucleotide polymorphisms (SNPs), and variations in natural splice sites in the genome (within and between species) to infer properties inherent to 5_ss, with important implications for human genetics. Recently at the institute level, efforts are being made to study the characteristic features of 5' splice sites of rice genome and compare them with other species.

**References**

Bajpai P.K. (1998). Some contributions to improved estimation of genotype x environment interaction and analysis of genotypic stability with special reference to sugarcane crop. *Unpublished Ph.D. Thesis*, P.G. School, IARI, New Delhi.

Bajpai, P.K. and Prabhakaran, V.T. (2000). A new procedure for simultaneous selection for high yielding and stable crop genotypes. *Indian J. Genet.* **60(2)**, 141-152.

Balakrishnan, K.A. (1986). Studies on some aspects of genotype-environment interactions in vegetable crops. *Unpublished Ph.D. thesis*, P.G. School, IARI, New Delhi.

Balakrishnan, K.A. and Jain, J.P. (1988a). A procedure for stability analysis with multivariate data in vegetable crops. *Indian J. Genet.* **48**, 201-205.

Balakrishnan, K.A. and Jain, J.P. (1988 b). Exponential-quadratic regression model for the analysis of non-linear G x E interactions in vegetable crops. *Ann. Agric. Res.* **9**, 229-233.

Balakrishnan, K.A., Jain, J.P. and Dutta, O.P. (1989a). Varietal regression analyses of incomplete data on onion and tomato by three different methods of fitting constants. *Indian J. Genet.* **49**, 203-207.

Balakrishnan, K.A., Jain, J.P. and Kataria, A.S. (1989b). Hanson's single parameter approach vs. two-parameter approaches in studying GE interaction in vegetable crops. *Ann. Agric. Res*. **10**, 125-130.

Behera, S.K. (2007). Study of heritability of threshold character. *Unpublished M.Sc. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Bhatia, V.K. (1984). Some aspects of yield survival relationships in dairy cattle. *Unpublished Ph. D. Thesis*, P.G. School, IARI, New Delhi.

Bhatia, V.K. (1991a) Some aspects of stability of crop varieties, Proceedings of National Symposium on Statistical Methodology for Dryland Agriculture held at CRIDA, Hyderabad, 232-248.

Bhatia, V.K. (1991b). On the modified regression analysis for assessing stability of crop varieties. *Ann. Agric. Res*. **12 (4)**, 372-377.

Bhatia, V.K. (1992). A comparison of two methods for assessing stability of crop varieties. *Ann. Agric. Res*. **13 (1)**, 44-48.

Bhatia, V.K. and Jayasankar, J. (1996a). On use of bootstrapping for estimation of confidence interval of heritability. *J. Ind. Soc. Agril. Statist.*, **48**, 217-223.

Bhatia, V.K. and Jayasankar, J. (1996b). A bootstrap look on the precision of Heritability estimators in unbalanced data. *J. Ind. Soc. Agril. Statist.*, **49**, 289-306.

Bhatia, V.K. and Jayasankar, J. (1998). Precision of heritability estimator – A computer intensive look. Proceedings of a Symposium on Computer technique in agricultural research and economic reforms in agriculture sector – A statistical assessment. Published by Indian Society of Agricultural Statistics held during 19-21 December, 1996 at IASRI, New Delhi (Golden Jubilee Conference): 72-81.

Bhatia, V.K. and Malhotra, P.K. (1995a). On some aspects of estimation and comparison of retention times of dairy cattle. *Ind. J. Dairy Sci.*, **48 (2)**, 103- 109.

Bhatia, V.K. and Malhotra, P.K. (1995b). Relationship between retention times and explanatory variables in dairy cattle. *Ind. J. Dairy Sci.*, **48 (6)**, 438-441.

Bhatia, V.K. and Mazumdar, D. (1990a). Study of the various characters in dairy cattle breeding. *Ind. J. Dairy Sci.*, **43 (4)**, 468-473.

Bhatia, V.K. and Mazumdar, D. (1990b). Studies on mixture of continuous and discrete distributed variables in dairy cattle breeding. *Ind. J. Dairy Sci*. **43 (4)**, 474-487.

Bhatia, V.K. and Prasad. R. (2005). Studies on data processing techniques for statistical analysis of large field variability in hilly and salt affected soil regions, Unpublished NATP project Report, IASRI, New Delhi.

Bhatia, V.K., Jayasankar, J. and Wahi, S.D. (1994). Use of bootstrap technique for variance estimation of heritability estimators. *Ann. Agric. Res.* **15(4)**, 476-480.

Bhatia, V.K., Mahesh Kumar, Malhotra, P.K. and Narain, P. (1987). Bayes discriminatory analysis in culling dairy cows for breeding. *J. Ind. Soc. Agril. Statist.*, **39(2)**, 205-215.

Bhatia, V.K., Narain, P. and Malhotra, P.K. (1992a). Comparison of culling patterns of different categories of dairy animals - Non-parametric approach. *Ind. J. Dairy Sci*., **45(3)**, 114-118.

Bhatia, V.K., Narain, P. and Malhotra, P.K. (1992b). Genetic parameters of stayability of different categories of dairy animals. *J. Ind. Soc. Agril. Statist.*, **44 (2)**, 122-137.

Dempster, E.R. and Lerner, I. M. (1950). Heritability of threshold characters. *Genet.* **35**, 212-236.

Garg, J. N., Narain, P., and Indra Singh (1994). Non-parametric density estimation for determining premium in all-risk crop insurance. Project Report, IASRI, New Delhi.

Garg, L.K., Narain, P., Jain, J.P., Chand, L., Rana, P.S. and Singh, I. (1984). Measurement of genetic improvement due to scientific breeding in cattle and buffaloes. IASRI Project Report, 1-146.

Garg, J. N., Narain, P., Shivtar Singh and Mahesh Kumar (1991). Statistical studies in relation to crop insurance. Project Report, IASRI, New Delhi.

Iqbal, M.A., Paul, A.K. and Bhatia, V.K. (2004). Empirical comparison of different estimates of heritability of herd life using unrelated auxiliary traits under non-normal situations. *Ind. J. Appl. Stat.*, **8**, 59-72.

Jaiswal, U. C. and Jain, J.P. (1989). Genetic group index on multiple traits for homogeneous covariance structure. *Biom. J.* **31**, 993-996.

Jaiswal, U. C. and Jain, J.P. (1990). Genetic group index on multiple traits for heterogeneous covariance structure. *Biom. J.* **32**, 357-360.

Jambhulkar, N.P. (2007). Statistical investigations on QTL detection in maize. *Unpublished M.Sc. Thesis*, PG School, IARI, New Delhi.

Jogendera Singh (1992). Some contributions to estimation of heritability. *Unpublished Ph. D. Thesis*, P.G. School, IARI, New Delhi.

Jugnu Ansari, Prabhakaran, V.T. and Rao, A.R. (1999). A bootstrap - variance approach to the choice of best mating design for heritability estimation. *Indian J. Appl. Statistics*, **5**, 29-40.

Kaur, Amrit Pal (1990). On some aspects of progeny testing and evaluation of dairy sires. *Unpublished Ph.D. Thesis*, PG School, IARI, New Delhi – 12.

Kaur, Amrit Pal and Bhatia, V.K. (1993). Empirical study of different estimators of heritability. *Ind. J. Dairy Sci*., **46(7)**, 296-301.

Kiran, P.S. (2003). On some statistical properties of estimates of heritability, *Unpublished Ph.D. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Kiran, P.S., Bhatia, V.K. and Rao, A.R. (2003). Sampling distributions of heritability. *J. Ind. Soc. Agril. Statist.* **56**, 294-301.

Kiran, P.S., Bhatia, V.K. and Rao, A.R. (2004). Robust method of estimation of heritability. *J. Ind. Soc. Agril. Statist.* **57**, 116-128.

Kolluru, R. (2000). On some aspects of growth patterns of cross bred cattle. *Unpublished M.Sc. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Kolluru, R. (2006). Statistical investigation on molecular marker based classification of crop varieties. *Unpublished Ph.D. Thesis*, Post Graduate School, IARI, New Delhi.

Kolluru, R., Rao,A.R., Prabhakaran,V.T., Selvi,A. and Mohapatra,T. (2007). Comparative evaluation of clustering techniques for establishing AFLP based genetic relationship among sugarcane cultivars. *J. Ind. Soc. Agril. Statist.* **61 (1)**, 51-65.

Kumar, S. (1997). On predictions of varietal performance under heterogeneous error variance structure. *Unpublished M.Sc. Thesis*, PG School, IARI, New Delhi.

Kumar, S., Rao, A.R. and Bhatia, V.K. (2004). Bayesian estimation of heritability in animal breeding experiments under 2-way nested classification. *J. Ind. Soc. Agril. Statist.* **58(3)**, 352-62.

Kundu, M. (2005). Genetic analysis of growth curves in pig. *Unpublished M.Sc. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Lal Chand (1988). Performance of crossbred cattle based on multiple traits. *Ind. J. Dairy Sci*., **41**, 208.

Lal Chand (1990). Use of discriminant function to study the relationship among different genetic groups based on multiple traits in dairy animals. *Ind. J. Dairy Sci.* **43(2)**, 214-220.

Lal Chand and Narain, P. (1991). Investigations on appropriate statistical methods for comparing genetic groups based on multiple traits in Dairy Animals. Project Report, IASRI.

Lal Chand, Wahi, S.D. and Bhatia, V.K. (1997). Relationship among different genetic groups for growth traits in goats. *Ind. J. Dairy Sci.*, **49(8)**, 511- 516.

Lal Chand, Wahi, S.D. and Parkash Lal (1998a). Estimation of heritability and genetic correlation in the parameter space. *Indian J. Animal Genetic and Breeding*, **20(1)**, 37- 40.

Lal Chand, Wahi, S.D. and Bhatia, V.K. (1998b). Growth performance based on multiple traits in goats. In: Advances in Agricultural Statistics and Computer Applications in Agriculture, IASRI Publication, 189-97.

Lal Chand, Wahi, S.D. and V. K. Bhatia (2002). Studies on growth pattern and heritability of fitness traits in Indian breeds of goats. Unpublished Project Report, IASRI, New Delhi.

Magnussen, S. and Kremer, A. (1995). The beta-binomial model for estimating heritability of binary traits. *Theor. Appl. Genet.* **91**, 544-552.

Malhotra, P.K. (1988). Statistical studies on the retention times of dairy cattle and buffaloes. *Unpublished Ph.D. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Mittal, P. (1989). Statistical aspects of crop insurance. *Unpublished Ph.D. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Muralidharan, K. (1990). Some studies on gene-flow technique for optimum selection strategies with overlapping generations. *Unpublished Ph. D. Thesis*, IASRI, New Delhi.

Muralidharan, K. and Jain, J.P. (1992a). Response to selection under non-random mating. I. Partitioning genetic variance. *Biom. J*., **34**(2), 147-152.

Muralidharan, K. and Jain, J.P. (1992b). Response to selection under non-random mating. II. Prediction. *Biom. J*., **34**(5), 633-637.

Muralidharan, K. and Jain, J.P. (1993). Gene flow technique for response to selection under non-random mating. *Biom. J*., **35**, 217-225.

Muralidharan, K. and Jain, J.P. (1995). Gene flow technique for stage-structured populations. *Biom. J*. **37**, 229-234.

Narain, P. (1985). Progeny testing with auxiliary traits, *Biometrics*, **41**, 895-907.

Narain, P. and Bhatia, V.K. (1984a). Statistical aspects of genotypes x environment interactions in plant breeding. Silver Jubilee Souvenir, IASRI Publication. pp. 148-160.

Narain, P. and Bhatia, V.K. (1984b). Some aspects of culling patterns in Indian herds of dairy cattle. Silver Jubilee Souvenir, IASRI Publication, pp. 161-173.

Narain, P. and Lal Chand (1994). Gene raised Heritability and its Estimation. *J. Ind. Soc. Agril. Statist.*, **46(3)**, 318-329.

Narain, P., Malhotra, P.K. and Bhatia, V.K. (1987). A review of research work done in statistical genetics at IASRI, Souvenir, IASRI Publication, New Delhi, 152-174.

Narain, P., Wahi, S.D., Malhotra, J.C. and Garg, L.K. (1991). Minimax linear procedure for comparing different grades of sheep in crossbreeding programmes. *Ind. J. Anim. Sci.*, **61(3)**, 305-310.

Paul Amrit Kumar (1997). On some aspects of estimation of Heritability of stayability for dairy cattle. *Unpublished Ph. D. Thesis*, P.G. School, IARI, New Delhi.

Paul, Amrit Kumar and Bhatia, V.K. (2000a). Heritability of stayability by path coefficient approach when herd-life is influenced by unrelated auxiliary characters. *Sankhya*, **B 62(3)**, 463-71.

Paul, Amrit Kumar and Bhatia, V.K. (2000b). Empirical comparison of different methods of estimation of heritability of stayability in dairy cattle. *J. Ind. Soc. Agril. Statist.*, **53(2)**, 196-206.

Paul, Amrit Kumar and Bhatia,V.K.(2001). Modification of beta-binomial method of estimation of heritability of stayability. *J. Ind. Soc. Agril. Statist.*, **54(3)**, 385-395.

Paul, Amrit Kumar and Bhatia,V.K.(2002a). Comparison of beta binomial with other methods for estimation of heritability of stayability. for different degree of unbalanced data in dairy cattle. *Ind. J. Anim. Sci.*, **72(4)**, 402-405.

Paul, Amrit Kumar and Bhatia,V.K.(2002b). Effect of auxiliary traits on estimation of heritability of stayability. *J. Ind. Soc. Agril. Statist.*, **55(1)**, 63-69.

Prabhakaran, V.T. (2001). Implications of curvilinear response in dairy cows with increasing level of exotic inheritance in relation to a plausible breeding strategy for cattle development. *Indian J. Genet.*, **61(2)**, 87-91.

Prabhakaran, V.T. and Jain, J.P. (1986). On estimation of variance of broad sense heritability. *Ann. Agric. Res.*, **7(2)**, 343-345.

Prabhakaran, V.T. and Jain, J.P. (1987a). Probability of inadmissible estimates of heritability from regression and half-sib analyses. *Biom. J.*, **29(2)**, 219-230.

Prabhakaran, V.T. and Jain, J.P. (1987b). Probability of obtaining negative estimates of heritability from full-sib analysis under a general model of gene action. *Biom. J.*, **29(4)**, 461-469.

Prabhakaran, V.T. and Jain, J.P. (1987c). Probability of heritability estimates from full-sib analysis exceeding unity under a general model of gene action. *Ann. Agric. Res.*, **8(2)**, 221-230.

Prabhakaran, V.T. and Jain, J.P. (1987d). On variance of heritability coefficient obtained from Intra-sire regression. *The Statistician*, **36**, 25-26.

Prabhakaran, V.T. and Jain, J.P. (1988). Probability of inadmissible estimates of heritability from full-sib analysis under a general model of gene action. *Biom. J*., **30**, 47-57.

Prabhakaran, V.T. and Jain,J.P. (1990). Logistic transformation regression estimators of heritability. *Biom.J.*, **32(3)**, 361-366.

Prabhakaran, V.T., Jain, J.P. and Ramesan, P.M. (1990). A note on the estimation of variance of intra-sire regression heritability. *J. Ind. Soc. Agril. Statist.*, **42 (2)**, 234-248.

Prabhakaran, V.T., Mahajan, V.K. and Uma. (1991). On an identity for finding moments of sample moments of bivariate normal random variables. *Austral. J. Statist.*, **33(1)**, 95-102.

Prabhakaran, V.T. and Rao, A.R. (2008). A new approach to the estimation of variance of sample heritability from full-sib analysis. *J. Ind. Soc. Agril. Statist.*, **62 (3)**, 208-213.

Prabhakaran,V.T. and Sharma,B.S. (1994a). Probability of heritability estimates from full-sib analysis exceeding unity based on underlying theoretical distribution. *Biom. J.*, **36(3)**, 341-352.

Prabhakaran V.T. and Sharma, B.S. (1994b). Modelling curvilinear response in crossbred dairy cows with increasing level of exotic inheritance. Unpublished Project Report, IASRI, New Delhi.

Prabhakaran, V.T. and Sharma, B.S. (1995). On improved estimation of heritability. *Indian Journal of Animal Genetics and Breeding*, **17**, 33-37.

Prabhakaran,V.T. and Seema Jaggi (1996). On the estimation of intra-sire regression heritability subject to linear constraint. *Biom. J.*, **38**, 455-460.

Raiger,H.L. (1997).On stability concepts, stability assessments and improved procedures for varietal selection. *Unpublished Ph. D. Thesis*, P.G. School, IARI, New Delhi.

Raiger H. L and Prabhakaran, V.T(2000) A statistical comparison between non-parametric and parametric stability measures. *Indian J. Genet*., **60(4)**, 417-432.

Raiger H. L and Prabhakaran, V.T. (2001). A study on the performance of a few non-parametric stability measures using pearl-millet data. *Indian J. Genet*., **61(1)**, 7-11.

Raju, B.M.K. (2002a). Some statistical aspects of assessing sensitivity of crop varieties. *Unpublished Ph.D. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Raju, B.M.K. (2002b). Study of AMMI model and its biplots. *J. Ind. Soc. Agril. Statist.*, **55**(3), 297-322.

Raju, B.M.K. and Bhatia, V.K. (2003). Bias in the estimates of sensitivity from incomplete G × E tables. *J. Ind. Soc. Agril. Statist.*, **56(3),** 294-30).

Raju, B.M.K., Bhatia, V.K. and Samanth, Kumar, V.V. (2006). Assessment of sensitivity with incomplete data. *J. Ind. Soc. Agril. Statist.*, **60(2),** 118-125.

Raju, B.K.M., Bhatia, V.K. and Bhar, L. M. (2009). Assessing stability of cropvarieties with incomplete data. *J. Ind. Soc. Agril. Statist.*, **63(1)**. To Appear.

Rana, P.S. and Narain, P. (1992). Statistical aspects of physiological kinetics in animal nutrition. Unpublished Project Report, IASRI, New Delhi - 12.

Rao, A.R. (1993). On procedures for studying non-linear genotype x environment interactions. *Unpublished M.Sc. Thesis*, P.G. School, IARI, New Delhi.

Rao, A.R. (1997). Some contributions to improved estimation of genetic parameters. *Unpublished Ph.D. Thesis*, P.G. School, IARI, New Delhi.

Rao, A.R. and Kumar, S. (2001). Bayesian estimation of heritability using Gibbs sampling for half-sib mating design. *Indian Journal of Applied Statistics*, **6**, 12-26.

Rao, A.R. and Prabhakaran, V.T. (2000). On some useful interrelationships among common stability parameters. *Indian J. Genet*., **60(1)**, 25-36.

Rao, A. R. and Prabhakaran, V. T. (2001). A bootstrap method of estimating heritability from varietal trial data. *Indian J. Genet.*, **61(2)**, 87-91.

Rao, A.R. and Prabhakaran, V.T. (2002). Empirical investigations on non-linear genotype-environment interactions applied to vegetable crops. *Indian Journal of Agricultural Sciences*, **72(5)**, 277-80.

Rao, A. R. and Prabhakaran, V. T. (2005). Use of AMMI in simultaneous selection of genotypes for yield and stability. *J. Ind. Soc. Ag. Stat.*, **59(1)**, 76-82.

Rao, A. R., Prabhakaran, V. T. and Singh, A.K. (2004). Development of statistical procedures for selecting genotypes simultaneously for yield and stability. Unpublished Project Report, IASRI, New Delhi – 12.

Roca, X., Olson,A.J., Rao,A.R., Enerly,E.R., Kristensen,V.N., Dale,A.B., Anderson, B.S., Krainer,A.R. & Schidanandam,R. (2008). Features of 5' splice site efficiency derived from disease causing mutations and comparative genomics. *Genome Research*, **18**, 77-87.

Samra, J.S., Bhatia, V.K. and Gill, H.S. (1987). Analysis of variability in afforestation studies in sodic soils. Proceedings of International Symposium on Afforestation of Salt Affected Soils held at CSSRI, Karnal, **3**, 49-63.

Samra, J.S., Gill, H.S., Abrol, I.P. and Bhatia, V.K. (1988a). Analysis of variability in the uniformity trial afforestation of Melia azedarach Lin planted in a sodic soil. *Soil Technology*, **1**, 209-222.

Samra, J.S., Gill, H.S., Abrol, I.P. and Bhatia, V.K. (1988b). Analysis of variability in the growth of a tree plantation in relation to soil sodicity, *Soil Technology*, **2**, 135-145.

Samra, J.S., Gill, H.S. and Bhatia, V.K. (1989). Spatial stochastic modelling of growth and forest resource evaluation, *Forest Sci.*, **35 (3)**, 663-676.

Sarika (2004). A study on the robustness of estimates of genetic correlation. *Unpublished M.Sc. Thesis*, P.G. School, IARI, New Delhi.

Sariaka, Wahi, S.D. and Rao, A.R. (2006a). Effect of outliers on the estimates of genetic correlation. *Indian Journal of Animal Genetics and Breeding*, **27(1, 2)**, 17-21.

Sariaka, Wahi, S.D. and Rao, A.R. (2006b). On the robustness of estimates of genetic correlation. *Ind. J. Animal Sciences*, **76(9)**, 759-763.

Sethi, I. C. and Jain, J.P. (1993a). Progeny testing with multiple ovulation and embryo transfer in cattle. *Ind. J. Anim. Sci.*, **63**, 257-262.

Sethi, I. C. and Jain, J.P. (1993b). Closed and open nucleus breeding schemes in genetic improvement of dairy cattle. *Ind. J. Anim. Sci.*, **63**, 863-868.

Sethi, I. C. and Jain, J.P. (1994). Preliminary selection of sires on partial records in dairy cattle. *Biom J*. **36**, 363-371.

Sharma, B.S. and Narain, P. (1986). A study of gene action in crossbred cattle. *J. Ind. Soc. Agril. Statist.*, **38**, 410-416.

Sharma, B.S. and Narain, P. (1988). On gene interaction in cross-bred dairy cattle *J. Ind. Soc. Agril. Statist.*, **40(1)**, 228-234.

Sharma, B.S. and Pirchner, F. (1991). Heterosis in Friesian x Sahiwal crosses. *J. Anim. Breed Genet.*, **108**, 241-252.

Sharma, B.S., Prabhakaran, V.T. and Pirchner, F. (2000) Gene action and heterosis in lifetime traits of Friesian x Sahiwal . *J .Anim. Breed. Genet.*, **117**, 319-330.

Shukla, R.K. (1993). Some investigations on the effects of genetico-statistical model assumptions on the estimates of genetic parameters. *Unpublished Ph. D. Thesis*, P.G. School, IARI, New Delhi.

Singh, N. Okendro and Wahi, S.D. (2001). Effect of master sample on the estimate of standard error of heritability by parent offspring regression method. *J. Ind. Soc. Agril. Statist.*, **54(3)**, 324-327.

Sreekumar, J. (1994). Studies on genetic gain on deleting and restricting various traits in different selection strategies. *Unpublished M.Sc. Thesis*, Post Graduate School, IARI, New Delhi – 12.

Thennarasu, K. (1995). On certain non-parametric procedures for studying genotype by environment interaction and yield stability. *Unpublished Ph.D. Thesis*, P.G. School, IARI, New Delhi.

Thomas,G., Mohapatra,T., Rao, A.R. and Sharma,R.P. (2006). Distinguishing Indian commercial wheat varieties using RAPD based DNA fingerprints. *Indian Journal of Biotechnology*, **5**, 200-206.

Verma, D. and Jain, J.P. (1989). Some improved estimators of heritability, *Biom. J.* **31**, 759-63.

Wahi S.D. (1994). Estimators of repeatability for perennial crops. *J. Ind. Soc. Agril. Statist.*, **46(2)**, 279-89.

Wahi, S.D. and Bhatia, V.K (1995). Use of bootstrap methods in comparing the performance of linear discriminant function. *Jour. Ind. Soc. Ag. Statistics*, **47(1)**, 12 - 20.

Wahi, S.D., Bhatia, V.K. and Lal Chand (1998). Some investigation on precision of heritability. In: Advances in Agricultural Statistics and Computer Applications in Agriculture, IASRI Publication, pp. 180-188.

Wahi, S.D., Bhatia, V.K. and Rao, A.R. (2006) "Study on statistical properties of genetics correlation using bootstrap technique". *Ind. J. Animal Sciences*, **76(9)**, 755-758.

Wahi, S.D., Dash, S and Rao, A.R. (2009). An empirical investigation on classical clustering methods. *The Icfai University Journal of Genetics & Evolution*, **2**(**3**), (in press).

Wahi,S.D. and Kher, K.K. (1991). A comparison of clustering procedures based on multiple traits in Gerbera and Dahlia. *Indian J. Genet*., **51(30)**, 335-41.

Wahi, S.D., Lal Chand and Bhatia, V.K. (2004). A growth pattern study in crosses and pure Indian breeds of goats. *Ind. J. Animal Science*, **74(9)**, 955-58.

Wahi, S.D. and Rao, A.R. (2001). Estimation of variance of repeatability estimates for perennial crops. *Ind. J. Applied Statistics*., **6**, 57-62.

Wahi, S.D. and Rao, A.R. (2004). Some empirical investigations to study the influence of fixed effects on the estimates of heritability. Unpublished Project Report, IASRI, New Delhi – 12.

Wahi, S.D. and Rao, A.R. (2005). Empirical investigations on the estimation of heritability in presence of non-genetic fixed effects under half-sib mating designs. *Ind. J. Anim. Genet. Breed.*, **26(1,2)**, 31-40.

Walke, R.D. (1993). Some investigations on inadmissible estimates of heritability. *Unpublished M.Sc. Thesis*, P.G. School, IARI, New Delhi.

# Information Systems and Software Development at IASRI: An Overview

**P.K. Malhotra and R.C. Goyal**

## 1. Introduction

Presently the society is being built on technology, knowledge and intelligence. Information technology (IT) empowers both people and machines with information, which is transformed into knowledge and intelligence. Appropriate use of the knowledge by both people and machines contributes to sustainable development. While informed and empowered people know their role as citizens in an environmentally sustainable society, empowered machines have the knowledge to minimize energy and material use, wastes, and pollutants.

The Internet, major source of IT exploration has gained immense popularity over the past decade and in this age of information explosion; no country, organization and individual can afford to be ignorant of this emerging technology. It is the easiest way to link the documents and their sections in a non-linear manner, over the different network paths. A very useful application of the Internet is to design and establish an on-line information system, where any one around the globe with authorized access permissions can do the data updating and retrieval of useful information. It is inevitable that ICAR should collect and publish data with regard to Agricultural Research, Development and Extension activities in the country and develop necessary infrastructure for the same purpose.

Data collected in agricultural research in laboratory and on-farm research are subjected to descriptive, inferential, diagnostic and analytical processing. The statistical packages cover most of the statistical procedures often required for data analysis, outstanding graphics capabilities, data handling tools, and are user friendly. The costs of these packages are very high and some of these require payment of annual license fee.

To meet the requirements of various stake holders in the National Agricultural Research System (NARS), IASRI has developed number of information systems as well as statistical packages for analyzing agricultural research data. An overview of these developments is presented here.

## 2. Information Systems

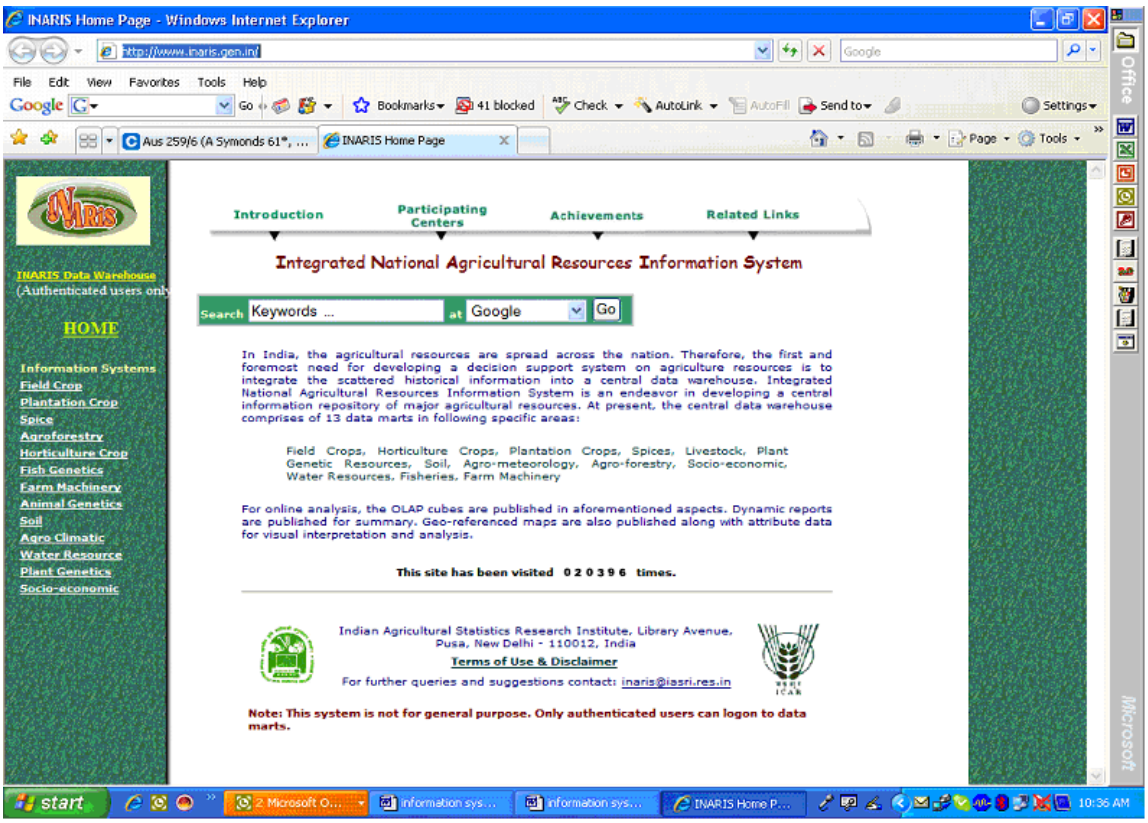
With the installation of B-4700 in 1977, the activity for development of information systems and data base applications was initiated. A group of scientists designed and implemented a data base comprising of the bio-data of agricultural research scientists under the ICAR institutes. A selective dissemination information service based on AGRIS data base available on tapes received from Vienna every month was in operation since 1978. This enabled the scientists in India to obtain references on selected topics of their interest in the field of agricultural science and technology. During 1980-84 necessary software were developed for the information storage and retrieval system for National Index of Field Experiments and National Index of Animal Experiments. During 1990-2000 a number of information systems were developed, to name a few National Information System on Agriculture Education (NISAGE); Project Information and Management System (PIMS); Agriculture Research Data Information System - ARDIS

(*Version 2.0*); File Monitoring System - FMS (*Version 1.0*); Inventory Resource Information System – IRIS; Leave Record Information System and a package for Scientists/Technical Personal Information on Computer Knowledge at IASRI etc. An overview of the information systems developed thereafter at IASRI is presented here.

### 2.1 Integrated national agricultural resources information system (INARIS)

A state of art Central Data Warehouse (CDW) of agricultural resources of the country has been developed and implemented **http://www.agdwh.iasri.res.in (**Rai *et al.*, 2002) at IASRI, New Delhi. This is probably the first attempt of data warehousing of agricultural resources in the world. This provides systematic and periodic information to research scientists, planners, decision makers and developmental agencies in the form of On-line Analytical Processing (OLAP) decision support system. INARIS has been implemented with active collaboration and support from 13 ICAR institutions. In all 59 databases on agricultural technologies generated by various institutes in the council, research projects in operation and related agricultural statistics from published official sources from the year 1990 onwards at districts/ state/national levels, population census including village level population data as well as tehsil level household assets and livestock census are included in CDW. Subject-wise data marts (Rai, *et al.*, 2004) have been designed, multi-dimensional data cubes (Chaturvedi *et al.*, 2004) developed and published in the form of on-line decision support system. The validation checks have been implemented wherever possible.

The information of this data warehouse are available to user in the form of decision support system where in the flexibility of the presentation of the information, its on line analysis including graphic is



inbuilt into the system. The system also provides facility of spatial analysis of the data through web using functionalities of Geographic Information System (GIS). Apart from this, subject wise information system has been developed for the general users. The user of this system has the access of subject wise dynamic reports through web. The facilities of data mining and generation of ad-hoc querying have also been extended to limited users.

### 2.2 Online information system on personnel management in ICAR

Personnel Management Information System Network (PERMIS*net*) for ICAR has been developed and implemented **http://www.iasri.res.in/permisnet** for providing information for systematic planning and development of human resources in ICAR (Balbir *et al*., 2004, 2005, 2008), (Shashi *et al*., 2004), (Alka *et al*., 2005), (Farooqi *et al*., 2005). The system provides information on scientists, technical personnel, administrative and supporting staff on personal and professional attributes of each employee, besides information on cadre strength and institutional structure to help the management in taking appropriate timely decisions. The information is available at the Institute, Divisions of ICAR and ICAR level.

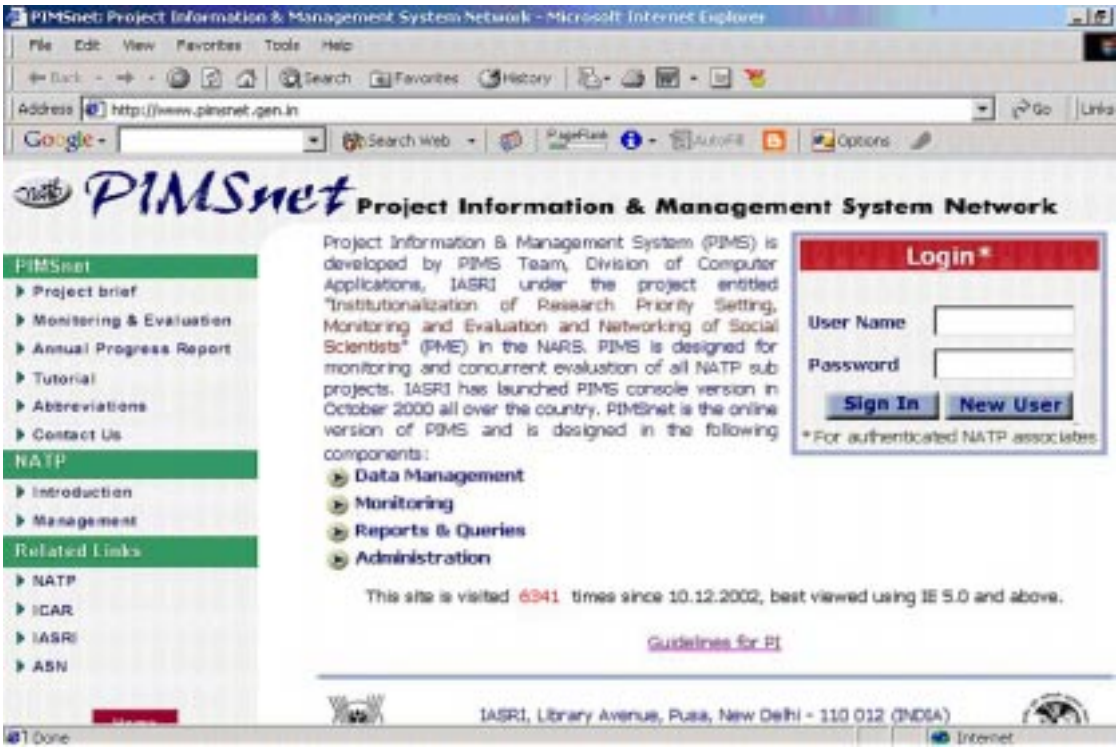
### 2.3 Project information management system network (PIMSNET)

PIMSNET was designed, developed and implemented under the World Bank funded National Agricultural Technology Project's (NATP) sub-project 'Institutionalization of Research Priority Setting, Monitoring and Evaluation and Networking of Social Scientists' under Project Monitoring and Evaluation (PME) Component of Organisational & Management (Goyal *et. al.*, 2002, Sharma *et al*., 2005). PIMSNET was used for Monitoring and Concurrent Evaluation (M & CE) of 845 research projects under NATP so as to effectively assess research efforts against well-defined targets, avoid duplication of research efforts and provide feedback to research planning process It had four modules viz., Data Management, Monitoring, Reports and Query and Administration.

The monitoring module of PIMSNET acted essentially as a monitoring system for NATP (Sharma *et al.*, 2001, 2002, 2003). Several monitoring quantifiable indicators were devised and implemented in PIMSNET. The module stored information on the progress of the projects for the different activities in terms of the expected targets achieved or not, and if not, the reasons thereof and the steps taken to achieve

the targets as reported by the Principal Investigators (PI). It was possible to monitor the half yearly and annual progress in terms of three types of indicators for physical progress, financial progress and scientific/ technical progress. The salient research achievements as well as shortcomings and difficulties faced by the researchers were captured by the system and the management could effectively monitor the progress of all the projects right from their desktop. The system also had an inbuilt algorithm to determine the projects performance status in terms of performing not well/ performing good / performing very good/ performing excellent. Provisions were also made to review the progress report of PI by the Site Committees, State Level Committees and the Peer Review Teams.

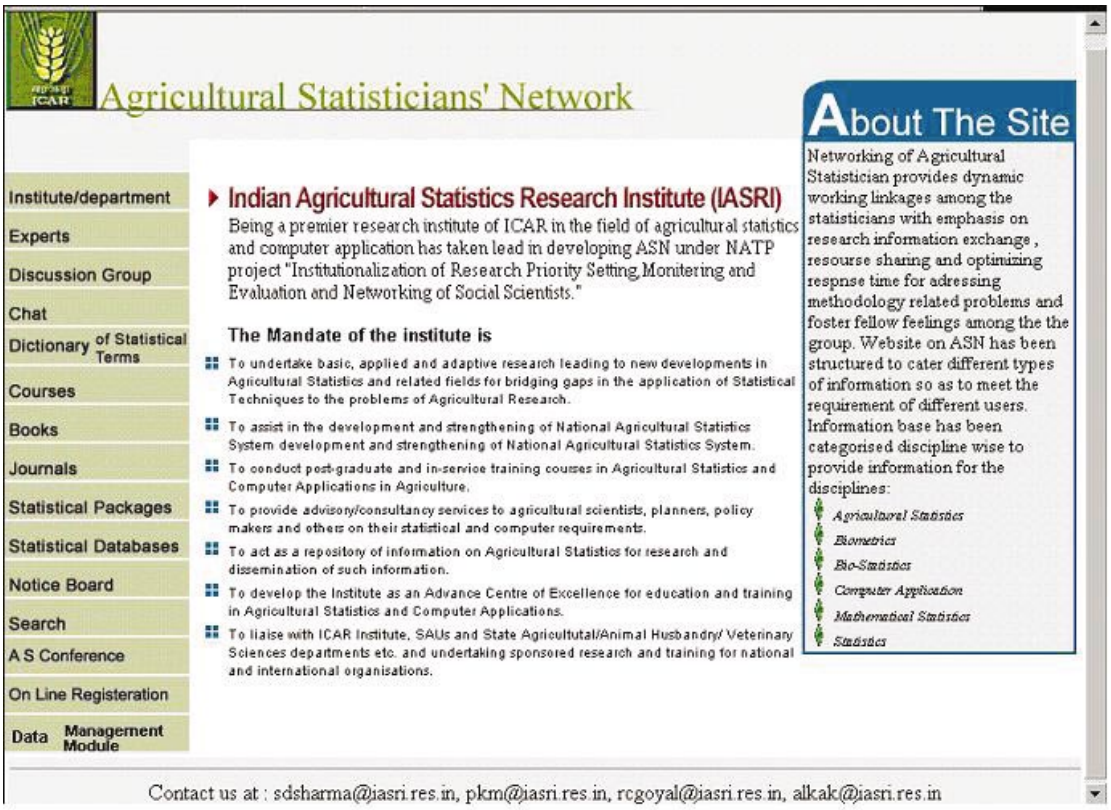
PIMSNET had output modules for the reports and queries through which top and middle level management may get reports and queries as per their specific needs. The system was able to produce General Reports which were ready to use reports at macro level. These reports were generated on the basis of a single criteria, Reports by Selection which are custom reports that meet the specific needs of the management and generated on the basis of options selected by the user based on multiple criteria, Specific Reports that were related to the details of a particular project which were of utmost importance when management needs to find out specific details of a single project and the Administrative Reports which act as a tool for the administration and management of the information storage and user management under the PIMSNET.




## 2.4 Agricultural statistician's network (ASN)

This web based information system has been developed and hosted from the IASRI server at the address **http://iasri.res.in/ASN (**Sharma *et al.*, 2001, 2005), **(**Alka *et al*., 2008). ASN provides dynamic working linkages among the statisticians with emphasis on research information exchange, resource

sharing and optimizing response time for addressing methodology related problems and foster fellow feelings among the group with cost effective communication media. The website provides an opportunity to enhance the image of organizations and the statisticians at the international platform. It lays emphasis on providing a unified base on various professional activities of the agricultural statisticians. Individual's research expertise is exposed to the international community thus providing opportunity for global collaboration. System has features like chat, discussion group, notice board, search, online registration and online data management for effective communication.




## 2.5 National information system on agricultural education network in India (NISAGENET)

NISAGENET has been designed, developed and implemented **http://www.iasri.res.in/Nisagenet** (Goyal *et al*., 2006, 2007) on the recommendations of the National Statistical Commission (NSC-2001) and the Department of Secondary & Higher Education, Ministry of Human Resource Development, Government of India so as to maintain and update the data regularly on parameters related to agricultural education in India. NISAGENET is currently coordinating with 42 agricultural organizations in the country for desired data collection on agricultural education in India. These include State Agricultural Universities, ICAR Deemed Universities, Central Universities in agricultural education and other private colleges/ centers affiliated to these organizations. The database contains the information on various aspects such as Academic data of the universities, Infrastructural facilities, Budget provision, Manpower employed and R&D activities. This on-line system provides an integrated platform for collection, analysis and dissemination of information on programs and activities of the agricultural education system in India through internet service mode.

This system has been designed and developed in two sets of the application software using .NET Framework. The Data Management Application Software is used for data entry/updating and uploading of the data from Local Server of each participating organization to the Central Server at IASRI, New Delhi. The Application Software is used for data integration and country level query/reporting from the Central Server. The Data Management Application Software at University Site has the facilities of Query/ Reports for University/Colleges management and Database Administration at Local Server (Sharma *et al.*, 2006). The Application Software at Central Server has the facilities of Query/Reports system at Country, State, University and College levels, Data Management of other colleges offering agricultural education and is affiliated to other Central Universities, Data Management of all the participating organizations, Database Administration at Central Server.

NISAGENET system generates reports to meet the requirements of Ministry of HRD and other users. The queries and reports made available at Local Server(s) broadly cover general aspects and academic data of the universities, infrastructural facilities in the universities, budget provision, manpower employed, personal information of the faculty and R&D activities pertaining to the respective university and its constituent/affiliated colleges. The queries and reports made available at Central Server broadly cover general aspects and academic data of the universities, infrastructural facilities in the universities, budget provision, manpower employed, personal information of the faculty and research & development activities pertaining to any of the participating university and its constituent/affiliated colleges and aggregated reports of all the universities at the country/state level. The reports have been designed to provide prompt access to information available at various states in the country. The information has also been provided under various categories, such as by location rural, urban, semi-urban; by various subject disciplines; by sex male, female; by longitude and latitude, by agro-climatic zones, and by class category General, OBC, SC, ST, PC, etc.

Other features included in NISAGENET are Directory Services, Search Engine (AgriKhoj) and a Discussion Forum.

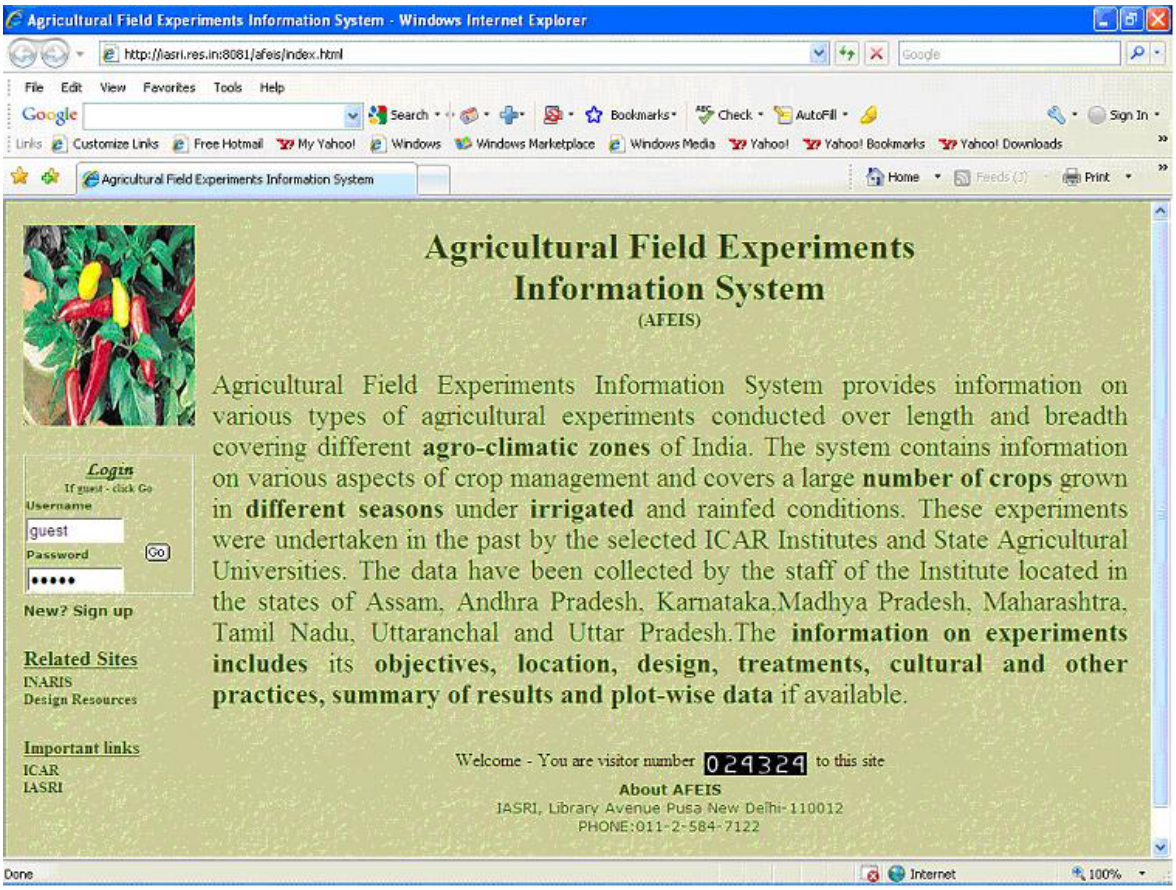
### 2.7 Agricultural field experiments information system (AFEIS)

AFEIS **http://iasri.res.in:8081/afeis/index.html** aims at systematic maintenance of data of field experiments conducted on various aspects of agricultural technology at a central place and retrieval of information on selective basis as per requirements. It was evolved over years as an outcome of National Index of Agricultural Field Experiments Scheme. The *Purely Varietal* trials have been excluded from this database system. Various items included in the database are Objective of the experiment, Details about treatments, Design used, Cultural and other practices followed, General crop conditions, Summary results and/or plot-wise observations.

Agencies engaged in Agricultural Research in India are Agricultural Universities, ICAR's Research Institutes, Project Directorates and All India Coordinated Research Projects, Directorates of Agriculture of State Governments etc. The staff of the Institute especially engaged for this purpose collects data of Agricultural experiments generated at various Agricultural Research Stations of these Organisations.

Though a large number of experiments are being conducted in National Agricultural Research System but this database contains only the information of those experiments that are collected under this Project. Presently the database contains the details of over 25,000 experiments.


NIS AgEnet
National Information System on Agricultural Education Network in India
Home
Objective
Org. Struct.
SAUs
Reports
Download
Nodal Off.
Contact Us
Division of Computer Application
Mandate to develop Databases & Information Systems for Agricultural Research and provide Teaching & Consultancy Services.
» More information
IASRI
To undertake basic, applied and adaptive research leading to new developments in Agricultural Statistics and related fields for bridging ....
» More information
Members Login
Username:
Password:
Login
Directory
For classified information of SAU's on Administartor's, Equipments, Extension Activities....
» Click Here
Search Engine
To Search Agricultural Education Activities in India
» Agri Khoj
Discussion Forum
For detailed information about Agricultural Discussion Forum.
» Click Here
Latest News
Ongoing Activities under Nisagenet.
» Read full article
7-8 August, 2006
NISAGENET Data Management Application Software Successfully Installed at Sher-e-Kashmir University of Agriculture Science & Technology, Jammu , J&K
4-5 August, 2006
NISAGENET Data Management Application Software Successfully Installed at Ch. Sarwan Kumar Krishi Vishwavidyalaya, Palampur, Himachal Pradesh
Nisagenet
National Statistical Commission (NSC) and the Department of Secondary & Higher Education of the Ministry of Human Resource Development, Government of India has desired that a national information system on agricultural education be prepared for various policy and planning purposes in the country. To implement these recommendations, the Indian Council of Agricultural Research (ICAR) has approved this project under its AP Cess Fund Scheme for a period of three years (01-01-2005 to 31-12-2007). With the support and supervision of the Education Division of ICAR, New Delhi, this project is being executed by the Indian Agricultural Statistics Research Institute (IASRI), New Delhi as Coordinating Unit and the Lead Center having collaboration with 42 participating organizations involved in imparting higher agricultural education in India. The major activities and the technical programme of the project are as follows:
• Collection and compilation of data on Agricultural Education
• Designing, Development and Implementation of NISAGENET software on the Internet
• Publications of annual statistical bulletins on Agricultural Education
• Supply of statistical data to Ministry of HRD as per their requirements.
The NISAGENET lays emphasis on providing a unified information base for collection, compilation and analysis of information about the activities of the agricultural education system in India. The information on Academic, Infrastructural facilities, Budget provision, Faculty and Manpower employed, Research and Development activities of all the State Agricultural Universities, five ICAR Deemed to be Universities including four of ICAR and Allahabad Agricultural Institute, three Central Universities and a Central Agricultural University would be provided. Besides these, the information of several other public funded and private aided and unaided colleges, which are offering higher agricultural education, would be made available.
Home | IASRI | Contact us | Div. of Comp. App.
Developed & Maintained by
Division of Computer Applications
Indian Agricultural Statistics Research Institute
New Delhi ©2005
Site has been visited 96519 times since 1st, Jan 2005.
objective org. structure major activities nodal officers download contact us



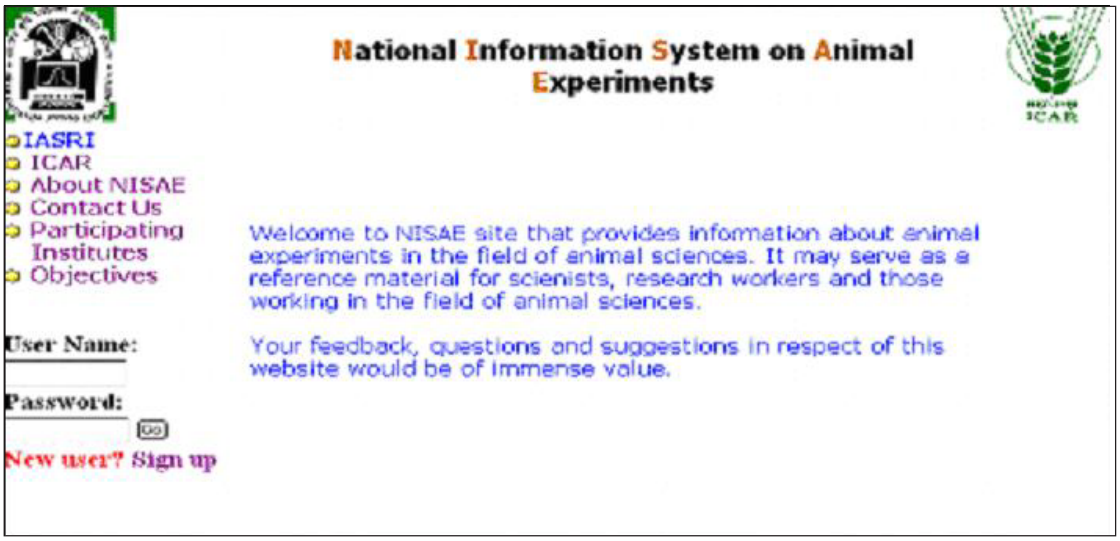

The database is updated annually and can answer user-based queries in respect of one or more of the data items like, state, crop, season, factors tried, design, irrigated/rainfed, soil type etc. Some of the reports that can be readily generated through the system are Location specific (based on state or research station), Treatment specific, Crop specific, Location and Crop specific and Crop and Treatment specific.

### 2.8 National information system on animal experiments (NISAE)

NISAE **http://iasri.res.in:8080/nisae22** (Batra *et al*., 2005) has been developed for analyzing data of the experiments in various disciplines of animal sciences that are conducted at various ICAR institutes dealing with Animal Sciences and Agricultural/Veterinary Universities. Information relating to the results of these experiments is not available in compatible form at one place to the scientific community. Keeping in view, National Information System on Animal Experiments (NISAE) has been developed wherein the information relating to the experiments conducted in the country in various fields of Animal Sciences can be stored at a central place in a compatible form.

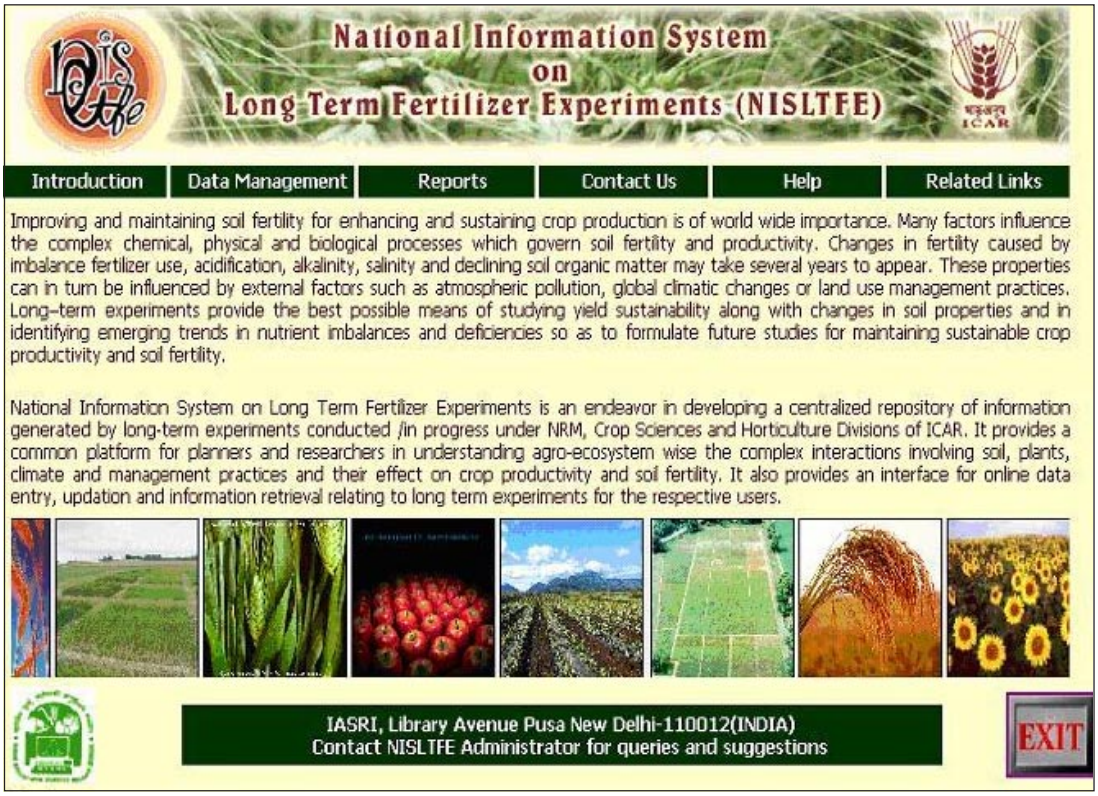
The key source of information has been the annual reports published by these organizations. The information of the experiment has been divided into data items viz. Title, Names of investigators, Subject, Source, Research Station, Species, Breed, Period of experiment, Year of conduct, Funding agency, and Salient results achieved. The information system has been designed keeping in view that the system is user friendly and can be used by individuals having no knowledge of computer based information system

and/or computer software background. The user can generate any query pertaining to a specific research station, subject, species, breed and investigator name to obtain the desired information.




## 2.9 National information system on long term fertilizer experiments (NISLTFE)

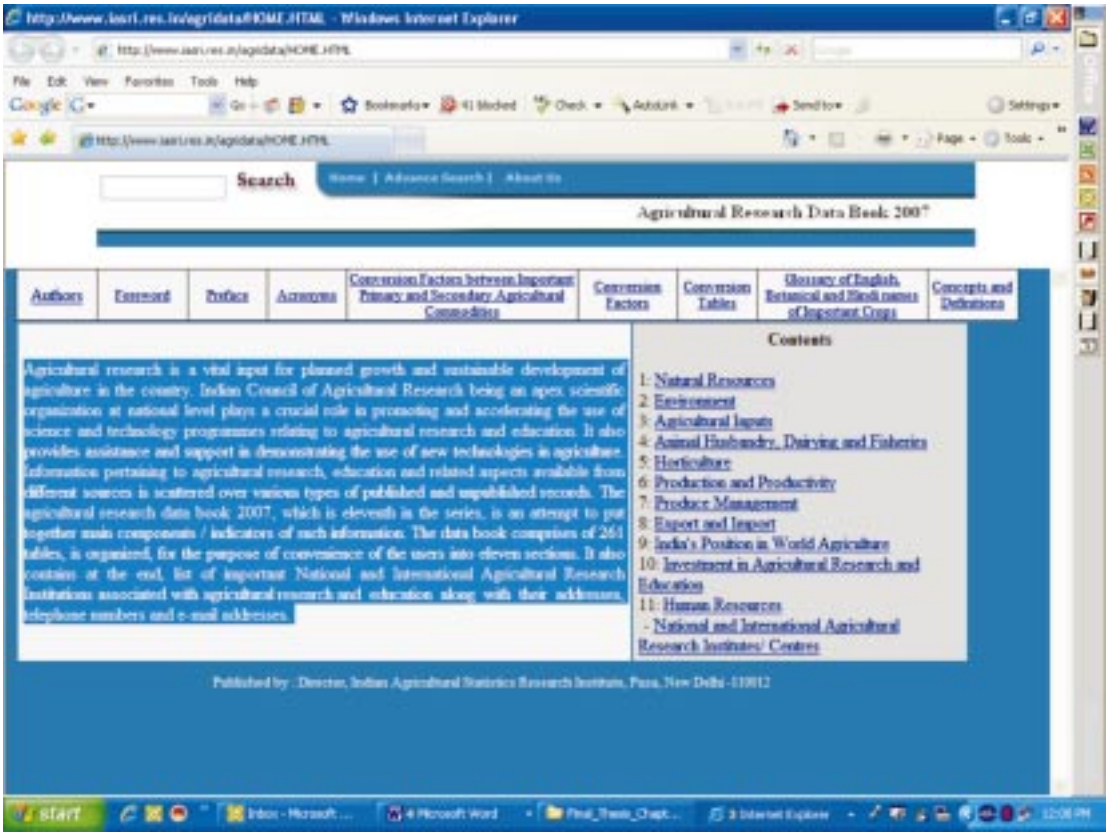
NISLTFE **http://iasri.res.in:8081/nisltfe/index.html** (Vats *et al*., 2005) has been designed and developed to provide information on the long term fertilizer experiments conducted/ in progress at various organizations under Horticulture, Crop Sciences and NRM divisions of ICAR. The information stored



includes (i) Experiment id information containing the names of ICAR division, organization type, centre and site. (ii) Centre information consisting of its ecosystem, soil taxonomy, longitude, latitude and altitude. (iii) Experiment general information like title, objectives, year of start/termination, statistical design, field layout plan, plot size etc. (iv) Crop related information like variety, standard week of sowing/ harvesting, plant spacing, crop condition/damage etc. (v) Treatment details information involving input doses, sources, methods etc. (vi) Data information containing plot wise/ mean data for different characters. (vii) Mid course modification information regarding changes carried out in the ongoing experiment. (viii) Superimposed treatments data for each character. (ix) Weekly weather data information and (x) Information relating to new user like creating password, electronic/ surface mail and fax of new user.

## 2.10 Agriculture research data information system (ARDISNET)

Agricultural research is a vital input for planned growth and sustainable development of agriculture in the country. ICAR 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 2007, (Bathla *et al*., 2007) which is eleventh in the series, is an attempt to put together main components/ indicators of such information. The data book available as web resource at http:// www.iasri.res.in/agridata comprises of 261 tables, is organized, for the purpose of convenience of the users into eleven sections. It also contains at the end, list of important National and International Agricultural



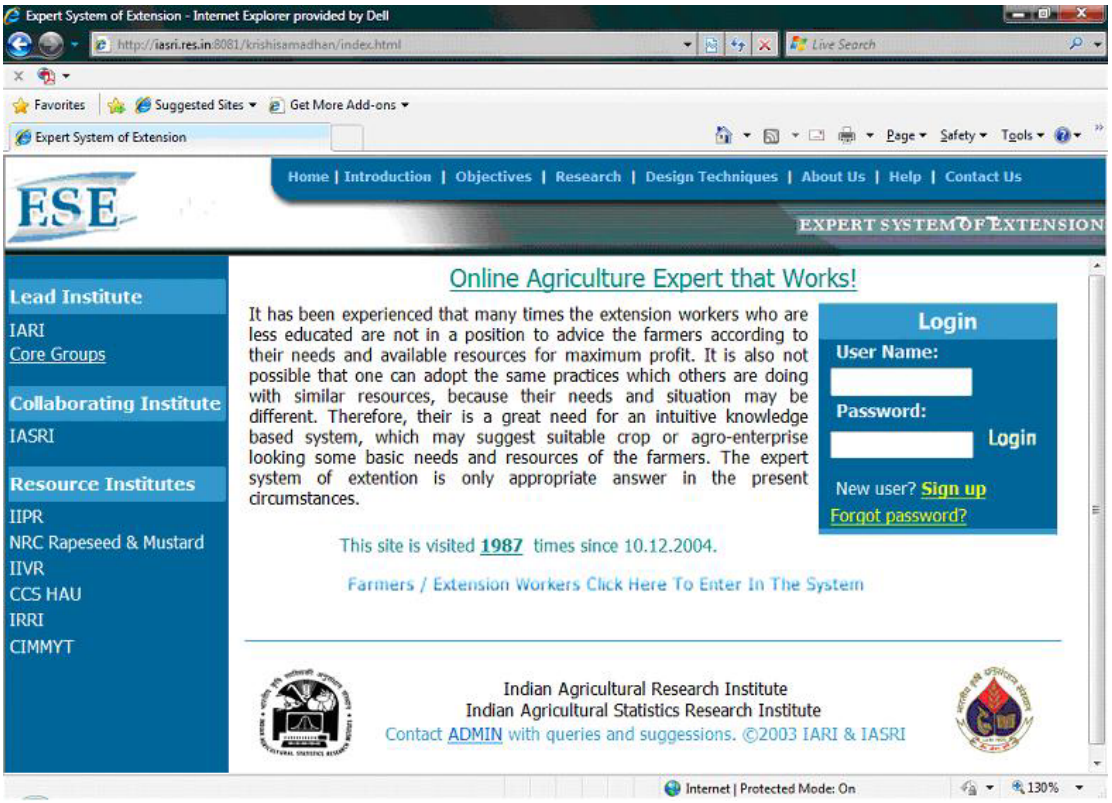
Research Institutions associated with agricultural research and education along with their addresses, telephone numbers and e-mail addresses.

## 3. Expert Systems

An Expert System (ES) is a computer program designed to simulate the problem-solving behavior of an expert in a narrow domain or discipline. In agriculture, expert systems unite the accumulated expertise of individual disciplines, e.g., agronomy, plant pathology, entomology, agricultural meteorology etc., into a framework that best addresses the specific, on-site needs of farmers. Expert systems combine the experimental and experiential knowledge with the intuitive reasoning skills of a multitude of specialists to aid farmers in making the best decisions for their crops. IASRI has taken initiatives and developed expert systems namely 'Expert System of Extension' and 'Expert System on Wheat Crop Management' to help the farmers in taking appropriate decisions for better management of their crop.

### 3.1 Expert system of extension

The Expert System of Extension was designed and hosted through IASRI server at http://www.iasri.res.in:8081/krishisamadhan/index.html.(Sudeep *et al*., 2002). This can also be reached through IASRI home page. It was developed under the NATP-CGP project in collaboration with IARI (Bahal *et al*, 2004 a, b; 2006). The main objective was to help farmers to take appropriate decisions and disseminate need based research findings to millions of the farmers at a time, which is neither possible nor practicable by conventional system of extension. The development of the system was quite relevant as the conventional extension system is lacking human and financial resources to meet the needs of millions of the farmers in diverse agro-climatic conditions. Farmers information needs, effectiveness of the technology, efficiency



of the website, its adoptability and cost has been considered in very fast changing electronic scenario. The system was launched in the Agro-eco-region-4 of the ICAR, which consists parts of Punjab, Haryana, Rajasthan, Gujarat, M.P., Delhi and U.P. and seven crops/technologies were identified.

The system uses the Java Expert System Shell (JESS). The shell allows the interaction with the latest web technologies and thus can be integrated seamlessly with other java based programs. It acts as a helping tool for the farmers and extension workers to take appropriate decision regarding the variety selection, insect and disease identification, nematode and physiological disorder identification and their control.

### 3.2 Expert system on wheat crop management (EXOWHEM)

EXOWHEM **http://iasri.res.in/expert** (Islam *et al.*, 2007) has been developed and implemented by the scientists of IASRI in collaboration with two premier research institutions on wheat namely Directorate of Wheat Research, Karnal and Indian Agricultural Research Institute, New Delhi. The system holds a collection of general principles that are potentially applied to solve a problem related to wheat crop management and is capable to extend expert advice to the researchers and wheat-growing farmers.

The system is designed to cover agricultural operations, variety selection, fertilizer application, and insecticide / pesticide application on one hand and economic benefits on the other hand. The system analyses the question given by the user and gives a piece of advice to the user.

This system has four modules for Variety selection, Plant protection, Cultural practices and Harvesting technology and one module for knowledge management. Variety selection module specifies the variety

from the farmer's point of view. Plant protection module is subdivided into pathological aspects, entomological aspects and weed management. In pathology, the system identifies micro diseases such as leaf rusts, blights and bunts etc. In entomology, the system identifies pest / insects affecting plants and recommends control measures. The cultural practices module specifies the process of cultivation of the crop. The harvesting technology module helps in advising the appropriate method, machinery and time for the harvest.

The developed system extends the research done on the wheat crop to the farmers in effective way to help in getting instant solution to the problems faced by them. It encourages the dissemination of research findings and helps farmer's where services of the experts are not available. It reduces the load from the experts and facilitates qualitative research on wheat supported by the knowledge base in the system.

## 4. Software Development - Statistical Packages

Since 1965 using IBM 1620 computer system the institute has been extending programming assistance for electronic data processing for the analysis of the research problems of students and agricultural scientists from various institutes under ICAR and SAUs (Mathur and Sathe, 1984). To meet the statistical data analysis requirements of the agricultural research data, it was essential to develop general purpose computer software. During eighties a number of computer programs for plant and animal breeding experiments were developed using FORTRAN language (Malhotra, 1990) and were ultimately integrated to give birth to a software package popularly known as Statistical Package for Agricultural Research (SPAR1) (Doshi *et al*., 1991) containing relevant modules like Diallel Cross, Partial Diallel Cross, Line x Tester or Hybrid Analysis, multi-factor analysis such as Path coefficient analysis, Discriminant function analysis, Mahalanobis D-square analysis and also estimation of genotypic, phenotypic and environmental correlations between the dependent and independent factors for analyzing plant breeding experiments.

The other type of experiments conducted by crop scientists is intended to compare several varieties or the effect of fertilizers at different levels of application and also in combination with two or more fertilizers on yield. The trials are conducted by choosing a suitable design. It may be randomized block design, factorial design, split-plot or split-split plot design, and confounded factorial design or replicated lattice design, etc. Appropriate application software for such situations were developed.

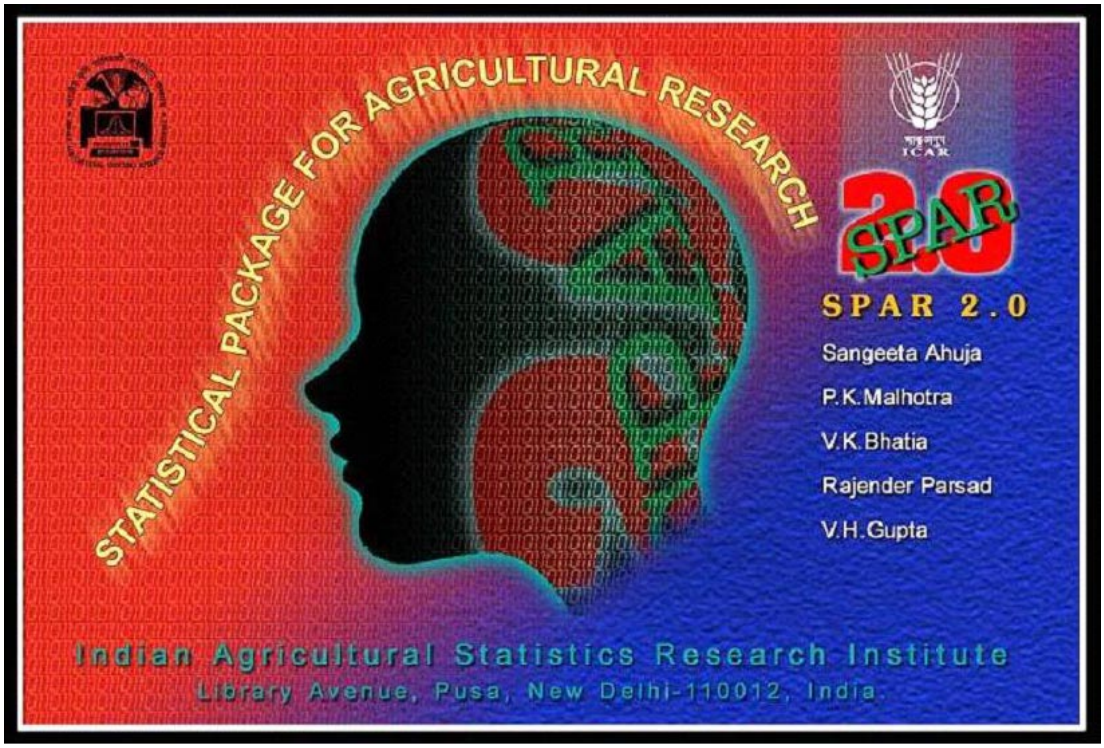
A number of computer programs were also developed to meet the requirements in the field of animal science (Dutta *et al*., 1985) such as general program for estimating heritability by (i) Intra sire regression method (ii) Half sib method (iii) Full sib method (iv) Estimation of repeatability (v) Least square analysis for unequal sub class numbers using Harvey's techniques.

With the evolution of computer science and different user friendly languages, continuous efforts are being made to meet the present day requirements, some in-house software developments of IASRI are presented here.

### 4.1 Statistical package for agricultural research data analysis (SPAR 2.0)

SPAR 2.0 (Ahuja *et. al*, 2005) is a Windows Version of SPAR1 with some additional modules developed for statistical analysis of experimental research data in Plant Breeding and Genetics. It is User-Friendly, Interactive, Password Protected, Menu-Driven Package and can also be operated using the Toolbars. A Context-Sensitive Help with Index, Contents and Search facility are available. This package consists of

all the necessary details to use the software and also the basic definitions including mathematical formulas related to the methods used in the package. Moreover, it includes sample inputs and outputs of all the modules. This is a helpful tool for quick understanding of this statistical software. This package is also useful for teaching the subject of genetical statistics to the post-graduate students and helpful for the researchers in statistics with special interest in plant and animal sciences.




The package consists of eight modules. Data Management module consists of two sub modules viz., (i) Editing of data and (ii) Transformation of data; Descriptive Statistics module consisting of six sub modules viz., (i) Measures of central tendency, (ii) Measures of dispersion (iii) Generation of moments (iv) Measures & coefficients of skewness (v) Measures and coefficients of kurtosis and (vi) Measures of partition values; Estimation of Breeding Values module can predict breeding value by fitting of Generations means and performing Scaling and Joint Scaling Tests; Regression and Path Analysis module, Variance and Covariance Components Estimation module consists of computation of ANOVA, components of variances bivariate analysis of variance and covariance components. In Stability Analysis module, analysis can be performed using Eberhart and Russell, Perkins and Jinks, Freeman and Perkins, Additive Main effects and Multiplicative Interaction, Site Regression (SREG) and Factorial Regression(Iti, et al., 2008). Multivariate Analysis Module has three sub modules, Cluster Analysis, Discriminant Analysis and Principal Component Analysis. Mating Design Analysis module can perform analysis for Complete Diallel, Partial Diallel, Line x Tester, Three way cross, Double cross and North Carolina Designs.

### 4.2 Statistical package for animal breeding (SPAB2)

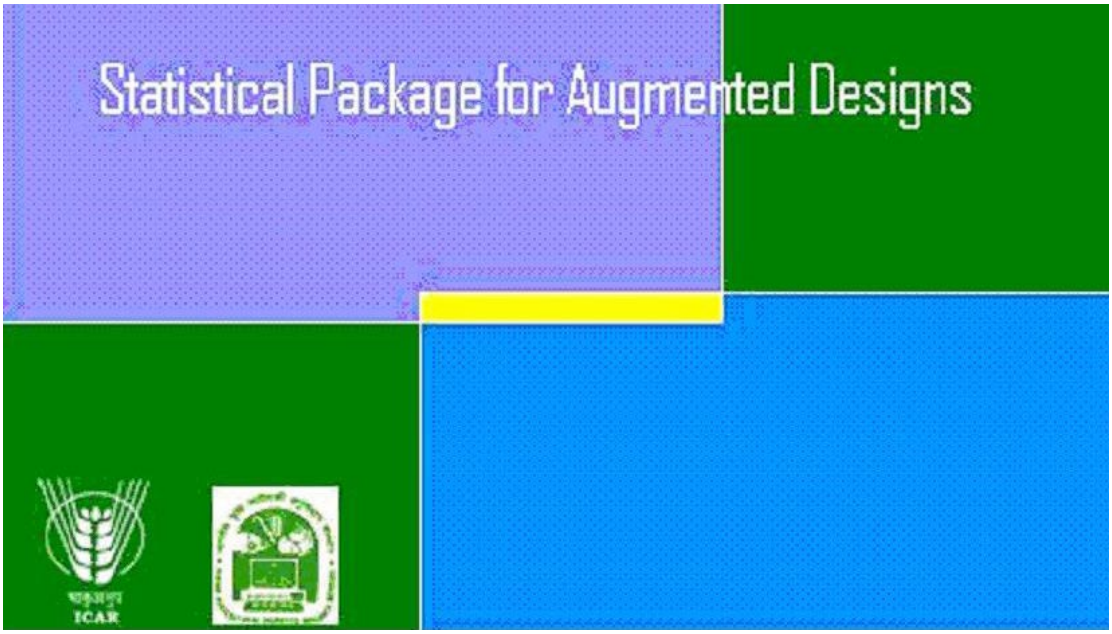
Statistical package for animal breeding (SPAB2) (Sethi, 2006) has been developed keeping in view, the computing requirements of scientists/students, mainly working in Animal Breeding and Animal Genetics

research. The package is Window based, Menu driven and works in a User friendly manner. In the present version of the package, 37 useful programs of maximum utility are included. These programs have been grouped into ten modules. It has provisions for Analysis of Mixed Model Data as provided in LSML package developed by Walter R. Harvey, Best linear unbiased prediction (BLUP) for single trait as well as for multiple traits, Adjustment for different non-genetic effects, Sire evaluation using SRLS and Sire evaluation using REML. Computation of Mean and SE for different classifications, Genetic parameters for Half sib data, Genetic parameters for Full sib data, Coefficient of repeatability and Producing ability. It provides computation of Selection index (Hazel's method), Restricted selection index, Sire indices for different models, Osborne's index, Cunningham's selection index. Diallel analysis can be performed for data with unequal classifications, for different modeling situations. One can calculate Inbreeding coefficient, Genetic gain and Genetic trend. Multivariate Analysis has programs for D square analysis, Multiple regression analysis (Step-up/Step-down methods, all possible combinations) and Principal component analysis. Most of the Non-Parametric tests are also provided in the package.

### 4.3 Statistical package for augmented designs (SPAD)

SPAD (Abhishek *et al.*, 2004) is useful for designing agricultural experiments conducted for comparing existing practices / check varieties, called controls, with new practices / varieties / germplasm collections, called tests, where the experimental material for the tests is limited and it is not possible to replicate them in the design. The package generates a randomized layout of an augmented randomized complete block (RCB) design and augmented complete block design with equal or unequal block sizes.

The optimal replication number of the control treatments in every block is obtained by maximizing the efficiency per observation for making tests *vs* controls comparisons. User has a flexibility to choose the replication number of the control(s) in each of the blocks. The package also performs the analysis of data generated from augmented block designs (complete or incomplete). The treatment sum of squares is partitioned into different components of interest *viz.* (i) among test treatments, (ii) among control treatments and (iii) among test treatments and control treatments. Multiple comparison procedures for making all



possible pair-wise treatment comparisons can also be employed through this package. A null hypothesis on any other contrast of interest can also be tested.

### 4.4 Statistical package for factorial experiments (SPFE 1.0)

SPFE 1.0 (Ahuja *et al.*, 2004) is a user friendly, interactive, password protected Software developed for generation of the designs for symmetrical and asymmetrical factorial experiments with or without confounding. It also generates the randomized layout of the designs for factorial experiments. The design for factorial experiments with confounding is generated on listing the independent interactions to be confounded. Different sets of interactions and different number of independent interactions can be assigned for confounding in different replications. It also generates regular fractional factorial plans for symmetrical factorial experiments. The data generated are analyzed as per procedure of unblocked/block 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 other contrast of interest can also be tested.


This package is also useful for teaching the subject of factorial experiments to the post- graduate students in the class. This can also be useful for the researchers in statistics with interest in experimental designs particularly in factorial experiments.

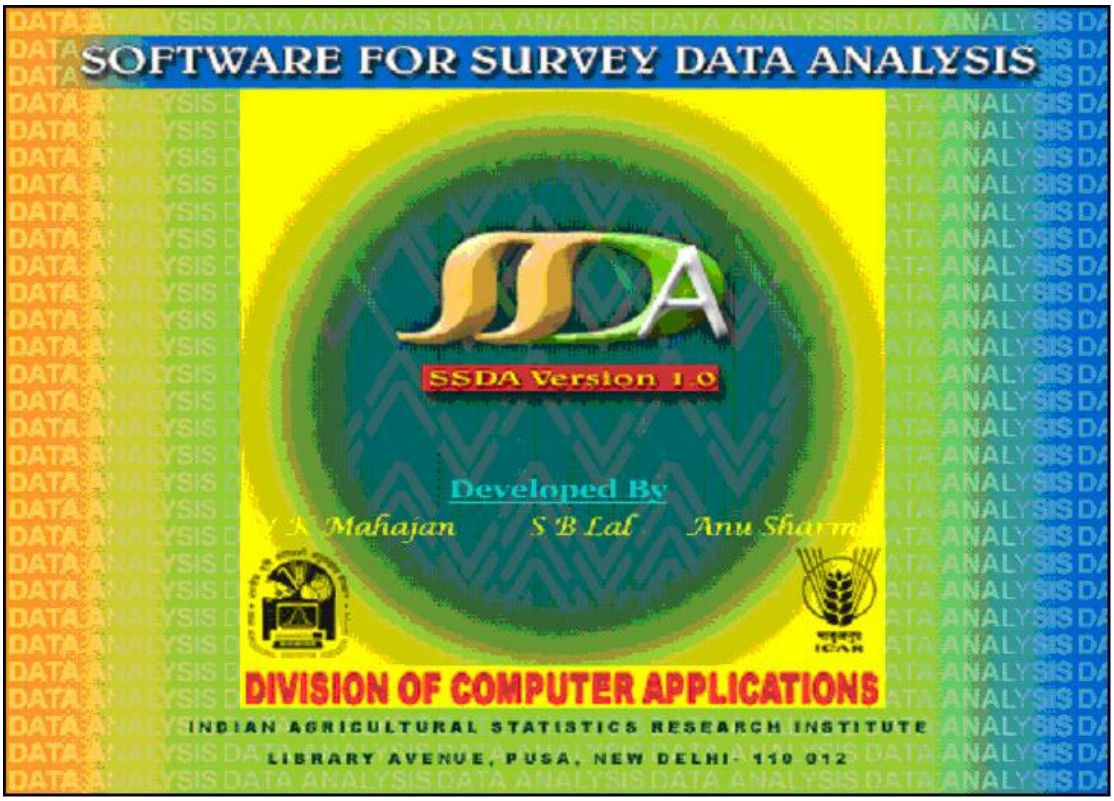
### 4.5 Software package for balanced incomplete block designs (SPBD 1.0)

SPBD (Parsad *et al*., 1997) enables a user to select and generate a randomized layout of Balanced Incomplete Block (BIB) Design. The package generates BIB designs with replication numbers up to a maximum of 20 for asymmetric BIB designs and 30 for symmetric BIB designs. The package also provides the analysis of variance with both treatments adjusted and blocks adjusted sum of squares, adjusted treatment means, variance of the estimated treatment contrasts and the contrast sum of squares, etc. The definitions of the terminology used are available on-line. The package is useful for the experimenters,

classroom teaching as well as for the researchers in statistics with special interest in design of experiments or computing genetic parameters for one-way and two-way classified data.

### 4.6 Software for survey data analysis (SSDA)

SSDA (Mahajan *et. al*., 2008) has been designed and developed to be used by survey statisticians for analysis of survey data and by academicians and teachers for the demonstration of survey data analysis methods. This software estimates the population parameters i.e. mean based on the sampling data collected using the important common sampling designs like simple random, straitified, systematic, cluster, two stage and stratified two stage. These schemes have been considered under the options of with/without replacement under equal probability of selection and with replacement under unequal probabilities. SSDA also has a module for impution of missing values using mean, zero and mean of neighbouring units methods. The results of the schemes have been presented in the form of Crystal Reports which can be imported into various common file formats.




## 5. E-advisory and E-Iearning Resources

### 5.1 Design resources server

Design Resources Server http://**www.isari.res.in/design** (Parsad and Gupta, 2009) has been developed to popularize and disseminate research in design of experiments among experimenters in agricultural sciences, biological sciences, animal sciences, social sciences and industry in planning and designing their experiments for making precise and valid inferences on the problems of their interest, generally treatment contrasts. It also provides support for analysis of data generated so as to meet the objectives of

the study. This server also aims at providing a platform to the researchers in design of experiments for disseminating research and also strengthening research in newer emerging areas so as to meet the challenges of agricultural research. This server attempts to spread advances in theoretical, computational and statistical aspects of design of experiments among the mathematicians and statisticians in academia and among the practicing statisticians involved in advisory and consultancy services.

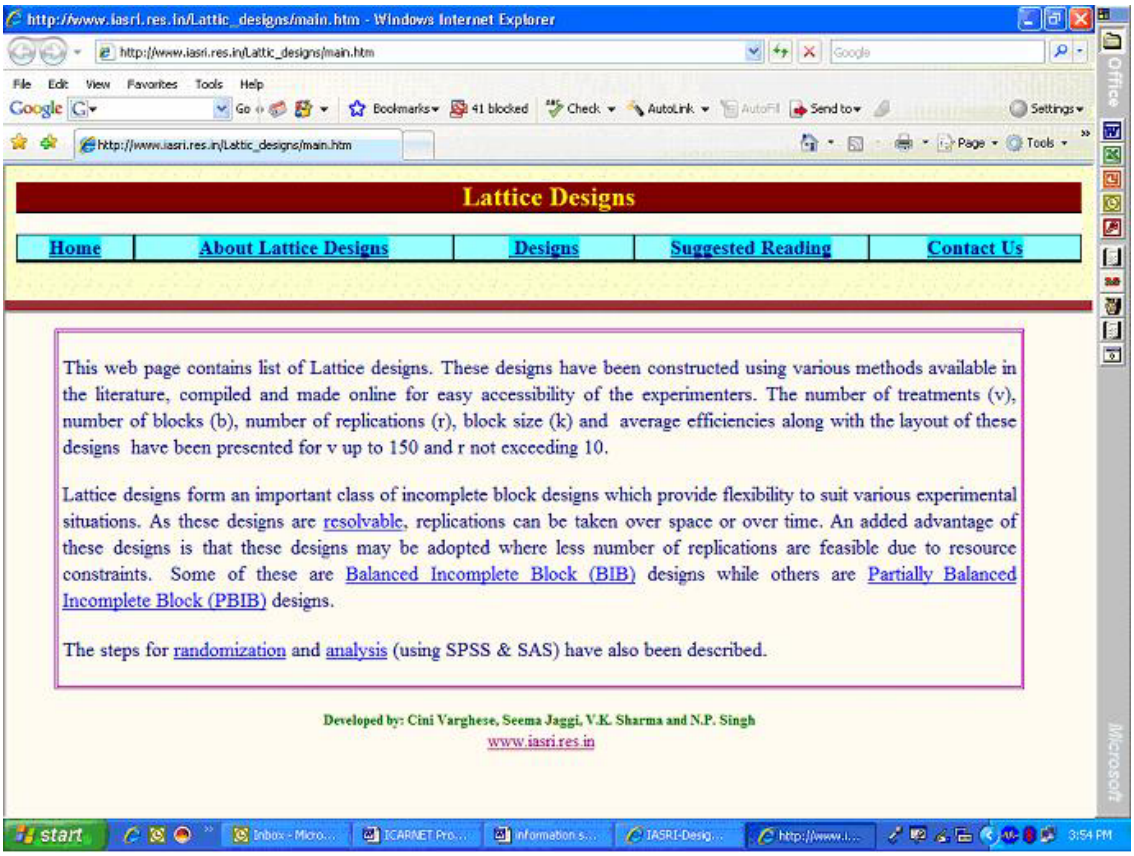
The material available on the server has been partitioned into four main components:

- **Useful for Experimenters:** Electronic Books, online generation of randomized layout of designs, online analysis of data, analysis of data using various softwares
- **Useful for Research Statisticians:** Literature and catalogues of BBB designs, designs for making test treatments-control treatment comparisons, supersaturated designs, online generation of Hadamard matrices, MOLS and orthogonal arrays
- **Other Useful Links:** Discussion Board, Ask a Question, Who-is-where
- **Site Information:** Feedback, How to Quote Design Resources Server, Copyright, disclaimer, contact us and site map


The Design Resources Server is like a mobile library on design of experiments in particular and statistics in general. It provides useful information both for active researchers in statistics as well as stake holders like scientists in NARS and others all over the globe. It is expected that the material provided at this server would help the experimenters in general and agricultural scientists in particular in improving the quality of research in their respective sciences and making their research globally competitive.

### 5.2 Web resources on lattice (http://www.iasri.res.in/lattic_designs/main.htm) and circular designs (http://iasri.res.in/circular_designs/)

This web page contains list of Lattice designs. These designs have been constructed using various methods available in the literature, compiled and made online for easy accessibility of the experimenters. The number of treatments ($v$), number of blocks ($b$), number of replications ($r$), block size ($k$) and average efficiencies along with the layout of these designs have been presented for $v$ up to 150 and $r$ not exceeding 10.

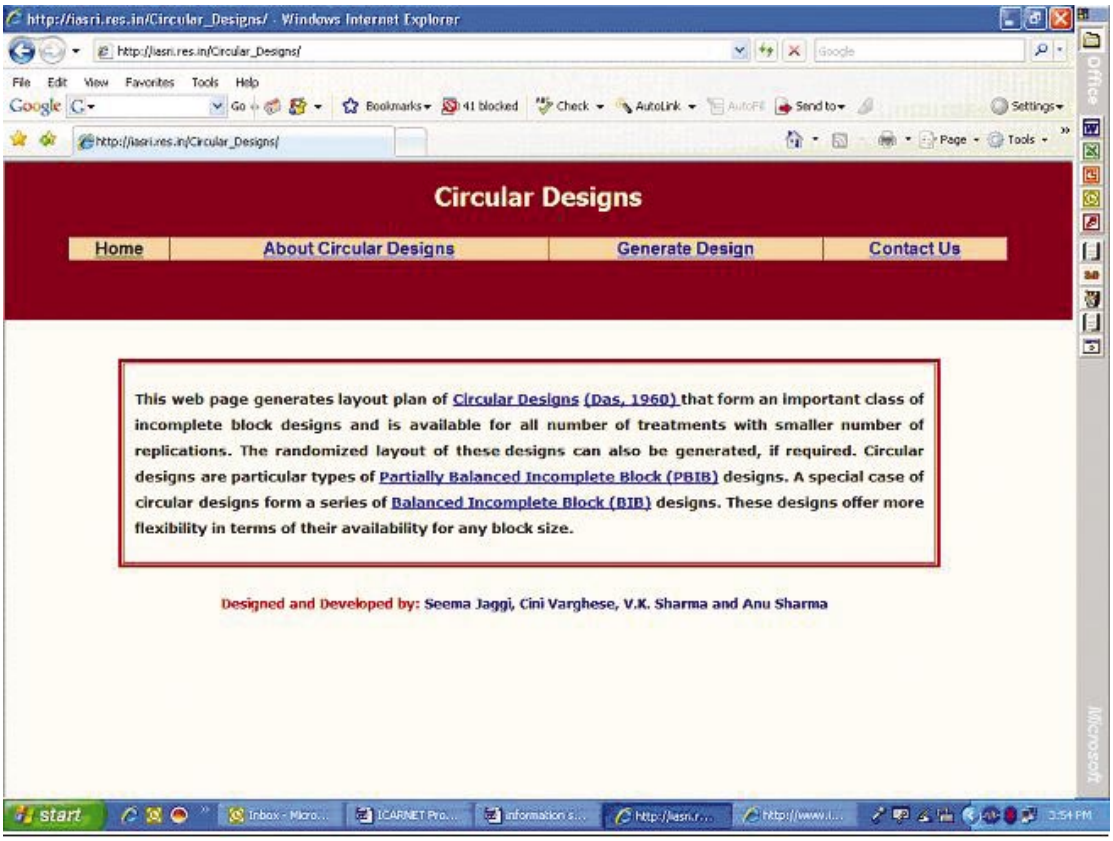



Lattice designs form an important class of incomplete block designs which provide flexibility to suit various experimental situations. As these designs are resolvable, replications can be taken over space or over time. An added advantage of these designs is that these designs may be adopted where less number of replications are feasible due to resource constraints. Some of these are Balanced Incomplete Block (BIB) designs while others are Partially Balanced Incomplete Block (PBIB) designs.

The steps for randomization and analysis (using SPSS & SAS) have also been described.

This web page generates layout plan of Circular Designs that form an important class of incomplete block designs and is available for all number of treatments with smaller number of replications.

The randomized layout of these designs can also be generated, if required. Circular designs are particular types of Partially Balanced Incomplete Block (PBIB) designs. A special case of circular designs form a series of Balanced Incomplete Block (BIB) designs. These designs offer more flexibility in terms of their availability for any block size.



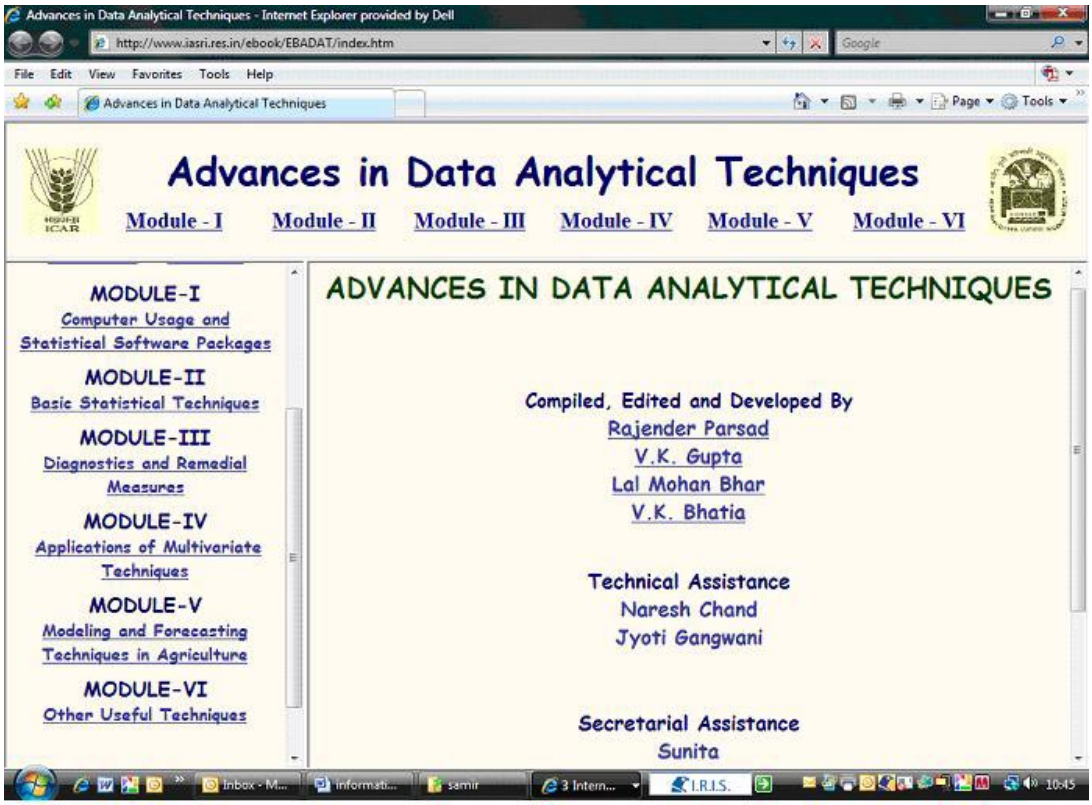

## 5.3 e-Books (http://www.iasri.res.in/ebook/ebooks.htm)

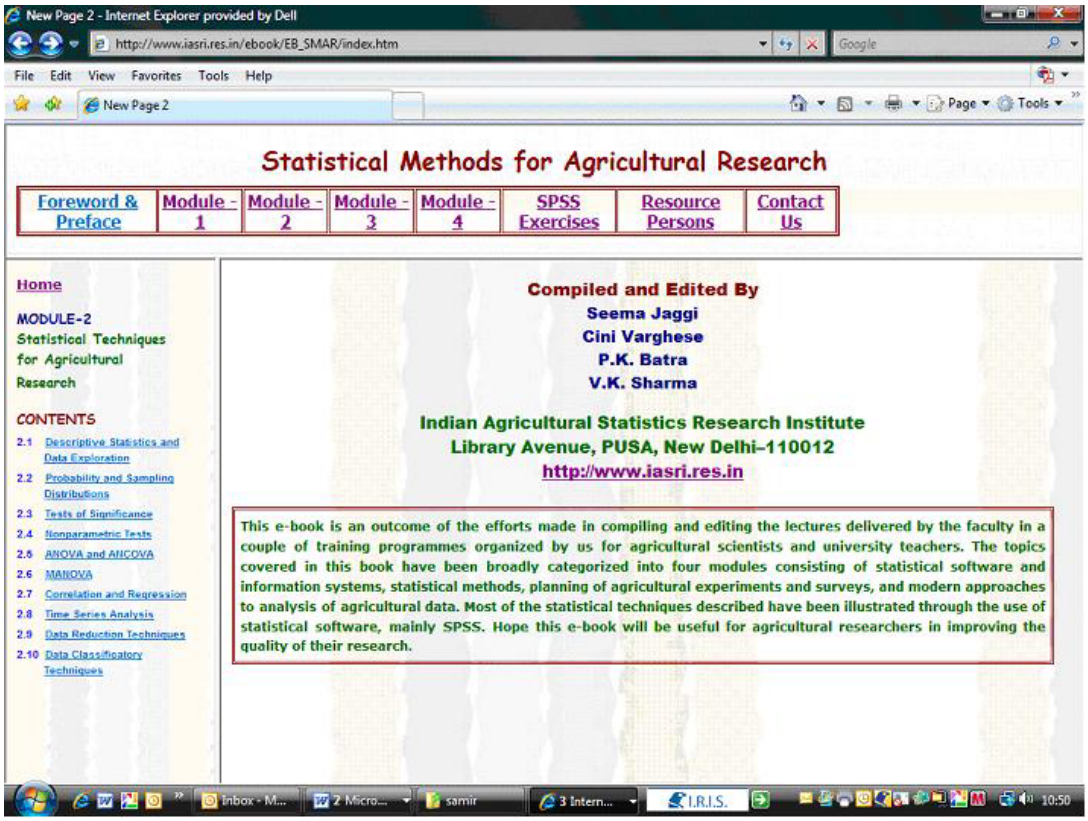
### 5.3.1 Advances in data analytical techniques

A training programme on "Advances in Data Analytical Techniques" was organized by the Institute. As a consequence of this training program, an electronic book has been developed for the benefit of not only the scientists in NARS, but for all the scientists throughout the world engaged in generation of data for their research and also subjecting the data to serious statistical analysis for drawing meaningful conclusions. The various modules included in the electronic book are (i) Computer Usage and Statistical Software Packages (ii) Basic Statistical Techniques (iii) Diagnostics and Remedial Measures (iv) Applications of Multivariate Techniques (v) Modelling and Forecasting Techniques in Agriculture and (vi) Other Useful Techniques such as Bio-informatics, Geoinformatics, Microarrays, etc.

### 5.3.2 Statistical methods for agricultural research

This e-book is an outcome of the efforts made in compiling and editing the lectures delivered by the faculty in a couple of training programmes organized by authors of this book for agricultural scientists and university teachers. The topics covered in this book have been broadly categorized into four modules consisting of statistical software and information systems, statistical methods, planning of agricultural experiments and surveys, and modern approaches to analysis of agricultural data. Most of the statistical techniques described have been illustrated through the use of statistical software, mainly SPSS.


Advances in Data Analytical Techniques - Internet Explorer provided by Dell
http://www.iasri.res.in/ebook/EBADAT/index.htm
File Edit View Favorites Tools Help
Advances in Data Analytical Techniques
Advances in Data Analytical Techniques
Module - I
Module - II
Module - III
Module - IV
Module - V
Module - VI
MODULE-I
Computer Usage and Statistical Software Packages
MODULE-II
Basic Statistical Techniques
MODULE-III
Diagnostics and Remedial Measures
MODULE-IV
Applications of Multivariate Techniques
MODULE-V
Modeling and Forecasting Techniques in Agriculture
MODULE-VI
Other Useful Techniques
ADVANCES IN DATA ANALYTICAL TECHNIQUES
Compiled, Edited and Developed By
Rajender Parsad
V.K. Gupta
Lal Mohan Bhar
V.K. Bhatia
Technical Assistance
Naresh Chand
Jyoti Gangwani
Secretarial Assistance
Sunita


New Page 2 - Internet Explorer provided by Dell
http://www.iasri.res.in/ebook/EB_SMAR/index.htm
File Edit View Favorites Tools Help
New Page 2
Statistical Methods for Agricultural Research
Foreword & Preface
Module - 1
Module - 2
Module - 3
Module - 4
SPSS Exercises
Resource Persons
Contact Us
Home
MODULE-2
Statistical Techniques for Agricultural Research
CONTENTS
2.1 Descriptive Statistics and Data Exploration
2.2 Probability and Sampling Distributions
2.3 Tests of Significance
2.4 Nonparametric Tests
2.5 ANOVA and ANCOVA
2.6 MANOVA
2.7 Correlation and Regression
2.8 Time Series Analysis
2.9 Data Reduction Techniques
2.10 Data Classificatory Techniques
Compiled and Edited By
Seema Jaggi
Cini Varghese
P.K. Batra
V.K. Sharma
Indian Agricultural Statistics Research Institute
Library Avenue, PUSA, New Delhi-110012
http://www.iasri.res.in
This e-book is an outcome of the efforts made in compiling and editing the lectures delivered by the faculty in a couple of training programmes organized by us for agricultural scientists and university teachers. The topics covered in this book have been broadly categorized into four modules consisting of statistical software and information systems, statistical methods, planning of agricultural experiments and surveys, and modern approaches to analysis of agricultural data. Most of the statistical techniques described have been illustrated through the use of statistical software, mainly SPSS. Hope this e-book will be useful for agricultural researchers in improving the quality of their research.

## Concluding Remarks

In the above, work related to the computer applications in agricultural research at IASRI, New Delhi has been reviewed. This work is continuing for the development of knowledge based systems, expert systems in more crops, decision support systems, information systems and statistical software. These efforts will help in strengthening the agricultural research, management and development under the NARS as well as providing the much needed information to the farmers.

## References

Ahuja, Sangeeta, Malhotra, P.K., Bhatia, V.K., Parsad, Rajender and Gupta, V.H. (2005). *Development of Statistical Package for Agricultural Research (Windows Version) - SPAR 2.0*, Project Report, IASRI, 1-40.

Ahuja, Sangeeta, Parsad, Rajender, Gupta, V.K. and Malhotra, P.K. (2004). Statistical package for factorial experiments (SPFE) 1.0. *J. Indian Soc. Ag. Statist*., **58(3)**,312-330.

Arora, Alka, Sharma, S.D., Malhotra, P.K. and Goyal, R.C. (2008). Agricultural statistician network (ASN), *J. Indian Soc. Ag. Statist.*, **62(1)**, 49-55.

Arora, Alka, Singh, Balbir, Dahiya, Shashi and Farooqi, Mohd. Samir (2005). Planning and distribution of manpower of ICAR using PERMISnet, *J. Indian Soc. Ag. Statist*., **59(2)**, 141-145.

Bahal, R., Wason, M., Sudeep, and Issar, D. (2006). Expert system for demand driven extension. *J. US-China Education Review*, **3(5)**, 58-65.

Bahal, R., Wason, M., Sudeep and Kumar, R. (2004 a). Expert system for effective extension. In *Communication Support for Sustaining Agricultural Extension Services* (Eds. Dipak De, Basavaprabhu Jirli and kalian Ghadei) Ganga Kaveri Publishing House, Varanasi, 91-102

Bahal, R., Wason, M., Sudeep and Issar, D. (2004 b). Expert system for effective extension. *Indian J. Extension Edu.*, **40(1 & 2)**, 8-11.

Bathla, H.V.L., Tyagi, K.K., Khatri, R.S., Singh, Jagbir, Agarwal. S.C., Sood, R.M. (2007). *Agricultural Research Data Book*, Joint Publication of IASRI and ICAR.

Batra, P.K., Goyal, R.C., Sehgal, D.K. and Sharma, V.K. (2005). *National Information System on Animal Experiments*. AP Cess Project Report, IASRI, New Delhi.

Chaturvedi, K.K., Rai, Anil, Dubey, V. and Malhotra, P.K. (2004). On-line analytical processing using multidimensional cubes. Paper presented in the conference of the 'Society of Statistics Computer and Application' held at Sri Venkateswara College, New Delhi during 22-24 December 2004.

Dahiya, Shashi, Singh, Balbir, Arora, Alka and Farooqi, Mohd. Samir (2004). Web based personnel management information system for Indian Council of Agricultural Research (ICAR). *J. Indian Soc. Ag. Statist*., **58(2)**, 244-252.

Doshi, S. P. and Gupta, K.C. (1991). *Statistical Package for Agricultural Research*, IASRI, New Delhi.

Dutta, O.P., Malhotra, P.K. and Mathur, S.N.(1985). *Development of Computer Program for Estimating Heritability of Threshold Characters by Different Methods*. Tech. Project Report, IASRI, New Delhi.

Farooqi, Mohd. Samir, Singh, Balbir, Dahiya, Shashi, Arora, Alka (2005). PERMISnet: e-solution for ICAR. *Agricultural Extension Review*, May-June 2005, 3-5.

Goyal, R.C., Malhotra, P.K., Arora, Alka, Lal, S.B., Gupta, V.H. and Dubey, Vipin (2007). *National Information System on Agricultural Education Network in India.* Final Report, AP-Cess Funded Project, ICAR, 1-173.

Goyal, R.C., Malhotra, P.K., Arora, Alka, Lal, S.B., Gupta, V.H. and Dubey, Vipin (2006). *Requirement Analysis Document of National Information System on Agricultural Education Network in India*. Technical Reference Manual. AP-Cess Funded Project, ICAR.

Goyal, R.C., Malhotra, P.K., Marwaha, Sudeep and Arora, Alka. (2002). *Project Information Management System (PIMS)*. Reference Manual, IASRI Publication.

Gupta, P.L., Malhotra, P.K. and Goyal, R.C. (2002). *Expert Management Information System.* Reference Manual, IASRI Publication.

Islam , Shahnawazul, Sikarwar , Harrnam Singh, Farooqi , Mohd. Samir, Dubey , Vipin K., Chaturvedi, K.K., Agarwal, Hari Om, Singh, Randhir, Sharma, A.K., Sharma, R.K., Srivastava, K.D., Sharma, Kirti, and Sharma, J.P. (2007). *Expert System on Wheat Crop Management.* Project Report. IASRI Publication.

Jha, Iti, Deb, D. Malhotra, P.K., Bhatia, V.K. and Parsad, Rajender (2008). AMMI, SREG and FREG models for stability analysis in SPAR 2.0. *J. Indian Soc. Ag. Statist.*, **62(3)**, 283-288.

Mahajan, V.K., Lal, S.B. and Sharma, Anu (2008). *Software for Survey Data Analysis*. Project Report. IASRI Publication.

Malhotra, P.K. (1990). *Programming With FORTRAN-77*. IASRI Publication.

Mathur, S.N., Sathe, K.V. (1984). Division of Computer Science and Numerical Analysis – 25 years of Data Processing. *Souvenir Silver Jubilee* 1959-1984, IASRI, New Delhi, 249-253.

Parsad, Rajender and Gupta, V.K. (2009). *Bulletin on Design Resources Server*. IASRI (ICAR), New Delhi.

Parsad, Rajender, Gupta, V.K. and Khanduri, O.P. (1997). *Statistical Package for Balanced Incomplete Block Designs: SPBD Release 1.0*. IASRI, New Delhi.

Rai, Anil, Madhavanand, V.N.B.S. and Malhotra, P.K. (2004). Design and development of data mart for animal resources in India. Paper presented at the Annual Conference of the Indian Society of Agricultural Statistics held at Govind Ballabh Pant University of Agriculture and Technology, Pantnagar (Uttaranchal) from February 05-07, 2004.

Rai, Anil, Sharma, S.D. and Malhotra, P.K. (2002). Development of data warehouse of agricultural resources in India- Status, Prospects and Issues in the Future. Paper presented in the 2nd International Agronomy Congress held at New Delhi from 26-30 November, 2002.

Rathore, Abhishek, Parsad, Rajender and Gupta, V.K. (2004). Computer aided construction and analysis of augmented designs. *J. Indian Soc. Ag. Statist*., **57**, 320-344.

Sethi, I. C. (2006). *Statistical Package for Animal Breeding (SPAB2).* Project Report, IASRI, New Delhi.

Sharma, Anu, Goyal, R.C., Gupta, V.H. and Grover, Rajni (2006). Design and on-line management of database on Indian aAgricultural education system. *J. Indian Soc. Ag. Statist.*, **60(3)**, 155-161.

Sharma S. D., Malhotra, P. K., Goyal, R. C., Arora, Alka, Sudeep, (2005). *Agricultural Statisticians' Network (ASN).* Reference Manual. IASRI, New Delhi.

Sharma, S.D., Malhotra, P.K., Goyal, R.C., Kumar, Ashok, Sudeep, Gupta, V.H., Arora, Alka and Ahuja, Sangeeta (2005). *Institutionalization of Research Priority Setting, Monitoring and Evaluation, and Networking of Social Scientists.* Final Report, National Agricultural Technology Project.

Sharma, S.D., Kumar, Ashok, Goyal, R.C. and Malhotra, P.K. (2003). *Study on existing monitoring mechanism at different institutions under National Agricultural Research System.* Project Report under NATP.

Sharma, S.D., Sudeep, Goyal, C., Malhotra, P.K. and Kumar, Ashok (2002). Online monitoring and concurrent evaluation system for national agricultural technology project. Paper presented at the Annual Conference of the Indian Society of Agricultural Statistics held at University of Agricultural Sciences, Krishi Nagar, Dharwad – 580 005 (Karnataka) from 18th to 20th December, 2002.

Sharma, S.D., Malhotra, P.K., Goyal, R.C. and Arora, Alka (2001). Networking of agricultural statisticians. Proceedings of the XIII National Conference of Agricultural Research Statisticians of ICAR and Agricultural Universities, held at Punjab Agricultural University, Ludhiana, Punjab from 6-8 November, 2001, 148-150.

Sharma, S.D., Malhotra, P.K., Goyal, R.C., Ashok Kumar, Sudeep, and Gupta, V.H. (2001). *Guidelines for Monitoring and Concurrent Evaluation of NATP Sub-projects*. NATP Publication.

Singh, Balbir, Arora, Alka, Dahiya, Shashi, Farooqi, Mohd. Samir. *Development of Software for Online Information System on Personnel Management in ICAR.* Project Report, 1-55.

Singh, Balbir, Arora, Alka, Dahiya, Shashi, Farooqi, Mohd. Samir (2008). *Development of PERMISnet-II*. Project Report, IASRI, New Delhi, 1-113.

Singh, Balbir, Arora, Alka, Dahiya, Shashi, Farooqi, Mohd. Samir (2004). Personnel management information system network (PERMISnet) for ICAR at www.iasri.res.in/Permisnet. In *ARIS News*, January-June 2004, 1-2.

Sudeep, Kumar, V. and Bahal, R. (2002). Web enabled expert system of extension: A need of time in agriculture for developing countries. 2nd International Agronomy Congress, Nov. 26-30, 2002, New Delhi, India.

# Teaching and Training Activities at IASRI

**Seema Jaggi, V.K. Bhatia and P.K. Malhotra**

The progress of statistical application to agricultural research has been vary rapid during the past fifty years, and the role played by the Indian Agricultural Statistics Research Institute (IASRI) has been very important, both in research as well as teaching and training in the field of Agricultural Statistics and Computer Application. The Institute earned world-wide recognition for its contribution to agricultural research and for evolving techniques for development of agricultural statistics. In addition to in-service training programmes at national and inter-national levels, the Institute has been contributing towards Human Resource Development Programme by way of conducting M.Sc. and Ph.D. programme in Agricultural Statistics since 1964-65 and M.Sc. in Computer Application since 1985-86 in collaboration with Indian Agricultural Research Institute (IARI), New Delhi. Many foreign nationals have also benefited from these programmes.

In this article, the contribution of IASRI in the area of teaching and training in Agricultural Statistics and Computer Application is reviewed.

## 1. Degree Courses at IASRI

An important landmark for the Institute was the signing of a Memorandum of Understanding with IARI in 1964. Courses leading to M.Sc. and Ph.D. degrees in Agricultural Statistics started in collaboration with Post Graduate (PG) School, IARI which has got a status of Deemed University. A course leading to M.Sc. degree in Computer Application in Agriculture was initiated from 1985-86 and was subsequently changed to M.Sc. (Computer Application) from the session 1993-94. The Institute was identified as Centre of Advanced Studies in Agricultural Statistics and Computer Application in 1983 by UNDP and subsequently by Education Division of ICAR in 1995.

### Course curriculum

The total number of courses in Agricultural Statistics at present is 39 consisting of 104 credits (76 theory and 28 practical). In the discipline of Computer Application, there are total 24 courses with 71 credits (49 theory and 22 practical). The details of all these courses can be seen in our institute's website http://www.iasri.res.in/iasriwebsite/degree_course.htm.

Trimester system is followed with three trimesters, each of twelve weeks duration, in an academic year. Besides the major field, every M.Sc. student has to take one minor field of study (two for Ph.D. students) from other disciplines with atleast 9 credits of course work in each. The students have to deliver seminars and complete a dissertation as a part of the degree requirement.

### Faculty strength

At present there is a dedicated team of 31 faculty members of PG School IARI in Agricultural Statistics with 27 members holding Ph.D. degree. 20 faculty members are eligible to guide M.Sc./ Ph.D. students. In the discipline of Computer Application, there are 18 faculty members with 6 having Ph.D. degree and 10 are eligible for guiding M.Sc. students. Some of the faculty members have received Best Teacher Award of IARI.

## Degrees awarded

The total number of students who have been awarded Ph.D. and M.Sc. degree in the discipline of Agricultural Statistics till 2009 convocation is 173 and 287 respectively including 10 foreign students (3 Ph.D. and 7 M.Sc.) from Indonesia, Burma, Sudan, Sri Lanka and Vietnam. The distribution of these students in different block years is shown in Fig. 1 and Fig. 2.

The total number of students who have been awarded M.Sc. degree in the discipline of Computer Application till 2009 convocation is 81 including 1 foreign student. The distribution of these students in different block years is shown in Fig. 3.

The students have to complete a dissertation as a part of the degree requirement. The distribution of dissertations in major research areas of Agricultural Statistics and Computer Application is given in Fig. 4 and Fig. 5 respectively.

At present, the number of students pursuing Ph.D. degree in Agricultural Statistics is 11, M.Sc. in Agricultural Statistics is 11 and M.Sc. in Computer Application is 18.

To commemorate the memory of the first Prime Minister of India Late Pt. Jawaharlal Lal Nehru and to encourage healthy competition among the students to excel in performance in their respective courses, the institute introduced two Nehru Memorial Medals, one for the best M.Sc. Agricultural Statistics student and one for the M.Sc. Computer Application every year. One V.V.R. Murthy Award is given to the best student of M.Sc. Agricultural Statistics.

The Board of Studies in the discipline of Agricultural Statistics and Computer Application monitors the progress of all the degree courses organized by the Institute. The members of the Board of Studies are drawn from different levels

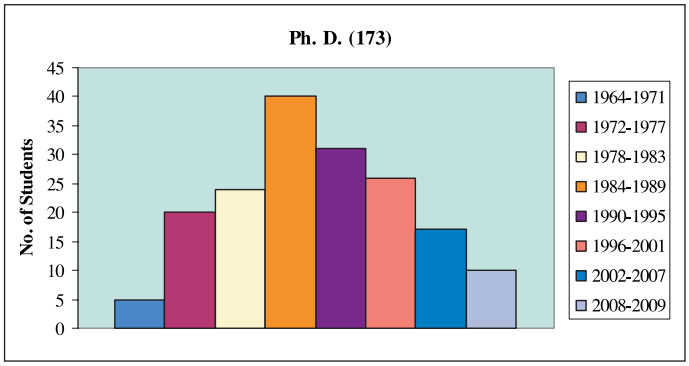



**Fig. 1:** Distribution of number of Ph.D. (Agricultural Statistics) students over different years

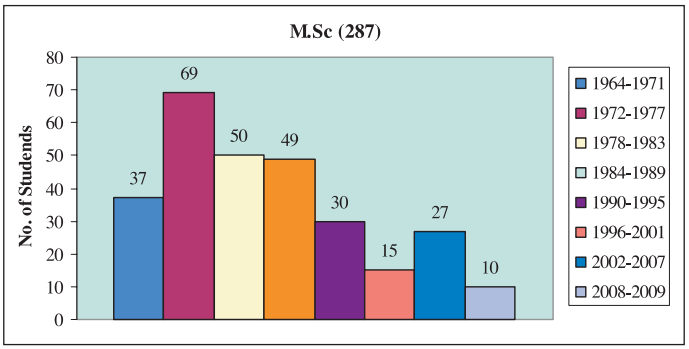



**Fig. 2:** Distribution of number of M.Sc. (Agricultural Statistics) students over different years

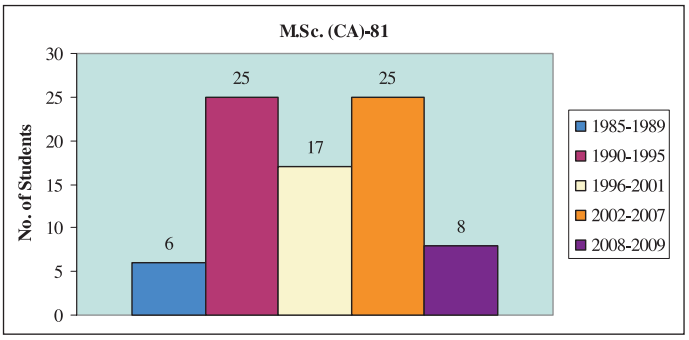



**Fig. 3:** Distribution of number of M.Sc. (Computer Application) students over different years

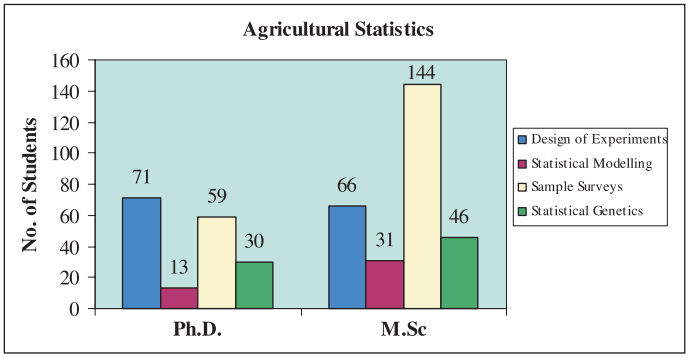



**Fig. 4:** Distribution of students' dissertations in major research areas of Agricultural Statistics

of scientists and include one member from students. All other matters regarding curriculum development, revision/ modification of courses or any other aspects concerning improvement in teaching are discussed in the Board of Studies.

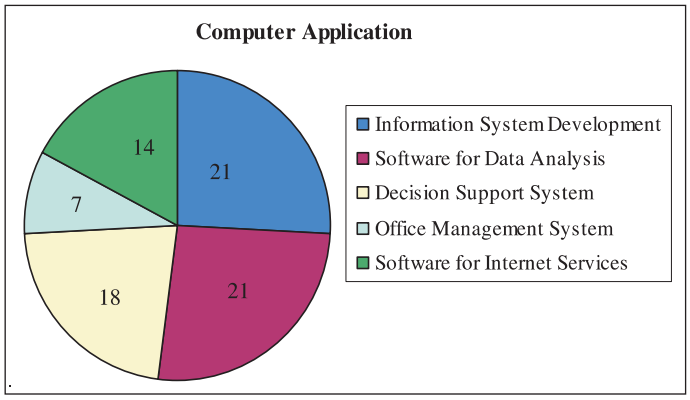



**Fig. 5:** Distribution of students' dissertations in major research areas of Computer Application

The Institute has achieved international recognition for its high quality research and teaching work in the field of Agricultural Statistics. A number of students of the Institute are at present occupying high positions in the universities and other academic and research institutions of the country as well as in other countries. Also, a number of research workers from the Institute have served as consultants and advisors in Asian, African and Latin American countries.

## 2. Other Courses at IASRI

The training activities of the Institute had its beginning in 1945 when two regular Post Graduate courses viz. the Certificate Course and Diploma Course were introduced for Professional Statisticians. Another in-service training of about 8 to 9 months duration called an application course was also regularized. These courses became extremely popular and with the increased demand for admission to these courses the Training Unit of the Institute was established in 1948. During the year 1951, a Short Duration Course of six months duration called Junior Certificate Course was started for the benefit of in-service Agricultural Research Workers from various Institutes/Universities under the Council. To meet the requirements of teaching load most of the appointed statisticians of the institute shared the responsibility in addition to the appointed staff in the Training unit. These courses meant for training professional statisticians and others, who dealt with statistical applications in fields of Agriculture and Animal Husbandry were re-organised into the following four courses:

(i) Junior Certificate Course of six months duration

(ii) Senior Certificate Course of one year duration

(iii) Professional Statisticians' Certificate Course of one year duration in Agricultural Statistics which was recognized as equivalent to Master's Degree course in Agricultural Statistics for employment purpose and

(iv) Diploma Course in Agricultural Statistics and Animal Husbandry of one year duration which was considered to be equivalent to one year research experience.

The Diploma students were required to present their research work in the form of a thesis which was evaluated and on approval of the thesis, the diploma was awarded.

The Institute had been conducting regularly adhoc Training Courses in Data Processing and Computer Programming for the benefit of Agricultural Research Workers from Institutes under ICAR and Central Agricultural Universities, since 1965 when an IBM 1620 Model 2 Electronic Computer was installed. After installation of a 3rd generation computer system Burroughs-4700 in 1977, the Institute felt that there

is a need of trained man power in computer programming. Since October 1981, a two years Diploma Course in Advanced Computer Programming was introduced. On the recommendations made by UNDP Key Consultant under the Project of Centre of Advanced Studies in Agricultural Statistics and Computer Application, the Diploma Course in Advanced Computer Programming was discontinued and a regular Master's Degree course in Computer Application in Agriculture was introduced. This was also in collaboration with PG School, IARI.

The total number of candidates who successfully completed the above given courses is detailed below:

| **Course** | **Year of start** | **Total no. of participants** |
|---|---|---|
| Junior Certificate Course | 1946 | 261 |
| Senior Certificate Course | 1955 | 630 |
| Professional Statisticians' Certificate Course | 1960 | 320 |
| Diploma in Agricultural Statistics and Animal Husbandry | 1948 | 175 |
| Diploma in Advanced Computer Programming | 1981 | 10 |

The above courses had to be discontinued on the advice of ICAR. The Junior Certificate Course was discontinued from the year 1981. Senior Certificate Course, Professional Statisticians' Certificate Course and the Diploma Courses were discontinued from 1985.

From October, 1983 to March, 1992 the Institute also functioned as the Centre of Advanced Studies in Agricultural Statistics and Computer Applications under the aegis of the United Nations Development Programme (UNDP). This programme was aimed at developing a Centre of Excellence with adequate infrastructure and facilities to undertake advanced training programmes and to carry out research on various aspects of Agricultural Statistics and Computer Application.

In 1985, the following new short term courses, along with the number passed out, were introduced that continued till 1996:

- Refresher Course in Statistics for Agricultural Scientists – 57
- Certificate Course in Statistics and Computing – 13
- Short Term Course on Use of Computers in Agricultural Research – 676
- NARP Trainings – 105
- Workshop-cum-Seminar on Software Development for Extension Personnel – 77

On the recommendation of QRT, the discontinued Senior Certificate Course was revived in 1997 in modified form having two modules with emphasis on Statistical Computing. The participants of this course have the flexibility to do either any one module or both the modules. Institute is getting good response for this course from State Government Departments, ICAR Institutes, State Agricultural Universities (SAUs), SAARC Countries, Commonwealth Secretariat and other Afro-Asian Countries. Since revival, 73 participants have completed both modules, 19 have completed module I (Statistical Methods and Official Agricultural Statistics, Use of Computers in Agricultural Research) and 11 have completed module II (Sampling Techniques, Econometrics and Forecasting Techniques, Design of

Experiments and Statistical Genetics). In Module II, a participant has the option to select any three of the said areas.

## 3. Training Programmes at IASRI

The institute is involved in the conduct of a number of training programmes for Statisticians/ Agricultural Research Workers. These programmes fall under the following categories:

### Programmes under revolving fund scheme

Information Technology (IT) has become a key component for success in this age of information explosion. There is great demand for the trained manpower in IT in view of the rapid developments taking place in IT in general and computer technology in particular. Keeping this in view, under Revolving Fund Scheme, Short Term Training Programs in IT were organised in the emerging technologies. High quality and professional standards at par with the international level were maintained. 908 scientists of National Agricultural Research System (NARS) were trained in the latest developments in IT through organisation of 101 training programmes under this scheme.

### Programmes under resource generation

The courses are designed to upgrade the professional skills of agricultural scientists, technical staff and administrative personnel in IT. This provides an opportunity to the participants to have an exposure to computer technologies and exchange of ideas in use of computers in agricultural research. The training also helps in increasing the productivity of participants. The group discussions provide fora to discuss and identify the gaps in the existing software for solving specific research problems.

### International programmes

The International training programmes are offered to share the technical expertise and experience of IASRI with the working professionals from the developing / under developing countries by offering a variety of courses in Agricultural Statistics with hands on experience of statistical data analysis on computers using statistical software packages.

### Training programmes under Centre of Advanced Studies

Centre of Advanced Studies (CAS) on Agricultural Statistics and Computer Applications was established at the Institute during VIII Five-Year Plan as a part of Human Resource Development Programme. The basic objective of the Centre is to develop trained man power in the field of Agricultural Statistics and Computer Application. These training programmes cover specialized topics of current interest in statistics, Computer Application and Agricultural Sciences. The courses under CAS are restricted to the Scientists of Indian Council of Agricultural Research Institutes and SAUs in India.

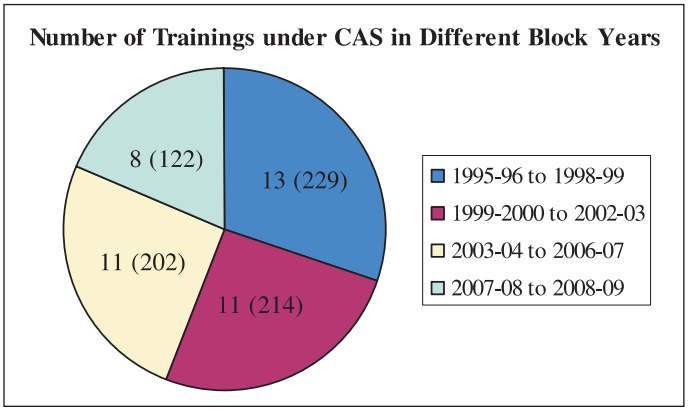



**Fig. 6:** Distribution of number of CAS trainings over different years along with the number of participants (in parenthesis)

The total number of trainings conducted under CAS in Agricultural Statistics and Computer Application are shown in Fig. 6. 767 scientists/ teachers/ researchers from different disciplines of Agriculture and from ICAR institutes and SAUs have successfully completed this training.

**Customized training programmes**

On specific request, the training courses are also organized suiting the customized needs of various organizations and institutions depending on the available expertise and the training facilities. The timing of such courses is decided keeping in view the availability of infrastructural facilities at the Institute. The course fee and other charges are finalized keeping in view the specific requirements of the programme. A series of training programmes have been organized for scientific personnel of Indian Council for Forestry Research and Education (ICFRE). Training have also been conducted for Indian Statistical Services probationers and senior officers of Central Statistical Organization. International training programmes on specific topics are arranged on request from the international agencies like FAO, AARRO, UNDP etc.

Students or participants undergoing training at Indian Statistical Institute, Calcutta and Central Statistical Organization, Govt. of India and a few other organizations visit the Institute for a couple of days as a part of their training programme every year. A suitable training programme is also organized by the Institute for them to acquaint with important research programs of the different divisions.

Seminars are one of the regular activities of the Institute. Every week seminars on relevant topics in Agricultural Statistics or Computer Application are delivered by staff and students of the Institute as well as occasionally by eminent statisticians or computer experts from India or abroad. This provides a good opportunity for students and research workers to exchange ideas on topics and problems of current research.

The Institute has several computer laboratories well equipped with latest hardware and software packages along with modern teaching aids. The e-library has rich collection of books and journals on Statistics, Computer Science and other related disciplines including on-line journals and bibliographic databases. The Institute has following two hostels:

- Panse Hostel: It caters to the requirements of training programmes as well as visiting scientists/ staff in the NARS.
- Sukhatme Hostel: It caters to the requirements of PG Programme. It has well furnished rooms, dining facility, common room, gymnasium and outdoor sports facility.

The hostels are managed under the supervision of a Warden who resides in the IASRI Campus. An International Students' Hostel will also be added very shortly in the list of hostels.

# Modernized Library Information System of IASRI

**P. Visakhi**

The Library of IASRI is a special library and it is well known for its specialized strong resources base in both print as well as electronic format in the field of Agricultural Statistics, Computer Applications, Agricultural Economics and allied sciences. It is the nerve center of educational, research, trainings, consultancy etc. of the Institute. Recently it has been recognized as one of the Regional Libraries of NARS (National Agricultural Research System) and best IT Agricultural Library under National Agricultural Technology Project (NATP).

Library of Institute was established in 1968 in small compact room with limited number of books in the Sample Survey Block. It kept on building up its resource base while serving its users to its best. Later in the year 1977, it was shifted to second floor of newly constructed Computer Centre Building. It was designed with openly wide Stack Rooms, Committee Room and Staff Rooms. The Period 2002-2005 was the golden period in the history of IASRI library. During this period, the library took ocean of change as it was renovated with separate computer lab. for staff with 10 Internet nodes, one user's computer lab for faculty and students, issue and return counter with Internet facility, separate air conditioned reading hall, new arrivals section etc. It was reorganized, into the following separate sections:

**Map of Library** has been signaged at the entrance of the library to direct the reader for various Sections.

**Entrance View of Library**

**Issue and Return Counter (Circulation Counter):** It is the face of the library and it was modernized with granite platform and glass fitted cabin as per library architectural specification. This counter is equipped with a latest computer, internet facility and barcode scanner to scan the borrowers' card and bar-coded documents.

**Reference Counter**: It is a help desk to assist the users in getting information through long and short range. Reference Assistant at the desk guides the users to locate the information from different print resources as well as electronic databases. It is equipped with a computer and internet connection.

**OPAC (On-line Public Access Catalogue):** Just adjacent to the entry gate of library there is a separate wing allocated for on-line catalogue. As soon as a reader enters the library, he/she can search and retrieve the details of documents of his/her interest available in library through this OPAC. In addition to the

above facility, user can also locate the position of the document on different shelves where all racks & shelves are signaged in bilingual.

**Reading Halls**: Apart from separate sitting arrangements on each corner of the wing, library is having three general reading halls. One is the books and back volumes stack area where readers are intended to read the English books, Hindi books, reference books, thesis, official reference publications, old reports (prior to 1995), reprints and back volumes of journals, etc. Another reading hall is located at the adjacent to the periodical wing where Indian, Foreign Periodicals and IASRI publications are displayed. The third reading hall is at the reports section where newspapers and magazines are displayed. The second and third central reading halls are air-conditioned and are covered with vanishing blinds to avoid glare of the sun light to the users. These reading halls are signaged for the convenience of the users.

**Reading Room of Periodical Section**

**Periodical Sections**: Both Indian and Foreign periodicals of currently subscribed journals are displayed here alphabetically. All the titles of periodicals are signaged with removable strips.

**Reports Sections**: Reports of National and International Institutes are displayed here chronologically. Publications of FAO, SAARC, NSSO, CSO, DES, Ministry of Agriculture, ICAR, etc. are arranged in separate shelves and classified chronologically. Each shelf has been signaged with name of the organization with multi strips in desktop frames. Reports of all Institutes of ICAR, State Agriculture Universities are arranged Institute wise and classified chronologically in different boxes. Current reports of different organizations are displayed here for awareness of users.

**Book Stack Sections**: Books on different area of specialization are placed as per UDC (Universal Decimal Classification) scheme. Name of subject arranged in each rack has been signaged in the form of shelf panel (shelf list) and fixed on front view of each rack. All shelves of each rack have been signaged with serial number. Number of the shelf where particular subject (Call No.) books are placed in the shelf has been indicated against the subject list in the shelf panel, so that user can see the shelf number of particular subject and can directly go to that particular shelf.

**Stack Room**

**Journals (back volumes) Sections**: Journals subscribed prior to 2008 are bounded and organized alphabetically in this wing. Shelves of the rack has been signaged with serial number, number of the shelf where title of the Journal placed in the shelf has been indicated against the title of that Journal signaged in shelf panel so that user finds very easy to locate that particular journal he wants.

**Current Books Display Section**: Currently purchased books are displayed in this wing after completion of technical processing, dressing up and bar-coding for the period of one month for the awareness of users on recent arrivals. This wing is located at the right side of the entrance of the library.

**Reference Section:** Located at the adjacent to the circulation counter and OPAC Section. Publications of reference nature like Encyclopedia, Dictionaries, Directories, Handbooks, Data Books, Year Books, Maps and Atlases, etc. are arranged in classified order and this section is also signaged and all shelves with title of the resource. These publications are only for reference purpose and not for issue.

**Thesis Section**: Dissertations of Diploma in Agricultural Statistics of IASRI, M.Sc. and Ph.D. in Agricultural Statistics and M.Sc. in Computer Application of PG School, IARI are arranged in classified order in this wing. All bibliographical details of these theses are entered in computerized catalogue and full text of Ph.D. thesis are digitized and given access to use through library web site for research and academic purposes.

**Hindi Publication Section**: Books and reports in Hindi are arranged and displayed in this wing in classified order in different subject areas.

**Official Reference Publications Sections**: Administrative and financial rules books published by Govt. of India etc. are arranged in this section for reference purpose of users.

Apart from above sections, library has got following facilities.

**Library Information System Lab.**: This lab is meant for data entry, updation of data base services, website and resources on regular basis by the staff of library. It is equipped with server and internet connectivity, CD hybrid mirroring technology server, one workstation, four computers, all-in-one printer-cum-scanner-cum-photocopier, barcode printer and flatbed scanner, etc.

**Users Computer lab**: This lab. is meant for users of library for utilizing electronic services and resources being provided by library. This lab is equipped with three computers and a printer.

**Photocopy Room:** Library extends photocopy facility to the users subject to the provision of copy-rights. Provides library literature on free of cost to the scientists of IASRI and charging nominal fee for students as well as users of NARS.

**Committee Room**: Library has got 20 seated committee room with air-conditioning facility and internet for presentation to the committee.

NATP played a vital role in modernization of Institute's library. It funded for strengthening

**Computer Lab**

of library in terms of hardware, Alice for Windows Library Management Software, ISM Multilingual Software, Retro conversion of library catalogues, automation of library house keeping activities, digitization of Ph.D. thesis right from beginning and old and valuable journals, networking of library under its Sub-Project ''**Strengthening of Library Improvement and Networking**". It has also taken care of empowerment of library professional by imparting 15 days training at INFLIBNET Centre at Ahmedabad. Under NATP, automation of library was taken up. Bibliographical details/Cataloguing information of Books, Reports, Journal, Grey literature, Thesis, M.Sc. Dissertation etc., were entered in LMS. These are continuously being updated and made available to its users through LAN and WAN i.e. Web OPAC (On-line Catalogue). The Bar-coding of publications completed, database of profile of library bona-fide members developed with photographs and issued bar-coded membership cards to all the members. The circulation section of library has also been computerized and bar-coded. The digitization of library in respect of all Ph.D. thesis right form 1970's and M.Sc. Dissertation both in Agricultural Statistics and M. Sc. Computer Application from 2005 were completed and made available to the users through LAN.

**Resources and Services**

**Collection:** Over the past four decades of its existence, the library has built up its collection of 26,182 number of subject books, 3086 Hindi books, 8561 back volumes of journals, 9324 gray literature and reports, 967 number of thesis and dissertation, 629 number of CD-ROMs etc. Currently library is subscribing 65 Indian and 63 Foreign Periodicals and 60 On-line Journals and also it is having 09 On-line Portals, 25 CD-ROM Databases, 3000+ e-Journals under CeRA Consortium (Consortium for electronic Resources of Agriculture).

**Services**: Following on-line services are provided through the home page of the library **(http://lib.iasri.res.in)**

**Home Page of Library**

- **Bibliographical Database:** Conventional catalogue of books, journals, Ph.D. thesis, M.Sc. and Diploma dissertations, some of Grey literature (Reports), CD-ROMS available in library have been computerized and given access (WEB- OPAC ) through Library website (http://lib.iasri.res.in) not only to IASRI users but also NARS. Any user on web can retrieve the details of documents of his interest by simple search and can see its physical availability in library and also can reserve the document if it is under loan after logging into his/her account (only for IASRI users).

- **Archival Database:** Ph.D. thesis of IASRI Students as well as some of old, fragile, rare and valuable publications of the library have been scanned and converted into PDF (full text ) and posted on library

website for users of IASRI. User can search and retrieve the full text of thesis/rare journals by simple search. It is retrievable only within the LAN.

- **Off-line Searches (CD-ROM Database)** 35 CD-ROM Databases were procured by library and made available to users for reference. Some of these are stand alone and some are network versioned.
- **Current Content Service (JCC)**: On-line weekly contents with abstracts from currently subscribed journals in library are being provided to the users through this service.
- **Reservation of Document:** Through this service users can make on-line enquires about the availability of a documents in library and also can make reservation of a document if it is under loan and user is intimated by e-mail after the book is received back in the library.
- **Patron Search**: User can know the history of his/her issue account by logging into his/her account.
- **Journals**: 60 journals subscribed by library are made available through library web site with links to full text.
- **Current Awareness Service (New Arrivals) plus Alert Service**: User is updated with monthly list of new additions of books, journals, reports through this service.
- **On-line Portals:** MATHSCINET, INGENTA, INDIASTAT, INDIAN HARVEST, Economic Intelligence Service, State Analysis System, DELNET

## Concluding Remarks

In the above, the development of modern library information services at the institute has been reviewed. This work is continuing and these efforts will help in strengthening the agricultural research, management and development under the National Agricultural Research System as well as providing the much needed information support to the stake holders.

## Saleable Technologies of Indian Agricultural Statistics Research Institute

**SPAR 2.0**: Statistical Package for Agricultural Research data analysis is a user-friendly, menu driven package meant for the analysis of experimental research data in Plant Breeding and Genetics. The package consists of eight modules, viz., Data Management, Descriptive Statistics, Estimation of Breeding Values, Correlation and Regression Analysis, Variance and Covariance Components Estimation, Stability Analysis, Multivariate Analysis and Mating Design Analysis. A Context-Sensitive Help with Index, Contents and Search facility are available.

**SPFE 1.0**: Statistical Package for Factorial Experiments generates randomized layout of designs for symmetrical factorial experiments with and without confounding. The package generates balanced confounded designs for asymmetrical factorial experiments. It also generates fractional factorial plans for symmetrical factorial experiments. The package analyzes the data generated from single factor and multi-factor experiments. This package, besides being useful for the experimenters, is also useful for teaching a course on Design and Analysis of Factorial Experiments.

**SPBD Release 1.0**: Statistical Package for Balanced Incomplete Block Designs enables a user to generate randomized layout of a Balanced Incomplete Block (BIB) design. The package generates asymmetric BIB designs with replication number smaller than 21 and symmetric BIB designs with replication number smaller than 31. The package provides the analysis of variance with both treatments adjusted and blocks adjusted sum of squares, adjusted treatment means, variance of the estimated treatment contrasts and the contrast sum of squares, etc. The definitions of the terminology used are available on-line. The package is useful for the experimenters, classroom teaching as well as for the researchers in Statistics with special interest in Design of Experiments.

**SPAD**: Statistical Package for Augmented Designs generates a randomized layout of an augmented randomized complete block (RCB) design and augmented complete block design with equal or unequal block sizes. The design uses optimum replication number of the control treatments in every block, obtained by maximizing the efficiency per observation for making tests vs controls comparisons. However, users have the freedom to choose replication number of the control(s) in each block. The package also performs the analysis of data generated from augmented block designs (complete or incomplete). The adjusted sum of squares due to treatments can be split up into three components, viz., (i) among test treatments, (ii) among control treatments and (iii) among test treatments vs control treatments. Multiple comparison procedures for making all possible pair wise treatment comparisons can also be employed through this package. A null hypothesis on any other contrast of interest can also be tested.

**SPAB2.0**: Statistical Package for Animal Breeding is useful for analyzing animal breading data with respect to estimation of breeding value of a sire, construction of selection indices including partial and restricted indices. The package is useful for estimation of genetic parameters, prediction of genetic merit [Best Linear Unbiased Prediction (BLUP)], non-parametric tests and for carrying out analysis of multivariate data through principal components, $D^2$ analysis, etc.

**SSDA 1.0**: Statistical Package for Analysis of Survey Data is useful for the analysis of survey data. SSDA analyzes the data collected using simple random sampling (SRS), systematic, probability proportional to size (PPS), stratified, cluster, two stage and stratified two stage sampling schemes. It provides the estimates of population mean, variance and design efficiency of the sampling scheme in comparison to the simple random sampling without replacement. It also provides descriptive statistics of the data without consideration of sampling design, i.e. measures of central tendency and measures of dispersion. The package also has the facility to impute missing data, if any, using commonly used imputation methods. This package is an aid in teaching the subject of analysis of sample survey data to the post-graduate students and is also helpful to the researchers in statistics with interest in sample surveys.

Published by Indian Agricultural Statistics Research Institute, New Delhi 110 012
Printed by M/s Royal Offset Printers, A-89/1, Naraina Industrial Area, Phase-I, New Delhi 110 028