Abstracts of

International Conference on Statistics & Big Data Bioinformatics in Agricultural Research

21–23 November 2016

ICRISAT-India

Organized by

ICRISAT
INTERNATIONAL CROPS RESEARCH INSTITUTE FOR THE SEMI-ARID TROPICS

ISAS

Platinum Sponsors

ICRISAT
INTERNATIONAL CROPS RESEARCH INSTITUTE FOR THE SEMI-ARID TROPICS

ICAR

Gold Sponsors

Integrated Breeding Platform
Today's tools for tomorrow's crops

Biosci

VSNi

Silver Sponsors

DUPONT PIONEER

Sociovent

S TECHNOLOGIES

Contents

Messages

Message from the Director General

Dear colleagues I welcome you to this international conference on Statistics and Big Data Bioinformatics to support Agricultural Development.

The theme is critically important for the world to achieve the sustainable development goals by unlocking the power of big data and bringing it to help us unlock the wealth and diversity we have in our food crops today. The world is facing many challenges: malnutrition, poverty, adaptation to climate change and living within the ecological boundaries of our planet when it comes to preservation of water and soil resources. In the absence of data that helps us manage that complexity, we will not achieve the SDGs.

I see three very large domains around big data for agriculture development.

The first one is around the use of all these wonderful tools to understand genetic diversity. With high-throughput genotyping facilities we are now able to sequence food crops in the course of a year or less. The most recent example for ICRISAT is the finger millet. We are also able to use markers to support high throughput genotyping that enables our breeding programs to use diagnostic markers in a very strategic way to integrate traits.

The second big opportunity we have is around phenotyping. The use of drones to support precision agriculture is accelerating. It also helps us to phenotype at a very large scale the wealth and diversity within our breeding programs. Phenotyping has become more precise with Internet of Things (IoT) and our ability to track moisture accurately, to capture photosynthesis in a precise manner so we can unpack the physiology of photosynthesis, the optimization of water resources for biomass accumulation, green production and of course under the ground, steadying the root system making sure that we can optimally design a plant that enables us to manage our water responsibly and for the planet to withstand periods of drought which will be an increasingly important issue.

The third very large and critically important domain is around geo-spatial information systems. Geo-spatial data can really help us unlock and prioritize our interventions around crop improvement. We can look at soil types, hydrology, etc., to optimize our crops for a particular environment. We can look at malnutrition, food habits and see which mix of varieties is required to deliver balanced nutrition. We can look at poverty and market enhancements that allows us to optimize market access for smallholder farmers to ensure that the varieties we develop meet the needs of processors, aggregators and ultimately society.

I am optimistic that through this workshop we will identify those unique opportunities for partnerships to unlock the power of big data that in turn enables us to unlock the wealth of genetic diversity. Currently we use about 1% of our genetic diversity. Can we move that up to 10% over the next decade? Thinking about how we integrate agriculture scientists with the broader community around health, nutrition and systems biology framework that allows us to really think about agriculture or agri-food systems within a broader frame will enable us to have sustainable equitable food systems for the future.

I wish you well in your deliberations and I look forward to the outcomes.

David J Bergvinson
Director General, ICRISAT,
Patancheru, India

Message from Prof M S Swaminathan

I am glad that an International Conference is being organised on the occasion of the 70th Anniversary of the Indian Society of Agricultural Statistics. This Society has a great tradition of service to the cause of agricultural statistics and had the privilege at one time of having President Dr. Rajendra Prasad as its President. I wish the Society continued success in helping to improve experimental rigour and scientific analysis of data.

Prof M.S. Swaminathan
Founder Chairman
M S Swaminathan Research Foundation
Chennai, India

Message from Director General, ICAR

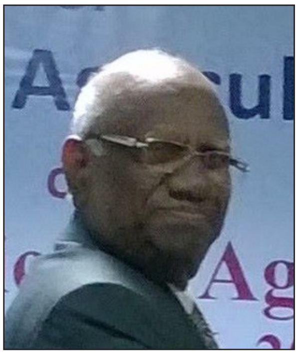
I am extremely happy to know that the Indian Society of Agricultural Statistics is organizing its 70” Annual Conference along with the International Conference on Statistics and Big Data Bioinformatics in Agricultural Research at the International Crops Research Institute for the Semi Arid Tropics, Patancheru, Hyderabad, Telangana during 21-23 November, 2016.

I am sure, this International Conference will provide a wider platform to younger scientists, eminent statisticians, computer scientists as well as experts in bioinformatics from different parts of the world to interact and deliberate on current challenges of research in agricultural statistics, computer applications, informatics and bioinformatics. I hope that the deliberations in this Conference would lead to identification of newer areas of research in agricultural statistics and bioinformatics.

I wish the Conference a grand success.

T. Mohapatra
Secretary, DARE &
Director General, ICAR,
New Delhi, India

Message from Padam Singh



I am glad that the 70th Annual Conference of the Indian Society of Agricultural Statistics as an International Conference is being organized at International Crops Research Institute for the Semi Arid Tropics, Patancheru, Hyderabad, Telangana during November 21-23, 2016.

The theme of this International Conference is Statistics and Big Data Bioinformatics in Agricultural Research. India has been a pioneer in the field of Agricultural Statistics. Dr. Rajendra Prasad Memorial Lecture during the Conference will be delivered by Dr. S Ayyappan, Former Secretary, DARE, Government of India and Director General, Indian Council of Agricultural Research. Also Dr. VG Panse Memorial Lecture will be delivered by Dr. K Alagusundaram, Deputy Director General (Engineering), ICAR, New Delhi.

The technical program comprises of a dozen technical sessions covering different themes together with sessions on contributed papers. It will provide an opportunity to debate and discuss challenging issues being faced by the country in wake of climate change with shifting focus from green revolution to evergreen revolution. The participation of distinguished delegates from all over the country and abroad in the international conference will be of great help in this endeavour.

I have great pleasure in conveying my best wishes for the grand success of this International Conference.

Padam Singh
Director-Invision Communications
and Research Private Limited,
New Delhi, India

Message from Dr. Umesh Chander Sud

It is a matter of pride for the Indian Society of Agricultural Statistics that its 70th Annual Conference is being organized as an International Conference on Statistics and Big Data Bioinformatics in Agricultural Research at International Crops Research Institute for the Semi Arid Tropics, Patancheru, Hyderabad, Telangana during November 21-23, 2016.

I am happy to know that during this International Conference twelve sub-themes Design of Experiments and Phenotyping, Bioinformatics, Data Science Teaching and Scientific Communication, Open Sources Statistical Computing, Big Data Analytics, Statistical Genetics and Genomics, Genomic Selection and Genome Wide Association Mapping, Bayesian Statistics, Geostatistics and Remote Sensing, Statistical Modelling and Forecasting, Data Management, Advances in Statistical Sampling would be organized. The themes are of current interest. I also note that Dr. Rajendra Prasad Memorial Lecture, Dr. VG Panse Memorial Lecture, Technical Address, Invited Talks would be delivered by internationally renowned personalities on emerging issues related to the theme of the Conference.

This Conference will provide a platform to younger scientists, eminent statisticians, computer scientists as well as experts in bioinformatics from different parts of the world to interact and deliberate on current challenges of research in Agricultural Statistics, Computer Applications and Bioinformatics. I hope that the deliberations in this Conference would lead to identification of newer problems of research in the discipline of Agricultural Statistics, Computer Applications and Bioinformatics.

I wish the Conference a grand success and wish the delegates a very fruitful discussion.

U.C. SUD
Director, IASRI
New Delhi, India

Message from V K Gupta

I am enormously fascinated to know that the Indian Society of Agricultural Statistics (ISAS) would be organizing its 70th Annual Conference as an International Conference at ICRISAT, Patancheru, Hyderabad, during November 21-23, 2016. The theme of the Conference "Statistics and Big Data Bioinformatics in Agricultural Research" encompasses within its gamete some important topics like Design of Experiments and Phenotyping, Bioinformatics, Data Science Teaching and Scientific Communication, Open Sources Statistical Computing, Big Data Analytics, Statistical Genetics and Genomics, Genomic Selection and Genome Wide Association Mapping, Bayesian Statistics, Geostatistics and Remote Sensing, Statistical Modelling and Forecasting, Data Management, Advances in Statistical Sampling. I also understand that two parallel pre-conference workshops on (a) Application of R in Bioinformatics, and (b) Breeding Data Management and Analysis using BMS and VSNi tools, would also be organized. The list of invited speakers for the conference is very impressive and highly reputed scientists would be making presentations.

The ISAS, ever since its inception, has been serving the cause of research in statistical sciences and its innovative application to agricultural sciences. It is quite inspiring to understand that the present International Conference would deliberate on the need of further developments in statistical sciences in the changing scenario and their fusion in agricultural sciences to help make the agricultural research globally acceptable. Data is now considered as the fourth paradigm in science; the first three paradigms being experimental, theoretical and computational science. Big data has now become a reality and data transfer, data security and privacy, data sharing, data storage are some of the important open issues. The volume, variety and velocity of data generations continue to scale unprecedented levels, and data processing, data management and data infrastructure are the computing challenges. Big data uses inductive statistics and concepts from nonlinear system identification to infer laws (regressions, nonlinear relationships, and causal effects) from large data sets to reveal relationships, dependencies, and to perform predictions of outcomes and behaviors. These challenges are monumental, but the opportunities are also vast. I am confident that the theme of the present Conference would motivate the young researchers to draw an agenda for future research on these emerging areas.

V.K. Gupta
Former ICAR National Professor
IASRI, New Delhi, India

Message from V K Bhatia

It is a matter of great pride for the Indian Society of Agricultural Statistics that its 70th Annual Conference is being organized as an International Conference at International Crops Research Institute for the Semi Arid Tropics, Patancheru, Hyderabad, Telangana during November 21-23, 2016.

I am glad to know that during the Conference deliberations will be focused on twelve sub-themes Design of Experiments & Phenotyping, Bioinformatics, Data Science Teaching and Scientific Communication, Open Sources Statistical Computing, Big Data Analytics, Statistical Genetics and Genomics, Genomic Selection & Genome Wide Association Mapping, Bayesian Statistics, Geostatistics & Remote Sensing, Statistical Modelling and Forecasting, Data Management, Advances in Statistical Sampling.. These topics are very important and timely. There will also be a series of very important lectures such as Dr. Rajendra Prasad Memorial Lecture, Dr. VG Panse Memorial Lecture, Technical Address and Invited Talks.

The Conference will also provide a platform to younger scientists, eminent statisticians and computer science experts from different parts of the world to interact and deliberate on current challenges of research in Agricultural Statistics, Informatics and Bioinformatics. I hope that the deliberations in this Conference would lead to emergence of newer problems of research in the discipline of Agricultural Statistics, Informatics and Bioinformatics.

I wish the Conference a grand success and wish the delegates a very good academic time.

V.K. Bhatia
Former Director, IASRI,
New Delhi, India

Programme

International Conference on Statistics & Big Data Bioinformatics in Agricultural Research

21st -23rd November 2016

Day 1: 21 November, 2016 (Monday)		
Venue: Ralph W Cummings Auditorium Lobby		
0800- 0900	Registration	ICRISAT Group
Venue: Ralph W Cummings Auditorium		
Inaugural Session		
0900-0910	Invocation Aarti & Lightening Knowledge Lamp	ICRISAT Group
0910-0915	Welcome Address	Dr. Peter Carberry, DDG, ICRISAT
0915-0920	Welcome Address	Dr. UC Sud, Director, ICAR-IASRI & Honorary Secretary, ISAS
0920-0925	Opening Remarks	Dr. Padam Singh, Former Member, National Statistical Commission & Executive President, ISAS
0925-0935	Technical & Keynote Address by Sessional President	Dr. David Bergvinson, DG, ICRISAT & Sessional President
0935-0945	Message from Chief Guest	Dr. Jeffrey Ehlers, Senior Program Officer, Bill & Melinda Gates Foundation, USA
0945-1005	Dr. Rajendra Prasad Memorial Lecture	Dr. S Ayyappan, Former Secretary, DARE, MoAFW, GoI & DG, ICAR and President, ISAS
1005-1025	Dr. VG Panse Memorial Lecture	Dr. K Alagusundaram, DDG Engineering, ICAR
1025-1035	Award Distribution	ICRISAT & ISAS Dignitaries
1035-1045	Vote of Thanks	Dr. Abhishek Rathore, Theme Leader (SBDM) & Conference Organizer
1045-1130	**Tea Break/ Group Photo**	**Academic Court**
Venue: Ralph W Cummings Auditorium		
Theme: Design of Experiments & Phenotyping		**Convener: Fred Van Eeuwijk; Co-Convener: Abhishek Rathore**
1130-1300	Dr. Fred Van Eeuwijk (The Netherlands)	The statistics of phenotyping
	Dr. Marcos Malosetti (The Netherlands)	Modelling spatial field trends by a two-dimensional P-spline mixed model
	Dr. Jens Möhring (Germany)	Do we need check plots in early generation field trial testing?
	Dr. Stuart W Gardner (USA)	Analyzing Field Experiments with GxE Spatial Mixed Models
1300-1400	**Lunch Break**	**204 Banquet Hall Area**

Venue: Ralph W Cummings Auditorium		
Theme: Bioinformatics		**Convener: Kelly Robbins; Co-Convener: Anil Rai**
1400-1600	Dr. Kelly Robbins (USA)	GOBII: Genomic Open-source Breeding Informatics Initiative
	Dr. David Marshall (UK)	With DNA based Variant Data - What You See is What You Get (hopefully)
	Dr. Ramil P Mauleon (Philippines)	Big data analysis in rice at IRRI: Challenges in reproducibility, discoverability, and interoperability
	Dr. Nickolai N Alexandrov (Philippines)	3,000 rice genome project: Big data and bioinformatics perspectives
	Dr. Dinesh Kumar (India)	Computational genomics approach for development of breed and variety signature system for agricultural germplasm management
1600-1630	**Tea Break & Poster Session**	**Academic Court**
Parallel Sessions		
Venue: Ralph W Cummings Auditorium		
Theme: Open Sources Statistical Computing		**Convener: Reinhard Simon; Co-Convener: Hukum Chandra**
1630-1800	Dr. Reinhard Simon (Peru)	Insect life cycle modeling (ILCYM) – an open source tool for entomologists
	Dr. Isaak Yosief Tecle (USA)	solGS: A Webtool for Genomic Selection Analysis
	Dr. Krishnan Sundara Rajan (India)	Satellite Imagery and Spatio-temporal Big-data Challenges in Agriculture – Use of Geospatial technologies
	Dr. Vincent Garin (The Netherlands)	mppR: an R-package for QTL mapping in multi-parent populations using linear mixed models
Venue: 212 CF Bentley Conference Center		
Theme: Advances in Statistical Sampling		**Convener: JP Singh Joorel; Co-Conveners: BVS Sisodia, Sanjeev Panwar**
1630-1800	Dr. UC Sud (India)	Calibration approach based estimators for two auxiliary variables
	Dr. Hukum Chandra (India)	Small Area Estimation of Proportions by Combining Survey Data with Different Levels of Auxiliary Data
	Dr. Sunil Kumar (India)	Study on Non-Response and Measurement Error in Sample Survey
1830-2100	**Dinner**	**IMOD Plaza**
Day 2: 22 November, 2016 (Tuesday)		
Venue: Ralph W Cummings Auditorium		
Special Talk		
0830-0900	Dr. Prashant Gupta (India)	Intelligent Cloud and Data Science to Transform Agriculture

Venue: Ralph W Cummings Auditorium		
Theme: Big Data Analytics		**Convener: Saumyadipta Payne; Co-Conveners: Rajender Parsad and Appavoo Dhandapani**
0900-1100	Dr. Saumyadipta Payne (India)	Big Data Analytics: Views from Statistical and Computational Perspectives
	Dr. P Krishna Reddy (India)	Building Efficient Weather-based Agricultural Decision Support Systems
	Dr. Sanjay Chaudhary (India)	Developing Big Data Analytics Architecture
	Dr. Siddhartha Mandal (India)	Statistical Analysis of High-dimensional Metagenomic Data with Applications in Agriculture
1100- 1130	**Tea Break & Poster Session**	**Academic Court**
Venue: Ralph W Cummings Auditorium		
Theme: Data Management		**Convener: Medha Devare; Co-Convener: Sudeep Marwaha**
1130-1330	Dr. Medha Devare (France)	Enhancing Discoverability and Re-use of CGIAR's Agricultural Data: Challenges and Progress
	Dr. Jawoo Koo (USA)	Managing geospatial database of harmonized, multi-disciplinary agricultural indicators for big data analytics in Africa South of the Sahara
	Dr. Rodrigo Sara (France)	TBD
	Dr. Lukas A Mueller (USA)	Cassavabase – a web-based enterprise-level breeding database for Cassava and other crops
	Dr. Marie-Angélique Laporte (France)	Planteome, a resource for reference ontologies and applications for plant biology
1330-1430	**Lunch Break**	**204 Banquet Hall Area**
Parallel Sessions		
Venue: Ralph W Cummings Auditorium		
Theme: Genomic Selection & Genome-Wide Association Mapping		**Convener: Dietrich Borchadt; Co-Convener: AR Rao**
1430-1630	Dr. Dietrich Borchardt (Germany)	Quantitative Trait Loci - From detection to application in breeding
	Dr. Valentin Wimmer (Germany)	Genome-based prediction in KWS breeding programs
	Dr. Keith Gardner (United Kingdom)	Exploring association mapping approaches in a multi-year, multi-location, multi-pathogen experiment of wheat disease resistance in NW Europe

	Dr. Patrick Thorwarth (Germany)	Application and optimization of genomic prediction models for barley breeding
	Dr. Christina Lehermeier (Germany)	GWAS and genomic prediction across multiple breeding cycles
Venue: 212 CF Bentley Conference Centre		
Theme: Bayesian Statistics		**Convener: Murari Singh; Co-Convener: Himadri Ghosh**
1430-1630	Dr. Jode Warren Edwards (USA)	Improving Decisions in Plant Breeding with a Bayesian Approach
	Dr. Athar Ali Khan (India)	Applications of Bayesian Computing in Modelling of Agricultural Data
	Dr. Satyanshu K Upadhyay (India)	Bayesian Modelling: A Few Related Issues
	Dr. Murari Singh (Jordan)	A Bayesian analysis of data from on-farm trials in legumes in Afghanistan
1630-1730	**Tea Break & Poster Session**	**Academic Court**
1800-2100	**Cultural Program & Gala Dinner**	**Anniversary Lawn**
Day 3: 23 November, 2016 (Wednesday)		
Venue: Ralph W Cummings Auditorium		
Theme: Geostatistics & Remote Sensing		**Convener: Chandrashekhar Biradar; Co-Convener: Tauqueer Ahmad**
0900-1100	Dr. Chandrashekhar Biradar (Jordan)	Earth observation and big-data analytics for building resilient agro-ecosystems
	Dr. Naga Manohar Velpuri (USA)	Analysis of Landsat-based Water Use Estimates (1984-2014) in the Southwestern United States: An Illustration of Big Data Analytics for Agricultural Remote Sensing
	Dr. Surender Varma Gadhiraju (India)	Image Stitching: Reliable base layer for field level studies
	Dr. Mukunda Dev Behera (India)	Geo-Statistics and Remote Sensing -Testing Plant Diversity and Environmental Heterogeneity Hypothesis in India
	Dr. Girish S Pujar (India)	Space Technology based Governance of Watershed Management: Perspectives and Issues in Indian Context
	Dr. Tauqueer Ahmad	Comparative evaluation of advanced classifiers for delineation of area under agroforestry from space-borne multispectral data
1100-1130	**Tea Break**	**Academic Court**
Venue: Ralph W Cummings Auditorium		
Theme: Statistical Modelling & Forecasting		**Convener: Peter Carberry; Co-Conveners: KN Singh**
1130-1300	Dr. Peter Carberry (India)	Some reflections on systems modelling and forecasting
	Dr. K N Singh (India)	Statistical models used for forecasting of food grain production in India

	Dr. Ramakrishna Lahu Shinde (India)	Statistical Modeling of Meteorological Data and its Applications in Weather Based Crop Insurance Schemes for Banana
	Dr. Jana Kholova (India)	Models in agriculture: use of mechanistic modeling approaches to accelerate breeding
1300-1400	**Lunch Break**	**204 Banquet Hall Area**
Parallel Sessions		
Venue: Ralph W Cummings Auditorium		
Theme: Statistical Genetics & Genomics		**Convener: Rajeev Gupta; Co-Convener: Lal Mohan Bhar**
1400-1530	Dr. Raman Babu (India)	Genomics based Next Generation Crop Improvement Approaches
	Dr. Ramesh Naidu (India)	A Machine Learning approach to find genetic diversity in Rice Genome Variants
	Dr. Ramana V Davuluri (USA)	Data Science Applications in Genomics
	Dr. J Sreekumar (India)	Mapping of quantitative trait loci in cassava using Markov chain Monte Carlo (MCMC) method
Venue: 212 CF Bentley Conference Center		
Theme: Data Science Teaching & Scientific Communication		**Convener: Joanna Kane-Potaka; Co-Convener: AK Choubey**
1400-1530	Dr. Joanna Kane-Potaka (India)	TBD
	Dr. Donghui Ma (Thailand)	TBD
	Dr. Madhavi Ravikumar (India)	TBD
	Ms. B Jayashree (India)	TBD
1530-1600	**Tea Break**	**Academic Court**
Venue: Ralph W Cummings Auditorium		
Plenary Talk		
1600-1630	Dr. Gary Atlin, BMGF (USA)	Accelerating genetic gains delivered by plant breeding; the role of data management and selection decision support
Closing Session		
1630-1700	Dr. Abhishek Rathore & ISAS70 Team	

Abstracts
(And biography of invited speakers)

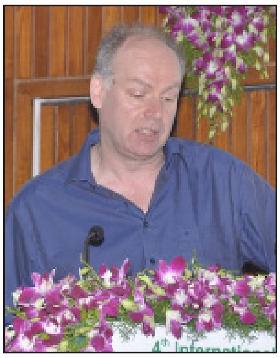
Session I
Design of Experiments and Phenotyping

Convener and Co-Convener

Professor Fred Van Eeuwijk
Wageningen University
Wageningen
The Netherlands

Dr. Abhishek Rathore
International Crops Research Institute for the
Semi-Arid Tropics (ICRISAT)
Hyderabad
India

The statistics of phenotyping



Professor Fred Van Eeuwijk
Professor
Wageningen University
Wageningen
The Netherlands
fred.vaneeuwijk@wur.nl

Dr. Fred van Eeuwijk is Professor in Applied Statistics, Wageningen University, The Netherlands. Dr. Eeuwijk studied Biology and Philosophy at Utrecht Univeristy in the Netherlands, where he graduated in 1985. He worked as a statistician at Plant Research International (1985-1996) , associate profeesor in Statistics at Wageningen University (1997-2000), associate professor in Plant Breeding at the same university (2001-2006) and in October 2006 full professor in Applied Statistics, again in Wageningen. He has more than 20 years of experience in the field of plant genetics and applied statistics. His main research topics include the statistical analysis of GXE and the development of QTL mapping methodology.

Abstract

In the past decade new genotyping techniques have dominated research in statistical genetics for plants. Recently, developments in phenotyping techniques created renewed interest in statistical methods for supporting and improving phenotyping strategies. Statistical methods can contribute at four levels to phenotyping strategies. At the most elementary level, the analysis of individual experiments, whether in the field or at phenotyping platforms, requires statistical design and analysis methods that minimize the influence of disturbing noise factors at genotypic comparisons. A second level of statistical contributions to phenotyping strategies occurs with respect to the choice of the genotypes for phenotypic evaluation. Different breeding populations and panels will have different properties with respect to the kind of genetic effects that can be detected and the power with which those can be detected. Statistical arguments can help in choosing the genotypes to be included in field trials and phenotyping platforms for phenotypic evaluations. Similar reasoning holds for the choice of environmental and management conditions under which sets of genotypes will be evaluated for phenotypic differences. Finally, breeding programs typically aim at improving yield, resistance/ tolerance against stresses, and quality. To improve the target trait, mostly yield, it may help to include component traits in prediction models for the target trait. Various statistical options are available to improve the prediction of a target trait by including component traits in multi-trait prediction models. In our presentation, we will illustrate how statistical methods can contribute to more efficient phenotyping strategies. We will also address some of the complications arising from the occurrence of genotype by environment interactions.

Modelling spatial field trends by a two-dimensional P-spline mixed model

Dr Marcos Malosetti
Researcher
Wageningen University,
Wageningen
The Netherlands
marcos.malosetti@wur.nl

Dr. Marcos Malosetti studied Agronomy in Universidad de la República, Uruguay where he graduated in 1996. He worked in the Plant Breeding group until 2000 where he started his postgraduate studies in the Netherlands. In 2002 obtained his MSc degree at Wageningen University, and in 2006 completed his PhD disertation at the same University. Since 1996 his activities involved research and teaching in the field of quantitative genetics applied to plant breeding, with a strong component on the development and use of statistical methods. The focus of the work has been on the analysis of complex data sets (i.e. multi-environment data, multiple traits data, etc), and its integration with molecular marker techniques with the objective of understanding the variation of quantitative traits.

Abstract

Large field trials are commonly used in agriculture. One example are plant breeding trials, in which breeders have the task of evaluating large numbers of candidate genotypes among which to select. A precise genotypic assessment is a key element for increasing the chances of genetic gain. However, precision can be compromised because of field heterogeneity. The two major allies for the breeders are experimental design and data analytical methods. Advanced experimental designs allow using prior information to impose restrictions to the allocation of treatments in the field, and thus reducing potential confounding with extraneous sources of variation such as row and column field trends. Besides the design, advanced analytical methods are required to obtain the adjusted means, which almost inevitably lead to some form of mixed model analysis of various degrees of complexity. The focus of this presentation will be on the latter, by introducing the methodology for spatial modelling recently proposed by Rodriguez-Alvarez et al (2016) who coined the acronym SpATS for it. The SpATS methodology uses two-dimensional P-splines to fit a bivariate smoothed function to the field trends. A further attractive feature of the methodology is that it is formulated within a mixed model framework offering a natural link with other mixed model approaches and facilitating interpretation of results. Besides, the methodology links naturally with other approaches since it allows the inclusion of experimental design features such as (incomplete) blocks, row and/or column effects as fixed or random terms in the mixed model. The methodology will be illustrated with simulated and real plant breeding trials, and a comparison with standard spatial mixed model analysis is made. The main conclusion of this presentation is that SpATS has a high potential in routine plant breeding trials, especially when the time and/or expertise to run elaborate mixed model analyses is restricted. The methodology is available as an R-package https://cran.r-project.org/package=SpATS.

Do we need check plots in early generation field trial testing?

Dr. Jens Möhring
Biostatistican
University of Hohenheim
Stuttgart
Germany
Jens.Moehring@uni-hohenheim.de

Dr. Möhring is associated scientists working in teaching and research in the Biostatistic unit at the University of Hohenheim, Stuttgart, Germany. After finishing his study of agriculture biology in 2003 he made his PhD in biostatistics in the work group of Prof. Piepho at the same university. His research interest are the optimization of field and experimental designs in cultivar and plant breeding science. He focuses on the application of mixed models on multi-environmental data and their sophisticated analysis as a pre-requirement for genetic (marker-based) analysis.

Abstract

In plant breeding, testing of entries in field trials always include some well-known checks. In replicated trials they are used as benchmark for selecting preferred entries or for checking environmental conditions like winter damage or pressure of infection with diseases. Additionally, in augmented designs with unreplicated entries checks are used for adjusting for block effects and estimating an error variance in each trial. Therefore, it is common to increase the number of check plots. For the latter design, simulations show that partially replicated (p-rep) designs are a preferred alternative (Cullis et al. 2006; Möhring et al. 2015). P-rep designs replace check plots by plots for entries already tested within the trial. As it is common to add checks in field trials, the amount of check plots and the benefit of additional check plots in partially replicated designs is unknown. We analyse five large one-year series of winter barley yield trials conducted as p-rep design with additional check plots within each incomplete block using a mixed model approach. We used four approaches fitting three different models to four subsets of the dataset to estimate entry main effects: (i) corresponds to a p-rep design using a model without check effects to data dropping observations from checks, (ii) corresponds to an augmented design using a model including check effects and dropping observations from half of the plots of replicated entries, (iii) the same model as in (ii) to all data or (iv) corresponds to an unreplicated design using a model with confounded entry-by-location and error variance and dropping both all observations from checks and half of the observations of replicated entries. Furthermore, we vary the assumption of error effects by fitting independent and spatially correlated error effects. We compared the approaches using the average s.e.d., the inter-location correlation of entry estimates and the correlation of selected entry estimates from the current year and their performance in the following year.

Analyzing Field Experiments with GxE Spatial Mixed Models

Dr. Stuart W Gardner
Research Scientist
DuPont Pioneer
IA
USA
stuart.gardner@pioneer.com

Dr. Stuart W Gardner received his BS in Forestry from the University of Maine in 1995 and his PhD in Statistics and Microbiology from Iowa State University (ISU) in 2010. Dr. Gardner served in the US Peace Corps in Nepal (1996-1998) and continued working in Nepal until 2000 in the environmental development field. At ISU, he worked in Small Grains Breeding developing QTL markers for oats (2001-2003) and then as Statistics Instructor (2003-2006). He has been a statistician at DuPont Pioneer since 2007 supporting breeding programs for maize, rice, canola, and sunflower. His expertise is in linear mixed spatial models and their application to field trials for breeding programs.

Abstract

The logistics of analyzing field trials necessitates modeling of field variation to remove noise when producing BLUPs for advancement decisions. Augmented single replicate designs have been employed for several years where the field variation is accounted for by spatial models with row and column effects on the G structure and autoregressive 1 effects on the R structure. Additionally these trials are conducted at multiple locations and several genotype-by-environment models such as compound symmetry, uniform correlation with heterogeneous variances, and factor analytic models can be fit to the data. In this talk I present the theory of these Mixed Spatial GxE models and some of the practical learnings I have gained over the years from fitting these models to various crops such as maize, rice, sunflower, and canola.

Session II
Bioinformatics

Convener and Co-Convener

Dr Kelly Robbins
Cornell University
Ithaca
USA

Dr Anil Rai
Centre for Agricultural Bioinformatics, Indian
Agricultural Statistics Research Institute (IASRI)
New Delhi
India

GOBII: Genomic Open-source Breeding Informatics Initiative

Dr. Kelly Robbins
Quantitative Geneticist
Cornell University
Ithaca
USA
krr73@cornell.edu

Dr. Kelly Robbins is the Director of the Genomic Open-source Breeding Informatics Initiative at Cornell University. Dr. Robbins studied Animal Breeding and Statistical Genetics at the University of Georgia where he graduated in 2007. He worked as a quantitative geneticist at Dow AgroSciences (2008-2015) where his research centered on the use of mixed models, Bayesian methodologies, and machine learning algorithms for the modeling of genomic data with applications in plant breeding and gene discovery. His current research focuses on the development of computation tools and databases for management and integration of genomic information into breeding programs focused on developing improved cultivars for South Asia and Sub-Saharan Africa.

Abstract

In the last ten years, genotyping costs have dropped significantly, making feasible powerful new breeding approaches that take advantage of the vast amounts of genomic data that have been generated. The Genomic & Open-source Breeding Informatics Initiative (GOBII) is the first large-scale public-sector effort to enable systematic application of high-density genotypic information to the breeding of staple crops in the developing world, through the development of a genomic data management system and analysis pipelines. The genomic data management system is comprised of a data warehouse (postgreSQL, Monet, and HDF5), a service layer with a BrAPI compliant RESTFUL API, and GUIs for loading and extracting data.

Big data analysis in rice at IRRI: Challenges in reproducibility, discoverability, and interoperability

Dr. Ramil P Mauleon
Bioinformatics Scientist
International Rice Research
Institute (IRRI)
Philippines
r.mauleon@irri.org

Dr. Ramil P. Mauleon is a bioinformatics scientist at the International Rice Research Institute (IRRI), based in the Philippines. He studied Genetics (and Computer Science) at the University of Philippines, working on transcriptome profiling of mutant rice lines, and graduating in 2006. His post-doctoral work at the Generation Challenge Program Bioinformatics subprogram focused on transcriptomics and bioinformatics analyses for candidate gene discovery. His current involvement at IRRI are with the International Rice Informatics Consortium, the 3,000 Rice Genomes project and the Genomic Opensource Breeding Informatics Initiative. His research interests include creating re-usable analysis workflows for high throughput datasets with high computing demand, and in-silico identification of candidate gene(s) causal to important traits in rice through analyses of heterogeneous datasets (ie. gene/genome sequences and variations, gene interaction network inferences, and meta-analyses of published QTLs).

Abstract

Numerous rice genome sequencing and high throughput genotyping projects (eg. 3k Rice Genomes, Rice High Density Array) have generated large datasets that are difficult to transfer, handle and analyse, and even more difficult to systematically capture the methods and output of analyses done. Analysis such as variant calling is computationally expensive, thus, outputs of this analysis become an important public good by itself. The challenge of data transfer and standardization of analyses methods is being addressed at IRRI by using publicly available bioinformatics analyses and workflow management systems (eg. Galaxy) that interoperates with external data sources. Galaxy allows easy integration of new analysis tools and extendibility to allow sharing of analyses workflows/input/output data on the system that are discoverable via data registry and persistent object identifier system. To address data/database interoperability challenges, database systems for breeding and genotyping at IRRI (eg. Breeding4Rice, Genotyping4Rice-GOBII) will also adhere to public standards for data description (eg. Rice Ontology) and Application Programming Interfaces via web services (eg. BrAPI) to allow interoperability of data and databases internally and with external systems that are BrAPI compliant.

With DNA based Variant Data - What You See is What You Get (hopefully)

Dr. David Marshall
Bioinformatician
The James Hutton Institute Invergowrie,
Scotland
United Kingdom
David.Marshall@hutton.ac.uk

Dr. David Marshall has worked on development of genetic marker technologies and DNA sequence diversity in both crop and wild plants and their deployment in applications from the analysis of the genetic structure of natural and breeding populations of crop plants to genome analysis and comparative genomics. The major focus of his interests is on genetic marker applications with barley, potatoes, wheat and soft fruit (blackcurrant and raspberry) in collaboration with Professor Robbie Waugh and other Geneticists at Dundee, in particular, the design and application of SNP-based genotyping arrays. Dr Marshall has developed of software tools to support the analysis of genotype and sequence data.

Abstract

The development of high though put technologies for both sequencing and genotyping are now leading to major changes in the approaches that are being adopted in the characterisation and utilisation crop germplasm at every stage from the characterisation of plant genetic resources to breeding applications. Hover the resulting high density data that it is not possible to generate in turn has led to a major problem in both quality control of the molecular data to our ability to browse and manipulate large complicated data sets. Within the James Hutton Institute we are developing a series of visualisation components which compliment the repertoire of bioinformatics and statistical tools we utilise at every stage in the process from variant calling and assay design to decision support based on the integration of trait, QTL, marker and pedigree information.

3,000 rice genome project: Big data and bioinformatics perspectives

Dr. Nickolai N Alexandrov
Senior Scientist
International Rice Research
Institute (IRRI)
Philippines
n.alexandrov@irri.org

Rice bioinformatics has exploded when the 3000 rice genome sequencing project lead by scientists from CAAS, BGI and IRRI was completed in 2014. Unprecedented amount of sequence data required immediate attention: it had to be reliably stored, carefully QCed, proper organized in data structures and thoroughly analyzed. Not everything went smoothly and there were unexpected delays and a few mistakes. Nevertheless, we have made the main results from this project available to rice community via friendly and interactive web portal. How this data helps breeders and other rice scientists in their work? What we have achieved and what new discoveries we should expect in the near future? How we should be prepared to meet new challenges in rice bioinformatics? More questions will be discussed and some of them will be answered.

Absract

Rice bioinformatics has exploded when the 3000 rice genome sequencing project lead by scientists from CAAS, BGI and IRRI was completed in 2014. Unprecedented amount of sequence data required immediate attention: it had to be reliably stored, carefully QCed, proper organized in data structures and thoroughly analyzed. Not everything went smoothly and there were unexpected delays and a few mistakes. Nevertheless, we have made the main results from this project available to rice community via friendly and interactive web portal. How this data helps breeders and other rice scientists in their work? What we have achieved and what new discoveries we should expect in the near future? How we should be prepared to meet new challenges in rice bioinformatics? More questions will be discussed and some of them will be answered.

Computational genomics approach for development of breed and variety signature system for agricultural germplasm management

Dr. Dinesh Kumar
Principal Scientist (Biotechnology)
Centre for Agricultural Bioinformatics (CABIN)
Indian Agricultural Statistics Research Institute (IASRI)
New Delhi
India
dineshkumarbhu@gmail.com

Dr. Dinesh Kumar is presently working as Principal Scientist (Biotechnology) at Centre for Agricultural Bioinformatics (CABIN), ICAR-Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, Pusa, New Delhi. Prior to this he has worked as Senior Scientist (Biotechnology) at ICAR-NBAGR, Karnal (1997-2012). He completed PhD in Biotechnology from Banaras Hindu University, Varanasi and PDF (2003-04 & 2007-08) from Iowa State University of Science & Technology, Ames, USA.

Abstract

India is one of the 12th mega biodiversity centres of the world. Due to climatic diversity, India has unique and vast agri-germplasm diversity having of crop varieties, animal and fish breeds each representing unique combinations of genes in its respective ecological niches. In post CBD era having globalisation sweep, the germplasm movement is imperative, rendering a great challenge in managing the sovereignty of Indian agri-germplasm. For crop DUS (distinctness, uniformity and stability) feature and for animal and fish breed descriptors have been developed to identify them. Such features or descriptors cannot address issues of admixture, essentially derived varieties (EDV) or germplasm like fish seed/milt/fingerlings, semen, ova, embryo, breed, crop seed and produce where there is no visible phenotypic descriptors. Advent of molecular markers like microsatellite and SNP have revolutionised breed/variety identification from even small biological tissue or germplasm. Microsatellite DNA marker based variety and breed tests have been reported but such methods have limitation of non-availability of DNA data in public domain, thus each time all reference breeds/varieties have to be genotyped which is neither logical nor economical. Even if such data is available but computational method needs expertise of data analysis and interpretation. Computational team of ICAR-IASRI has developed animal breed and crop varieties identification web servers having readily available reference breed and varieties data thus unknown germplasm can be conveniently compared with existing referral breeds/varieties for identification using DNA data. Such computational approach can address the issue of trans-boundary movement of germplasm with bio piracy, issues of trace-ability of agri-produce as well as effective legal protection/dispute resolution especially of crop germplasm varieties which has status of intellectual property having issues of sovereignty or ownership.

Session III
Open Sources Statistical Computing

Convener and Co-Convener

Dr. Reinhard Simon
Bioinformatics Scientist
CIP
Peru

Dr. Hukum Chandra
ICAR- National Fellow & Principal Scientist
Indian Agricultural Statistics Research Institute
(IASRI)
New Delhi
India

Insect life cycle modeling (ILCYM) – an open source tool for entomologists

Dr. Reinhard Simon
Bioinformatics Scientist
CIP
Peru
r.simon@cgiar.org

Dr. Reinhard Simon is Data scientist with a biological background starting with farming to conducting field trials to working with farmers groups in seed production to applying molecular biology to breeding and genetic resources management and, as of late, using informatics for more effective use of field and lab data. Experience in data mining, machine learning, high performance computing, parallel computing and cloud computing. Google Scholar: http://scholar.google.com.pe/citations?user=NVEFblESICwJ.

Abstract

The Insect Life Cycle Modeling software (ILCYM) assists in the development of phenology models. It also provides analytical tools for studying pest population ecology. The software consists of three modules. The "model builder" facilitates the development of insect phenology models based on experimental temperature-dependent life-table data of a specific pest. The 'Validation and simulation module' provides tools for validating the developed phenology model using insect life-tables established at fluctuating temperature conditions. It also provides tools for deterministic and stochastic simulations. The 'Potential population distribution and risk mapping module' implements the phenology model in a GIS environment and allows for spatial - global or regional - simulation of pest establishment, population growth and activity ("pest risk mapping"). In comparison to other generic pest risk modeling tools ILCYM allows a better biological understanding and a higher modeling precision using different climatic input data (global, regional, or from local weather stations. Other applications of the software include assessment of climate change impacts. Recently, a virus transmission model has been developed which will be integrated in the software. On a data and software engineering level, ILCYM is an example of open source tools using 'big' spatial data sets in combination with traditional entomological tools. The current version builds on two main open source platforms: R for all calculations and Java/Udig for building the interface. However, we are transitioning to a version that will use R also for building the user interface. This reduces complexity and facilitate its use by researchers and students using advanced modeling techniques.

solGS: A Webtool for Genomic Selection Analysis

Dr. Issak Yosief Tecle
Bioinformatics Scientist
Cornell University
Ithaca
USA
iyt2@cornell.edu

Dr. Isaak Yosief Tecle is trained in plant breeding and genetics (PhD, Cornell '06). His research focus is developing bioinformatics tools that help plant breeders make efficient decisions in their crop improvement programs. This includes developing intuitive web-tools for statistical analysis, visualization and sharing 'omics' data. He is currently working on a genomic selection web-tool for the project.

Abstract

Genomic selection relies on high-density genotyping and computationally intensive statistics, which presents significant challenges in data management, analysis and results sharing. solGS, a web-based tool, is designed to alleviate these challenges for breeders. It has a database to store phenotype and genotype data and a web-interface for statistical analyses and interactive data visualization. It uses RR-BLUP for the statistical modeling and GBLUP method for breeding values prediction. It performs also descriptive statistics, population structure, phenotypic and genetic correlations, and selection index analyses. solGS is part of the NextGen Cassava Breeding Project (http://nextgencassava.org) and implemented on http://cassavabase.org/solgs. solGS is organism-agnostic, and thus GS breeders can adapt it for any organism.

mppR: an R-package for QTL mapping in multi-parent populations using linear mixed models

Mr. Vincent Garin
PhD candidate
Wageningen University
Wageningen
The Netherlands
vincent.garin@wur.nl

Mr. Vincent Garin is a PhD candidate coming from Switzerland. He earned a Master degree in Criminology from University of Lausanne and a Master degree in Statistics from University of Geneva. He is actually doing is PhD at Wageningen University in collaboration with KWS saat AG. His project is on statistical methods for QTL detection in multi-parent populations. Vincent is interested in agriculture in general especially in small farmers issues. He had several working experiences in Swiss small farms.

Abstract

Multi-parent populations (MPPs) are an interesting resource in plant genetics that have grown in popularity. In this talk we will present MppR, an R-package for QTL mapping in MPPs. The package can handle various types of populations involving more than one cross such as nested association mapping (NAM), half-dialleles, and incomplete factorial designs. MppR include features for all common steps in QTL experiment analysis: data pre-processing, significance threshold determination, cofactors selection, multi-QTL model search, cross-validation, genetic effect estimation and visualisation of the results.

A central question in MPP QTL analysis is the definition of genetic relatedness between the genotypes coming from different crosses. In MppR we can limit relatedness to the cross level and consider that QTL effect are different in each cross. A second model integrates connection between crosses due to shared parents. A third possibility is to consider that parents are not independent but are related via common ancestors. Based on parental genetic similarity a reduced number of ancestral alleles can be inferred using for example the R package clusthaplo (Leroux et al. 2012). Then relatedness between lines can be redefined through the share of a common ancestor above the parental lines. A last possibility is to define relatedness between genotypes based on identical by state (IBS) marker information like in association mapping analysis.

The estimation of genetic relatedness within MPPs designs can also be realised in a more unified and sophisticated way by methods like the reconstructing ancestry block bit by bit (RABBIT) algorithm (Zheng et al. 2015). This technology is based on a general hidden Markov model (HMM) framework allowing to model recombination events in many MPPs designs. Given mating scheme, reference ancestor and final lines genotypes, RABBIT produces identical by descent (IBD) probability allowing to trace origin of the genotypes allele back to the reference ancestors. RABBIT outputs could be later used in mppR models.

A second important question for QTL detection in MPPs is to define appropriately the form of the residual polygenic term. mppR proposes several mixed models allowing to relax the linear assumption of an homogeneous variance residual terms by the use of cross-specific residual variance terms or by the a direct modelling of the polygenic term using a random term based on pedigree relationships. The computation of the different mixed models in mppR use functions from the asreml-R package.

Satellite Imagery and Spatio-temporal Big-data Challenges in Agriculture – Use of Geospatial technologies

Dr. Krishnan Sundara Rajan
Associate Professor
International Institute of
Information Technology,
Hyderabad, India
rajan@iiit.ac.in

Dr. KS Rajan is an Associate Professor at IIIT, Hyderabad and leads the institute's Lab for Spatial Informatics (LSI). Dr. Rajan is a multi-disciplinarian, with major interests in Geo-Spatial Technologies - GIS and Remote Sensing; Land use modelling and Environmental Policy; and gGovernance. He has taken a key interest in the areas that overlap computer science and geospatial technologies and through his research works has helped focus on bridging this gap be it in developing spatio-temporal data mining algorithms; Spatio-temporal big data analytics; Web-based Geospatial technologies; or New algorithms to help convert satellite imagery to useful satellite based thematic products. Dr. KS Rajan is also an active proponent of OpenSource in India and for Geospatial technologies in particular. His Lab released two Open Source tools – LSIViewer and VRGeo.in and actively contribute to a few others.

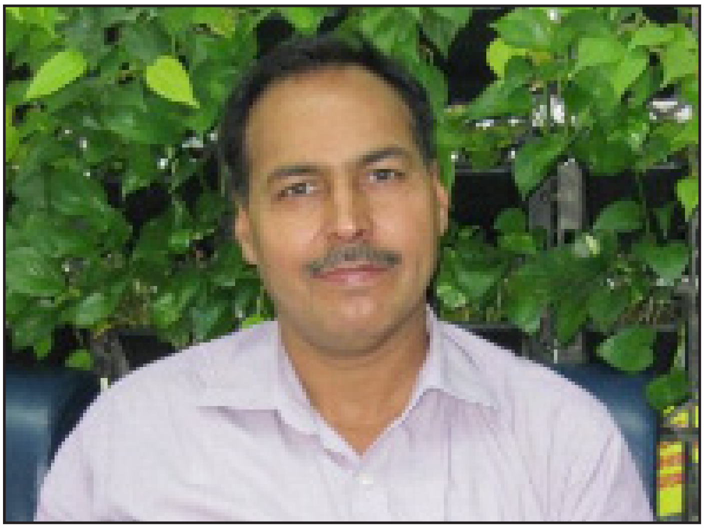
Abstract

Satellite imagery and its products have been used in a wide variety of applications in agriculture over the last few decades, but the challenge of monitoring crop growth, understanding land cover history and providing a user-driven understanding of the agricultural landscape has still eluded the field. While on one hand, a range of IoT devices are being employed in Agriculture for monitoring specific parameters and building response activities – these still do not provide the systemic view needed for assessing, monitoring and response modeling the agricultural practices. In addition to the spatial variability of the factors that influence agricultural processes and practices, the temporal dynamics of such changes are also important clues.

Some of the questions that these pose are - Can an open data infrastructure help provide for a multi-disciplinary work and enhance the understanding in the field? Are technologies useful for a collaborative decision making process to be adopted? What kind of platforms do we need to achieve smart decision making? In this talk, we would like to present and share our work on how a change in satellite data processing paradigm can help better understand the agricultural practices and some insights on how technological developments in both ICT – bigdata, IoT, data analytics etc., and Geospatial sciences will help to understand such complex interactions and enable to build a platform for a participatory and community led management of the resources.

Session IV
Advances in Statistical Sampling

Convener and Co-Convener



Dr. J P Joorel
Dean, Faculty of Mathematical Sciences University of Jammu,
Jammu
India

Dr. B V S Sisodia
Professor & Head,
Department of Agricultural Statistics,
N.D. University of Agriculture & Technology
Faizabad
India

Calibration approach based estimators for two auxiliary variables

Dr. Umesh Chander Sud
Director
Indian Agricultural Statistics Research Institute (IASRI)
New Delhi
India
ucsud@iasri.res.in

Dr. Umesh Chander Sud is presently working as Director ICAR-Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, Pusa, New Delhi. Prior to joining as Director, ICAR-IASRI, he worked as Head, Division of Sample Survey, IASRI. His areas of interest include small area estimation, use of calibration approach for estimation of finite population parameters, methodology development for estimation of parameters relating to crop and livestock sectors.

Abstract

The calibration approach has been used to develop estimators when information on two auxiliary variables is available. Complete information is assumed available on variable which is less correlated with the study variable while no information is available on variable which is highly correlated with the study variable. Expressions for bias and variance of the estimator are obtained. Using real data as well as simulated data, it has been shown that the calibration approach based estimator beats many available estimators in terms of the criterion of MSE of an estimator.

Small Area Estimation of Proportions by Combining Survey Data with Different Levels of Auxiliary Data

Dr. Hukum Chandra
ICAR- National Fellow & Principal Scientist
Indian Agricultural Statistics Research Institute
(IASRI)
New Delhi
India
hchandra12@gmail.com

Dr. Hukum Chandra is ICAR- National Fellow & Principal Scientist at Indian Agricultural Statistics Research Institute, New Delhi, India. He did his PhD from the University of Southampton, United Kingdom and Post Doctorate Research from University of Wollongong, Australia. His main research areas include Sample Survey Design and Analysis, Small Area Estimation, Bootstrap Methods, Statistical Modelling and Data Analysis, Statistical Methodology for Improvement of Agricultural Statistics. Dr. Hukum Chandra has made outstanding research contributions to the discipline of Statistics in general and Sample Survey in particular for which he has been recognized globally. He has received number of awards and appreciations for his excellent research contribution such as National Award in Statistics by the Ministry of Statistics and Programme Implementation, Govt of India, ICAR-National Fellow award, Cochran-Hansen Award by the International Association of Survey Statisticians, Lal Bahadur Shastri Outstanding Young Scientist Award of ICAR, P.V. Sukhatme Gold Medal Award and Dr. D.N. Lal Memorial Award of Indian Society of Agricultural Statistics, Young Researcher/Student award of the American Statistical Association. He is recipient of British Government Commonwealth Scholarship offered by Commonwealth Scholarship Commission in the United Kingdom. He has published 75 research papers in number of reputed national and international journals of high impact factor. He has also published three books, several technical bulletins, book chapter etc. He also has delivered number invited talks in many national and international platforms of repute. He is council member of International Association of Survey Statisticians, elected member of International Statistical Institute and International Consultant Sampling for the Food and Agricultural Organization of the United Nations.

Abstract

Binary data are often of interest in many small area applications. The application of standard small area estimation (SAE) method based on linear mixed model becomes problematic for such data. The empirical best predictor (EBP) under unit level generalized linear mixed model (GLMM) with logit link function is used for the estimation of small area proportions and is proven to be efficient. However, the EBP estimator requires the availability of unit level population information for auxiliary data which may not always be available. As a consequence, in many practical situations, the EBP approach cannot be applied. Based on availability of auxiliary data at different levels, we describe different estimators for estimation of small area proportions. We also develop two approaches of MSE estimation, the analytical MSE and the parametric bootstrap based MSE. Monte Carlo simulations based on both simulated and real dataset show that the proposed small area estimators work well for generating the small area estimates of proportions and represent practical alternative to the EBP. Developed estimator is also applied to debt investment survey 2012-13 data conducted by national sample survey office for rural areas of the state of Bihar in India to generate reliable district level estimates and spatial mapping of incidence of indebtedness for farm households.

Study on Non-Response and Measurement Error in Sample Survey

Dr. Sunil Kumar
Assistant Professor
General Management
Alliance University
Bengaluru
India
sunilbhougal06@gmail.com

Dr. Sunil Kumar is a Statistics Professor, Scholar and Researcher working with Alliance University, Bangalore, India (since 2013). Dr. Kumar hold graduate and post graduate degrees in Statistics from University of Jammu and Master of Philosophy and Doctorate degree in Statistics in the year 2007 and 2010, respectively, from Vikram University, Ujjain, M. P. (India). He worked as Visiting Scientist/Professor at Indian Statistical Institute, Kolkata (2012-2013), Visiting Faculty at Department of Statistics, University of Jammu (2010-2012). Dr. Kumar has published many articles in International Journals. His main research interests are: sample survey, non-response, measurement error, consumer behavior, latent class analysis and applications, Statistical Analysis, Applied Statistics, Statistical Inference, Data Analysis, Qualitative Research, E-Commerce, Mobile Applications, and Commerce.

Abstract

This study concerns about the quality of survey research. Non-response and measurement error are the main possible sources of error in sample survey research. In this paper, the problem of estimation of finite population mean of the study variable is discussed in the presence of non-response and measurement error by using the auxiliary variable. Some realistic conditions have been obtained under which the proposed estimator is more efficient than usual unbiased estimator, ratio estimators and product estimators. Empirical study is also carried out to support the theoretical findings in different situations.

Session V
Big Data Analytics

Convener and Co-Convener

Dr. Saumyadipta Pyne
Professor
Indian Institute of Public Health
Hyderabad
India

Dr. Rajender Parsad
Principal Scientist
Division of Design of Experiments,
Indian Agricultural Statistics
Research Institute (IASRI)
New Delhi
India

Dr. Appavoo Dhandapani
Principal Scientist
National Academy of Agricultural
Research Management
Hyderabad,
India

Big Data Analytics: Views from Statistical and Computational Perspectives

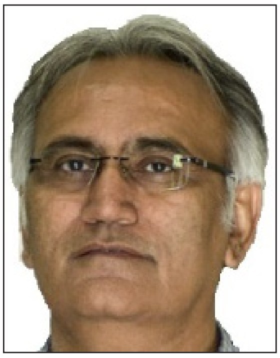
Dr. Saumyadipta Pyne
Professor
Indian Institute of Public Health
Hyderabad
India
spyne@broadinstitute.org

Dr. Saumyadipta Pyne joined as Professor at the Indian Institute of Public Health, Hyderabad, in late 2015. Formerly, he held the prestigious PC Mahalanobis Chair while being Professor and Head of Bioinformatics, in CR Rao Advanced Institute of Mathematics, Statistics and Computer Science. He is also Remote Research Associate of Broad Institute of MIT and Harvard University, USA. Dr. Saumyadipta Pyne research interests include Big Data in Life Sciences and Health Informatics, Computational Statistics and High-dimensional Data Modeling. He is actively engaged in promoting Big Data research and training activities in India and abroad. He is the founder Chairman of Computer Society of India's Special Interest Group on Big Data Analytics, and also leads the Health Analytics Network. In 2016, Springer will publish the text, 'Big Data Analytics: Methods and Applications' edited by S. Pyne, et al.

Abstract

Without any doubt, the most discussed current trend in computer science and statistics today is BIG DATA. Different people think of different things when they hear about big data. For the statistician, the issues are how to get usable information out of datasets that are too huge and complex for many of the traditional or classical methods to handle. For the computer scientist, big data poses problems of data storage and management, communication, and computation. For the citizen, big data brings up questions of privacy and confidentiality. Big Data appears to have some unique properties, which lead to new challenges for both statistics and computer science. In this talk, I will discuss some of the differences in terms of "cultural perspectives" of statisticians and computer scientists towards tackling the common Big Data problem, and illustrate how some of the new approaches are trying to bridge such differences.

Developing Big Data Analytics Architecture



Dr. Sanjay Chaudhary
Associate Dean
School of Engineering and Applied Science
Ahmedabad University
India

Dr. Sanjay Chaudhary is Professor and Associate Dean at School of Engineering and Applied Science, Ahmedabad University. His research areas are Distributed Computing, Cloud Computing, Data Analytics, and ICT Applications in Agriculture and Rural Development. He has authored four books and six book chapters. He has published more than ninety research papers in international conferences, workshops and journals. Dr. Sanjay Chaudhary is an active member of program committees of leading International conferences and workshops as well as review committees of leading journals. He has received research grants from leading organizations including IBM, Microsoft and Department of Science and Technology, Govt. of India. Four PhD candidates have completed PhD under his supervision and four PhD students are currently working under his guidance. He loves literature, music, travel and wild life.

Abstract

In this fast, growing modern view of data generation and consumption, traditional methods may fail to handle high-volume, high-speed, and variety of complex data. Ubiquity of complex data from new data sources has evolved into a new technology called big data analytics. Modern trends in the agriculture domain have made people realize the importance of big data. The key challenge of big data in agriculture is to identify the effectiveness of big data analytics. Moreover, how big data analytics can be used to improve the productivity in agricultural practices and bridge the gap between rural communities and information through recommendation and decision support system. Geo-spatial data is always big data. Spatial data are playing important role in many application areas such as Agriculture, finance and banking, disease surveillance, retail and E-commerce, etc. There is an urgent need to manage spatial-temporal data at scale with specialized systems, techniques and algorithms. We are in the process of developing a robust and efficient framework/architecture for massive scale Geo-spatial data management using state-of-the-art big data tools and technology in real time and batch processing. It uses OGC standards for big data processing for spatial data analytics, develop a prototype to showcase development and execution of various analytical services related to agriculture domain.

Building Efficient Weather-based Agricultural Decision Support Systems

Prof. P. Krishna Reddy
Head
IT for Agricultural Research Center
International Institute of Information Technology Hyderabad (IIIT-H)
Hyderabad
India
pkreddy@iiit.ac.in

Prof P. Krishna Reddy is the head of IT for Agricultural Research Center at IIIT Hyderabad, India. During 2013 to 2015, he has served as a Program Director, ITRA-Agriculture & Food, Information Technology Research Academy (ITRA), Government of India. From 1997 to 2002, he was a research associate at the Center for Conceptual Information Processing Research, Institute of Industrial Science, University of Tokyo. From 1994 to 1996, he was a faculty member at the Division of Computer Engineering, Netaji Subhas Institute of Technology, Delhi. During the summer of 2003, he was a visiting researcher at Institute for Software Research International, School of Computer Science, Carnegie Mellon University, Pittsburg, USA. His research areas include data mining, database systems and IT for agriculture. He has published about 130 refereed research papers which include 17 journal papers and three book chapters. He has delivered several invited/panel talks at the reputed conferences and workshops in India and abroad. He was a proceedings chair of COMAD 2008, a workshop chair of KDRS 2010, organizing chair of BDA2015, media and publicity chair of KDD 2015 and a member of sub-Group 3 for IT/ICT in Agricultural Insurance, NITI Aayog. He has organized the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2010) and the Third National Conference on Agro-Informatics and Precision Agriculture 2012 (AIPA 2012). He has got several awards and recognitions. He is a steering committee member of pacific-asia knowledge discovery and data mining conference series since 2010. Since 2004, he has been investigating the building of the eSagu system, which is an IT-based personalized agro-advisory system, to provide scientific agricultural information to farming community. He has also built the eAgromet system which is an IT-based agro-meteorological advisory system to provide risk mitigation information to farmers. He has received two best paper awards. The eSagu system has got several recognitions including CSI-Nihilent e-Governance Project Award in 2006, Manthan Award in 2008 and finalist in the Stockholm Challenge Award 2008.

Abstract

Weather influences almost all occupations. The systems for weather observation and forecast are being operated to deal with adverse weather. For any domain (agriculture, livestock, human health and so on), the weather-based decision support systems (DSSs) provides appropriate suggestions based on the weather condition of the given duration. In the literature, the notion of reuse is being employed to improve the efficiency of DSSs. Normally, in a weather-based DSS, similar suggestion will be provided for a similar weather condition. So there is an opportunity to improve the performance of DSS, if there exists a similarity among the weather conditions. In this talk, I will present a framework to analyze the extent of similarity among the weather conditions to exploit reuse in weather-based agricultural DSSs. Using the proposed framework, we have computed the extent of similarity for weather conditions of each year to the weather conditions of preceding years by considering 30 years of weather data of Rajendranagar, Hyderabad, Telangana State, India. We have also computed the extent of similarity among season-wise weather conditions and phenophase-wise weather conditions related to rice and cotton crops of the given year to the corresponding season-wise and phenophase-wise weather conditions of the preceding years respectively. The data analysis results show that the weather conditions of the given year are about 80 per cent similar to the weather conditions of preceding five years. Also, in the agricultural domain, the season based weather conditions of the given year are similar to about 70 percent of weather conditions of preceding five years and the phenophase-wise weather conditions of the given year for rice and cotton crops are similar to about forty percent to sixty percent of corresponding weather conditions of preceding five years. Overall, the results show that there is a similarity among the weather conditions. So, after developing the content (advisory or suggestions) for a given weather condition, there is a significant scope to reuse the content for the future occurrences of the weather condition. As a result, the efficiency of weather-based agricultural DSSs could be improved.

Statistical Analysis of High-dimensional Metagenomic Data with Applications in Agriculture

Dr. Siddhartha Mandal
Research Scientist
Public Health Foundation of India (PHFI)
New Delhi
India
siddhartha.mandal@phfi.org

Dr. Siddhartha Mandal completed his Bachelors and Masters in Statistics from Presidency College, Kolkata and IIT Kanpur respectively. He obtained his PhD in Biostatistics from University of North Carolina at Chapel Hill. During his postdoctoral tenure at the Norwegian Institute of Public Health, Dr. Mandal worked on human gut microbiome and its associations with infant health outcomes and the effects of environmental toxicants in breastmilk on health. Among his publications is a novel statistical methodology for Illumina sequencing data which provides statistical frameworks that correctly analyzes microbial composition data. In another recent publication, he studied the effects of diet during pregnancy on maternal gut microbiome. Currently he is employed as a research scientist at Public Health Foundation of India working on antibiotic resistance and effects of air pollution on cardiometabolic health, alongside ongoing microbiome projects.

Abstract

Microbial communities are significant components of our ecosystem and play numerous roles, both beneficial and harmful. Specifically in agriculture, both the plant and soil microbiome can be important modulators of quality and quantity of production, plant diseases, reducing the uptake of poisonous chemicals, control of greenhouse gases etc. Similar to the human scenario, the microbiome represents a parallel ecosystem alongside the plant genome thereby producing interactions between the host and the the microbes. Modern culture independent technologies involving next generation sequencing have opened new horizons to study a broad spectrum of microbial species and their functions. PCR amplification of the 16S rRNA allows large scale identification of microbial species, while global analyses such as metagenomics, metatranscriptomics and meta-proteomics provide an insight into the functional potential, offer a snapshot of the community-wide gene expression and protein abundances of the microbial communities. In this talk we shall explore the various approaches to study the microbiome, the data generated by these technologies and pipelines that can be used to extract and process information from these large datasets. In addition, we shall take a look at some examples to explain how these datasets can be used to arrive at biologically significant conclusions. Finally, we shall explore areas with agricultural applications where such techniques can be applied in the Indian context.

Session VI
Data Management

Convener and Co-Convener

Dr. Medha Devare
Data and Knowledge Manager
CGIAR System Office
France

Dr. Sudeep Marwaha
Principal Scientist
Division of Computer Applications
Indian Agricultural Statistics Research Institute
(IASRI)
New Delhi
India

Enhancing Discoverability and Re-use of CGIAR's Agricultural Data: Challenges and Progress

Dr. Medha Devare
Data and Knowledge Manager
CGIAR System Office
France
m.devare@cgiar.org

Dr. Medha Devare is the Data and Knowledge Manager at the CGIAR System Office. She is a scientist with expertise in agronomy, microbial ecology, and bioinformatics tools, but is equally comfortable in the data and knowledge management arena, with experience in the development and application of standard vocabularies, ontologies, and semantic web tools and approaches. Medha works across CGIAR's fifteen Centers and Research Programs to coordinate a consistent approach to Open Access and Open Data, and facilitate management and dissemination of CGIAR research outputs via FAIR principles (Findable, Accessible, Interoperable, Reusable). Prior to working at the System Office, Dr. Medha was a Scientist and Knowledge Management Specialist at the International Maize and Wheat Improvement Center (CIMMYT) in Kathmandu, Nepal, where she led the USAID-funded initiative "Cereal Systems Intensification for South Asia in Nepal".

Abstract

CGIAR's 15 Centers and other entities involved in agricultural research and development are charged with tackling complex challenges at a variety of scales, but research outputs are too often not easily discoverable or reusable. CGIAR is attempting to enhance discovery and reuse of its data for further analysis and visualization through the Open Access and Open Data initiative and the Big Data Platform to harness the power of big data and ICTs. The contents of most repositories at CGIAR Centers are not generally easily discoverable or inter-linked (e.g., agronomic trial data with socioeconomic or adoption data in the same geography). In the absence of such interoperability-mediated integration, "open" is of limited utility. OA/OD efforts focus on creating a culture of data sharing, and provide support for harmonized data through the development of common metadata, ontologies, and strengthened collaboration and coordination around tools and approaches. This foundational work aims to render CGIAR outputs interoperable, ensuring they are discoverable via integrated and contextualized views across Centers and programs, type (e.g., publications, data, etc.), and discipline (e.g., genetic/genomic; agronomy; breeding; socioeconomic, and other sectors). The overall objective, then, is to manage CGIAR's trove of research data and associated information, making it more accessible for indexing and interlinking and ensuring that research outputs are Findable, Accessible, Interoperable and Re-usable (FAIR) for simulation, analytical, and visualization tools to enhance innovation and impact. This presentation will review the challenges to sharing and mining CGIAR data effectively, and the progress so far towards addressing these.

Managing geospatial database of harmonized, multi-disciplinary agricultural indicators for big data analytics in Africa South of the Sahara

Dr. Jawoo Koo
Senior Research Fellow
International Food Policy Research Institute (IFPRI)
Washington-DC
USA
J.Koo@cgiar.org

Dr. Jawoo Koo is a Senior Research Fellow at International Food Policy Research Institute (IFPRI), Washington, DC, working on the modeling of crop production systems in developing world in the Environment and Production Technology Division. Jawoo leads the Spatial Data and Analytics Theme. Collaborating with various research and development partners in Africa, Jawoo is working to design and develop Technology Platform, which will deliver up-to-date information on agricultural technologies to agricultural research managers and policy/decision makers. Jawoo also serves as the Global Coordinator of CGIAR-CSI, a Community of Practice of geospatial scientists in CGIAR centers.

Abstract

Recent progress in large-scale georeferenced data collection is widening opportunities for combining multi-disciplinary datasets from biophysical to socioeconomic domains, advancing our analytical and modeling capacity. Granular spatial datasets provide critical information necessary for decision makers to identify target areas, assess baseline conditions, prioritize investment options, set goals and targets and monitor impacts. However, key challenges in reconciling data across themes, scales and borders restrict our capacity to produce global and regional maps and time series. This presentation provides overview, structure and coverage of CELL5M—an open-access database of geospatial indicators at 5 arc-minute grid resolution—and introduces a range of analytical applications and case-uses in CGIAR Centers and partner institutes. CELL5M covers a wide set of agriculture-relevant domains for all countries in Africa South of the Sahara and supports our understanding of multi-dimensional spatial variability inherent in farming landscapes throughout the region.

Cassavabase – a web-based enterprise-level breeding database for Cassava and other crops

Dr. Lukas Mueller
Associate Professor
Cornell University
Ithaca
USA
lam87@cornell.edu

Dr. Lukas Mueller is a biochemist and bioinformatician who has been involved with large plant genome databases for many years. Initially working at Arabidopsis database TAIR, he has been directing the SGN database for more than a decade. His group is mainly interested in describing plant genomes and phenomes, and how the two can be linked to help scientists and breeders. He has also been involved in the International Tomato Genome sequencing project and the larger initiative to characterize the Solanaceae, called the SOL project. He is delighted to be part of the Next Generation Cassava Breeding project because it has the potential to improve the lives of so many people.

Abstract

Cassava (Manihot esculenta) is the main source of calories for 250 million people in Africa and therefore indispensable for food security. Cassavabase (https://www.cassavabase.org/) is the central datastore for the NextGen Cassava Project, aiming to leverage cassava genomic breeding and accelerate variety development in Africa (http:// www.nextgencassava.org/) . Cassavabase aims to provide a complete bioinformatic toolset to the cassava breeding and research community. From the field to the lab to the office, it offers integrated solutions for tracking of information, streamline management of genotypic and phenotypic data, and genomic selection analyses. As Cassavabase is web-based, it is a globally accessible resource, enhancing data sharing and communication within cassava research community in Africa and overseas. Close interaction with the user community and Cassavabase staff drives new development and improvements. Current tools cover phenotyping data (trait ontology dictionary, fieldbook app), sequencing data (GBS pipeline, Jbrowse), breeding management (breeding program, trial, trait, pedigree search) and molecular breeding (trait descriptive statistics, genetic maps, genomic selection). Since its inception, 9 million phenotypic datapoints and 18,000 genotyped accessions have been stored from 500+ users across half dozen breeding programs. The system is open source and available to the community (https://github.com/solgenomics/HYPERLINK "https://github.com/solgenomics/)"). The Cassavabase software is species independent and can be applied to any breeding program of any crop.

Title

Dr. Rodrigo Sara
Senior Legal Officer
CGIAR System Office
France
R.Sara@cgiar.org

Dr. Rodrigo Sara works is legal Officer at the CGIAR System Office with expertise in intellectual asset management, genetic resources, PPPs and compliance related issues. His support across CGIAR's fifteen Centers and Research Programs helps ensure a consistent approach in resolving legal issues that arise in the operationalization of Open Access and Open Data via FAIR principles (Findable, Accessible, Interoperable, Reusable), including in relation to ethics and confidentiality considerations.

Planteome, a resource for reference ontologies and applications for plant biology

Dr. Marie-Angélique Laporte
Post-Doctoral Fellow
Bioversity International France
marieangelique.laporte@gmail.com

Dr. Marie-Angélique Laporte is a Bioversity International post-doctoral fellow working on the NFS funded project Planteome, which aims to develop a set of reference ontologies for Plant as well as common annotation standards, and the Crop Ontology project. Her research interests include metadata standards and ontology engineering aiming at the agricultural data representation, integration and analysis on the semantic web.

Abstract

The Planteome project is a centralized online plant informatics portal which provides semantic integration of widely diverse datasets with the goal of plant improvement. Traditional plant breeding methods for crop improvement may be combined with next-generation analysis methods and automated scoring of traits and phenotypes to develop improved varieties. The Planteome project (www.planteome.org) develops and hosts a suite of reference ontologies for plants associated with a growing corpus of genomics data. Data annotations linking phenotypes and germplasm to genomics resources are achieved by data transformation and mapping species-specific controlled vocabularies to the reference ontologies. Analysis and annotation tools are being developed to facilitate studies of plant traits, phenotypes, diseases, gene function and expression and genetic diversity data across a wide range of plant species. The project database and the online resources provide researchers tools to search and browse and access remotely via APIs for semantic integration in annotation tools and data repositories providing resources for plant biology, breeding, genomics and genetics.

Session VII
Genomic Selection & Genome-Wide Association Mapping

Convener and Co-Convener

Dr. Dietrich Borchardt
Sugarbeet Breeder
KWS SAAT SE
Einbeck
Germany

Dr. A R Rao
Principal Scientist
Centre for Agricultural Bioinformatics
Indian Agricultural Statistics Research Institute
(IASRI)
New Delhi
India

Quantitative Trait Loci – From detection to application in breeding

Dr. Dietrich Borchardt
Dr. Dietrich Borchardt
Sugarbeet Breeder
KWS SAAT SE
Einbeck
Germany
dietrich.borchardt@kws.com

Dr. Dietrich Borchardt is working as Sugarbeet Breeder at KWS in Einbeck, Germany, and specialized in applied molecular breeding. Dr. Dietrich studied Agriculture in Kiel and Plant Breeding in Stuttgart Hohenheim. Since his PhD in 1994 on selection theory and optimizing breeding schemes, his main interest is to translate modern molecular techniques into breeding application. A strong technical development in combination with statistical tools in trait mapping strongly increased the power and efficiency of plant breeding.

Abstract

The breeding process is focused mainly on yield, more generally yield components as quantitative traits. In addition complementary traits like disease resistance are required; inheritance is mostly less complex. Several methods of trait mapping have been described, mainly QTL mapping in structured populations and genome-wide association mapping (GWAS). The main principle is to combine phenotypic and genotypic information in order to detect regions of interest for the respective trait. Selective Sweep mapping as an approach using only genotypic data will be described. It is applicable in breeding populations that undergo several cycles of strong phenotypic selection. Application of marker assisted selection has been limited in breeding programs for quantitative traits. In this regard technical developments and cost reduction in high-throughput genotyping have been driving factors for applications of genome-wide prediction. Depending on trait architecture different scenarios to integrate prediction approaches are possible.

Genome-based prediction in KWS breeding programs

Dr. Valentin Wimmer
Central Statistic Unit
KWS R&D
Einbeck
Germany
Valentin.Wimmer@kws.com

Dr. Valentin Wimmer is working within the central statistic unit for KWS R&D in Einbeck, Germany. His working areas include marker statistic applications (genomic prediction, genome-wide association studies, QTL mapping) as well as field trial analysis. He received B.Sc. and M.Sc. degrees in statistics from the University of Munich and a Ph.D. in plant breeding from the Technische Universität München. His dissertation was concerned with the efficiency of statistical methods for genome-based prediction. During his PhD, he designed and contributed to the synbreed R package providing a framework for the analysis of genomic prediction data.

Abstract

Genome-based prediction is an attractive method for plant breeding programs. Cycle length can be reduced and cost-intensive phenotyping can be alleviated by genotyping a large number of selection candidates. A training set with genotyped and phenotyped individuals is used to construct a genome-wide regression model. From this, predictions of genetic values of genotyped only selection candidates are derived. Suitable trainings sets will be characterized in terms of marker density, training population size, and genetic relatedness to the prediction set. Choice of appropriate statistical models will be discussed. For traits influenced by non-additive effects, it will be shown how models incorporating epistatic interactions can enhance prediction performance. The implementation of these technologies within KWS breeding programs will be illustrated.

Exploring association mapping approaches in a multi-year, multi-location, multi-pathogen experiment of wheat disease resistance in NW Europe

Dr. Keith Gardner
Statistical Geneticist
NIAB
United Kingdom
Keith.Gardner@niab.com

Dr Keith Gardner is a statistical geneticist at NIAB in the UK. His work focuses on the genetic analysis of complex data sets in crops, including QTL mapping in multi-parent populations, association genetics and genomic selection. He is also involved in teaching quantitative and population genetics, as they apply to plant breeding. His PhD and his previous research experience were in evolutionary genetics, where he used large genetic datasets to address natural history and conservation questions.

Abstract

Evaluation of resistance to multiple diseases over multiple sites and years using the same breeding pool is a powerful method for identification of resistance loci as well as to potentially obtain understanding of the interactions between loci, between loci and environments and between diseases. This total information can then be used to improve the way in which markers for resistance can be incorporated into breeding programs. Furthermore, such a data set provides an opportunity for comparing and experimenting with different analytical approaches. To identify new sources of disease resistance in NW European wheat (*Triticum aestivum* L.), we performed a genome-wide association study (GWAS) using a collection of 500 UK and related NW European wheat varieties, genotyped with >26000 variable SNPs. In a series of more than 30 field trials over 3 years in 12 locations across 5 countries, adult plants were evaluated for resistance to four of the most important fungal diseases of wheat in NW Europe: yellow rust (caused by Puccinia striiformis f. sp. tritici), brown rust (*Puccinia triticina*) septoria tritici blotch (*Mycosphaerella graminicola*) and powdery mildew (*Blumeria graminis*). Using a standard mixed linear model GWAS approach, between 20 and 78 unique resistance loci per disease were identified with a false discovery rate (fdr) threshold of 0.05. For stripe rust (SR), a distinctive temporal shift in the pattern of hits was evident, correlated with dramatic race changes in the pathogen population over the course of the field trials. Almost all the most breeder-useful SR marker-trait associations were subsequently validated in a selection of 20 breeder's biparental crosses; in addition several "borderline significant" marker-trait associations were also validated in this way. We explored several approaches for detecting interactions between detected hits and other unlinked markers, with different approaches largely confirming the same patterns. In this specific breeding pool, many of these putative interactions were of the "suppressor" type – presence of a minor frequency "suppressor" allele at one locus reduced the effectiveness of a resistance allele at another locus. We then investigated patterns of cross-disease resistance across the data set, including attempting to detect loci with pleiotropic disease affects. Finally, as we also have similar phenotypic results from our NIAB 8-parent wheat multiparent advanced generation inter cross (MAGIC) population, which was genotyped with the same array, and contains >80% of the marker variation observed in the GWAS panel, we tried to see if we could predict from our GWAS results to results from the MAGIC population and vice versa, with not very successful, but nevertheless interesting, results.

Application and optimization of genomic prediction models for barley breeding

Dr. Patrick Thorwarth
Research Scholar Scientist
University of Hohenheim
Stuttgart
Germany
Patrick.Thorwarth@uni-hohenheim.de

Dr. Patrick Thorwarth studied Crop Sciences with specialization in Plant Breeding and Seed Sciences at the University of Hohenheim. After obtaining his M.Sc. degree, he started his work as a PhD student in the group Crop Biodiversity and Breeding Informatics led by Prof. Karl Schmid at the University of Hohenheim. His research focus on the application and optimization of molecular breeding methods such as Genome wide association studies and Genomic Selection using empirical and simulated data, with an emphasis on self-fertilizing plants such as barley (*Hordeum vulgare* L.).

Abstract

To ensure breeding progress under changing environmental conditions the implementation and evaluation of new breeding methods is of crucial importance. In the framework of the BARSELECT project, we assessed the potential of genomic prediction in a training population of 750 individuals, consisting of multiple six-rowed winter barley (Hordeum vulgare L.) elite material families and old cultivars which reflect the breeding history of winter barley of the last 30 years in Germany. Selected crosses of parents out of the training set were used to create a validation set of double haploid families of 750 individuals in total, which was used to confirm the results obtained from a cross-validation in the training set using 11 different models for genomic prediction. A crucial component for genomic prediction is the available number of markers. To test if the marker spacing in this barley population is sufficient linkage disequilibrium and the persistence of the linkage phase between families was assessed. We can show that even though the number of markers is not sufficient for long term selection, relatively high prediction abilities ranging from 0.31 to 0.71, for yield related traits can be obtained. A further question we try to answer is the optimal composition of a training population and if a minimal marker set can be derived from a large set of markers, which can help to increase the prediction ability and support the biological interpretation of results obtained using genomic prediction.

GWAS and genomic prediction across multiple breeding cycles

Dr. Christina Lehermeire
Research Scientist
Technical University of Munich
Germany
christina.lehermeier@tum.de

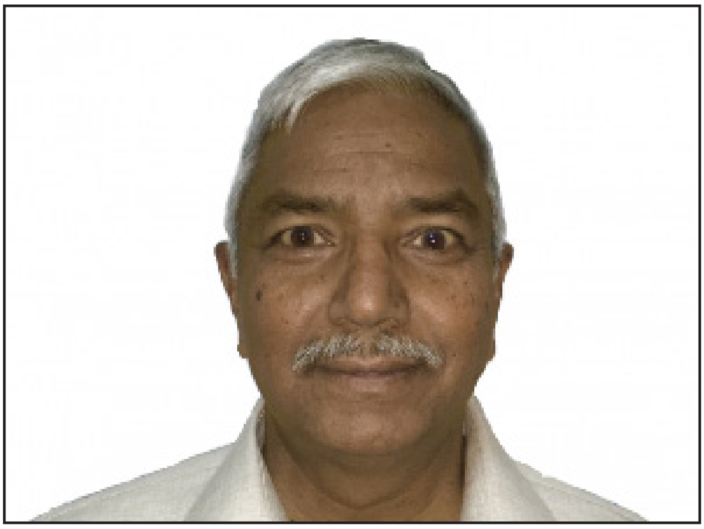
Dr. Christina Lehermeier studied statistics at University of Munich, where she graduated with a M.Sc. degree in 2011. She is now working as a research scientist at the plant breeding group of the Technical University of Munich, where she finished her Ph.D. in 2015. Her research involves investigation of genomic prediction in structured plant populations, optimization of designs for genomic selection, as well as the development and application of statistical methods for identification of QTL and prediction of quantitative traits.

Abstract

Hybrid breeding is expected to gain substantially from the use of genome-based selection, where lines are selected based on predicted testcross performance using high-density marker profiles. Mainly two prediction scenarios arise: i) prediction of the genetic value of individuals from the same breeding cycle in which model training is performed and ii) prediction of individuals from subsequent cycles. It is the latter from which a reduction in cycle length and consequently a strong impact on selection gain is expected. Based on experimental data from multiple selection cycles of hybrid breeding programs in maize and rye, we investigate the efficiency of genome-based prediction for agronomic traits within and across selection cycles. Factors driving the prediction accuracy besides ancestral relatedness will be analyzed. Further we discuss the effect of integrating knowledge on large and stable QTL discovered by GWAS for improving prediction accuracy across cycles.

Session VIII
Bayesian Statistics

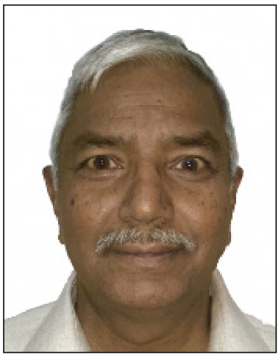
Convener and Co-Convener



Dr. Murari Singh
Senior Biometrician
ICARDA
Amman
Jordan

Dr. Himadri Ghosh
Principal Scientist
Division of Statistical Genetics
Indian Agricultural Statistics Research Institute
(IASRI)
New Delhi
India

A Bayesian analysis of data from on-farm trials in legumes in Afghanistan



Dr. Murari Singh
Senior Biometrician
ICARDA
Amman
Jordan
m.singh@cgiar.org

Dr. Murari Singh is the Senior Biometrician and Executive Assistant to the Deputy Director General – Research at the International Center for Agricultural Research for Dry Areas (ICARDA). He holds a Ph.D. in Agricultural Statistics, with a Major in Design of Experiments and Minors in Genetic Statistics and Agricultural Economics, from the Indian Agricultural Research Institute, New Delhi, India. He is a Professional Statistician accredited by Statistical Society of Canada. Over the past 38 years, he has served in various capacities, including assistant Professor, Associate Professor, Scientist, Statistician and Senior Biometrician – at three Indian institutions, three North American Universities and two CGIAR Centers. Dr. Singh has been rendering statistical consultancy in the areas of agro-biological, environmental and socio-economic research, and has developed effective biometrics and statistics services through the implementation of advanced applied biometrical and statistical techniques.

Abstract

On-farm demonstration (OFD) are conducted to provide a real-time comparison of the recommended technology and the farmer practice in farmer's field, and are effective and easy way to convince the farmers on the potential advantages of adopting new technologies or "seeing is believing". Since the OFDs are routinely conducted under a crop improvement program and their data analyzed, such OFDs can provide prior information on various parameters of the model for analyzing a current data. In the present study, data from 2009-2011 were analyzed and the estimates provided the prior information. While such information remain unutilized in a frequentist approach, they can be used under a Bayesian approach. A Bayesian approach is discussed for analysis of such demonstrations for evaluating the improved varieties of chickpea (Sehat and Madad) and mung bean (Mai-2008 and Maash-2008) in selected provinces of Afghanistan in 2012. Posterior means and risks to meet the target productivity levels were compared for the improved packages and the farmer practice and found the former practices superior.

Improving Decisions in Plant Breeding with a Bayesian Approach

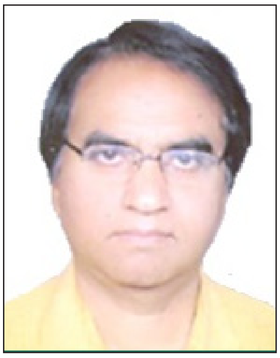
Dr. Jode Warren Edwards
Research Geneticist
U.S. Department of Agriculture-Agricultural
Research Service
USA
Jode.Edwards@ARS.USDA.GOV

Dr. Jode Edwards is a Research Geneticist with the U.S. Department of Agriculture-Agricultural Research Service. Dr. Edwards' research is in the area of Maize Quantitative Genetics with primary efforts in the areas of maize plant density response, genotype by environment interaction, and application of Bayesian Statistics to analysis of multi-environment plant breeding trials. Dr. Edwards' work in Bayesian statistics is focused on using the Bayes paradigm to obtain optimal estimators and comparisons in data sets with heterogeneous data quality, heterogeneous stability of cultivars, and with unbalanced replication. He has done primary research on Bayes approaches in addition to regularly applying Bayesian approaches to his own research on plant density response and genotype by environment interaction.

Abstract

Stability can be highly variable among crop varieties and breeding lines. Although numerous methods exist to quantify variation in stability, most methods do not integrate variation in stability with the estimate of mean performance. The objective of this approach was to formally model variation in error and genotype by environment variance in analysis of multi-environment variety trials. In example data sets in maize and oat, substantial variation among environments and varieties was found in precision of performance estimators. By formally modeling the variation in error and genotype by environment variance using a Bayesian approach, optimal performance estimators were obtained such that varieties with more variable performance across environments (less stable) had statistically optimal penalty (via shrinkage estimation) applied to their performance estimators. A simulation study suggested that based on typical levels of heterogeneity in genotype by environment variance in maize data in the central corn belt in the United States of America, selection response could be increased by 15% by modeling the variation in stability. Bayesian modeling of heterogeneity in variances thus had direct impact on the estimators of performance, in particular, when rank order of means is important for decision making as in plant breeding.

Bayesian Modelling: A Few Related Issues



Dr. Satyanshu K Upadhyay
Professor
Banaras Hindu University
Varanasi
India
skupadhyay@gmail.com

Dr. Satyanshu K. Upadhyay is a Professor at the Department of Statistics, Banaras Hindu University. He is also linked with DST Centre for Interdisciplinary Mathematical Sciences at Banaras Hindu University as a core group researcher. He obtained his Ph.D. degree in 1984 from Banaras Hindu University and became Professor of Statistics in the year 2005. Dr. Upadhyay received a number of distinctions in his career, and worked as EEC research bursary at Imperial College, London, during the year 1992-93 with Adrian Smith. He is associated with a number of academic bodies and also the fellow of the Royal Statistical Society. His research interests include preliminary test procedures, life testing and reliability, biostatistics and Bayes inferences. He has published several research papers, books, and other academic materials in his career.

Abstract

Bayesian modelling is nothing but the combination of information in the form of prior and likelihood joined together using Bayes theorem. This results in to what we call the posterior distribution or, going a step ahead, the predictive distribution. The posterior and the predictive distributions form the basis of Bayesian inferences although drawing the inferences are computationally challenging unless one is enriched with the sophisticated tools of Bayesian computation. Moreover, the appropriateness of the final inferences does depend on the appropriateness of the chosen models and, therefore, judging and choosing the appropriate model is also an important aspect in any posterior or predictive based inferences. The present talk focuses on some such issues using the tools of Bayesian paradigm.

Applications of Bayesian Computing in Modeling of Agricultural Data

Dr. Athar Khan
Professor
Aligrah Muslim University
Aligarh
India
atharkhan1962@gmail.com

Dr. Athar Khan is currently working as Professor Statistics, Dept. of Statistics and Operations Research Aligrah Muslim University, Aligarh, India. Worked as Principal Investigator in project, "Bayesian Modeling of Soil Nutrients in Kashmir", worth Rs. 14.23 lakhs funded by the ICAR in Competitive Grant II during 2001-2003 during stay at S K University of Agricultural Sciences of Kashmir, Srinagar India. Developed a software for data analysis in 2004. Published around 48 research papers in national and international journals. His current area of research is Bayesian Statistics and a member of advisory committee of Indian Bayesian Society and Program in Chair of Indian Chapter of International Society for Bayesian Analysis (ISBA), Duke University Durham, USA. Receiving invitations from several Universities to deliver lectures in workshops and conferences on Applications of R software package for the last many years. Guided many students in the field of Bayesian Computing. R and JAGS are his favorite.

Abstract

Due to availability of computing languages like R and JAGS, implementation of Bayesian approach of data analysis could be used in modeling agricultural data. In this talk, an attempt has been made to use both analytical and simulation tools to model the data generated from designed agricultural experiments. Bayesian parallel of linear and generalized linear models are attempted for modeling the data generated from designed experiments. Simple as well as advanced designs are considered and data is taken from standard books and published papers. Graphic analyses are given equal importance. Simulation tools like MCMC and analytical tools like Laplace Approximation are used extensively throughout the document. R and JAGS are used for implementations.

Session IX
Geostatistics and Remote Sensing

Convener and Co-Convener

Dr. Chandrasekhar Biradar
Principal Scientist
Head-Geoinformatics Unit
ICARDA
Amman
Jordan

Dr. Tauqueer Ahmad
Head, Division of Sample Surveys
Indian Agricultural Statistics Research Institute
New Delhi, India

Analysis of Landsat-based Water Use Estimates (1984-2014) in the Southwestern United States: An Illustration of Big Data Analytics for Agricultural Remote Sensing

Dr. Naga Manohar Velpuri
Senior Scientist
ASRC Federal
USA
manohar.velpuri.ctr@usgs.gov

Dr. Naga Manohar Velpuri is a Senior Scientist for ASRC Federal, contractor to the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center and adjunct professor with South Dakota State University, USA. He earned his PhD for his research on the operational use of remote sensing and hydrologic modeling techniques for monitoring surface water bodies in the remote pastoral regions of Africa. His research interests are focused on the overlapping areas of agriculture, surface hydrology, remote sensing, and climate change. Current research projects broadly cover understanding (a) multi-scale watershed hydrologic processes (b) variability in surface water storage using multi-source satellite data (c) climate and human impacts on water resources availability (d) analysis of water availability and use in agriculture and (e) the use of big data analytics in remote sensing hydrology. Earlier he worked in several national and international organizations (SDSU, UNEP, IWMI, University of Cambridge, INCOIS, and NIRD) in various capacities. He has more than 40 publications in the form of peer-reviewed international journals, book chapters and research reports.

Abstract

Remote sensing-based field scale evapotranspiration (ET) maps are useful for characterizing water use patterns and assessing crop performance. However until recently, modeling evapotranspiration data using Landsat has been limited to small spatial and temporal scales due to the limitations in model parametrization and large data volume. The new Operational Simplified Surface Energy balance (SSEBop) model that integrates weather and remotely sensed images eliminates the need for selection of hot and cold anchor pixels for each imagery. Coupled with the improvements in high performance computing and SSEBop model, it is now possible to produce seamless estimates of evapotranspiration over large spatial and temporal scales. In this study, a total of 3,396 Landsat images (with nearly 36 million pixels per band) were processed to produce historical (1984-2014) Landsat-based ET maps for nine major irrigation districts in the southwestern US. Model output evaluation and validation using point-based eddy covariance flux tower, gridded-flux data and water balance ET approaches indicated relatively strong association between SSEBop ET and validation datasets. Results from the Seasonal Mann-Kendall trend analysis of 7 agro-hydrologic variables grouped into two categories (a) climate influenced (atmospheric water demand, ETo; air temperature, Ta; and precipitation, P) and (b) management influenced variables (actual evapotranspiration, ETa; land surface temperature, Ts; and basin runoff, Q) are presented. Our results indicate that while all climate variables showed positive trends, only air temperature showed a consistent and significant increase (up to 1.2 K) across all 9 irrigation districts/sub-basins during the 31 years. In a pair wise comparison, management influenced variables were found to be more variable than their corresponding climate counterparts, responding to the impacts of management decisions. District-wide ETa estimates were used to compute historical crop water use volumes and monetary savings for the Palo Verde Irrigation district (PVID). During the peak crop fallowing program in PVID, the water savings reached a maximum of ~ 85,000 ac-ft which is equivalent to a dollar amount of ~$600 million. This study demonstrate the use of big data analytics and has many applications in planning water resource allocation, managing water rights, sustaining agricultural production, and quantifying impacts of climate and land use/land cover changes on water resources. With increased computational efficiency and model development, similar big data analytical studies can be conducted in other parts of the world.

Image Stitching: Reliable base layer for field level studies

Dr. Surender Varma Gadhiraju
Research Scientist
Dupont Pioneer
Hyderabad
India
surender.gadhiraju@pioneer.com

Dr. Surender Varma Gadhiraju is a Research Scientist at Dupont Pioneer, Dupont Knowledge Centre, Hyderabad. He is a researcher with expertise in image processing, geoinformatics and machine learning. Formerly, he is associated with Indian Institute of Technology, Bombay as research scholar and research associate. His current interests span IOT (Internet of Things) solutions for agriculture, precision farming, big data analytics and cloud computing. He has developed novel methodologies to detect changes from time series imagery.

Abstract

UAS offer relatively cheap, flexible, high resolution imagery becoming the platform of choice for field based high throughput phenotyping. Image stitching transforms hundreds of raw images in to one large orthomosaic covering whole field in a single view. Consistent high quality orthomosaics are difficult to achieve due to complications in image matching, blending and generation of uniform 3D point cloud that may adversely affect downstream analysis tasks. This talk presents few agricultural applications of UAS imagery and highlights key challenges of image stitching, bringing them to the attention of research community.

Title

Dr. Mukunda Dev Behera
Associate Professor, Indian Institute of
Technology Kharagpur, India
mukundbehera@gmail.com

Dr. Mukunda Dev Behera is an Associate Professor of Ecological Remote Sensing at the Centre for Oceans, Rivers, Atmosphere and Land Sciences (CORAL) in Indian Institute of Technology Kharagpur. With a PhD in Botany from Indian Institute of Remote Sensing (ISRO) (GK University, Hardwar) and MS degree in Advanced Geo-information from University of Paris VI (France), Dr. Behera has been working in ecosystem ecology and climate, biodiversity, forest biomass and carbon sequestration, land/ vegetation cover dynamics studies using Geoinformation and modelling techniques. Dr. Behera has been an active member of an Indian national team on 'Biodiversity characterization at landscape level using remote sensing and GIS' and was very active in bringing out the maiden forest vegetation type map of India using satellite data at 1:50K scale that has been validated with >15,000 ground data points. During this study, he has provided validation protocols to several landscape ecological parameters such as fragmentation, juxtaposition and biological richness index. He has analysed the (i) plant diversity along India`s latitudinal and longitudinal gradients and brought out the maiden pattern, (ii) plant dispersal pattern across India using species commonality as an index, (iii) forest resilience pattern in India with respect to trees, shrubs and grasslands in lieu of climate change and disturbance, (iv) 'Plant diversity and environmental heterogeneity' hypothesis evaluation for India, (v) transformation of vegetation type-wise field sampling sufficiency on to a grid scale. He has published more than 50-peer reviewed articles in scientific journals and guest edited issues of 'Biodiversity and Conservation' and 'Current Science' journal.

Space Technology based Governance of Watershed Management: Perspectives and Issues in Indian Context

Dr. Girish S Pujar
Scientist
NRSC-ISRO
Hyderabad
India
pujargs@gmail.com

Dr. Girish Pujar is currently working as Scientist with NRSC-ISRO, Hyderabad. His research focus is in the field of remote sensing applications especially for Watershed monitoring and rural development. His Interests include trees outside forests, OBIA, dynamic vegetation models, terrestrial laser scanner applications apart from applications using IRS datasets.

Abstract

Impending trends in changing global patterns of water availability and use, have made it imperative for developing countries to move towards holistic natural resource management approach. Government of India through Department of Land Resources, Ministry of Rural Development is actively pursuing restoration of ecological balance in rural landscapes through national initiative of Integrated Watershed Management Programme (IWMP), now brought under PMKSY Pradhan Mantri Krishi Sinchayee Yojana. Approach involves principles of soil and water conservation through participatory approach at national level. In this effort, National Remote Sensing Centre, ISRO is supporting the effort through setting up a web GIS through open source Bhuvan interface. Interface, named Srishti involving a high resolution image based service coupled with smart phone based application called Drishti is enabling monitoring of watershed projects implemented since 2009-10 till 2014-15 for all the 8000 plus micro- watershed projects across the country. Effort relies on sequentially catering to the requirements of monitoring and evaluation of project implementation in varied agroclimatic and terrain contexts of India, by setting up each component of the process balancing the stakes of various institutions involved. Every state level nodal agency under the directives of Central ministry is collaborating through this platform and has been successfully uploading point of interest data for every activity done. Till date upto 4.45 Lakh POIs with precise geolocations have been uploaded consisting of a pair best resolution photos of the field asset created as well the attributes explaining the entity, rendered against Orthorectified natural color IRS high resolution data. Process throws in enough challenges in automated data analysis, especially in terms of quality of the inventory data collected, high resolution image based change detection, clear and unambiguous reporting as well as the day to day tracking of the progress of the data compilation being accomplished. As part of the first stage effort, more than 1200 high resolution merged images have been uploaded for the first year of observation (sites corresponding to 2009-10 implementation). Initial semi-automated image processing efforts have been implemented for 2.5 m color images, on experimental basis to harness the geolocation as seeds to delineate the change contexts using segmentation based routines. Results showed that for contrasting material such as water surface and structures detection has been quite reliable. Segmentation facilitates unambiguous segregation of material with respect to the noisy neighbors typical of semi-arid contexts of degraded land cover elements.

Session X
Statistical Modelling and Forecasting

Convener and Co-Convener

Dr. Peter Carberry
Deputy Director General-Research
ICRISAT
India

Dr. K N Singh
Head, Forecasting and Agriculture Systems Modeling, Indian Agricultural Statistics Research Institute (IASRI)
New Delhi
India

Some Reflections on Systems Modelling and Forecasting

Dr. Peter Carberry
Deputy Director General-Research
ICRISAT
India
P.Carberry@cgiar.org

Dr. Peter Carberry, an Australian national, is the Deputy Director General- Research, ICRISAT. Dr. Carberry received his PhD in Agriculture from the University of Sydney. Before joining ICRISAT, Peter was Chief Research Scientist and Partnership Leader (CSIRO-DFAT Africa Food Security Initiative), Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. His expertise is in crop physiology and in the development and application of systems models. He is a key developer of the Agricultural Production Systems Simulator (APSIM) cropping systems model. Dr. Peter has held several important scientific positions, including Board Director of the Australian Institute of Agricultural Science and Technology; President, Australian Society of Agronomy; Grains Research & Development Corporation, Senior Fellowship 2007; Board Member, International Crop Science Congress (2004). He is Associate Editor, Food Security journal and member of the Editorial Board of the Journal of Integrative Agriculture. He has 87 journal papers, 29 book chapters, four guest editorships, and has contributed to over a hundred conference papers, newsletters and other publications.

Abstract

In this digital age, the tools of *modelling and forecasting* are moving to the forefront of everyday practice. Although seen as an emergent capability and opportunity in the 21stC, such capability has been around for decades with much learnt experience. Will the rush to capitalise on the ICT revolution draw on such learnings or has there been a paradigm shift into a digital age with new tools, applications and practices? This paper does look back, drawing on the experience in Australia's dryland agriculture, to suggest past lessons are indeed worth contemplating as we look forward with positive opportunities for the application of modelling and forecasting approaches and tools.

In 1990 in Australia, the Agricultural Production Systems Research Unit (APSRU) was established as a partnership born out of competition in the domain of systems analysis and modelling between research agencies in north-east Australia. APSRU's core technology is the APSIM model that not only provided a true systems modelling capacity but also adopted rigorous software engineering practices (Holzworth et al., 2014). Today, APSIM is regarded as the standard for the modelling of crops and cropping systems in Australia (Robertson & Carberry, 2010). APSRU has been a leading participant in the development and application of seasonal climate forecasting systems relevant to Australian industry and internationally (Meinke and Stone, 2005). The Southern Oscillation Index phase system was developed and promoted by APSRU as a management tool throughout Australia. This work on climate prediction has helped make this phenomenon part of everyday language in Australia such that farmers nationally are acutely aware of the implications of SOI trends on their enterprises' potential productivity.

A key, and possibly unique, attribute of APSRU was in being both a developer of decision support systems for use by farmer and industry practitioners as well as their harshest critic (McCown et al., 2002). As a consequence, APSRU has been an active contributor to both the theory and practice of systems research internationally (Carberry et al., 2002). Today, Yield Prophet® offers farmers in Australia access to APSIM on-line as a subscriber service designed with theory and history's learnings in mind (Hochman et al., 2009).

A challenge to the current generation of developers of modelling and forecasting tools is both to continue to innovate within the very active and fast-moving ICT domain but also to be mindful of the likely cognitive limitations to decision support within human activity systems.

Statistical Models Used For Forecasting Of Food Grain Production In India

Dr. K N Singh
Head, Forecasting and Agriculture Systems
Modeling, Indian Agricultural Statistics Research
Institute (IASRI)
New Delhi
India
knsingh@iasri.res.in

Dr. K. N. Singh is Head of Forecasting and Agriculture Systems Modeling division at ICAR-IASRI, New Delhi. He pursued Ph.D degree in Agricultural Statistics from ICAR- Indian agricultural Research Institute, New Delhi in 1991. Recently completed an international collaborative research project under EU-India framework of social research. Developed the first “Soil Test Crop Response Based Fertilizer Recommendation System” for Targeted Yield(s) of Crops, Soil test data generated through soil testing laboratories of the country have been used to prepare soil fertility maps of different states. Developed methodology for developing geo-referenced soil nutrient maps using GPS and GIS. Dr. K. N. Singh has developed models to Estimate the Soil Nutrients at Field Level using Remote Sensing. Currently undertaking research in the area of Forecasting, Modeling, Application of RS, GIS and GPS in natural resource management. Among other engagements, currently handling the major study in the area of forecasting for sustainable agricultural development.

Abstract

In India, about 70 percent of the population is directly or indirectly engaged in agricultural practices for meeting up their daily livelihood. The agriculture sector provides employment to 58.4 percent of the country’s workforce and is the single largest private sector occupation. Several important industries in India obtain their raw materials from the agricultural sector. In brevity, agriculture and its allied sector contribute about 38 per cent if we talk about the total export of the country. In short, agriculture occupies a central place in the Indian economy. In such a country where agriculture is the life blood of the country’s economy and the livelihood of the people the importance of crop yield forecasting well in advance of harvest is undeniable.

In a country like India reliable forecast of crop production before the harvest is important for advance planning, formulation and implementation of a number of policies dealing with food procurement its distribution, pricing structure, import and export decisions and storage and marketing of the agricultural commodities. Weather plays an important role in crop growth. Therefore model based on weather parameters can provide reliable forecast in advance for crop yield. Sometimes weather variability causes the losses in the yield. Weather variable can be used for crop production forecast by making appropriate model. Statistical method is used to forecast the crop yield using weather variables. Crop yield forecast under the prevalent system in India are being issued by the Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi. These forecasts are however subjective in nature, since these are based on the judgment of agricultural officials. Several organizations in India and abroad are also engaged in developing methodology for pre-harvest forecasting of crop yield using different approaches. The statistical technique employed for forecasting purposes should be able to provide objective, consistent and comprehensible forecasts of crop yield with reasonable precisions well in advance before harvest. The main factors affecting crop yield are inputs and weather. Weather affects crop differently during different stages of crop growth. The weather influence on crop yield depends not only upon the magnitude of weather variables but also on the distribution pattern of weather over the crop season. Fisher (1924) suggested a technique based on relatively smaller number of manageable parameters that takes care of entire weather distribution over the crop season, assuming that the effect of change in weather variable in successive weeks would not be abrupt or erratic but an orderly one that follows some mathematical law and assumed that these effects are composed of the terms of a polynomial function of time. Hendricks and Scholl (1943) have modified Fisher's technique. They divided the crop season into *n* weekly intervals and assumed

that a second degree polynomial in week number would be sufficiently flexible to express the relationship. Baier (1973) and Robertson (1974) expressed crop-weather analysis models as the product of two or more factors, each representing the functional relationship between a particular plant response (e.g. yield) and the variations in selected variables at different plant developmental phases. The overall effects, as expressed by the numerical values of the factors modify each other but are not additive as in the case of a multivariate linear regression model. Agrawal *et al.* and Jain *et al.* (1980) modified the model suggested by Hendricks and Scholl by expressing effects of changes in weather variables on yield in the different week as second degree polynomial in respective correlation coefficients between yield and weather variables. Agrawal *et al.* (1986) further modified the model by expressing the effects of changes in weather variables on yield in different weeks as a linear function of respective correlation coefficients between yield and weather variables.

Statistical Methodology Employed:

Suppose we assume that m denotes weeks (w=1(1)m) at which the pre-harvest forecast of the crop yield need to be released. If we use the weekly data on m weeks in p variables, now new weather variables and interaction components can be generated with respect to each of the weather variables using the below mentioned procedure. Forecast model has been developed by considering all the generated variables simultaneously including the time trend (T).

In order to study the individual effect of each weather variables, two new variables from each variable can be generated as follows:

Let X_{iw} be the value of the i^{th} (i=1(1)p) weather variable at the w^{th}week (w=1(1)n), r_{iw} be the simple correlation coefficient between weather variable X_i at the w^{th} week and yield over a period of k years. The generated variables are given by:

$$Z_{\boldsymbol{j}} = \frac{\sum_{w=1}^{n} r_{iw}^{j} x_{iw}}{\sum_{w=1}^{n} r_{iw}^{j}} ; j = 0,1$$

For j=0 we have unweighted generated variable as:

$$Z_{i1} = \frac{\sum_{w=1} r_{iw} X_{iw}}{\sum_{w=1}^{n} r_{iw}}$$

And weighted generated variables as:

$$Z_{i1} = \frac{\sum_{w=1}^{n} r_{iw} X_{iw}}{\sum_{w=1}^{n} r_{iw}}$$

The following model is then fitted to study the effect of individual weather variables.

$$Y = a_0 + a_1 Z_{i0} + a_2 Z_{i1} + cT + \mathcal{E}$$

where, *Y* is yield, *T* is variable expressing time effect, a_0, a_1, a_2 and *c* are constant entities known as the parameters that needs to be evaluated from the model and ε is the error term which is supposed to be distributed with a null expectation and a constant dispersion σ^2. Thus for each of the weather parameter two variable will be obtained which along with the time component and the intercept term.

For generation of state level crop yield forecast, the districts contributing towards 92% production of the state are identified for development of yield forecast models of respective districts. Subsequently, State level crop yield forecast is obtained based on the area weighted District level yield forecasts. The general equation used for generation of crop yield forecast at state level is as follows.

$$Y = \sum_{i=1}^{n} \frac{A_i X_i}{A_i}$$

where A_i and X_i are the area and predicted yield of crop for ith district.

National level crop yield forecast is prepared by the Mahalanabis National Crop Forecasting Centre (MNCFC) based on the state level forecast generated in IMD. The yield forecast at different stages are generated on different dates for different dates for different crops, mutually agreed between IMD and MNCFC.

During the last 4-5 years researchers have worked on improving the yield forecast by making use of weather indices.

A certain modification was done where the yield was detrended by obtaining the parameter estimates of the model and subsequently the detrended yield was used to forecast the yield of the crop using ARIMA model. The proposed method of obtaining Pre-harvest forecasting of yield of crops was compared with the traditional approaches of forecasting and the proposed method was evaluated in terms of criteria's such as goodness of fit of the model which was evaluated using standard statistical measures. It was concluded that the proposed approach is better and more suitable as compared to the traditional approach for forecasting the wheat yield in the five districts of Uttar Pradesh.

Further, in order to select significant weather variable affecting the yield of crop LASSO technique was as well as stepwise regression methodology was employed. Lasso is a regularization technique which reduces the number of predictors in a regression model and identifies important predictors. Lasso is a shrinkage estimator with potentially lower predictive errors than ordinary least squares and it also includes a penalty term that constrains the size of the estimated coefficients. Therefore, it resembles ridge regression. It generates coefficient estimates that are biased to be small. Nevertheless, a lasso estimator can have smaller mean squared error than an ordinary least-squares estimator when you apply it to new data. Unlike ridge regression, as the penalty term increases, lasso sets more coefficients to zero. This means that the lasso estimator is a smaller model, with fewer predictors. As such, lasso is an alternative to stepwise regression and other model selection and dimensionality reduction techniques. The lasso technique solves this regularization problem. For a given value of λ, a nonnegative parameter, `lasso` solves the problem by:

$$\min_{\beta_0 \beta} \left(\frac{1}{2N} \sum_{i=1}^{N} (Y_i - \beta_0 - X_i^T \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

Where

N is the number of observations.

y_i is the response at observation *i*.

x_i is data, a vector of *p* values at observation *i*.

λ is a nonnegative regularization parameter corresponding to one value of Lambda.

The parameters β_0 and β are scalar and p-vector respectively.

As λ increases, the number of nonzero components of β decreases.

It has been shown empirically that the modified models where LASSO was used instead of stepwise regression are performing better compared to the traditional models.

Further, in recent years it has been seen that there is high fluctuations in the yield of crops. These abrupt variations in yield between the years cannot be captured properly by just applying regression models. To this end, GARCH model was employed to the errors of the regression model to capture these sudden fluctuations.

The proposed model is given by

$$Y = a_0 + a_1 Z_{i0} + a_2 Z_{i1} + cT + \mathcal{E}$$

$$\varepsilon_t = \xi_t h_t^{1/2},$$

and

$$h_t = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j},$$

where $\xi_t \sim IID(0,1),\ a_0 > 0, a_i \geq 0,\ i = 1,2,\ldots,q.\ \ b_j \geq 0,\ \ j = 1,2,\ldots,p$

Though, all these models work in isolation, we are trying to develop a model which takes care of all nonlinear characteristics and other situations. And finally if possible the developed model will be admissible among certain class of model.

Statistical Modeling of Meteorological Data and its Applications in Weather Based Crop Insurance Schemes for Banana

Dr. Ramakrishna Lahu Shinde
Professor and Head
Department of Statistics,
North Maharashtra University
Jalgaon
India
rlshinde@nmu.ac.in,
shinde.stat@gmail.com

Dr. Ramakrishna Lahu Shinde is Professor in Department of Statistics, North Maharashtra University. He holds a Ph.D. in Statistics from Sardar Patel University, Vallabh-Vidyanagar, Anand. He has handled many statistical projects and guided several PhD students. His research interests are in Statistical Process Control, Life Insurance, Crop Insurance, Health Statistics, Reliability Theory, Statistics in Clinical Trials, Distribution Theory of Runs, Scans and Patterns and their applications.

(Joint work with Mr. Rajendra N. Chavhan, Ph.D. student)

Abstract

Banana is a popular and important commercial fruit crop grown in tropical and sub-tropical part of world. Jalgaon district of Maharashtra state of India contributes about 3% in the world's banana production. Also, productivity of banana in Jalgaon district is about 3.5 times of world's productivity. Jalgaon is often known as the Banana Capital of India. There are some adverse climatic conditions for banana crop such as high temperature(above 400C) with bright sunshine causes sun scorching, very low temperature (below 80C) causes chilling injury and high wind speed which exceeds 40 km/hr damages the crop. As the banana crop is more likely to be damaged due to such bad weather conditions hence Government of Maharashtra as per the guidelines of Agricultural Insurance Company of India introduced Weather Based Crop Insurance Schemes (WBCIS), WBCIS-1 and WBCIS-2 to cover expected losses of banana cultivators in 2011-12 and 2015-16 respectively.

This paper investigate probability distribution of daily minimum temperature (Tmin) during winter (November to February) based on meteorological data of Jalgaon weather station for the period of agriculture years 1973-74 to 2015-16. We define event A as the event that daily minimum temperature is below 80C for consecutive three or more days during the period of November to February. To model the occurrence of extreme low temperature events, in particular event A, we define stochastic process, Yn taking value 1 if daily minimum temperature less than 80C on nth day during November to February and 0 otherwise. Degenerate sequence of zeros and Markov chains of different orders are fitted to yearly sequences of Yn. The probability of occurrence of the event A is estimated using distributions of number of success runs in the sequence of Markov Bernoulli trials. Distributions of three random variables as the functions of daily minimum temperature are also reported.

Considering the risk factors with adverse climate conditions, we have studied the WBCIS-1 and WBCIS-2 for banana based on weather data for Jalgaon. Loss years are identified as per risk events specified under WBCIS-I and WBCIS-2 for banana and estimated the annual premium per hector. We have also suggested modified WBCIS-3 for banana in which modification is suggested by using Markov chain modeling for Tmin and percentile of probability distribution of monthly mean Tmin and mean Tmax. Concluding table gives comparison of estimated premium rates under WBCIS-1, WBCIS-2 and WBCIS-3 for Jalgaon. Our results may be useful for pricing of weather derivatives in Weather Based Crop Insurance Scheme (WBCIS) for banana.

Models in agriculture: use of mechanistic modeling approaches to accelerate breeding

Dr. Jana Kholova
Scientist
ICRISAT, Hyderabad, India
j.kholova@cgiar.org

Dr. Jana Kholová is currently working as Scientist in ICRISAT working on Plant Biochemistry, Molecular Biology, Interdisciplinary research along crops value-chain. She has done her Ph.D with specialization in Plant Genetics & Physiology in Charles University in Prague, Faculty of Science, The Czech Republic.

Abstract

Use of modelling tools is experiencing a boom across biological disciplines. Models allow us to reconstruct the changes in system by knowing its initial conditions and the processes underlying changes in the system in time. Theoretically, modelling should provide us an enormous power to capture variability in the system in time across wide range of conditions which cannot be covered experimentally otherwise. In the case of agricultural systems, a wide range of modeling tools exist. These agri-system models allow us to reconstruct the crop behavior (and production) in the system based on the knowledge of meteorological and soil parameters, crop management practices and crop characteristics. Obviously, reliable predictions of crop behavior requires an understanding of the mechanistics and quantification of biological processes underlying crop growth and development and its responsiveness to environmental cues; i.e. interactions between crop-environment-management (GxExM).

Since far, there are only few successful examples where agri-system modelling has been well-incorporated into advanced breeding programs and has resulted in acceleration of crop improvement. In these cases, the used modelling tools managed to integrate the multidisciplinary information (met-soil-crop-system) in a mechanistic manner therefore allowed for reliable predictions in situations on which the model has not been calibrated. In the lecture, we will go through examples of "mechanistic" and "statistical" functions used in current models and successful examples where crop model was used to assist the strategic decisions in breeding. Finally, we will review some of the agri-systems modelling activities at ICRISAT and their future projections.

Session XI
Statistical Genetics & Genomics

Convener and Co-Convener

Dr. Rajeev Gupta
Principal Scientist
Theme Leader-Genomics & Trait Discovery
ICRISAT
India

Dr. Lal Mohan Bhar
Head, Division of Statistical Genetics
Indian Agricultural Statistics Research Institute
New Delhi
India

Genomics based Next Generation Crop Improvement Approaches

Dr. Raman Babu
Technology Deployment Lead (Asia-Pacific)
Dupont Pioneer
Hyderabad
India
raman.babu@pioneer.com

Dr. Raman Babu holds a PhD (Genetics) from Indian Agricultural Research Institute, New Delhi. From 1999 to 2005, Babu served as maize breeder with ICAR in Almora and is credited with developing first generation MAS products in maize that have double the amount of lysine and tryptophan, popularly referred to as Quality Protein Maize. Babu was awarded a post-doctoral fellowship by Japanese Science Promotion Society (JSPS) in 2006 to pursue research on positional cloning of disease lesion mimic genes that offer broad spectrum resistance to blast pathogen in rice. Babu joined CIMMYT in 2008 and steered the maize molecular breeding activities in the global maize program for improving abiotic stress tolerance and nutritional quality traits in the tropical germplasm. Since 2016, Babu is with Dupont Pioneer as Technology Deployment Lead (Asia-Pacific) overseeing molecular breeding, doubled haploids, precision phenotyping and informatics components across crops.

Abstract

High density genotyping platforms, high throughput phenotyping techniques, doubled haploids and breeding informatics form the core of modern plant breeding programs nowadays. Integrating these modern tools in time and resource efficient manner is critical to accelerating genetic gains especially in the regions that are experiencing intense and adverse climate change effects such as rising temperatures, frequent droughts, excess rainfall as well as evolving biotic stresses. Both specific marker- based approaches such as QTL mapping and genome-wide association studies (GWAS) as well as whole genome-based prediction strategies offer unprecedented opportunities for improvement of simple as well as complex traits under challenging array of growing environments. Genome editing is another new tool that promises great potential and opens up a lot of newer avenues. Here, the recent developments with respect to various molecular breeding strategies and tools and their applicability to plant breeding programs with few illustrations from tropical corn are reviewed. Various challenges and infrastructural bottlenecks that impede the adoption of such modern breeding tools in the tropics and opportunities to overcome them are highlighted

A Machine Learning approach to find genetic diversity in Rice Genome Variants

Mr. Ramesh Naidu Laveti
Senior Technical Officer
C-DAC
Bangalore
India
rameshl@cdac.in

Mr. Ramesh Naidu Laveti is working as Senior Technical Officer at “Centre for Development of Advanced Computing (C-DAC, Bangalore)” in the Big Data Analytics Team. He have more than ten-years of professional experience in the implementation of Parallel Math Kernel Libraries for Scientific Applications on HPC, Grid and Cloud Infrastructures. Currently working in the area of Big Data Analytics and Machine Learning. His areas of interest include: Big Data Analytics, Tools and Technologies & Machine Learning, Heterogeneous Programming, Data Intensive Computing, Parallel Algorithms & Parallel Programming Paradigms, Scientific Grid & Cloud computing.

Abstract

It is essential to understand genetic diversity of rice within-species to enhance crop production and quality. The 3000 Rice Genome Project, unprecedented resource, gave us more challenges and opportunities to explore the sequence data and its variants. Systematic mining of this huge open data to analyze genotypic variation and link it to functional variation can be a step forward in creating new and sustainable rice varieties. One of the challenges is to analyze the rice single nucleotide polymorphisms (SNPs) which were discovered when aligned to the reference genome, Nipponbare to understand genetic diversity.

In this paper, Oryza sativa Germplasm “DJ 123”- Bangladesh origin, Oryza sativa Germplasm “kasalath”- Indian origin, Oryza sativa Germplasm “IR64” - Philippines origin and Oryza sativa Germplasm “93-11”- China origin are considered for analysis. SNPs of these four different countries origin rice genomes are analyzed and visualized to capture the level of genetic variations among them by creating clusters using machine learning. The model built is able to predict the origin of new population structures based on genetic variants. Since input files (VCF files) are very huge, we used distributed framework compute engine, Apache Spark and advanced machine learning algorithms which otherwise traditional technologies may not handle.

Data Science Applications in Genomics

Dr. Ramana V Davuluri
Professor of Preventive Medicine –
Health and Biomedical Informatics
Northwestern University - Feinberg
School of Medicine
Chicago
USA

Dr. Raman Babu holds a PhD (Genetics) from Indian Agricultural Research Institute, New Delhi. From 1999 to 2005, Babu served as maize breeder with ICAR in Almora and is credited with developing first generation MAS products in maize that have double the amount of lysine and tryptophan, popularly referred to as Quality Protein Maize. Babu was awarded a post-doctoral fellowship by Japanese Science Promotion Society (JSPS) in 2006 to pursue research on positional cloning of disease lesion mimic genes that offer broad spectrum resistance to blast pathogen in rice. Babu joined CIMMYT in 2008 and steered the maize molecular breeding activities in the global maize program for improving abiotic stress tolerance and nutritional quality traits in the tropical germplasm. Since 2016, Babu is with Dupont Pioneer as Technology Deployment Lead (Asia-Pacific) overseeing molecular breeding, doubled haploids, precision phenotyping and informatics components across crops.

Abstract

Given that the biomedical research is rapidly acquiring the character of BIG DATA, with rapid accumulation of datasets on genes, proteins and other molecules, data science applications are increasingly playing an important role in the analysis and interpretation of these large files ranging from discovery phase to clinical applications. Bioinformaticians have been successfully mastered the application of data science skills since the early days of human genome sequencing; for example, prediction of genes in the assembled genomes or genomic contigs. I will discuss some of those applications our group has successfully applied in the prediction of (a) gene promoters in the human genome (b) gene regulatory signals that are altered in breast cancer, (c) molecular grouping of brain tumor patients and (d) functional roles of germline single nucleotide variants that are associated with prostate cancer. I will also discuss various data science issues; for example – (a) processing of unstructured data to prepare the data matrices; (b) clustering of samples based on gene expression data; (c) feature selection; and (d) classification algorithms, etc.

Mapping of quantitative trait loci in cassava using Markov chain Monte Carlo (MCMC) method

Dr. J Sreekumar
Principal Scientist
Central Tuber Crops Research Institute,
Thiruvananthapuram, India
sreekumarctcri@gmail.com

Dr. J.Sreekumar is working as a Principal Scientist at ICAR-Central Tuber Crops Research Institute, Thiruvananthapuram, Kerala. He did his MSc. in Agricultural Statistics from ICAR-Indian Agricultural Research Institute (IARI), New Delhi and PhD is Statistics from Mahatma Gandhi University, Kottayam, Kerala. He has got post-doctoral work experience in Bioinformatics at Plant Research International, Wageningen University, The Netherlands. He is a Faculty Member of IISER, Thiruvanathapuram and Research guide of MSc (Integrated) Biotechnology students of Kerala Agricultural University, Vellayani. His main research interests are Statistical Genetics & Genomics, Bioinformatics and Computational Biology. Dr. J. Sreekumar has extensive research contributions in the field of Agricultural Statistics in general and application of Statistical and machine learning techniques in predicting protein-protein interaction using multiple sequence alignments and also in gene interaction network prediction using gene expression data. He has handled various research projects with grants received from Department of Biotechnology, Ministry of Electronics and Information Technology etc. He published many research articles in national and international journals of repute and guided many students.

Abstract

Genetic markers have been used to map quantitative trait loci (QTL) for a long time to associate phenotypic traits with morphological markers. The availability of suitable phenotypic markers was the main practical limitation. But with the availability of inexhaustible supply of molecular markers there is considerable progress in detection of QTL for many agricultural importance traits. Many of these traits are of quantitative in nature. Many statistical methods have been well developed for QTL mapping such as ANOVA where the genotypes at particular location is grouped into two or three groups depending on the type of cross and the phenotypic means are compared using standard hypothesis test (t-test or F-test). There were several limitations with this method. The method of Interval mapping proposed was a significant development in this area which addresses many important issues in the previous method. The advantage of this method is that apart from searching for QTL in markers it can look into putative QTL between two adjacent flanking markers. Interval mapping is biased when nearby linked QTLs influence the putative QTL in question. Composite interval mapping and multiple QTL models overcome this issue by simultaneous search for multiple QTL in multiple intervals. However these methods were mainly for traits with continuous distributions. Many of the agricultural important traits are of discrete in nature for example disease score, which are also polygenic. Bayesian methods of QTL mapping are preferable for such traits with non-normal distribution. A real life QTL mapping experiment in cassava for cassava mosaic disease, where the disease score, which is of discrete in nature was used for the study. Different methods of QTL mapping were applied. To apply the Bayesian method of QTL mapping, the disease score was modelled as following poisson distribution and the expectation is connected to QTL effects through a log link function. The MCMC procedure in SAS particularly designed for Bayesian analysis using Markov chain Mote Carlo algorithm was used for carrying out the computations. In this lecture the results from the different methods applied for the QTL detection will be discussed in detail.

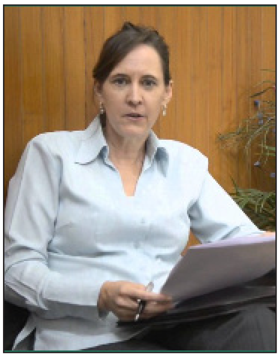
Session XII
Data Science Teaching & Scientific Communication

Convener and Co-Convener

Dr. Joanna Kane-Potaka
Director-Strategic Marketing & Communication
ICRISAT
India

Dr. A K Choubey
Head, Division of Computer Applications, Indian Agricultural Statistics Research Institute
New Delhi
India

Title



Ms. Joanna Kane-Potaka
Director-Strategic Marketing & Communication
ICRISAT
India
J.Kane-Potaka@cgiar.org

Ms. Joanna Kane-Potaka, an Australian national, is the Director of Strategic Marketing and Communication, ICRISAT. Ms. Potaka has a Master of Science degree in Global Marketing. Prior to joining ICRISAT, she was Director, Communications and Marketing Information and Knowledge Group, IWMI. She was also Head, Information Management and Marketing, Bioversity International and Information and Communications Leader, WorldFish Center. She began her career as an agricultural economist with the Australian Bureau of Agricultural and Resource Economics and later moved into market research in the agribusiness area of the Queensland Department of Primary Industries. Ms. Joanna Kane-Potaka she has worked in a wide variety of marketing-related areas including strategic marketing, communications, fundraising, knowledge management, and uptake of scientific research. She has worked for government, private industry and with non-profit. She has more than 25 years of professional experience. She is a certified practicing marketer in Australia, an Associate of the Chartered Institute of Marketing in the UK and a fellow of the Australian Marketing Institute.

Title

Dr. Donghui Ma
Chief Executive Officer, VSN China / Biosci
Thailand
Thailand
India
donghui.ma@vsni.co.uk

Dr. Donghui Ma has scientific background is in biostatistics, plant breeding and business administration. As CEO of VSN China and Biosci Thailand he is responsible to identify market opportunities and build as well as lead teams to implement our market entry strategy. As the VSN and IBP product and services are both very competitive in plant breeding, his work experience in leading European plant breeding company give him many insights in the R&D market from this sector. Working as part of the VSN and IBP collaboration (PPP) makes him also well connected in the global network, which is important to achieve our commercial goal and sustainable growth.

Title

Dr. Madhavi Ravikumar
Assistant Professor
Department of Communication,
Hyderabad Central University
madhaviravik@gmail.com

Dr. Madhavi Ravikumar, a journalist, academician and researcher for over 20 years. She had obtained PhD in Journalism and Mass Communication. Prior to joining the present position as Assistant Professor, Department of Communication, S N School of Arts and Communication, Hyderabad Central University, she was with Asian College of Journalism, Chennai. She had been associated with media organisation and NGO's like The Indian Express, The Hindu, The Week, NDTV, All India Radio and M S Swaminathan Research Foundation (MSSRF) in various capacities. Dr. Madhavi Ravikumar had published books and articles in both national and international publications on various subjects. Also done few projects with BBC, German institute InWent and Unicef. Her teaching interests include print and new media, broadcast journalism, development journalism and convergence journalism. Dr Madhavi research interest has been in areas like media studies, development communication and policy studies.

Title

Dr. B Jayashree
Principal Scientist and Head
The Hindu Media Resource
Centre at M S Swaminathan
Research Foundation
jayashreeb@mssrf.res.in

B Jayashree is Principal Scientist and Head – The Hindu Media Resource Centre at M S Swaminathan Research Foundation. Her work in media and communication for over two decades has traversed sectors of HIV/AIDS prevention, Child Health, Nutrition, Water and Sanitation, Agriculture, Food Security besides corporate and media experience. At MSSRF she has taken forward the Foundation's goal of linking science and society, through various contemporary communication approaches. Prior to this, she worked with UNICEF and with Bill and Melinda Gates Foundation's HIV project as Communication Officer. In her media career, she was Bureau Chief South India, with India Today Television. Her work has been carried out in English, Hindi and in Tamil, reaching different sections of people. An academic researcher in media science, she has Masters' degrees in Mass Communication, Human Rights Law and English Journalism and is recipient of the UK Chevening scholarship, One World Broadcast Fellowship, London and RNTC Fellowship for Media and Development, Netherlands. A national certified trainer in CSR, she is also visiting faculty of development journalism. She has written and broadcast extensively in popular media besides engaging in strategic communication.

Plenary Talk

Accelerating genetic gains delivered by plant breeding; the role of data management and selection decision support

Dr. Gary Atlin
Senior Program Officer
Bill & Melinda Gates Foundation,
Seattle, USA
Gary.Atlin@gatesfoundation

Dr. Gary Atlin is research manager, agricultural development specialist, and plant breeder with 30 years of experience in cultivar development, genetic analysis, and the management and optimization of breeding programs in the Americas, Africa, and Asia. Currently Senior Program Officer within the Agricultural Research and Development team at the Bill and Melinda Gates Foundation. As a Senior Program Officer he is working on developing and managing grants on crop improvement and agricultural productivity within the Agricultural R&D Initiative of the Global Development Program. His specialties: Rice, maize, and wheat cultivar development, breeding for abiotic stress tolerance, quantitative genetics and applications of genomic selection, optimization of cultivar and trait development pipelines, agricultural development in Canada, South and Southeast Asia, Africa, and Latin America. A leader in the field of management of genotype x environment interaction and the detection and deployment of genes for drought tolerance in plant breeding programs. Extensive experience in management of plant breeding programs, and in planning and coordinating agricultural research projects in developing countries. Guiding the development of accountability and breeding and genomic information management systems for public plant breeding programs in Africa and South Asia.

Abstract

The rate of genetic gain delivered by public plant breeding programs since the end of the Green Revolution has been slow, and must be increased to meet increasing global demand for food in the face of a changing climate. Plant breeders must make optimal use of the data they generate to accelerate breeding cycles, increase selection intensity, improve phenotyping accuracy, and integrate molecular markers in selection decisions. Breeding pipelines will need to be re-designed and optimized to allow efficient use of resources. Breeders will need to be supported in the adoption of integrated data management systems that support population development and the integration of phenotypic and genotypic data on selection candidates. Data collection, validation, and analysis functions need to be automated. Predictions need to be based on mixed models that account for selection bias and allow breeders to weight traits according to repeatability in phenotyping. Biometricians and data management professionals need to be fully integrated into breeding teams for effective population development and selection decision support.

Special Talk

Intelligent Cloud as Transformative Force

Dr. Prashant Gupta
Principal Director
Microsoft Cloud & Enterprise
Division, India
prgup@microsoft.com

Dr. Prashant Gupta works as Principal Director in Microsoft Cloud & Enterprise Division. He has contributed to company's cloud computing strategy since its early days incl. work on introduction of Linux in Windows Azure IaaS and acquisition of Revolution R for SQL Server and Azure Machine Learning. Prashant co-founded Microsoft Accelerator for startups in May 2012 that later became Microsoft Ventures and at present, driving Digital India opportunities in Machine Learning – especially in education, healthcare and agriculture space and most recently, architected first machine learning conference in Bangalore. Prashant started his career with Motorola in Chicago & worked on early mobile internet innovations incl. VIAMOTO – one of the first voice assistance for GPS navigation on cellphone, RAZR Maxx V6 – first 3.5G touch device and & then helped setup Ingram backed Silicon Valley startup in mobile publishing before joining Microsoft. He has an MS degree from University of Florida, MBA in Marketing from Kellogg and is a recipient of IEEE Larry K. Wilson Award, Lt. Governor Gold Medal from Delhi College of Engineering.

Abstract

Azure cloud platform supports the broadest selection of operating systems, programming languages, frameworks, tools, databases and devices - run Linux; build apps with JavaScript, Python, .NET, PHP, Java and Node.js; build back-ends for iOS, Android and Windows devices. Azure easily integrates with existing IT environment through the largest network of secure private connections, hybrid database and storage solutions and data residency and encryption features — so enterprise assets can stay right where you need them. Microsoft has made an industry-leading commitment to the protection and privacy of your data. It was the first cloud provider recognized by the European Union's data protection authorities for commitment to rigorous EU privacy laws. Microsoft was also the first major cloud provider to adopt the new international cloud privacy standard, ISO 27018. Microsoft has also launched Azure Government, a stand-alone version of Azure designed to meet the rigorous compliance requirements of U.S. public agencies. Azure runs on a worldwide network of Microsoft-managed datacenters across 26 regions and now has one of the most comprehensive suite comprising of Big Data Analytics and Machine Learning.

Contributory Sessions

Recent Advances in Statistics (Session-I)

Mixture Distribution Approach for Identifying Differentially Expressed Genes in Microarray Data of *Arabidopsis Thaliana*

Arfa Anjum[1]*****, **Seema Jaggi**[1], **Eldho Varghese**[1], **Shwetank Lall**[1], **Anil Rai**[1], **Arpan Bhowmik**[1], **Dwijesh Chandra Mishra**[1] and **Sarika**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi – 110 012. India
*Email: anjum.arfa@gmail.com

The basic aim of analyzing gene expression data is the need to identify genes whose expression patterns differ according to phenotype or experimental condition. A microarray is a tool for analyzing simultaneous relative expression level of thousands of genes within a particular cell population or tissue in single experiment through the hybridization of RNA. Determining the genes that are differentially expressed in two different types of tissues or samples obtained under two experimental conditions is the matter of interest in analyzing the microarray data. To identify differentially expressed genes between two conditions, it is important to consider the distributional property of the data. Fitting of appropriate distribution to gene expression data provides statistically sound cutoff values for identifying differentially expressed genes. In this paper, the focus is mainly to investigate the differential gene expression analysis for sequence data based on mixture distribution model. Mixture model provides a mechanism to incorporate extra variation and correlation in the data, add model flexibility and are a natural approach for modelling data that arise in multiple stages or when populations are composed of sub populations. Two-component mixture normal model was fitted to microarray data of Arabidopsis thaliana and was found to be more capable of capturing the variability as compared to single component normal distribution. It was concluded that two-component mixture normal model identified differentially expressed genes more accurately in microarray data.

Role of Big Data in Rice Crop Management through Integration of Geospatial Technologies, Crop and Climate Models

B. Sailaja[1]***, S. Gayathri**[1]**, D. Subrahmanyam**[1]**, SR Voleti**[1]**, R. Nagarjuna Kumar**[1] and **V Ravindrababu**[1]
[1]Indian Institute of Rice Research, Rajendranagar, Hyderabad-30, Telangana, India
*Email: bandasailaja@gmail.com

Recent weather conditions force us to look for future challenges in the agricultural systems due to increasing climate variability. Rice is the most important crop in India and its diversity in growing environments, management levels and production constraints vary spatially and temporally. Growing suppression in ground water levels, increase in irrigation demand, inefficient use of fertilisers and pesticides warn agriculture systems to utilise precision technologies. The complexity of modern farming practices need a system approach which is an interaction between weather-soil-genotypes and crop management to evaluate their impact on production and environmental outcomes. Big data hold tremendous values, but until they are properly analyzed, their impact in agriculture remains limited. We are now at a critical point to promote the adoption and utilization of precision technologies due to new developments in big data through yield mapping and advancement in the spatial, spectral and temporal resolutions of remote sensing data, GIS technologies, improved crop and climate modelling and Information and Communication Technology (ICT) systems. This paper will describe the latest advancement in spatial technologies linked with crop modelling and weather forecasts to enable growers to implement spatially variable management across the field to increase profit and reduce environmental impact.

Genealogical Studies of High Yielding Rice Varieties of India through Big Data Analytics

Jyothi Badri[1]***, N Shobha Rani**[1]**, S Srikanth**[1]**, VP Bhadana**[1]**, RM Sundaram**[1]**, P Senguttuvel**[1]**, LV Subba Rao**[1]**, AS Hariprasad**[1]**, GSV Prasad**[1]**, BC Viraktamath**[1]**, T Ram**[1] and **V Ravindra Babu**[1]
[1]ICAR-Indian Institute of Rice Research, Rajendranagar, Hyderabad-500030, Telangana, India
*Email: jyothirishik@gmail.com

Introduction of plant type based high yielding rice varieties since early 1990s transformed India from a net importing country to a potential exporter of rice. More than 1000 high yielding rice varieties (HYRVs) including hybrids were released in the country replacing thousands of traditional land races, thus bringing down the genetic diversity. A detailed genealogical study was undertaken to assess the genetic diversity that has been utilized in breeding HYRVs. The information on parents involved, breeding methods, procedures adopted in the development of HYRVs released from 1966-2012 was collected from various sources viz., national database (ICAR-IIRR), breeders concerned and literature. The information was curated with Integrated Breeding platform (IBP) and International Rice Information System (IRIS). The huge data set was analysed using Genealogical Management System (GMS) of the International Crop Information System (ICIS). GMS captures data describing any conceivable method that plant breeder might use and accommodates the complete selection history as well as novel methods for generating variability. The analysis revealed the use of 2750 unique parents (10.3%) of Indian (51%) as well as exotic origin (49%) in the development of HYRVs. However, only 92 genotypes of Indian origin were most frequently used in the development of HYRVs of which 66 (71.7%) were landraces indicating the richness of Indian germplasm which could be 15-20% of the available worldwide rice genetic diversity. Enormous contribution of Indian ancestors to the Indian breeding programmes are due to 15 most frequently used genotypes that include Latisail, GEB 24, Co 18, SLO 17, Ptb 21, CR 55-35, T141, Mudgo, Nagina 22, Eswarakora, Basmati 370, Bhogjira 1, ARC 6650 and ARC 5984. Latisail has served as starting material in breeding of 567 HYRVs and virtually had gone into the background of most of the HYRVs developed at IRRI and elsewhere. Among the exotic sources, the germplasm contribution from IRRI collection was 71.42% of which 84 elite lines were most often used. Besides India and IRRI, 37 other countries contributed germplasm to the Indian HYRVs. Japan, USA, Taiwan, China, Bangladesh, Columbia, Thailand, Philippines and Indonesia were the top contributors of germplasm. The Taiwanese donor Dee-Geo-Woo-Gen (DGWG) was involved in 936 cross combinations resulting in 646 Indian HYRVs. Decadal wise trends revealed maximum utilization of well known traditional landraces during 1961-70 that accounted up to 94 with the landmark release of two miracle rice varieties 'Jaya' and 'IR8'. The next decade (1971-80) witnessed 5-fold increase in the use of 489 diverse unique parents and the number further increased to 742 during 1981-90 and to 1285 in 1991-2000, however majority of them were not the terminal landraces but were intermediate derivatives. This indicates that most of the selections were carried from crosses made between released varieties or with elite lines which share common ancestry in the last thirty years (1981-2012). The apparent increase in genetic diversity in Indian HYRVs is due to the increase in generative activities in breeding but the obvious narrow genetic diversity is due to reduction in the use of terminal landraces in breeding programmes.

Microarray Data Expression Study for Better Identification of Differentially Expressed Genes

Neeraj Budhlakoti[1]* and **Ravi Shankar**[1]
[1]ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012
*Email: neeraj35669@gmail.com

As whole genome sequencing technologies led to a boom in the availability of genetic information. Microarray has been emerged as one of the most powerful tool in the field of transcriptomics. It analyses the expression of gene in a cell or tissue in given moment of time. It allows the scientist to understand the molecular mechanism in normal and modified conditions. Here our main focus is to study variation among the results of existing technologies and also suggest some protocol to add more reliability in the result. In this study we have compared the different methods of expression analysis. We found that there is significant difference among the results, as every gene has its own expression level range and they cannot express beyond that level. Existing methods does not take care of this, which may result in filtering of some important differentially expressed genes due to logFC and p-value criteria, applied generally. Here we suggest to use gene expression data for control condition selected across the different experiment from same tissue and species following similar protocol. In this way we can have wide range of gene expression for controlled condition which may add some more confidence in the results.

Role of Big Data in Agriculture-A Statistical Prospective

Sukanta Dash[1*], **Prakash Kumar**[1], **Kanchan Sinha**[1] and **Mrinmoy Ray**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi - 110 012, India
*Email: sukanta.iasri@gmail.com

Data are playing an important role making good planning and policies for agricultural growth and development. Population growth and climate change are worldwide trends that are increasing the importance of using big data science to improve agriculture. Add to that land degradation increasing marginal land and loss of biodiversity are better deals with study of big data science. Crop data can be break down into bits and bytes it will give better study about the crop development by using advance data analytics tools for betterment of agriculture. Here, talk about some important tools and techniques to handle and study the big data.

In silico Prediction of Off-Target Minimized Efficient RNA Interference Molecules for Management of Insect Pests

G Sharath Chandra[1]***** and **R Asokan**[1]
[1]Division of Biotechnology, Indian Institute of Horticultural Research, Hesaraghatta Lake (PO),
Bangalore – 560 089, India.
*Email: sharathgsc@gmail.com

Agriculture is the backbone of Indian economy as it plays a prominent role in providing food, feed and fodder and generating jobs and income for the country. Agricultural crops are facing systematic annual losses due to pests and diseases. Among the insect pests, *Helicoverpa armigera* is a global polyphagous pest inflicting many important crops such as chickpea, pigeonpea, cotton, tomato, etc. and causing billions of rupees of loss in the form of crop losses and pesticide input. Although, this pest is not amenable to the current management strategies, viz. transgenic plants expressing *Bacillus thuringiensis* derived Cry proteins (toxins) and chemical insecticides due to resistance development against these. In view of this, the cutting edge tool, RNA interference (RNAi) holds tremendous potential in the management of insect pests. Currently, double-stranded RNA (dsRNA) is been used to achieve the RNAi, hence, the potency of RNAi is governed by the dsRNA. Though, the efficacy of RNAi is highly influenced by the selection of region for dsRNA designing. In this regard, we have compared different online freeware software's such as dsCheck, Sfold and SnapDragon tools for identifying the efficient dsRNA region. Moreover, the effective dsRNA design perhaps fraught with 'off-target' effects such as unintentional silencing of non-targeted genes. To overcome this impediment, the selected dsRNA regions were further fragmented into 21 nucleotides substrings then these were subjected to global blast analysis using NCBI-BLASTn online tool. Therefore, using the above software's we have designed off-target minimized effective dsRNA regions for three important genes, chymotrypsin, cytochrome P450 (P450) and juvenile hormone acid methyl transferase (*jhamt*) which are involved in digestion, detoxification and metamorphosis in the insect. We have isolated these three genes from *H. armigera* larva, cloned and sequence characterized using NCBI-BLASTn. Finally, off-target minimized effective dsRNA regions of 146bp of chymotrypsin, 144bp of P450 and 121bp of *jhamt* were identified. These sequences were used for the development of chimeric dsRNA constructs to achieve efficient RNAi. Additionally, we have compared the efficiency of the above selected dsRNA with the randomly (without any criteria) selected dsRNA in silencing of target genes. Results revealed that there was a significant variation in the extent of silencing and the selected effective dsRNA regions (using above tools) were found to cause higher silencing than the randomly selected dsRNA regions. The demonstrated strategy aids in the identification of effective and eco-safer target regions for effective RNAi mediated management of insect pests.

A New Algorithm for Multiple Pattern Matching using Pair Count of Pattern

B. Manadhata[1], **N.N. Jambhulkar**[1*], **R. Jena**[1] and **L.K. Bose**[1]
[1]ICAR-National Rice Research Institute, Cuttack – 753006
*Email: nitiprasad1@gmail.com

Pattern matching is an important and active research area with large applications now a day. Pattern matching in a DNA sequence or searching a pattern from a large data base is a major research area in computational biology. DNA is the basic building block of life composed of nucleotides of four types i.e. Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). Pattern matching is the finding of occurrences of a particular pattern from a large sequence. To extract the pattern match from a large sequence it takes more time, in order to reduce searching time we have proposed an approach that reduces the search time with accurate retrieval of the matched pattern in the sequence. As performance plays a major role in extracting patterns from a given DNA sequence or from a database independent of the size of the sequence. When sequence databases grow, more efficient approaches to multiple matching are becoming more important. Hence the aim of this paper is to develop a pattern matching algorithm which is more efficient than the available algorithm. The algorithm with least number of comparisons for determination of the pattern in the sequence is said to be more efficient algorithm. The algorithm works in two phases, pre-processing phase and processing phase. Pre-processing of index occurrence table is done by taking each paired characters one by one and its corresponding index value is stored in the array. The pre-processing method is used for constructing the array of index table till all the paired characters of the pattern completes i.e. till the end of paired characters. The second phase is processing phase. In this phase the sequence alignment is done using minimum number of occurrences of pattern pair with that of the sequence. Matching is done using the maximum number of occurrence of pattern pair in the pattern. If it matches then the next lowest occurrence of pair in the pattern and the process is repeated till all pairs in the pattern exhausted. The results have been compared with several other available algorithms in literature and the proposed algorithm reduced the number of comparisons as compared to other algorithms.

Comparative Modeling, Molecular Docking and Simulation Studies of a Chickpea Multiple Stress Responsive WRKY Transcription Factor

Aravind Kumar Konda[1*], **Aravind Setti**[2], **M.A.Iquebal**[3], **Sarika**[3], **Khela Ram Soren**[1], **Shanmugavadivel PS**[1], **Sushil K. Chaturvedi**[1], **Pallavi Singh**[1], **Priyanka Gangwar**[1] and **N P Singh**[1]
[1]ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, Uttar Pradesh, 208024, India; [2]Osmania University (OU), Hyderabad, Telangana, 500007, India; [3]ICAR-Indian Agricultural Statistics Research Institute (IASRI), Posa, New Delhi-110012, India;
*Email: aravindbio@gmail.com

WRKY transcription factors (TFs) are one of the major TF families in the plant kingdom characterized by the presence of heptad amino acid sequence WRKYGQK in the DNA binding domain extending 70-80 amino acids in length (Rushton et al. 1995, 1996, Eulgem et al. 2000). The WRKY domain specifically binds to cis-acting element: W box (5'-(T) TGAC [CT]-3') and regulate gene expression influencing plethora of cellular processes like defense signaling, biotic and abiotic stresses, seed development, seed dormancy, germination, hormonal signaling and senescence etc., and thus act as a hub of cell signaling pathways (Agarwal et al. 2011, Duan et al. 2007, Ciolkowski *et al.* 2008 Yamasaki *et al.* 2012, Chen et al.2012, Eulgem *et al.* 2007). A *Cicer arietinum* WRKY TF (Car-WRKY TF) that is differentially expressed in response exposure against to multiple stress conditions like drought, salinity, cold and herbicide has been identified (Mantri *et al.* 2007, Soren *et al.* 2016, unpublished). Presumably, the Car-WRKY TF may play a central role in stress signal transduction pathways in chickpea. The present study has been carried out with an objective to predict the Car-WRKY TF structure and decipher protein-protein networks under stress conditions.

Analysis of Expression Profiles of Genes Specific for Wild Abortive Cytoplasmic Male Sterility (WA-CMS) and Fertility Restoration Trait in Rice

Pranathi K[1*], Virakthamath BC[1], Hari Prasad AS[1], Neeraja CN[1], Balachandran SM[1], Kulkarni SR[1], Koti P[1], Hajira SK[1], Balachiranjivi CH[1], Bhaskar S[1], Anila M[1], Rekha G[1], Dilip T[1], Koushik MBVN[1], Harika G[1], Kale RR[1] and R.M. Sundaram[1]

[1]Biotechnology lab, Indian Institute of Rice Research, Hyderabad, Telangana, India
*Email: karnatipranathi@yahoo.co.in

Large-scale cultivation of hybrid rice is one of the feasible options for significantly enhancing India's rice production. Hybrids based on WA-CMS system have become popular among most of the rice growing countries like India, China, Vietnam, Philippines etc. They have 15-20% more yield advantage over inbred varieties and are produced using three-line system of breeding, involving a wild abortive cytoplasmic male sterile line (WA-CMS), its cognate iso-nuclear maintainer line and a genetically different restorer line, which possess fertility restorer (Rf) genes in the nuclear genome. CMS/Rf interactions play a key role in exploiting heterosis. Manifestation of sterility in CMS lines is associated with chimeric mitochondrial genes that co-transcribe along with other functional mitochondrial genes and sterility is restored by interaction of dominant nuclear fertility restoring gene products (Rf) with chimeric gene/gene products. Fertility restoration in WA-CMS system is reported to be associated with Rf4 on chromosome 10 and Rf3 on chromosome 1. In present study, a set of genes/ORFs reported to be associated with trait of CMS and fertility restoration in rice were analyzed for their expression using RNA extracted from young premature rice florets of IR58025A (WA-CMS), IR58025B (maintainer), KMR3R (restorer) and KRH2 (hybrid). With respect to WA-CMS associated genes/ORFs, only WA352 and rpl5 genes expressed only in sterile mitochondria possessing lines (WA-CMS and the hybrid). WA352 is a reported chimeric gene that co-transcribed with rpl5 mitochondrial gene, WA352-rpl5 transcripts unregulated a reactive oxygen species (ROS) pathway in tapetum cells leading to abortion of pollen. Among the genes specific for Rf4, PPR9 (pentatricopeptide repeat containing protein) gene showed expression only in restorer lines. PPR9 gene products were reported to reduce WA352-rpl5. Through our expression analysis we confirmed the candidacy of WA352 and PPR9 genes and implicated these genes for their role with respect to WA-CMS and fertility restoration traits respectively.

Global Gene Expression Profiling in Response to Classical Swine Fever Vaccination in Indigenous and Crossbred Pigs using RNA Seq Analysis

Shalu Kumari Pathak[1], **Amit Kumar**[1]*****, **Bhuvana G**[1], **Vaishali Sah**[1], **Vikramaditya Upmanyu**[1], **A K Tiwari**[1], **Aditya Prasad Sahoo**[1], **Amit Ranjan Sahoo**[1], **Sajad A W**[1], **Manjit Panigrahi**[1], **N R Sahoo**[1] and **Ravi Kumar**[1]
[1]Animal Genetics Division, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh-243122, INDIA
*Email: vetamitchandan07@gmail.com

In present study global gene expression profiling was performed at the cellular level after classical swine fever (CSF) vaccination in crossbred and indigenous piglets. After RNA Seq analysis 5222, 6037 and 6210 common differentially expressed genes (DEGs) were identified by expectation maximization (RSEM) using EBSeq, DESeq2 and EdgeR in three different combinations i.e indigenous post-vaccinated verses pre-vaccinated, crossbred post-vaccinated verses pre-vaccinated and post-vaccinated crossbred verses indigenous pigs, respectively. Functional annotation of differentially expressed genes showed enrichment of antigen processing-Cross presentation, B cell receptor signalling, T cell receptor signalling, NF-κB signalling and TNF signalling pathways. The protein-protein interaction network among the immune genes included more number of genes with greater connectivity in vaccinated crossbred than the indigenous piglets. From RNA Seq analysis it was observed that expression profile of IRF3, TAP1, IL1β, SLA2, SLADR, CSK, PAG, IL-6ST, NFκB, Lyn, Fyn, Jun, Fos, Syk and BLNK were important genes were ascertaining to breed differences in humoral immune response of two different genetic group of piglets under investigation. Hence, it was concluded that immune response processes are more enriched in crossbred piglets and they show better adaptive immune response compared to indigenous (desi) following CSF vaccination.

Prediction of lncRNAs from Capra hircus Transcriptome

Himanshu Kumar[1]***, Sarika Sahu**[1] and **A.R.Rao**[1]
[1]Centre for Agricultural Bio-Informatics (CAB-in), ICAR-Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, PUSA Campus, New Delhi-12, India
*Email: himanshu.genetics@gmail.com

Large scale genome analysis of different organisms revealed that most of the genomic components are non-expressive and referred as non-coding RNAs. Different non coding RNAs are miRNA, siRNA, piRNA, snoRNA etc. Long non coding RNAs (lncRNA) are longer than 200 nucleotides and play a significant role in gene regulations of goat. However, exploration of lncRNAs in the context of primary and secondary hair follicles is limited. In this study genome sequence of *Capra hircus* has been retrieved from Goat Genome Consortium and considered as reference genome whereas transcriptome sequences of the primary hair follicles (PHF) and secondary hair follicles (SHF) of *capra hircus* goat are taken from Yang, *et al* (2013). Reference genome of the *Capra hircus* has been indexed with the help of Bowtie2, and the transcriptome sequences of the *Capra hircus* have been trimmed for the purpose of quality check. TopHat aligner was used to align the trimmed sequences against the indexed reference genome and sequences having less than 200 nucleotides have been excluded. The selected sequences have been aligned with the swissprot database to filter out the potential non protein coding RNAs. Out of 402456 PHF and 68784733 SHF sequences, a total of 3000 PHF and 8108 SHF were filtered out as likely long non coding RNAs. These were subjected to the estimation of coding potential by running Coding Potential Calculator (CPC) to infer the quality and completeness of lncRNA on the basis of protein coding score. A total of 2300 and 2750 sequences of PHF and SHF respectively were finally filtered as lncRNAs. The tRNAs, snRNAs, and snoRNAs were then eliminated from the filtered lncRNAs by scanning through their respective available databases and the rest of the sequences were considered as predicted lncRNAs. The selected lncRNAs were further analyzed for their function as well as prospective roles in gene regulations of goat.

In Silico Identification and Functional Annotation of Bread Wheat lncRNAs

Sarika Sahu[1*], **Indra Singh**[1], **Monika Dalal**[2], **Kishor Gaikwad**[2], **Sabari Ghosal**[3] and **A.R.Rao**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi – 110 012; [2]ICAR – National Research Centre on Plant Biotechnology; [3]Amity Institute of Biotechnology, Amity University, Noida
[*]Email: sahusarikaiiita@gmail.com

Long non coding RNAs (lncRNAs) are a class of non protein coding RNAs of length more than 200 base pairs. lncRNAs are regulated by development and play important role in biotic and abiotic stresses in plants. In the recent past, whole genome sequence of wheat was made available in public domain. Most of the in silico work in bread wheat has been focused on coding regions. However, very limited work has been done on non coding region of DNA elements. Therefore the aim of the present study was to (i) predict the lncRNAs from cDNA available in plant ensemble data base of wheat (ii) identify lncRNAs as target mimics (TM) for miRNAs and (iii) study the involvement of the lncRNA in gene expression. To accomplish these objectives a computational approach involving characteristic features of lncRNAs like length, coding potential and ORF has been followed to predict the probable lncRNAs. Further these lncRNAs were mapped on to transcripts from RNA-seq data of control and drought stressed tissues of wheat seedlings. The mapped transcripts were annotated and analysed for their functionality. Besides, through miRNA target prediction the target mimic lncRNAs were identified and their involvement was studied in various stress mechanisms. The results revealed that 45% of the available wheat cDNAs, available at NCBI, are having length between 200-1000 base pairs. Among them, 17% of transcripts have exhibited non coding potential. Out of these, 26% of transcripts have ORF length < 100 base pairs. All these transcripts were filtered out for housekeeping genes and potential lncRNAs. The identified potential lncRNAs have been then mapped on to the drought stress transcripts of wheat varieties: HD 2329 and RAJ3765. Out of total predicted lncRNAs, 3.5 % of lncRNA have been predicted to serve as a decoy for 10 different miRNAs of wheat. The annotations of mapped lncRNAs show their involvement in various metabolic processes like transcription factor, disease resistance RPM1 and elongation factors.

In silico Identification and Characterization of Enhanced Disease Susceptibility1 (EDS1) in *Solanum melongena*

Soumya Sharma[1*] and **Sunil Archak**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute
[*]Email: Sharmasoumya92@gmail.com

Brinjal (*Solanum melongena*) assumes a status of important vegetable crop in India under solanaceous family. Despite the huge significance of brinjal in Asian and African countries, there is still little molecular and genetic information available for this crop. Brinjal crop witnesses a greater amount of loss due to pest and disease incidences. Recent advances in genetic engineering technologies for example target specific genome editing using CRISPER/cas9 have empowered the scientific community with the most important weapon for modulating genomes for desired phenotypes. But, it requires detailed information of the gene and its function for being a suitable candidate for genetic editing. In this aspect, Enhanced Disease Susceptibility1 (EDS1), a key regulator of plant defence can be a suitable target for gene editing to manipulate host resistance. Arabidopsis thaliana EDS1 (AtEDS1) is the most deeply studied EDS1 protein with crystal structure in PDB (4NFU). In the present study an attempt to extract detailed information regarding this protein in brinjal has been done. Brinjal EDS1 (SmEDS1) protein coding sequence is extracted from brinjal genome database using tblastn taking AtEDS1 as a query sequence. Prediction of the coding sequence refined with the help of transcriptome assembly data. *SmEDS1* gene has been found in the contig Sme2.5_09498.1 of eggplant draft genome assembly. The gene has 1806 nucleotide long coding sequence encoding for 602 amino acid long protein. The gene has the most common architecture with three exons as compared to EDS1 protein coding genes from other families. The comparative analysis of SmEDS1 protein along with 46 other species EDS1 proteins proved the strong sequence and structural conservation of this protein among plants. The incongruence in the sequence and structure based phylogenetic trees was observed that could be attributed due to the influence of difference between global alignment and conservation of sequence signatures or it could be possibly explained by the fact that sequences are related by phylogeny whereas structures by constraints on their functions and regulations. Deeper analyses of phylogenetic relationships based on EDS1 will facilitate its genetic manipulation for agronomic purposes.

Application of Hadoop Map Reduce Framework to Detect Tandem Repeats in DNA Sequence

Vandanababu T[1]***** and **Dr Raju Bhukya**[2]
[1]Research Scholar, Department of CSE, National Institute of Technology Warangal
[2]Assistant Professor, Department of CSE, National Institute of Technology Warangal
*Email: vandana9925@gmail.com

Repetitive structure of DNA reveals many secrets to discover. Simple sequence repeats(SSRs) or tandem repeats or micro satellites which are widely distributed in genomes have been used extensively for genetic studies such as diversity evolution, species identification, gene mapping; these have been associated with complex diseases. Variation in number of repeat units provides information to understand genetic diversity and its importance; it also provides information for gene mapping and also reveals diseases and consequences. These repeats in DNA have been provided pathways of diseases and mechanism of human diversity. Special tandem repeats like mono nucleotides were used to play a regulatory role in some instances. This study is continues one and grows day by day hence this field need to be updated to current most effective computing technologies. Now the field of bioinformatics is now to face up to and to deal with analysis of large scale data sets with the advent of next generation sequencing technologies, by these new technologies the DNA data is being doubled for every 5 to 7 months, the cost of DNA sequencing reduced from million dollars to few thousand dollars since 2000. At the same time there is need of new tools to process this much of huge data at a higher rate. All existing tools in this field were single processor based and run on traditional setup. Most recently there are very few tools in the field of bioinformatics developed by Hadoop frame work developed for sequence analysis for read alignment in map phase and variant calling in reduce phase, mining repetitive sequences in a map reduce setting proposes a method for finding various kinds of repeats by scanning the document, other map reduce framework for mining maximal contiguous frequent patterns in large DNA sequence Datasets. Tools for SSR were not there developed by Hadoop to get the advantage of parallel programming and to meet the current demands in field of bioinformatics.

Prediction of Heat Shock Proteins in Plants based on Amino Acid Compositions and Machine Learning Methods

V. Radhika
Division of Plant Genetic Resources, ICAR-Indian Institute of Horticultural Research,
Hesaraghatta Lake PO, Bangalore – 560030, Karnataka, India
Email: radhika.vanama@gmail.com

Heat shock proteins (HSPs) are an important class of proteins which are expressed in cells during extreme biotic or abiotic stress situations. The HSPs protect the cells during stress conditions and helps the organism to survive. Several HSPs are known to function as "molecular chaperones," preventing the aggregation and promoting the proper refolding of denatured proteins. They help the cells to cope up from damage to polypeptides in two ways – firstly by promoting degradation of abnormal proteins and secondly by reactivating stress-damaged proteins. In the plant system they are broadly divided into 5 families – HSP60, HSP70, HSP90, HSP100 and sHSP (small HSPs). Presently, agricultural scientists across the globe are investigating the role of these proteins in conferring stress tolerance to hybridized plants which can help in combating drought and poor soil conditions. Rapid identification of these HSPs is crucial in studies related to inducing plant tolerance to abiotic stresses using biotechnological approaches. Many methods of function prediction rely on identifying similarity in sequence and/or structure between a protein of unknown function and one or more known proteins. Other methods include identification of conserved patterns in members of a functionally uncharacterized family for which many sequences and structures are known. However, this kind of approach failed to work when the query protein does not have significant sequence similarity to any attribute-known proteins. Thus, various non-sequential or discrete models were proposed by different workers. The simplest discrete model used to represent a protein sample is its amino acid compositions and the di-peptide compositions. In the present study, utility of the features of protein sequences namely sequence length along with (i) amino acid compositions (ii) di-peptide compositions and (iii) in combination, with the help of four machine learning based classifiers viz. neural networks, decision trees, nearest neighbour and Naïve Bayes have been evaluated for the identification of the heat shock proteins so as to come out with the best approach. The protein sequences available in SwissProt protein database of National Centre for Biotechnology Information (NCBI) were utilized for the study. A classifier for classification of each class of heat shock proteins (HSP70, HSP90, HSP100 and sHSP) from the remaining sequences has been able developed. 10-fold cross-validation method has been followed for generating the training and testing sets and evaluation of classification models. For each classification algorithm, the performance of the classifier was obtained based on the measures of sensitivity, false positive rate, accuracy, precision, recall, F-measure values and area under the receiver operating characteristic curve (AUC). Based on the AUC measure, the Naïve Bayes algorithm has been found to be superior in identifying the heat shock proteins in all the classes. The results of the studies will have direct bearing on the outcome of various researches involved in HSP studies.

Modelling Breakpoint Data for Development of Intensity-Duration-Frequency Curves for Different Districts of Uttarakhand, India

N.M. Alam[1*], **P.K. Mishra**[1], **D.R. Sena**[1], **C. Jana**[1] and **G.C. Sharma**[1]
[1]ICAR-Indian Institute of Soil and Water Consecration, Dehradun
*Email: alam.nurnabi@gmail.com

Changes in the hydrologic cycle due to increase in greenhouse gases cause variations in intensity, duration, and frequency of precipitation events. Since rainfall characteristics are often used to design soil and water conservation structures, reviewing and updating rainfall characteristics (i.e., Intensity–Duration–Frequency (IDF) curves) for recent and future climate scenarios is necessary. Rainfall Intensity Duration Frequency (IDF) curves are developed using historical rainfall time series data. Annual extreme rainfall is fitted to a theoretical probability distribution from which rainfall intensities, corresponding to particular durations, are obtained. In the use of this procedure an assumption is made that historic hydro meteorological conditions can be used to characterize the future scenario (i.e., the historic record is assumed to be stationary). However, this assumption may not be valid under changing climatic conditions. Potential shifts in extreme rainfall at the local level demand revisions of the existing IDF curves which are generally followed while designing various soil and water conservation structures. Hence, quantifying the potential effects of climate variability on rainfall characteristics and adapting to them are very demanding to reduce the vulnerability of extreme rainfall events on infrastructures developed for soil and water conservation. The characteristics of rainfall storm amount, intensity and duration play and important role in determining the rate of soil erosion. Greater in the intensity of rainfall, greater kinetic energy it processes (Wischmeier and Smith, 1958). Rainfall intensity-duration-return period equation and nomographs for various stations are required for design of soil conservation and runoff disposal structures and for planning flood control projects. Such relationship and nomographs have been developed earlier by Gupta et al. (1968), Raghunath et al. (1969), Khullar et al. (1975), Senapati et al. (1976), Ram Babu et al. (1979, 2001a,b and 2002).

Given, the importance of IDF curves in the design of such hydraulic structures, an effort was made to develop IDF curve using rainfall breakpoint data of last 30 years (1985-2014) of different districts of Uttarakhand. A visual Basic based programme was developed was and used to calculate rainfall intensity at different interval (5 minutes to 24 hours). Return period of rainfall (2 years to 100 years) has been calculated using Generalized Extreme value distribution and IDF curves for different districts of Uttarakhand have been developed using non-linear modeling technique. This developed IDF curves for different district of Uttarakhand will help in strengthening the robust design criteria of various hydraulic structures.

Modelling on the Establishment of Zygogramma Bicolorata for the Control of Parthenium

Yogita Gharde[1*] and **Sushil Kumar**[1]
[1]ICAR-Directorate of Weed Research, Maharajpur, Adhartal, Jabalpur- 482004, MP, India
[*]Email: yogitagharde@gmail.com

Parthenium hysterophorus L., known as feverfew, ragweed or *Parthenium* is a weed of world significance. The origin of *Parthenium* is considered to be from Mexico and about 20 countries are considered in its native range from where it has been introduced into other 34 countries (Adkin and Shabbir 2014) around the globe, including major continents except Europe and many islands. In Australia and India, *Parthenium* has achieved the status of "worst weed". The spread of *Parthenium* has been reported from all states of India in varying intensity. Among many methods, biological control by *Zygogramma bicolorata* is one of the most promising, suitable, cost effective and environmentally safe methods for the control of *Parthenium*. Initially, it was thought that *Z. bicolorata* will work effectively in India only in moderate climate and will not be able to establish well in the areas having low and high temperature extremes below and above 15 and 35°C, respectively (Jayanth and Bali, 1993a), but recent work betrayed this assumption as beetle was found to cause large-scale defoliation in many parts of such extremes (Sushil Kumar 2012, 2014).

A survey was made at 107 places covering whole India and level of establishment (for e.g. negligible, moderate and high) of beetles were recorded after releasing them in these places. Monthly weather data of 10 years from 2001-2010 was collected from Indian Meteorological Department (IMD), Pune on different weather parameters like average maximum temperature, average minimum temperature, rainfall, relative humidity for 107 cities of India where bioagents were released frequently during the period. Data on different weather parameters for the months July, August, September and October (favourable for good establishment of beetle) was considered for the study as mentioned by Jayanth and Bali (1993b) and Sushil Kumar (2010). Using different weights to these four months, composite weather variables were obtained as the weighted sum of the weather variable in different year from 2001-2010. Multinomial ordered logistic model was used for regression analysis. Results of logistic model yielded the different probabilities to predict the level of establishment in a particular place based on its weather conditions. Decision tree was also obtained using J48 algorithm which classified 81.7% instances correctly and finally after validation, rules have been obtained to predict the establishment of beetles. It is expected that this rules will be highly useful to decide the most suitable places for release and establishment of Z. bicolorata not only in India but in South Asia and other parts of the world also.

Recent Advances in Statistics (Session-II)

Genotype x Environment Interaction and Identifying High Yielding Stable Lines in Wheat under Afghanistan Environments

Darya Khan Akbarzai[1]*****, **Yashpal Saharawat**[1], **Lina Mohammadi**[1], **Abdul Rahman Manan**[1], **S. Tavva**[1], **Swain Nigmananda**[1] and **Murari Singh**[1]
[1]International Center for Agricultural Research in the Dry Areas (ICARDA), Kabul, Afghanistan
*Email: d.akbarzai@gmail.com

Wheat is a key cereal for human consumption in Afghanistan but has low productivity due to various biotic and abiotic stresses and soil factors. With a view to increase yield productivity, a number of wheat genotypes were evaluated in a four set of trials in incomplete block designs over locations in three to four provinces of Baghlan, Herat, Kabul, Mazar and Nangarhar during 2014-15. Variation due to genotype (G) and genotype x environment interaction (GEI) were examined and stability analysis of genotypes was carried out. Significant differences in genotypes effects were found for grain yield and days to following (DF) in each of the four locations and non- significant GEI ($P<0.01$) for one (14th-SBWYT2014-15) out of 4 trials. The stability of the genotypes over locations, the genotype SER.1B*2/3/KAUZ*2/BOW//KAUZ/4/KAUZ/FLORKWA-1 (G11) and SER.1B*2/3/KAUZ*2/BOW//KAUZ/4/PFAU/MILAN (G12) were two top yielders at 5.65t/ha and 5.44 t/ha overall means respectively. G11 being unstable with slope 1.47±0.06 ($P=0.02$ for differing from 1) in the sense of Type-2 stability as slope of regression line on the environment means differ significantly from unity, showed specific adaptation to Herat using a GGE-biplot, while G12 was stable with slope 0.937±0.32 ($P=0.86$ for differing from 1). In other trials (14th-ISBWYT 14-15), the top two high yielding genotypes were G8 [VEE#7//MT773/EMUS/3/SAFI-1] and G5 [SHUHA-8/DUCULA] with yield mean of 6.3 t/ha (slope of 1.13 ± 0.48, the $P=0.55$) and 6.2 t/ha (slope of 1.03 ± 0.03, the $P=0.23$). For trial 14th-ESBWYT 2014-15, the best high yielded genotypes were G6 [Solh 02 (an improved check)], specifically adapted to Mazar and G2 [VMP/MOS] with yield mean of 6.3 t/ha and 5.8 t/ha. Also in trial 15th-ESBWYT2014-15, the genotypes G17 and G22 were the top yielded with average mean of 6.7 t/ha and 6.4 t/ha. Further, the selected genotypes will be evaluated in diverse different in Afghanistan.

Investigation on Optimum Size and Shape of Plots for Field Experiment on Rustica Tobacco (*Nicotiana Rustica*, L)

BH Prajapati[1]***** and **Gajendra Beniwal**[1]
[1]SD Agricultural University, Sardar Krushinagar, Gujarat
*Email: bhp_1956@yahoo.com

A uniformity trial was conducted on rustica tobacco (Nicotiana rustica, L) crop at Agricultural Research Station, SD Agricultural University, Ladol to estimate optimum size and shape of plot along with number of replications using Dantiwada calcutti tobacco-4 (DCT-4). The data were analyzed using (1) Maximum curvature method (2) Fair-field Smith variance law and (3) comparable variance method. Based on coefficient of variation values, it was recommended that a plot of 8 basic units size with shape of 2 rows each of 4 plants (1.8 m. x 1.2 m. = 2.16m2) was considered as optimum plot size (net plot) with four minimum replications at 5% S.Em for precise experiment on rustica tobacco at ARS, Ladol (Gujarat).

Web Generation of Factorial Experiments with Minimum Level Changes in Run Sequences (*Web*FMC)

Arpan Bhowmik[1]***, Eldho Varghese**[1]**, Seema Jaggi**[1] and **Cini Varghese**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi - 110 012
*Email: arpan.stat@gmail.com

Factorial experiments, wherein two or more factors each at two or more levels are used simultaneously, have profound applications in many fields of agricultural and allied sciences. These experiments allow studying the effect of each individual factor as well as the effects of interactions between factors on the response variable. In order to avoid any kind of bias in the estimation of these effects, it is always advisable that the order of execution of runs in a factorial design is random. However, experimentation under factorial setup may become expensive, time-consuming and difficult due to a large number of changes in factor levels induced by randomization. The number of level changes is a serious issue to experimenters in many agricultural and allied experiments that involve hard-to-change factors. In such situations, an effective solution to adopt factorial experiments with minimum number of changes in the factor levels. For potential use of these designs, an online software for generation of factorial designs with minimum number of changes in the factor levels has been developed. This paper describes the web solution called *web*FMC (http://webfmc.iasri.res.in) developed using client–server architecture along with an online catalogue of the designs within a permissible range of parameters. *web*FMC is accessible any time from arbitrary platforms through internet. This software provides a freely available solution for the researchers and students working in this area.

Assessing Unrealised Yield Potential of Groundnut Producing Districts in India Using Multivariate Cluster Model

B.M.K. Raju[1*], **C.A. Rama Rao**[1], **Ch. Srinivasarao**[1], **K.V. Rao**[1], **M. Osman**[1], **M. Srinivasa Rao**[1], **R. Nagarjuna Kumar**[1], **V.V. Sumanth Kumar**[2] and **N. Swapna**[1]
[1] ICAR-Central Research Institute for Dry land Agriculture, Santoshnagar, Hyderabad-500059, Telangana, India;
[2] International Crops Research Institute for Semi-Arid Tropics, Patancheru, Hyderabad-502324, Telangana, India
[*]Email: bmkraju@yahoo.com

India is currently depending upon imports for about 50 % of its oil requirement. Groundnut is an important edible oilseed crop. Groundnut was cultivated in an area of 5.51 Mha during 2013-14 in India which produced 9.71 Mt with an yield of 1764 kg/ha. Any effort towards increasing production would relieve the pressure on imports. Scope for area expansion being limited, there is need for enhancement of yield. It calls for identifying areas where huge unrealised yield potential exists. With a view to address the issue, the present paper delineates homogeneous agro-climatic zones for groundnut production system in India using multivariate cluster model. District was considered as a unit of analysis and key factors of production such as climate, soil, irrigated area under the crop were used as clustering variables. There are 79 districts in India growing groundnut as a major crop (with area sown more than 10,000 ha). These districts were divided into 21 zones using multivariate cluster model. As the resources for raising the crop within a zone are more or less same, the district producing highest yield in a zone is regarded as potential target for the remaining districts in the zone. Unrealised yield potential (yield gap) for a district is computed as the difference between yield of the district and yield of potential target district (highest yield in the zone to which the district of interest belongs to). Unrealised yield potential was assessed at district level in each of the 21 zones. In each zone a comparison was made between two groups of districts, namely high unrealised yield potential districts and low unrealised yield potential districts, with respect to adoption of technology and other crop management factors. This analysis gave some idea of where the districts of high unrealised yield potential are lagging behind and scope for improving the yield in those districts. Findings of the study will be relevant for district level administrators.

Algorithmic Construction of A-optimal Balanced Bipartite Block Designs

B.N. Mandal
ICAR-Indian Agricultural Statistics Research Institute Library Avenue, New Delhi-110012
Email: mandal.stat@gmail.com

The class of balanced bipartite block (BBPB) designs is an important class of designs for comparing several test treatments with several standard or control treatments. Here, we present an algorithm to construct A-optimal BBPB designs given the number of test treatments (v1), number of control treatments (v2), number of blocks (b), and block size (k). We illustrate the construction of A-optimal BBPB designs using the proposed algorithm with an example. We obtain a list of A-optimal BBPB designs in the parametric range v1 ≤ 30, v2 ≤ 3, b ≤ 50, k ≤ 10. Many of the designs are possibly new.

Balanced Bipartite Generalized Row-Column Designs

Anindita Datta[1]***, Seema Jaggi**[1]**, Cini Varghese**[1] and **Eldho Varghese**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi – 110 012. India
*Email: aninditaiasri@gmail.com

This article deals with the generalized row-column design setting when there are two sets of treatments that are disjoint, one set consisting of test treatments and the other of control treatments. The two sets are disjoint in the sense that there are no common treatments between the two. The interest here is to estimate the contrasts pertaining to test treatments vs. control treatments with as high precision as possible. Balanced Bipartite Generalized Row-Column (BBP-GRC) designs have been defined and series of BBP-GRC designs for comparing a set of test treatments to a set of control treatments have been developed. The designs obtained are variance balanced in the sense that all the contrasts among test treatments are estimated with same variance and all the contrasts pertaining to test vs. control are estimated with the same and less variance.

On the Relationship between the Continuous Linear and Complete Reducible Orthogonal Representation

Diah Junia Eksi Palupi
Mathematics Department, Gadjahmada University, Yogyakarta, Indonesia
Email: diahju55@yahoo.com

A continuous linear representation is a homomorfism from a topological (G, μ) into $(GL_c(V), \tau_{GL})$ that is a set of bijection transformation which are continuously on topological vector space V. A continuous linear representation is said an orthogonal if on the topological vector space (V, σ) there is a simetric bilinear function that is a positif definite, invariant under the continuous linear representation. Futhermore, a continuous linear representation is said complete reducible if every invariant topological subspace U of V under the continuous linear representation, there is a topological W that is invariant under the continuous linear representation such that $V = U \oplus W$. Every orthogonal representation is complete reducible.

Modified Champernowne Transformation for Nadaraya-Watson Estimator

Kartiko
Statistics Department, Institute Science and Thechnology AKPRIND, Yogyakarta, Indonesia
Email: ook_kartiko@yahoo.com

In terms of the general nonparametric regression, Nadaraya-Watson estimator, depends on the whole observed sample $\{X_i\}_{i=1}^{n}$ through the kernel density estimation $\hat{f}_h(x)$. Kernel density estimation has a boundary problem, when it has limited supporting data. This condition will also affects Nadaraya-Watson estimator. This paper will propose a kernel transformation using modified Champernowne transformation as a substitute of the kernel in the Nadaraya-Watson estimator. A simulation study will be performed to make a comparison between nonparametric regression using the original Nadaraya-Watson estimator and its modified version.

Estimation of Missing Values in Experimental Designs using EM Algorithm

Kuncham Srinivas[1]* and **N.Ch. Bhatra Charyulu**[1]
[1]Department of Statistics, University College of Science, Osmania University, Hyderabad, Telangana-500007, India
*Email: kunchamsrinu86@gmail.com

Even the experiment is conducted in well planned manner some observations / response values are missing, then the designed experiment are no longer balanced and loses its symmetry and cannot proceed further to carry out the analysis. In some experiments the missing observation(s) affects the distribution of the remaining ones, and the distribution of the observed random variables of the model is the same with or without the missing observations. The most common approach to solve this problem estimate the missing values and placing in the proper place. Several authors made attempts on estimation of missing values in various experimental designs.

In this paper an attempt is made to estimate the missing observations using Expected Maximization algorithm for any experimental design model. Let $Y = X\beta + \varepsilon$ be a general linear model for any experimental design, where Y be the vector of responses, X be the design matrix, β be the vector of parameters and ε be the vector of random error. Assume the responses follows N (μ, σ^2). Partition the model with respect to the known and unknown response values. Recursive normal equations for the parameters μ, σ^2, β and Y_m can be obtained to estimate the same parameters and missing observations using expected maximization algorithm. The properties of estimated parameters and missing values are presented with suitable illustrations.

Data Analysis of Chlorophyll Fluorescence for Water Stress Detection in Tomato Plants

Manish K Patel[1*], **J. Adinarayana**[1] and **Surya S Durbha**[1]
[1]Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India
*Email: manishpatel9074@gmail.com

With the advent of new tools and techniques, it is much easier to understand the crop damage caused by biotic and/or abiotic stresses. It is possible to detect the stresses, which alters the crop Photosynthetic activity, afore the significant reduction in chlorophyll content and leaf area index appear in the visible spectrum of electromagnetic radiation as photosynthesis is inversely cognate with the chlorophyll fluorescence (CF). This research work aims to detect water stress, induced due to the deliberately engendered water deficiency in the tomato plants, by data analysis of solar induced chlorophyll fluorescence emission spectrum (CFES). CFES was generated by using FluorMOD model and Inverse PROSPECT model. The reflectance data of tomato plants, collected by ASD FieldSpec3 spectro-radiometer for 15 days, has been used into Inverse PROSPECT model and their biophysical parameters have been retrieved. Further, the artificial solar spectrum is gathered on per day basis from PV Lighthouses solar spectrum calculator by inputting atmospheric parameters. The study unveils significant difference in indices like F685+F730, F690/F740, F684/F700, (F684*F735)/F700 and F684*F700*F735 with time for healthy and stressed plants. Strong correlation; r=-0.92, r=-0.73, r=-0.74 for F685/F730; r=-0.94, r=-0.93, r=-0.96 for F690+F740 and r=-0.80, r=-0.79, r=-0.93 for F684/F735 in Cab content of healthy, moderately stressed and stressed plants respectively, have been unveiled. The results from the integration of FluorMOD and Inverse PROSPECT model show a good demarcation between the healthy and stressed tomato plants. The PROSPECT model was not able to accommodate the water stress condition as the changes in chlorophyll content and leaf water content was minimal in the time series analysis for 15 days. The FluorMOD model is able to predict only the shape of the CF spectra, not the magnitude. In future, the precision agriculture would be based on CF spectra since this is directly related to and more sensitive to the changes in photosynthetic activity. The stress detection will not only be based on CF alone but also to incorporate the heat dissipation to increase stress detection and prediction efficacy in manifold as energy in photochemical and non-photochemical process remain constant for a specific set of variables such as plant's health and environmental condition.

Genotype × Environment Interaction and Identifying High Yielding Stable Lines in Chickpea under Afghanistan Environments

Lina Mohammadi[1*], **Yashpal Saharawat**[1], **Darya Khan Akbarzai**[1], **Abdul Rahman Manan**[1], **S. Tavva**[1], **Swain Nigmananda**[1] and **Murari Singh**[1]
[1]International Center for Agricultural Research in the Dry Areas (ICARDA), Kabul, Afghanistan
[*]Email: lina.mohammady@yahoo.com

To meet increasing demand of protein for the growing population in Afghanistan, winter-sown chickpea is an important contributor and its productivity needs to be increased. Therefore this study was conducted to identify high yielding genotypes with wide adaptation and suitable for winter season in Afghanistan. Three field experiments, denoted as Chickpea International Elite Nursery – winter-sown (CIEN-W), Chickpea International Adaptation Yield Trial (CIAYT), and Winter Type Chickpea Yield Trial (WTCYT), were conducted in alpha designs at three out of five provinces, Baghlan, Herat, Kabul, Mazar and Nangarhar, during 2014-15. The result showed that, in CIEN-W trial, the genotypic variation was highly significant in all the three locations but the genotype x environment interaction (GEI) was low (P-value = 0.07). In CIAYT trial, the genotypes variation was significant at two of the three locations; also the GEI was non – significant, while in WTCYT trial, the GEI was significant.

Stability analysis of the genotypes from three locations, an extremely small number, showed that in CIEN-W trial, ILC482 with mean yield of 2.23t/ha was the highest yielding genotype and stable with slope 1.75±1.20 (P=0.65), specifically adapted to Baghlan based on a GGE-biplot, followed by FLIP09-423C with mean yield of 1.91 t/ha (slope of 0.89 ± 0.68, P=0.89) and FLIP88-85C with mean yield of 1.87t/ha. In trial CIAYT, the FLIP98-121C with mean 1.91 t/ha was the top yielding genotypes then FLIP93-58C with 1.86 t/ha and WTCYT trial, FLIP07-34C was the high yielding genotypes with mean 1.87t/ha followed by the location check with mean 1.80 t/ha. The selected high yielding materials will be further evaluated in replicated trials at much larger number of diverse locations in Afghanistan.

Super Saturated Designs for Mixed Factorial Experiments

Kashinath Chatterjee[1], **Anindita Datta**[2] and **Mohd Harun**[2*]
Department of statistics, Visva-Bharti University, Shantiniketan, India; [2]ICAR-Indian Agricultural Research Institute, New Delhi, India
[*]Email: harun20264@gmail.com

The purpose of any design involving factorial experiment is to provide a set of level combinations, but with increasing number of factors it is not possible to have a complete factorial and in this case we go for fractional factorial. There may be situations like factor screening experiments and many problems of scientists and engineers regarding primary stage of scientific experiments where the number of runs are even smaller than what needed to estimate the main effects. A super saturated design is a design involving factorial experiments that does not have enough runs to estimate all the main effects but it can be used as cost-effective designs to identify the few potential or active factors from a large set of factors. There are two main area in which research has been done regarding supersaturated designs; construction of E (s^2) optimal super saturated designs and improving the existing lower bound. In this paper a review work of existing methods of construction of super saturated design and various lower bonds for the designs has been done. Besides, a new concept of super saturated designs involving mixed factorial experiments has been introduced in this paper. The existing lower bound for any super saturated designs involving symmetrical factorial experiments $E_g(s^2)$, has been extended for super saturated designs involving mixed factorial experiments and thus a new lower bound has been derived.

ANCOVA Model for Agricultural Field Experiments with Split Plot Set-Up

Dikeshwar Nishad[1*], **Anurup Majumder**[1] and **Hiranmoy Das**[1]
[1]Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya
Mohanpur, Nadia – 741 252 West Bengal
*Email: dikeshwarnishad89@gmail.com

Analysis of covariance is an analysis technique which consists of analysis of variance and regression technique, together. In analysis of covariance, measurements of character of primary interest of the experimenter is recorded for analysis as is done for analysis of variance (ANOVA), in addition measurement of one or more characters are also recorded for analysis. These additional character (or characters) or variable (or variables) is (are) known as 'concomitant' variable (or variables) or simply covariate (or covariates). Here, the variable of primary interest is the yield or response (y) and the additional variable (or variables) (x) is the covariate (or covariates). The analysis of experimental designs by the above model has been used by several workers. But the application of the method in agriculture is still very much restricted to single factor experiments. However, some authors had discussed the method of analysis of covariance in multi factor experiments, especially in split plot set up, e.g. Federer and Meredith (1992). The present paper analyses the experiment on Linseed crop in split plot set up using the yield of just preceding rice crop as covariate. The experiment was repeated for consecutive two years 2014-15 and 2015-16. In split plot design for ANCOVA model, two error terms and two regression coefficients are involved for two levels of blocking. Here the experiment has been analyzed by following the method given by Federer and Meredith (1992) with two separate regressions of the dependent variate on the concomitant variable at two different levels for the whole plot factor and the sub plot factor. However, some authors assumed that there exists only one regression coefficient for a split plot design with single covariate, e.g. Bartlett (1937) used only the sub plot regression to adjust all sum of squares (including whole plot). This piece of study examines the methods mentioned above on the basis of reduction of error mean squares, F ratios, relative gain in efficiencies etc. by following the procedure as shown by Truitt and Smith (1956). Results of the analysis of the present study reveal that using two separate errors are not always better.

Nonparametric Bootstrap Approach for Constructing Prediction Intervals of VAR Model

Mrinmoy Ray[1*] and **K. N. Singh**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: mrinmoy.ray@icar.gov.in

Time series is a sequence of observations taken sequentially in time. Understanding, description of the generating mechanism and the forecasting of future values based on past value are the prime objectives of time series. Vector Autoregressive (VAR) models are one of the most important time series models used in forecasting over the past several decades which is an extension of univariate autoregressive model to multivariate time series data. One of the assumption of this model is that errors follow multivariate normal distribution and if this assumption is violated it is observed that the prediction intervals become width with the increase of forecast horizon. In this article, to overcome this problem a nonparametric bootstrap approach is proposed to construct the prediction interval of VAR model. Proposed method was implemented in simulated as well as real datasets. The proposed method was compared with existing approach results showed that the proposed approach perform superior as compared to the existing approach.

On the Generation of Cost-effective Response Surface Designs

Eldho Varghese[1]***, Arpan Bhowmik**[1]**, Seema Jaggi**[1] and **Cini Varghese**[1]
[1]Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi-110 012.
*E-mail: eldhoiasri@gmail.com

Run order consideration for mixed factorial, fractional factorial and confounded factorial have been studied by several authors in depth, but is lacking for Response Surface Designs (RSDs) except the results obtained by Quinlan and Lin (2015) for Plackett-Burman Design, a commonly used first order response surface design for screening purpose. Second Order Response Surface Designs (SORDs) are used to explore relationship between the response variable and the input variables and to find out the optimum input combinations to achieve a desired response. In this paper, we aim to find out optimal run orders with respect to minimizing level changes in SORD. Minimizing the level changes implies the minimization of experimental cost. Generation of four classes of designs viz., Plackett-Burman Design, Cost-effective Central Composite Design (CCD) with full factorial as well as fractional factorial points and Cost-effective Box-Behnken Design have been described through Macros developed using SAS IML.

Assessment of SPI Drought Class Transitions using Markov Chain and Log-Linear Models for Bundelkhand Regions of India

C. Jana[1*], **N.M. Alam**[1] and **G.C.Sharma**[1]
[1]ICAR-Indian Institute of Soil and Water Conservation, Dehradun, India
[*]Email: chayna_4503@yahoo.co.in

Present study attempts to analyse 20th Century rainfall of Bundelkhand, a drought prone region of India to model drought class transitions using markov chain model and 3-dimensional log-linear model. Standardized Precipitation Index (SPI) series were computed in a 12 month time scale for the 13 districts of the study region to derive the draught class transitions. Total study period was divided into three periods of 33/34 years to examine the differences related to drought class transitions using log-linear modelling. Markov chain modelling provides the details of the drought class probabilities, which represent proneness to drought formation; the expected residence time in each class of severity; the recurrence time to a particular drought class and the expected time for the SPI to change from a particular class to another. From the SPI series it is clear that all the 13 districts were often affected by moderate to severe droughts. Result from non-homogeneous Markovian chains is more essential since it reflects the seasonality of precipitation and how it affects the potential retreat of drought. Quasi-association and Quasi-symmetry log-linear models have been fitted to the drought class transitions derived from SPI-12 time series. The estimates of odds and respective confidence intervals were obtained in order to understand the drought evolution and to estimate the drought class transition probabilities. However, it is found that the probabilities for the non-drought and mild drought classes have little differences between the 13 districts. For the initial months, the calculated odds show the lower value as the drought severity increases and the same decreases in the subsequent months. This signifies that with the increasing of drought severity of the current class, the odds of a transition to the non-drought class versus a transition to any drought class decrease. Conversely, 3-dimensional log-linear model shows that the drought probability has been increased for almost all the 13 districts of the Bundelkhand region. These statistical characteristics of Markov chain provides information which can be used for drought hazard assessment. The overall results may act as essential components for monitoring the transitions drought occurrence and provide early warning to better cope up with the adverse situation.

Pre Harvest Forecasting Of Crop Yield with Non-Linear Modelling: A Concept

Sanjeev Panwar[1*], **K N Singh**[2], **Anil Kumar**[2], **Susheel Kumar Sarkar**[2] and **Bishal Gurung**[2]
Indian Council of Agricultural Research, Krishi Bhawan, New Delhi, India
*Email: scientist1775@gmail.com

This paper deals a concept of pre-harvesting of crop yield using nonlinear growth models and use of non-linear regression analysis for developing Wheat yield forecast model for Uttar Pradesh. A novel approach attempted in this study to use nonlinear models with different weather variables and their indices and compare them to identify a suitable forecasting model. For that, trend analysis has been done through linear and non-linear approaches. In which for each weather variable two indices have been developed, one as simple total of values of weather parameter in different weeks and the other one as weighted total, weights being correlation coefficients between detrended yield and weather variable in respective weeks. Weather indices based regression models were developed using weather indices as independent variables while detrended yield (residuals) was considered as dependent variable. The approach provided reliable yield forecast about two months before harvest.

Pilot Study for Developing State Level Estimates of Crop Area and Production on the Basis Of Sample Size Recommended by Prof. Vaidyanathan Committee - Observations on Field Data Collection Work

KK Tyagi[1], **G Mohan Rao**[1*], **Kaustav Aditya**[1], **Hukum Chandra**[1] and **UC Sud**[1]
[1]Indian Agricultural Statistics Research Institute, New Delhi, India
[*]Email: gmrao_56@rediffmail.com

During the year 2009, Directorate of Economics & Statistics (DES), Department of Agriculture & Cooperation (DoAC), Ministry of Agriculture (MoA), Government of India (GoI) constituted a Committee under the chairmanship of Prof. A. Vaidyanathan for reviewing methodologies used in TRS/EARAS/GCES/ICS in the preparation of crop statistics and suggesting institutional framework for improving the system of Agriculture Statistics in India. The Committee submitted its report during 2011 concluding that deficiencies in the existing system are not due to its design but due to its poor implementation and a large scale of efforts involved. Therefore, for generating State and National level estimates, the Committee recommended to conduct sample survey with smaller sample sizes to be conducted through an independent agency. The committee also suggested a sample of size 15,000 villages for estimation of crop acreage and 90,000 CCEs for estimating yield rates of major food grain crops respectively. In view of these recommendations, DES, DoAC, Ministry of Agriculture & Farmers Welfare (MoAFW), GoI approached ICAR-IASRI, New Delhi for looking in the adequacy of the sample sizes recommended by the committee, by undertaking a sample survey in few States.

Accordingly, ICAR-IASRI had prepared a project, funded by DoAC, MoAFW, GoI with the specific purpose of examining the reliability of estimates of crop area, yield, and crop production at State level on the basis of sample sizes recommended by Prof. Vaidyanathan Committee. The pilot survey under the project has been conducted in Assam, Karnataka, Orissa, and
Uttar Pradesh States, for collecting requisite data on area and yield for both Kharif and Rabi seasons of Agricultural Year (AY) 2015-16, while for Gujarat State, it could not be conducted during AY 2015-16 because of some unavoidable reasons, however, in Gujarat State, the data is being collected during AY 2016-17.

The data received so far indicates that there are omissions and commissions in collecting data, recording entries in the schedules and also in data entry. The data is being cleaned by removing different kinds of errors observed, in order to do statistical analyses for generating estimates of area, yield and production of major food grain crops at District and State levels.

Recent Advances in Statistics (Session-III)

Evaluation of Alternative Nonlinear Mixed Effects Models for Estimating Pig Growth Parameters

Pankaj Das
ICAR- Indian Agricultural statistics Research Institute, New Delhi-110012
Email: pankaj.iasri@gmail.com

The variability of growth rate is important aspect for meat producing animals like pig, goat, sheep etc. Animal growth models are used to identify alternative strategies to improve the efficiency of livestock production and to estimate daily nutrient requirements for the animal of different age and sex group. In literature it has been found that nonlinear growth model *i.e.* the fixed effect models are used to define the growth pattern of animals. In this study the efficiency of nonlinear mixed effects models were explored and a comparison was made between the predictive power of fixed effects models and the mixed effects models. 300 Body weight (BW) data including male and female pigs were used for model fitting. One pig specific random effect was included in each of the models. The random function was a random deviation of mature BW of the individual from average mature BW of its genotype. Logistic, Gompertz and Von-Bertalanffy fixed and mixed models have been explored for these data. The goodness of fit statistics like i.e. Mean Square Error (MSE) and Root Mean Square Error (RMSE) of the fitted models has been computed. It has been found that Logistic mixed effects model is performed better than the other nonlinear mixed effects models.

Communicating Science to Farmers: A Case Analysis of Direct to Farm

Reshma Gills[1]***, J P Sharma**[2] and **Ravi Khethrapal**[3]
[1]Indian Agricultural Research Institute, PUSA Campus, New Delhi-12; [2]Indian Agricultural Research Institute, PUSA Campus, New Delhi-12; [3]Strategic Science Partnerships, South Asia, CABI, NASC Complex, New Delhi – 110012
*Email: reshma1818@gmail.com

Direct2Farm is one of the flagship programmes of CABI. It is a mobile-enabled agriculture informatory service aimed at making high quality information readily accessible to farmers, thus empowering them to solve their everyday farming problems through a series of operations that turns factsheets into short SMS and voice messages that are delivered straight to smallholder farmers via mobile phones. A case study was carried out to find the strength, opportunities, weakness and threats of the D2F programme. D2F content for covering the different aspects of the selected crops is being prepared using authentic sources and validated by experts retired from IARI working as consultants with CABI. Dissemination is done as per crop calendar and weather bulletins received from Indian Meteorological Department. SOWT analysis of D2F showed opportunities and strength of mobile technologies for data dissemination along with some weakness and threats like, lack of any scientific methodology for the beneficiary selection, difficulty in developing and validation of content in the colloquial language etc.

Design of Field Phenotyping Experiment with Data Assimilation Techniques for Identifying Water-Stress Resistant Breeds

Soumyashree Kar[1]***** and **Adinarayana J**[1]
[1]Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India
*Email: ksoumya2301@gmail.com

High Throughput Plant Phenotyping (HTPP) is an indispensable fast, time-saving, non-destructive and automatic (or semi-automatic) method for understanding the dynamics among the plant, water, soil and atmosphere that result in different plant traits in different growth stages, to be able to focus on high-yielding, stress-tolerant and disease-resistant breeds. Such an initiative towards sustainable agriculture demands thorough analysis of the interplay and establishment of a niche that can parametrically define (and model) a phenotype. Since the conceptualization of the phenotyping process, there has been constant endeavor in evolving towards field phenotyping from lab phenotyping. While several studies focus on modeling various phenotypic behaviors, several other experiment with different platforms that can both scale up and help understand the regional phenology better. Whatever be the approach, the idea behind all these studies has always been finding the best stress-resistant breed that can yield high; the most important being, selection of water stress tolerant plant variety particularly in the semi-arid ecosystems. The breeding and agronomic programs world-wide are concentrating on finding ways to improve water-use efficiency (WUE) as plant water stress induces stomatal closure, which leads to decreased photosynthesis and growth rate, further resulting in heat stress due to reduced transpirational cooling. Hence, this review emphasizes on discussing the approaches adopted by researchers to study plant water stress as a cue to identify water-stress tolerant species through phenotyping and thereby proposing a design and methodology for implementing HTPP. Historically, the use of HTPP technologies have alleviated the impediment to simultaneous, effective and rapid evaluation of expressed phenotypes under field conditions, by exploiting both datadriven and process-based methods. The experimental set up and design in such studies aims at early identification of plant water stress using certain platforms and sensors or cameras [mostly Thermal, Infra-Red (IR), IR-Thermal, Hyperspectral and RGB]. For HTPP (in field conditions) the unmanned armed vehicles (UAVs) have best served the purpose by providing timely data with customized temporal, spectral and spatial resolutions. Hence, the proposed methodology tries to understand plant phenology for identifying plant water stress using UAVs, RGB and hyperspectral cameras. While it has been observed that RGB cameras have mostly been implemented for deriving metrics from visible morphology, the physiological processes could be circumvented using hyperspectral cameras. The importance of clustering techniques, statistical analysis and multivariate-relationship modeling for such studies has also been defined. This review also focuses on excerpts that state - for phenotyping purposes, the relative differences among genotypes are more important than the absolute data (and hence calibration). This therefore draws more attention towards iterative modeling and thresholding at each phenological stage until the least validation error is obtained. The design of such an experiment that integrates statistical and data assimilation techniques for modeling relationships (among plant, water, soil and weather) can help find both dominant and latent morphological as well as physiological parameters that govern plant behavior, both visibly and internally.

Geographically Weighted Regression Methods with a Tricube Kernel Weighting Function

Sri Haryatmi Kartiko[1]* and **Sifriyani**[1,2]
[1]Mathematics Department, Gadjahmada University, Yogyakarta, Indonesia; [2]Mathematics Department, Mulawarman University, Samarinda, Indonesia;
*Email: s_kartiko@yahoo.com

Geographically Weighted Regression incorporates local spatial relationship into the regression framework. This paper discuss the model estimation by likelihood method, with the weight that vary spatially. The GWR model is yi = β0 (ui, vi) + $\sum_{k-1}^{p} \beta_k$ (ui, vi) Xki + εi, where it needs the dependent variables, the independent variables and the variables describing location to calibrate. Here tricube kernel weighting function is proposed. The result is then applied to model the Human Development Index Data in East Kalimantan of Indonesia.

Rank Based Simultaneous Index for assessing Stable Crop Varieties with high yield

Prakash Kumar[1,*], **B.M.K. Raju**[2], **L. M. Bhar**[1], and **A. K. Paul**[1]
[1]Indian Agricultural Statistics Research Institute, New Delhi-110012; [2]Central Research Institute for Dry-land Agriculture, Hyderabad - 500 059
*Email: prakash289111@gmail.com

India is a developing country where, large numbers of small and marginal farmers have small land holding. The farmer's risk can be minimized through selection of stable genotypes and improve their economic condition by using a suitable simultaneous index which includes both stability and performance. In the present study an attempt has been made to identify/ develop a reliable simultaneous index for selecting genotypes stable to environmental variation with more yields. Evaluation of genotypes on the basis of stability and high yield is essential for yield trials in different environments. Though large numbers of simultaneous measures are available in literature, the problem of farmers has been to decide which of the stability measure is to be chosen for the purpose of selecting genotypes yield with stability. Here we used ranked based approach to deals with interaction residual having normal/non-normal distribution of quantitative/qualitative characters. Here aim of this study is to develop a composite index (simultaneous index) for selecting stable genotypes with high performance.

Identification of Outliers in incomplete Multi-Response Experiments

Raju Kumar[1*] and **L.M. Bhar**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi
[*]Email: raju.iasri@gmail.com

Cook-statistic has been developed for detecting any t outliers in incomplete multi-Response experiment when p (=p1+p2+ p3) response variables are observed from n experimental units of a block design D (say). These n experimental units are further divided into three sub-designs namely D1 comprising n1 experimental units, D2 comprising n2 experimental units and D3 comprising n3 experimental units such that n1+ n2+n3 = n. From n1 experimental units, all p response variables are observed, p1+p2 response variables are observed from n2 experimental units and p3 response variables are observed from n3 experimental units. Outlying observations, as we assumed that t in number may occur in D1 or in D2 or in D3. Or they may occur in all sub designs or any two sub designs. Again, an outlier means that all observations pertaining to all response variables are outliers, i.e, we will have t vectors of outliers. These vectors may be either of p component or (p2+ p3) component or p3 component. Cook-statistic has been developed keeping this in mind. Developed Cook-statistic has been applied to real experimental data for identifying outliers, if any. It has been shown that presence of outliers can distort the overall conclusion from an experiment.

Massive Open Online Courses (MOOCs) and Dynamic Learning: Experiences from ICRISAT

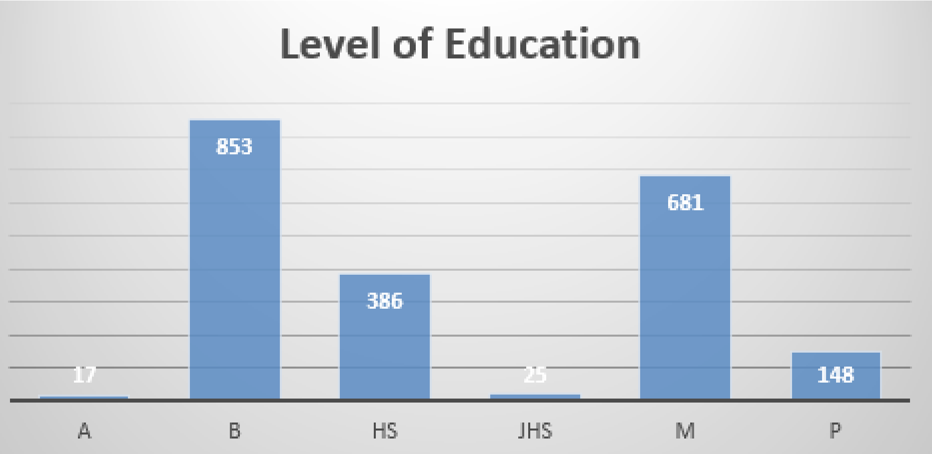
VV Sumanth Kumar[1]***, Nilesh Mishra**[1]**, Kanika Singh**[2] and **T Srinivasa Rao**[3]
[1]ICRISAT, Hyderabad-502324, Telangana, India; [2]Centre for Research on Innovation and Science Policy (CRISP), Banjara Hills, Hyderabad-500034; [3]Vivisol Technologies Pvt Ltd, Hyderabad, India
*Email: V.sumanth@cgiar.org

Agriculture education along with other complex factors plays an important role that influence food security and sustainable agriculture development. Current and future scenario requires an improved quality and relevance of agricultural education to create a strong pool of human resources to make productive contributions. As per a survey on higher education by the Ministry of Human Resource Development (MHRD),Government of India percentage of students pursuing agriculture as an undergraduate/ post graduate degree or PhD remains amongst the lowest and some of key reason for this can be attributed to outdated curriculum, lack of good faculty, etc. (Makwana, 2013). The expectations of fast paced techno savvy learners demands the current teaching methods be improved by looking for ways to make agricultural education more attractive like other subjects and apparently the one way to do that could be by utilizing information and communication technologies (ICTs). Advances in ICTs have promoted enormous amount of Virtual Learning Resources, Learning Management Systems (LMSs) and most recently Massive Open Online Courses (MOOCs) to address the above challenge. All these demands for a systematic ICT mediated learning approaches and tools to revitalize global agricultural education system.

To work on this challenge scientists at ICRISAT made an attempt by launching its first MOOCs titled “Diseases of Horticultural Crops and their Management” on January 26, 2015, in 2 parts on an Indian MOOCs platform coordinated by IIT Bombay and ICRISAT. The course followed the outline of the curriculum of PATH 372 that is offered at the State Agricultural Universities in India. The topics of this course cover economic importance and distribution, symptoms, cause, disease cycle and management of diseases of important horticultural crops. The main aim of this MOOCs course is to share the best teachers across National Agricultural Research & Education System of India and to democratize the agricultural education system by open online courses free of cost for the benefit of masses. This follows “create once use many times” approach, which saves time of a teacher handling this course.

Learning’s from the first MOOCs

The first set of MOOCs ran in 3 parts: part I for all interested Indian nationals and the part II & Re-run of Part-I were focused specifically for the students from 4 agricultural universities of Telangana (PJTSAU), Andhra Pradesh (ANGRAU) and 2 universities from Maharastra (VNMAU & MPKV) along with general public. In total the MOOC witnessed registration of 2,220 learner participants across the country with participants ranging from research



scholars, graduates to junior high school students and progressive farmers. Among the total participants 348 (15.67%) have received Honour Code Certificate on successful completion of course and evaluation. Among the 2,220 learner participants 853 have Bachelor's, 681 have Master's, 386 have High school, 148 have Primary school, 25 have junior high school and 17 associate degrees/certifications.

Gender analysis depicts 1682 male participants and 482 female participants were registered with the MOOCs system and the platform has witnessed with participants from varied age groups. The MOOCs received feedback through Focused Group Discussions [FGDs] and online exit survey responses. The feedback is positive and very encouraging with responses from across the country and agricultural students are excited to see more of such MOOC on agriculture in future as well.

Way forward

Learning from the experiences of first MOOCs we can say that there was increased interest from the agricultural professionals who pursued the course and they wanted to see more such MOOCs on agriculture in future as well. Bringing the contemporary proven successful models to agricultural education will surely enable better knowledge sharing and learning exchanges. This will also address the challenge of exploiting the potential of ICT's in agricultural education and make agriculture education more interesting for the future tech savvy generation. There need to be more supportive policy environment and funding support for undertaking such initiatives and exploring the possibilities to recognize the MOOCs course as credit course in future.

View and structure of the MOOC platform

On Construction of D-Optimal Designs for Logistic Regression Model

Shwetank Lall[1]***, Seema Jaggi**[1]**, Eldho Varghese**[1]**, Cini Varghese**[1] and **Arpan Bhowmik**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi -110 012. India.
*Email: shwetanklall@gmail.com

Logistic model is one of the most extensively studied nonlinear model in biological and social sciences. Though different aspects of fitting logistic regression model for various situations are very well explored in literature, the aspect of designing experiments based on logistic model is sparse. Technically, constructing design for logistic model is difficult because of two reasons. First being the determination of number of support points and the second is solving the optimization problem based on specific model, optimal criterion and design objectives. In general, for a nonlinear model, one has to solve and construct optimal designs for combination of the model, criterion and objectives separately. Fortunately, recent theoretical proof for existence of de la Garza phenomenon in different nonlinear models has narrowed down the algebraic cumbersome. In the present study, logistic regression model to construct D-optimal designs has been explored. The number of support points has been obtained using Caratheodory's and de la Garza theorems. D-optimal designs have been generated using the modified Fedorov-Wynn algorithm.

Geographically Weighted Logistic Regression Model with a Gaussian Kernel Weighting Function

Nissa Meilani[1*] and **Kartiko**[2]
[1]Statistics Department, Islamic University of Indonesia, Yogyakarta, Indonesia; [2]Statistics Department, AKPRIND Institute of Technology, Yogyakarta, Indonesia
[*]Email: meilaninissa92@gmail.com

Geographically Weighted Regression Logistic (GWLR) Model is a developed GWR design to predict model of the data set that has a binary response variable which takes into account the spatial factor. As an expansion to the standard logistic regression, logistic GWR incorporates a set geographical location to the models, taking form as the following equation:

Logit $(P_i) = \log\left(\frac{Pi}{1-Pi}\right) = \beta_0 (u_i, v_i) + \beta_1 (u_i, v_i)x_1$ where β_{0i}, β_{1i} are local model parameters specific to location at (u_i, v_i) coordinate. Parameter estimation for this model is done by adding a location factor using Gaussian kernel weighting function and is solved using Newton-Rapson iteration since it does not have a closed form. The result of this research is used to model the data Kemiskinan in East Java province.

G X E Interaction Studies in Multi-Location Yield Data of Rice Hybrids by AMMI and GGE Biplot Analysis

Revathi Ponnuswamy[1]*****, **Abhishek Rathore**[2], **Anilkumar Vemula**[2], **Roma Rani Das**[2], **Arun Kumar Singh**[1,] **Divya Balakrishnan**[1], **Hari Prasad A.S**[1], **V. Ravindra Babu**[1] and **T. Ram**[1]
[1]ICAR-Indian Institute of Rice Research, Rajendranagar, Hyderabad, India; [2]Breeding Informatics Unit, ICRISAT, Patancheru, Hyderabad, India
*Email: revathi.ponnusamy@gmail.com

The All India Coordinated Rice Improvement Project of ICAR- Indian Institute of Rice Research, Hyderabad organises multi-location testing of elite lines and hybrids to test and identify new rice cultivars for the release of commercial cultivation in India. Data obtained from Initial Hybrid Rice Trials of three years were utilized to understand the genotype x environment interaction patterns among the test locations of five different agro-ecological regions of India. The combined analysis of variance and AMMI ANOVA for yield of rice hybrids was highly significant for GEI. The GGE biplots first two PC's explained 54.71%, 51.54% and 59.95% of total G + GEI variation during 2010, 2011 and 2012 respectively, whereas AMMI biplot PC1 and PC2 explained 46.62% in 2010, 36.07% in 2011 and 38.33% in 2012 of the total GEI variation. Among the seventeen locations tested in 2010, GGE biplot showed that Coimbatore and Maruteru were the two most diverse and discriminating locations with strong crossover interaction, whereas in 2011 out of eighteen location tested Jammu, Raipur and Ludhiana were identified as more discriminative locations and during 2012 among the 17 locations, Karjat, Raipur and Dabhoi were identified as more discriminative. The GGE biplot identified rice hybrids viz., PAN1919, TNRH193, DRH005, VRH639, 26P29, Signet5051, KPH385, VRH667, NIPH101, SPH497, RH664 Plus, TNRH222 as stable rice hybrids. The discriminative environments identified in different test years were Coimbatore, Maruteru, VNR seeds, Jammu, Raipur, Ludhiana, Karjat and Dabhoi. The AMMI1 biplot identified the adaptable rice hybrids viz., CNRH102, DRH005, NK6303, NK6320, DRRH78, NIPH101, Signat5050, BPH115, Bio452, NPSH2003 and DRRH83. The present study demonstrated that AMMI and GGE biplots analyses were successful in assessing GEI of rice hybrids in different environments and aided in identification of stable and adaptable rice hybrids with higher mean and stable yields. The results suggested that rice hybrids promotion must consider yield stability across test environments based on genotype by environment interaction effects along with superior mean yield over checks for identifying superior potential rice hybrids.

Near Resolvable Multi-Session Sensory Designs Balanced for Carryover Effects

Sumeet Saurav[1]***, Cini Varghese**[1]**, Eldho Varghese**[1] and **Seema Jaggi**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
*Email: sumeet4927@gmail.com

Sensory trials play a vital role in food and nutrition experiments in establishing certain sensory facts about agricultural/animal produce. To draw definite conclusion from the study, it is important to eliminate or minimize all sources of error and control all factors that may influence the inference. Hence, in addition to the potential sources associated with the preparation of the test products, variability due to measurement or assessment process, order effects, carryover effects and assessor fatigue are to be considered. An experimental design for sensory evaluation should be capable of accommodating all these variations. However, when there are a large number of products two operational constraints, viz. assessor constraint and preparation constraint, may limit the choice of experimental designs. Assessor constraint sets a maximum number of products that an assessor can evaluate within a session before onset of sensory fatigue and preparation constraint limits the number of products that can be prepared for a given session without loss of experimental control. Therefore, many times it may become necessary to split sensory evaluation into sessions. Here, a general method is developed based on initial sequences to construct designs for multisession sensory trials balanced for carry over effects. In the proposed designs, all panelists will have to evaluate only a subset of samples in each session and they will not have to taste the same product more than once during different sessions.

Genotype by Environmental Interaction and Yield Stability of Potato Cultivars for Differential Harvesting Stage under Tropical Conditions

Raja Shankar[1]***, Med Ram Verma**[1]**, P. C. Sathpathy**[1]**, L.M. Yadav**[1]**, Raj Kumar**[1]**, Zafar Ullah**[1]**, Rajiv Khaiwal**[1]**, R. K. Dubey**[1]**, Shambhu Kumar**[1]**, Dherender Singh**[1]**, M.R. Deshmukh**[1]**, Dharminder Verma**[1] and **P.M. Govindakrishnan**[1]

[1]All Indian Coordinated Research Project on Potato, ICAR-Central Potato Research Institute, Shimla-171001, Himachal Pradesh, India,

*Email: rajascientists@gmail.com

Potato (*Solanum tuberosum ssp. tuberosum*) cultivars grown under tropical conditions exhibit comparatively poor yields than the temperate conditions. In addition, a wider yield variation between tropical locations is the greater concern to increase yield potential across growing environments for which the cultivars recommended are expected to express a stable level of yield with minimal environment flexibility. The cultivars interaction with ten environments was assessed using 2-year field experiments carried out under All India Coordinated Research Project (Potato) across the country for evaluating tuber yield for 2 harvesting stage (75 and 90 DAP). The results of single site analysis revealed that the cultivars tested were significantly different among themselves for total and marketable yield at seven environments for 75 DAP and at eight environments for 90 DAP. The pooled analysis of variance over ten environments showed a significant difference for genotypic and environmental main effect for all the traits observed indicated the existence of genetic variability among the genotypes justified from the heterogeneity of environments. The significant difference observed for G x E interaction for total and marketable yield at 90 DAP, as against non-significance for the same traits at 75 DAP clearly indicated that the genotypes interacted with differential environments at later harvesting stage only. Portioning of G x E interaction in to linear (E + G x E) and non-linear (Pooled deviations) components reflected a highly significant difference for Environment (Linear) for all the traits studied strongly suggesting the real differences in the genotypes for regression over environmental means and the response of genotypes to environment was controlled genetically. Except total yield at 75 DAP, the other traits exhibited a linear interaction G x E due to its significance value suggesting the major portion of interaction was linear in nature and prediction over environments may be possible. The variances due to non-linear component (pooled deviation) were significant for total as well as marketable yield at 75 and 90 DAP indicating that the major components for differences in stability were due to deviation from linear function and therefore relatively unpredictable components of the interaction may be more important than the predictable components. The cultivar K. Pukhraj is proven early bulking and stable cultivar for total and marketable yield at 75 DAP and predictable due to its x, bi and S^2di values. Whereas the cultivars K. Khyati found to be a stable cultivar for total and marketable yield at 90 DAP across growing environments of tropical conditions.

Sudoku Square Design

Nivetha Sri
S.D.N.B Vaishnav College for Women, Chromepet, Chennai, Tamil Nadu, India
Email: nivi97vocal@gmail.com

Sudoku is a popular combinatorial puzzle. Sudoku puzzle has become a regular feature in many newspapers and magazines all over the world inviting readers from all walks of life to solve this entertaining and addictive combinatorial puzzle. The basic properties of a Sudoku as an experimental design have been presented by Saba and Sinha (2014). The measure of uniformity plays a key role in the construction of uniform designs. It is well known that LSD is viewed as an experimental design which goes beyond CRDs and RBDs in simultaneously eliminating external sources of variation in the experimental units in an ANOVA set-up. Likewise, Sudoku squares and their generalizations as experimental designs, go one step beyond LSDs and their generalizations, and provide an extra dimension of utility as experimental designs. Sudoku, as experimental designs, build upon the LSDs and accommodate one more component of variation through the concept and formation of ’internal blocks’. The peculiarity of a SDK lies in the fact that in addition to incorporating variations due to Row, Column and Treatment components it also includes an additional component of variation, referred to as ‘Internal Block Classification’. In the context of an agricultural experiment, for example, these internal blocks might represent’ differential fertility situations’, besides the two external sources represented by row and column components. Sudoku square can remove three sources of variation from experimental error. Sudoku square design may be applied to multifactor experiments also. A detailed statistical analysis is provided along with the underlying ANOVA Table for the designs based on Sudoku for agricultural data.

Reduction of Response Surface Design Model – Bayesian Approach

T. Deepthi[1]***** and **N.Ch. Bhatra Charyulu**[1]
[1]Department of Statistics, University College of Science, Osmania University, Hyderbad-500007, Telangana
*Email: deepthitogercheti@gmail.com

The factor-response relationship is called the 'Response surface' and the designs used for fitting the response surface are termed as Response Surface Design Model. When the degree and order of the polynomial increases, only few factors are significant and underlying phenomena of interest and it is possible to eliminate the insignificant factors which are not affecting the response much, results in reduction of size of the Response Surface Design Model. Hence the analysis of the Response Surface Design model is relatively in accurate in higher dimensions, which leads to problem of dimensionality.

In this paper an attempt is made to estimate the parameters of first and second order response surface design models in Bayesian approach and evaluated the probabilities for all possible models: M_1, M_2, ..., M_k (i.e. $k = 2^p$ models) to select the best subset of predictor variables for the models having highest posterior probability. The method is illustrated with suitable examples.

2k Factorial Experiments with Logistic Error Distribution

Sunil Kumar Yadav[1]*****, **Krishan Lal**[1], **Rajender Parsad**[1] and **V.K. Gupta**[1]
[1] I.A.S.R.I., Library Avenue, New Delhi-110 012, India
*Email: kr.sunil.yaad@gmail.com

In designed experiment, it is not always true that the error of the generated data follows the normal distribution which is one of the basic assumptions of analysis of variance. Under such situations, the maximum likelihood equations may not be linear and so are not solvable. The equations obtained from the first derivative of log likelihood function with respect to parameters under study do not yield the explicit solutions for the estimates due to non-linearity of the function. Solving these equations by iterations is indeed problematic for reasons of (i) multiple roots, (ii) non-convergence of iterations, and (iii) convergence to wrong values. Therefore, methods have been developed using the modified maximum likelihood estimates (MMLE) in which the maximum likelihood equations are linearized by using the Taylor's expansion and estimates of the parameters are obtained. In the present investigation, the error is assumed to follow logistic distribution and the model of factorial experiment is assumed to be fixed effect model and design considered is completely randomized design for equal number of observations per cell. The modified maximum likelihood estimates for 2^3 factorial experiments for all the effects (main effects and interaction effects) and the estimate of the error have been obtained. *F*-statistics have been developed for all the treatment effects for testing the significance of parameters. These results have been generalized for the factorial experiments with *k* factors each at two levels. The size of the test is computed using Monte Carlo simulation technique for different values of the parameter of the distribution.

Detection of Outliers in INAR (1) models

H. S. Roy[1*], **R. K.Paul**[2], **L. M. Bhar**[2] and **P. Arya**[2]
[1]ICAR-National Academy of Agricultural Research Management, Hyderabad, Telangana;
[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
*Email: himadriiasri@gmail.com

In many data analysis a large number of variables are being recorded or sampled. One of the important parts of data analysis is detection of outlaying observations, if any. Although outliers are often considered as an error or noise, they may carry important information. Detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification, biased parameter estimation and incorrect results. It is therefore important to identify them prior to modeling and analysis. In real world phenomenon, some situations often occur where observations are not continuous. Such situations in time series may be number of road accidents, number of patients hospitalized, number of phone calls, the number of products sold, etc. In agriculture, it may be number of pests emerged in a crop season, number of crops infected by a certain type of disease. In these situations, usual autoregressive (AR) model where errors follow Gaussian distribution is not applicable. Integer valued autoregressive (INAR) model is applied for prediction of count observations. INAR model of first order, contaminated with additive outliers is studied. Parameter estimation with Yule-Walker and conditional least squares (CLS) is also described. Finally the detection of outlier observations is done.

On the Study of Sample Sizes for Estimation of Area and Production of Food Grain Crops

KK Tyagi[1*], **UC Sud**[1], **AK Gupta**[1], **VK Jain**[1] and **Kaustav Aditya**[1]
[1]Indian Agricultural Statistics Research Institute, New Delhi, India
[*]Email: kkanttyagi@gmail.com

In India, estimates of yield rates of principal food and non-food crops are obtained on the basis of crop cutting experiments (CCEs) conducted in majority of States /UTs through General Crop Estimation Survey (GCES) conducted by State agencies following scientific techniques of random sampling.

At present, over 95 per cent of the total food grains production is estimated on the basis of yield rates obtained from the CCEs conducted on scientific basis spread over various States/UTs.

The Directorate of Economics and Statistics (DES) releases estimates of area, production and yield in respect of principal crops of food grains, oilseeds, sugarcane, fibers and important commercial and horticulture crops. The estimates of crop production are obtained by multiplication of area estimates by corresponding yield estimates. Therefore, the estimates of area and yield rates assume prime importance in the entire gamut of agricultural statistics.

Field Operations Division (FOD) of the National Sample Survey Office (NSSO) under the Ministry of Statistics and Programme Implementation (MoSPI), Govt of India (GoI) has the overall responsibility for providing technical guidance to States /UTs in developing suitable survey techniques for obtaining reliable estimates, assistance in training of staff and exercising supervision. Under the Improvement of Crop Statistics (ICS) Scheme which was taken up during 1973-74 with the objective of locating, through the joint efforts of NSSO and State Governments, the deficiencies in the system of crop statistics by exercising technical supervision over the primary field work and suggesting remedial measures for improving the system. To achieve this, sample checks on area enumeration, and area aggregation are carried out in a sample of about 10,000 villages in each season and on conduct of around 31,000 CCEs in a year.

Data pertaining to CCEs for different crops (having smaller sample sizes) pertaining to good number of States under ICS scheme were obtained from NSSO for AY 2010-11. Data files pertaining to different crops pertaining to these States were prepared and statistically analysed through computer program prepared in R. Estimates of average yield pertaining to various crops along with their percentage standard errors were worked out. It was observed that estimates of average yield for wheat and paddy crops were obtained with suitable degree of precision, however, for crops like maize, barley, jowar, ragi etc., these were obtained with high percentage standard errors. Sample sizes at the primary stage of sampling i.e. villages were worked out for estimation of average yield of different crops for different levels of margin of errors.

Data pertaining to area for different crops pertaining to different districts (20 survey numbers in a village) of some States for AY 2010-11 under ICS scheme were obtained from NSSO and analysed. Estimates of total area under different crops were obtained with very high percentage standard errors. However, these were also worked out had the sample size would have been increased from 20 survey numbers to 100 survey numbers. In that case, the percentage standard errors decreased significantly for number of crops.

The main finding in this study is that on the basis of smaller sample sizes, the estimates of average yields of paddy and wheat crops may be estimated with suitable degree of precision in those States in which these crops are being grown as major crop. However, for other crops, these sample sizes may not be adequate for estimating the average yields of these crops with suitable degree of precision. For estimating these with permissible margin of errors, the sample sizes for these crops would have to be increased suitably.

Recent Advances in Statistics (Session-IV)

Software to Analyse Performance of Rice Genotypes to Insect Pests

A.P. Padmakumari[1*], **B. Sailaja**[1], **G. Katti**[1] and **V. Ravindra Babu**[1]
[1]ICAR-Indian Institute of Rice Research, Rajendranagar, Hyderabad- 500 030
[*]Email: padmakumariento@gmail.com

All India Coordinated Rice Improvement Project (AICRIP) is a major programme in rice involving several locations all over the country for testing of various rice production. Every year numerous improved elite breeding lines are being screened for their reaction to insect pests both in field and greenhouse conditions at many locations across the country to select the best performing genotypes. The database is huge with above 50 locations and nearly 2000 breeding lines. The data includes various pest specific damages and population levels which is dependent on the pest incidence at a location. The input data would be the genotypes their cross combinations, IET Nos, designations, pest damage and the locations where they were evaluated. Till date the data was analysed in Excel format. Analysing this data for different trials is more cumbersome when the genotypes involved are more than 50 in a particular trial when the locations are more. A software package has been developed for a quick and efficient analysis of screening trials data on the reaction of genotypes to insect pests. The software was developed in Microsoft Access with database structure and visual basic (VB) script as a front end.

User-friendly screen prompts are available for entering multiple screening-set data, details of genotypes, locations and pests. Package output screen consists of different combo boxes to query for required information and command buttons to generate various reports. Provision has been created to send the generated reports to spread sheet format for the benefit of the user. This software contains modules for analysing reactions of test genotypes across different locations by generating reports on the performance of the genotype at a location against one or many pests, across locations for one or many pests which helps in identification of multiple pest resistance in the test entries, identifying the best entries and also serving as a database for future use. This software was validated on identifying the best performing genotypes using the phenotypic data generated from pest screening trials for 4 years from All India coordinated rice improvement project. Apart from analysis of the reactions to pests, this software can also be used to assess the data on reactions of genotypes to all abiotic or other biotic stresses with variation in the input data.

Mutations Associated with Genetic Instability in Peanut (*Arachis hypogaea* L.)

Hake A[1], **Yadawad A**[1], **Shirasawa K**[2], **Gowda MVC**[3] and **Bhat RS**[1]*
[1]Department of Biotechnology, University of Agricultural Sciences, Dharwad – 580 005, India;
[2]Department of Frontier Research, Kazusa DNA Research Institute, Chiba 292-0818, Japan;
[3]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad 580 005, India
*E-mail: bhatrs@uasd.in

Genetic instability resulting from hybridization followed by EMS mutagenesis displayed several unusual inheritance behaviors in an earlier study in peanut. VL 6, a mutant recovered from this study, continued to exhibit genetic instability in the advanced generations. We observed transposition of *Arachis hypogaea* miniature inverted-repeat transposable element (*AhMITE1*) leading to somatic mutations in VL 6. DNA profiling in VL 6 and its mutants showed hyper-activity of *AhMITE1*. Mutants from VL 6 showed polymorphism due to transposition at unique genomic sites, indicating that they were independent in origin. Genome sequence of the mutants were compared with that of VL 6 for the frequency of single nucleotide polymorphisms (SNPs) and variations in microsatellites or simple sequence repeats (SSRs) using the NGS big data generated by restriction site associated DNA-sequencing (RAD-Seq). Efforts are being made to identify any association between the mutant phenotypes and the DNA polymorphisms. However, the genetic and/or epigenetic factors contributing for the hyper-activity of *AhMITE1* need to be worked out.

Self-Organizing Mapping Networks (SOM) for Selection of Water Stress Tolerant Safflower Genotypes

C. Sarada[1]***, Lakshmi Prayaga**[1]**, K. Alivelu**[1] and **P. Lakshmamma**[1]
[1]Indian Institute of Oilseeds Research, Rajendra Nagar, Hyderabad
*Email: saradac@yahoo.com

Among multivariate data analysis techniques, clustering is a set of methods capable of creating groups (clusters or classes) of objects based on their degree of association. Clustering techniques are purely empirical classification methods relying upon inductive approaches, thus influenced by the nature of the input data to be analyzed. Errors during cluster analysis contribute to ineffective decisions. Choosing an appropriate clustering method for the given data set is a difficult task or even finding a most relevant approach for a given data set is very hard. The most commonly used grouping methods PCA and other classical clustering methods. All these classical clustering techniques perform better and give good approximation for well characterized data sets with compact, isolated clusters. Each method has its own advantages and disadvantages. But in real world situation the concerned empirical data such as crop genotypes data, does not confirm to the ideal conditions of distinct cluster structures required for accurate cluster definition. So an efficient clustering algorithm/method is to be chosen to overcome these limitations. In the present study an artificial neural network method viz., Self-Organizing Map (SOM) neural network clustering method is applied for grouping safflower genotype for water stress tolerance. This SOM clustering method can handle the data set with various levels of imperfections that include data dispersion, outliers, irrelevant variables, and non-uniform cluster densities very efficiently.

A lab experiment was conducted at Directorate of Oilseeds Research, Hyderabad to study the response of 28 safflower genotypes to water stress induced by PEG to delineate the tolerant genotypes from susceptible ones. The observations on germination percentage, Days to minimum germination, seedling vigour, for different stress levels were recorded. The genotypes germinated under high stress conditions also recorded. Thus the main aim of the experiment is to classify the genotypes based on these parameters in to different groups. Genotypes were grouped based on PCA as well as SOM. PCA grouped the genotypes into two groups good and poor genotypes, While SOM could group genotypes more distinctly when compared to PCA. It has formed eight groups which distinguished the genotypes as1. Very good germination with very good seedling vigour. 2. Good germination with good seedling vigour 3. Moderate germination with moderate seedling vigour 4. Good germination and moderate seedling vigour 5. Moderate germination with good seedling vigour 6. Poor germination and poor seedling vigour and 8. No germination and no seedling vigour.

Differential Response in Sheep and Goats to Virulent Peste Des Petits Ruminants Virus Infection: Key Antiviral Molecules and miRNAs Identified in PBMCs

Amit Ranjan Sahu[1], **Sajad Ahmad Wani**[1], **Alok Khanduri**[1], **Kaushal Kishor Rajak**[1], **Aditya Prasad Sahoo**[1], **Ashok Kumar Tiwari**[1], **Aruna Pandey**[1], **Sonam Kanchan**[1], **Shikha Saxena**[1], **D Muthuchelvan**[1], **Bina Mishra**[1], **Bishnu Prasad Mishra**[1], **Raj Kumar Singh**[1] and **Ravi Kumar Gandham**[1]*
1Computational Biology & Genomics Facility, Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, UP-243122
*Email: gandham71@gmail.com

Peste des petits ruminants (PPR) is one of the highly contagious viral disease, primarily affecting sheep and goat. It is characterized by erosive stomatitis, mucopurulent nasal discharge, pneumonia, gastro-enteritis, fetid diarrhea etc. Though the epidemiological data suggests that goats are more susceptible to PPR, reports on host susceptibility to PPR are contrasting and this variable host response has never been investigated at cellular and molecular level. In a regular PPRV vaccine potency testing experiment at IVRI, Mukteswar, Uttarakhand, blood samples were collected at regular intervals from the unvaccinated control group (both sheep and goats) infected with the Izatnagar/94 virulent virus. The animals developed symptoms and succumbed to death in 11 days post infection (dpi). High-throughput RNA Sequencing data was generated from PBMCs isolated from blood of sheep and infected with virulent PPRV at 9 dpi. A total of 3387 commonly differentially expressed genes (DEGs) were identified through DE packages - EBSeq, DESeq2 and edgeR using RSEM counts. Ingenuity pathway analysis (IPA) revealed B cell receptor signaling to be the top canonical pathway in both the species. Cellular growth and proliferation, cellular function and maintenance, cell death and survival, cellular development and cellular movement were the key molecular and cellular functions in both species. For all the DEGs, MIF and MyD88 governing the adhesion of immune cells and chemotaxis of myeloid cells respectively were predicted to be the key regulatory molecules that were activated in both the species. However, C3 governing the inflammatory response was predicted to be activated only in goat and MAPK14 governing adhesion of phagocytes was predicted to be activated only in sheep. miR - 19 and miR - 21 - 5p were identified to be the possible downregulated miRNAs leading to organism death and recruitment of cells in both the species, respectively. Differentially expressed highly connected (DEHC) genes (Fold change > 3 and degree > 5) were identified in goat (218) and sheep (226) on network analysis using customized Perl scripts. In the DEHC network, antiviral molecules - ISG15, IFIT1, IFIT3, IFIT5, MX1, and HERC5 were found to be upregulated in both the species with goat showing greater expression than sheep. The upstream regulators of these antiviral molecules that were dysregulated in our study in both the species were identified through IPA. DDX58, DOCK8, TLR7, STAT1, STAT2, CD14, STAT3, SOCS3, SOCS1, TRIM24, PTPN22 and TLR4; and STAT4, DNASE2 and IFN8 were among the upregulated and downregulated upstream genes in goat, respectively. DDX58, TLR7, DOCK8, SOCS1, SOCS3, IL1RN and MAPK1; and TRIM 24 and DNASE2 were among the upregulated and downregulated upstream genes in sheep, respectively. Further studies are being carried out to confirm the predicted differences between the two species.

Development of Marker Resources for Fine-Mapping of Late Leaf Spot and Rust Resistance in Peanut (*Arachis hypogaea* L.)

Gayathri M[1]***, Shirasawa K**[2] and **Bhat R S**[1]*****
[1]Department of Biotechnology, University of Agricultural Sciences, Dharwad – 580 005, India;
[2]Department of Frontier Research, Kazusa DNA Research Institute, Chiba 292-0818, Japan
*Email: gaya1293@gmail.com

Peanut exhibits very low DNA polymorphism owing to its narrow genetic basis and genetic bottlenecks. Developing a large number of new markers might help fine-mapping of important traits. In order to fine-map the resistance to late leaf spot and rust in peanut, an effort is being made to develop Cleaved Amplified Polymorphic Sequences (CAPS) and *Arachis hypogaea* miniature inverted-repeat transposable element (*AhMITE1*) markers. NGS big data obtained from whole genome re-sequencing of four diverse peanut genotypes were analyzed for single nucleotide polymorphism (SNP). Till now, restriction sites containing these SNPs were identified at 8,670 genomic regions belonging to B01, B02 and B03 chromosomes. Also, the genomic sequences were compared to identify the variation in the sites of *AhMITE1* insertion. Primers are being developed to amplify the region containing the polymorphic restriction sites and *AhMITE1* insertion sites from the parents of the various mapping population available at our laboratory for fine-mapping of late leaf spot and rust resistance in peanut.

Assessment of Vegetation Seasonality in Semi-Arid Tropics using Massive Archives of Satellite based Remote Sensing

Jayantrao Mohite[1]*****, **Suryakant Sawant**[2], **Minal Sawant**[1], **Rajesh Urkude**[1] and **Srinivasu Pappula**[1]
[1]TCS Innovation Labs, Tata Consultancy Services Mumbai, India; [2]Indian Institute of Technology, Mumbai, India
*Email: jayant.mohite@tcs.com

Climate change has huge impact on socio-economic and natural systems of semi-arid areas. Vegetation dynamics plays a crucial role in natural resources, land use planning and regional policy decisions. Increasing remote sensing platforms like satellites, airborne surveys, unmanned aerial vehicles, etc. facilitates the utilization of growing massive archives of remote sensing observations (big data) to understand the trends in land use / land cover and vegetation growth stages. The main objective of this study is to understand the phenological stages of major land use / land cover (i.e. forest, scrub forest and agriculture) in semi-arid region using multi-sensor (Landsat 5, 7 and 8) remote sensing time series observations. A study area considered for this work is Melghat tahsil of Amravati District, Maharashtra, India which comes under semi-arid region of central India. In this study freely available non-commercial satellite imagery store Google Earth Engine (GEE) is used to extract the time series of remote sensing observations over the selected area. The time series of vegetation indices derived from satellite based reflectance has been analyzed to estimate the phenological stages of forest, agriculture and scrub forest. Weather variables like rainfall and temperature were used to validate the estimated vegetation growth stages. The results confirmed time lag between vegetation growth stages and rainfall events which results into crop water stress in rain-fed agricultural systems. In addition to this, an assessment of surface water storage structures has been carried out using the time series remote sensing observations, to evaluate the need for micro water conservation structures. The proposed methodology based on time series of vegetation indices provides information on region specific subsistence irrigation schedules which attempts towards crop yield improvement. In future crowd source based application will be used to validate crop growth stages and to provide information on subsistence irrigation schedules.

Assessing the Impact of Drought Stress on Simultaneously Occurring Pathogen Infection in Field Grown Chickpea using Simulation Models

Ranjita Sinha[1], **Basavaiah Mohan Raju**[2], **Suganthi Angappan**[3] and **Muthappa Senthil-Kumar**[1]*
[1]National Institute of Plant Genome Research, JNU Campus, Aruna Asaf Ali Marg, New Delhi-110067;
[2]Department of Crop Physiology, University of Agricultural Sciences, GKVK, Bangalore, Karnataka - 560 065;
[3]Krishi Vigyan Kendra, Tamil Nadu Agricultural University, Virinjipuram, Vellore, Tamil Nadu, - 632 104
*Email: skmuthappa@nipgr.ac.in

Independently occurring drought and pathogen stresses reduce chickpea yield in the field. Owing to the nature of simultaneous occurrence of these two stresses, the severity of stress impact on crop growth and productivity is further aggravated. In this regard, the interaction of these two stresses, their individual and combined effect and impact on plant growth and productivity traits was assessed. We conducted field experiments for two consecutive years during 2015 and 2016 at three different locations in India and investigated the effect of different irrigation regime on pathogen incidence and the net effect of combined stresses on chickpea yield. Different irrigation regimes were maintained to impose mild to severe moisture stress. Drought stress levels were assessed as soil moisture content and the abilities of the plants to maintain the tissue water status through measuring the leaf relative water content. Diseases were identified based on symptoms and the pathogens were further confirmed by pure culture studies. Effect of drought stress on pathogen incidence was depicted as percent disease incidence. The drought stress incidentally increased the occurrence of three fungal diseases namely, dry root rot caused by Rhizoctonia bataticola, stem rot caused by Sclerotinia sclerotiorum and grey mold caused by Botrytis cinerea and the disease incidence correlated with stress level. Further, we used decision support system for agrotechnology transfer (DSSAT) simulation model to know the outcome of simultaneous occurrence of drought and pathogen stress on various physiological parameters. The simulation outcome was authenticated by comparing with actual field data obtained for a few selected yield attributed traits namely, specific leaf area and chlorophyll content. Overall, our study indicated that, the drought stress enhances the incidence of three fungal diseases and the combined stresses cause significant yield reduction than the individual stresses. In this regard, we propose a further systematic study to assess the effect of stress interactions and the combined effect and impact of abiotic and biotic stresses on chickpea yields.

Genome Wide Mining of Microsatellites in Sesame (*Sesamum indicum* L.)

P. Supriya[1]*, S. Sarika[1], R. Saurabh[1], AR Rao[1], and **KV Bhat[2]**
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi;
[2]ICAR-National Bureau of Plant Genetic Resources, New Delhi
*Email: puramsupriya@gmail.com

Microsatellites are short tandem repeat motifs that exhibit complex patterns in their abundance, reproducibility and are usually associated with high frequency of length polymorphism. Sesame (*Sesamum indicum* L.) is an important oil seed crop belonging to family Pedaliaceae. Genetic diversity studies in sesame are challenging due to limited number of polymorphic SSR markers available. In the present study, we developed a comprehensive information system of sesame SSRs consisting of information regarding SSRs, their PCR primers, physical map and a genome browser etc. from draft genome sequence of Swetha variety. The SSRs have been identified from sesame genome using MISA tool. The SSRs detected were classified into di, tri, tetra, penta, and hexa nucleotide repeats, and their distribution in genome was studied. The study reveals that the distribution of di, tri, tetra, penta and hexa nucleotide repeats on 16 putative linkage groups is 33%, 16%, 23%, 5%, and 3% respectively. Compound SSRs account for nearly 18%. The highest number of SSRs were located in linkage group 3 followed by linkage group 6 and 8. The least number of SSRs were found in linkage group 16. A similar pattern was observed for compound repeats also. Di nucleotide repeats were more abundant followed by tetra nucleotide repeats and the least abundant were hexa nucleotide repeats. In-house *perl* scripts were written to extract the flanking sequences of SSRs for designing primers. Batchprimer3 was used for designing primers from the flanking sequences with the criteria of GC content between 20-80%, primer length 18-30 bp and annealing temperature between 57-63 ^{o}C. Five sets of primers are provided in the database for the convenience of user. The primer information consists of forward and reverse primer sequences, GC content, melting temperature, length of primer and product size etc. The flanking sequences of SSRs are also provided in the database. We developed physical map of SSRs using whole genome linkage group wise assembly of sesame. We also developed genome browser of sesame which shows the distribution of microsatellites on the 16 linkage groups. The database developed will be highly useful in cultivar identification and diversity analysis so as to improve the genetic potential of the crop. The genome browser can be used to select markers across the linkage groups and to avoid possible linkages which generally result in ambiguous results.

Assessment of Genetic Divergence in Rice (*Oryza sativa* L.) Genotypes for Biometrical Characters

Parmar DJ[1]*, **Patil AS**[1], **Patel SG**[2] and **Makwana MG**[2]
[1]Department of Agricultural Statistics, BACA, AAU Anand – 388 110;
[2]Main Rice Research Station, AAU, Nawagam
*Email: djparmar.a@gmail.com

Genetic divergence analysis estimates the extent of genetic diversity which plays an important role in plant breeding either to exploit heterosis or generate productive recombinants. The choice of parents is of paramount importance in breeding programs. Thus, the knowledge of genetic diversity and relatedness in the genotypes is a pre-requisite for crop improvement programs. The aim of this study was to identify genetically divergent rice genotypes with desirable traits for hybridization particularly for yield. Sixty three genotypes of lowland rice varieties developed at various research stations in Gujarat were grown in randomized block design with 3 replications during *kharif* 2014. The analysis of variance revealed significant differences among the 63 genotypes for all the seventeen characters under studied. The quantum of genetic divergence was also assessed by cluster analysis using Mahalanobis's Euclidean squared distances. The Euclidean squared distance grouped the material into 18 clusters. Among the clusters, cluster I was the largest comprising of 20 genotypes. The intra cluster distance analysis revealed that the maximum divergence was observed in cluster V. Inter cluster D^2 value was the highest between cluster V and XVII (5890.36) while, lowest divergence was noticed between cluster VII and VIII (95.40). Thus hybrids developed from these genotypes may produce high magnitude of heterosis or desirable transgressive segregates for successful breeding of rice. The results showed that the contribution of iron content towards genetic divergence was highest followed by zinc content, grain length breadth ratio, plant height and grain breadth. Values of R^2 were found highest for iron followed by zinc, grain breadth, grain length breadth ratio and plant height indicated that these biometrical characters would contribute more in clustering of genotypes.

Genetics of Iron Deficiency Chlorosis Resistance in Groundnut

Santosh K. Pattanashetti[1,2]***, Gopalakrishna K. Naidu**[1]**, Prakyath Yashoo**[1]**, Omprakash Kumar Singh**[1]**,**
Basavaraj D. Biradar[1], and **Mruthyunjay C. Wali**[1]
[1]College of Agriculture Vijayapur - 586 101, University of Agricultural Sciences, Dharwad, Karnataka;
[2]Present Address: International Crops Research Institute for the Semi-Arid Tropics, Patancheru – 502 324, Telangana state, India
*Email: S.Pattanashetti@cgiar.org

Groundnut is an important food legume which is sensitive to Fe deficiency in alkaline and calcareous soils and displays iron deficiency chlorosis (IDC). IDC is prevalent in Gujarat, Maharashtra, Rajasthan, Tamil Nadu, and Karnataka states of India as well as in northern China and Pakistan causing considerable reduction in pod yield. Although application of Fe fertilizers in inorganic, chelated, and organic forms have been suggested for correcting Fe deficiency, they are either ineffective or economically not feasible because groundnut is mainly grown as a rainfed subsistence crop in the semi-arid tropics. Considerable genetic variability for IDC response has been reported earlier in groundnut. It is necessary to understand the genetics of IDC resistance trait in groundnut towards development of IDC resistant cultivars.

The present study investigated the genetics of IDC resistance among four crosses of groundnut. Towards this, four released cultivars of groundnut (Dh 86, TAG 24, GPBD 5, G 2-52) with varying degree of IDC susceptibility as female parents were crossed with IDC resistant male parent (ICGV 86031) to generate F1's of four crosses. The F1, F2 and F3 generations of four crosses were evaluated for IDC response in iron-deficient calcareous soils at College of Agriculture, Vijayapur during summer 2013-14, Rainy 2014 and Rainy 2015 seasons, respectively. IDC response was assessed based on associated traits like (i) visual chlorosis rating (1 to 5 scale: 1 – highly resistant, 5 – highly susceptible) proposed by Singh and Chaudhari (1993) based on severity and coverage of interveinal chlorosis in entire plant, and (ii) SPAD chlorophyll meter values. Higher VCR score and lower SPAD values indicate IDC susceptibility, while lower VCR score and higher SPAD values indicate IDC resistance. All the plants of parents, F1, F2 and F3 generations were scored for VCR and SPAD values across three stages i.e., 30, 60, and 90 days after sowing (DAS). Since IDC severity was highest at 60 DAS, VCR and SPAD values at this stage were considered for categorizing the plants as IDC resistant or susceptible. All the four F1's were resistant to IDC indicating the dominance nature of IDC resistance over susceptibility. The F2 populations of all the four crosses showed a good fit to the ratio of 15 (IDC resistant): 1 (IDC susceptible) indicating that duplicate dominant genes govern IDC resistance in groundnut. This was further confirmed based on the behavior of F2-derived F3 families (F2:3) among three crosses which showed perfect fit to the expected ratio of F2:3 families i.e., 7 (breeding true for resistance): 4 (segregating in 3:1 ratio): 4 (segregating in 15:1 ratio): 1 (breeding true for susceptibility), except the cross G 2-52 × ICGV 86031 which had less population size. The outcome of the present study will be useful in designing appropriate breeding methodology for developing high yielding IDC resistant/tolerant cultivars for adoption in the relevant ecologies.

Role of Soft Computing Tools in Predicting Plantation Crop Yields: A Case Study with Cashew

P D Sreekanth
ICAR-National Academy of Agricultural Research Management, Hyderabad, Telangana State
Email: sreekanth@naarm.ernet.in

Cashew is a very important and dollar earning crop for India. In the present study, cashew yield prediction model was developed for different states *viz.,* Maharashtra, Kerala, Andhra Pradesh, Karnataka and Tamil Nadu. Odisha, Goa and India as a whole by using soft computing tools with the help of secondary data. The performance of Elman Artificial Neural Network was examined for predicting cashew yield. The model efficiency and accuracy was measured based on the Root Mean Square Error (RMSE) and Regression Coefficient (R^2). The model provided the best fit and the predicted trend followed the observed data very closely. The accuracy of these models in predicting cashew production for different states is around 96%.

Estimation of Breeding Value using Longitudinal Data

U.K. Pradhan[1]***, P.K. Meher**[1]**, A.R. Rao**[1] and **A.K. Paul**[1]
[1]ICAR-I.A.S.R.I., Library Avenue, Pusa, New Delhi-110 012
*Email: upen4851@gmail.com

Now a day different type's data are collected by breeding associations for the purpose of genetic evaluation. These data are based on the biological characteristics of individual animals that can be measured at multitude times over an animal's lifetime commonly called as longitudinal study. Genetic analysis of longitudinal data that incorporate temporal variations is also important for understanding genetic architecture and biological variations in animal genetics. Longitudinal data are occasionally collected over time; hence the observations within subject may tend to be serially correlated. For analysis of longitudinal data, most works in the literature have assumed that the observations within each subject are uncorrelated. Again, idea of reducing the number of days required to reach their desired endpoint (culling) is not new. Given this economic relevance, genetic evaluation research for reducing these required days has received very little attention. A new methodology has been developed using random regression model (RRM) by taking into account the correlated errors. The variance components has been estimated and optimized using Average Information (AI) algorithm. Further the breeding value has been estimated for different heritability level.

Mapping Suitable Areas for Growing Pulses in Rice-Fallows using Multi-Criteria Spatial Decisions

Raj Kumar Singh[1]* and **Chandrashekhar Biradar**[1]
[1]International Center for Agricultural Research in the Dry Areas, New Delhi-110012, India
*Email: Raj.Singh@cgiar.org

Food legumes (Pulses) plays an important role in ensuring food, nutritional as well as environmental security through sustainable intensification and crop diversification in India and elsewhere. India being one of the largest producer of the grain legumes but its present production is not meeting the domestic demand and imports nearly 4 metric tons of pulses. There is a need to increase both production and productivity to bridge the supply-demand gap. One such untapped opportunity lies in the potential use of crop-fallows such as rice-fallows. The rice paddy fields in eastern and central India remain left fallow in the winter month due to several limitations. The most limiting factors are lack of understanding of the fallow dynamics, soil moisture regimes and suitable area specific to crops and crop varieties. Information on the spatio-temporal dynamics provide a window of opportunity for farmers to grow sort duration pulses like lentil, grasspea and chickpea during fallow periods which brings them additional income, feed for livestock plus addition nitrogen to soil. Therefore, a study has been undertaken to map and quantify the suitable area for growing pulses in rice-fallow periods. The rice-paddy area, rice-fallows and associated vegetation and crop phenological parameters such as start and end of fallow periods, length of the fallow periods, etc. from the existing database (Biradar, et al., 2016) were used as key baseline for assessing the rice-fallow dynamics across the growing seasons and years. The resultant outputs were further analyzed in geostatistical domain with various causative factors to determine suitable range.
The range of the resultant values were further normalized with fuzzy fuzzy-gamma end members to produce hotspot index values). Hotspot values has be refined with 16 factors of lentil suitability parameters, including climatic condition, soil types, soil moisture, nutrients, topography, etc. The fuzzy set model, analytic hierarchy process (AHP) method, GIS technique and MCDM were used to create final land suitability map for growing lentil in fallow areas. The results showed that 20.14%, 45.14%, 25.26% and 9.46% of the total area was most, moderate, marginal and least suitable for Lentil production respectively. The produced map was verified by compared with national statistical data. These maps and statistics are preliminary results of the ongoing study.

Geo-Statistical Approach to Model Soil Properties and Nutrient Status in Different Agro-Climatic Regions of Bihar

Nidhi Sinha[1]***, S.P. Singh**[1]**, Ranjan Laik**[2] and **Mahesh Kumar**[1]
[1]Dept. of Statistics, Mathematics & Computer Application, Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar – 848 125; [2]Dept. of Soil Science, Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar – 848 125
*Email: nidhi_96@yahoo.co.in

Spatial variability of soil properties and fertility status across different agro-climatic regions of Bihar was captured. The study was carried out in nineteen districts spread over different agro-climatic regions. Soil sampling is done using Global Positioning System (GPS) during 2015 and samples were analyzed for soil chemical properties viz. pH, EC and OM and soil fertility status (N, P, K, Zn and B). Geo-statistical techniques like semi-variogram analysis have been used to describe the spatial dependence. The spatial distribution model and spatial dependence level are observed to vary both between and within agro-climatic regions. Spherical, exponential and Gaussian models are fitted to the empirical semi-variograms. The data on various measured soil properties and the semi-variogram models developed have been used to estimate the soil test values at unsampled locations using geo-statistical technique of kriging. The knowledge of spatial dependence of soil properties and nutrient status is expected to help in region specific agricultural practices in terms of fertilization schemes.

Effect of Moderate and Extreme Heat Stress on Physiological and Yield Attributes of Elite x Wild Introgression Lines and Nagina 22 Mutants of Rice Reveals Delay of Flowering is An Important Adaptive Trait

V. Vishnu Prasanth[1]*****, **T. Vishnu Kiran**[1], **D.V.N. Chakravarthi**[1], **Y. Venkateswara Rao**[1], **M. Suchandranth Babu**[1], **D. Subrahmanyam**[1], **S.K. Mangrauthia**[1], **V. Ravindra Babu**[1], **S.R. Voleti**[1] and **N. Sarla**[1]
[1]ICAR - Indian Institute of Rice Research (Directorate of Rice Research), Rajendranagar, Hyderabad- 500030, India
*Email: vishnu.vnu@gmail.com

A diverse set of 31 rice genotypes including 12 stable BC2F8 Swarna/*O. nivara* introgression lines (Swarna ILs), 11 stable BC2F5 KMR3/*O. rufipogon* introgression lines (KMR3 ILs), 2 Nagina22 (N22, heat tolerant) EMS induced mutants along with parents KMR3 (restorer line of popular hybrid KRH2), Swarna (mega variety of irrigated lowlands), N22 (heat tolerant variety) and 3 improved varieties Vandana (drought tolerant), Krishna Hamsa and Madhukar (semi deep water rice) were evaluated for heat tolerance in pots kept in two different heat stress conditions. One set of lines were grown in wet season 2011 (WS2011) under normal conditions (June to November) and considered as control set. Two sets were grown under late sown conditions (January to June) during dry season 2012 (DS2012) and one of these sets was shifted to custom made polyhouse 60 days after sowing and maintained till maturity. Polyhouse set was considered as extreme heat stress (EHS) and the late sown set outside the polyhouse as Moderate heat stress (MHS). Each set contains 3 replications and each pot contains 3 individual plants. The mean day/night temperatures were recorded as 31.1±0.2/17.8±0.4 °C during control conditions, 37.8±0.3/24.1±0.3°C during MHS and 44.1±0.5/24.8±0.3°C during EHS conditions. Overall, the late sown set (MHS) and cover house sets (EHS) were exposed to 6.5°C and 9.9°C higher temperatures compared to the control set of lines. All the 3 sets were phenotyped for 24 traits including days to 50% flowering (DFF), flag leaf SPAD, thickness, temperature, leaf gas exchange, chlorophyll fluorescence and yield attributes. In general, there was a significant increase in mean DFF and decrease in photosynthetic efficiency but genotypes differed in their response to heat stress. . Heat susceptibility index (HSI) was calculated for each trait and genotypes were ranked based on HSI value. Combining the results of both MHS and EHS sets, five lines 175-2K (Swarna/*O. nivara* IL) was best followed by N22, 142S (Swarna/*O. nivara* IL), NH787 (N22 mutant) and 377-24 (KMR3/ *O. rufipogon* IL) were identified as the most heat tolerant or heat adaptive and 90-5 (KMR3/ *O. rufipogon* IL), Madhukar, Krishna Hamsa 166-2 (Swarna/*O. nivara* IL) and Vandana as the most heat susceptible or least adaptive to heat stress. Results revealed that the genotypes with ability to alter their flowering time/period, maintain photosynthetic efficiency and recovery ability after prolonged heat stress can withstand both moderate and extreme heat stress best.

Estimation of Finite Population Regression Coefficient Involving Two-Stage Sampling Design

Pradip Basak[1*], **U.C. Sud**[1] and **Hukum Chandra**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[*]Email: pradipbasak99@gmail.com

Generally, survey data are of multivariate in nature and therefore, many a times, the objective of the survey is to establish the relationship pattern between variables rather than estimation of simple parameters like means or totals. When the variables are quantitative in nature and the interest is to find causal relationship then regression analysis may be an appropriate method. The well-known ordinary least squares estimation of regression coefficient is based on the assumption that sample elements are independent and identically distributed. This assumption of independence does not hold good if the data are collected through sampling design other than simple random sampling with replacement. At the present time, with the availability of enhanced computationally facility, most of the survey designs are complex in nature involving stratification, unequal probabilities of selection, clustering, multi-stages and multi-phases etc. From the regression analysis point of view, any deviation from independence assumption leads to complications in the form of error variance-covariance model. In addition to this, the classical finite population sampling theory also requires survey weight of sample elements to be incorporated in the regression analysis. In this study, three different calibrated estimator of population regression coefficient has been proposed based on the availability of auxiliary information at the PSU and SSU level under two-stage sampling design. The expression for its variance and variance estimator has also been obtained using Taylor series linearization. We demonstrated the improved performance of the proposed estimators as compared to existing estimator using a real data based simulation study.

MGARCH Models for Modelling Price Volatility under Bayesian Framework

Achal Lama[1*], **Girish K. Jha**[3], **Bishal Gurung**[2], **Alolika Mondal**[2] and **Ranjit Kumar Paul**[2]
[1]National Academy of Agricultural Research Management (NAARM), Hyderabad-500030, Telangana, India; [2]ICAR -Indian Agricultural Statistics Research Institute, New Delhi, India; [3]ICAR - Indian Agricultural Research Institute, New Delhi, India
[*]Email: chllm6@gmail.com

India is one of the largest importer, producer, processor and consumer of pulses in the world. Thus pulses are an integral part of Indian agricultural system. Indian agriculture till date is largely dependent on monsoons. Pulses being no exception are also largely dependent on monsoons. To aggravate the situation further, pulses are grown mainly on marginal lands. These factors together contribute to the high level of fluctuations in pulse production and prices. In this paper an attempt has been made to model the volatile pulses prices using the VAR-MGARCH approach. The parameters of MGARCH models are estimated through classical Maximum Likelihood Estimation (MLE), but in past decade the Bayesian method for estimating the parameters have been developing fast. The very basic assumption of a Bayesian framework is that the data is not exhaustive to explain all the underlying behavior of the series. Thus priors are to be assigned to the parameters of the model and then estimated under that prior information. Bayesian analysis also provides the density of the parameters of the model unlike the point or interval estimates provided by the classical approaches. To deal with the presence of kurtosis in the series Student -t distribution innovation was also considered. Further, it was empirically found that the MGARCH-DCC model with Student -t distribution under Bayesian framework was the best among all for modelling the pulses series.

Recent Advances in Statistics (Session-V)

Transfer Function Model for Crop Yield Forecast using Weather Variables

Md. Wasi Alam[1*], **Mrinmoy Roy**[1], **Santosha Rathod**[1] and **KN Singh**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: wasi@iasri.res.in

An effort has been made to apply transfer function model with exogenous variables like minimum temperature, maximum temperature and rainfall along with rice yield for the period of 1975 to 2013 of Aligarh district of Uttar Pradesh. As a candidate models ARIMA (1, 1, 0), ARIMA (1, 1, 1) and ARIMA (0, 1, 1) models were fitted for the yield data. On the basis of goodness of fit criterion, significance of parameters estimated and white noise tests, we selected ARIMA (1, 1, 0) model as best one model. We got forecast values of rice yield of Aligarh district by 2020 through ARIMA (1, 1, 0) model. We also got the forecast of residuals by ANN approach and corrected the forecast values of yield. On the basis of goodness of fit criterion, ARIMAX model with ARIMA (0, 1, 1) and rainfall as exogenous variable was found as suitable transfer function model for the yield data. We found 15.904% MAPE through ARIMAX as compared to 18.72% obtained by ARIMA (1, 1, 0) model.

An Improved Method of Estimation of Hedge Ratio using Kalman Filter

Prawin Arya[1*] and **Bishal Gurung**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
[*]Email: prawin_arya@yahoo.co.in

Risk in agriculture is omnipresent due to the sudden fluctuations in prices of commodities. To cope up with these fluctuations, hedging is usually employed to reduce risk. Studies of hedging carried out in developed markets like the United States and Europe have finally been employed in emerging markets such as China, India, Brazil, Russia and many other Asian countries. Our study deals with estimating the optimal hedge ratio, which is defined as the relationship between the price of the spot commodity and that of the hedging commodity. Hedging is done in order to minimize the price risk of agricultural commodity due to volatility in prices. Several techniques to estimate the hedge ratio have been proposed in literature. The conventional approach is based on fixed coefficient model which gives a constant optimal hedge ratio. Previous studies have shown that the expected relationship between economic or financial variables may be better captured by a time varying parameter model rather than a fixed coefficient model. Therefore, the optimal hedge ratio, can be one that is time varying rather than constant. To this end a new and powerful technique called the Kalman filter has been applied for estimating time varying hedge ratio. An illustration using three contracts of soybean market was also carried out. It was seen that time varying hedge ratio performed better than constant hedge ratio for the data under consideration.

Pre-Harvest Forecasting of Irrigated Wheat Yield on the Basis of Weather Variables in Banaskantha District of Gujarat

G K Chaudhary[1*] and **B H Prajapati**[1]
[1]C. P. College of Agriculture, S. D. Agricultural University, Sardarkrushinagar- 385 506, Gujarat, India
[*]Email: c_gk2000@yahoo.com

To suggest most suitable pre-harvest forecasting model for irrigated wheat yield in Banaskantha district of Gujarat state past 30 years (1982-83 to 2011-12) weather data (weekly average of maximum and minimum temperature, morning and evening relative humidity, bright sunshine hours/day from 46th to 6th meteorological standard weeks (MSW) and annual rainfall (mm) were collected from the Agro-Meteorological Observatory, Agronomy Instruction Farm, S. D. Agricultural University, Sardarkrushinagar. The time trend was also included as independent variable. The data on average irrigated wheat yield (dependent variable) of Banaskantha district were obtained from the Directorate of Agriculture, Gujarat state, Gandhinagar. The step-wise regression procedure was employed by using 30 years data. The simulation of the model was done using the data from 2010-11 to 2013-14. The equation of 11 weeks crop period of 28 years provided suitable pre-harvest forecasting model predicting irrigated wheat yield 5 weeks before actual harvest and explained 83.50 % variation in the yield with minimum standard error (132.26).

Hectareage Response Study of Wheat Crop using Nerlovian Model for Gujarat State

A.S. Dudhat[1*] and **N.J. Rankja**[1]
[1]College of Agriculture, Junagadh Agricultural University, Amreli, Gujarat
[*]Email: dudhatashok1963@gmail.com

The present study on hectareage response of wheat has been carried out using Nerlov's model for Gujarat state. The state level data relating to area, production, productivity and farm harvest prices of wheat were obtained from the published and compiled information by Directorate of Agriculture, Gujarat State, Gandhinagar for the period starting from 1980-81 to 2007-08. On the basis of the correlation co-efficients of selected independent variables with current hectareage under wheat crop, single equation, linear as well as log-linear models were formed. The partial regression co-efficient of expected yield of wheat crop was significant at different levels in all the single equation models. Yield factors like lagged price and expected price played an important role in hectareage change for wheat crop, while, non-price factors like hectareage of competing crop and expected yield little influenced the hectareage of wheat crop.Risk factors like price risk, return risk and yield risk had not showed a significant role in hectareage change for wheat crop. According to R 2 and adjusted R 2 , model HEWH = -13186.99 + 0.0108HEWHL -5.1574PWHL + 0.3095EPWH+ 7.6970**EYWH +12046.26REPWH + U 1 were found to be the best fitted model for prediction of hectareage of wheat crop in Gujarat state.

Specification of variables: Let X: crop and WH: Wheat.

Hectareage variables: HEX: Current hectareage under X crop in 00' ha. HEXL: Lagged hectareage of X crop in 00' ha.

Price variables: PXL: Lagged price of X crop in rupees per quintal. RPXL: Lagged relative price of X crop calculated as, **RPXL = PXL / PCL** where, PCL: Lagged price of competing crop EPX: Expected price of X crop calculated as the average of the last three year's price. REPX: Relative expected price of X crop. Calculated as, **REPX = EPX / EPC** where, EPC: Expected price of competing crop.

Yield variables: EYX: Expected yield of X crop calculated as average of the last three year's yield in kg/ha.

Nonparametric Time-Series Regression Modeling and Forecasting under Correlated Errors using Simple Data-Driven Bandwidth

Himadri Ghosh[1*] and **Soumen Pal**[1]
[1] ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: hghosh@gmail.com

This paper considers nonparametric regression analysis in time-domain for modelling and forecasting of country's growth rate and production of agricultural commodities. The methodology for estimation of nonparametric regression function has been developed and applied under the set-up of correlated errors. This lends to obtain appropriate minimization criteria for the choice of data driven band-width under cross validation. This is based on criterion of minimizing squared error between observation at time t and its stochastic conditional expectation given all other observations to minimize mean integrated square error for estimation of regression function. Another data-driven choice of bandwidth estimation is developed based on bootstrap technique. The methodologies are used while estimating the nonparametric regression function by moving averages and kernel smoothing of observations with bimodal kernel density. Finally, interval estimates are obtained for true regression function and prediction intervals are developed based on criteria of unconditional and conditional prediction error variance.

An Insight into Technology Diffusion of Tractor through Weibull Growth Model

Bishal Gurung[1*], **Yeasin Md**[1], and **K.N. Singh**[1]
[1]ICAR- Indian Agricultural Statistics Research Institute, New Delhi, INDIA
[*]Email: vsalrayan@gmail.com

Most of the technological innovation diffusion follows an S-shaped curve. But, in many practical situations this may not hold true. To this end, Weibull model was proposed to capture the diffusion of new technological innovation, which does not follow any specific pattern. Nonlinear growth models play a very important role in getting an insight into the underlying mechanism. These models are generally 'mechanistic' as the parameters have meaningful interpretation. The nonlinear method of estimation of parameters of Weibull model fails to converge. Taking this problem into consideration, we propose the use of a powerful technique of Genetic Algorithm (GA) for parameter estimation. The methodology is also validated by simulation study to check whether parameter estimates are closer to the real value. For illustration purpose, we model the tractor density time-series data of India as a whole and some major states of India. It is seen that fitted Weibull model is able to capture the technology diffusion process in a reasonable manner. Further, comparison is also made with Logistic and Gompertz model; and is found to perform better for the datasets under consideration.

Shifting Area from Major Cereal Crops to Other Crops in Saurashtra and Kutch in Gujarat

A.D. Kalola[1*] and **R.L. Yadav**[1]
[1]Department of Agricultural Statistics, B. A. College of Agriculture,
Anand Agricultural University, Anand- 388110 (Gujarat)
*Email: adkalola@aau.in

The present study was made to examine the shift in area of major cereal crop in last five, five year plan periods (VIIIth: 1992-93 to 1996-97 to XIIth: 2012-13 to onwards) in Saurashtra and Kutch zone. In order to calculate shift in area, average area of eighth five year plan was taken as base and percentage change was worked out for successive five year plans for different crops. For the horticultural crops the average area of tenth five year plan was taken as base, due to unavailability of data for previous plans. The data on area of major cereal crops from 1992-93 to 2013-14 and for horticultural crops 2002-03 to 2013-14 were collected from Directorate of Agriculture, Gujarat state, Gandhinagar. Five year plan period wise shift in area of major cereal crops of Saurashtra and Kutch zone was calculated and the results showed that most of the cereal crops like rice, wheat, sorghum and bajra were sown 36.36%, 18.60%, 64.35% and 19.60%%, respectively decrease in area over VIIIth five year plan. The area of all the cereals were continuously decreased in XIIth plan. And the area of cotton and horticultural crops (flowers, spices, and fruit crops) were continuously increased after Xth plan. It means the farmers switched over to the crops like cotton and horticultural crops, especially flower and spices crops from cereal crops in Saurashtra and Kutch zone.

Forecasting of Prices of Major Pulses in Andhra Pradesh- An Application of ARIMA Model

K.Divya[1*], **S. Rajeswari**[1], **I. Bhavani Devi**[2], and **P.Sumathi**[3]
[1]Dept. of Agricultural Economics, S.V. Agricultural College, Tirupati, Andhra Pradesh, India; [2]Institute of Agri-Business Management; S.V. Agricultural College, Tirupati, Andhra Pradesh, India; [3]Dept. of Statistics and Mathematics, S.V. Agricultural College, Tirupati, Andhra Pradesh, India
[*]Email: divya.kathula@gmail.com

India is bound to be the global leader in terms of production and consumption of pulses. Since, India is leading importer of pulses, production of pulse/ legume crops has been stagnant over the years. Consequent upon this there is widening gap between demand and supply. About 20 % of the total pulses demands are met by imports only. India being the largest producer (18.5 million tonnes) and processor of pulses in the world. It also imports around 3.5 million tonnes annually on an average to meet its ever increasing consumption needs of around 22.0 million tonnes (Patel 2015). Foodgrains output in 2014-15 crop year (July-June) fell by 3.2 per cent to 257.07 million tonnes due to 12 per cent deficit rains and unseasonal rains during February-March (Singh 2015). As a consequence the per capita availability of pulses which was already declining would further decline. Prices have gone up substantially due to such anticipation in recent period. Being the chief source of protein for average Indian a fall in its availability per capita is a cause of concern. Forecasting prices for agricultural commodities, which are perishable, is many a times risky. Forecasting involves making estimates of the future values of variables of interest using past and current information. There are a number of methods to generate prediction ranging in intuitive judgments through time-series analysis to econometric models. Weiss (2000) concluded that an Auto Regressive Integrated Moving Average (ARIMA) is most common for forecasting prices of commodity. This paper attempts forecasting the monthly prices of major pulse markets of Andhra Pradesh. Two markets were selected for each crop i.e., bengalgram (Koilakuntla and Kurnool), blackgram (Tenali and Ponnur), greengram (Suryapeta and Thandur) and redgram (Thandur and Kurnool) based on maximum arrivals. The secondary data pertained to the period from 2000 to 2014 has been used for the study through fitting of univariate Auto Regressive Integrated Moving Averages (ARIMA) models. The results of ARIMA model for bengalgram indicated that the prices from January to March, 2015 would be ranging from Rs.2788 to Rs.2830 per quintal in Koilakuntla market and Rs.2565 to Rs.2583 per quintal in Kurnool market. For blackgram Rs.5966 to Rs.6009 per quintal in Tenali market and Rs.5986 to Rs.6037 per quintal in Ponnur market. For greengramRs.6137 to Rs.3187 per quintal in Suryapeta market and Rs.6183 to Rs.6277 per quintal in Thandur market. In the case of redgram Rs.4933 to Rs.4969 per quintal in Thandur market and Rs.3680 to Rs.3813 per quintal in Kurnool market. When the forecasts were compared with the real time prices, it was observed that there was less deviation in the case of bengalgram, blackgram, greengram and redgram.

Trends and Decomposition Analysis of Pigeonpea in India

Hemant Kumar[1*] and **Devraj**[1]
[1]ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh-208024
[*]Email: rushtohemant@rediffmail.com

The time series secondary data on area, production and yield of pigeonpea from 1950-51 to 2013-14 were broadly partitioned into seven decades in order to demonstrate the trend of pigeonpea production in more convincing and simple manner in the present study. The partition of the data in decade's terms could clearly show the pigoenpea production status and growth pattern in the country. The area under peigeonpea in 1950-51 was 2.18 million ha which showed an increase trend and recorded as 3.9 million ha in 2013-14. It was highest (4.42 million ha) in 2011-12. The production and yield of the crop also witnessed an increasing trend. The production of pigeonpea was 3.17 million tons in 2013-14 as its production was 1.72 million tons in the base year. Similarly, the yield of pigeonpea showed an increasing trend and recorded 813 kg/ha in 2013-14 as against 788 kg/ha in 1950-51. The farmer could achieve this increasing trend in production and yield mainly introduction to resistant varieties against different diseased and insects and pests, better management and matching improved production and protection technologies. Decomposition analysis was also performed and the percentage contribution of area, yield and their interaction in increasing or decreasing the production of pigonpea for each decades from 1950-51 to 2011-12 and the overall period was calculated. The productivity effect has a greater say in pigeonpea each decades separately. Response to increase in production because of increase in acreage is evident during eighties. In nineties and first decades of this century, the production is decreased due the reduction in acreage. The interaction of area and yield is not much except the fifties. Overall the production is increased mainly due to area affect.

Sugarcane Productivity in Bihar- A Forecast through ARIMA Model

Mahesh Kumar[1*], **Rohan Kumar Raman**[2], **Nidhi**[1], **S.P.Singh**[1] and **N.K.Azad**[1]
[1]Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar-848125; [2]ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata-700120, India
[*]Email: mahesh_smca@yahoo.co.in

This study is undertaken to attempt forecasting the sugarcane productivity of Bihar through fitting of well-known Box Jenkins univariate Auto Regressive Integrated Moving Average (ARIMA) model. Time series data on sugarcane productivity in Bihar from 1939-40 to 2014-15 were taken from Sugarcane Research Institute (SRI)*Pusa, Bihar and Indian sugar** for the study. The data on sugarcane productivity in Bihar from the year 1940 to 2010 were utilized to build an ARIMA model and validated through five year productivity data from 2011 to 2015. Akaike information criterion (AIC) was selected for best model selection criteria. ARIMA (0, 1, 1) model found best suitable model for sugarcane productivity in Bihar based on AIC model selection criteria. The performances of models are validated by comparing with actual values of sugarcane productivity in Bihar data. Using developed ARIMA (0, 1, 1) model, two year ahead, year 2016 and 2017 sugarcane productivity in Bihar forecasted showing increasing productivity with 4.22 % and 5.15 % prediction standard error.

Forecasting the Hazards of Agriculture with Statistical Tools

Madhura.M[1*] and **Indhra.P**[1]
[1]S.D.N.B Vaishnav College for Women, Chennai, Tamil Nadu, India
[*]Email: madhura.aish@gmail.com

India is mainly an agricultural country. Agriculture is the most important occupation for most of the Indian families. In India, agriculture contributes about sixteen percent (16%) of total GDP and ten percent (10%) of total exports. Over 60 % of India's land area is arable making it the second largest country in terms of total arable land. Agricultural products of significant economic value include rice, wheat, potato, tomato, onion, mangoes, sugar-cane, beans, cotton, etc.

A number of different factors can cause agricultural productivity to increase or decrease. It is important to note that productivity is not an absolute measure, but rather a reflection of the ratio between inputs and outputs.

The factors that affect agriculture in India are physical and economic factors. The physical factors such as climate, richness of soil and topography which are not in the hands of farmers. It is very necessary to forecast these physical factors to manage the forthcoming loss.

This paper discusses on forecasting rainfall and climatic changes to manage crop irrigation, forecasting disaster to manage agricultural loss, to promote adaption and preparedness for meeting climatic change challenge and evolve mechanisms for effective drought and food management and soil survey & land use planning.

Sensor Fusion for Normally Distributed Noise Estimation by Snapshots Techniques using Foot and Mouth Diseases Massive Data Sets in Karnataka State

Basavarajaiah D M[1*]**, H N Narasimha Murthy**[1]**,** and **B N Nagaraju**[1]
[1]Veterinary College, KVAFSU (B), Hebbel, Bangalore-560024
[*]Email: sayadri@gmail.com

The foot and mouth disease (FMD) is one of the highly contagious viral diseases causing illness in cloven footed animals. The outbreak of this disease can grave deleterious economic burden on national GDP and global economic level. FMD life data sets are often sparse and unable to guide the policymakers on outbreak, epidemic attack, and vaccination status, impact of vaccination, economic loss and mortality rate. Many intrinsic factors have influenced FMD epidemic *viz*., Climatic, Intrasectoral transmission, disease residual effect, hygiene of the heard, food, shelter, drinking water *etc*. Over a period of disease occurrence, impact factors, different variants and confounders have aroused from the FMD real life datasets. Due to paucity of literature, no modified Statistical and Mathematical tools have been availed for reduction of high dimensional massive data sets. In this context of the research gap, the present study aims to fit Sensor fusion noise estimation by Snapshots techniques for the reduction of high dimensional massive real life FMD data sets. The FMD high dimensional life datasets studied pertained to the accrual period from 2005-2014. The data on demographic profile, epidemic factors and economic loss or constraints faced by the farmers were collected through questionnaires. Sensor fusion for noise estimation was simulated by Mathmatica-16.50 version. The fitted model estimates incidence rate, outbreak and economic losses within subject noise attributes. Present study recommends the improvement of life quality domain, and also nurtures how best the model would be applied in high dimensional data sets with respect to traditional method.

Information Management of Prices and Arrivals of Castor in Major APMCs of India

P. Madhuri[1*] and **S.V. Ramana Rao**[1]
[1]Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad-30, Telangana, India
[*]Email: p.madhuri@icar.gov.in

India is the global leader in castor contributing to around 85% of world's total production and dominating the global trade with a share of more than 9% from the country. India produces around one million tonnes of castor seed and around 0.5 million tonnes of castor oil. During 2014-15, the total castor production was 1.7.million tonnes from an area of 1.1 million ha with a productivity of 1.57 tonnes/ha. In India Castor is chiefly confined to Gujarat and Rajasthan and also cultivated in parts of Andhra Pradesh, Telangana, Karnataka, Tamil Nadu and Odisha. The asymmetry in market information leads to less profitability due to pre-determined prices by the middlemen who dominate the castor trading. This necessitates developing the Market Information System on castor to provide access to the farmers on the prices and arrivals scenario in major APMC's trading castor. In this direction, an attempt is made in this paper to design a customized data management system to minimize the price asymmetry of castor farmers. The data on prices and arrivals from major APMCs trading castor in the states of Gujarat, Rajasthan and Andhra Pradesh for the period 2001-02 to 2014-15 were collected and a query based market information system was developed using PHP and MySQL for easy retrieval of information based on market, date and arrivals. The customized information from the relevant APMC's in a given agro ecological region / district were disseminated to the stake holders thro SMS service using the ICT platform. This enabled farmers to access prices in the APMC's of their jurisdiction/ district and helped them sell at a higher price over the pre quoted price by the middlemen/trader.

Modeling of Diagnostic Test Performances using the Summary Indices of Lorenz Curve

R. Anjana[1*] and **R. Aishwarya**[1]
[1]SDNB Vaishnav College for Women, Chennai, Tamil Nadu, India
[*]Email: anjana.rajagopal97@gmail.com

The Lorenz Curve, developed by Max O. Lorenz in 1905 is a graphical representation of inequality of distributions. The Gini Index and Pietra Index both are measures of statistical heterogeneity and are derived from Lorenz curve and the diagonal line. In medicine, a diagnostic test is any kind of medical test performed to aid in the diagnosis or detection of disease. For example, such a test may be used to confirm that a person is free from disease, or to fully diagnose a disease, including sub-classifying it regarding severity and susceptibility to treatment. The main challenge of these tests is to interpret whether a person is prone to a particular disease or not. The probabilistic interpretation of performances of diagnostic tests has already been found using several indices based on the Receiver Operating Characteristic Curve (ROC curve). However, in this paper a new approach has been established using the Lorenz curve. The summary indices of Lorenz curve i.e. Gini index and Pietra Index can be interpreted in several ways as the probability of correct diagnosis and are used as a good measure of test characteristics. The probabilistic model using Lorenz curves is based on real diagnostic scenarios.

Trade, Market and Market Intelligence with Relation to Pulses in Major Pulse-Hub Region (Bundhelkhand) of India

Deepak Singh
ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
Email: deepaksingh2112@gmail.com

The pulses constitute an important crop in the drought prone areas like Bundhelkhand region (known as bowl of pulses) but the risk and uncertainty in its yield, high price volatility and low price elasticity of demand leads to insecurity among pulse growers. An attempt is made to study the pulses market, trade, infrastructure and price forecast by considering five pulses in the study viz. urd, mung, chickpea, pigeonpea and lentil in Pulses growing region. The data of variables with relation to trade, storage and processing and support along with monthly wholesale price and arrivals for the period January, 2007 to March, 2015 from nearby bundhelhkand region mandis were collected i.e. Aligarh, Allahabad, Etawah, Fatehpur, Jhansi, Kanpur, Lalitpur and Lucknow. The information was also collected from farmers, traders and market officials to carry out the study.

The Kanpur mandi was the largest mandi in terms of arrival followed by Jhansi, Lalitpur and Allahabad with respect to pulses. The Index methodology was used to access the Mandis infrastructure with relation to sub-indices of Trade, storage and processing, support and overall comparison with “overall index” of three-subindices. The lucknow mandi stood first with all the indices while Kanpur mandi stood second for all the indices except trade index. From index and pulses arrival in the market it is deduced that Kanpur mandi works as the networking-hub for inter and intra state trade of Pulses. It was found that the average price variation for pulses from producer (farmer) to consumer was around 95 % with maximum change at level of Pulse processors i.e. 27%.

Price forecast was made by checking the efficiency of few models -Trend analysis, Decomposition, Moving Average (MA), Single Exponential Smoothing (SEM), Double Exponential Smoothing (DES), Winters, Autoregressive integrated moving average (ARIMA) and Seasonal autoregressive integrated moving average (SARIMA). Auto Correlation Function (ACF), Partial Auto Correlation Function (PACF), Maximum absolute percentage error, Akaike Information Criteria (AIC) and Baysian Information Criteria (BIC) were used to test the efficiency and validity of the models. Among the different models tested ARIMA was found to be the best suited maximum times for price forecast in Pulses as for chickpea (ARIMA -1, 1, 0), Urd (winter additive), Mung (ARIMA-0,1,1) , lentil (ARIMA-0,1,1) and pigeonpea (ARIMA-1,1,1) were suited to price forecast.

Statistical Analysis and Forecasting of Rice Gall Midge Based on Climate Factor in Spsr Nellore District of Andhra Pradesh

B. Ravindra Reddy[1*], **G. Mohan Naidu**[1] and **P. Sumathi**[1]
[1]Dept. of Statistics & Mathematics, S.V. Agricultural College, Tirupati (A.P,)
[*]Email: balamrr72@gmail.com

This paper presents the rice Gall midge severity and the influence of climatic factors in rice growing seasons through descriptive statistics, correlation, regression and logistic regression studies in SPSR Nellore district of Andhra Pradesh. The data analysis on Gall midge incidence with the weather factors in standard weeks of rice growing seasons from 2003 to 2014 revealed that the rainfall distribution varied greatly within rice growing seasons and also over years. The average minimum temperature 23^{0}C, average maximum temperature 29^{0}C, morning relative humidity (81 – 97%) and evening relative humidity (66-90%) observed during crop seasons over years. The gall midge severity during the years 2003 to 2008 was mild to moderate and during 2009 to 2014 was high. Analysis of 11 years (2003-2014) weather data revealed that the days with RH > 91%, temperature (22^{0}C - 31^{0}C) are the most critical factors in the development of gall midge incidence. Overall for the years 2003 -2014, the results of correlation studies revealed that, among the climate factors rainfall, relative humidity's exhibits positive correlation and the temperatures were negatively correlated. The MLR model explained the 80% variability on gall midge with respect to climatic factors.

The MLR models were developed and different combinations of climate factors were found to be useful in the prediction of rice gall midge. The Logistic regression models were useful for prediction of probabilities of occurring/ non-occurring of the gall midge incidence in a particular standard week in rice growing seasons subject to the quantitative climate factors. The ANOVA carried out for testing the significance between standard weeks/varieties/dates of planting with respect to gall midge incidence.

Cost Benefit – Profit Analysis of a System which Needs Operation of their Main Unit and Sub Unit Concurrently

Sandeep Kumar Sharma[1*], **Vishal D. Pajankar**[2] and **P. G. Khot**[3]
[1]Sir Sunderlal Hospital, Institute of Medical Science, Banaras Hindu University, Varanasi-221 005 (India);
[2] Educational Survey Division, National Council Educational Research Training, New Delhi – 110 016 (India);
[3] Department of Statistics, RTM Nagpur University, Nagpur – 440 010 (India)
[*]Email: sandeep.stats@gmail.com

In real life, there may be situation where working of main unit depends upon on its dependent subunit. Similarly, operation of dependent sub unit also required for the repair of main unit. Considering this idea in mind, system with two types units (main unit and sub unit) is taken to consideration. The main unit works only if it dependent subunit is in operative condition and which also required during the repair/maintenance of main unit. On the failure of unit, the repairman attends immediately and priority for repair is given to one of the sub unit on the basis FFFR (first failed first repair).

Recent Advances in Statistics (Session-VI)

Calibration Estimators under Two Stage Sampling Design When Auxiliary Variable is Inversely Related with Study Variable

Ankur Biswas[1*], **Kaustav Aditya**[1] and **U. C. Sud**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
[*]Email: ankur.bckv@gmail.com

When auxiliary variable is inversely related to study variable, product estimators are used to make better use of auxiliary information. In this study, product type calibration estimators of the population total were proposed using the calibration approach (Deville and Särndal, 1992) under the assumption that auxiliary variable is inversely related to study variable. Different cases of availability of complex auxiliary variable under two stage sampling design have been considered for development of product type calibration estimators. The variance and the Yates-Grundy estimator of the variance of all the proposed estimators were also developed. A simulation study has been carried out for empirical evaluation and the proposed product type calibration estimators perform better than the usual product estimators and Horvitz Thompson estimators under two stage sampling design.

A Multiple Frame Approach for Estimating Population Total using Ranked Set Sampling under Finite Population Framework

Pratyush Dasgupta[1*], **Tauqueer Ahmad**[1], **Ankur Biswas**[1] and **Prachi Misra Sahoo**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi – 110 012, India
[*]Email: pratyushiasri@gmail.com

Ranked set sampling (RSS) is a cost-efficient alternative to simple random sampling (SRS), when measuring an observation is expensive, but ranking a small subset of observations is relatively easy. It is well known that use of more than one frame often results in a cost efficient survey. Thus use of RSS for obtaining independent random samples from each frame may lead to more precise estimation of population parameters. However, estimating the variance in this case can be found to be cumbersome. Therefore, in this paper, we propose a multiple frame ranked set (MFRSS) estimator for population total and variance estimation procedures using Jackknife method for MFRSS estimator under finite population framework.

Raising the Voices of the Poor: Village Dynamics Studies Knowledge Bank

Anupama GV[1*]**, Uttam Kumar Deb**[2] and **Cynthia Bantilan**[1]
[1]ICRISAT, Patancheru, Hyderabad, India; [2]Arkansas University, United States of America
[*]Email: G.Anupama@cgiar.org

The longitudinal Village Level Studies (VLS) of ICRISAT (now called the Village Dynamics Studies in South Asia – VDSA[1]) have been providing profound insights into the social and economic changes in the village and household economies in the semi-arid tropics of India since 1975. The World Bank recognized the ICRISAT VLS dataset as an "International Public Good" and the "goose laying golden eggs", while others consider it as "Gene Bank of Social Scientists" and "One of the Jewels of ICRISAT". ICRISAT's longitudinal VLS have proved to be one of the most valuable contributions of the CGIAR to the knowledge base on rural household economies.

The ICRISAT-VLS longitudinal panel is a premier source of high quality longitudinal micro-household and meso-macro level data that is effectively disseminated through **VDSA Data Warehouse (also referred as Knowledge Bank).** The **Data Warehouse** is an open access platform to the academic and research world with online analytical processing (OLAP) features established through latest ICT tools -Microsoft SQL Server and Business Intelligence (BI). Two component databases were developed:

1. **Household** Level Panel Database (micro-level dataset). The VLS/VDSA micro-level dataset include household-level survey data collected from: (a) six traditional villages of the Indian Semi-Arid Tropics for more than three decades period 1975 to 1984, 1989, and 2001-2008; and (b) data collected from 42 villages (including the above-said six villages) in India and Bangladesh for the succeeding period 2009 to 2014.
2. **Meso-macro** level (district and regional level) database on agriculture. The meso-macro level dataset covers data pertaining to the performance, structure and behaviour of a national or regional economy at a disaggregated district, and state level in India (1966 to 2012) and Bangladesh (1952 to 2012).

The conference paper highlights the technical design of the VDSA Data Warehouse. It provides a systematic documentation of the archiving and validation as well as the retrieval process which has been facilitated by advanced ICT tools. The Data Warehouse has been successfully developed for the massive longitudinal panel and time series dataset and has been harnessed by large number of researchers, graduate students and development analysts/investors all over the world. Currently, 1450 unique users from 50 countries of Asia, Africa, Australia, Europe and North America have downloaded datasets released through the VDSA website. The VDSA Data Warehouse has enabled ICRISAT to effectively and efficiently share the highly demanded massive databases through the website (http://vdsa.icrisat.ac.in) and the VDSA Knowledge Bank (http://vdsakb.icrisat.ac.in/).

The citation of the VLS and VDSA data through the Data Warehouse continues. The binding constraint which for a long time muted the statistical "voices" of the rural poor has been lifted. It has effectively resonated household and village level agricultural statistics and informed decision making on policy. ICT aided tools indeed facilitated reliable and timely data on consequences of change on the rural poor. The VLS Data Warehouse (now well established and encompassing more than 40 years of socioeconomics and statistical panel data) is a unique source of valuable quantitative and qualitative insights on individual, household and agricultural village economies. They are currently harnessed as essential inputs for informed policy support, particularly on opportunities for agricultural growth and alternative pathways in moving out of poverty.

An Information System on Buffalo (Bubalus bubalis) Genome

Amit Kairi[1*], **Tanmaya Kumar Sahu**[1] and **Dr. Atmakuri Ramakrishna Rao**[1]
[1]Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: amit.kairi90@gmail.com

India contains the largest herds of buffalo and cattle and produces highest amount of milk in the world. Among the livestock species, buffalo has remained an integral part of the Indian rural economy. With the advent of genome sequencing technologies, the whole-genome sequence assembly of Murrah breed of water buffalo has been done recently in India. Also, significant amount of information on different functional elements of various breeds of buffalo genome are available in public domain. However, the annotations of functional elements on the genome are not available. In addition, the 3D structures of buffalo proteins are not available and there exist no browser available to visualize the genic information on buffalo genome. Hence, a study has been taken up with the objectives to (i) retrieve and the nucleotide and protein sequence information available in public domain and thereby develop a database (ii) develop user-friendly Web-based information system (iii) develop a genome browser to map gene information on genome. Initially, both the nucleotide and protein sequence information has been retrieved from NCBI and parsed. Later on, the 3D structures of proteins were predicted, validated, refined and stabilized. The parsed sequence information and protein sequence as well as structure information was imported to a database. An information system on buffalo genome with 3-tier architecture has been developed with MySQL database as bottom layer, Personal Home Page (PHP) as server side application-middle layer, HTML, CSS and JavaScript at top layer as client side application layer. The developed information system contains 930 complete CDS, 1154 partial CDS, 656 Exon sequences, 237 Intron sequences, 1709 Mitochondrial DNAs, 73 sequences of Promoter Region and 67 sequences of UTR region. Out of 930 buffalo genes, 837 have been found to be mapped onto *Bubalus bubalis* (female Murrah breed). The developed genome browser shows that maximum number of genes are distributed on chromosome 4 followed by chromosome 18. The total number genes present on chromosome 4 and chromosome 18 are 324 and 155 respectively. The results from the study also reveal that 837 out of 930 genes are mapped onto *Bubalus bubalis* genome. Whereas, 561 buffalo genes are mapped onto *Bos taurus* genome. Further, 202 genes are found to be predicted as orthologues between cattle and buffalo genomes. The developed information system provides several user-friendly facilities like search, filter, download, print, visualisation and browsing. The information system on buffalo genome can be accessed at
http://cabgrid.res.in:8080/bgis.

Effect of Parity, Calving Season and Diseases on Modelling of Lactation Curves of Vrindavani Cattle

Shashank Kshandakar[1*], **Med Ram Verma**[1], **Yash Pal Singh**[1] and **Sanjay Kumar**[1]
Division of Livestock Economics, Statistics and Information Technology,
ICAR – Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh
*Email: skshandakar@gmail.com

Vrindavani cattle are recently developed synthetic crossbred cattle strain of India (50-75% exotic inheritance of Holstein-Friesian, Brown Swiss, Jersey and 25-50% indigenous inheritance of Hariana cattle). The present study was conducted on milk yield data of Vrindavani cattle maintained at Cattle and Buffalo Farm (LPM section), Indian Veterinary Research Institute, Izatnagar over a period of 5 years (2010-2014). The aim of this study was to determine suitable lactation models explaining the pattern of milk yield of Vrindavani cattle for different parities, season of calving and disease conditions. Different lactation curve models such as Ali & Schaeffer (AS), Cobby and Le Du (CL), Sikka (SK), Mitscherlich x Exponential (ME), Mixed log (ML), Wilmink (WL) and Wood (WD) models were fitted to daily test day milk yield (DTDMY) records of 404 Vrindavani cattle. The goodness of fit of different models was judged by R^2 (adj.), AIC and BIC. Durbin-Watson test was used to test autocorrelation and Shapiro-Wilk test was used to test normality present in the residuals. The results of the study indicated that Ali & Schaeffer model was best fitted to DTDMY records of Vrindavani cattle in LO-2, LO-3, LO-4 and Vrindavani cattle calving in rainy season. Mitscherlich cum Exponential model was best fitted to DTDMY records of Vrindavani cattle in LO-1 and Vrindavani cattle calving in summer season. Mixed log model was best fitted to DTDMY records of Vrindavani cattle in diseased condition (Lameness, Mastitis) and Vrindavani cattle calving in winter season. The estimated value of coefficient of heritability (±standard error) from half-sib data for Vrindavani cattle milk production is 0.28578(± 0.29712).

Energy-Growth Linkage in Indian Agriculture: A Comprehensive Analysis using Panel Co-Integration

Rajeev Ranjan Kumar[1*], **Girish Kumar Jha**[1] and **K. N. Singh**[1]
[1]Division of Forecasting and Agricultural System Modeling, ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), Library Avenue, Pusa Campus, New Delhi-110012, India.
[*]Email: rrk.uasd@gmail.com

This paper empirically examined the long-run co-movement and the causal relationship between electricity consumption and real GSDP from agriculture and allied sector for 17 major states of India during the period 1990-2010. Since the time series analysis may yield unreliable and inconsistent results with the short time spans of datasets, we employed new heterogeneous panel co-integration and panel-based error correction models techniques to investigate the relationship between two variables. The empirical results fully supported a positive long-run co-integrated relationship between GSDP and electricity consumption when the heterogeneous states effect is taken into account. It is found that although agricultural growth and electricity consumption lack short-run causality, there is a long-run unidirectional causality running from electricity consumption to agricultural growth. This implies that reducing electricity consumption does not adversely affect agricultural growth in the short-run but would affect in the long-run.

Development of Information System on Livestock Epigenetics

Sayanti Guha Majumdar[1*] and **A.R. Rao**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: sayanti23gm@gmail.com

To meet the demand for livestock products, it is essential to understand molecular mechanisms of livestock species. With the advent of new technologies, it has become possible to understand the underlying mechanisms present in the growth and development of important traits of livestock species. Epigenetics is an emerging field which deals with the study of mitotic and meiotic (or both) heritable changes in gene function that cannot be explained by changes in DNA sequence. In general, three epigenetic mechanisms are available in nature and they are (i) DNA methylation (ii) Histone modification and (iii) RNA interference (RNAi). However, the information related to the said epigenetic mechanisms in livestock species is not available at one place. Moreover, analysis of epigenetic information is required for improvement in production traits and controlling diseases in livestock. Hence, the aim of the paper is to (i) parse and analyze data of different epigenetic mechanisms in livestock species, (ii) populate a database on epigenetic mechanism and make it available through web interface. Initially, the gene and protein sequence information related to epigenetic mechanisms of livestock species has been retrieved from NCBI, UCSC and CABin. Besides, the microRNA information of cattle and sheep has been retrieved from miRBase. Subsequently, *Sequence Manipulation Suit: CpG Islands* tool is used to obtain the probable methylation sites present in the 1K upstream regions of genes in cattle. The results reveal that the CpG islands are found in 8033 genes. In addition, the three dimensional structures of histone proteins of cow, sheep, goat, camel have been predicted, validated, refined and stabilized. Also, the probable genomic regions of histone proteins in buffalo were predicted. Besides, the secondary and tertiary structures of microRNAs of cattle and sheep were predicted. In addition, the miRNA information in buffalo species has been predicted by using miRNAs of cattle. For ready use of the generated information, a web-based Information System has been developed with MySQL database as bottom layer, PHP as server side application-middle layer and HTML, CSS and JavaScript at top layer. The developed information system can be accessed at http://bioinformatics.iasri.res.in/edil/.

Development of Online Database System for All India Coordinated Research Project on Irrigation Water Management

Ashok K. Nayak[1*], **P. Nanda**[1], **R.C. Srivastava**[1], **M. Das**[1], **S. Ghosh**[1], **A. Kumar**[1], **S.K. Ambast**[1], **R.G. Patil**[2], **U.M. Khodke**[3] and **B.S. Yadav**[4]
[1]ICAR-Indian Institute of Water Management, Bhubaneswar, Odisha; [2]AICRP on Irrigation Water Management Centre, NAU, Navasari, Gujarat, India; [3]AICRP on Irrigation Water Management Centre, MAU, Parbhani, Maharashtra, India; [4]AICRP on Irrigation Water Management Centre, SKRAU, Sriganganagar, Rajasthan, India
[*]Email: aknayak75@rediffmail.com

The cost effective agriculture water management technologies are being developed through 26 different All India Coordinated Research Projects on Irrigation Water Management (AICRP-IWM) centres spread over fourteen agro-ecological regions across the country under the supervision of ICAR-Indian Institute of Water Management, Bhubaneswar as coordinating unit. In order to strengthen research administration and information sharing among the coordinating unit and other cooperating centres, it is essential to have an interactive virtual platform. This would provide a database of the solutions to different problems concerning water management in different agro-ecological regions and geo-hydrological situations of the country. Using the information communication technology application, a web based database system was developed for managing AICRP-IWM data. The web portal was created with different pages like login window, manpower details, technologies developed, expenditure details, submission of reports and other information related to each cooperating centres. The information submitted by each centre was maintained at a central server at the coordinating unit. This unit receives the data from all cooperating centres, compile and generate reports under the project for smooth functioning of the scheme. The system being online and regularly monitored at Institute level has made it an effective tool for decision making in release of funds and other guidelines for research activities. Use of open source software, PHP and MySQL in development of the system has made it a user friendly platform with enhanced security. The customized different access rights have made it an easy and specific access to information from the system to the end user. Options were created to disseminate achievements and technologies for implementation among the farmers and other stakeholders. The successful use of the system would enable the policymakers to take appropriate decision to execute and manage AICRP-IWM centres in a comprehensive manner.

KVK Knowledge Network: Data Management of Krishi Vigyan Kendras in India

Pal Soumen[1*], **Arora A**[1], **Choubey A. K**[1], **Marwaha S**[1], **Rao, N. S**[1], **Islam S. N**[1], **Adhiguru P**[1], **Saravanakumar R**[1], **Gupta C**[1], **Suhag V**[1], **Tyagi S**[1] and **Pratibha S**[1]
ICAR-Indian Agricultural Statistics Research Institute, New Delhi – 110012, India
*Email: soumen.4345@gmail.com

In India, the research output for farming community is disseminated through 645 Krishi Vigyan Kendras (KVKs) under Indian Council of Agricultural Research (ICAR). KVKs, an integral part of the National Agriculture Research and Education System (NARES) aim for promotion of location specific technology modules in agriculture and allied enterprises through extension activities viz. technology assessment, refinement and demonstrations. Krishi Vigyan Kendra Knowledge Network is an online portal (http://kvk.icar.gov.in/) to disseminate knowledge and information from KVK's to farmers. It is a single window platform which provides information about KVK, the kind of facilities it provides to the farming community, the details of events organized to propagate knowledge among rural people and contingency plan for each of the districts of India. The portal facilitates KVKs to update and upload all types of information including KVK profile details of employees, posts, finance, soil health cards, appliances, resources, crops and animals. A dashboard to provide the overview of KVK activities at a glance has also been developed. Agricultural Technology Application Research Institutes (ATARIs) are using this online application for monitoring the activities of their respective KVKs. To this end, Agricultural Extension Division of ICAR which is at the top of the hierarchy can monitor both the ATARIs and KVKs under them through this portal. The portal is developed using .NET framework and web responsive. The development of an android app has also been initiated to access the functionalities of the portal through mobile interface.

Wavelet Methodology for Capturing Volatility in Agricultural Commodity Prices

Ranjit Kumar Paul[1*] and **Priyanka Anjoy**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute
*Email: ranjitstat@gmail.com

Agricultural commodity markets often exhibit a marked extent of volatility or instability in price behaviour imposing a greater concern for the farmers, stakeholders and governments. In order to deal with the forecasting the prices of volatile commodities, traditional autoregressive integrated moving average (ARIMA) and Generalized Autoregressive conditional heteroscedastic (GARCH) model has not been found to be sufficiently extensive in many cases. Present study establishes the potential performance of wavelet based combinatory model as a proficient alternative to the traditional statistical approaches. Wavelet-GARCH model offers a unique strategy for volatility forecasting through transformation and representation of original series at various scales. Decomposition makes it possible to describe the useful pattern of the series from both global and local aspects and subsequently minimizes the signal noise level making the transformed series amenable to easy analysis and evaluation.

Insights into the Infochemicals of Insect Community through Big Data Analytics

Kamala Jayanthi P.D.[1*], **Raghava T.**[1], **Sai Pavan G.**[2] and **Sreekanth P.D.**[3]
[1]Indian Institute of Horticulture Research, Department of Entomology and Nematology, Hessargahatta, Bangalore -560089;[2] Forsys Inc, Hyderabad;[3]Information and Communication Management Division, National Academy of Agricultural Research Management, Rajendranagar, Hyderabad 500030
[*]Email: jaiinsect@gmail.com

Phytophagous insects depend on volatile chemical signals to find mates, oviposition sites and their host plants. The infochemicals have been identified based on the response and stimulus reaction in insects and have been termed as Pheromones, Arrtractants, Allomones, Synomones and Kairomones. There is a steadily growing interest to exploit these chemicals as environment friendly components of pest management.

The Pherobase host the infochemicals information on its site with ~3500 semiochemicals and ~8000 organic compounds which influence the insect behavior across the insect kingdom encompassing different orders, families, genera and species. The vast information of infochemicals needs the better understanding in context of insect evolution. Nevertheless, such huge data handling becomes difficult to process using on-hand database management tools or traditional data processing applications. Consolidating the nature of infochemicals being used across these taxonomic hierarchies to uncover the hidden patterns and unknown correlations using a systematic approach will help the chemical ecologists more precisely to gain insights in to similarity and functional diversity of these infochemicals that are being used by insects. In other words, there still exists a hidden big picture of how these infochemicals were chosen, adapted and deployed by each order, family and genus of insect in to as Big data analytics can help to better understand such information contained within the data.

In this paper, we report on our preliminary results of infochemicals usage pattern in order Diptera (True flies) using Big data analytics. Currently there are reports of infochemicals for 42 families across 218 genera in order Diptera and a further, deeper analytics in to this database will unveil better understandings of true flies chemical communication.

Database Application for RFD and Publications Data Management

N S Raju[1*], **R Nagarjun Kumar**[1], **Sumanth V**[1], **M Osman**[1] and **S S Balloli**[1]
[1]Central Research Institute for Dryland Agriculture, Hyderabad, India
[*]E-mail: nsraju@crida.in

As we are facing problem to collect monthly RFD data from around 250 NICRA partnering institutions, we thought to have a web application, so that we can easily manage the information related to monthly RFD. This web application was designed and developed to collect the monthly RFD data from NICRA partnering institutions viz., around 100 KVK's, ICAR intuitions, ACRIPAM centers, AICRPDA centers, SAU's and NGO's. This application was developed using open source tools PHP and MySQL. In this application user management was done in hierarchal passion. We have developed another Database application for data management of Publications, Symposia, Seminars, workshops, trainings attended by the scientists and also developed Reports module to generate frequently used reports on above subjects.

Space Time Autoregressive Moving Average (STARMA) Model for Modelling and Forecasting Spatio-Temporal Time Series Data

Santosha Rathod[1*] and **K.N. Singh**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[*]Email: santosha.rathod@icar.gov.in

The univariate Box-Jenkins models proven to be very useful in broad spectrum of time series analysis. Since these models are univariate they are applicable only in single series of data. The Box-Jenkins ARIMA models deals only with successive observation on phenomenon on particular point or region in space. An alternative to the univariate time series is multivariate time series analysis in which several correlated exogenous variables are considered in a model. A further improvement in multivariate time series can be expected if one includes the systematic dependencies between the observations at each region and the observations at neighbouring regions also. The phenomenon of spatial correlation is modelled over a period of time is called as space-time modelling. The autoregressive and moving average components of univariate time series lagged in both space and time is referred as space time autoregressive moving average (STARMA) model. In this work an attempt has been made to implement the three stage STARMA model. The stationary homogenous spatial weight matrix has been considered in covariance component. The results of STARMA models are compared with univariate Box-Jenkins ARIMA models and the results reveals that STARMA model performed better as compare to univariate ARIMA model.

Decomposition Analysis of Agricultural Production in Rajasthan

R.S. Shekhawat[1*] and **K N Singh**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110 012
[*]Email: shekhawat.raviraj93@gmail.com

Rajasthan is an important state for pulses and oilseeds production. The state occupied first position in production of rapeseed & mustard and second in gram in the country. Rajasthan has made significant progress in the production of agricultural produces during last two decades (1993-94 to 2012-13). In this paper an attempt was made to analyze the relative contribution of area and productivity to the change in the production of oilseed, pulses. It was calculated the contribution of various components to the change aggregate agricultural production of the state. Data related to area, production and productivity of different crops was collected from Directorate of Economics and Statistics, Jaipur. To measure the relative contribution of area and productivity to the change in production, decomposition model given by Minhas (1963 and 1964) was used. To measure the relative contribution of various components to the change in aggregate production of the state, seven factor decomposition model given by Minhas and Vaidyanathan (1965) was used. Results were observed very interesting and will help for policy makers to make appropriate policy to further enhance the production of agricultural output.

Genome-Wide SNP Discovery and Genotyping Driven by Next-Generation Sequencing in Cultivated Peanut

Kenta Shirasawa[1*], **Chikara Kuwata**[2], **Manabu Watanabe**[2], **Masanobu Fukami**[2], **Hideki Hirakawa**[1] and **Sachiko Isobe**[1]
[1]Kazusa DNA Research Institute, Japan; [2]Chiba Prefectural Agriculture and Forestry Research Center, Japan
[*]Email: shirasaw@kazusa.or.jp

In the breeding process, rapid and accurate genome-wide genotyping techniques are required to carry out selections based on the genome data. We initially considered that double-digest restriction-site associated DNA sequencing technology would be helpful for our peanut breeding program; however, this method was not suitable because few SNPs were available due to low genetic diversity in our breeding materials. To overcome this problem, since the genome sequences of the probable diploid ancestors of peanut has been available, we performed whole genome resequencing analysis on the breeding lines to identify genome-wide SNPs. Then, we performed target amplicon sequencing analysis to genotype a subset of the SNPs, which cover the entire peanut genome. Using this technique, genome-wide genotypes of the breeding populations were easily and rapidly determined. This strategy made it possible to achieve genomics-based breeding in peanut.

An Economic Study on Future Trading in Castor seeds

S.P.Bhardwaj[1*], **Bishal Gurung**[1], **Kanchan Sinha**[1] and **K.N.Singh**[1]
Indian Agricultural Statistics Research Institute. New Delhi-12, India
[*]Email: satyajai55@gmail.com

Commodity Futures market has a very vital role to play in any economy as the futures contracts perform two important functions of price discovery and price risk management. It also promotes integrated price structure throughout the country and ensures price stability. At micro-level it provides several economic benefits to the market participants. Farmers can get an idea of the future price likely to prevail and therefore can decide between various competing crops. Theoretically, trading in commodity futures helps the market discover prices based on fundamentals (i.e. based on the strength of the supply and demand forces) while allowing the physical market participants to hedge (or lock in) their returns. As the futures markets percolate into the physical markets, it also lead to several beneficial effects indirectly by means of boosting the infrastructure, increasing the credit flow, providing efficient alternative marketing opportunities, etc. With the futures discovered prices being transparent it would also help the producers through empowering them with the price information.

Functional Characterisation of Genes for Drought Tolerance from GWAS and GS Models in Maize

T. Nepolean[1*], **Shikka Mittal**[1], **Kanika Arora**[1], **AR Rao**[2], **Abhishek Rathore**[3], **MG Mallikarjuna**[1], **T Mohapatra**[4] and **HS Gupta**[5]
[1]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India; [2] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India; [3] International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India; [4] Indian Council of Agriculture Research, New Delhi, India; [5] Borlaug Institute for South Asia, New Delhi, India
[*]Email: tnepolean@gmail.com

We identified genetic loci and their association with functional mechanisms in 240 accessions of subtropical maize using a high-density marker set under drought stress. Out of 61 significant SNPs identified across agronomic traits, models, and locations by subjecting the accessions to water stress at flowering stage, 48% were associated with drought-tolerant genes. Maize gene models revealed that SNPs mapped for agronomic traits were in fact associated with number of functional traits as follows: stomatal closure, 28; flowering, 15; root development, 5; detoxification, 4; and reduced water potential, 2. Interactions of these SNPS through the functional traits could lead to drought tolerance. In addition to that prediction accuracies of seven genomic selection models –ridge regression, LASSO, elastic net, random forest, reproducing kernel Hilbert space, Bayes A and Bayes B were tested across traits and locations. Our analysis showed Bayes B outperformed other six GS models with highest prediction accuracy of 0.97. From the Bayes B a set to top 1053 significant SNPs with higher marker effects were selected across all datasets to validate the genes and QTLs. Out of these 1053 SNPs, 77 SNPs mapped ten drought-responsive transcription factors in their vicinity. SNPs associated with genes involved in strategic biological functions identified from GWAS and GS models will be useful to understand the mechanisms of drought tolerance and development of tolerant maize hybrids.

An Integrated Modelling for Micro-Energy Management in Rural Sector

V. Bhushana Babu[1*], **Sandip Gangil**[1] and **Harsha M Wakudkar**[1]
[1]ICAR-Central Institute of Agricultural Engineering, Bhopal- India
[*]Email: bhushanciae@gmail.com

Energy management in Indian agriculture is challenging task due to diversity of energy sources and rural activities. The activities occurring in any rural energy system can be broadly divided in four major groups of activities (sectors) namely, crop production, post-harvest, domestic and livestock rising. A energy resource assessment was conducted in Ganiyari village of Madhya Pradesh where the cropping pattern was soybean-wheat. The assessment revealed that 46% of the population had a family size of 5 members, about 42 % of the population was only educated up to metric. Most of the farmers (72%) were small farmers with an average land holding of 3 acres. The important machine inputs used for crop production were 5 hp electrical pumps for irrigation (90%), 35 hp to 50 hp tractors along with attachments for carrying various agricultural operations. More than 90% of harvesting was carried by hired combines. Post-harvest storage and marketing was found to be very insignificant. Management of domestic sector was done mainly by female members (12 h/day) which included cooking, washing and cleaning, fetching water, baby care, fuel gathering and transport. Energy requirement in livestock rising sector involved bringing grass, chaff cutting, feeding, maintenance of hygiene of animals, milking, etc. It was observed that a family with 3-5 cattle requires at least 3 hrs for proper management.

The intermingling nature of different rural activities is the basic theme of the integrated approach for rural energy planning and management. The existing available models are mostly for the production agriculture relating the yield and energy input. These relations are not directly applicable in domestic and livestock raising sector. It was found that the inter-linkages of different rural activities in domestic and animal raising sectors of rural energy system make it difficult to use the crop production based energy-models in domestic and animal raising sectors. The major limitation of energy-yield relations available for crop production is that it is based on single output factor. In domestic and livestock sectors the output is not directly in single term but multiple output factors. Therefore, the study concluded that the relations obtained for crop production sector between the energy and yield of domestic and livestock sectors cannot be directly applied.

Methodological Investigations in Harvest and Post-Harvest Loss Estimation in Fisheries

V. Geethalakshmi[1*], **V. Radhakrishnan Nair**[1] and **Nikita Gopal**[1]
[1]ICAR- Central Institute of Fisheries Technology, Cochin
[*]Email: geethasankar@gmail.com

Unlike agricultural commodities, fish reaches the end consumer through a chain of distributors and the highly perishable nature of the commodity leads to appreciable loss due to various reasons at each stage. Losses occur in marine fisheries due to discards in good condition and spoilage makes it unavailable and unacceptable for human consumption. Global demand for fish is growing and reduction in harvest and post-harvest losses can make a major contribution to satisfying this demand, improving quality and quantity for consumers and increasing income for primary producers. A study conducted at CIFT, Cochin attempted to estimate harvest and post-harvest losses in marine fisheries covering Ernakulam and Alleppey districts. Multiday fishing by the mechanized trawlers reported maximum loss due to capture of juveniles and their discards. Therefore for mechanized fishing sector, estimation was carried out by using both the standard two-stage stratified random sampling design and a modified approach using post-stratification. The new approach can be regarded as a two stage random sampling with post-stratification at the second stage with first stage units as the landing centre and second stage units as the boats of specific gear type. In this paper, the results are presented after estimating fish harvest losses using both the approaches. The study revealed that though the percentage loss at harvest stage using both the approaches is more or less of the same order, percentage standard errors estimated by second approach were significantly less than that estimated by the first approach. The post-harvest losses in processing and marketing sector were also computed. At the processing stage, losses occurred due to discoloration, broken tentacles, and black spot and at time loss during glazing. Few units reported rejections at export destination due to heavy metal detection. Losses in the marketing sector were due to damage during transportation, spoilage due to delay in transport and weather. The harvest and post-harvest loss estimation in marine fisheries was carried out for various seasons and it was found that harvest losses in mechanized fishing varied from 11.8% to 15.7%, whereas post-harvest losses in fish processing varied from 0.3 % to 36.9% depending on season and type of post-harvest activity. The estimated post-harvest losses in fish marketing varied from 2.5% to 8.2%.

Manpower Assessment and Forecasting of Fisheries Professionals in India

Ramasubramanian V[1*], **Ananthan P S**[1], **Krishnan M**[1], **Mary Josephine P**[1], **Hino Fernando E**[1] and **Ubair Nisar**[1]
[1]ICAR-Central Institute of Fisheries Education, Mumbai – 400 061 Maharashtra, India
*Email: ram.vaidhyanathan@gmail.com

The dynamics in supply and demand of professionals who are skilled, trained and certified solely in fisheries sector, is yet to be understood and brought into equilibrium to ensure comprehensive development of the sector. In India, few attempts have already been made to document the aforesaid dynamics in agriculture and allied sectors including fisheries but not in depth exclusively for fisheries. Keeping this in view, the present study has been carried out to analyze the gap existing between the present day fisheries education and the actual needs of fisheries job sector. Thus while bridging this gap between the technical/ skilled and scientific/ teaching personnel available at present in the fisheries system and fisheries education in particular and the need and capacity of the fisheries sector to absorb the educational system's output in general, the manpower has been assessed by forecasting the demand and supply separately. Moreover, relevant database on trained manpower in fisheries education has been documented for some states which may help the decision making process easier. Attempts have been made to look at the professional fisheries human capital requirements in India using standard manpower models. Structured interview schedules were prepared for elicitation of information. Different stake holders like students, teachers and alumni of fisheries colleges were the key respondents of the study. The forecasts have been made about fisheries manpower under different plausible scenarios within different sets of conditions. The results will help in intervention of policy makers to correct and connect the graduates' demands job market's requirements.

List of Posters

S.No.	Presenter's Name	Affliation	Abstract Title	E-mail
1	Anjoy, Priyanka	Indian Agricultural Statistics Research Institute, New Delhi, India	Wavelet mediated transformation: Some recent developments in time series analysis	anjoypriyanka90@gmail.com
2	AS, Amaresh	University of Agricultural Sciences, Bengaluru, India	Application of microsoft excel Addins for analysing agricultural research data	amaresh.as13@gmail.com
3	Banakara, Kanthesh	Navsari Agricultural University, Navsari India	Pre-harvest crop modelling of kharif rice using weather parameters in Valsad district of South Gujarat	kantheshbanakara@gmail.com
4	Chaudhary, Sunil	International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India	Prospects of genomic selection for disease resistance and nutritional quality traits in Groundnut	schoudhary612@gmail.com
5	Ganjeer, Pradeep Kumar	Indira Gandhi Krishi Vishwavidyalaya, Raipur, Raipur, India	Assessment of distribution of area, production and productivity of Wheat in Sarguja district of Chhattisgarh	pkganjeer7@gmail.com
7	Godara, Poonam	Chaudhary Charan Singh Haryana Agricultural University, Hisar, India	Variable Selection for classification and discrimination of Indian Mustard (Brassica juncea) genotypes for oil content	poonamsinghsinghmar@gmail.com
8	Hafeez, Afrah	Aligarh Muslim University, Aligarh, India	Optimal orthogonally blocked designs for an additive quadratic mixture model in three components	afrahhafeez987@gmail.com
9	Husain, Bushra	Aligarh Muslim University, Aligarh, India	F-squares based efficient uniform designs for mixture experiments in four components	bushra_husain@rediffmail.com
10	J, Yashoda	University of Agricultural Sciences, Raichur, India	Role of statistical genetics in the genomic era	yj31990@gmail.com
11	J, Prudviraj	National Institute of Technology, Warangal, India	An automatic plant leaf type detection, extraction and classification using primitive features	rajprudvi87@gmail.com
12	Joon, Shikha	Jawaharlal Nehru University, New Delhi, India	A network biology approach to Decipher stress response in bacteria using Escherichia coli as a model	shikhasriv86@gmail.com
13	K J, Yoogesh	Banaras Hindu University, Varanasi, India	Robust estimation and outlier detection for farm and non-farm income - An empirical study	kjyogeesh.agri@gmail.com
14	Karadi, Ashwini	University of Agricultural Sciences, Raichur, India	Role of genome editing in crop improvement	ash5557@gmail.com

S.No.	Presenter's Name	Affliation	Abstract Title	E-mail
15	Kolekar, Rohini	University of Agricultural Sciences, Dharwad, Karnataka, India	Development of a consensus map from three RIL populations using SSR and AhTE markers for mapping late leaf spot and rust resistance in peanut (*Arachis hypogaea* L .)	rohinimkolekar@gmail.com
16	Kumar, Sudhir	Bihar Agricultural University, Sabour, Bihar, India	Phenotyping - a parallel requirement to utilize genomic information	sudhirk121@gmail.com
17	Kumar, Sunil	Sri Venkateswara Agricultural College, Tirupathi, India	Recombinant analysis of Tobacco streak virus replicasae gene among TSV isolates of different hosts and locations	sunilkmr8617@gmail.com
18	Kumar, Vikas	National Bureau of Plant Genetic Resources, New Delhi, India	Expression data analytics of eggplant sequencing reads	vikshyd@hotmail.com
19	Kurup, Smitha	Mahyco Research Centre, Dawalwadi, Maharastra, India	Infrared imaging: A non-invasive screening for estimation of salt tolerance in Rice	smitha.kurup@mahyco.com
	KV, Vijaya Kumar	University of Agricultural Sciences, GKVK, Bengaluru; Interrnational Crops Research Institute for the Semi-Arid Tropics (ICRISAT)-Hq, Patancheru;	Characterization of Resistance to Blast Disease in Finger Millet Though Double Digest-Restricted Site Associated DNA Sequencing (Dd-Radseq) Approach	K.Vijaya@cgiar.org
20	M, Sukruth	University of Agricultural Sciences, Dharwad , Karnataka, India	Genetic variability among peanut genotypes as revealed by RAD-Seq and whole genome re-sequencing data	sukruth.crp@gmail.com
21	Priyadarshi, Rahul	Indian Institute of Rice Research, Hyderabad India	Introgression of wide compatibility gene (S5n) into IR 58025B, an elite maintainer line of rice through marker-assisted backcross breeding	rhl.priyadarshi@gmail.com
22	R, Ellango	Indian Institute of Horticultural Research, Bengaluru, India	Insilico Prediction and Characterization of microRNAs from large milkweed bug (Hemiptera: Lygaeidae) Genome	ellango85@gmail.com
23	Rudra, Ajanta	Osmania University, Hyderabad, India	Some studies on estimation of missing values in repeated measurement designs – Bayesian approach	ajanta.rudra@gmail.com

S.No.	Presenter's Name	Affliation	Abstract Title	E-mail
24	Saha, Gopal	Indian Agricultural Statistics Research Institute, New Delhi, India	Multiple outlier detection in ARIMA model: An application to agricultural data	sahagopal390@gmail.com
25	Sahu, Ghananand	Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India	Influence of area and yield of the production of maize in Chhattisgarh plain	ghananand52@gmail.com
26	Sebastian, Abin	University of Hyderabad, Hyderabad, Telangana, India	Diversity of iron homeostasis genes in rice	abin@uohyd.ac.in
27	Seram, Devina	Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India	Optimization of bruchid resistance phenotyping method in Vigna radiata X V. umbellata F9 RILs mapping population	devnah@gmail.com
28	Sircar, Sanchari	Center for Computational Natural Sciences and Bioinformatics, IIIT, Hyderabad, India	Meta-analysis of drought-tolerant Rice genotypes: A network-based approach	sanchari.sircar@research.iiit.ac.in
29	Sundaram, Prity	International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India	Mating designs for rapid initial screening of genetic stocks	pritys28@gmail.com
30	Tiwari, Anshuman	Mahyco Research Centre, Dawalwadi, Maharastra, India	Genetic variability - A basis for line and hybrid development	anshuman.tiwari@mahyco.com
31	Vemuri, Hindu	Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India	Genetic architecture of kernel- Iron and kernel-Zinc content in tropical maize (*Zea mays* L.) germplasm as revealed by genome-wide association analysis	v.hindu@cgiar.org
32	Verma, Suman	Chaudhary Charan Singh Haryana Agricultural University, Hisar, India	Linear mixed modeling for crop yield prediction in Haryana (India): An application	sumanjai27@gmail.com
	Yadav, Satyapal	ICAR-Directorate of Poultry Research, Hyderabad, India	Comparative Analysis of Cytotoxic T Lymphocytes (CD8+) in Different Breeds of Chicken	yadav.satyapal@gmail.com

Wavelet Mediated Transformation: Some Recent Developments in Time Series Analysis

Priyanka Anjoy[1*] and **Ranjit Kumar Paul**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
[*]Email: anjoypriyanka90@gmail.com

An agriculture-dominated developing country like India has been always in need of efficient and reliable time series forecasting methodologies to describe various agricultural phenomenons; it may be either seasonality of production or volatility in price, otherwise periodicity in rainfall and underlying basic features of such occurrence being mainly nonstationarity, asymmetric volatility and complex form of nonlinearity. There are several possibilities of trying out such cases with traditional statistical approaches like Autoregressive integrated moving average (ARIMA), Generalized autoregressive conditional heteroscedastic (GARCH), Artificial neural network (ANN) or Periodic autoregressive (PAR) model. But, performance of these models may be hampered due to certain chaotic nature of the data series and hence there is a need to develop a methodology to respond such situations more competently. In this aspect, wavelet analysis tool is employed for smoothing the series and minimizing the signal noise level. Wavelet based transformation offers a unique strategy for forecasting through effective decomposition and representation of data series at various scales which are more amenable to analyze and processing. Decomposition makes it possible to describe the useful pattern of the series from both global as well as local aspects and found to be highly proficient in denoising and capturing the inherent pattern of the series through a distinctive approach. For the present study, Wavelet-GARCH and Wavelet-ANN architecture are envisaged to depict the volatility and nonlinearity pattern of the temporal price data series, whereas Wavelet-regression and Wavelet-PAR models are implemented mainly to interpret the nonstationarity and nonlinearity feature of climate data series. Wavelet mediated transformation approach has proven to be more powerful in providing forecasts and information regarding the component of a time series than trying to predict a single time series.

Application of Microsoft Excel Add-ins for Analysing Agricultural Research Data

Amaresh[1*] and **R. Manjula**[1]
[1] University of Agricultural Sciences, GKVK, Bengaluru-560065
[*]Email: amaresh.as13@gmail.com

Agriculture is the backbone of India, the rural India depends on the agriculture and agricultural research is required for sustainable and modern agriculture. For sustainable agriculture, statistical analysis on the agricultural research data is most important. Statistical analysis becomes easier and fast with the use of different software in agriculture sector. One of powerful and widely used software for the data computing is the Microsoft Excel which was not originally designed for statistical analysis. Although Excel contains a number of statistical functions, it is often difficult to implement in Excel a number of statistical tests to the level required by many researchers and students. However, addins over comes this shortcoming of Excel. Addins for Excel performs a range of statistical analysis within the Microsoft Excel environment. These software are very easy, accurate and gives fast results. Many statistical addins which can be used in agricultural research for better analysis. Some of the statistical addins are Real-Statistics, RegressIt, XLSTAT, Analyse-It, Statistician, StatEL, NumXL, SigmaXL, WinSTAT etc. Real-Statistics is a free addin software for variety of statistical analysis in Excel. RegressIt is an Excel addin that performs multivariate descriptive data analysis and multiple linear regression analysis with presentation quality output. The XLSTAT statistical analysis addin offers a wide variety of functions to enhance the analytical capabilities of Excel, making it the ideal tool for data analysis and statistical requirement. Analyse-It brings powerful statistical analysis and data visualisation into Excel and there are four editions in this. They are Standard Edition, Quality Control & Improvement Edition, Method Validation Edition and Ultimate Edition. Statistician is a comprehensive and simple to use addin works in a unique way as it allows to store a data set and perform multiple analysis on it. These addins for Microsoft Excel performs high quality statistical analysis in the field of agricultural research.

Pre-Harvest Crop Modelling of Kharif Rice using Weather Parameters in Valsad District of South Gujarat

K.B. Banakara[1*], **Y. A. Garde**[2], **B. K. Bhatt**[3], **s. Ojha**[1] and **H. R. Pandya**[1]
[1]Department of Agricultural Statistics, Navsari Agricultural University, Navsari, Gujarat – 396 450;
[2]Department of Agricultural Statistics, College of Agriculture, Navsari Agricultural University, Waghai, Dang, Gujarat – 396 450; [3]Department of Agricultural Statistics, ASPEE College of Horticulture and Forestry, Navsari Agricultural University, Navsari, Gujarat – 396 450
[*]Email: kantheshbanakara@gmail.com

India is mainly based on agricultural country. Agriculture is the most important occupation for most of the Indian families so; it is the soul of Indian economy. Rice is the most important staple food in India which play crucial role in daily requisite of diet. In the Gujarat state, rice occupies about 7-8 per cent of the gross cropped area and accounts for about 14.00 per cent of the total aliment grain production. Pre harvest forecast provide utilizable information to farmers, regime offices and vendors. In the current study statistical crop modelling were employed to provide forecast before harvest of crop for taking timely decisions by the farmers. In this paper Multiple Linear Regression (MLR) Technique and Discriminant function analysis were utilized for estimating average rice production in Valsad district of South Gujarat. The weather indices were developed utilizing correlation coefficient as weight by weekly weather parameters for the year 1975 to 2014. The cross-validation of the developed forecast models were tested their precision utilizing data of the year 2010 to 2014. It is observed that value of Adj. R^2 varied from 0.363 to 0.756 in different models. The study revealed that high value of Adj. R^2 in the Model which designated that it is opportune forecast model than other models. Based on the findings in the present study it is observed that Discriminant function analysis techniques found to be better than MLR for pre harvest forecasting of rice crop yield.

Prospects of Genomic Selection for Disease Resistance and Nutritional Quality Traits in Groundnut

Sunil Chaudhari[1, 2*], **D. Khare**[2] and **P. Janila**[1]
[1]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India (502324);
[2]Jawaharlal Nehru Krishi Vishwavidyalya, Jabalpur, India (482004)
[*]Email: schoudhary612@gmail.com

Breeding for disease resistance is one of the prime focus of plant breeding programs, as a successful variety should combine high yield, disease resistance, agronomic performance, and quality. Consumer health needs and industry demand for quality has put nutritional and processing quality of food on fore-front. Marker assisted selection has been proven to be a successful breeding method for improving disease resistance involving major resistance genes. However, this method usually ignores the effect of minor genes that produce more durable and effective resistance. Conventional breeding accomplishes this to some extent, but it is prone to uncertainties and laborious phenotyping procedures over the generations. Techniques such as marker assisted recurrent selection (MARS), multi-parent advanced generation intercross (MAGIC) etc are also available to introgress QTLs with minor effects into elite backgrounds but require more expenditure on time and resources. Moreover, complex nature of inheritance of yield, biotic/abiotic stresses and nutritional quality in groundnut poses additional constraints, especially when multiple traits are simultaneously targeted in breeding program. Backcross introgression is an option for combining multiple traits into an elite background, but introgression of major effect QTLs for quantitative traits are sometimes unable to achieve desired level of expression in introgression lines due to background effect and environmental factors. Also the introgressed varieties very often fail to deliver over the recurrent parent in yield evaluation trials due to trait biased selection approach. Genomic selection is an emerging approach to increase selection intensity, accuracy and genetic gains in breeding program, for improving complex polygenic traits through increasing frequency of favorable alleles in advance generation with the help of genomics estimated breeding value (GEBV) predicted using whole genome marker profile data and multi-environmental phenotypic data. Considering genotype × environment interaction in construction of genomic selection model additionally helps to identify wider adaptable end product at early generations. The inheritance studies of rust and late leaf spot resistance and nutritional quality in groundnut showed that the gene action involved in inheritance of these traits varies from population to population. Looking towards the low heritability and bottlenecks in improvement of complex polygenic traits through conventional breeding methods, genomic selection in groundnut for improving disease resistance, yield and nutritional quality will be a powerful approach to increase genetic gain per unit of time and cost through selection of promising lines and parents on the basis of GEBVs.

Assessment of Distribution of Area, Production and Productivity of Wheat in Sarguja District of Chhattisgarh

Pradeep Kumar Ganjeer[1*], **M. L. Lakhera**[1] and **Vaibhav Mishra**[1]
[1]College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, 492012 (CG)
[*]Email: pkganjeer7@gmail.com

The present study has been undertaken to study the trends, prediction model of area, production and productivity of wheat crop in Northern Hills of Sarguja district of Chhattisgarh in India. Time series data for wheat crop were collected for the period from 1979-80 to 2012-13 from various governmental publications and web sites. The whole period divided into three sub periods (Period-I, 1979-80 to 1986-87; Period-II, 1987-88 to 1997-98 and Period-III, 1998-99 to 2013-13) for trends analysis and for prediction model. Besides these, compound growth rate, coefficient of variation (CV), instability index of area, production and productivity of the crop have also been estimated. The growth rate for area, production and productivity for wheat crop are positive and instability indices are very low which also indicates less risk in growing of this crop in future too.

Variable Selection for Classification and Discrimination of Indian Mustard (Brassica Juncea) Genotypes for Oil Content

Poonam Godara[1*], **B.K. Hooda**[1] and **Nitin Tanwar**[1]
[1]Department of Mathematics, Statistics and Physics, CCS Haryana Agricultural University, Hisar, Haryana, India
[*]Email: poonamsinghsinghmar@gmail.com

A number of procedures have been proposed for selection of variables in discriminant analysis (DA) and classification problems. The aim is to find from among a set of variables a smaller subset which enables an efficient classification of cases. In DA, variables are selected from the available variables, the variables which discriminate on account for differences between groups with minimum error of classification. In the present study, technique of variable selection in discriminant analysis has been applied to find best discriminators among high and low oil content genotypes of Indian Mustard.

Optimal Orthogonally Blocked Designs for an Additive Quadratic Mixture Model in Three Components

Bushra Husain[1] and **Afrah Hafeez**[1*]
[1]Department of Statistics & O.R, Women's College, Aligarh Muslim University-Aligarh, Uttar Pradesh (INDIA) 202002.
[*]Email: afrahhafeez987@gmail.com

Scheffé (1958) introduced models and designs for experiments with mixtures. Scheffé (1963) discussed the problem of mixture experiments involving process variables. Orthogonal block designs for Scheffé's quadratic model in three and four components were given by John (1984), Czitrom (1988, 1989, 1992), Draper *et al.* (1993), Chan and Sandhu (1999) and Ghosh and Liu (1999). Prescott *et al*. (1993) obtained optimal orthogonal block designs for five mixture components. Aggarwal *et al*. (2002) obtained D-, A- and E- optimal orthogonal block designs for Becker's model in three and four components. Singh (2003) considered optimal orthogonal designs in two blocks for Darroch and Waller's (1985) quadratic mixture model in three and four components. Husain and Parveen (2016) obtained F- square based four component D-, A-, and E- optimal orthogonal block designs for an additive quadratic mixture model. In this paper, we have considered the class of design proposed by John (1984) to obtain D-, A- and E- optimal orthogonal block designs for the additive quadratic mixture model.

F-Squares Based Efficient Uniform Designs for Mixture Experiments in Four Components

Bushra Hussian[1*] and **Sanghamitra Sharma**[1]
[1]Department of Statistics & O.R, Women's College, Aligarh Muslim University-Aligarh, Uttar Pradesh (INDIA) 202002.
[*]E-mail: bushra_husain@rediffmail.com

Scheffé (1958) introduced models and designs for experiments with mixtures. Uniform mixture designs were given by Wang and Fang (1990). This paper gives two classes of uniform designs for mixture experiments and investigates the properties of these classes of designs using various models. The two classes of designs are (i) mixture designs constructed using uniform designs based on cyclic F-squares for two designs with different runs and (ii) mixture designs constructed by projecting the uniform designs based on good lattice point method based on two designs with different runs. For restricted exploration of mixtures, we have used the transformation given by Saxena and Nigam (1977) with some modifications and constructed designs for each of these two classes of designs. The design efficiencies of the most uniform designs of the two classes of designs are also computed and compared.

Role of Statistical Genetics in the Genomic era

Yashoda[1&3]***, K.P.Viswanatha**[2]**, P. Janila**[3] and **Santosh Khadakabhavi**[1]**&**[3]
[1]University of Agricultural Sciences, Raichur-584102
[2]MPKV Rahuri, Maharastra-431722
[3]Internal Crops Research Institute for the Semi-Arid Tropics, Patancheru, Hyderabad-502324
*Email: yj31990@gmail.com

Statistical genetics is a scientific field deals with the development and application of statistical methods for drawing inferences from large genomics and plant breeding data. In recent years, a large number of advance genomic techniques to deal with complex traits put forward to the need of advance statistical techniques to analyze large genomic data and to detect genetic variations among individuals of same as well as different species and genera that helps us to understand the complex puzzle of evolution and syntenic/phylogenetic relationships. The application of statistical methods to genetics has a long history, starting with Fisher's contributions and now extending to the recent achievements of bioinformatics. The interchange of knowledge among statistician and molecular/plant breeder has been regarded as crucial for ensuring the progress of molecular biology and plant breeding practices. The rapid progress in genomic techniques such as 3rd generation of genetic markers (SNPs), genome wide association mapping, genotyping by sequencing, the DNA sequence of numerous organisms and gene-statement information (transcriptional profiling) produces large data set. Statisticians play an important role in the analysis of big data and the design of genetic studies. There are number of selected statistical methods for population-based association mapping, from single-marker tests of association to multiple marker data mining techniques for gene x gene interaction detection as well as several statistical models to deal with genotype x environment interaction in plant breeding. Therefore, success of agriculture in general, and molecular plant breeding in particular would depend on advances in statistical genetics, development of powerful software packages, databases, and networking systems. Comprehensive computational tools will be needed to integrate information regarding genotypic performance, pedigree relationships, and germplasm diversity so that genomic data can be interpreted in ways that are useful to agricultural scientists. Hence, choosing a proper statistical method is extremely critical and important for diverse application in the era of plant breeding.

An Automatic Plant Leaf Type Detection, Extraction and Classification Using Primitive Features

J.Prudviraj
NIT Warangal, Telangana-506004, India
Email: rajprudvi87@gmail.com

In the earth's ecology plants play vital role in many domains such as agriculture based industries, health care industries. Automation on identification of plant species is today's challenging scenario due to its overlapping leaves, weeds, and plant diseases. This paper discusses different approaches to focus on identifying individual leaves from sample images taken by camera or mobile phone. These methods are aimed at dimension of the problem than solution due to space constraint. Leaf area is distinguished from the background by marking a threshold green component value. Mostly weed texture is different from the leaf structure, there are many texture extraction methods are proposed in last four decades. Individual leaf identification dealt with many constraints such as holes on leaves, leaf diseases, insect pests, missing regions , image captured method, luminance and so on. As histogram solves some of the problems but it is not enough to identify individual leaves. Leaf specie identification and classification can achieve by identifying vein structure, central lines (both x and y direction), head and tail structure. Combination of above methods with texture extraction and histogram based methods may achieve better results.

A Network Biology Approach to Decipher Stress Response in Bacteria Using Escherichia coli as a Model

Shikha Joon [1, 2*], **Shashwat Deepali Nagar** [1], **Sonika Bhatnagar**[1], and **Rakesh Bhatnagar**[2]
[1]Computational and Structural Biology Laboratory, Division of Biotechnology, Netaji Subhas Institute of Technology, Dwarka, New Delhi – 110078; [2]Laboratory of Molecular Biology and Genetic Engineering, School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India.
*Email: shikhasriv86@gmail.com

The non-judicious usage of hazardous drugs has led to the emergence of multidrug resistance in bacteria posing a threat to global health and a challenge for their treatment. The extreme environmental fluctuations are well survived by these drug resistant bacteria owing to stress responses whose mechanism is poorly understood. The stress response is attributable to the presence of specific modules. The phenomenon of cross-stress protection in bacterial species is suggestive of the presence of the central proteins likely to control diverse stress responses. With this notion we aimed to target these central controllers whose destruction may further lead to the breakdown of the entire stress response machinery. In this work, we have used *Escherichia coli* to model the bacterial stress response. We generated a Protein-Protein Interaction Network by integrating differentially expressed genes known to be involved in eight stress conditions of pH, temperature, and antibiotics with relevant gene ontology terms. Topological analysis identified 24 central proteins. Out of identified 24 central proteins, 16 central proteins have a well-documented role in stress responses. This is indicative of the central control mechanism of the response so generated. The remaining eight proteins may have a novel role as central mediators of stress response in *E. coli* which needs to be further investigated. Cluster analysis of the generated network implicated RNA binding, flagella assembly, ABC transporters, and DNA repair as important processes during response to stress. Taken together our study revealed that energy and substrate metabolism, and the translational machinery are key definers of the Stress Response-Protein Protein Interaction Network. Pathway analysis showed crosstalk of Two Component Systems with metabolic processes, oxidative phosphorylation, and ABC transporters for induction of stress response. Analysis of an independent cross-stress protection dataset confirmed the results. Further this work explores the common nodes between stress response, bio film formation, and response to antibiotics. Our study provides a comprehensive view of bacterial stress response and identifies its mechanisms. Extension of these findings to pathogenic organisms can help identify similar mechanisms in them. In conclusion, we suggest an extensive exploitation of these stress responses to combat antibiotic tolerance and multidrug resistance in future.

Robust Estimation and Outlier Detection for Farm and Non-Farm Income - An Empirical study

Shivashankar Kadam, Dr. M.Gopinath Rao, Yogeesh, K.J.*
Department of Agricultural Statistics, Applied Mathematics and Computer Science,
University of Agricultural Sciences, GKVK, Bangalore-65
*Email: kjyogeesh.agri@gmail.com

The present study is an effort towards detecting outliers in farm and nonfarm income using the data obtained from the Cauvery command area of Mandya district in Karnataka. The area is divided into three reaches as upper, mid and lower reach. In each reach three hundred household data is used to detect outliers and estimation is done reach wise and for entire study area. There are several variables which influence the non-farm income and farm income in the command area, relationship of response variables, non-farm income and farm income with other socio economic variables are estimated. Outliers are detected by using robust procedure, 14, 16, 16 and 51outliers are detected based on Mahalanobis distance by considering farm income and for nonfarm income as dependent variables. 8, 6, 7 and 25 outliers are detected for upper, mid and lower reach and for entire study area respectively. Regular MLR method and robust regression methods like MM and LTS methods are used, estimation and Comparison is done between all these three methods for with and without outliers, reach wise and for entire study area. After removing outliers we can see standard error has decreased and parameter estimates have changed. Finally LTS method found to be the best method compared to MM and MLR method.

Role of Genome Editing in Crop Improvement

Ashwini Karadi[1, 2*]**, Prity Sundaram**[1]**, Sobhan B. Sajja**[1] and **K.P. Vishwanatha**[3]
[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru - 502324, India; [2]UAS Raichur, Karanataka-584104; [3]MPKV, Rahuri, Maharastra- 413722
[*]Email: ash5557@gmail.com

Genome editing is a novel technique which overcomes the constraints of classical plant breeding and transgenic technology. It creates the genetic variation by using sequence specific nucleases without introducing foreign genes into the plant genome. It has a great potential for crop improvement to meet the increasing global food demands and to provide sustainable productive agriculture system. Crops in near future will be genome edited *via* a technology called “Genome Editing with Engineered Nucleases”, a technology in which DNA is inserted, replaced or removed from genome using artificially engineered nucleases. This technology helps to develop resistance to biotic and abiotic stresses, tolerance to herbicide and to improve the nutritional quality in crop plants. These nucleases cut DNA and create specific double stranded breaks at desired locations in the genome, harnessing the cell’s endogenous mechanisms to repair the induced breaks by natural processes of homologous recombination and non-homologous end Joining. Mainly Zinc finger nuclease (ZFN) and Transcription Activator Like Effectors Nuclease (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) /CRISPR associated system (Cas) are used as modern tools for genome editing. ZFNs zinc fingers typically cleaves in the repeats that are 3bp apart and TALENs have one-to-one recognition ratio between the amino acids and the recognized nucleotide pairs. Both ZFNs and TALENs have FokI nuclease, which performs nonspecific DNA cleavage with sequence specific DNA binding domain. ZFNs can efficiently cleaves at an endogenous target gene for example, in *Arabidopsis* with ABA-INSENSITIVE4 (ABI4) gene as a target, which plays a major role in regulating the concentration of abscisic acid (ABA), that controls important traits like plant responses to abiotic stress and seed development. TALENs technology was also exploited to edit a specific gene in rice to thwart the virulence strategy of *Xanthomonas oryzae* and thereby engineer heritable genome modifications for resistance to bacterial blight. The latest genome-editing technologies, particularly CRISPR/Cas, promise to be more efficient and precise to edit genes when the genome sequences of target genes are known. The CRISPR/ Cas system mechanism requires two components like Cas9 and guide RNA. Endonuclease Cas9 introduces double strand break on a site directed by the guide RNA. Direct delivery methods of Cas9 and gRNA using *Agrobacterium* and viral replicons by using nanoparticles can be very useful for simplifying the genome editing technology. This technology has some challenges and obstacles, but has a greater role in future research and development by expanding benefits to both basic and applied biological sciences. Increased threat of climate change has already a substantial impact on agricultural production worldwide as biotic and abiotic stresses cause significant yield losses with greater risks to future global food security. Development of climate resilient crop varieties can be possible by integrating advanced biotechnological and genomic techniques like genome editing with traditional breeding methods.

Development of a consensus map from three RIL populations using SSR and AhTE markers for mapping late leaf spot and rust resistance in peanut (Arachis hypogaea L .)

Kolekar, R.M[1*], **Asha, B.,**[1] **Shirasawa, K.**[2], **Sujay, V.**[3], **Khedikar Y.P.**[3], **Cholin, S.**[3], **Mondal, S.**[4], **Chougule D.B.**[1], **Yadawad A.**[1], **Sukruth M.**[1], **Badigannavar A.M.**[4], **Pandey M.K.**[5], **Varshney R.K.**[5], **Moger N.B.**[1], **Gowda M.V.C.**[3] and **Bhat R.S.**[1]
[1]Department of Biotechnology, University of Agricultural Sciences, Dharwad – 580 005, India; [2]Department of Frontier Research, Kazusa DNA Research Institute, Chiba 292-0818, Japan; [3]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad 580 005, India; [4]Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400 085, India; [5]Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502 324, India
[*]Email: rohinimkolekar@gmail.com

Peanut (*Arachis hypogaea* L.) is an important oilseed and legume crop worldwide. Peanut breeding for biotic and abiotic stress resistance is being complemented with the development of genomic resources. Recombinant inbred line populations (TAG 24 × GPBD 4, TG 26 × GPBD 4 and VG 9514 × TAG 24) differing for foliar disease resistance were considered for developing a consensus map and localizing the QTL regions. The consensus map of 1876.10 cM with 20 linkage groups had an average inter-marker distance of 4.59 cM. The mapped markers included -294 simple sequence repeats (SSR) and 114 *Arachis hypogaea* transposable element (AhTE) markers. Using the extensive phenotypic data on late leaf spot (LLS) and rust disease across 12 seasons for TAG 24 × GPBD 4 and TG 26 × GPBD 4 RIL populations, and only rust disease across five seasons for VG 9514 × TAG 24 RIL population, common QTL regions were identified. These included the region(s) on linkage groups AhXII and AhXV. Some of these QTL regions are being transferred to two elite but LLS and rust susceptible varieties, TMV 2 and JL 24 from GPBD 4, ICGV 86699 and ICGV 99005 using marker-assisted backcrossing to breed for disease resistant and productive genotypes.

Phenotyping - A Parallel Requirement to Utilize Genomic Information

Sudhir Kumar[1, 2*], **Santosh K**[1]
[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502324, Hyderabad, Telangana, India; [2]Department of Plant Breeding and Genetics, Bihar Agricultural University, Sabour, 813210 (Bihar)
*Email: sudhirk121@gmail.com

Phenotyping as a breeding tool has been successfully utilized in breeding programs worldwide to develop several trait specific varieties. However, most of the phenotyping studies are conducted in advanced generation population which necessitates large expenditure on time, labour and resources. The development of genomic tools such as molecular markers in breeding programs has made it possible to screen early generation progenies for specific traits especially those under the effect of major QTLs/few genes. However, most of the economically important traits are controlled by several QTLs with minor effects and show high genotype x environment interaction effects. In such a scenario, the development of molecular tools is linked to the accuracy of phenotyping data. Unlike genotyping which has benefited from the development of high throughput genotyping tools, phenotyping has lagged behind. Robust phenotyping systems are needed to characterize the full suite of genetic factors that contribute to phenotypic variation across the environments and species. The objective of modern phenotyping is to increase the accuracy, precision and throughput of phenotypic estimation at all levels of biological organization while reducing costs and minimizing labor through automation, remote sensing, improved data integration and experimental design. Next-generation phenotyping generates significantly more data than previously and requires novel data management, access and storage systems, increased use of ontologies to facilitate data integration, and new statistical tools for enhancing experimental design and extracting biologically meaningful signal from environmental and experimental noise. High throughput phenotyping tools are catching the interests of plant breeders and are expected to revolutionize the prediction of complex traits, including growth, yield and adaptation to stress but their applicability is limited by their high cost and maintenance needs.

Recombinant Analysis of Tobacco Streak Virus Replicasae Gene among TSV Isolates of Different Hosts and Locations

M. Sunil Kumar[1*], **M. Krishna Reddy**[2], **P. Hema Chandra Reddy**[2], **Akshata Gad**[2], **Salil Jalali**[2], **D.K. Samuel**[2]

[1] Department of Plant Pathology, S.V. Agricultural College, Tirupathi, Andhra Pradesh – 517 502; [2]Virology lab, Division of Plant Pathology, Indian Institute of Horticultural Research, Bangalore – 560089

[*]Email: sunilkmr8617@gmail.com

RNA recombination is one of the major factors that create RNA genome variability. Assessing its evidence in plant RNA viruses helps understand the formation of new isolates and evaluate the effectiveness of crop protein strategies. *Tobacco streak virus* (TSV) belongs to genus Ilarvirus of *bromoviridae* family with tripartite genome of 8.5 kb (RNA1, 3.4Kb; RNA2, 2.9 Kb; RNA3, 2.2Kb). To search for recombination in TSV from different crops *viz.*, groundnut, sunflower, cucumber and parthenium, Recombination Detection Programme (RDP 4.0) was used. Suspected TSV infected samples exhibiting characteristic symptoms of severe necrosis of leaf lamina, petioles and other plant parts were collected and full length genome sequence of RNA1, RNA2 and RNA3 was done. The complete nucleotide sequence of RNA1 (Replicase gene) of TSV isolates (Groundnut, sunflower, cotton and cucumber) along with previously sequenced RNA1 from different crops and countries from Genebank, NCBI was subjected to recombinant analysis using RDP 4.0. Recombination analysis identified a total of 14 recombination events (groundnut 1; sunflower 4; cotton 5 and cucumber 3) among all isolates of RNA1 of TSV. To investigate whether recombination events might have occurred among the 4 TSV isolates, we performed recombination detection using only the RNA1 sequences for the 4 TSV isolates. As expected there were atleast seven recombination events in all isolates. To confirm the reliability of unusual frequent recombination events among TSV in this study, phylogenetic trees based on recombination breakpoints were constructed to assess the relationship between the receptor and donor sequence. Most parents of detected TSV recombinants originated from foreign countries. For single stranded RNA viruses recombination is a major evolutionary way for an isolate to adopt to new environmental conditions and host. Moreover, reassortment frequently occurs in natural populations of animal viruses with segmented genomes and in natural populations of plant viruses with multipartite genomes. For multipartite RNA viruses, both recombination and reassortment can be important for genetic exchange.

Expression Data Analytics of Egg Plant Sequencing Reads

Vikas Kumar[1*], **Gaikwad AB**[1], **Archak S**[1] and **Bhat KV**[1]
[1]Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi-110 012, INDIA
*Email: vikshyd@hotmail.com

The illumina sequencing read libraries from three samples containing 88 million, 70 million and 76 million reads of egg-plant (*Solanum melongena*) were assembled and analysed using different bioinformatics tools and techniques. Two methods were used for the analyses i.e. (i) *de novo* assembly based and (ii) reference genome based. The *de novo* assembly based method was carried out in CLC Genomics Workbench V7.5, and reference genome based analysis was carried out with *Tuxedo protocol* toolkit V2.1 using tomato (*Solanum lycopersicum* - 900mb, 12 chr and 31,063 predicted genes) as reference genome.

In *de novo* assembly, the merged transcriptome assemblies of first pair (125,502 contigs) and second pair (59,598 contigs) with minimum contig size 200bp, were locally blasted (BLASTx) to customised non-redundant (nr) database (*Viridiplantae*, taxid-33090) using HPC. As many as 60% and 67% potential homologs were discovered respectively and GI identifiers were further functionally annotated to gene ontology classes. The annotations were merged to expression data of respective read libraries. Differentially expressed genes were filtered (first pair- 937 and second pair- 136) and plotted with GO Plot in R software.

In reference genome, the *Tuxedo protocol* toAolkit, was used for RNA-seq expression analysis with reference genome perspective and also to discover the common genes and transcripts between two species. The *Tuxedo protocol* pipeline is an open-source software tools for gene discovery, splice variants (isoforms) of known ones and comprehensive expression analysis of high-throughput mRNA sequencing data. The accessory tools and utilities aids were used in data management and visualisation of RNA-seq analytics. The reads libraries of three samples were mapped (46%, 61% and 60%) to reference genome using *Tophat*. The *Cufflink* was used with BAM file and gene model annotation file (gtf/gff) to prepare the transcript assembly for each samples. *Cuffdiff* was used to calculate the expression level between sample pairs, using merged assembly and initial read sample read mappings, as a result statistically significant, differentially expressed genes in first pair - 165 and second pair - 44 genes were filtered out from total expressed genes i.e. 19,450, 21030 and 20,512 of respective sample. The expression data of genes were plotted and visualised with *CummeRbund*. The results are expected to provide additional genomic resources for crop improvement.

Infrared Imaging: A Non-Invasive Screening for Estimation of Salt Tolerance in Rice

Abhijeet Shillak[1], **Smitha Kurup**[1*], **Sajid Khan**[1], **Bharat Char**[1] **and Usha Zehr**[1]
[1]Mahyco Research Centre, Dawalwadi, Jalna-Aurangabad Road, Jalna, Maharashtra – 431203, India
[*]Email: smitha.kurup@mahyco.com

Salinity is one of the most common abiotic stresses affecting rice (*Oryza sativa*) productivity; and rice is classified as a salt sensitive crop. Plants in saline conditions struggle to take up water from the soil. Moreover, stomata of leaves are closed so as to prevent water loss; which ultimately results in an increase in leaf temperature. This physiological phenomenon is captured by infra red thermography (heat sensitive sensor) to obtain valuable information regarding the tolerance or susceptibility level of plants through canopy temperature (CT) measurement. CT is directly related to the rate of evapo-transpiration from the canopy surface. In the present study, varying levels of salt concentration were imposed on rice cultivars of two different categories *ie*., tolerant and susceptible groups. Infrared image processing along with appropriate statistical analysis explicitly demonstrated a significant difference between tolerant and susceptible cultivars at both 15 mM and 30 mM saline condition. This non-destructive screening method will be helpful to identify salt tolerant as well as susceptible cultivars before the onset of symptoms. Next generation phenotyping technology is emerging as a promising tool for screening of large populations in crop improvement programs.

Characterization of Resistance to Blast Disease in Finger Millet Though Double Digest-Restricted Site Associated DNA Sequencing (Dd-Radseq) Approach

Vijaya Kumar K.V[1, 2*]**, Siva Subbramani**[2]**, Sujay V**[3]**, M.V.C. Gowda**[3, 1]**, Malali Gowda**[4]**,** and **Santhosh Deshpande**[2]
[1] University of Agricultural Sciences, GKVK, Bengaluru-65; [2]Interrnational Crops Research Institute for the Semi-Arid Tropics (ICRISAT)-Hq, Patancheru-502324; [3] All India Coordinated Small Millets Improvement Program (AICSMIP), ICAR, GKVK, Bengaluru-65; [4]Center for Cellular and Molecular Platforms (C-CAMP), NCBS, GKVK campus, Bengaluru-65
[*]Email: K.Vijaya@cgiar.org

Finger millet is one of the important crops of traditional farming system of dryland agriculture in peninsular India, providing food and fodder security. It has a unique nutritional status of high calcium, iron, potassium and zinc along with many health benefits. The blast disease is one of the major biotic constraints that limit production in all finger millet growing areas. Conventional breeding has not been completely successful in developing resistant varieties adapted to different agro-climatic situations. Marker Assisted Breeding (MAB) is still in its infancy with very few reports on markers associated with blast resistance. The advancement of Next Generation Sequencing (NGS) technologies offers rapid, high-throughput and cost effective solutions in genomics. In the present investigation an attempt has been made to evaluate GPU 28 (blast resistant) × KM 252 (blast susceptible) derived F_5 mapping population for reaction to blast disease across Bengaluru, Coimbatore and Vizianagaram locations during *rainy* seasons of 2013 and 2014. Along with 16 (8+8) extremely resistant and susceptible RILs, 8 known (4 resistant + 4 susceptible) accessions including GPU 28 and KM 252 were sequenced with Illumina HiSeq paltform. The high quality sequence reads were assembled and annotated. The total genome size of 24 genotypes was in the range of 403 to 2042 Mb. A total of 352631 SNPs from 45618 tags were identified and among them 144 SNPs were found common between resistant and susceptible groups, 56 were synonymous and 88 were non-synonymous SNPs. A total of 30 selected reads were subjected to functional annotation in a protein database, 26 reads were with Blastx having hits to the Jasmonate inducible protein, Mitogen-activated protein kinase 11, Luc7-like protein 3-like, Serine/threonine protein kinase 18, proline rich protein and NBS type disease resistance protein. Further validation of these SNPs in population and or in the germplasm pool segregating for blast reaction will help fast tracking blast disease resistance breeding for finger millet.

Genetic Variability among Peanut Genotypes as Revealed by RAD-Seq and Whole Genome Re-Sequencing Data

Shirasawa K.[2], **Sukruth M.**[1*], **Hake A.A.**[1], **Yadawad A.**[1], **Patil Malagouda**[1], **Gowda M.V.C.**[3] and **Bhat R.S.**[1]
[1]Department of Biotechnology, University of Agricultural Sciences, Dharwad – 580 005, India; [2]Department of Frontier Research, Kazusa DNA Research Institute, Chiba 292-0818, Japan; [3]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad 580 005, India
[*]Email: sukruth.crp@gmail.com

A set of ten genotypes of peanut differing for late leaf spot (LLS) resistance was subjected for restriction site associated DNA sequencing (RAD-Seq) to study the extent and nature of variability. The genotypes showed considerable variability. Single nucleotide polymorphisms (SNPs) were very common across the genomes. These SNPs were found in isolations or in clusters over a large genomic region. The latter feature was also very frequent, and such a region appeared as a fragment resulting from introgression. No clear-cut association could be observed between the DNA variation and resistance to LLS among these genotypes. Selected four genotypes were also subjected for whole genome re-sequencing. The data revealed SNP hotspots within the genome. The SNP data and other marker data are being analyzed to identify their association with disease resistance using a recombinant inbred line (RIL)-population and two mutant-populations.

Introgression of Wide Compatibility Gene (S5n) into IR 58025B, an Elite Maintainer Line of Rice through Marker-Assisted Backcross Breeding

Rahul Priyadarshi[1*, 2, 3], **Hari Prasad A.S.**[1], **Akhilesh Kumar Singh**[2,4], **Durga Khandekar**[2,5], **K. Ulaganathan**[3], **Vinay Shenoy**[2,6]

[1]ICAR-Indian Institute of Rice Research, Hyderabad, India; [2] Barwale Foundation Research Centre, Hyderabad, India; [3]Osmania University, Hyderabad, India; [4]Rasi Seeds Private Limited, Hyderabad, India; [5]AgriGenome Labs Private Limited, Hyderabad, India; [6]Kaveri Seed Company Limited, Secunderabad, India.

[*]Email: rhl.priyadarshi@gmail.com

For exploitation of *indica-japonica* heterosis in hybrid rice breeding programmes and to overcome sterility problem, it is essential to develop parental lines having wide compatibility genes. In this study, marker-assisted backcross breeding was employed to introgress wide compatibility (*WC*) gene into the genetic background of IR 58025B, a maintainer line of wild-abortive cytoplasmic male sterile (WA-CMS) IR 58025A. IR 58025B with long-grain type was used as recurrent parent and Dular possessing (*S5*^{n} allele) was used as donor parent. PCR-based molecular marker S5-InDel tightly linked to *WC* was used for selection of positive wide compatibility trait in selected lines. Further a closely linked marker *nksbadh2* was used for identification of backcross – derived plants devoid of aroma. Arduous phenotypic followed by background selection strategy was adopted which led to the fast recovery of recurrent parent genome (RPG). Ninety-one parental polymorphic microsatellite markers covering across the genome were utilized for background selection. At BC_3F_5, ten promising, stable, backcross-derived lines possessing wide compatibility gene, high yield, long-slender grain type and possessing the range of recurrent parent genome recovery 94.51% to 98.90, have been identified; their maintainer behavior was validated through test crosses with IR 58025A. These improved lines are being converted to CMS lines through marker-assisted backcross breeding. The introgressed wide compatibility could significantly improve group mean pollen percentage with *indica*, *japonica* and *tropical japonica* testers by 2.50%, 13.93% and 36.99% respectively; and for spikelet fertility by 2.22%, 77.11% and 15.38% respectively.

Insilico Prediction and Characterization of MicroRNAs from Large Milkweed Bug (Hemiptera: Lygaeidae) Genome

Ellango R.[1*], **Asokan R.**[1], **Ramamurthy V.V.**[2]
[1] Division of Biotechnology, ICAR-Indian Institute of Horticultural Research, Bangalore, India; [2] Division of Entomology, ICAR-Indian Agricultural Research Institute, New Delhi, India.
[*]Email: ellango85@gmail.com

For studies on functional genomics, small RNAs, especially microRNAs (miRNAs), have emerged as a hot topic due to their importance in cellular and developmental processes. The priority is for finding conserved and less conserved miRNA from model insect species and identifying their relationships with closely and distinctly related species. Identification of insect miRNAs largely depends on the availability of genomic sequences in the public domain. The large milkweed bug, *Oncopeltus fasciatus* (Dallas) is a hemimetabolous insect which has become a model hemipteran system for various molecular studies. In Total 96 candidate mature miRNAs in 66 miRNA genes identified from *O. fasciatus* genome using a blast search with the previously reported animal miRNAs. The secondary structure of predicted miRNA sequences was determined online using "mfold" web server and verified by calculating the minimal free energy index (MFEI). Six miRNAs viz., let-7e, miR-133c, miR-219b, mir-466d, mir-669f, and mir-669l are reported for the first time in Insecta. Comparison of *O. fasciatus* mir-2 and mir-71 family clusters to those of diverse insect species showed that they are highly conserved among hemipterans. The phylogenetic analysis of miRNAs revealed the evolutionary relationship of conserved miRNAs of *O. fasciatus* with other insect species. Using a classical rule-based algorithm method, the possible targets of the new miRNAs were predicted. This study not only identified the list of miRNAs in *O. fasciatus* but also provides vital information about evolutionarily conserved regulators among hemipterans and for pest management through transgenic approaches.

Some Studies on Estimation of Missing Values in Repeated Measurement Designs - Bayesian Approach

R. Ajantha[1*] and **N.Ch. Bhatra Charyulu**[1]
[1]Department of Statistics, University College of Science, Osmania University, Hyderabad, Telangana-500007, India
*Email: ajanta.rudra@gmail.com

In any experimental design, if an observation is missing, the data is incomplete to carry out the analysis as per the original plan of the experiment, due to the affect of orthogonality in the data. Several authors made attempts in estimating the missing observations in least square and maximum likelihood approach in both iterative and non-iterative approaches.

In this paper, an attempt is made to estimate the missing responses in Bayesian approach if the experiment is conducted under the principles of randomized block design with repetition of measurements. The method is illustrated in case of single, two and m- missing observations using Bayesian approach by developing a code in R and WinBUG softwares.

Multiple Outlier Detection in ARIMA Model: An Application to Agricultural Data

Gopal Saha[1*], **Ranjit Kumar Paul**[1] and **L. M. Bhar**[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012, India
[*]Email: sahagopal390@gmail.com

The proper analysis and accurate interpretation of time series data related to agriculture like yield, price or export data is an important concern among both farming community and policy makers. Presence of outliers is most common in these data series. The conclusions drawn through analyzing the data series contaminated with outliers may be erroneous. It is, therefore, important to identify the time points where outliers are present and then remove the effect of the outliers from the corresponding residual series. In this study, an iterative method based on the procedure proposed by Chang and Tiao (1983) along with use of robust estimate of error variance is presented. The power of this iterative procedure in detecting outliers is also investigated. The methodology is illustrated using monthly wholesale price index data of several agricultural commodities like onion, tomato etc. The result of the study clearly indicates that this outlier detection technique using the robust estimate of error variance can successfully detect all the outliers present in the data series. Using the procedures a final residual series can be obtained which is adjusted after removing the effects of outliers and it can be used for further analysis.

Influence of Area and Yield on the Production of Maize in Chhattisgarh Plain

Ghananand Sahu[1*] and **Dr. M.L. Lakhera**[1]
[1]Department of Agricultural Statistics and Social science (L.), Indira Gandhi Krishi Vishwavidyalaya, Raipur 492012, Chhattisgarh, INDIA
[*]Email: ghananand52@gmail.com

Cropped area is decreasing day to day because of increasing population and civilization. So it is necessary to increase production through increase in productivity of crops. The present study has been undertaken to study the trends, prediction model of area, production and productivity of cereal crop maize (*Zea maize* L.) in plain zone of Chhattisgarh in India. The study has been conducted based on secondary data. Predictive models for the maize crop (*Zea maize* L.) of Chhattisgarh plain and its constituent districts have been made. Models have been fitted for the area, productivity and production of the crop separately for above region. Based on these models prediction of area productivity and production of maize have been made year wise between 1998-99 to 2013-14. The partial compound growth rates of the area, production and productivity of the crop have been also estimated and discussed. It was observed that Chhattisgarh plain had registered statistically significant increasing periodic partial compound growth rate in productivity (15.770 percent) and production (18.212 percent) at 5 percent level. Periodic effect of five years as well as annual effects was found to be working in most of the districts/region based on a postulated and estimated production function of area and productivity. It was found that the major influencing factor on the production of maize was its area and productivity for different districts. This influence of area and productivity was around 7 and 22 percent for the Chhattisgarh plain respectively.

Diversity of Iron Homeostasis Genes in Rice

Abin Sebastian[1*], **Ashwini Nangia**[1], **M.N.V. Prasad**[2]
[1]School of Chemistry, University of Hyderabad, Prof.CR. Rao Road, Central University PO, Hyderabad, 500046, India; [2]Department of Plant sciences, University of Hyderabad, Prof. CR. Rao Road, Central University PO, Hyderabad, 500046, India
*Email: abin@uohyd.ac.in

Iron is essential for plant growth. This element have critical role in redox reactions in the cell. But more accumulation of Fe in the plant leads oxidative stress. Hence control of Fe homeostasis is critical for plant growth. Gene families responsible for Fe homeostasis in rice are Yellow Stripe (YS), Fe3+-Chelate Reductase Oxidase (FRO), Zinc Regulated Transporter (ZIP), Natural Resistance-Associated Macrophage Protein (NRAMP), and Ferritin (Fer). These gene families are implicated in uptake, intracellular targeting and storage of Fe. Homologous sequences studies revealed occurrence of 18 genes that belong to YSL family in rice genome. OsFRO1 and OsFRO2 are the main FRO genes in rice genome which have central role in Fe uptake via chelate reductase activity. Studies have shown that OsFRO2 have more sequence similarity with non graminaceous plants. ZIP proteins are involved in the uptake of Fe2+ ions, and the studies revealed existence of 13 genes belong to ZIP family in rice genome. NRAMP are critical for control of uptake of divalent cations, and the rice genome pose 8 genes belong to NRAMP family. Ferritin play critical role in buffering of cellular Fe by acting as a Fe storage protein. Rice genome has two functional genes that code for ferritin. It must also notice that metalloprotein take part in metabolic pathways such as photosynthesis and respiration require Fe and incorporation of Fe into these proteins also help in buffering of intracellular Fe. It is conclded that Fe homeostasis in rice is combined effect of many genes and the present study aim to disclose diversity of these genes especially with respect to expressed sequence tag.

Optimization of Bruchid Resistance Phenotyping Method in Vigna radiata X V. Umbellata F9 RILs Mapping Population

Devina Seram[1*], **Senthil Natesan**[2], **Pandiyan Muthaiyan**[3] and **Kennedy John Samuel**[1]
[1]Department of Agricultural Entomology, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu; [2]Department of Biotechnology, Agricultural College & Research Institute, Madurai, Tamil Nadu; [3]Department of Plant Breeding and Genetics, Agricultural Research Station, Virinjipuram, Tamil Nadu, India
[*]Email: devnah@gmail.com

Development and standardization of screening methods is the key to an effective resistance breeding program. Screening techniques for host plant resistance studies vary with crops and insect pests. In case of storage insects like bruchids (e.g. *Callosobruchus* spp.), resistance is of rare natural occurrence in cultivated pulse species (e.g. *Vigna radiata*) for direct exploitation, yet resistance sources have been identified in wild relatives (*V. radiata* var. *sublobata*) and (or) other related species (e.g. *V. umbellata*). In conventional breeding, inter- or intra-specific hybridization is being carried out for crop improvement (e.g. insect resistance) with the intention that desirable genes can be transferred from wild to cultivated species through generation advancement. Subsequently, such mapping populations are required to be evaluated for identification of insect resistance source(s), marker(s) linked with the resistance and QTL mapping essential for development of markers which can be employed in molecular breeding approaches such as marker assisted selection. Generally, identification of bruchid resistance source requires artificial infestation on seeds of each sample with replications and monitoring different insect biological parameters which is a tedious process and also resulting in favourable amount of seed wastage. In lieu of this method, an “innovative model” was especially designed and developed for conducting “free-choice” insect resistance screening. To maximize the identification and measurement of resistance, the use of both free-choice and no-choice methods is suggested to provide reliable results. "Free-choice” test is more practical when large numbers of samples are to be tested and can be conveniently used to eliminate obviously susceptible lines or varieties. For this study, seeds of 187 F_9 recombinant-inbred lines (RILs) of a highly bruchid susceptible mung bean [*var*. VRM (Gg)1] and a completely resistant rice bean (TNAU Red) were used for bruchid resistance phenotyping, taking parental seeds as check. Ninety samples were assessed at once. Lines showing more than 50 per cent seed damage were considered susceptible and eliminated in first round of screening by means of this model. For further resistance verification, less than half (47.06%) of the total RILs with ≤ 50 per cent damage were subjected to the standard “no-choice” test thus optimizing seed usage. The lines were classified into five categories as completely resistant (0%), resistant (1-9%), moderately susceptible (10-69%), highly susceptible (70-99%) and completely susceptible (100%) based on seed damage which provided feasible results. Combined phenotypic data from both the tests can be employed for QTL analysis and mapping studies. It is expected that this “innovative model” design (or concept) will be exploited in future for allied studies in different insect pests since QTL mapping for insect resistance in crops is meagre, yet to be investigated and documented hitherto.

Meta-analysis of Drought-Tolerant Rice Genotypes: A Network-Based Approach

Sanchari Sircar[1*] and **Nita Parekh**[1]
[1]CCNSB, IIIT Hyderabad, Telangana, India
[*]Email: sanchari.sircar@research.iiit.ac.in

Adverse environmental conditions such as drought pose a severe threat to the yield of rice, an important food crop for over half the world's population. Apart from conventional breeding strategies to develop drought-tolerant crops, use of innovative computational approaches may provide insights into the underlying molecular mechanism of stress response. With the advancement in high-throughput technologies, a large number of microarray studies are being performed to monitor the global transcriptomic changes in plants such as Arabidopsis, rice and maize. Meta-analysis of related microarray studies can be a powerful way to validate the hypotheses and identify novel underlying mechanisms.

In this study we present meta-analysis of seven drought-tolerant rice genotypes under drought stress using network-based approach. The objective of the study is to identify gene signatures and processes that are a hallmark of drought tolerant species. Co-expression networks enable large-scale interpretation of gene-pair associations and the co-expressed gene modules may represent coordinated biological processes. A weighted co-expression network is constructed using WGCNA (an R-based package) across 57 samples and ~14,270 genes which are clustered into co-expressed modules. Differentially expressed genes are mapped to identify drought adaptive modules. It is observed that modules associated with regulation of cellular processes, transcription, amino acid metabolism, transport and localization respond most significantly to drought stress. We used network-based topological measures, degree and eigengene-based centrality, to identify and prioritize potential candidate genes in the drought-adaptive modules. Using these candidate genes, a drought-specific protein-protein interaction network is generated to identify tightly regulated processes in which proteins may physically interact to bring about stress-induced signalling and metabolic cascades. Analysis of this network revealed crosstalk between ABA signalling (via bZIP23 and PP2Cs) and biosynthesis of secondary metabolites associated with sugars and amino acids. Moreover, clusters associated with molecular chaperons (Dnak and Dnaj proteins) and OsHSFs known to be induced during heat and oxidative stress are identified. An interesting observation made is the interaction between cell wall associated genes (OsUGE1, MUR4 and GHMP kinases) and kinases involved in proline synthesis which hints at increased tensile strength of cell walls, an important drought-tolerant trait. Clusters associated with ubiquitin-mediated post-translational modifications and protein degradation are identified which can be associated with a myriad of processes such as degradation of misfolded proteins due to stress, trafficking, signal transduction and cell division. Metabolic processes such as enzymes associated with TCA cycle, starch degradation and mitochondrial ATP synthesis are also been identified.

Combining sequence-based methods and conserved co-expressed neighbours in Arabidopsis, we show how network-based approach can help in functional annotation of stress-induced uncharacterized genes. This can help in improving the functional annotations of genes and identify novel candidates associated with abiotic stress.

Mating Designs for Rapid Initial Screening of Genetic Stocks

Prity Sundaram[1, 2, *], **Ashwini Karadi**[1], **Sobhan B. Sajja**[1] and **S. P. Singh**[2]
[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India;
[2]Bihar Agricultural University, Sabour, Bhagalpur, India
[*]Email: pritys28@gmail.com

Classical Mendelian genetics explained the inheritance of qualitative traits. However, to study the inheritance of quantitative traits biometrical approaches were developed, subsequently mating designs were developed to estimate the genetic composition of genotypes. For initial screening as well as for intensive study of genetic stocks, selection of efficient mating designs is important for successful conventional Plant Breeding programme. The initial screening of a large collection of genotypes is necessary because most of the genotypes needed to be analyzed rapidly to shed the load of large proportion of undesirable stocks as early as possible and for shortlisting of some promising stocks for further intensive study. In all mating designs, randomly selected individuals are crossed in different combinations. The progenies thus produced could be half-sibs or full-sibs. The bi-parental (paired crossing) design is the simplest one to execute. But the simplicity of this design is counterbalanced by its inability to yield sufficient information to estimate all parameters required by the model because the progenies from this design are either full-sibs or unrelated. In topcross design progenies comprise half-sibs and is probably the simplest model for preliminary screening of genotypes as it involves the lowest crossing load besides simple statistical analysis. A single tester may not offer wide genetic background for testing the inbreds and the number of crosses become large if the test inbreds are more are the two shortfalls of the topcross design. Polycross design is the natural substitute of topcross with same genetic basis with wider genetic background of pollen sources. Polycross and topcross designs provide an idea about *gca* in terms of breeding value of parents. However, more precise estimates of *gca*, and *sca*, and other parameters are provided by line × tester (L×T) design which is basically an extension of topcross involving more number of crosses than topcross. The major drawback of L× T design is that the significance of *gca* and *sca* variances obtained cannot be tested. In partial diallel design (circulent design), slightly more number of crosses per parent can tested as compared to L× T design. Furthermore, this design provides precise estimates of *gca* and *sca* variances, and *gca* effects. Increase in standard error due to less number of crosses per parent and involvement of inversion matrix which is tedious are the two major drawbacks of this design. These problems are mitigated to a good extent by adopting triangular designs (TD I and TD II) or fractional diallel design. Among all the models triangular designs are the most useful ones because of their simplicity in statistical treatment of data; much precise estimates of *gca* effects and variances, and large number of genotypes can be screened rapidly. A comparative evaluation of mating designs for rapid initial screening of genotypes, in terms of coverage of population and amount of information, the mating designs can be ordered as: trangular design > partial diallel > L× T > topcross > polycross > paired crossing design.

Genetic Variability - A Basis for Line and Hybrid Development

Anshuman Tiwari[1*], **Jayesh Warade**[1], **Jayendra Padiya**[1], **Vijay Dalvi**[1], **Bharat Char**[1], **Usha Zehr**[1] and **Venugopal Mikkilineni**[1]
[1]Mahyco Research center, Maharashtra Hybrid Seeds Company Private Limited, Dawalwadi, Jalna-431203
[*]Email: anshuman.tiwari@mahyco.com

Selection of parental lines is the most critical step in any hybrid breeding program. In hybrid wheat program, obtaining a high degree of heterosis has been a major challenge. Achieving the desirable level of heterosis depends upon the combining ability of parental lines which depends on various factors. One key factor that may contribute to heterosis is the genetic variability between the parental lines. By utilizing genetic variability as one of the basis for selecting the parental lines, we hope to achieve the desired level of heterosis as well as develop newer recombinant lines. Wheat is typically grown in six agro-climatic zones, but the major are the Central and Peninsular Zones and North West and North East Plain Zones. In the presented study, a total of 188 wheat lines, 94 lines from each Central and North West plain zone were selected. These 188 wheat lines were genotyped with 576 wheat SNP markers. Genetic distance was calculated using Jaccard Coefficient and clustering of genotypes was done based on the Ward algorithm using the R- language. Our analysis showed a diversity range between 0 to 44% among all wheat line across both the zones. Genotypes from two different zones broadly grouped into two separate clusters which agree with the zonal classification of wheat lines into Central Zone (CZ) and North West plain zone (NWPZ). To maximize the potential for heterosis and to develop new recombinant lines two criteria were used for selecting the crosses within each zone, the genetic distance between the parental lines and their position on the cluster of the dendrogram. Further validation of these crosses and for selecting the potential recombinant lines is being done through field evaluations.

Genetic Architecture of Kernel- Iron and Kernel-Zinc Content in Tropical Maize (Zea Mays L.) Germplasm as Revealed by Genome-Wide Association Analysis.

Hindu Vemuri[1, 4*], **Willy B**[2], **Palacios-Rojas N**[3], **Sudha Nair**[1], **R Usha**[4] and **R Babu**[5]
[1]International Maize and Wheat Improvement Center (CIMMYT), c/o ICRISAT, Patancheru, India-502234; [2]Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Indonesia-16680; [3]International Maize and Wheat Improvement Center (CIMMYT), Mexico-56130; [4]Department of Biotechnology, Sri Padmavati Mahila Visvavidyalayam, Tirupati, India- 517502; [5]Multi-Crop Research Center (MCRC), DuPont Pioneer, Hyderabad, India-500078
*Email: v.hindu@cgiar.org

Micronutrient malnutrition, due to low zinc (Zn) and iron (Fe) in diets, is widely reported across globe. Bio-fortification through genetic improvement contributes to solve global nutrient malnutrition.Maize is grown worldwide, as a staple crop as well as valuable commodity and has the potential to be a useful tool for addressing dietary Fe and Zn deficiency among the undernourished poor. Herein, we evaluated a diverse maize association mapping panel of 923 inbred lines across three environments using high density SNPs obtained through genotyping-by-sequencing (GBS) and implemented GWAS to identify genomic regions and potential candidate genes influencing kernel-Fe and kernel-Zn concentrations. GWAS was performed through GLM, MLM & MLMM approaches. Principal components and kinship matrix were used in mixed models for minimizing population structure and cryptic relatedness effects. The physical intervals identified by the significant haplotype blocks and individual SNPs contained several candidate genes such as myb156 and bZIP (on Chr.1), platz7 and 15(on Chr. 4&10), sal1 (on Chr. 9) and Zn/Fe ion binding transporters (on chr.10). Highest prediction accuracies for kernel-Fe (54%, 64% and 83%) and kernel-Zn (48%, 56% and 90%) were achieved when few GWAS-identified genomic regions used as predictors in the genomic selection strategies .A set of donor lines from the tropical maize germplasm that is consistently high in kernel-Fe and kernel-Zn has been identified and high throughput SNP markers are being developed to facilitate marker-assisted introgression of these genomic regions.

Linear Mixed Modelling for Crop Yield Prediction in Haryana (India): An Application

Suman Verma[1*] **and Urmil Verma**[1]
[1]Department of Maths, Statistics and Physics, CCS Haryana Agricultural University, Hisar – 125 004, Haryana, India
*Email: sumanjai27@gmail.com

The aim of this paper is to show the usefulness of mixed modelling approach for crop yield prediction. The study deals in obtaining pre-harvest crop yield forecasts of major sugarcane growing districts in Haryana (India). The common approach to crop yield forecast is linear regression with constant coefficients over time. This may be restrictive and of limited prediction power since it does not account for the year-to-year dependence in the yield variable. However, a mixed model procedure provides a flexible way to fit multi-level models for crop yield prediction. The linear mixed effects models with random time effects at district, zone and state level were fitted for sugarcane yield estimation in Haryana. The percent deviation(s) of district-level crop yield forecasts from the real time yield(s) data show a preference for using linear mixed models.

Comparative Analysis of Cytotoxic T Lymphocytes (CD8+) in Different Breeds of Chicken

S.P. Yadav[1*], **TR Kannaki**[1], **MR Reddy**[1], **TK Bhattacharya**[1], **RK Mahapatra**[1], **SK Sarkar**[2] and **RN Chatterjee**[1]
[1]ICAR-Directorate of Poultry Research, Hyderabad, India-500030; [2]ICAR- Indian Agricultural Statistics Research Institute, New Delhi, India.
[*]Email: yadav.satyapal@gmail.com

Disease resistance is a complex trait and the mortality/disease susceptible in poultry is a great concern for economic loss in poultry industry. Moreover, the heritability of immune traits is also very low in poultry. The differences in genetic response to different diseases have been well established in chickens and selection of chickens which shows better genetic resistance for diseases is an important way to better disease protection. Cell-mediated immunity plays an essential role against pathogens particularly against viral infection via presentation of peptides bound to MHC class I molecules to CD8+ cytotoxic T lymphocytes. To observe the Immune-competence traits in different breeds of chicken (viz. Ghagus, Dahlem Red and Nicoabri) total count of T lymphocyte bearing CD8+ molecule in the peripheral blood of chicken was carried out by flowcytometry assay. In this study the antibody used was mouse anti-chicken CD8α-PE. The study revealed a significant difference ($P<0.05$) between the three breeds. The T cell number bearing CD8+ was found to be highest in Ghagus (2132.34±69.526) followed by Nicobari (2023.53±75.14) and least in Dahlem Red (1649.49± 64.54). Sex-wise results showed that males (1989.41±65.47) were having high Tc cell compare to females (1880.83±47.14). In Dahlem Red, the number of cytotoxic T lymphocytes ranges between 1522.60 to 1776.38 while in Ghagus and Nicobari it ranges between 1995.65 to 2269.03 and 1875.51 to 2171.25, respectively. Sex-wise, the value ranges for female and male were 1788.16 to 1973.51 and 1860.68 to 2118.13, respectively. Therefore, it is observed that the native germplasm viz. Ghagus and Nicobari were having higher CD8+ cytotoxic T lymphocytes compared to Dahlem Red.